# In silico investigation of xenobiotic interactions with lipid bilayers and ABC membrane transporters, the case of ABCC4/MRP4

Benjamin Chantemargue

**Ph.D thesis**

Palacký University Olomouc

Université de Limoges

# Université de Limoges – Palacký University Olomouc

## ED 615 - Sciences Biologiques et Santé (SBS)

### Inserm UMR1248 - IPPRITT

Submitted for the degrees of

Doctor of the University of Limoges

Discipline / Speciality: Theoretical Chemistry

Doctor of the Palacký University

Discipline / Speciality: Physical Chemistry

Presented and defended by

**Benjamin Chantemargue**

On 18 décembre 2018

# *In silico* investigation of xenobiotic interactions with lipid bilayers and ABC membrane transporters, the case of ABCC4/MRP4

Thesis directed by Patrick Trouillas and Michal Otyepka

JURY :

Referees

Ms. **Lucie Delemotte**, Ph.D. Associate Prof., PI, KTH Royal Institute of Technology

Ms. **Catherine Vénien-Bryan**, Prof., IMPMC (UMR7590) CNRS/Sorbonne Universités UPMC/IRD/MNHN

Examiners

Ms. **Kateřina Valentová**, Ph.D. PI, Institute of Microbiology, Czech Academy of Sciences

M. **Pavel Banáš**, Ph.D. Associate Prof., Carbon nanostructures, biomacromolecules and simulations, RCPTM, Palacký University

M. **Gabin Fabre**, Ph.D. Associate Prof., PEIREINE, Faculté de Pharmacie, Université de Limoges

M. **Michal Otyepka**, Ph.D. Prof., PI, Carbon nanostructures, biomacromolecules and simulations, RCPTM, Palacký University

M. **Patrick Trouillas**, Ph.D. Prof., INSERM U1248, Faculté de Pharmacie, Université de Limoges

**Ph.D thesis**

Palacký University Olomouc

Université de Limoges

# Université de Limoges – Palacký University Olomouc

## ED 615 - Sciences Biologiques et Santé (SBS)

### Inserm UMR1248 – IPPRITT

Submitted for the degrees of

Doctor of the University of Limoges

Discipline / Speciality: Theoretical Chemistry

Doctor of the Palacký University

Discipline / Speciality: Physical Chemistry

Presented and defended by

**Benjamin Chantemargue**

On 18 décembre 2018

# *In silico* investigation of xenobiotic interactions with lipid bilayers and ABC membrane transporters, the case of ABCC4/MRP4

Thesis directed by Patrick Trouillas and Michal Otyepka

JURY :

Referees

Ms. **Lucie Delemotte**, Ph.D. Associate Prof., PI, KTH Royal Institute of Technology

Ms. **Catherine Vénien-Bryan**, Prof., IMPMC (UMR7590) CNRS/Sorbonne Universités UPMC/IRD/MNHN

Examiners

Ms. **Kateřina Valentová**, Ph.D. PI, Institute of Microbiology, Czech Academy of Sciences

M. **Pavel Banáš**, Ph.D. Associate Prof., Carbon nanostructures, biomacromolecules and simulations, RCPTM, Palacký University

M. **Gabin Fabre**, Ph.D. Associate Prof., PEIREINE, Faculté de Pharmacie, Université de Limoges

M. **Michal Otyepka**, Ph.D. Prof., PI, Carbon nanostructures, biomacromolecules and simulations, RCPTM, Palacký University

M. **Patrick Trouillas**, Ph.D. Prof., INSERM U1248, Faculté de Pharmacie, Université de Limoges

*À mes grands-mères*

*"The best currency, the most valuable of all is gratitude. When you're dead, you're dead, but you're not quite so dead if you contribute something"*

John Dunsworth

## Acknowledgements

My first and biggest thanks are addressed to Patrick Trouillas. Patrick, it is thanks to you that I was able to enter in research and in this field. I feel beyond lucky that I met you to have you as my supervisor, advisor and mentor. You allowed me to push my limits and improve myself not only professionally, but also personally, as a person. You shared with me all your kindness, your patience, and your passion. You taught me everything you know, you showed me your academic and scientific skills. I will never be thankful enough to you for everything you taught me and for your generosity. I am also grateful for your support and for sacrifices you made to help me. I like to think that sometimes you have to encounter some people to help you take the right path and push you to extract the best of you. I am grateful, you are one of these persons.

I would like to express my sincere gratitude to Michal Otyepka for your co-supervising this thesis. I truly appreciated your warm welcome in your lab, your city and your country. I really want to thank you for everything you taught me also, the stimulating and fruitful discussions that allow me to always push further.

My next thanks go to all members of the jury. I am truly honored that you accepted to be part of this jury. I truly appreciate your kindness, your expertise and your insightful discussions.

I am very grateful to Lucie Delemotte. Thank you for accepting to review this work. It is an honor to have you as a member of my jury.

Catherine Vénien-Bryan, it is an honor to have you as a jury member. Your expertise in the experimental procedures have always brought interesting discussions.

Kateřina Valentová, I am really grateful to have you as a member of my jury.

I would like to thank Pavel Banáš and Gabin Fabre for evaluating this work. Your expertise in MD simulations, and biased methods have always brought interesting discussions. Thank you for your constant help and kindness during my studies.

My next thanks are addressed to Florent Di Meo. Thank you very much Florent. Sadly, there are no words that are strong enough to express my gratitude towards you. You allowed me to structure more myself and you shared with me your experience and advices. You have made me grown, and I am truly happy that I met you.

My next thanks are going to all the Ph.D students in Olomouc, Martin, Martin, Vašek, Veronika, Petr, Petra, and Petra, Marie, Sylva and all others. I really enjoyed every one of my stay, it was truly a pleasure. I will always keep all the memories with all of you. These encounters are the one of a kind, I discovered a country, a city, and friends. You really are the best.

I would also like to thank all my colleagues from the lab, Gabin again with all your kindness, Bibi, Seb, Bastien, Danko, Hassan, Quentin, CC, Tom, but also my buddy BrOlivier ! with who I have a wonderful time, especially in the Mexx. I thank also everybody from the laboratory as well. Thank to all of you.

I would also like to thank all the French PhD students I have encountered during this Ph.D, Martial, Florian, Sylvain, Hugo, Robin, Alexis, JM and plenty others (I don't forget you). I made wonderful encounters and had a really great time with all of you.

I would also like to thank all my family, Papa, Maman, Tonton and Tata that did not minimize their efforts supporting me during this PhD, with all my possible moods. From my first days in

life till now, I am not grateful enough for everything you have done for me. I would also like to thank Papi, that always supported me in everything I have done and during my studies.

My very sincere thanks to my friends Joe, Max, Fab for all your support during this PhD, you helped and supported me through this Ph.D with everything you had, thank you very much again it really counted and will always count for me.

A tremendous thank to the love of my life Marie, thank you for being immensely kind and generous, supporting me. I am a lucky guy to have you.

Finally, thank to everybody that I was lucky to encountered. I am the today me thank to you.

## Copyright



## Table of Contents

## Table of Figures

## List of Tables

## List of Schemes

## Introduction

Xenobiotics are molecules that are not naturally present or produced in a living organism. The term xenobiotics includes numerous molecules of various origins, namely i) nutrients *e.g.*, vitamins or antioxidants present *e.g.*, in vegetables, fruit, oils, chocolate; ii) toxins, *e.g.*, pollutants, pesticides; or iii) drugs of various therapeutic purposes. These molecules circulate in our body. Their specific location, if any, is highly xenobiotic-dependent. Many xenobiotics play a (positive or negative) role in the organism at specific targets. To reach their target, xenobiotics must cross several lipid bilayer membranes, which constitute physical barriers of solid organs, skin, tissues, cells, or cell organelles. Membranes are xenobiotic-dependent selective barriers. This selectivity is partially attributed to the different compositions of membranes, which depend on the organs, or the cell organelles, but also depending on the inner or outer leaflets. To give only one typical example, the blood-brain barrier is a crucial selective barrier to prevent some xenobiotics from directly affecting the nervous system. Some xenobiotics cannot cross lipid bilayers; others only partially cross, *i.e.*, accumulating in the bilayers as it is known for some antioxidants. Membrane crossing is possible through different ways, mainly passive permeation or through membrane proteins.

Pharmacokinetics is studying the fate of drugs after their intake. The passage of drugs through cell membranes (*e.g.*, through ABC or SLC membrane transporters) is key in absorption, distribution, metabolism and excretion (ADME). Any dysfunction in these transporters may dramatically impact the patient response. For instance, drugs can accumulate abnormally in cells and induce toxicity. Some of these membrane transporters, *i.e.*, ABC transporters, are involved in multidrug resistance. This is particularly critical in clinical situations requiring multi-drug treatments, *e.g.*, cancer or organ transplantation. Questions then emerged: How do xenobiotics interact with lipid bilayers? What is the transporter role? Can we rationalize the intimate details of these interactions? Does the chemical structure of xenobiotics influence their interaction with membranes or membrane proteins? How does the transport cycle of these membrane proteins work? Can multidrug resistance be prevented or counterbalanced? We have tackled some of these questions in this manuscript.

Experimental studies conducted to rationalize interactions between xenobiotics and membrane or membrane proteins require long and expensive investments to often obtain fragmented information. *In silico* methods such as molecular dynamics (MD) have proven to be powerful tools that open many perspectives to shed light on xenobiotic-membrane interactions. These methods can provide an atomic description of molecular interactions while supporting experimental data. They are also able to predict xenobiotic-protein interactions or protein structure when there is a lack of experimental data.

This manuscript is subdivided into four chapters allowing a progressive description of interactions between xenobiotics, membranes and ABC membrane transporters. Chapter I describes: i) the pharmacokinetics and ADME of xenobiotics (Section I.1); ii) lipid bilayers, their compositions and biophysical properties (Section I.2); and iii) membrane proteins, *i.e.*, metabolizing enzymes, SLC and ABC transporters (Section I.3). Chapter II describes: i) molecular mechanics principles and the notion of force field (Section II.1); ii) MD simulations and underlying equations and parameters (Section II.2); iii) metadynamics principles and the simulation parameters (Section II.3); iv) the construction of a protein model by homology modeling and related analysis (Section II.4); v) molecular docking and principles (Section II.5). Chapter III lists the publication that I co-authored during my PhD studies. After a short

introduction on the importance of xenobiotic-membrane interactions, four publications (two published and two submitted) on drug penetration in lipid bilayers are presented. The first two publications, in Sections III.1 and III.2, are dealing with antioxidants and how lipid peroxidation inhibition is affected by location in membranes. MD simulations are a valuable tool to determine the position and orientation of the antioxidants in membranes. First, we assessed the antioxidant activity of curcumin (Section III.1), a natural polyphenol which exhibits nerve functional recovery and regeneration activities. Section III.2 investigates the antioxidant properties of a novel flavonolignan hybrid formed of ascorbic acid and silybin linked by a dodecanedioate ($C_{12}$) linker. Section III.3 reports on the insertion and positioning of the beta-lapachone (B-lap), a promising anticancer drug, in membranes made of various lipid compositions. Section III.4 investigates the antioxidant properties of α-phenyl-*N*-*tert*-butylnitrone (PBN) derivatives, linked to Trolox and perfluorinated chains. Chapter IV is dedicated to the study of ABC transporters. Section IV.1 investigates the construction and the structural features of the human ABCC4/MRP4 membrane protein. We assessed the influence of cholesterol on the protein and the impact of a mutation known to completely stop the transport activity of this protein. Section IV.2 reports on the docking study of several xenobiotics in the human ABCC4/MRP4 protein that allowed to identify crucial regions and residues involved in substrate binding into the cavity. Section IV.3 is an attempt to investigate the transport cycle of ABC transporters using a well-elucidated crystallographic transporter, namely Sav1866. We used metadynamics simulations to elucidate intermediates characteristic of the xenobiotic transport.

## I.1. ADME: absorption, distribution, metabolism and excretion

Over the past decades, human beings have got into the habit to take many drugs during a lifetime, often to cure disease but not only. When one (orally, topically, intravenously…) takes a drug, it starts a long trip through the organism, which is governed by the science of pharmacology, mainly pharmacokinetics. It is commonly divided into four steps, namely absorption, distribution, metabolism and excretion (ADME)[1]. These four processes can be seen by pairs, based on drugs' concentration. The first pair consists in absorption and distribution, when the drug has the highest concentration, *i.e.*, just after being taken. The second pair is metabolism and excretion, *i.e.*, from its action at its target to its excretion from the organism. Drugs can move randomly by diffusion processes or being engaged in directional transports. Drugs can go either through blood flow, or through the different biological membranes, mediated or not by proteins.

Drug pharmacology is a complex process, in which membrane crossing is a key event. Indeed, to reach its biological target, drugs have to cross several membranes from one to another aqueous compartment (Figure 1). This pharmacological context can be pictured as an organization of boxes separated by membranes which are interconnected to favor drug transport.

The membranes, which can be of several types, constitute barriers which are more or less selective. The gastrointestinal mucosa is an epithelial barrier, which made up of multiple layers and cavities that must be crossed by drugs. Another example is the endothelial cells, also acting as filters for molecules. Another typical example of barriers is the blood-brain barrier, which precludes most of drug crossing to protect the brain against harmful molecules. The same situation also exists between blood and the placenta.

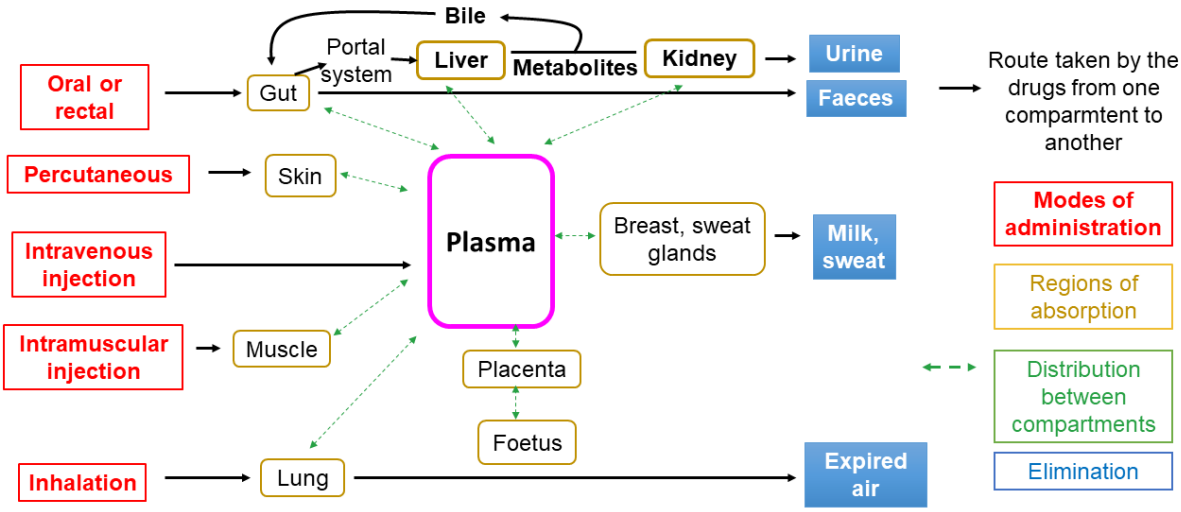



Figure 1: Schematic representation of the routes taken by drugs from the different administration to the absorption, distribution and finally elimination, adapted from Rang & Dale's Pharmacology, 7th edition.

### I.1.1. Absorption

Drug absorption depends on the mode of administration, which can be oral, sublingual, rectal, topical application, inhalation or (intravenous, intramuscular, percutaneous) injection. The mode of administration is chosen according to the drug itself but also to its capacity to be

absorbed and bioavailable. It defines a route through the different organs and compartments in the body (Figure 1). The most common mode is orally (75%)[1], which in principle leads the drug to the small intestine where it is eventually absorbed. Oral absorption mainly depends on the gastrointestinal mobility, but also the splanchnic blood flow, the molecular weight, the drug formulation, and various physical chemical factors. Sublingual administration is used when a rapid response is required, as it allows a direct access to the systemic circulation through the numerous blood vessels under the tongue, a short-circuit to avoid degradation in the stomach, the liver... Rectal administration is used when a local effect is desired, or when another mode of administration is precluded due to the patient (*e.g.*, vomiting). Local effects are expected in topical application, *i.e.*, cutaneous, nasal or eye administration without avoiding a possible systemic effect (*e.g.*, steroid hormones). Inhalation is useful for a quick local response in the ear, nose, and throat (ENT) region. The most direct mode of administration to the systemic circulation is injection. The rate of absorption however depends on the site of injection, so the neighboring blood flow. Injection can be delayed by using a sub-cutaneous implant.

A xenobiotic is bioavailable when its concentration (or the concentration of its metabolites) is sufficient at its biological target, which is often related to a sufficient concentration in the systemic circulation. Whatever the mode of administration, to be absorbed and so bioavailable, xenobiotics will cross several membranes, either i) by passive permeation, or ii) through membrane proteins, *i.e.*, by diffusion through pores (*e.g.*, through aquaporins), or by active transport through ATP-binding cassette (ABC) and solute carrier (SLC) transporters, or iii) by endocytosis.

The capacity for passive permeation is drug-dependent, mainly due to xenobiotic hydro/lipophilicity, but also to other parameters including molecular weight and pKa of xenobiotics. In principle, negatively charged xenobiotics hardly insert and partition into lipid bilayers[2]. However, re-protonation can occur at the surface of the membrane and facilitate insertion and permeation[3]. Permeability coefficients are often correlated to partition coefficients (P), but specific interactions like hydrogen boding should be considered with care.

In active transport, the SLCs mediate passive movement of the xenobiotics down their electrochemical gradient, while the ABC transporters are active (pump motion) using a power stroke from ATP hydrolysis in their catalytic domains (see section I.3.3 for more details). These transporters can influx or efflux a broad range of xenobiotics and they are located everywhere in the organism. These transporters may exhibit numerous polymorphisms, defining families but also related to inter-individual variability.

Endocytosis (or pinocytosis for liquid phase endocytosis) for influx, or its counterpart for efflux, exocytosis, allows a dynamic transport of xenobiotics by wrapping them into lipid bilayers, creating vesicles, and releasing them into the aqueous compartment. This process may transport large polar molecules that are not able to cross lipid bilayers passively.

At therapeutic concentrations in plasma, many drugs exist in bound form. The number of free drugs in bulk water is often very small, the other drugs are bound to plasma proteins. For instance, albumin can bind many drugs, with a competition existing at the binding sites. Drug treatment can increase or decrease drug mobility. For instance, drugs used against migraine facilitate absorption of analgesic by increasing gastrointestinal mobility. Many other factors such as diseases can affect drug absorption.

### I.1.2. Distribution

The second step after absorption is distribution. The distribution of a xenobiotic is mainly based on the distribution of water in the body, the major compartments being the plasma, the interstitial fluid, intracellular fluid, and transcellular fluid. An equilibrium is established between these compartments, which depends on the exchanges through membranes, binding to (circulating or membrane) proteins, and pH.

A special case is the brain blood barrier which is inaccessible to most of drugs. However, the central nervous system possesses zones where this barrier is more permeable (chemoreceptor trigger zone) to some drugs (*e.g.*, doperidone). Xenobiotics do not obviously reach aqueous compartments. Fat compartments (*i.e.*, adipose tissues) also constitute a special case, which can be reached only by lipophilic (apolar) compounds. Moreover, fat compartments are not well vascularised, which explains that drugs slowly reach these regions. However, drugs given chronically can accumulate in fat tissues. Accumulation in membranes is also likely with lipophilic or amphiphilic xenobiotics.

An important factor in distribution is the volume of distribution ($V_d$). $V_d$ is defined as the total amount of a drug at the same concentration observed in plasma. This volume should be measured with caution based on the targeted compartment, to allow an efficient drug activity. $V_d$ may change due to binding to proteins, which differs from one to another compartment. For instance, hydrophilic drugs are mainly confined into the plasma and interstitial fluids and do not enter the brain, whereas lipophilic drugs reach all compartments and may accumulate in fat compartments.

The absorption and distribution can be improved by designed delivery systems. For instance, prodrugs are inactive precursors that are metabolized into active metabolites. Prodrugs can be useful to enhance the bioavailability, or to overcome the instability of a drug.

### I.1.3. Metabolism

During their time in the organism, xenobiotics are highly metabolized. The purpose of metabolism is to prepare xenobiotics, especially lipophilic xenobiotics (more difficult to eliminate), for their excretion. The major site for metabolism is at the liver, but also at the hepatobiliary system and the intestine. Drug metabolism, also called biotransformation of drugs[4], occurs in two phases, which are catalysed by specific enzymes. Phase I is a catabolic process *i.e.*, hydrolysis (bond breaking to break down large molecules), oxidation (addition of an OH group), or reduction, which often leads to more polar metabolites than parent compounds, this is called functionalization. The purpose of this phase is to prepare the xenobiotics for the phase II. It is mainly catalyzed by cytochrome P450 (see section I.3.1). Drugs metabolized *via* phase I reactions have longer half-lives. The metabolites can be more active, neutral or toxic. Phase II corresponds to conjugation reactions, in which glucuronate, acetate, sulfate or methyl moieties are added, to inactivate xenobiotics. Phase II reactions are catalyzed by 5'-diphospho-glucuronosyltransferases (UGTs), N-acetyltransferases (NATs), sulfotransferases (SULTs) and glutathione S-transferases (GSTs)[5], as well as the S-Adenoslyl methionine[6] and the Acetyl-CoA carboxylase (ACC)[7]. In principle, the corresponding metabolites are easily excreted.

### I.1.4. Elimination

The elimination or excretion of drugs or metabolites is generally achieved by the kidney (Figure 1). Most of drugs cross the glomerular filter freely (exception made for the protein-bound

drugs). Many drugs, especially weak acids and weak bases are actively secreted into the renal tubule and thus more rapidly excreted. Lipophilic drugs are passively reabsorbed by diffusion across the tubule and thus they are not efficiently eliminated in the urine. The rate of drug excretion also depends on pH; alkaline urine will favor weak acid excretion and *vice-versa*. The elimination of drugs by the kidneys is measured by the renal clearance, defined as the plasma concentration of xenobiotics eliminated by the kidneys by unit of time. Any dysfunction of the kidneys may lead acute toxicities.

## I.2. Lipid bilayer membranes

Membranes are barriers separating inner and outer compartments of cells, as well as subcellular compartments (*e.g.*, mitochondria, endoplasmic reticulum). Membranes are mainly constituted of lipids arranged in bilayers (about 50% of the membrane for most of cells), in which other constituents are embedded including proteins and carbohydrates. Membrane lipid bilayers are highly dynamics while being highly regulated. There exists a huge diversity in membranes, mainly according to the chemical structures of the constituting lipids and the composition. The latter differs according to species, tissues, cells, organelles. It can even differ from one to another bilayer leaflet and to bilayer subdomains. This variability in composition may dramatically affect physical properties of membranes (fluidity, segregation, membrane curvature, lipid phase), thus the related biological properties (permeation, interaction with other components as proteins)[8,9].

There exist many biophysical techniques to tackle membranes including solid-state NMR (nuclear magnetic resonance), high-resolution spectroscopy, atomic force microscopy (AFM), mechanical properties of Langmuir films. Molecular dynamics has become an attractive tool to rationalize the behavior of a lipid bilayer as well as the interaction with drugs at a fentosecond and nanometer time and space scales. This chapter will detail the major lipid types found in mammal cells as well as composition and diversity to understand the dynamism of lipid bilayers.

### I.2.1. Lipid types

The physical properties of lipid bilayers strongly depend on the chemical structures of lipids. The chemical nature of lipids can be seen as a construction made of Lego® bricks (Figure 2). There exist four main classes of lipids, namely glycerophospholipids, sphingolipids, glycolipids and sterols, mainly cholesterol in mammals. An additional class of biological lipids not constituting lipid bilayers (*i.e.*, glycerolipids) will also be described in this section. The first three classes exhibit a head (polar) group, a glycerol backbone and acyl (fatty) chains, as three different bricks. One or two fatty chains constitute membrane lipids; in the latter case, the two chains should be distinguished (*sn-1* and *sn-2* chains) as being different in most cases (Figure 2).

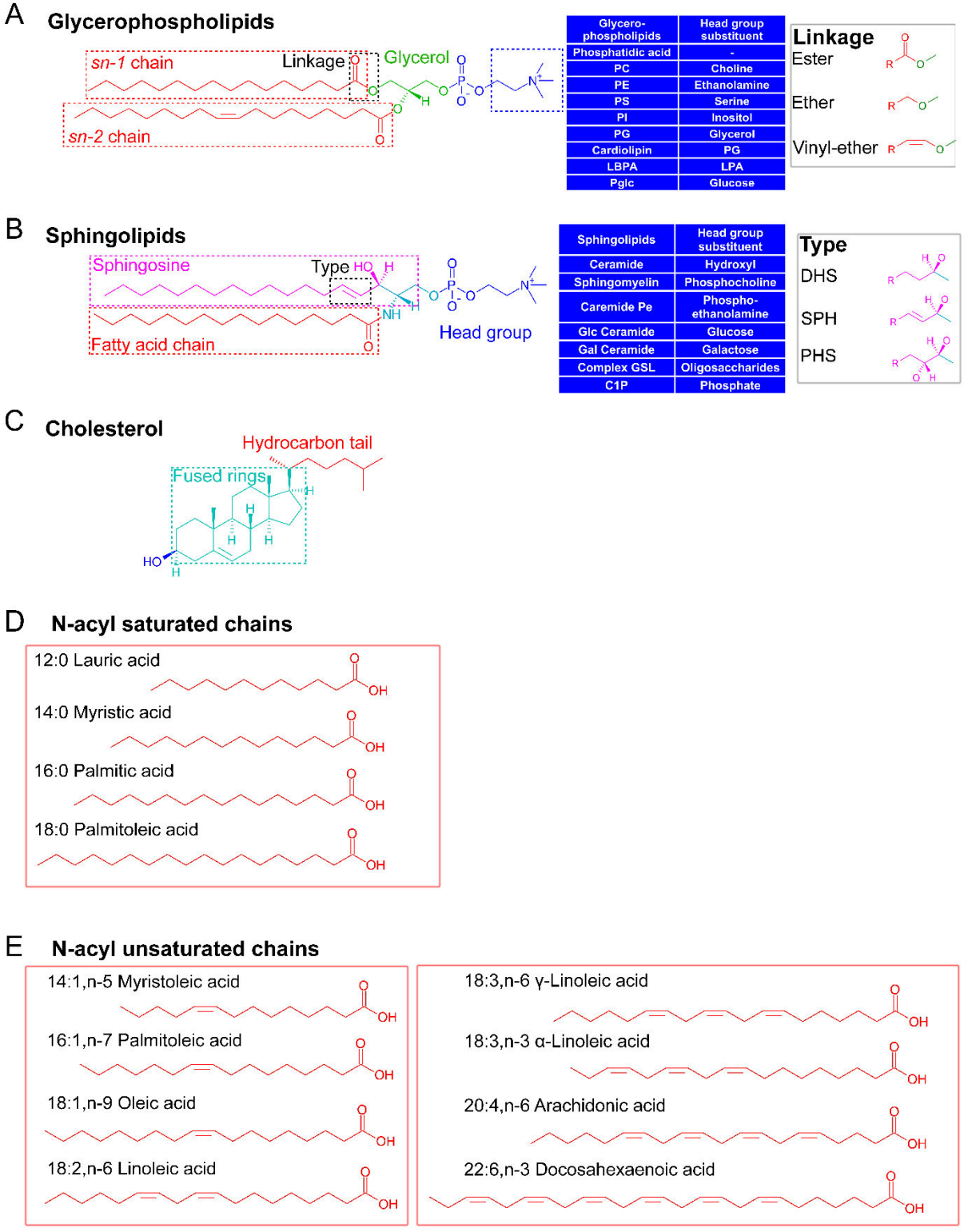



Figure 2: (A) Representation of a characteristic glycophospholipid, the head group substituents and the linkage between glycerol and fatty chain. (B) Representation of a characteristic sphingolipid, the head group substituent and the different types in the sphingoid base backbone. (C) Representation of a typical sterols, *i.e.*, cholesterol. (D), (E) N-acyl chains with various length and unsaturation, and position of the double bond. The key is of the form (XX: Y, n-Z), with XX, Y, and Z the number of carbons, double bonds and position of the first double bond from the omega end, respectively. This key describes also the two fatty acyl chains with the form (XX: Y, n-Z): (XX: Y, n-Z). C1P: ceramide-1-phosphate, Complex GSL:

complex glycosphingolipids, DHS: sphinganine, Gal Ceramide: calatosylceramide, Glc Ceramide: glucosylceramide, LBPA: lysobiphosphatidic acid, LPA: lysophosphatidic acid, PC: phosphatidylcholine, PE: phosphatidylethanolamine, PG: phosphatidylglycerol, PHS: 4-hydroxy-sphinganine, PI: phosphatidylinositol, PS: phosphatidylserine, SPH: sphingosine. Adapted from [10].

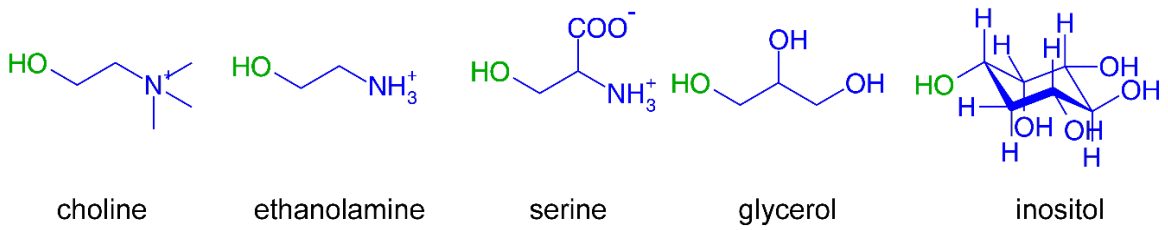



Figure 3: Glycerophospholipids' head groups. The hydroxyl group is bound to the phosphate moiety as indicated in Figure 2.

#### I.2.1.1. Glycerophospholipids

Glycerophospholipids differ from one to another by the length of the N-acyl chains (fatty acids), the number of double bond(s), their position, the type of head group, and the link between the acyl chains and the head group (Figures 2A, D, E). The polar head is a phosphate moiety linked to an alcohol group (Figure 2A). The polar head provides different charges to the phospholipids. It is either zwitterionic (choline in phosphatidylcholine, PC or ethanolamine in phosphatidylethanolamine, PE) or negatively charged (serine in PS – phosphotidylserine - glycerol in PG - phosphatidylglycerol - or inositol in PI - phosphatidylinositol) (Figure 3). The polar head charge and its distribution across the leaflets may induce a surface potential that drive biological processes (Figure 5)[9,11–13]. Fatty acid chains are often made of 16 or 18 carbons (Figures 3D, E). The *sn-1* chain is more likely saturated or mono unsaturated, whereas the *sn-2* chain is often mono- or polyunsaturated. The double bonds in unsaturated fatty acid chains are separated by a saturated methylene group. They are mainly *cis*, but *trans* conformers also exist (synthetized by dehydrogenation). The chains are usually labelled by a notation in which the first number stands for the total number of carbon atoms in the chain, and the second number is the number of unsaturation, *e.g.*, 16:0 for a 16-atom long saturated chain (no double bond), or 18:2 for a 18-atom long chain containing two double bonds (Figures 2D, E) [1]. The nomenclature of the head groups and the tails allows composing the acronyms of the glycerophospholipids, the first two letters describing the tails and the last two letters describing the head group (Table 1).

[1] Polyunsaturated chains have a less-used alternative labeling in which the ending carbon in the chain is named omega and the double bond, if any, is numbered from this carbon. A typical example is omega-3 lipids, in which the first double bond is located at the third carbon atom from carbon omega. There also exist the omega-6 and omega-9 lipids.

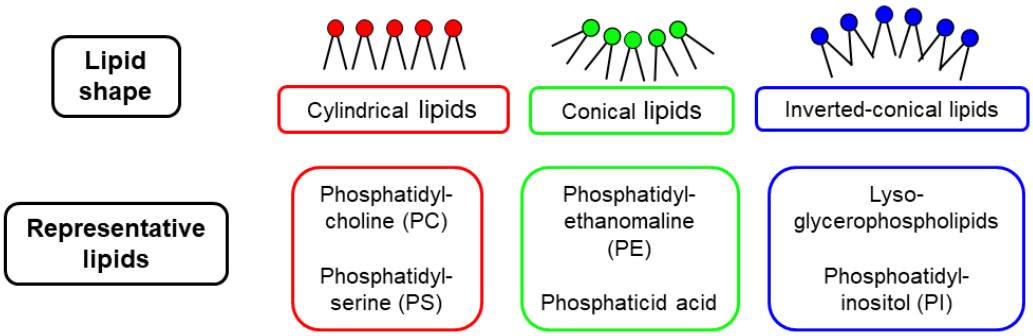



Figure 4: Schematic representation of the lipids shape with their representative lipids for each shape.

The size of both the polar head and the lipid tails affect the shape of the lipid bilayer[14] (Figure 4). Indeed, cylindrical lipids will more likely form flat lipid bilayers, as for PC and PS. Conversely, lipids with small head group and long unsaturated lipid chains can adopt negative curvature, as PE. This may induce disruption of lipid bilayers, PE being crucial for membrane protein embedding as well as modifying proteins through their ethanolamine moiety *i.e.* they anchor signaling proteins to the surface of the plasma membrane[15]. The lipids with large head and short lipid tails more likely result in positive curvature, as PI that control various processes including the formation of membrane trafficking vesicles[16,17] and protein regulation[18]. Biological membranes contain several types of lipids mixed together with various roles, *e.g.*, for vesicle fusion and fission during exocytosis, and membrane fluidity or protein regulation. Overall, the composition of lipid bilayers in glycerophospholipids is complex and it regulates complex pathways. They are the most prevalent lipid type in biological membranes, which may partly rationalizes their involvement in several diseases (*e.g.*, Alzheimer[19]).

Cardiolipin (Figure 5) is a specific glycerophospholipid; it is the dimer diphosphatidylglycerol or 1,3-*bis*(*sn*-3'-phosphatidyl)-*sn*-glycerol). In eukaryotes, cardiolipin is specific to mitochondria, mainly of the inner membrane[20], constituting up to 20% of mitochondria membranes. The acyl chain in cardiolipin varies from one to another organism or tissue. In humans, cardiolipin contains mainly di-unsaturated acyl chains (18:2)[21]. It has a specific function to regulate the structure of aggregates; it helps to build quaternary structure, *i.e.*, respiratory enzyme electron transport chain of the mitochondria; it triggers apoptosis when distributed to the outer mitochondrial membrane and interacts with cytochrome c[22]; it serves as proton trap for oxidative phosphorylation; it can also be involved in cholesterol translocation from outer to inner membrane of mitochondria. Alterations or default of cardiolipin can lead to several diseases, *e.g.*, in Barth syndrome[23] in which the enzyme responsible for the synthesis of cardiolipin is impaired, resulting in cardiomyopathy.

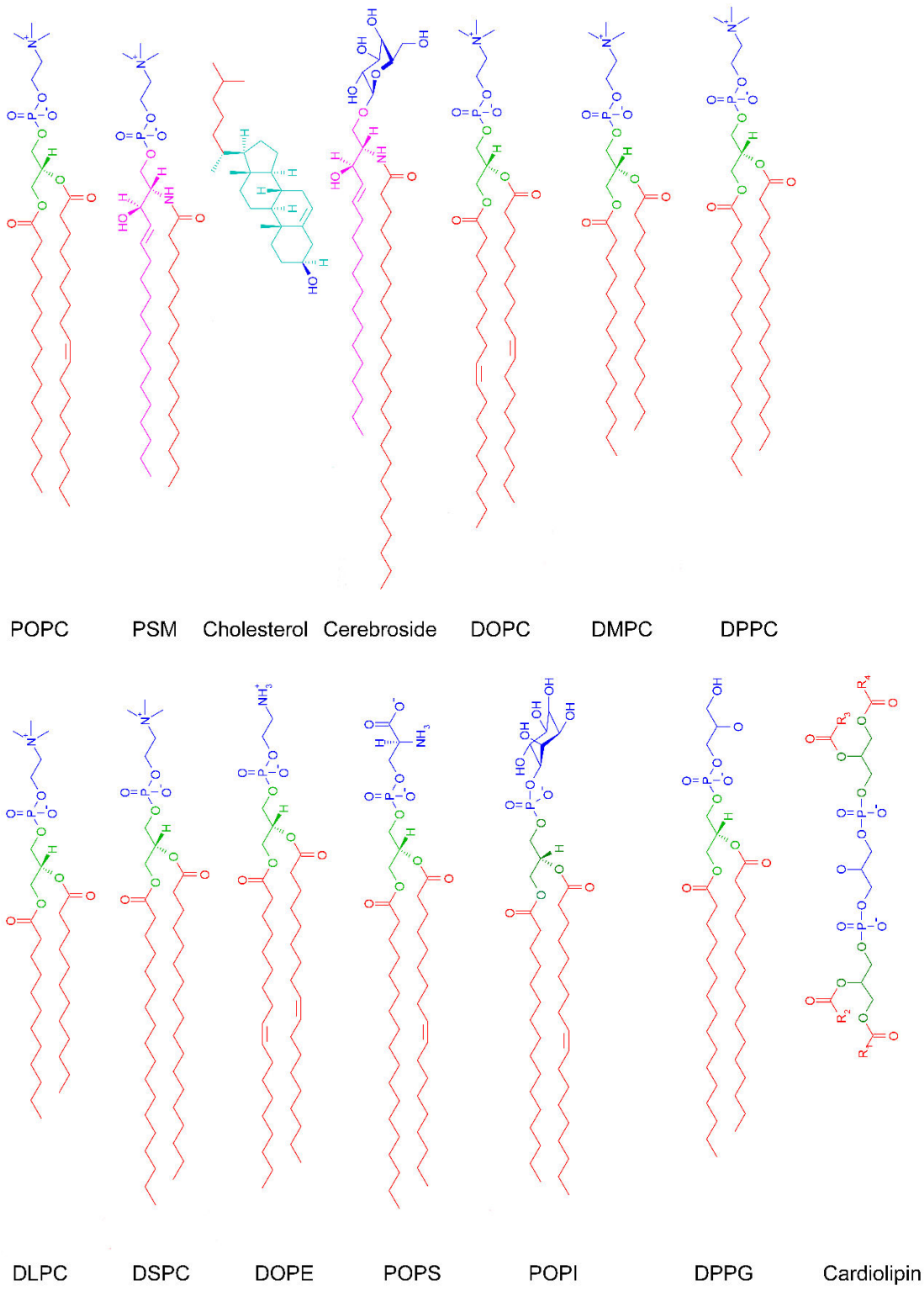



Figure 5: Chemical structures of representative lipids.

| Abbreviation | Chemical name |
|---|---|
| DOPC | 1,2-dioleoyl-*sn*-glycero-3-phosphatidylcholine |
| POPC | 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine |
| DPPC | 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine |
| DMPC | 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine |
| DLPC | 1,2-dilauroyl-*sn*-3-phosphocholine |
| DSPC | 1,2-distearoyl-*sn*-glycero-3-phosphocholine |
| DOPE | 1,2-dioleoyl-*sn*-glycero—phosphoethanolamine |
| POPS | 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-L-serine |
| POPI | 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoinositol |
| DPPG | 1,2-dipalmitoyl-*sn*-glycero-3-phospho-(1'-rac-glycerol) |
| PSM | N-palmitoylsphingomyelin |

Table 1: Commonly found phospholipids and their abbreviations.

#### I.2.1.2. Sphingolipids

The different sphingolipids differ from one to another by the length and the type of the sphingosine moiety (sphingoid-based backbone, equivalent to *sn-1* in glycerophospholipids), the substitution in the head group and the N-acyl chain (Figure 2B). The N-acyl chain is more likely saturated and longer than that of the glycerophospholipid acyl chains (with however some exceptions). As for glycerophospholipids, different substituents of the head group provide the diversity of this class of lipids (Figure 2B). The main mammalian sphingoid bases are the dihydrosphingosine and sphingosine. Ceramides are major components of the sphingolipid class of lipids. Sphingolipids play a crucial role in the signaling cascades involved in *e.g.*, apoptosis, stress response, inflammation[24–26]. They are also involved in cell signaling and recognition[27]. Sphingolipids have also been involved in several diseases, *e.g.*, in neurodegenerative disorders by modulating lipid-protein interactions[28], or in tumor progression due to differences in expression of ceramides and sphingosine-1-phosphate[29].

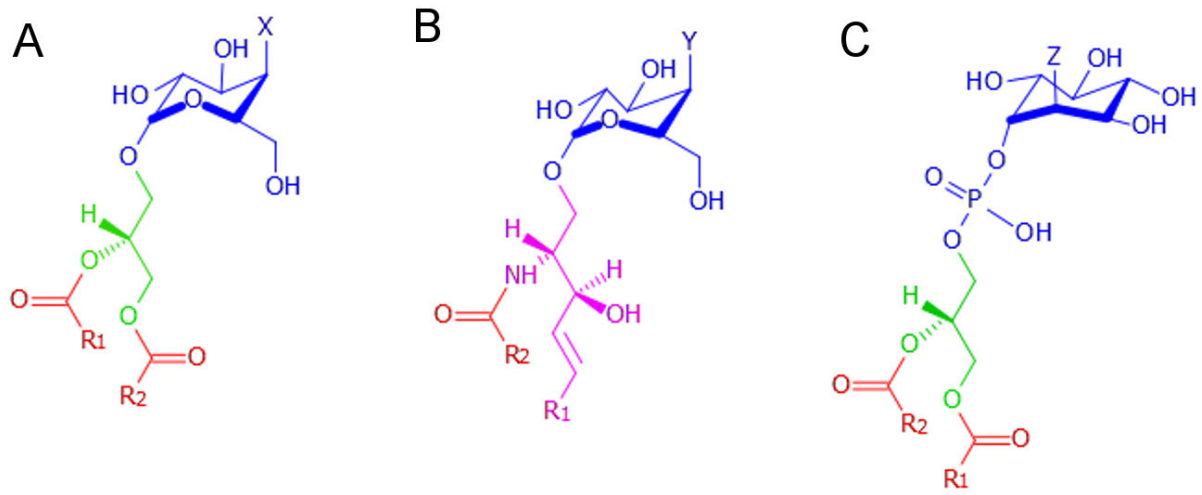



Figure 6: Chemical structures of the three sub-classes of glycolipids: A) glycoglycerolipid, B) glycosphingolipid and C) glycophosphatidylinositol. X is a hydroxyl group, one/two/three sugar moieties (*e.g.*, galactose), or a sulfoquinovosyl ($SO_3^-$ moiety). Y and Z are sugars moieties. $R_1$ and $R_2$ are N-acyl chains (Figures 2D and E).

#### I.2.1.3. Glycolipids

Glycolipids are divided in three main sub-classes: i) glycoglycerolipids made of two N-acyl chains (Figures 2D and E) linked to a monosaccharide or oligosaccharides (sugar chains) by a glycerol; ii) glycosphingolipids made of a sphingosine and a N-acyl chain (Figures 2D and E) linked to one (or more) sugar unit by an ether (*e.g.*, cerebroside) (Figure 6); and iii) glycophosphatidylinositol which is a PI with inositol moiety (Figure 3) linked to several sugar moieties. These oligosaccharides possess up to seven units, forming complexes like gangliosides. As sphingolipids, glycolipids are largely found in cellular membranes and are asymmetrically distributed between the inner and outer leaflets of the lipid bilayers, the carbohydrate (or saccharide) moiety being mainly found in the outer leaflet. Among other roles, glycolipids are responsible for ABO (blood type) antigens; intra and inter-cells interactions[30] (*e.g.*, in cell growth or apoptosis); immune response[31].

#### I.2.1.4. Sterols

Sterols are important constituent of lipid bilayers; in some cases, they constitute up to half of the lipid content in membranes. They are made of fused rings, a hydrocarbon tail and a hydroxyl group (Figure 2C). In membranes, the hydroxyl group of the sterols is oriented toward the water bulk (Figure 8C), embedded in the polar head groups of the glycophospholipids and sphingolipids. One typical case of sterols, only found in eukaryotic cells, is cholesterol (Figures 2C and 5); it is the most abundant sterol in animals. Sterols affect ordering in lipid bilayers, *e.g.*, cholesterol decreases ordering in gel phase membranes (*e.g.,* DPPC membrane at 298 K, Figure 8B), whereas the opposite effect is observed in fluid phase membranes (leading to $L_O$ phase, Figure 8C).

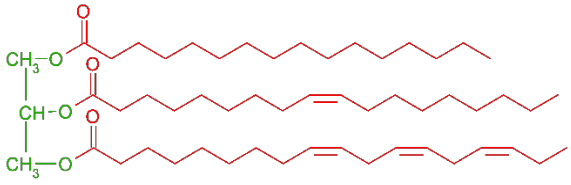


Figure 7: Representation of a triglyceride. The glycerol moiety and the N-acyl chains are depicted in green and red, respectively.

### I.2.1.5. Glycerolipids

Glycerolipids are a class of lipids composed of mono, di or tri-substituted glycerols (no polar head). The most known are the triesters of glycerol (R-COO-R'), the triglycerides, also known as triacylglycerol (Figure 7). Triglycerides are composed of three glycerol linked by an ester moiety to three fatty acid chains with different lengths (usually 16, 18 or 20 carbons) and unsaturation. They are the main form of stored energy in mammals. They are used for long-term storage of energy in humans; they are also used for heat production in new-born children[32]. They are mostly found in very-low-density lipoprotein (VLDL) that enables fats to move within the bloodstream, and they contain more than the double of energy compared to carbohydrates. However, high concentration of triglycerides in the bloodstream may lead to heart diseases[33], mainly atherosclerosis[34]. Glycerolipids are not components of lipid bilayer membranes.

### I.2.1.6. Lipid phase

The lipid composition in lipid bilayers drives their physical properties (*e.g.*, organization, phase, thickness, curvature, area per lipid, order parameter). The phase of lipid bilayers is related to its organization and fluidity[35,36]. There are three major phases in living organisms (Figure 8C):

i) the liquid disordered phase, also known as fluid phase or liquid crystalline phase ($L_d$ or $L_\alpha$, respectively, Figure 8A), is characterized by low order and high fluidity; this phase is favored by short or unsaturated lipids with *cis* double bonds *e.g.*, in pure PC lipid bilayers above their gel-to-liquid transition temperature Tm (Figures 8A and D);

ii) the liquid ordered phase ($L_O$) is characterized by high flexibility and high order; this phase is characteristic of mixtures containing cholesterol;

iii) the gel phase, also referred as crystalline phase, labelled $S_O$ or $L_\beta$, is characterized by a low fluidity and high order (Figure 8B); it is favored by long and saturated lipids (with high Tm) that align to form crystals.

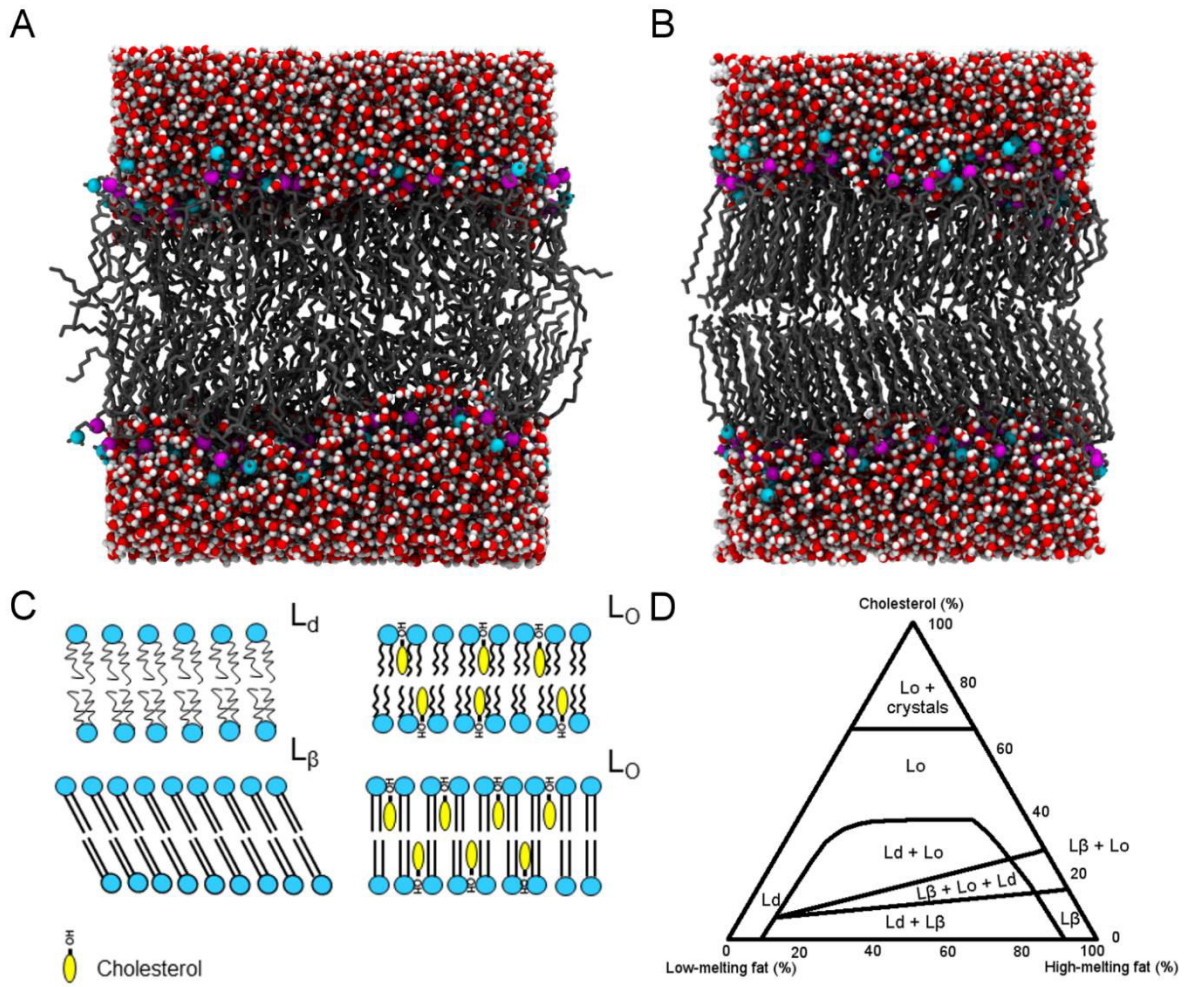



Figure 8: (A), (B) Representation of two lipids phases: liquid disordered $L_d$ and gel phase $L_\beta$, respectively. (C) Schematic representation of the organization of $L_d$ and $L_\beta$ and the effect of cholesterol on these phases. (D) Lipid phase diagram depending of the amount of low and high-melting lipids and cholesterol content. Adapted from [37].

Lipid bilayers are often inhomogeneous, which is crucial for cellular function, *i.e.*, lipid segregation from one to another leaflet but also within a given leaflet. “Patches” of specific lipids may occur, which are raft[38] or nanodomains (~10-200 nm). Cholesterol is a major components of these domains, as well as sphingolipids[39]. The existence of nanodomains has been a matter of intense debate mainly due to their highly dynamic character[40–42]. The cholesterol-enriched domains are involved in key biological processes including: signal transduction (*e.g.*, by recruiting proteins to allow activation of T-cells[43]); membrane trafficking (*e.g.*, rafts help trafficking from the Golgi apparatus to the cell surface by acting as a selective carrier and play on the membrane curvature[44]); membrane protein function (*e.g.*, proteins adapt to the physical properties of the membranes by binding to specific lipids and, in turn, influence their lipid environment[45,46]); protein-protein and protein-lipid interactions[47]. Due to their importance in several biological processes, disorders in the metabolism or synthesis of sterols can lead to severe diseases (*e.g.*, Parkinson or Lemli-Opitz - multiple malformation - syndrome)[48–50].

### I.2.2. Membrane structural analyses

#### I.2.2.1. Membrane thickness

The thickness is a highly lipid composition dependent parameter of lipid bilayers. For instance, a lipid bilayer is thicker with than without cholesterol because of the modulatory effect of cholesterol. Although the definition of membrane thickness is intuitive, boundaries of the lipid bilayers should be carefully defined[51], which has led to several modes of thickness calculation including:

i) the distance between the peaks of electron density between the head groups of the two leaflets, *i.e.*, Head-to-Head distance $D_{HH}$[51–55]; $D_{HH}$ can be obtained both experimentally and theoretically.

ii) by measuring the distance between the center of mass of the lipid phosphate moieties; this can be obtained only theoretically.

iii) by using the partial head group thickness $D_{H1}$ measured as the distance between the phosphates and the average lipid chain boundary (estimated experimentally[51–53]) , then to deduce the (hydrocarbon chain) thickness $2D_C$ with the half thickness $D_C$ defined as:

$$D_C = D_{H1} - \frac{D_{HH}}{2} = \frac{V_C}{A} \qquad (\mathrm{Eq.}\,1),$$

where $V_C$ is the volume of hydrocarbon chains of a monolayer (without including the carbonyl carbon), and $A$ is the averaged area per lipid (*i.e.*, averaged over both leaflets); it can be obtained both experimentally and theoretically[51–53,56].

iv) by using the (steric) head group thickness $D_{H'}$ of a monolayer and the half thickness $D_C$ (steric bilayer thickness, $D_{B'}$):

$$D_{B'} = 2(D_c + D_{H'}) \qquad (\mathrm{Eq.}\,2),$$

v) by using the ratio of volume per lipid $V_L$ of a monolayer and area per lipid $A$ (Luzzati thickness, $D_B$):

$$D_B = 2\frac{V_L}{A} \qquad (\mathrm{Eq.}\,3),$$

$D_B$ can also be evaluated from MD simulations, as a function of the box volume $V_B$ and the water volume:

$$D_B = 2\frac{(V_B - n_W V_{1W})}{A} \qquad (\mathrm{Eq.}\,4),$$

where $n_w$ is the number of water molecules and $V_{1W}$ is the volume of water molecules in the simulation box.

Various softwares can be used[57–59], which have different ways to define the head groups. For instance, APL@voro[59] uses the Voronoï diagram and the Delaunay triangulation.

#### I.2.2.2. Area per lipid

The area per lipid, $A_L$, is a fundamental property for lipid bilayers to describe the lipid phase and the effect of the lipid composition on the lipid bilayer behavior. As for the thickness, there are multiple ways and software to calculate the area per lipid.

The simplest way to calculate $A_L$ is to divide the area of the simulation box in the $xy$ plane by the number of lipids in one layer as for example:

$$A_L = \frac{\mathrm{Box}_X.\mathrm{Box}_Y}{N_L} \qquad \text{(Eq. 5)},$$

where $Box_X$ and $Box_y$ are the $x$ and the $y$ sides of the simulation box, $N_L$ is the number of lipids per leaflet.

$A_L$ can also be calculated using the Luzzati thickness $D_B$ and the volume of lipids $V_L$. It is worth noting that experimental area per lipid are error-prone while force fields of lipids have often been developed to fit with experimental $A_L$ values. Over the past years, careful attention has thus been paid in force field improvements to correctly predict these values[60].

### I.2.2.3. Order parameter

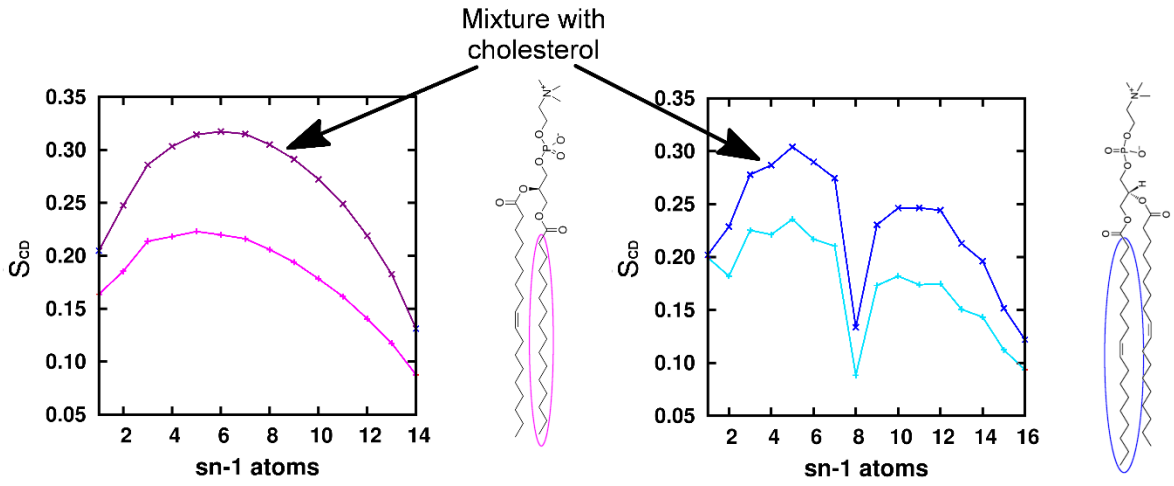



Figure 9: Difference of order parameter between a saturated chain of POPC lipid on the left and an unsaturated lipid chain of DOPC lipid on the right and the difference with membrane mixture with cholesterol.

The order parameter (Figure 9), $S$, is a dimensionless quantity related to the orientation and the flexibility of lipids in a lipid bilayer (*i.e.*, mainly related to entropy), it is defined as:

$$S = \frac{3}{2}\langle \cos^2 \theta \rangle - \frac{1}{2} \qquad \text{(Eq. 6)}.$$

The brackets average the angle over the simulation time and over all lipid molecules. The order parameter is defined by the fluctuation of the angles $\theta$ defined by the C-H bond vector with respect to the bilayer normal (*e.g.*, in membrane simulations typically $z$ axis). $S$ is in the range from -0.5 to 1. A lipid with no preferential orientation (equivalent to a lipid tumbling in homogeneous solvent), the $S$ value equals 0. However, this 0 value corresponds to the perfect order as well with $\theta = 54.7356°$ (*i.e.*, magic angle). This angle is obtained by solving the above equation with $S = 0$ and considering no average (*i.e.*, perfect order). A value of -0.5 indicates perfectly ordered acyl chains with all C-atoms in trans conformation. An orientation can only be translated in terms of order parameter from dynamical collection of data obtained by sufficient sampling.

### I.2.2.4. Lateral diffusion

Lipid bilayers are highly dynamic systems whether it is a simple lipid bilayer or a complex mixture of lipids with various phases (*e.g.*, nanodomains), or even protein complexes floating

on membranes. In this context, lateral diffusion is an important parameter that can be experimentally measured (by fluorescence correlation spectroscopy (FCS), fluorescence recovery after photobleaching (FRAP) or Single particle tracking (SPT)[61]) as well as computed from MD simulations. In the latter case, the most common way is to follow the mean-squared displacement (MSD) of lipids and to estimate the lateral diffusion coefficients $D_{lat}$ according to the Einstein approach[62] (assuming a homogeneous environment):

$$D_{lat} = \lim_{t\to\infty} \frac{1}{2n}\frac{d}{dt}\langle|\Delta r_{xy}(t)|^2\rangle = \lim_{t\to\infty} \frac{1}{2n}\frac{d}{dt} MSD(t) \qquad (\text{Eq. } 7),$$

where

$$MSD \equiv \langle (x - x_0)^2 \rangle = \frac{1}{N}\sum_{n=1}^{N} \big(x_n(t) - x_n(0)\big)^2 \qquad (\text{Eq. } 8),$$

where $N$ is the number of particles, $x_n(0) = x_0$ is the position of reference (initial position) of particle $n$ and $x_n(t)$ is its position at time $t$. The calculation of the diffusion coefficient allows for instance to check if the fluidity of the simulated lipid bilayer is realistic.

#### I.2.2.5. Membrane curvature and lipid packing

Local membrane curvature is often observed in membranes, mainly due to asymmetric distribution of lipids in the inner and outer leaflets, which is however highly dynamic[63,64]. The shape and charge of the constituent lipids affect the curvature[65], *e.g.*, lipids with short tails and large heads induce a positive curvature whereas lipids with long tails induce negative curvature[63]. Curvature can be evaluated using the lipid packing parameter $P$[66]:

$$P = \frac{v}{al} \qquad (\text{Eq. } 9),$$

where $v$ is the molecular volume, $l$ is the lipid length and $a$ is the head group cross section area. The curvature of lipid bilayers is closely correlated with lipid packing[67]. It can be obtained experimentally (*e.g.*, NMR[68], X-ray SAXS[69,70]) as well as from MD simulations by clustering[67] a minimum of four lipids (using a distance cut-off between the lipids) and a minimum of a few ns to consider the cluster as stable in time (not too long as the clusters can be physically highly dynamic, *i.e.*, transients).

#### I.2.2.6. Inter-lipid interactions

Interactions between lipids, *i.e.*, between lipid tails or lipid tails and cholesterol, are directly related to mechanical properties of membranes (*e.g.*, fluidity, order, thickness). These interactions not only affect the lipid bilayer physical properties but also drug partitioning[35]. They can be evaluated as the average of the short-range Coulomb and the Lennard-Jones potentials ($V_{tails}$) between all pairs of atoms belonging to the lipid tails; the lipid tails are defined by all atoms of the tails and the three glycerol carbon atoms for the glycerophopholipids, or by all atoms but the hydroxyl group for cholesterol. The $V_{tails}$ per atom values show that the higher the cholesterol concentration, the stronger the interactions, confirming the ordering (packing) effect of cholesterol in lipid bilayers.

The mass density of groups of atoms can be given as distribution profiles along an axis or in a plane. In a plane, the RDF pictures interactions between a group of atoms and another one (*e.g.*, a protein domain and lipids of a surrounding bilayer). The RDF, or $g_{AB}(r)$, between the groups of particles $A$ and $B$ is given as follows:

$$g_{AB}(r) = \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle_{local}} = \frac{1}{\langle \rho_B \rangle_{local}} \frac{1}{N_A} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \frac{\delta(r_{ij} - r)}{4\pi r^2} \qquad \text{(Eq. 10)},$$

where $\rho_B(r)$ is the density of particles $B$ at a distance $r$ around particles $A$, and $\langle \rho_B \rangle_{local}$ is the density of particles $B$ that is averaged over all spheres around particles $A$ with the radius $r_{max}$ (with usually half the box length). The RDF is averaged over the simulation time.

### I.2.3. Lipid composition in tissues and organelles

#### I.2.3.1. Composition diversity in tissues

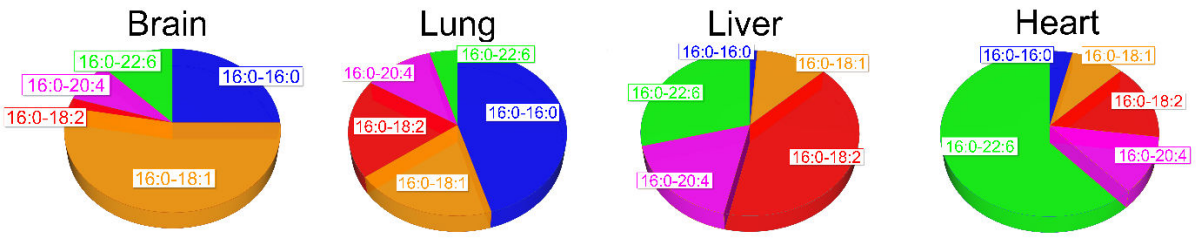



Figure 10: Distribution of phosphatidylcholine in brain, lung, liver and heart based on the sn-1 and sn-2 chains length and unsaturation.

The distribution of membrane lipids (*e.g.*, lipid type, chain length or the degree of unsaturation) differs from one organ to another (Figure 10), depending mainly on its function. For instance, the 16:0-16:0 chains are more important in the brain and in the lung. In the lung, they play the role of pulmonary surfactants, reducing surface tension on alveolus to help breathing[71]. Another example is the 16:0-18:2 PC which is a major component in the liver, as these lipids (mainly provided from dietary sources) are highly secreted in the bile with cholesterol and bile acids, they are essential as they are converted in arachidonic acids and derivatives for the organism needs[72].

The lipid composition may differ from one tissue to another in the same organ. For instance, in nerves, myelin sheath is rich in sphingomyethrocytes (associated with a change in function)[73]. This can be due to i) the formation of SM and cholesterol enriched domains, ii) insulation of the sheath, or iii) SM participation in the metabolism of particular lipids (*i.e.*, ceramide).

Another interesting feature is that brain tissues are enriched in PE (~45% of phospholipids) and PS, this can be explained because PE is the precursor for the synthesis of the ligand for cannabinoid receptors[74].

A typical example of membrane lipid composition is in the erythrocytes (red blood cells)[75,76]. The proportions of PC and SM are higher in the outer leaflet compared to PE and PS, preferentially found in the inner leaflet. This distribution may be explained by the main lipid feature which is the presence of lipid domains composed of SM and cholesterol. Their presence induces the reshape of the erythrocytes (associated with a change in function)[73].

The composition of the external layer of the skin (also called *stratum corneum*) is often less studied, however the importance of this bilayer is important as for permeation[77]. This external layer of skin is composed of piles of dead cells (corneocytes) surrounded by a lipid matrix composed of fatty acids, cholesterol and mainly ceramides[78]. This external layer of skin acts as a barrier against external molecules and microorganisms, mainly due to the ceramides.

### I.2.3.2. Lipid composition in organelles

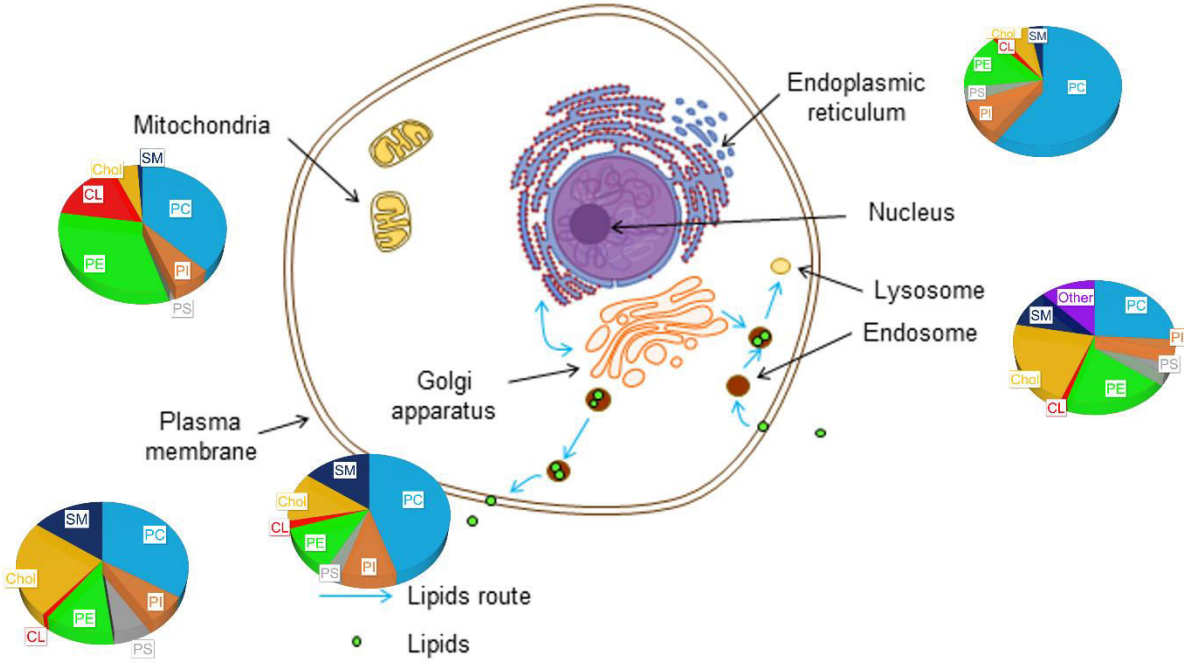



Figure 11: Distribution of lipids in organelles: phosphatidylcholine, PE: phosphatidylethanolamine, PS: phosphatidylserine, PI: phosphatidylinositol, CL: cardiolipin, Chol: cholesterol, SM: sphingomyelin. Adapted from [79].

The lipid distribution in organelles is related to their function (Figure 11)[80]. The lipid route begins at the Endoplasmic Reticulum (ER). Lipids are then transferred reversibly to the Golgi apparatus. They are then taken by the endosomes, which transport them either to the plasma membrane or to the lysosomes (for degradation). The plasma membrane can reversibly transfer lipids to endosomes. The mitochondria is not connected to this vesicular transport system; in this case lipids are directly issued from the cytoplasm. Here, lipids can insert in the mitochondria membranes either by passive diffusion or through membrane proteins.

Independently of the organelles, membranes are mainly constituted of PC (~40-50%) and PE (~35%), which are essential components for the lipid bilayer structuration. There are however specificities according to the organelles. The ER is the first secretory step because it is the site for lipid synthesis. It produces several lipids including phospholipids and cholesterol, among others. However, the cholesterol is quickly transported to other organelles explaining the low cholesterol content in ER. Consequently, lipid packing is decreased in the ER membrane, driving insertion and transport of newly synthetized lipids; this stands also true for facilitated protein insertion. The Golgi apparatus is a lipid sorting platform. This sorting is achieved thanks to rafts (made of cholesterol) and sphingolipids. Endosomes and lysosomes are involved in lipid-mediated signaling[17] (*e.g.*, to orient the transport to the right organelle for the endosome) and recognition for lipid-protein interactions. Endosome and lysosome functions are ensured because of lipid composition (in PC and cholesterol) close to that of the plasma membrane. The plasma membrane is a highly selective barrier for cells. It is a platform for the intra- and extra-cellular signalization (often through membrane proteins). These roles are related to the high content of cholesterol and sphingolipids, *e.g.*, by reinforcing the role of barrier (more packed membrane so less permeable to xenobiotics) or by being involved in rafts participating in cell signalization. In plasma membrane, PS is another example of crucial lipid which are involved in apoptosis as signaling agents (*i.e.*, they are located in the inner leaflet

and their flip-flop signals the apoptosis). Among others, the mitochondria contains a specific lipid *i.e.*, cardiolipin, which constitutes up to 20% of mitochondria lipid composition. It is essential for normal functioning of this organelle. PE is also a crucial component of mitochondria membranes, supporting oxidative phosphorylation. The high content of PE is explained by the reminiscence of its presence in bacteria.

## I.3. Membrane proteins

### I.3.1. Metabolizing enzymes

The CYP proteins are found in most of the living organisms[81]. While bacterial CYPs are soluble, eukaryotic CYPs are membrane proteins, mainly bound to the different organelles, and to a less extent to plasma membranes. They are heme-containing proteins, heme being the catalytic site where metabolic reactions occur. They are anchored to the membrane of either endoplasmic reticulum or mitochondria, and to a much lower extent in the Golgi apparatus, peroxisomes or plasma membranes[82–85]. The sequence length of mammalian CYPs is around 500 amino acids. A ~20–25 amino acid long N-terminal transmembrane helix is anchoring CYP to the membrane. This α-helix is attached to the catalytic domain, which is partially in contact with the polar head group region of the lipid bilayer (Figure 12). CYPs catalyze mostly hydroxylation or dealkylation reactions. As a result, phase I degradation products tend to have lower average molecular weight by 20–25 Da and they tend to be more polar; only a minority of metabolites (4–8%) are more lipophilic than their parents due to some dealkylation reactions[86]. Moreover, affinity of phase I metabolites towards lipid bilayer is lower whereas the barriers for passive membrane crossing are higher than for substrates[87,88]. The excretion of the metabolites is, in principle, faster than for the parent molecule.

Several groups of CYPs have been defined, including CYP a, b, c and CYP450[89,90]. The CYP450 group constitutes an ensemble of proteins that have been highlighted as of crucial clinical and pharmacological importance[91,92], as being responsible for the metabolism of many xenobiotics. Subsequently CYP450 is related to drug action and toxicity. The access of the xenobiotics to the catalytic site (heme) is in the lipid bilayer, in contact with the polar head group surface (Figure 12). In this membrane-access channel, the xenobiotic goes through specific regions in the CYPs, based on its chemical nature. The lipid bilayer plays an important role in the function and kinetics of the CYP450 proteins, by accumulating the xenobiotics in the lipid bilayer thus enabling them to enter the CYP active site (buried inside the enzyme) from the membrane. The active site can conformationally accommodate to a wide range of

substrates[93–96]. After transformation in the active site, the metabolites are then release through an outward-facing channel.

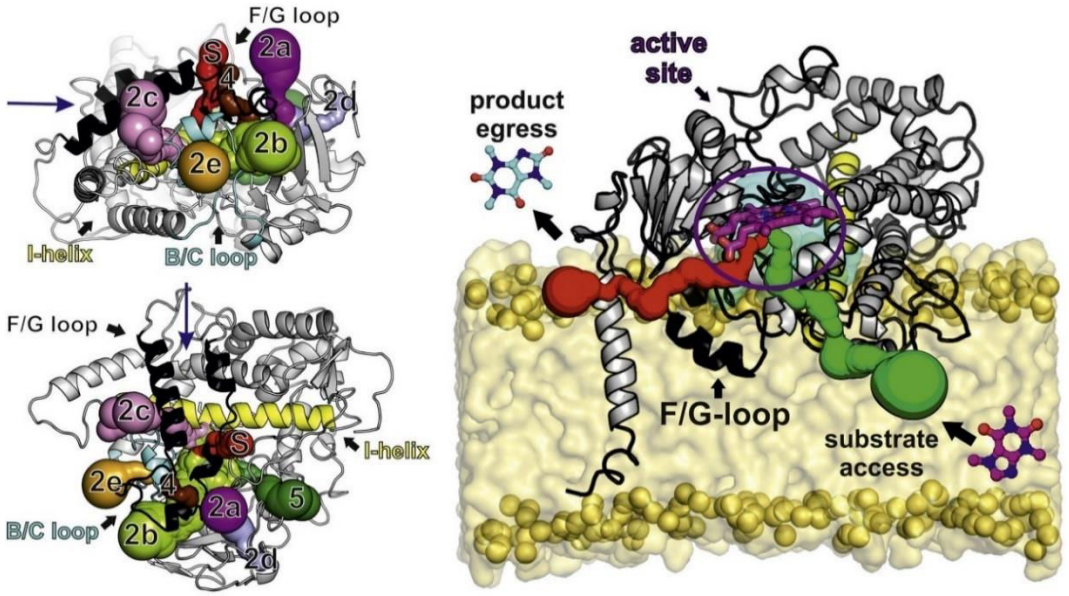



Figure 12: Localization of gating channels in the structure of CYP3A4 (top and side view in the left and right panels, respectively). The F/G, B/C loops and I-helix are known regions of the CYP where the channels go through. On the right panel, the channels (depicted in red and green) suggest the metabolic pathway for amphiphilic compounds from the membrane to the CYP active site. The heme moiety is depicted in violet. Reprinted from [92], with permission from Elsevier.

Cytochromes P450 are important especially in the context of inter-individual variability in drug response. This variability is particularly sensitive in polymedication (*e.g.*, cancer or organ transplantation), in which drug-drug interactions are likely to occur. Genetic polymorphisms modulating CYP450 proteins is a major factor to rationalize inter-individual variability[97]. Many CYP450 isoforms have been identified so far, and the genetic map of several of them have been investigated in the context of organ transplantation, for which there exists a strong interest in strategies to adapt the treatment to each patient. The CYP3A4 isoform is responsible for the metabolism of more than half of drugs that are metabolized. This isoform is also the most abundant in humans, especially in the liver and in the small intestine[98]. Another example is the CYP2D6 isoform which is widely studied as this protein is involved in the metabolism of morphine, codeine, and neuroleptics[99]. The CYP2C9 isoform, found in the liver, has the hypertensor losartan tamoxifen, and fluvastatin as substrates[100,101],.

The most important enzymes catalyzing phase II reactions are UGTs, GSTs, NATs, and SULTs [5]. Of those, the major drug metabolizing enzymes are from UGT1 and UGT2 subfamilies. UGT enzymes transfer a glucuronic acid to the drug forming a glycosidic bond[102,103]. A specific transporter among the UGT1 family, the UGT1A1 is known to metabolized bilirubin.

The complete human UGT crystal structure is not available, but the crystal structure of the UGT2B7 C-terminal domain was published[104]. This structure together with templates for the missing N-terminal domain from various sources (plant UGT enzymes, bacterial glycosyltransferases and macrolide transferases) have served for construction of several atomistic UGT models[5,105–108]. Albeit none of the models was built as cast in a membrane, they were used to identify the effects of amino acids in UGT1A1, whose mutations are known to play a role in the Crigler-Najjar type I or II syndromes or the Gilbert syndrome[105], or to

assess substrate specificity between UGT1A9 and UGT1A10[107]. The full atomistic membrane model of UGTs is however still missing.

Statistical analysis over the phase II metabolites showed rather systematic modifications of the physicochemical properties of bio-transformed drugs and drug-like molecules. Conversely to phase I reactions, phase II conjugation reactions lead to a substantial gain in molecular weight (145 amu on average). Furthermore, phase II reactions lower logP values of drugs by 1.4 log units on average[86]. This agrees with previous analysis of membrane positioning of quercetin metabolites by means of MD simulations[2]. The conjugated substituent groups pulled quercetin moiety toward membrane surface. Overall, biotransformation reactions usually weaken drug-membrane interactions, leading to the excretion of drug in a final phase.

### I.3.2. SLC transporters

Drug membrane crossing (influx or efflux) can occur by facilitated diffusion, which is mediated by a superfamily of membrane transport proteins named solute carrier (SLC)[109,110]. This transfer is a secondary active transport, which means that it only uses the energy stored in electrochemical gradients. More than 380 different SLC proteins, divided into 52 families, have been annotated in the human genome[111–116]. Most of the SLC transporters can be grouped into two large structural folds in membrane transporters, namely major facilitator superfamily (MFS) and neurotransmitter:sodium symporter (NSS) or LeuT-like fold[114]. Here a particular attention is paid to the MFS superfamily since only those influx transporters are described as being of "emerging clinical importance" in drug discovery[110] by the International Transporter Consortium[117].

*In silico* studies on human MFS transporters are rather limited, mainly due to the absence of X-ray crystallographic structures. However, homologous models (mostly bacterial) have provided substantial insights in SLC transporter functions[115,118–120]. It is important to note that the surrounding membrane is not necessarily involved directly in the large-scale events; however, it naturally restrains protein movements to some extents. The use of protein models embedded in lipid bilayers is thus mandatory to simulate the transport cycle. In this section, we will only focus on *in silico* studies in which the membrane is explicitly included.

MFS transporters all adopt similar secondary structure consisting of twelve transmembrane helices (TM) organized into two N-terminal (TM1-TM6) and C-terminal domain (TM7-TM12) substructures (Figure 13A). The X-ray crystallographic structures supported by computational studies described two main conformations, namely inward- or outward-facing (IF and OF, respectively, see Figure 13B). The drug/substrate transport across the membrane is thus driven by large-scale conformation changes from one to the other state, following a "clamp-and-switch" model of function[121]. This suggests the existence of intermediate occluded conformations in IF- and/or OF-states, later confirmed by the X-ray structure of the Xylose:H+ symporter (XylE)[122,123]. Such mechanical pathway was also suggested in several transporters (namely, xylose XylE[124], glucose[2] GLUT1[126] and lactose LacY[127–129] transporters). It is worth noting that the "clamp-and-switch" model is actually a recent enhancement of the "rocker-switch" model[121]. Particular attention should be paid to a recent computational study on the glycerol-3-phosphate:phosphate transporter (GlpT)[130]. Although this transporter drives efflux, this study provides useful insights in the deep understanding of

[2] It must be stressed that this model was built by homology modeling using XylE as reference[124]. The theoretical results agreed with the recently crystallized human GLUT1 in the IF-state[125].

influx since all MFS transporters are very likely to share the "clamp-and-switch" mechanical pathway.

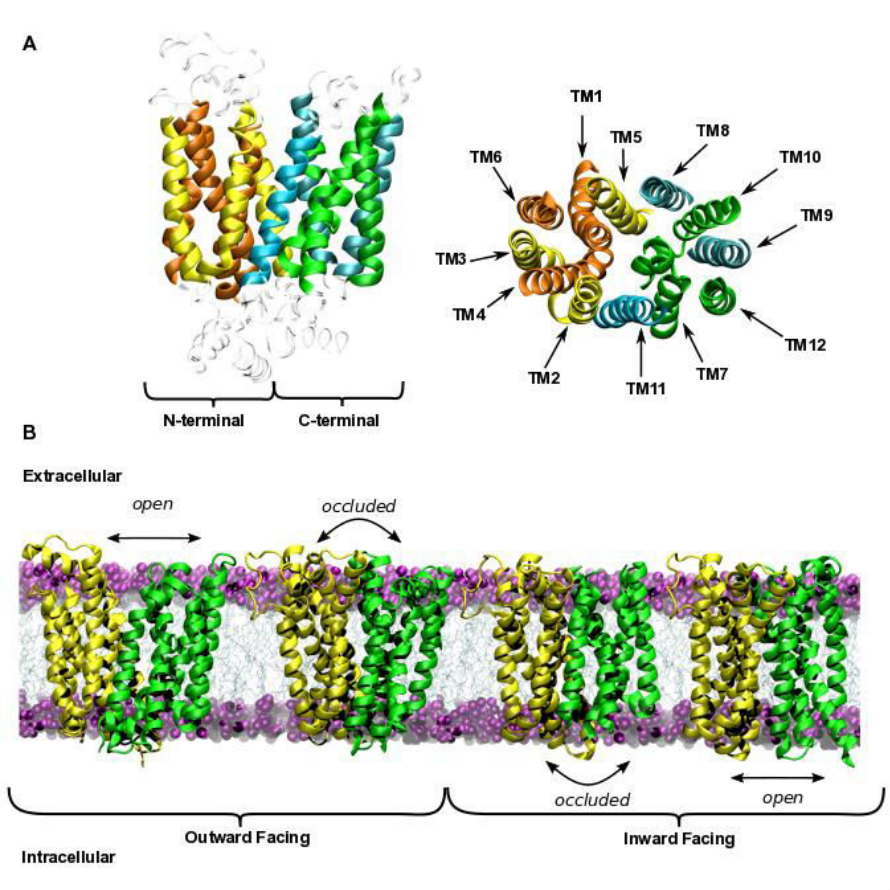



Figure 13: (A) X-ray structure of Glucose transporter type 3 (PDB-ID: 4ZW9 - GLUT3[131], a detailed structural description is also available in ref. [121]). (B) Examples of the several conformations detailed in the “clamp-and-switch” model from X-ray structures. Occluded-OF and occluded/open-IF structures were taken from X-ray structures of XylE transporters (PDB-ID: 4GBY[123], 4JA3[122] and 4JA4[122], respectively); open-OF structure was taken from the GLUT3 transporter. N- and C-terminal domains are depicted in yellow and green, respectively.

The complete Gibbs energy landscape of transport cycle was then calculated with and without substrate underlining the "catalytic" role of the substrate. The existence of occluded-IF and -OF states was also confirmed to be thermodynamically stable states if and only if the substrate is bound to the transporter. This should explain why occluded states are so challenging to observe experimentally, since the substrate is not always present in crystal structures. Furthermore, details about the role of specific TMs along the transport cycle were investigated to correlate local and global conformational changes. The role of TM1 and TM7 to the periplasmic gating process was stressed as observed in GLUT1 influx transporter[126], whereas TM4, TM5, TM10 and TM11 were shown to be involved in the cytoplasmic gating. Such observations have strengthened the hypothesis of the asymmetrical global conformational changes along the transport cycle of MFS.

Joint experimental and computational studies are likely to elucidate more local events. For example, in $H^+$-cotransporters (*e.g.*, LacY and XylE), the $H^+$-conduction (electrochemical gradient) from the periplasmic to cytoplasmic side is shown to provide the energy required to the transport cycle. Such conduction along amino acid side chains (mainly Glu, Asp and His) proceeds by proton transfer reactions that locally change charge states of protonatable

residues, leading to salt-bridge breaking/formation that trigger the IF-to-OF transition[124,128,129,132,133].

Although MFS transport cycles have been extensively investigated theoretically, only a few studies have tackled the direct influence of the surrounding membrane. Yet, studying lipid-protein supramolecular assemblies[129,134] as well as considering membrane composition [129] have repeatedly been suggested to be of crucial importance. Another currently missing aspect is the direct study of human transporters of clinical interest. Only three human MFS transporter structures have been elucidated so far[125,135]. To the best of our knowledge, either homolog (*e.g.*, with PEPT-1/2, see [136]) or homology-modeling-made (*e.g.*, with PEPT-1 and OAT1, see refs. [137,138]) transporters were used so far. Interestingly, the dynamic of the overall mechanism of PEPT-1 was found to be very close to that of LacY although they belong to two different distant SLC families. This supports the hypothesis that MFS transporters exhibit large-scale conformational transitions during transport cycle regardless substrates. MFS transporters mainly differ in local binding events that govern their substrate selectivity[136].

A particular attention should also be paid to the MATE (multi-antimicrobial extrusion protein) transporter, which represents the only efflux SLC-type transporters considered of emerging clinical importance by the International Transporter Consortium[110]. Although a few theoretical investigations have been achieved[139], they can hardly be transposed to the human homolog functions (MATE-1 and MATE-K). This is complicated by the fact that no X-ray crystallographic structure is available for these human transporters so far.

### I.3.3. ABC transporters

ABC (ATP-binding cassette) transporters are membrane proteins which are involved in the transport of a broad range of xenobiotics, including drugs, nutrients, endogenous compounds and toxic agents. They are ATPase, *i.e.*, ATP-dependent substrate transporter. They locate in plasma membranes but also in the membranes of cell organelles (endoplasmic reticulum, inner mitochondrial membrane, peroxisomal and vacuolar membranes). ABC transporters are found both in prokaryotes and eukaryotes. Phylogenetics suggests that eukaryotic ABC transporters have been originated from bacterial ABC transporters, and high level of conservation is indeed often observed. There has been growing clinical interest in ABC transporters, as their dysfunctions (*e.g.*, due to polymorphisms or rare mutations) have been related to various diseases, which is critical in multidrug treatments (*e.g.*, cancer and organ transplantation). About 18 of 48 ABC having been implicated in the clinical manifestation of various diseases[140].

#### I.3.3.1. The ABC superfamily

ABC transporters constitute a superfamily of membrane proteins. Sav1866 was the first elucidated ABC structure isolated from *Staphylococcus aureus*[141]; it has been extensively studied as a prototypical ABC. MsbA is an ABC involved in the transport of lipid A, which is responsible for the toxicity of gram-negative bacteria; it has also been widely studied as being homolog of human multidrug resistance proteins[142]. Mammalian ABC transporters have been classified into several sub-families. These sub-families have been designed mainly according to their biological function and to their structural features. High level homology, as well as similarity in the global structural feature, can be seen between members of the different sub-families[143].

The ABCA sub-family contains twelve members which are mainly expressed in plasma membranes. ABCA1 is a lipid translocase, known as the Cholesterol Efflux Regulatory Protein (CERP); it also regulates the efflux of phospholipids[144].

The ABCB sub-family contains ABCB1, the P(Permeability)-GlycoProtein 1 (P-gp or Pgp) also known as MultiDrug Resistance protein 1 (MDR1). This transporter has gained much clinical interest because it has many drugs as substrates[3], including anticancer drugs. ABCB2 and B3 are also known as Transporter associated with Antigen Processing (TAP 1 and 2, respectively). They belong to the Peptide-Loading Complex (PLC), which is constituted by TAP 1 and 2, the oxidoreductase ER p57, the MHC-I heterodimer, and the chaperones tapasin and calreticulin[146]. This supramolecular complex is expressed in the endoplasmic reticulum and it coordinates the translocation of peptides between the endoplasmic reticulum lumen and the cytosol. ABCB6-8 and 10 are more specific to the mitochondria.

The ABCC sub-family contains twelve members. ABCC1-6 and ABCC10-12 are known as Multidrug Resistance Proteins (MRP1-6 and MRP7-9, respectively), they transport a wide range of organic anions and hydrophobic drugs. The ABCC7 is a ATP-gated ion channel, known to transport chlorides. It also regulates other channel proteins. It is named Cystic Fibrosis Transmembrane conductance Regulator (CFTR). Indeed, CFTR dysfunctions in the lung, the pancreas, the liver, the kidney, and the intestine result in cystic fibrosis, a hereditary disease that causes the body to produce thick mucus clogging the organs as induced by hyperactivation of the chloride channel. ABCC8 and C9 are also named SUlfonylurea Receptors (SUR 1 and 2, respectively); they regulate the permeability potassium channels and they are part of the supramolecular hetero-octameric transmembrane complexes, named inward-rectifier potassium (K) ion channels (Kir6.2 and 6.1, respectively)[147].

The ABCD sub-family are half-transporters expressed in the peroxisomal membrane. There exist four members of this family, *i.e.*, ABCD1/ALDP (AdrenoLeukoDystrophy Protein), ABCD2/ALDR (ALDP-related), ABCD3/PMP70 (70kDa Peroxisomal Membrane Protein) and ABCD4/PMP70R (PMP70-related). They are named according to the disease in which there are involved, namely, chromosome X-linked disorders responsible for a neurodegeneration and adrenal deficiency disease.

The ABCE sub-family is made of a single member ABCE1, it is an Organic Anion-Binding Protein (OABP), which should not be confused with SLC transporters. As all members of the ABCF sub-family, ABCE1 is not a transmembrane ABC protein.

The ABCG sub-family contains several members including ABCG5 and 8, which have been involved in a rare recessive genetic disease (sitosterolemia), characterized by the retention of dietary cholesterol and abnormal retention of non-cholesterol sterols (phytosterols) in the body. The main member of this sub-family is ABCG2, which is a multidrug resistance protein named

[3] To exemplify how broad is the range of P-gp substrates here is a non-exhaustive list: Actinomycin D, Aldosterone, ALLM peptide, ALLN peptide, Amitriptyline, Amprenavir, Atorvastatin, β-amyloid, Bromperidol, Calcein acetoxymethylester, Carbamazepine, Celiprolol, Chloropromazine, Clopidogrel, Cimetidine, Citalopram, Colchicine, Corticosterone, Cortisol, Cyclosporine A, Daunorubicin, Dexamethasone, Digoxin, Diltiazem, Docetaxel, Domperidon, Doxycycline, Doxorubicin, Erythromycin, Etoposide, Fexofenadine, Grapefruit juice, Gramacidin D, Gramacidin S, Imatinib, Indinavir, Irinotecan, Itraconazole, Ivermectin, Ketoconazole, Lamotrigine, Lansoprazole, Levetiracetam, Levofloxacin, Loperamide, Losartan, Lovastatin, Melphalan, Methylprednisolone, Mevinolin, Mitomycin C, Mitoxantrone, Morphine, Nelfinavir, Omeprazole, Ondansetron, Paclitaxel, Pantoprazole, Pentazocine, Phenobarbital, Phenothiazine, Phenytoin, Propanolol, Quinidine, Ranitidine, Rifampicin, Ritonavir, Saquinavir, short chain lipids, Simvastatin, Sirolimus, Sparfloxacin, Tacrolimus, Talinolol, 99mTc-MIBI, Teniposide, Terfenadine, Tetracycline, Topotecan, Valspodar, Vecuronium, Verapamil, Vinblastine, Vincristine, from [145].

BCRP (Breast Cancer Resistance Protein). BCRP is considered as one of major transporters causing drug resistance in mammalian cells, together with P-gp and MRP transporters.

### I.3.3.2. Structural features of ABC transporters

Despite highly variable amino acid sequences due to the large number of variants that appeared with phylogeny, most ABC transporters exhibit surprising structural similarities. ABC transporters can be full transporters (one amino acid sequence, *e.g.*, P-gp and MRP4), as well as homo- or hetero-dimers (non-covalent combination of two half-transporters, *e.g.*, BCRP).

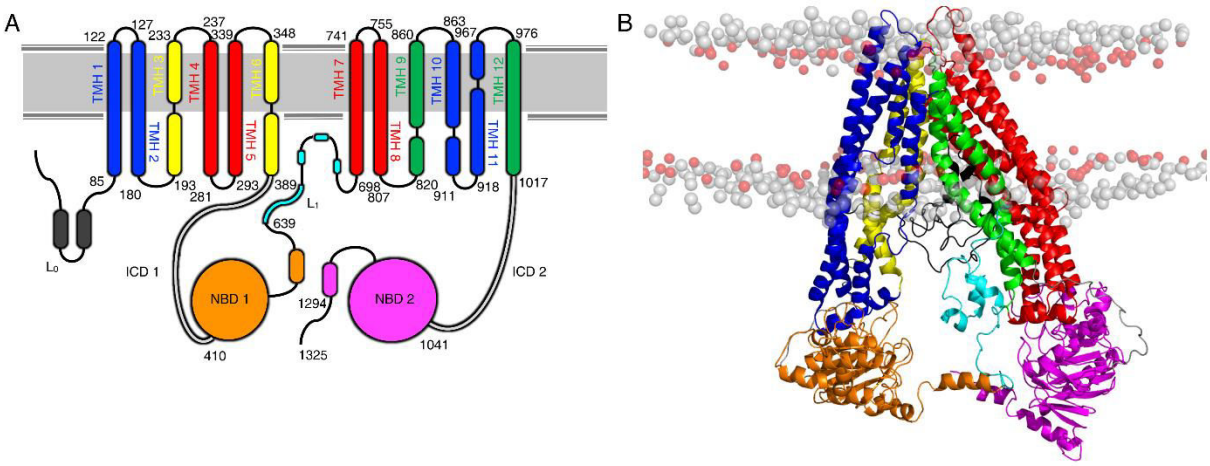



Figure 14: A) Topology of the human MRP4 exporter, highlighting the two TMDs made of twelve TM (TM1-12); the two NBDs (NBD1 and NBD2); ICD 1-2; the linker $L_1$; the zipper helices; the $L_0$ motif. If any, the irregularities in helices are depicted. B) Fully relaxed 3D-structure of the molecular WT MRP4 model. Figure and caption taken from [148].

Most ABC transporters are made of twelve (*i.e.*, 2 times 6) TransMembrane alpha-helices (TM), which constitutes four different bundles (Figure 14), but some are made of five to eleven helices. TMs have been poorly conserved during evolution. Although their membrane-anchoring role is similar for almost all membrane ABC transporters, their amino sequence many dramatically modulate their substrate binding capacity and specificity, as well as their mechanism of xenobiotic translocation. Indeed, TMs define the protein chamber where substrates dynamically bind during their transport (see below for more details). Some discontinuities in TMs may impact the global conformational feature thus transport function. For instance, discontinuity in the TM8 of CFTR, which makes two sharp breaks inside the lipid bilayer membrane, modifies hydrogen bonding capacity, in turn modifying interactions with ions and water along the ion-conduction pathway[149]. Single mutations in this region (*e.g.*, L927P) may create flexible hinges, thus modifying the gating, hence impacting on cystic fibrosis.

The ABC transporters are made of two nucleotide binding domains (NBDs), see Figure 14, which are highly conserved regions. NBDs contain the catalytic sites, where ATP molecules can bind and be hydrolyzed. In homo-dimers, the two NBDs are symmetrical and they operate in the exact same way, *i.e.*, random binding to one, the other, or both. In hetero-dimers or single-polypeptide-chain ABC transporters, evolution has created asymmetry between the two NBDs. In such cases, it is suggested that the ATP molecule preferentially binds the consensus NBD domain rather than to the degenerate NBD. Competitive inhibitors have been identified, which preclude ATP-binding and impair the ABC function (*e.g.*, verapamil, cyclosporine A, cnidiadin in P-gp).

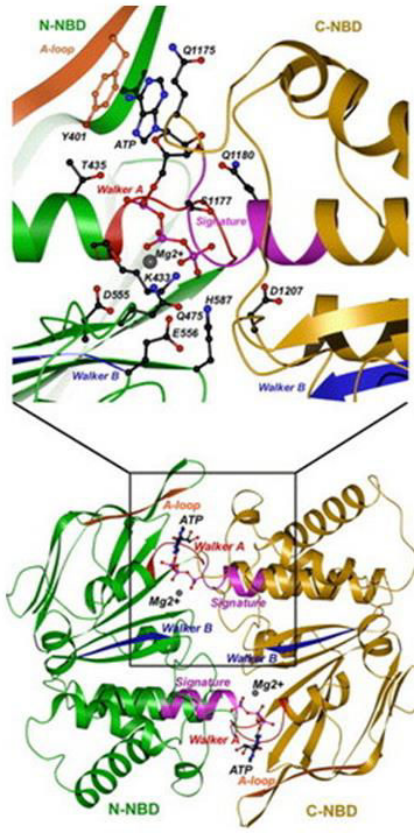



Figure 15: Model structure of the dimeric Pgp nucleotide-binding domain showing conserved residues involved in ATP-binding. The atomic model of the dimeric NBDs of Pgp was based on the coordinates from the crystal structure of the MJ0796 E171Q dimer (PDB code: 1L2T)[150] by mapping the N- and C-terminal NBD sequence to either subunits, respectively. The model was built with the program O[151]. The resulting dimeric NBD was subjected to a few cycles of molecular dynamics in the CNS program to remove unfavorable interactions present in the initial model and to conform to the standard stereochemistry values of bond angles, bond lengths, and dihedrals angles and the final dimeric NBD model was checked with the program PROCHECK. The N-terminal NBD is depicted as a ribbon diagram in green; and the C-terminal NBD is shown in gold. Functionally important sequence motifs are highlighted in different colors. The Walker A motif appears in red, the Walker B motif blue, the signature motif magenta and the A-loop brown. The modeled ATP is shown as a ball-and-stick model with carbon atoms in black, oxygen red, nitrogen blue and phosphorus magenta. The $Mg^{2+}$ ion is shown as a metallic ball. The conserved residues that are part of the ATP-binding environment are shown in the expanded panel. The Y401 residue of the conserved subdomain, A-loop (shown in brown) stacks against the adenine ring of ATP. The figure was created with GLR software (ww.convent.nci.nih.gov/glr). This figure and caption were taken from [140].

The different sub-domains of NBDs are well characterized; they are made of the Walkers A and B, the ABC signature (also called LSGGQ motif), the C-region, and the A-, H-, Q-, D-loops. Although some modulations are possible according to some polymorphisms, the ATP binding mode is common among ABC transporters. The phosphate moiety of ATP bind to i) the Walkers A and B of one NBD and ii) the signature and D-loop of the other NBD (Figure 15). This moiety also coordinated with one $Mg^{2+}$ cation present in the ATP-binding site. The adenine moiety of ATP makes a $\pi$–$\pi$ stacking interaction with a conserved residue which is almost always tyrosine in the consensus NBD (the C-terminal NBD 2), whereas it can be tyrosine, tryptophan or phenylalanine in the N-terminal (degenerate) NBD 1 domain. Hydrogen bonding and cation–$\pi$ interaction with the adenine base are also likely to contribute to ATP binding in this region[140]. When ATP binds NDBs, the two NBDs are locked in a head-to-tail

arrangement (Figure 15). In some cases, only one ATPase site is catalytically competent (*i.e.*, allowing ATP hydrolysis), *e.g.*, Walker A and B of NDB 2 and signature of the degenerate NBD 1, the other being deficient in ATP hydrolysis[149].

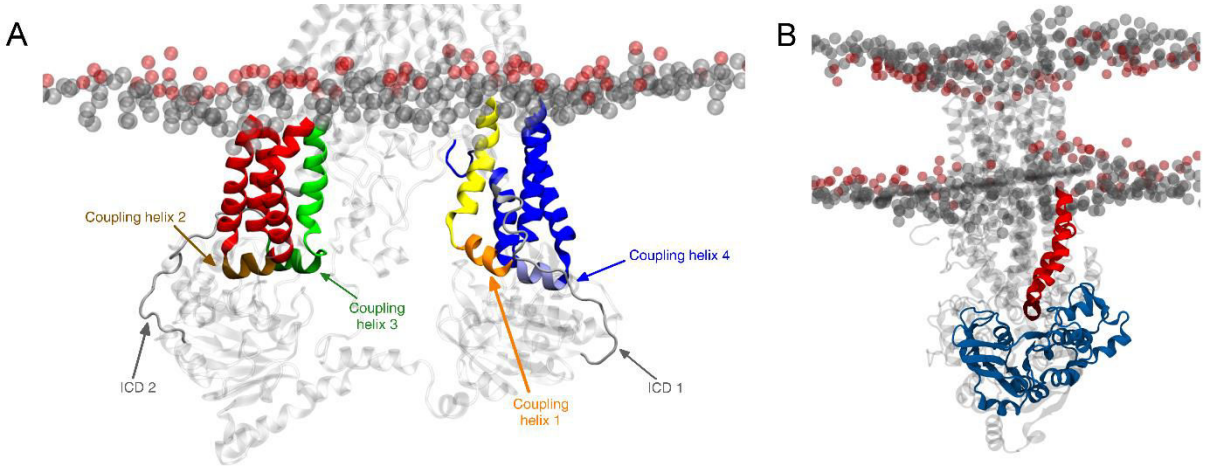



Figure 16: A) Zoom in on ICLs 1 and 2 picturing the connection between NBDs and TMDs in human MRP4 model. B) Side view of the coupling helix 2 and ICL 1 with the ball-and-socket assembly with NBD 1.

The NBDs are connected to the TransMembrane Domains (TMDs) either covalently through the (Intra)Cytoplasmic Domains (ICDs), one in each half of the transporter, or non-covalently to the IntraCellular Loops (ICLs). The ICDs are unstructured external loops, which are tightly attached to the two NBDs, as driven by electrostatic interactions. The ICLs connect the different TMs to each other following the amino acid sequence. They have been described to make a ball-and-socket assembly with the two NBDs (Figure 16). This noncovalent NBD/TMD joint transfers the ATP-hydrolysis energy into the large conformational changes that occur in the TMDs to transport drugs (see below for more details on the mechanism of transport).

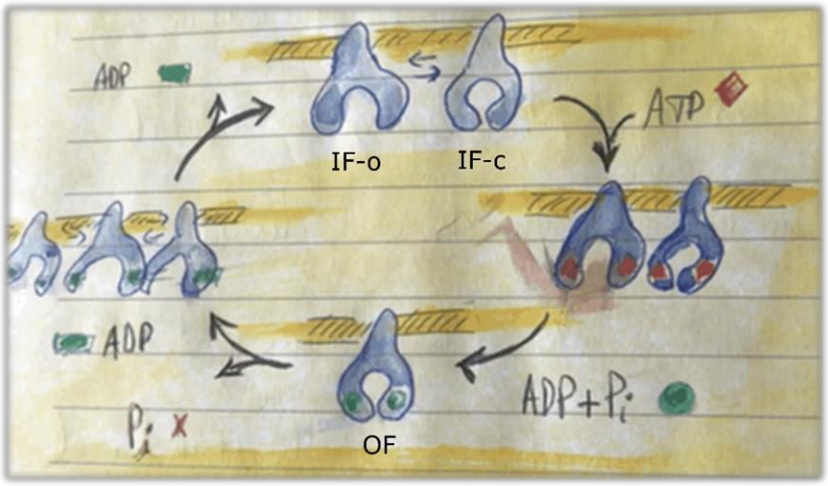
There exist domains which are specific of some ABC transporters. The R (regulatory) domain is unique in CFTR. It links the two halves of the transporter and has been suggested to regulate the channel activity. It is made of about 200 residues including 10 potential phosphorylation sites. In this case, not only ATP binding but also phosphorylation of the R-domain is required for functioning. As observed by targeted-mutagenesis *in vitro* experiments, the unphosphorylated R-domain lies in the protein chamber, obstructing the channel[149]. The phosphorylation releases the R-domain from the channel, thus activating ATPase activity. The binding of the R-domain to the transporter induces intramolecular re-arrangements which are transmitted to the NBD/TMD interface[152]. There is a counterpart of the R-domain in MRPs, which is usually referred as the $L_1$-linker (see Chapter IV for more details with MRP4).

An extra domain (TMD0) made of five transmembrane helices packed together is present in MRP1-3, MRP6 and MRP8-9. It has been described relatively far from the ABC transporter to which it is linked. Its role is still unclear as no protein partner has been reported so far. No effect of mutations in TMD0 have been reported in drug transport. TMD0 is linked to the N-terminal of the MRP by the so-called lasso motif in CFTR, which named $L_0$ in, *e.g.*, MRP1[153]. In MRP4-5 and CFTR (MRP7), the TMD0 domain is absent, although the $L_0$ motif remains present. $L_0$ has appeared crucial for drug translocation through the related ABC transporter[152]. In CFTR, it has been suggested to regulate the channel gating by interacting

with the R-domain ($L_1$ linker)[149]. Despite the completely different functions of CFTR and MRP1, the lasso motif of CFTR and the $L_0$ linker of MRP1 are very similar. Due to its flexibility, it has been poorly characterized so far. It has been partially resolved in both CFTR and MRP1[149,153]. It is made of two helices, one amphiphilic which inserts diagonally into the lipid bilayer, the other one is likely to interact with the core of the protein (TMDs).

#### I.3.3.3. ABC dynamical functioning

Although most ABC structures have been elucidated in their Inward-Facing (IF) open conformer, various other conformers have been identified, mainly by recent cryo-EM studies[154]. It has been shown that the apo-state (conformer bound neither with ATP nor with substrate) of the human P-gp co-exists as an IF-open state and an IF-close state[155]. Outward-Facing (OF) conformers have also been elucidated, *e.g.*, for the bovine MRP1 transporter[156]. Based on these structural data and the experimental conditions used to stabilize these different conformers, various mechanisms for drug translocation have been suggested. More recently, MD simulations, mainly biased MD simulations, have got new insights into these mechanisms[157,158].




Scheme 1: Representation of the transporter cycle of ABC transporters involving ATP hydrolysis and the large conformational changes between IF-o, IF-c and OF conformations.

ATP binding is a mandatory step of the whole transport process through ABCs. ATP hydrolysis into ADP is also a major part of the process (Scheme 1), although this stage depends on the transporter (symmetric or asymmetric ATP hydrolysis exists due to symmetry or asymmetry between the two NBDs). The transport of the xenobiotics from one side to the other side of the membrane requires large conformational changes, which must be thought for proteins embedded in lipid bilayer membranes. In the IF-open, the NBDs are far from each other (Figure 17); this is considered as the resting state. Although the two NBDs can approach to mimic an IF-close conformer by thermal fluctuation, the stabilization of this conformer is mandatory to pursue the process. Its stabilization requires ATP binding into NBD 1 and 2 and a head-to-tail lock (NBD dimerization), see Figure 15. At this stage, large conformational changes occur in the transmembrane domains, during which bundles interchange to lead to the OF-conformer (Figure 17). The energetic landscape of the IF-open → OF and the reverse has been studied on the apo-state of MsbA by non-equilibrium MD simulations, following key collective variables[158]. Although the distance between the two NDBs is an obvious variable, the twisting between the two NBDs has here been shown as a key parameter to trigger these large

conformational changes. This highlighted the importance of the NBD-NBD interface (predominantly electrostatic and highly modified during twisting) and the highly cooperative NBD-TMD interactions. By elucidating the minimum free-energy path, this MD-based study has highlighted several intermediates states, which were connected to confirm an alternative-access process, as also shown by cryo-EM[154].

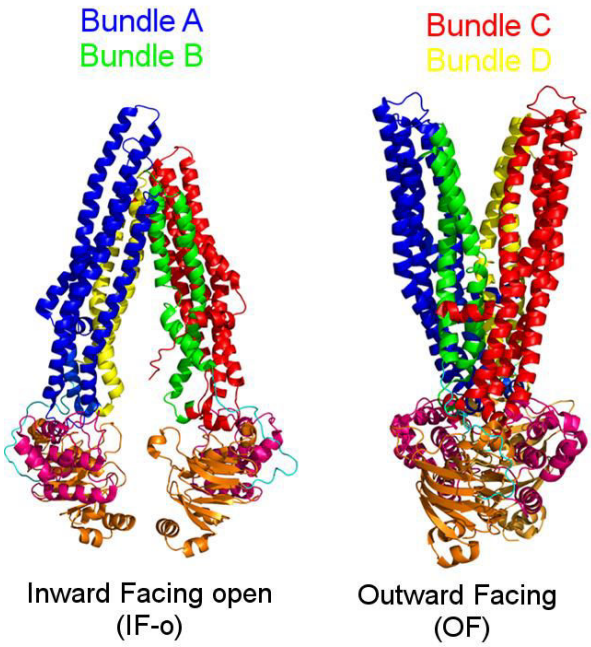



Figure 17: Representation of two PDB structures in two conformations. The IF-o conformer on the left was taken from 3G5U PDB (P-gp from *Mus musuclus*) and the OF was taken from 2HYD PDB (Sav1866 transporter from *Staphylococcus aureus*). This representation shows the exchange of bundles C and D with bundles A and B between the two IF-o and OF conformations.

The underlying question of understanding the mechanism of IF-open → OF large conformational changes is how substrates are translocated and released outside the cell and the organelles. In some cases, the binding of substrates into the protein chamber is not mandatory to provide a basal ATPase activity, it is however somehow coupled to the transport activity. In some cases, it is even thought to be the driving force. Various models have been proposed so far. In Model 1, the xenobiotics goes to the binding site of the IF-conformer with high-affinity, and it is transferred to a lower affinity site in the OF conformer, as driven by the large conformational changes catalyzed by ATP hydrolysis. In Model 2, ATP hydrolysis releases the energy required to break the substrate-binding forces, which triggers the xenobiotic release. In Model 3, recently proposed for MRP1, the xenobiotic is released by the local conformational re-arrangements in the binding site that are coupled to NDB dimerization before ATP hydrolysis occurs[153]. In this case, ATP hydrolysis would only be used to somehow reset the transporter, *i.e.*, to drive it back into the IF-open conformer, allowing a new cycle to occur. Model 3 means, for MRP1 but most probably for other ABC transporters, that both the ATP and the xenobiotic binding are mandatory to initiate the process which will be responsible for translocation. In MRP1, the substrate binding has been shown to drive the TM-bundles closer to each other and lining up the structure to trigger NDB dimerization (prior to ATP binding), this constitutes, in this case, the first step of the whole process.

#### I.3.3.4. Substrate binding, polymorphism and importance to understand functioning

The binding of the substrate into the pocket is a crucial event in its own translocation from inside to outside the cell or the organelles; in this process, the substrate is not passive, but it is an active player. Therefore, a thorough characterization of substrate binding is mandatory to gain into the knowledge of drug efflux by ABC transporters, hence to reach a step forward in the clinical usage of ABC transporters. Key residues in substrate recognition may serve as target to new drugs to limit the impact of ABC dysfunctions.

ATP-binding into the catalytic site has been widely characterized as it can be easily resolved because it only allows a highly specific binding. Conversely, the mode of substrate binding is more delicate to characterize because of high flexibility of the substrate-binding region, also known as the protein chamber, or vestibule. The ABC protein chamber strongly depends on the ABC structure. Therefore, the specificity of substrate-binding is also ABC-dependent.

In CFTR, the protein chamber forms a highly positive electrostatic surface made of polar and charged residues. Some of them have appeared critical for CFTR functioning, namely Lys95, Gln98, Arg303, Ser341, Arg352 and Lys1060[149]. Such a cavity allows translocation of anionic xenobiotics, it is considered as an ATP-gated ion channel. In CFTR, the N-tail has also been suggested to interact with the translocation path, *e.g.*, with the acidic residues Asp47, Glu51, Glu54 and Asp58[159].

P-gp transports hydrophobic and some weakly cationic xenobiotics, as the chamber surface is hydrophobic (Leu65, Ile306, Ile340, Phe343, Phe728, Val982)[160,161] with acidic patches (*e.g.*, Asp188, Glu353, Glu782, Asp997)[153,162].

MRPs transport a broad range of amphiphilic organic xenobiotics, *e.g.*, MRP1 transports organic acids as the chamber is basic, or MRP4 transports amphiphilic compounds. In MRP1, the binding site of the protein chamber can be divided into two regions, namely a positively charged one and a hydrophobic one (P- and H-pockets, respectively), which provides an ideal situation to accept amphiphilic acids (*e.g.*, Leukotriene C4)[153]. In MRP4, hydrophobic and polar residues have been highlighted in substrate recognition in the protein chamber (Phe368, Phe369, Glu374, Arg375, Glu378, Trp995 and Arg998)[163,164], see Chapter IV for more details.

Thorough identification of residues constituting the protein chamber and involved in the recognition of xenobiotics that can be substrates of P-gp and MRPs has already suggested that the binding mode proceeds through hydrogen bonding and $\pi$–$\pi$ stacking interactions. This thorough identification has also suggested that the protein chamber is a multiple allosteric binding sites[165,166], which may allow better understanding multidrug resistance processes and drug-drug interactions related to ABC transporters. Another issue that is partially addressed by the knowledge about the protein chamber is the xenobiotic membrane access. In P-gp, xenobiotics can enter into the chamber directly from the lipid bilayer. Conversely, in MRP1, the access from the lipid bilayer to the translocation path is completely clogged, and only access from the cytoplasm is likely; in other words, conversely to P-gp, the substrate recognition in MRP1 seems not to depend on membrane partitioning.

So far, the identification of key residues in substrate recognition has been highlighted by targeted-mutagenesis. A few theoretical studies have better characterized the protein chamber, however mainly based of docking procedures performed in a static way, *i.e.*, not accounting for the highly dynamic character of the translocation path. Further dynamical

docking studies should enable better understanding of selectivity in transport through ABC and possible dysfunction.

## References


[1] J. Ritter, F. Rod, G. Henderson, Humphrey Rang, Rang & Dale's Pharmacology, 7th edition, Churchill Livingstone, 2011.

[2] P. Košinová, K. Berka, M. Wykes, M. Otyepka, P. Trouillas, Positioning of Antioxidant Quercetin and Its Metabolites in Lipid Bilayer Membranes: Implication for Their Lipid-Peroxidation Inhibition, J. Phys. Chem. B. 116 (2012) 1309–1318. doi:10.1021/jp208731g.

[3] J.L. MacCallum, W.F.D. Bennett, D.P. Tieleman, Distribution of amino acids in a lipid bilayer from computer simulations, Biophys. J. 94 (2008) 3393–3404. doi:10.1529/biophysj.107.112805.

[4] P. Anzenbacher, U.M. Zanger, Metabolism of Drugs and Other Xenobiotics, John Wiley & Sons, 2012.

[5] S.P. Tripathi, A. Bhadauriya, A. Patil, A.T. Sangamwar, Substrate selectivity of human intestinal UDP-glucuronosyltransferases (UGTs): in silico and in vitro insights, Drug Metabolism Reviews. 45 (2013) 231–252. doi:10.3109/03602532.2013.767345.

[6] J.M. Mato, M.L. Martínez-Chantar, S.C. Lu, S-adenosylmethionine metabolism and liver disease, Ann Hepatol. 12 (2013) 183–189.

[7] L. Shi, B.P. Tu, Acetyl-CoA and the Regulation of Metabolism: Mechanisms and Consequences, Curr Opin Cell Biol. 33 (2015) 125–131. doi:10.1016/j.ceb.2015.02.003.

[8] A. Dickey, R. Faller, Examining the Contributions of Lipid Shape and Headgroup Charge on Bilayer Behavior, Biophysical Journal. 95 (2008) 2636–2646. doi:10.1529/biophysj.107.128074.

[9] S. Pöyry, I. Vattulainen, Role of charged lipids in membrane structures - Insight given by simulations, Biochim. Biophys. Acta. 1858 (2016) 2322–2333. doi:10.1016/j.bbamem.2016.03.016.

[10] T. Harayama, H. Riezman, Understanding the diversity of membrane lipid composition, Nat. Rev. Mol. Cell Biol. 19 (2018) 281–296. doi:10.1038/nrm.2017.138.

[11] E. Goormaghtigh, P. Chatelain, J. Caspers, J.M. Ruysschaert, Evidence of a specific complex between adriamycin and negatively-charged phospholipids, Biochim. Biophys. Acta. 597 (1980) 1–14.

[12] T.M. Fong, M.G. McNamee, Stabilization of acetylcholine receptor secondary structure by cholesterol and negatively charged phospholipids in membranes, Biochemistry. 26 (1987) 3871–3880.

[13] P.W. van Dijck, Negatively charged phospholipids and their position in the cholesterol affinity sequence, Biochim. Biophys. Acta. 555 (1979) 89–101.

[14] R. Ernst, C.S. Ejsing, B. Antonny, Homeoviscous Adaptation and the Regulation of Membrane Lipids, J. Mol. Biol. 428 (2016) 4776–4791. doi:10.1016/j.jmb.2016.08.013.

[15] A.K. Menon, V.L. Stevens, Phosphatidylethanolamine is the donor of the ethanolamine residue linking a glycosylphosphatidylinositol anchor to protein, J. Biol. Chem. 267 (1992) 15277–15280.

[16] J.L. Gallop, A. Walrant, L.C. Cantley, M.W. Kirschner, Phosphoinositides and membrane curvature switch the mode of actin polymerization via selective recruitment of toca-1 and Snx9, Proc Natl Acad Sci U S A. 110 (2013) 7193–7198. doi:10.1073/pnas.1305286110.

[17] G. Di Paolo, P. De Camilli, Phosphoinositides in cell regulation and membrane dynamics, Nature. 443 (2006) 651–657. doi:10.1038/nature05185.

[18] Y. Yoon, P.J. Lee, S. Kurilova, W. Cho, In Situ Quantitative Imaging of Cellular Lipids Using Molecular Sensors, Nat Chem. 3 (2011) 868–874. doi:10.1038/nchem.1163.

[19] J. Mett, T. Hartmann, M.O.W. Grimm, Chapter 13 - The Effects of Glycerophospholipids and Fatty Acids on APP Processing: Implications for Alzheimer's Disease, in: R.R. Watson, F. De Meester (Eds.), Handbook of Lipids in Human Function, AOCS Press, 2016: pp. 377–421. doi:10.1016/B978-1-63067-036-8.00013-5.

[20] R.H. Houtkooper, F.M. Vaz, Cardiolipin, the heart of mitochondrial metabolism, Cell. Mol. Life Sci. 65 (2008) 2493–2506. doi:10.1007/s00018-008-8030-5.

[21] F.L. Hoch, Cardiolipins and biomembrane function, Biochim. Biophys. Acta. 1113 (1992) 71–133.

[22] G. Paradies, V. Paradies, V. De Benedictis, F.M. Ruggiero, G. Petrosillo, Functional role of cardiolipin in mitochondrial bioenergetics, Biochimica et Biophysica Acta (BBA) - Bioenergetics. 1837 (2014) 408–417. doi:10.1016/j.bbabio.2013.10.006.

[23] N. Ikon, R.O. Ryan, Barth syndrome: connecting cardiolipin to cardiomyopathy, Lipids. 52 (2017) 99–108. doi:10.1007/s11745-016-4229-7.

[24] Y.A. Hannun, L.M. Obeid, The Ceramide-centric universe of lipid-mediated cell regulation: stress encounters of the lipid kind, J. Biol. Chem. 277 (2002) 25847–25850. doi:10.1074/jbc.R200008200.

[25] A.J. Snider, K.A.O. Gandy, L.M. Obeid, Sphingosine Kinase: Role in Regulation of Bioactive Sphingolipid Mediators in Inflammation, Biochimie. 92 (2010) 707–715. doi:10.1016/j.biochi.2010.02.008.

[26] S. Spiegel, S. Milstien, Sphingosine 1-Phosphate, a Key Cell Signaling Molecule, J. Biol. Chem. 277 (2002) 25851–25854. doi:10.1074/jbc.R200007200.

[27] F.-X. Contreras, A.M. Ernst, P. Haberkant, P. Björkholm, E. Lindahl, B. Gönen, C. Tischer, A. Elofsson, G. von Heijne, C. Thiele, R. Pepperkok, F. Wieland, B. Brügger, Molecular recognition of a single sphingolipid species by a protein's transmembrane domain, Nature. 481 (2012) 525–529. doi:10.1038/nature10742.

[28] E. Posse de Chaves, S. Sipione, Sphingolipids and gangliosides of the nervous system in membrane function and dysfunction, FEBS Letters. 584 (2010) 1748–1759. doi:10.1016/j.febslet.2009.12.010.

[29] B. Oskouian, J.D. Saba, Cancer Treatment Strategies Targeting Sphingolipid Metabolism, Adv Exp Med Biol. 688 (2010) 185–205. doi:10.1007/978-1-4419-6741-1_13.

[30] J. Vogel, G. Bendas, U. Bakowsky, G. Hummel, R.R. Schmidt, U. Kettmann, U. Rothe, The role of glycolipids in mediating cell adhesion: a flow chamber study1This paper is dedicated to Professor H.W. Meyer (University of Jena) on the occasion of his 65th birthday.1,

Biochimica et Biophysica Acta (BBA) - Biomembranes. 1372 (1998) 205–215. doi:10.1016/S0005-2736(98)00058-3.

[31] V.F. Vartabedian, P.B. Savage, L. Teyton, The processing and presentation of lipids and glycolipids to the immune system, Immunol. Rev. 272 (2016) 109–119. doi:10.1111/imr.12431.

[32] M.J.R. Dawkins, J.F. Stevens, Fatty Acid Composition of Triglycerides from Adipose Tissue, Nature. 209 (1966) 1145–1146. doi:10.1038/2091145a0.

[33] K.E.. Harchaoui, M.. Visser, J.J.. Kastelein, E.. Stroes, G.. Dallinga-Thie, Triglycerides and Cardiovascular Risk, Curr Cardiol Rev. 5 (2009) 216–222. doi:10.2174/157340309788970315.

[34] J.S. Lee, P.-Y. Chang, Y. Zhang, J.R. Kizer, L.G. Best, B.V. Howard, Triglyceride and HDL-C Dyslipidemia and Risks of Coronary Heart Disease and Ischemic Stroke by Glycemic Dysregulation Status: The Strong Heart Study, Diabetes Care. 40 (2017) 529–537. doi:10.2337/dc16-1958.

[35] W. Cl, van der S. D, H. Js, Large influence of cholesterol on solute partitioning into lipid membranes., J Am Chem Soc. 134 (2012) 5351–5361. doi:10.1021/ja211929h.

[36] M. Luxnat, H.J. Galla, Partition of chlorpromazine into lipid bilayer membranes: the effect of membrane structure and composition, Biochim. Biophys. Acta. 856 (1986) 274–282. doi:10.1016/0005-2736(86)90037-4.

[37] G.W. Feigenson, Phase diagrams and lipid domains in multicomponent lipid bilayer mixtures, Biochim. Biophys. Acta. 1788 (2009) 47–52. doi:10.1016/j.bbamem.2008.08.014.

[38] K. Simons, E. Ikonen, Functional rafts in cell membranes, Nature. 387 (1997) 569–572. doi:10.1038/42408.

[39] D.A. Brown, E. London, Structure and Function of Sphingolipid- and Cholesterol-rich Membrane Rafts, J. Biol. Chem. 275 (2000) 17221–17224. doi:10.1074/jbc.R000005200.

[40] K. Simons, W.L.C. Vaz, Model systems, lipid rafts, and cell membranes, Annu Rev Biophys Biomol Struct. 33 (2004) 269–295. doi:10.1146/annurev.biophys.32.110601.141803.

[41] D. Lingwood, K. Simons, Lipid rafts as a membrane-organizing principle, Science. 327 (2010) 46–50. doi:10.1126/science.1174621.

[42] K. Simons, M.J. Gerl, Revitalizing membrane rafts: new tools and insights, Nat. Rev. Mol. Cell Biol. 11 (2010) 688–699. doi:10.1038/nrm2977.

[43] E.C. Jury, F. Flores-Borja, P.S. Kabouridis, Lipid rafts in T cell signalling and disease, Semin Cell Dev Biol. 18 (2007) 608–615. doi:10.1016/j.semcdb.2007.08.002.

[44] B. Diaz-Rohrer, K.R. Levental, I. Levental, Rafting through traffic: Membrane domains in cellular logistics, Biochimica et Biophysica Acta (BBA) - Biomembranes. 1838 (2014) 3003–3013. doi:10.1016/j.bbamem.2014.07.029.

[45] P.S. Niemelä, M.S. Miettinen, L. Monticelli, H. Hammaren, P. Bjelkmar, T. Murtola, E. Lindahl, I. Vattulainen, Membrane proteins diffuse as dynamic complexes with lipids, J. Am. Chem. Soc. 132 (2010) 7574–7575. doi:10.1021/ja101481b.

[46] H.J. Sharpe, T.J. Stevens, S. Munro, A comprehensive comparison of transmembrane domains reveals organelle-specific properties, Cell. 142 (2010) 158–169. doi:10.1016/j.cell.2010.05.037.

[47] M. Javanainen, H. Martinez-Seara, I. Vattulainen, Nanoscale Membrane Domain Formation Driven by Cholesterol, Sci Rep. 7 (2017) 1143. doi:10.1038/s41598-017-01247-9.

[48] G.E. Herman, L. Kratz, Disorders of sterol synthesis: beyond Smith-Lemli-Opitz syndrome, Am J Med Genet C Semin Med Genet. 160C (2012) 301–321. doi:10.1002/ajmg.c.31340.

[49] C.-J. Lin, C.-K. Lai, M.-C. Kao, L.-T. Wu, U.-G. Lo, L.-C. Lin, Y.-A. Chen, H. Lin, J.-T. Hsieh, C.-H. Lai, C.-D. Lin, Impact of cholesterol on disease progression, Biomedicine (Taipei). 5 (2015). doi:10.7603/s40681-015-0007-8.

[50] I. Eriksson, S. Nath, P. Bornefall, A.M.V. Giraldo, K. Öllinger, Impact of high cholesterol in a Parkinson's disease model: Prevention of lysosomal leakage versus stimulation of α-synuclein aggregation, European Journal of Cell Biology. 96 (2017) 99–109. doi:10.1016/j.ejcb.2017.01.002.

[51] J.F. Nagle, S. Tristram-Nagle, Structure of lipid bilayers, Biochim. Biophys. Acta. 1469 (2000) 159–195. doi:10.1016/S0304-4157(00)00016-2.

[52] N. Kucerka, S. Tristram-Nagle, J.F. Nagle, Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains, J. Membr. Biol. 208 (2005) 193–202. doi:10.1007/s00232-005-7006-8.

[53] N. Kučerka, J.F. Nagle, J.N. Sachs, S.E. Feller, J. Pencer, A. Jackson, J. Katsaras, Lipid Bilayer Structure Determined by the Simultaneous Analysis of Neutron and X-Ray Scattering Data, Biophys J. 95 (2008) 2356–2367. doi:10.1529/biophysj.108.132662.

[54] K. Mitra, I. Ubarretxena-Belandia, T. Taguchi, G. Warren, D.M. Engelman, Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol, PNAS. 101 (2004) 4083–4088. doi:10.1073/pnas.0307332101.

[55] J.D. Perlmutter, J.N. Sachs, Experimental verification of lipid bilayer structure through multi-scale modeling, Biochimica et Biophysica Acta (BBA) - Biomembranes. 1788 (2009) 2284–2290. doi:10.1016/j.bbamem.2009.07.006.

[56] J.P.M. Jämbeck, A.P. Lyubartsev, An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes, J. Chem. Theory Comput. 8 (2012) 2938–2948. doi:10.1021/ct300342n.

[57] W.J. Allen, J.A. Lemkul, D.R. Bevan, GridMAT-MD: A grid-based membrane analysis tool for use with molecular dynamics, Journal of Computational Chemistry. 30 (2009) 1952–1958. doi:10.1002/jcc.21172.

[58] S. Buchoux, FATSLiM: a fast and robust software to analyze MD simulations of membranes, Bioinformatics. 33 (2017) 133–134. doi:10.1093/bioinformatics/btw563.

[59] G. Lukat, J. Krüger, B. Sommer, APL@Voro: a Voronoi-based membrane analysis tool for GROMACS trajectories, J Chem Inf Model. 53 (2013) 2908–2925. doi:10.1021/ci400172g.

[60] N. Kučerka, M.-P. Nieh, J. Katsaras, Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature, Biochim. Biophys. Acta. 1808 (2011) 2761–2771. doi:10.1016/j.bbamem.2011.07.022.

[61] R. Macháň, M. Hof, Lipid diffusion in planar membranes investigated by fluorescence correlation spectroscopy, Biochimica et Biophysica Acta (BBA) - Biomembranes. 1798 (2010) 1377–1391. doi:10.1016/j.bbamem.2010.02.014.

[62] M. Allen, D. Tildesley, Computer Simulation of Liquids, Oxford University Press, 1987. doi:10.1007/BF01532020.

[63] M.M. Kamal, D. Mills, M. Grzybek, J. Howard, Measurement of the membrane curvature preference of phospholipids reveals only weak coupling between lipid shape and leaflet curvature, PNAS. 106 (2009) 22245–22250. doi:10.1073/pnas.0907354106.

[64] I.K. Jarsch, F. Daste, J.L. Gallop, Membrane curvature in cell biology: An integration of molecular mechanisms, J Cell Biol. 214 (2016) 375–387. doi:10.1083/jcb.201604003.

[65] J. Bigay, B. Antonny, Curvature, Lipid Packing, and Electrostatics of Membrane Organelles: Defining Cellular Territories in Determining Specificity, Developmental Cell. 23 (2012) 886–895. doi:10.1016/j.devcel.2012.10.009.

[66] J.N. Israelachvili, D.J. Mitchell, B.W. Ninham, Theory of self-assembly of hydrocarbon amphiphiles into micelles and bilayers, J. Chem. Soc., Faraday Trans. 2. 72 (1976) 1525–1568. doi:10.1039/F29767201525.

[67] H. Koldsø, D. Shorthouse, J. Hélie, M.S.P. Sansom, Lipid Clustering Correlates with Membrane Curvature as Revealed by Molecular Simulations of Complex Lipid Bilayers, PLOS Computational Biology. 10 (2014) e1003911. doi:10.1371/journal.pcbi.1003911.

[68] R.L. Thurmond, G. Lindblom, M.F. Brown, Curvature, order, and dynamics of lipid hexagonal phases studied by deuterium NMR spectroscopy, Biochemistry. 32 (1993) 5394–5410. doi:10.1021/bi00071a015.

[69] S.H. Alley, O. Ces, M. Barahona, R.H. Templer, X-ray diffraction measurement of the monolayer spontaneous curvature of dioleoylphosphatidylglycerol, Chem. Phys. Lipids. 154 (2008) 64–67. doi:10.1016/j.chemphyslip.2008.03.007.

[70] B. Kollmitzer, P. Heftberger, M. Rappolt, G. Pabst, Monolayer spontaneous curvature of raft-forming membrane lipids, Soft Matter. 9 (2013) 10877–10884. doi:10.1039/C3SM51829A.

[71] T. Harayama, M. Eto, H. Shindou, Y. Kita, E. Otsubo, D. Hishikawa, S. Ishii, K. Sakimura, M. Mishina, T. Shimizu, Lysophospholipid acyltransferases mediate phosphatidylcholine diversification to achieve the physical properties required in vivo, Cell Metab. 20 (2014) 295–305. doi:10.1016/j.cmet.2014.05.019.

[72] A.M. El-Badry, R. Graf, P.-A. Clavien, Omega 3 - Omega 6: What is right for the liver?, J. Hepatol. 47 (2007) 718–725. doi:10.1016/j.jhep.2007.08.005.

[73] C. Leonard, L. Conrard, M. Guthmann, H. Pollet, M. Carquin, C. Vermylen, P. Gailly, P.V.D. Smissen, M.P. Mingeot-Leclercq, D. Tyteca, Contribution of plasma membrane lipid domains to red blood cell (re)shaping, Scientific Reports. 7 (2017) 4264. doi:10.1038/s41598-017-04388-z.

[74] W.A. Devane, L. Hanus, A. Breuer, R.G. Pertwee, L.A. Stevenson, G. Griffin, D. Gibson, A. Mandelbaum, A. Etinger, R. Mechoulam, Isolation and structure of a brain constituent that binds to the cannabinoid receptor, Science. 258 (1992) 1946–1949. doi:10.1126/science.1470919.

[75] A.J. Verkleij, R.F.A. Zwaal, B. Roelofsen, P. Comfurius, D. Kastelijn, L.L.M. van Deenen, The asymmetric distribution of phospholipids in the human red cell membrane. A combined study using phospholipases and freeze-etch electron microscopy, Biochimica et Biophysica Acta (BBA) - Biomembranes. 323 (1973) 178–193. doi:10.1016/0005-2736(73)90143-0.

[76] J.T. Dodge, G.B. Phillips, Composition of phospholipids and of phospholipid fatty acids and aldehydes in human red cells, J. Lipid Res. 8 (1967) 667–675.

[77] M. Paloncýová, R.H. DeVane, B.P. Murch, K. Berka, M. Otyepka, Rationalization of reduced penetration of drugs through ceramide gel phase membrane, Langmuir. 30 (2014) 13942–13948. doi:10.1021/la503289v.

[78] G.K. Menon, G.W. Cleary, M.E. Lane, The structure and function of the stratum corneum, Int J Pharm. 435 (2012) 3–9. doi:10.1016/j.ijpharm.2012.06.005.

[79] G. van Meer, A.I.P.M. de Kroon, Lipid map of the mammalian cell, J Cell Sci. 124 (2011) 5–8. doi:10.1242/jcs.071233.

[80] G. van Meer, D.R. Voelker, G.W. Feigenson, Membrane lipids: where they are and how they behave, Nat. Rev. Mol. Cell Biol. 9 (2008) 112–124. doi:10.1038/nrm2330.

[81] F.P. Guengerich, Human Cytochrome P450 Enzymes, in: Cytochrome P450, Springer, Cham, 2015: pp. 523–785. doi:10.1007/978-3-319-12108-6_9.

[82] S.D. Black, Membrane topology of the mammalian P450 cytochromes, FASEB J. 6 (1992) 680–685. doi:10.1096/fasebj.6.2.1537456.

[83] P.A. Williams, J. Cosme, V. Sridhar, E.F. Johnson, D.E. McRee, Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity, Mol. Cell. 5 (2000) 121–131. doi: 10.1016/S1097-2765(00)80408-6.

[84] E.E. Scott, Y.Q. He, J.R. Halpert, Substrate Routes to the Buried Active Site May Vary among Cytochromes P450: Mutagenesis of the F−G Region in P450 2B1, Chem. Res. Toxicol. 15 (2002) 1407–1413. doi:10.1021/tx020057u.

[85] M. Seliskar, D. Rozman, Mammalian cytochromes P450--importance of tissue specificity, Biochim. Biophys. Acta. 1770 (2007) 458–466. doi:10.1016/j.bbagen.2006.09.016.

[86] J. Kirchmair, A. Howlett, J.E. Peironcely, D.S. Murrell, M.J. Williamson, S.E. Adams, T. Hankemeier, L. van Buren, G. Duchateau, W. Klaffke, R.C. Glen, How do metabolites differ from their parent molecules and how are they excreted?, J Chem Inf Model. 53 (2013) 354–367. doi:10.1021/ci300487z.

[87] K. Berka, M. Paloncýová, P. Anzenbacher, M. Otyepka, Behavior of Human Cytochromes P450 on Lipid Membranes, J. Phys. Chem. B. 117 (2013) 11556–11564. doi:10.1021/jp4059559.

[88] M. Paloncýová, K. Berka, M. Otyepka, Molecular insight into affinities of drugs and their metabolites to lipid bilayers, J Phys Chem B. 117 (2013) 2403–2410. doi:10.1021/jp311802x.

[89] T. Hendrychova, K. Berka, V. Navratilova, P. Anzenbacher, M. Otyepka, Dynamics and hydration of the active sites of mammalian cytochromes P450 probed by molecular dynamics simulations, Curr. Drug Metab. 13 (2012) 177–189. doi:10.2174/138920012798918408.

[90] J.L. Baylon, I.L. Lenov, S.G. Sligar, E. Tajkhorshid, Characterizing the membrane-bound state of cytochrome P450 3A4: structure, depth of insertion, and orientation, J. Am. Chem. Soc. 135 (2013) 8542–8551. doi:10.1021/ja4003525.

[91] F.P. Guengerich, D.H. Kim, M. Iwasaki, Role of human cytochrome P-450 IIE1 in the oxidation of many low molecular weight cancer suspects, Chem. Res. Toxicol. 4 (1991) 168–179. doi:10.1021/tx00020a008.

[92] M. Šrejber, V. Navrátilová, M. Paloncýová, V. Bazgier, K. Berka, P. Anzenbacher, M. Otyepka, Membrane-attached mammalian cytochromes P450: An overview of the membrane's effects on structure, drug binding, and interactions with redox partners, Journal of Inorganic Biochemistry. 183 (2018) 117–136. doi:10.1016/j.jinorgbio.2018.03.002.

[93] P.R. Wilderman, J.R. Halpert, Plasticity of CYP2B Enzymes: Structural and Solution Biophysical Methods, Curr Drug Metab. 13 (2012) 167–176. doi:10.2174/138920012798918417.

[94] J. Hritz, A. de Ruiter, C. Oostenbrink, Impact of Plasticity and Flexibility on Docking Results for Cytochrome P450 2D6: A Combined Approach of Molecular Dynamics and Ligand Docking, J. Med. Chem. 51 (2008) 7469–7477. doi:10.1021/jm801005m.

[95] C. Hayes, D. Ansbro, M. Kontoyianni, Elucidating Substrate Promiscuity in the Human Cytochrome 3A4, J. Chem. Inf. Model. 54 (2014) 857–869. doi:10.1021/ci4006782.

[96] F.P. Guengerich, Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity, Chem. Res. Toxicol. 14 (2001) 611–650. doi:10.1021/tx0002583.

[97] C.F. Samer, K.I. Lorenzini, V. Rollason, Y. Daali, J.A. Desmeules, Applications of CYP450 Testing in the Clinical Setting, Mol Diagn Ther. 17 (2013) 165–184. doi:10.1007/s40291-013-0028-5.

[98] F.P. Guengerich, Cytochrome p450 and chemical toxicology, Chem. Res. Toxicol. 21 (2008) 70–83. doi:10.1021/tx700079z.

[99] Y.J. He, J. Brockmöller, H. Schmidt, I. Roots, J. Kirchheiner, CYP2D6 ultrarapid metabolism and morphine/codeine ratios in blood: was it codeine or heroin?, J Anal Toxicol. 32 (2008) 178–182. doi:10.1093/jat/32.2.178.

[100] D. Van Booven, S. Marsh, H. McLeod, M.W. Carrillo, K. Sangkuhl, T.E. Klein, R.B. Altman, Cytochrome P450 2C9-CYP2C9, Pharmacogenet Genomics. 20 (2010) 277–281. doi:10.1097/FPC.0b013e3283349e84.

[101] N. Miroševic Skvrce, N. Božina, L. Zibar, I. Barišic, L. Pejnovic, V. Macolic Šarinic, CYP2C9 and ABCG2 polymorphisms as risk factors for developing adverse drug reactions

in renal transplant patients taking fluvastatin: a case-control study, Pharmacogenomics. 14 (2013) 1419–1431. doi:10.2217/pgs.13.135.

[102] P.I. Mackenzie, K.W. Bock, B. Burchell, C. Guillemette, S. Ikushiro, T. Iyanagi, J.O. Miners, I.S. Owens, D.W. Nebert, Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily, Pharmacogenet. Genomics. 15 (2005) 677–685. doi:10.1097/01.fpc.0000173483.13689.56.

[103] Y. Ishii, A. Nurrochmad, H. Yamada, Modulation of UDP-glucuronosyltransferase activity by endogenous compounds, Drug Metab. Pharmacokinet. 25 (2010) 134–148. doi:10.2133/dmpk.25.134.

[104] M.J. Miley, A.K. Zielinska, J.E. Keenan, S.M. Bratton, A. Radominska-Pandya, M.R. Redinbo, Crystal structure of the cofactor-binding domain of the human phase II drug-metabolism enzyme UDP-glucuronosyltransferase 2B7, J. Mol. Biol. 369 (2007) 498–511. doi:10.1016/j.jmb.2007.03.066.

[105] L. Laakkonen, M. Finel, A molecular model of the human UDP-glucuronosyltransferase 1A1, its membrane orientation, and the interactions between different parts of the enzyme, Mol. Pharmacol. 77 (2010) 931–939. doi:10.1124/mol.109.063289.

[106] R. Fujiwara, M. Nakajima, T. Yamamoto, H. Nagao, T. Yokoi, *In silico* and *in vitro* approaches to elucidate the thermal stability of human UDP-glucuronosyltransferase (UGT) 1A9, Drug Metab. Pharmacokinet. 24 (2009) 235–244. doi:10.2133/dmpk.24.235.

[107] S.P. Tripathi, R. Prajapati, N. Verma, A.T. Sangamwar, Predicting substrate selectivity between UGT1A9 and UGT1A10 using molecular modelling and molecular dynamics approach, Molecular Simulation. 42 (2016) 270–288. doi:10.1080/08927022.2015.1044451.

[108] C.W. Locuson, T.S. Tracy, Comparative modelling of the human UDP-glucuronosyltransferases: Insights into structure and mechanism, Xenobiotica. 37 (2007) 155–168. doi:10.1080/00498250601129109.

[109] K.M. Giacomini, S.-M. Huang, Transporters in Drug Development and Clinical Pharmacology, Clinical Pharmacology & Therapeutics. 94 (2013) 3–9. doi:10.1038/clpt.2013.86.

[110] K.M. Giacomini, S.-M. Huang, D.J. Tweedie, L.Z. Benet, K.L.R. Brouwer, X. Chu, A. Dahlin, R. Evers, V. Fischer, K.M. Hillgren, K.A. Hoffmaster, T. Ishikawa, D. Keppler, R.B. Kim, C.A. Lee, M. Niemi, J.W. Polli, Y. Sugiyama, P.W. Swaan, J.A. Ware, S.H. Wright, S.W. Yee, M.J. Zamek-Gliszczynski, L. Zhang, Membrane transporters in drug development, Nat Rev Drug Discov. 9 (2010) 215–236. doi:10.1038/nrd3028.

[111] M.H. Saier, C.V. Tran, R.D. Barabote, TCDB: the Transporter Classification Database for membrane transport protein analyses and information, Nucleic Acids Res. 34 (2006) D181–D186. doi:10.1093/nar/gkj001.

[112] M.H. Saier, V.S. Reddy, D.G. Tamang, Å. Västermark, The Transporter Classification Database, Nucl. Acids Res. 42 (2014) D251–D258. doi:10.1093/nar/gkt1097.

[113] A. Schlessinger, S.W. Yee, A. Sali, K.M. Giacomini, SLC classification: an update, Clin. Pharmacol. Ther. 94 (2013) 19–23. doi:10.1038/clpt.2013.73.

[114] A. Schlessinger, P. Matsson, J.E. Shima, U. Pieper, S.W. Yee, L. Kelly, L. Apeltsin, R.M. Stroud, T.E. Ferrin, K.M. Giacomini, A. Sali, Comparison of human solute carriers, Protein Sci. 19 (2010) 412–428. doi:10.1002/pro.320.

[115] A. Schlessinger, N. Khuri, K.M. Giacomini, A. Sali, Molecular modeling and ligand docking for solute carrier (SLC) transporters, Curr Top Med Chem. 13 (2013) 843–856. doi:10.2174/1568026611313070007.

[116] M.A. Hediger, B. Clémençon, R.E. Burrier, E.A. Bruford, The ABCs of membrane transporters in health and disease (SLC series): introduction, Mol. Aspects Med. 34 (2013) 95–107. doi:10.1016/j.mam.2012.12.009.

[117] S.-M. Huang, L. Zhang, K.M. Giacomini, The International Transporter Consortium: A Collaborative Group of Scientists From Academia, Industry, and the FDA, Clinical Pharmacology & Therapeutics. 87 (2010) 32–36. doi:10.1038/clpt.2009.236.

[118] J. Li, P.-C. Wen, M. Moradi, E. Tajkhorshid, Computational characterization of structural dynamics underlying function in active membrane transporters, Current Opinion in Structural Biology. 31 (2015) 96–105. doi:10.1016/j.sbi.2015.04.001.

[119] D. Dwyer, Molecular Modeling and Molecular Dynamics Simulations of Membrane Transporter Proteins, in: Q. Yan, Q.Y. MD (Eds.), Membrane Transporters, Humana Press, 2003: pp. 335–350. doi:10.1385/1-59259-387-9:335.

[120] F. Khalili-Araghi, J. Gumbart, P.-C. Wen, M. Sotomayor, E. Tajkhorshid, K. Schulten, Molecular dynamics simulations of membrane channels and transporters, Curr. Opin. Struct. Biol. 19 (2009) 128–137. doi:10.1016/j.sbi.2009.02.011.

[121] E.M. Quistgaard, C. Löw, F. Guettou, P. Nordlund, Understanding transport by the major facilitator superfamily (MFS): structures pave the way, Nat Rev Mol Cell Biol. 17 (2016) 123–132. doi:10.1038/nrm.2015.25.

[122] E.M. Quistgaard, C. Löw, P. Moberg, L. Trésaugues, P. Nordlund, Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters, Nat Struct Mol Biol. 20 (2013) 766–768. doi:10.1038/nsmb.2569.

[123] L. Sun, X. Zeng, C. Yan, X. Sun, X. Gong, Y. Rao, N. Yan, Crystal structure of a bacterial homologue of glucose transporters GLUT1-4, Nature. 490 (2012) 361–366. doi:10.1038/nature11524.

[124] G. Wisedchaisri, M.-S. Park, M.G. Iadanza, H. Zheng, T. Gonen, Proton-coupled sugar transport in the prototypical major facilitator superfamily protein XylE, Nat Commun. 5 (2014) 4521. doi:10.1038/ncomms5521.

[125] D. Deng, C. Xu, P. Sun, J. Wu, C. Yan, M. Hu, N. Yan, Crystal structure of the human glucose transporter GLUT1, Nature. 510 (2014) 121–125. doi:10.1038/nature13306.

[126] M.-S. Park, Molecular Dynamics Simulations of the Human Glucose Transporter GLUT1, PLoS ONE. 10 (2015) e0125361. doi:10.1371/journal.pone.0125361.

[127] M.Ø. Jensen, Y. Yin, E. Tajkhorshid, K. Schulten, Sugar transport across lactose permease probed by steered molecular dynamics, Biophys. J. 93 (2007) 92–102. doi:10.1529/biophysj.107.103994.

[128] Y. Yin, M.Ø. Jensen, E. Tajkhorshid, K. Schulten, Sugar binding and protein conformational changes in lactose permease, Biophys. J. 91 (2006) 3972–3985. doi:10.1529/biophysj.106.085993.

[129] M. Andersson, A.-N. Bondar, J.A. Freites, D.J. Tobias, H.R. Kaback, S.H. White, Proton-coupled dynamics in lactose permease, Structure. 20 (2012) 1893–1904. doi:10.1016/j.str.2012.08.021.

[130] M. Moradi, G. Enkavi, E. Tajkhorshid, Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate antiporter, Nat Commun. 6 (2015) 8393. doi:10.1038/ncomms9393.

[131] D. Deng, P. Sun, C. Yan, M. Ke, X. Jiang, L. Xiong, W. Ren, K. Hirata, M. Yamamoto, S. Fan, N. Yan, Molecular basis of ligand recognition and transport by glucose transporters, Nature. 526 (2015) 391–396. doi:10.1038/nature14655.

[132] J. Holyoake, M.S.P. Sansom, Conformational change in an MFS protein: MD simulations of LacY, Structure. 15 (2007) 873–884. doi:10.1016/j.str.2007.06.004.

[133] D. Wang, G.A. Voth, Proton Transport Pathway in the ClC Cl−/H+ Antiporter, Biophysical Journal. 97 (2009) 121–131. doi:10.1016/j.bpj.2009.04.038.

[134] A.-N. Bondar, C. del Val, S.H. White, Rhomboid protease dynamics and lipid interactions, Structure. 17 (2009) 395–405. doi:10.1016/j.str.2008.12.017.

[135] F. Gruswitz, S. Chaudhary, J.D. Ho, A. Schlessinger, B. Pezeshki, C.-M. Ho, A. Sali, C.M. Westhoff, R.M. Stroud, Function of human Rh based on structure of RhCG at 2.1 Å, PNAS. 107 (2010) 9638–9643. doi:10.1073/pnas.1003587107.

[136] S. Doki, H.E. Kato, N. Solcan, M. Iwaki, M. Koyama, M. Hattori, N. Iwase, T. Tsukazaki, Y. Sugita, H. Kandori, S. Newstead, R. Ishitani, O. Nureki, Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT, PNAS. 110 (2013) 11343–11348. doi:10.1073/pnas.1301079110.

[137] I.F. Tsigelny, D. Kovalskyy, V.L. Kouznetsova, O. Balinskyi, Y. Sharikov, V. Bhatnagar, S.K. Nigam, Conformational Changes of the Multispecific Transporter Organic Anion Transporter 1 (OAT1/SLC22A6) Suggests a Molecular Mechanism for Initial Stages of Drug and Metabolite Transport, Cell Biochem Biophys. 61 (2011) 251–259. doi:10.1007/s12013-011-9191-7.

[138] D. Meredith, R.A. Price, Molecular Modeling of PepT1 — Towards a Structure, J Membrane Biol. 213 (2007) 79–88. doi:10.1007/s00232-006-0876-6.

[139] Y.M. Leung, D.A. Holdbrook, T.J. Piggot, S. Khalid, The NorM MATE transporter from N. gonorrhoeae: insights into drug and ion binding from atomistic molecular dynamics simulations, Biophys. J. 107 (2014) 460–468. doi:10.1016/j.bpj.2014.06.005.

[140] S.V. Ambudkar, I.-W. Kim, D. Xia, Z.E. Sauna, The A-loop, a novel conserved aromatic acid subdomain upstream of the Walker A motif in ABC transporters, is critical for ATP binding, FEBS Lett. 580 (2006) 1049–1055. doi:10.1016/j.febslet.2005.12.051.

[141] R.J.P. Dawson, K.P. Locher, Structure of a bacterial multidrug ABC transporter, Nature. 443 (2006) 180–185. doi:10.1038/nature05155.

[142] A. Ward, C.L. Reyes, J. Yu, C.B. Roth, G. Chang, Flexibility in the ABC transporter MsbA: Alternating access with a twist, PNAS. 104 (2007) 19005–19010. doi:10.1073/pnas.0709388104.

[143] G.E. Tusnády, B. Sarkadi, I. Simon, A. Váradi, Membrane topology of human ABC proteins, FEBS Letters. 580 (2006) 1017–1022. doi:10.1016/j.febslet.2005.11.040.

[144] S.-L. Liu, R. Sheng, J.H. Jung, L. Wang, E. Stec, M.J. O'Connor, S. Song, R.K. Bikkavilli, R.A. Winn, D. Lee, K. Baek, K. Ueda, I. Levitan, K.-P. Kim, W. Cho, Orthogonal lipid sensors identify transbilayer asymmetry of plasma membrane cholesterol, Nat Chem Biol. 13 (2017) 268–274. doi:10.1038/nchembio.2268.

[145] L.M. Hodges, S.M. Markova, L.W. Chinn, J.M. Gow, D.L. Kroetz, T.E. Klein, R.B. Altman, Very important pharmacogene summary: ABCB1 (MDR1, P-glycoprotein), Pharmacogenet Genomics. 21 (2011) 152–161. doi:10.1097/FPC.0b013e3283385a1c.

[146] A. Blees, D. Januliene, T. Hofmann, N. Koller, C. Schmidt, S. Trowitzsch, A. Moeller, R. Tampé, Structure of the human MHC-I peptide-loading complex, Nature. 551 (2017) 525–528. doi:10.1038/nature24627.

[147] G.M. Martin, C. Yoshioka, E.A. Rex, J.F. Fay, Q. Xie, M.R. Whorton, J.Z. Chen, S.-L. Shyng, Cryo-EM structure of the ATP-sensitive potassium channel illuminates mechanisms of assembly and gating, ELife. 6 (2017). doi:10.7554/eLife.24149.

[148] B. Chantemargue, F. Di Meo, K. Berka, N. Picard, H. Arnion, M. Essig, P. Marquet, M. Otyepka, P. Trouillas, Structural patterns of the human ABCC4/MRP4 exporter in lipid bilayers rationalize clinically observed polymorphisms, Pharmacol. Res. 133 (2018) 318–327. doi:10.1016/j.phrs.2018.02.029.

[149] Z. Zhang, J. Chen, Atomic Structure of the Cystic Fibrosis Transmembrane Conductance Regulator, Cell. 167 (2016) 1586-1597.e9. doi:10.1016/j.cell.2016.11.014.

[150] P.C. Smith, N. Karpowich, L. Millen, J.E. Moody, J. Rosen, P.J. Thomas, J.F. Hunt, ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer, Mol. Cell. 10 (2002) 139–149.

[151] T.A. Jones, J.-Y. Zou, S.W. Cowan, M. Kjeldgaard, Improved methods for building protein models in electron density maps and the location of errors in these models, Acta Crystallographica Section A. 47 (1991) 110–119. doi:10.1107/S0108767390010224.

[152] E. Biemans-Oldehinkel, M.K. Doeven, B. Poolman, ABC transporter architecture and regulatory roles of accessory domains, FEBS Letters. 580 (2006) 1023–1035. doi:10.1016/j.febslet.2005.11.079.

[153] Z.L. Johnson, J. Chen, Structural Basis of Substrate Recognition by the Multidrug Resistance Protein MRP1, Cell. 168 (2017) 1075-1085.e9. doi:10.1016/j.cell.2017.01.041.

[154] B. Verhalen, R. Dastvan, S. Thangapandian, Y. Peskova, H.A. Koteiche, R.K. Nakamoto, E. Tajkhorshid, H.S. Mchaourab, Energy transduction and alternating access of the mammalian ABC transporter P-glycoprotein, Nature. 543 (2017) 738–741. doi:10.1038/nature21414.

[155] G.A. Frank, S. Shukla, P. Rao, M.J. Borgnia, A. Bartesaghi, A. Merk, A. Mobin, L. Esser, L.A. Earl, M.M. Gottesman, D. Xia, S.V. Ambudkar, S. Subramaniam, Cryo-EM Analysis

of the Conformational Landscape of Human P-glycoprotein (ABCB1) During its Catalytic Cycle, Mol Pharmacol. 90 (2016) 35–41. doi:10.1124/mol.116.104190.

[156] Z.L. Johnson, J. Chen, ATP Binding Enables Substrate Release from Multidrug Resistance Protein 1, Cell. 172 (2018) 81-89.e10. doi:10.1016/j.cell.2017.12.005.

[157] P.-C. Wen, E. Tajkhorshid, Conformational Coupling of the Nucleotide-Binding and the Transmembrane Domains in ABC Transporters, Biophys J. 101 (2011) 680–690. doi:10.1016/j.bpj.2011.06.031.

[158] M. Moradi, E. Tajkhorshid, Mechanistic picture for conformational transition of a membrane transporter at atomic resolution, PNAS. 110 (2013) 18916–18921. doi:10.1073/pnas.1313202110.

[159] J. Fu, H.-L. Ji, A.P. Naren, K.L. Kirk, A cluster of negative charges at the amino terminal tail of CFTR regulates ATP-dependent channel gating, The Journal of Physiology. 536 (2001) 459–470. doi:10.1111/j.1469-7793.2001.0459c.xd.

[160] A.W. Ravna, G. Sager, Molecular model of the outward facing state of the human multidrug resistance protein 4 (MRP4/ABCC4), Bioorg. Med. Chem. Lett. 18 (2008) 3481–3483. doi:10.1016/j.bmcl.2008.05.047.

[161] A.W. Ravna, I. Sylte, G. Sager, Binding site of ABC transporter homology models confirmed by ABCB1 crystal structure, Theor Biol Med Model. 6 (2009) 20. doi:10.1186/1742-4682-6-20.

[162] H. Omote, M.K. Al-Shawi, Interaction of Transported Drugs with the Lipid Bilayer and P-Glycoprotein through a Solvation Exchange Mechanism, Biophys J. 90 (2006) 4046–4059. doi:10.1529/biophysj.105.077743.

[163] A.A.K. El-Sheikh, J.J.M.W. van den Heuvel, E. Krieger, F.G.M. Russel, J.B. Koenderink, Functional Role of Arginine 375 in Transmembrane Helix 6 of Multidrug Resistance Protein 4 (MRP4/ABCC4), Mol Pharmacol. 74 (2008) 964–971. doi:10.1124/mol.107.043661.

[164] H.G.M. Wittgen, J.J.M.W. van den Heuvel, E. Krieger, G. Schaftenaar, F.G.M. Russel, J.B. Koenderink, Phenylalanine 368 of multidrug resistance-associated protein 4 (MRP4/ABCC4) plays a crucial role in substrate-specific transport activity, Biochemical Pharmacology. 84 (2012) 366–373. doi:10.1016/j.bcp.2012.04.012.

[165] F.G.M. Russel, J.B. Koenderink, R. Masereeuw, Multidrug resistance protein 4 (MRP4/ABCC4): a versatile efflux transporter for drugs and signalling molecules, Trends in Pharmacological Sciences. 29 (2008) 200–207. doi:10.1016/j.tips.2008.01.006.

[166] M.S. Jin, M.L. Oldham, Q. Zhang, J. Chen, Crystal structure of the multidrug transporter P-glycoprotein from C. elegans, Nature. 490 (2012) 566–569. doi:10.1038/nature11448.

# Chapitre II. Basic principles of molecular dynamics

## II.1. Principle of molecular mechanics

### II.1.1. Introduction

Molecular Mechanics (MM) is based on classical physics. Unlike Quantum Mechanics (QM), in which the motion of electrons is considered, MM does not consider explicitly this motion and the related quantum effects. Within this approximation, MM allows to study systems with an important number of atoms including DNA, membranes, (membrane) proteins. Although structural properties of these systems can be evaluated, MM does not allow studying electronic properties and reactivity.

MM methods consider atoms as balls and bonds as springs, with a focus given to bonded interactions (*i.e.*, bond stretching, angle bending and torsion) and non-bonded interactions, mainly electrostatic and van der Waals, see Figure 1. All mathematical data describing these different contributions are collected in a force field, which is the essence of any MM method.

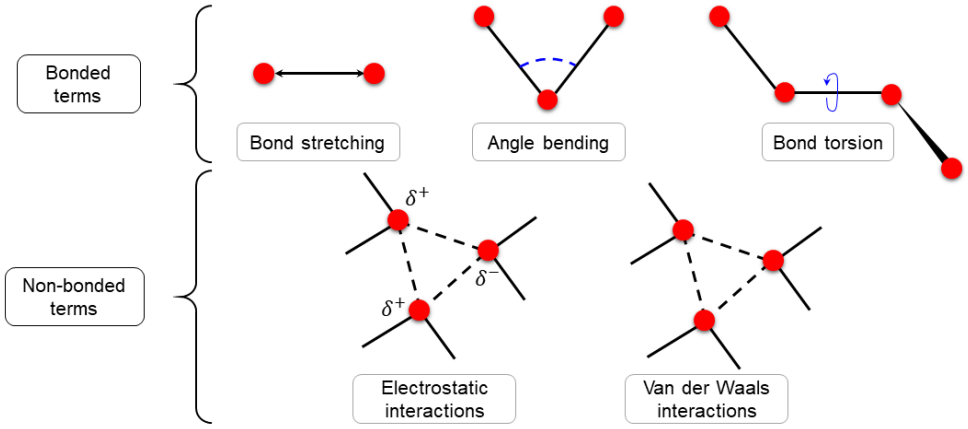



Figure 1: Representation of the basic bonded and non-bonded terms included in a force field.

The total potential energy $V$ is given by:

$$V_{total} = V_{bonded} + V_{non-bonded},$$

where:

$$V_{bonded} = V_{bond} + V_{angle} + V_{torsion}$$

$$V_{non-bonded} = V_{van\,der\,Waals} + V_{electrostatic}$$

Precisely, for a given force field, the potential energy $V(r^N)$ is given as a function of the position $(r)$ of the $N$ particles constituting the system. A typical form of equation for a force field is:

$$V(r^N) = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4_{\varepsilon ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right) \quad \text{(Eq. 1)},$$

where the forms of the different terms and their advanced versions are described below.

### II.1.2. Force field potentials

#### II.1.2.1. Bonded interactions

##### II.1.2.1.1. Bond stretching

In most of force fields, the bond stretching potential is considered within the harmonic approximation (Figure 2), and it is expressed by the Hooke's law:

$$v(l) = \frac{k}{2}(l - l_0)^2 \qquad (\text{Eq. } 2),$$

with $l_0$ the equilibrium, or reference, value of a given interatomic length and $l$ this length at a given time; $k$ is the force constant, describing how rigid the bond is. Atom types are defined by specific $k$ and $l_0$ parameter values. These parameters have been refined to fit with experimental bond values.

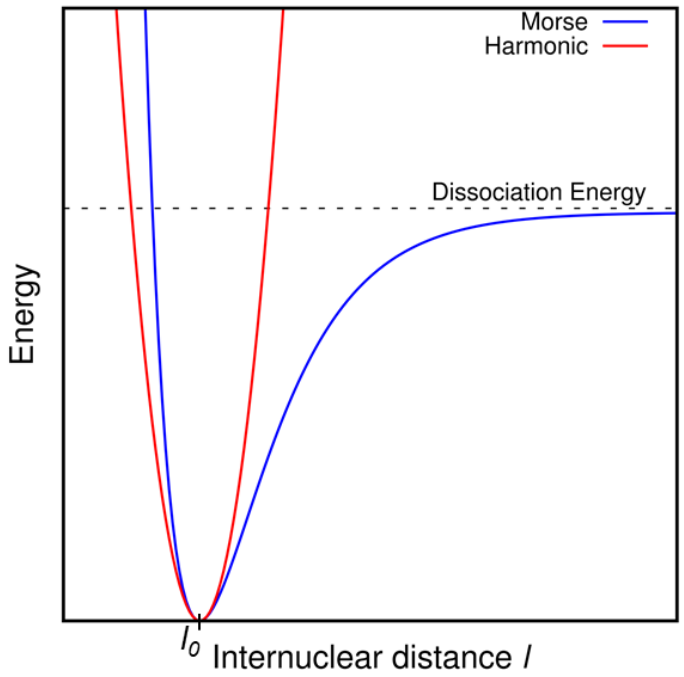



Figure 2: Representation of the harmonic and Morse potentials for bond stretching.

As a better approximation than the harmonic one, which provides a weak description far from the equilibrium of interatomic distance, one can consider a Morse potential too (Figure 2), expressed as follows:

$$v(l) = D_e\{1 - \exp[-a(l - l_0)]\}^2 \qquad (\text{Eq. } 3),$$

where $D_e$ is the depth of the potential energy minimum, $a = \omega\sqrt{\mu/2D_e}$ with $\mu$ the reduced mass and $\omega$ the frequency of the bond vibration. $\omega$ is related to $k$, the bond stretching constant, by:

$$\omega = \sqrt{k/\mu} \qquad (\text{Eq. } 4).$$

This Morse potential form is more accurate than the harmonic one, however, it is not commonly used in force fields, because of its greater computational cost due to the use of three parameters for each bond with respect to two in harmonic potentials and the use of an exponential function.

Another way to address the weakness of the harmonic potential description, and fit better with Morse potential, is to use cubic or higher terms:

$$v(l) = \frac{k}{2}(l - l_0)^2[1 - k'(l - l_0) - k''(l - l_0)^2 - k'''(l - l_0)^3 \dots] \qquad (\text{Eq. } 5),$$

As this approximation is mathematically based on Taylor series, it should be used only when the geometry of the molecule is close to equilibrium. Indeed, far from the reference bond value, the bond stretching could lead to an unrealistically huge stretching of the bond length.

#### II.1.2.1.2. Angle bending

Angle bending is often described by using a harmonic potential too:

$$v(l) = \frac{k}{2}(\theta - \theta_0)^2 \qquad (\text{Eq. } 6).$$

As for bond stretching, this term is described by a force constant $k$ and an equilibrium (reference) angle value $\theta_0$. The force constants for angle bending is often smaller than those used for bond stretching, mainly because deviation of an angle from its equilibrium value is easier, *i.e.*, it requires less energy than bond stretching.

The description of the angle bending potential in force fields can be improved, as for bond stretching, by including higher order terms:

$$v(\theta) = \frac{k}{2}(\theta - \theta_0)^2[1 - k'(\theta - \theta_0) - k''(\theta - \theta_0)^2 - k'''(\theta - \theta_0)^3 \dots] \qquad (\text{Eq. } 7).$$

#### II.1.2.1.3. Torsions

The term related to torsions is crucial to tackle conformation of molecules. Two torsion potential terms are actually described, namely the proper torsion (Figure 1) and the improper torsion (out-of-plane bending, see Figure 20). The most characteristic form of the proper torsion term uses a periodic cosine function:

$$v_{proper}(\omega) = \sum_{n=0}^{N} \frac{V_n}{2}[1 + \cos(n\omega - \gamma)] \qquad (\text{Eq. } 8),$$

where $\omega$ is the torsion angle, $V_n$ is the barrier height, $n$ is the multiplicity and $\gamma$ is the phase factor.

Improper torsions are defined by four atoms $(i, j, k, l)$, three of them forming a covalent bond with the forth (central) atom. The corresponding potential is given by:

$$v_{improper}(\omega_{ijkl}) = \frac{1}{2}k_\omega(\omega_{ijkl} - \omega_0)^2 \qquad (\text{Eq. } 9).$$

The improper torsion is not always an explicit term in force fields (*i.e.* it is mostly found in united-atom forcefield to maintain the stereochemistry at chiral centers), it can be replaced by an 1,4 non-bonded interaction.

The out-of-plane bending term can be described in two ways (Figure 3). This torsion can be considered as an out-of-plane bend calculating the angle between a bond from the central atom and the plane defined by the central atom and two other atoms (Figure 20A). The other way is to calculate the height of the central atom above a plane defined by the other three atoms (Figure 20B). These descriptions are ascribed to a harmonic potential as follows:

$$v(\theta) = \frac{k}{2}\theta^2 \qquad \text{(Eq. 10a)}, \qquad\qquad v(h) = \frac{k}{2}h \qquad \text{(Eq. 10b)}.$$

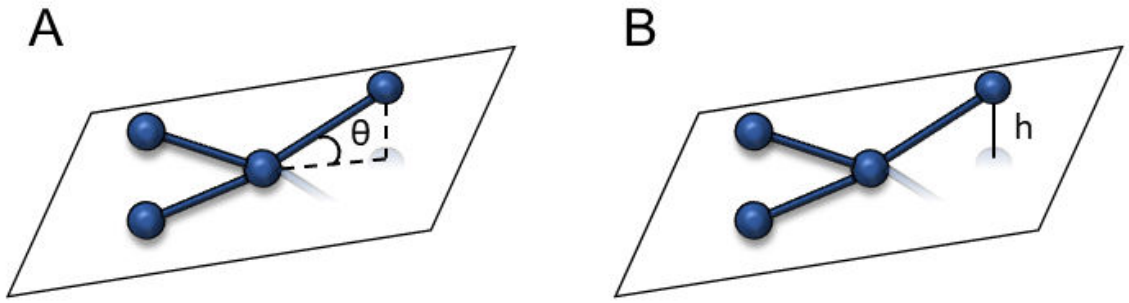



Figure 3: Representation of the out-of-plane torsions adapted from Molecular Modelling. Principles and Applications, by A.R. Leach, 2001.

#### II.1.2.2. Non-bonded interactions

Non-bonded interactions are weaker than bonded ones (which are related to covalently bonded atoms) but they are long-ranged. They play a key role in conformational arrangements including protein folding or lipid bilayer assembly. The main three types of non-bonded interactions will be described in this section, namely the electrostatic (or Coulomb), van der Waals and hydrogen bond interactions.

##### II.1.2.2.1. Electrostatic interactions

The charge distribution can be represented as (fractional) point charges distributed on the structure of the molecule, which allows calculating the electrostatic interactions between atoms of a molecule. In most cases, the point charges are atom-centered charges. These charges are called partial atomic charges or net atomic charges. The following equation describes the electrostatic interaction between all atoms (point charges) of two molecules:

$$v_{electrostatic\,(r_{ij})} = \sum_{i=1}^{N_A}\sum_{j=1}^{N_B}\frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \qquad \text{(Eq. 11)},$$

where $N_A$ and $N_B$ are the numbers of point charges in the molecules, $q_i$ and $q_j$ are the charges of atoms of the two molecules $i$ and $j$, $\varepsilon_0$ is the dielectric constant or vacuum permittivity, $\varepsilon_r$ is the relative dielectric constant.

##### II.1.2.2.2. van der Waals interactions

The van der Waals interactions are mainly ascribed to dipole-dipole interactions, which include dispersion forces. In force fields, they are often derived from by the Lennard-Jones 12-6 function:

$$v(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \qquad \text{(Eq. 12)},$$

where $\sigma$ is an adjustable collision diameter and $\varepsilon$ is the depth of the potential well.

For pairs of atoms $(ij)$, it is written as follows:

$$v_{Lennard-Jones}(r_{ij}) = \left(\frac{C_{ij}}{r_{ij}}\right)^{12} - \left(\frac{C_{ij}}{r_{ij}}\right)^{6} \qquad \text{(Eq. 13)},$$

where the $C_{ij}$ parameters are defined for the different atom types, *e.g.*, aliphatic atoms or aromatic atoms which both exhibit different van der Waals interactions. Depending on the force field, the combination of these terms can be obtained geometrically or arithmetically.

#### II.1.2.2.3. Hydrogen bonding

As hydrogen bonding is often mainly driven by electrostatic interactions, it is usually sufficiently described by the electrostatic potential. Nonetheless, some force fields (*i.e.*, old all-atom force fields like AMBER force fields[1]) propose an explicit hydrogen bonding term, which is usually described by the following potential:

$$v(r) = \left(\frac{A}{r^{12}}\right) - \left(\frac{C}{r^{10}}\right) \qquad (\text{Eq. } 14).$$

The aim of this hydrogen bonding description is to enhance the description of its specificity (*e.g.*, halogens or polar hydrogens), and subsequently its effect on geometrical features. Other force fields even use more complicated terms to consider the deviation of the geometry between hydrogen bond donor and acceptor.

#### II.1.2.2.4. Cut-off distances

An important factor to consider is that the number of bonded interactions is almost proportional to the number of molecules, whereas non-bonded interactions scale differently to $N(N-1)/2$ ($N$ is the number of atoms). To avoid an overwhelming number of calculations, cut-off values are used, beyond which non-bonded interactions between two atoms can be neglected. The cut-off values are force field dependent, but they range from 9 to 14 Å for both Coulomb and van der Waals interactions. Above the cut-off, non-bonded interactions are not explicitly evaluated, and long-range corrections are applied (*i.e.*, adding an additional term to the energy which depend on internal coordinates). The most common long-range correction is the Particle-Mesh Ewald (PME) method[2].

#### II.1.2.2.5. Cross terms

Cross terms can be added in force fields (*e.g.*, in CHARMM force field[3]), which are describing coupling between stretching, bending and torsion. For instance, a bending (*e.g.*, the strongly bent angle in the $H_2O$ molecule) affects the stretching (*e.g.*, the O–H bonds of the $H_2O$ molecule). Stretch-bend, stretch-stretch, bend-bend, stretch-torsion and bend-torsion correlation terms can be added. An example of stretch-bend term in the CHARMM force field is the Urley-Bradley term, which is a harmonic potential of the 1,3 inter-atomic distance to implicitly establish a correlation with the angle bending description:

$$\sum_{Urey-Bradley} k_u(u-u_0)^2 \qquad (\text{Eq. } 15),$$

where $u$ is the 1,3 inter-atomic distance. These terms are used to enhance the accuracy, but it increases computational cost.

### II.1.3. Comparison between force fields

Differences between force fields rely in both the basic equation and all related parameters (*e.g.*, force constants, equilibrium values). Some force fields differ only from some parameters, which partially holds true for the AMBER[4], CHARMM[5] and GROMOS[6] force fields, which are all based on a similar equation of the potential energy except a few extra terms (see below) [7]:

$$U = \sum_{bonds} k_r(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{Urey-Bradley} k_b(b - b_0)^2 + \sum_{torsions} k_\emptyset\big(1 + cos(n\emptyset - \delta)\big) + \sum_{non-bonded} 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) + \sum_{non-bonded} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad (\text{Eq. } 16).$$

The force fields can be either non-polarized (classical), *e.g.*, CHARMM[5] (*e.g.*, CHARMM36), AMBER[4], OPLS[8], GROMOS[6], COSMOS[9], MARTINI[10], or polarizabled, *e.g.*, AMEOBA[11], CHARMM[12] (*e.g.*, CHARMM Drude). It is worth mentioning that classical force fields calculate electrostatic interactions as fixed charges, which are centered on the nuclei without considering electrostatic polarization, and charge transfer. Polarization may strongly impact the energetics of a molecular structure. Polarizable force fields allow to modify the partial charges along the simulations, depending on the environment. It may appear critical with the dynamic of polar species dynamic (*e.g.*, ion channels)[13]. However, the drawback is that the computational cost is at least doubled, explaining the more general use of non-polarizable force fields. This section mainly focuses on the former type, as they are those mainly used in current MD simulation studies.

From a general point of view, force fields have been designed to reach the best accuracy with a given molecular system, however any extra term may affect the time allocated to a MD simulation. In MM, the computational time increases linearly with the number of atoms, as pair-wise interactions are considered. With the increase of computational power more extra terms should be included if they appear critical for the biological system and the process under investigation. There are however molecular systems, aiming at mimicking realistic situations, for which the size and the need of long-term sampling (*e.g.*, membrane crossing or protein folding) are critical. Several options are available to decrease simulation time, by *e.g.*, i) decreasing the number of atoms (force field resolution), ii) increasing time step of the MD simulations (decrease of the resolution time), see section II.2.1, iii) adding restrictions (*e.g.*, geometrical constraints), and iv) changing the way the solvent is treated, see section II.2.1.4.

Three types of force fields have been developed, namely, all-atom (AA), united-atom (UA) and coarse-grain (CG) force fields, the last two artificially decrease the number of atoms numbers. In the AA force fields (*e.g.*, AMBER, CHARMM), all atoms of the molecular system are explicitly considered to evaluate interatomic interactions. In the UA force fields (*e.g.*, GROMOS), not all hydrogens are explicitly described, *e.g.*, the non-polar hydrogens (*e.g.*, in lipid tails of membranes). To do so while keeping relevancy, the hydrogens are grouped with the linked carbon within a single atom type, *e.g.*, an atom type for the $CH_2$ or the $CH_3$ group. This description reduces the local and specific description of hydrogen bonding (*e.g.*, polar hydrogens involved in hydrogen bonds). However, depending on the studied process, the atom type and its chemical environment, the UA force field can be particularly relevant, *e.g.*, aliphatic chains. The use of UA force fields allows to speed up the calculations up to three times compared to AA forcefields. Another option which dramatically reduces the number of atoms and speed up the simulations relies on CG force fields (*e.g.*, MARTINI). CG force fields regroup not only the hydrogen atoms with the linked carbons on an aliphatic chain, but several heavy atoms (*e.g.*, C, N, O) in a single bead (~ 4 atoms/bead). The atoms are grouped together based on their chemical nature and their common role in chemical surrounding, *e.g.*, mimicking a polar/hydrophobic moiety. The number of atoms is thus considerably decreased by using such

force fields, making the computational cost dramatically lower and allowing tackling events in a range of the millisecond to second timescale[10,14]. A less explicit description of the atomic interactions allows faster simulations but a decrease in the atomistic detailed description of structural features. Moreover, independently of its resolution, the accuracy of a force field relies on parametrization. To exemplify differences between force fields, the importance of each term and parameter is discussed hereafter based on a few representative force fields.

The AA AMBER force field, more precisely the ff94, a version named “helical friendly”. Its equation is given by[15]:

$$E_{total} = \sum_{bonds} K_r\left(r - r_{eq}\right)^2 + \sum_{angles} K_\theta\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos(n\emptyset - \gamma)\right] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right] + \sum_{H-bonds}\left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right] \qquad (\text{Eq. } 17).$$

In this equation, both bond stretching and angle bending are represented with a harmonic expression, the van der Waals is represented by a Lennard-Jones 12-6, electrostatic interactions are expressed by a Coulomb potential with atom-centered point-charges, and an explicit term for hydrogen bonding is present. Neither improper dihedrals nor cross terms are described. The name “helical friendly” was given because of an over-stabilization of α-helices with this force field. In the further implementation, the AMBER ff96 version, the Φ/Ψ torsion parameters were adjusted based on QM calculations, but the opposite effect (destabilization of α-helices) was observed. While the AMBER ff99 force field was developed to optimize several other parameters, it also definitely fixed this helicity error.

The UA GROMOS force fields[16] include an improper torsion term:

$$v_{har}\left(r;\, k_\xi, \xi_0\right) = \sum_{n=1}^{N_\xi} k_{\xi_n}\left[\xi_n - \xi_{0_n}\right]^2 \qquad (\text{Eq. } 18),$$

where $\xi_n$ is the value of the dihedrals angle between atoms i, j, k and l, $\xi_0$ is the equilibrium value, and $k_\xi$ is the force constant. The dihedral angle can be calculated using the atomic position $r$ with:

$$\xi_n = sign(\xi_n) arccos\left(\frac{r_{mj}.r_{qk}}{r_{mj}r_{qk}}\right) \qquad (\text{Eq. } 19),$$

where $r_{mj} \equiv r_{ij} \times r_{kj}, r_{qk} \equiv r_{kj} \times r_{kl}$ , $m$ and $q$ are defined through the cross products and the sign of the angle $\xi_n$ is given by the expression $sign(\xi_n) = sign\left(r_{ij}.r_{qk}\right)$. The angle $\xi_n$ is not defined if $r_{mj} = 0$ or $r_{qk} = 0$. This term is not required in AA force fields as all torsions are explicitly described.

Another specific term has been added to the GROMOS force fields which is defined as follows:

$$V(r;s) = V_{phys}(r;s) + V_{special}(r;s) \qquad (\text{Eq. } 20),$$

where $V_{phys}(r;s)$ is the bonded and non-bonded terms of the classical force fields, and $V_{special}(r;s)$ explicitly represents a user-defined restriction on *e.g.*, position, distance, angles. There exist several versions of the GROMOS force field. For proteins, the difference between the GROMOS53a6[16] and GROMOS54a7[6] versions is that the former under-estimates the stability of α-helices while the latter slightly over-estimates it. The difference between these two versions relies in the torsion energy term (Table 1[6] and Table 5[16]), with a different $C_{12}$

value for the O,N-atom type pair. This modification stabilizes the hydrogen bonding in the backbone between N-H and C=O to fit $\varphi$ and $\psi$ of proteins in the PDB (protein data bank).

The AA Slipids force field[17], which is dedicated to lipids, was developed from the (protein-friendly) CHARMM force field, namely keeping from CHARMM36[18] the bond and angle parameters for all atoms, as well as the torsion and Lennard-Jones parameters of the head group and glycerol. All the other parameters (*i.e.*, partial atomic charges for the entire lipid, Lennard-Jones and torsion parameters for lipid tails) were revisited. The Lennard-Jones parameters were fitted using experimental data (*i.e.*, heats of vaporization and densities). For instance, for saturated lipids, the charges for the lipid tails were fitted from hexadecane ($C_{16}H_{34}$), as averaged on more than 50 different conformations obtained from several MD simulation runs. The torsion potential for lipid tails was obtained using a scan of the potential energy surface (PES) of octane ($C_8H_{18}$). The energies between the conformations were obtained using high level *ab initio* methods (CCSD(T)/cc-pVQZ)[7] and reference herein. The atomic charges for the headgroups were derived from the DMPC lipid, as averaged on numerous different (randomly chosen) conformations obtained from several MD simulation runs. For each conformation, the charges were computed using by DFT (B3LYP/cc-pVTZ) and the restricted electrostatic potential approach (RESP)[19]. Further Slipids' versions have been elaborated, including specific re-parameterization for polyunsaturated lipids, by combining QM and MD calculations[20].

The GROMOS 45A3[21] force field was also developed for lipid bilayers. It showed errors in the description of some properties including area per lipid (APL), and it failed at describing gel phases[22]. An updated force field, *i.e.*, G43A1-S3[23] was then proposed with new parameters based on *ab initio* computations (*i.e.*, re-evaluating the PC head group torsion parameters), and on thermodynamic data for liquid alkanes improving the agreement with experiments. The GROMOS family has been continuing to improve, with the G53A6 version[16] proposing a full re-parametrization. This parametrization has allowed to better describe the free enthalpy of hydration of common phospholipids, which is critical for several biomolecular processes.

An interesting point related to the differences between force fields, is the compatibility between them. Indeed, based on the common formalism, Slipids (for lipids) is compatible with AMBER force fields (for proteins), mainly due to similar parametrization of partial charges. Similarly, Lipid14[24] and GAFF[25] are compatible, as Berger lipids[26] and GROMOS are. Mixing force fields having a different formalism may lead to artifacts and inconsistencies[27].

### II.2. Principle of molecular dynamics (MD)

MM computes the potential energies $(V)$ of atoms in the surrounding of local minima. MD accounts for atomic motions (either small fluctuations around the minimum or large-scale conformational changes). Along the same line with MM formalisms, the Newton's laws of motion are applied. In MD, successive atomic configurations are studied along time, which constitutes a trajectory. The trajectory results from modifications of atomic positions and velocities over time. Newton's law states that:

- a particle moves parallel to its velocity until a force is applied on it;
- the sum of forces on a particle equals its mass times its acceleration ($F = ma$)

- any action of one particle on another makes that the second particle has the opposite action on the first one

The trajectory is obtained by solving the $F = ma$ equation, which can be written for a particle $i$ as:

$$\frac{d^2x_i}{dt^2} = \frac{F}{m_i} \qquad (\text{Eq. } 21),$$

where $m_i$ is the mass of the particle, at coordinate $x_i$.

### II.2.1. Integrators, periodic boundary conditions, thermostats and barostats

For a given molecular system, the potential energy and the forces are calculated for each atom at a given time. This leads the atoms to move at a certain speed to reach a new position. The system of atoms thus moves from one to another time step. The evolution of the speed and the position of each atom is calculated by a so-called integrator, an algorithm which solves the simplest form of Newton's equation:

$$F(t) = m\,\frac{\Delta v}{\Delta t} \qquad (\text{Eq. } 22).$$

Several integrators have been developed, however in this section a focus is given only on the most two commonly used, namely the leap-frog and velocity verlet algorithms. In the former algorithm, the equations allow linking the successive time steps as follows:

$$v_i\left(t+\frac{1}{2}\Delta t\right) = v_i\left(t-\frac{1}{2}\Delta t\right) + \frac{\Delta t}{m}F(t) = \; v_i\left(t-\frac{1}{2}\Delta t\right) + \; a_i(t)\Delta t \qquad (\text{Eq. } 23),$$

$$r_i(t+\Delta t) = r_i(t) + v_i(t)\Delta t\left(t+\frac{1}{2}\Delta t\right) \qquad (\text{Eq. } 24).$$

In this algorithm, while the new coordinates $r$ are computed at every step and the velocities $v$ at $t+\frac{1}{2}\Delta t$ (Figure 21), the velocities are calculated halfway between position steps. The positions and velocities are not known simultaneously.

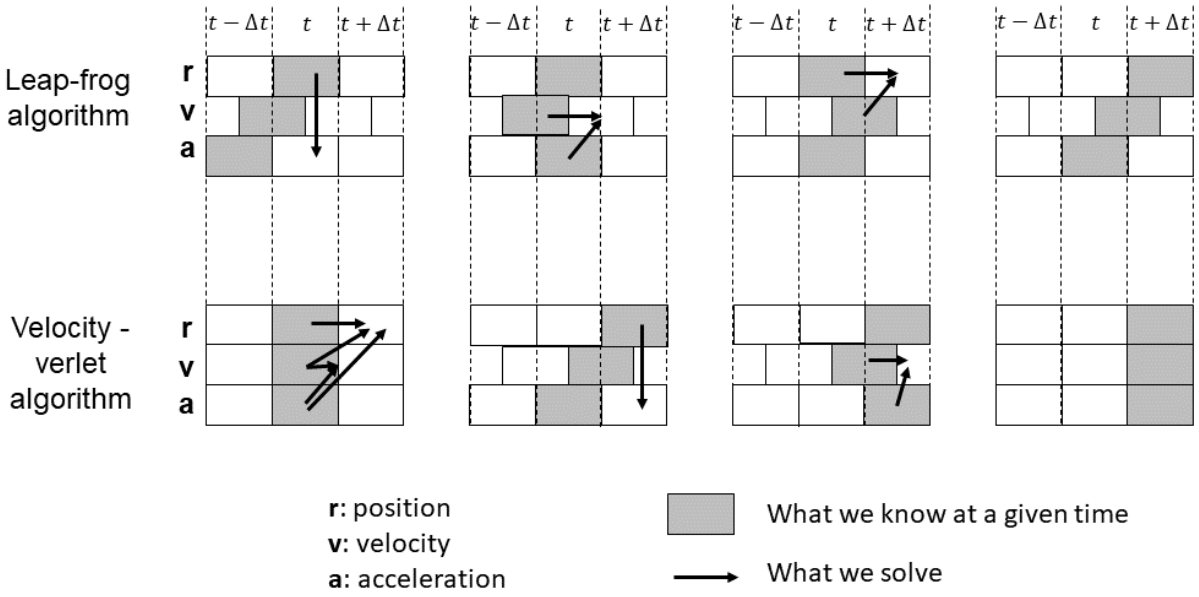



Figure 4: Schematic explanation of the propagations of the motions based on two algorithms: leap-frog and velocity-verlet algorithms

Another efficient algorithm, however less used in classical MD simulations, is the Velocity Verlet algorithm. Compared to the leap-frog algorithm, it allows to know the positions and velocities simultaneously (Figure 4) as follows:

$$v_i(t+\Delta t) = v_i(t) + [a_i(t) + a_i(t+\Delta t)]\frac{\Delta t}{2} \qquad (\text{Eq. } 25),$$

$$r_i(t+\Delta t) = r_i(t) + v_i(t)\Delta t + \frac{1}{2}a_i(t)\Delta t^2 \qquad (\text{Eq. } 26).$$

To accurately describe molecular motions, the (user-defined) time step $\Delta t$ is a crucial parameter. It should be smaller than the fastest atomic motion. Knowing that the vibration of hydrogen atoms is around 13 fs, a most common time step used in MD simulations is 2 fs.

#### II.2.1.1. Periodic boundary conditions

In MD, the studied molecular system and its environment (*e.g.*, water) is embedded in a box, which is usually a rectangle. The boundaries of the box are artificial limits, at which the molecular behavior or intermolecular interactions should be carefully treated. To solve this issue, periodic boundary conditions (PBC) have been developed, which consist in (virtually) placing the same box at each face of the primary one (see Figures 5A for a 2D schematization). When a molecular system (a solute) reaches, and crosses a boundary, it reappears at the opposite side of the box, mimicking an infinite environment (Figure 5B). Big molecules (say a protein) can partly cross the boundaries, which may induce a critical artifact by inducing non-bonded self-interaction between one side of the molecule and its opposite side (Figure 5C). To avoid this error, boxes should be constructed sufficiently big; however, the size of the box may dramatically affect the computational time, mainly because of the increasing number of water molecules.

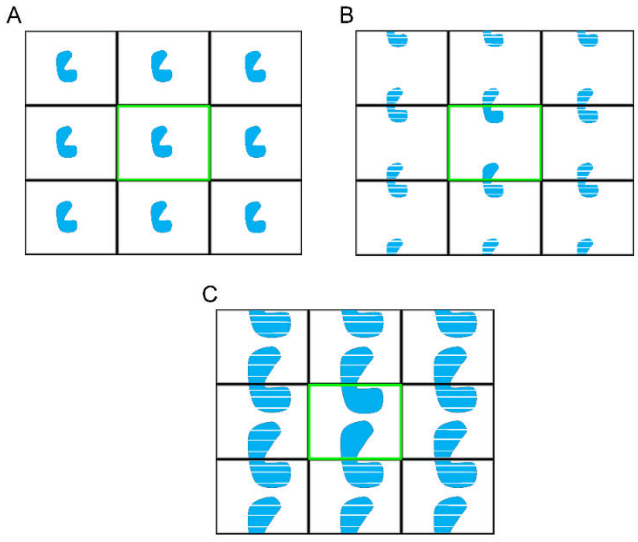



Figure 5: A) 2D representation of the PBC conditions. B) 2D representation of the PBC conditions when a particle (here represented by a bean) is going outside of the box. C) 2D representation of a case of self-interaction, which is the case with a too small box.

#### II.2.1.2. Temperature regulation

The temperature $T$ captures the molecular agitation and it thus influences the velocities, so the kinetics of the molecular system. There are several ways to control $T$, depending on the system

and on the force field parametrization. Among a plethora of algorithms (*e.g.*, Velocity-rescaling[28], Berendsen[29], Nosé-Hoover[30], Nosé-Hoover chain[31], Langevin[32], Andersen[33]) only the three most common thermostats are detailed in this section, namely Berendsen, Nosé-Hoover and Langevin. The following equation is used to control $T$:

$$E_{kin} = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 = \frac{1}{2} N_{df} kT \qquad (\text{Eq. } 27),$$

where $E_{kin}$ is the kinetic energy, $v_i$ is the velocity of each atom, $k$ is the Boltzmann's constant and $N_{df}$ is the number of degrees of freedom (equals to 3 in MD simulations).

In the NVE (constant number of particles, volume and total energy) ensemble, $T$ can fluctuate, which makes the simulation of biological conditions impossible. The NVT (constant temperature instead of constant total energy) ensemble seems more adapted to fit with these conditions. In the NVT ensemble, $T$ is regulated by scaling velocities. A rough approach would be to scale the velocities of all atoms directly when $T$ deviation is measured, which may induce artifacts, because preventing $T$ fluctuations. To appropriately and slowly correct $T$, the Berendsen thermostat describes deviation in $T$ by an exponential decay [29]:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \qquad (\text{Eq. } 28),$$

where $T_0$ is the reference temperature, and $\tau$ is the decay time constant. This mathematical evolution mimics the presence of one external temperature bath, which is a very efficient approach to relax the system at a desired temperature. This method is however less adapted to extensively sample molecular systems as the fluctuations in the kinetic energy are artificially quenched.

Another way to regulate $T$ is the Nosé-Hoover's thermostat[30]. This thermostat allows a proper canonical ensemble simulation by introducing an external heat bath variable $p_\xi$ in the equation of motion:

$$\frac{d^2 x_i}{dt^2} = \frac{F_i}{m_i} - \frac{p_\xi dx_i}{Q\ dt} \qquad (\text{Eq. } 29),$$

where $p_\xi$ is the heat bath variable defined as:

$$\frac{dp_\xi}{dt} = (T - T_0) \qquad (\text{Eq. } 30),$$

and:

$$Q = \frac{\tau_T^2 T_0}{4\pi^2} \qquad (\text{Eq. } 31).$$

Conversely to the Berendsen thermostat that allows a rapid convergence to the reference temperature $T_0$, Nosé-Hoover's thermostat induces temperature oscillations that will slowly converge to $T_0$. Both Berendsen and Nosé-Hoover thermostats are recommended at different stages of MD simulation. The former is recommended for pre-equilibrated simulations, while Nosé-Hoover is recommended for production (long-term simulations). It is worth mentioning that in MD simulations, multiple thermostats can be used with different temperatures for different groups of atoms (*e.g.*, to prevent a substrate to boil when performing simulated annealing on a protein to force secondary structure to re-organize). An issue that should be considered is how to define $T$ when starting with a set of fixed atomic coordinates (before using

thermostat or because of a constrain). For instance, in the leap-frog integrator, the velocities of atoms at $t_0 - \frac{1}{2}\Delta t$ can be generated by a Maxwell-Boltzmann distribution:

$$p(v_i) = \sqrt{\frac{m_i}{2\pi kT}}\, exp\left(-\frac{m_i v_i^2}{2kT}\right) \qquad (\text{Eq. } 32).$$

Another thermostat used with AMBER force fields[4] is the Langevin thermostat[32]. In this case, the particles receive a random force and their velocities are lowered by a constant friction at each step. This is comparable to a system being embedded in a sea of frictional smaller particles. The friction and dissipation are related, and they ensure a proper statistics. It is described in the fluctuation-dissipation theorem in which a system in thermodynamic equilibrium responds to a small force applied as the response to a spontaneous fluctuation. The equation applied to follow the movement of a given particle is as follows:

$$m\ddot{r}_i = -\nabla_i U(r) - \gamma\dot{r} + \sqrt{2\gamma k_B T}\, R(t) \qquad (\text{Eq. } 33),$$

where $\ddot{r}_i$ is the particle acceleration, $\dot{r}$ is the velocity, $U(r)$ is the particle interaction potential, $\nabla$ is the gradient operator, $-\nabla \mathrm{U}(r)$ is the force applied on the particle, $k_B$ the Boltzmann's constant, $\gamma$ is the friction coefficient, and $R(t)$ is the random force due to the random collision of the surrounding molecules.

However, due to the random nature of the applied force, the transfer of the momentum is not ensured, which makes *e.g.*, this algorithm not recommended when computing diffusion coefficients.

#### II.2.1.3. Pressure regulation

The regulation of pressure is based on the same principle than that of thermostats. The correlation between $T$ and velocities is replaced by the correlation between pressure and an adaptation of box size (box scaling). The "bath" principle is also used. The most common barostats (*i.e.*, Berendsen, Parinello-Rahman) are described in this section. The Berendsen barostat[34] is based on the following mathematical expression of the pressure evolution:

$$\frac{dP(t)}{dt} = \frac{P_0 - P(t)}{\tau_P} \qquad (\text{Eq. } 34),$$

where $P_0$ is the reference pressure, $P(t)$ is the instantaneous pressure and $\tau_P$ is the relaxation time constant. The pressure leads to scale the box at every step by a scaling factor $\mu$ given for a cubic box by:

$$\mu = 1 - \frac{\beta\Delta t}{3\tau_P}\big(P_0 - P(t)\big) \qquad (\text{Eq. } 35),$$

where $\beta$ is the isothermal compressibility of the system, and $\Delta t$ is the integrator time-step.

The barostat counterpart to Nosé-Hoover is the Parinello-Rahman barostat[35–37]. This barostat allows the box to change size and shape.

To choose the barostat depends on the system, and the expected behavior. For instance, Berendsen thermostat and barostat, being easy to define and quick, are often used for the equilibration stage. The Nosé-Hoover thermostat and Parinello-Rahman barostat are usually used, in a second stage, to produce a proper NPT statistical ensemble.

### II.2.1.4. Water models

Surrounding solvent (mainly water in biological applications) can be described explicitly, implicitly or by using a hybrid approach. The implicit solvents (*e.g.*, Poisson-Boltzmann, Generalized Born[38]) dramatically decrease simulation time, however, they are not adapted to water description as they totally ignore specific intermolecular interactions, such as hydrogen bonding. In MD simulations, an explicit description is preferred, in which water molecules are explicitly added to the simulation box and their interaction with the solute are explicitly described, according to the model. Several water models have been developed[39] to fit with physical parameters (*e.g.*, density, radial distribution function). Three categories of models can be distinguished, namely the rigid models (*i.e.*, the atomic positions are fixed, and water molecules fit to a known geometry of water), the flexible water models (*i.e.*, described as a force field with bond stretching and angle bending) and polarizable models.

The rigid models (SPC[40], SPC/E[41], TIP3P[42], TIP4P[42] and TIP5P[43]) are the most often used. There are different models based on geometric parameters, size and location of partial charges. Each model describes differently the physical properties (*e.g.*, diffusion constant, dielectric constant).

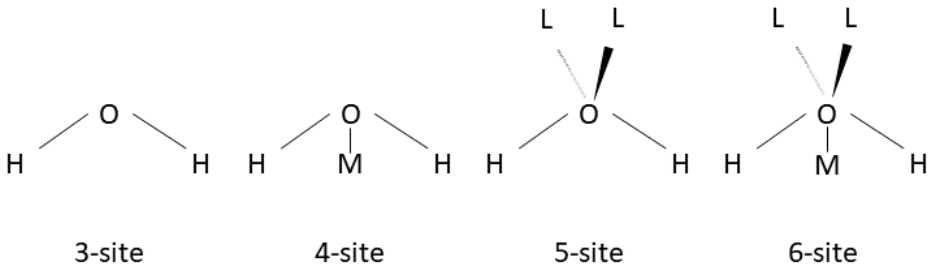



Figure 6: Schematic representation of the different water rigid models.

The most common are:

- the 3-site model (*e.g.*, SPC, SPC/E, TIP3P), which is simple but efficient, and widely used. Here, each atom of the water molecule is assigned to a point charge (Figure 6) and only oxygen is described by its own Lennard-Jones parameters (*e.g.*, SPC and SPC/E compatible with GROMOS force fields). The differences between the three 3-site models is that the SPC/E model adds a polarization correction to the potential. The TIP3P water model (popular in AMBER and CHARMM force fields) differs from the other two in the Lennard-Jones and Coulomb terms. The corresponding differences result in a different description of bulk properties including self-diffusion coefficients or mean-square displacement (MSD)[44].
- the 4-site model (*e.g.*, TIP4P), in which a negative charge is placed on a dummy atom M instead of oxygen (Figure 6). This model improves the electrostatic description around the molecule.
- the 5-site model (*e.g.*, TIP5P), in which the negative charges are split in two lone pairs (L) of the oxygen atoms with a tetrahedral-like geometry.
- the 6-site model[45], which combines the 4 and 5-site models.

For each rigid model, a flexible variant exists (*e.g.*, Flexible SPC, TIP4PF). Hybrid solutions [46] are also possible in which the two explicit and implicit approaches are considered, *e.g.*, explicit for the studied molecule and the first solvent shell, and implicit for the rest. Hybrid approaches can be used in QM/MM simulations (simulations combining QM and classical treatments).

#### II.2.1.5. Free molecular dynamics

Free MD simulations mean simulations without applied constrains, conversely to biased MD simulations (see section III). To exemplify a classical procedure, one can consider the study of the depth of penetration of a molecule in a lipid bilayer (Figure 7). In this procedure, the lipid bilayer is supposed to be pre-equilibrated (the below procedure can also be used to equilibrate the lipid bilayer itself) and the topology of the molecule properly created (with the proper initial geometry and charges).

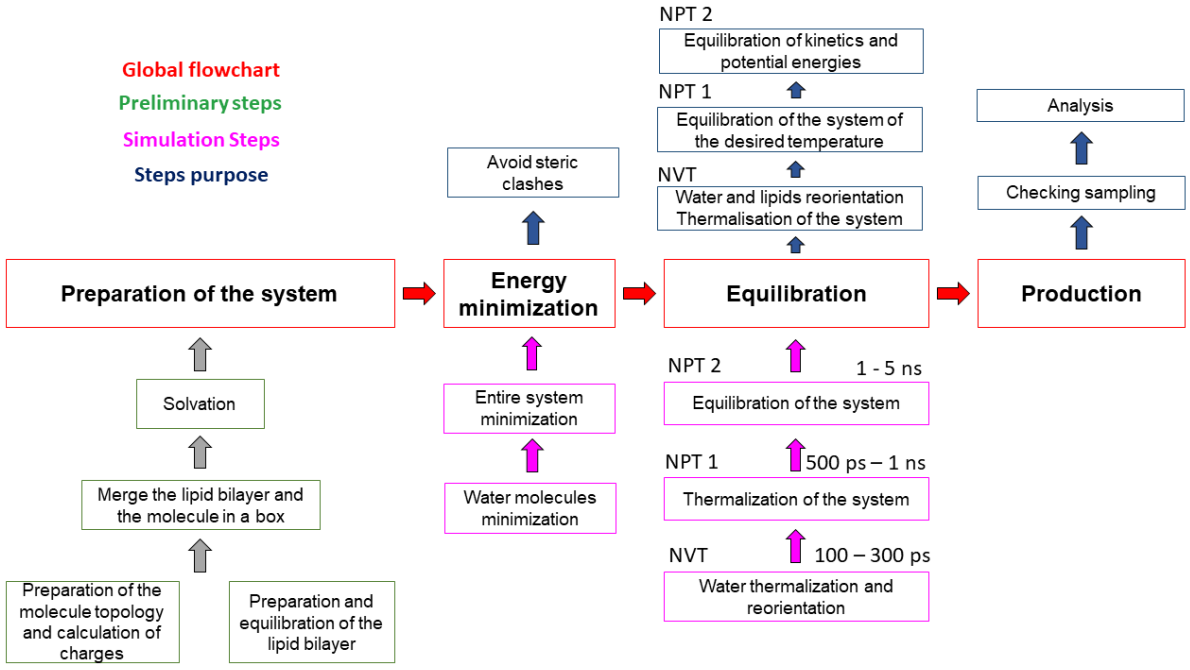



Figure 7: Step-by-step schematic explanation of a classical free MD simulation.

Both the solute and the lipid bilayer are put together and solvated (addition of water and ions) in a simulation box. The $x, y$ coordinates of the box are defined by the lipid bilayer, whereas the $z$ coordinate (box size) should be tuned to avoid self-interaction, according to the cut-off values chosen for non-bonded interactions.

The first step is the energy minimization, first of the water system only, then of the whole system (water, ions, lipids, solute), which only aims at avoiding steric clashes (*e.g.*, between water and solute, or between lipids and solute, water and lipid head groups, ion and lipid head groups), and not finding global minimum of energy. After checking the decrease in potential energy, the next steps can be performed.

The second step consists generally in a short NVT simulation (100-300 ps) to thermalize the water system, mainly allowing reorientation of water molecules. The whole system slowly and linearly reaches the targeted temperature. After checking temperature stabilization, two longer NPT simulations (500 ps – 1ns, and 1 – 5 ns, respectively) are carried out to thermalize the

entire system at a constant temperature and to equilibrate the system (*i.e.*, equilibration of the kinetic and potential energies) under NPT conditions. The pressure evolution (particularly for lipid bilayers) and box vectors are analyzed to ensure a stable lateral (xy) area of the lipid bilayer. These equilibration stages at the desired temperature and pressure are critical, especially with a protein embedded in a lipid bilayer, the protein-lipid-water system being more heterogenous. A longer equilibration is required for such heterogeneous system in which each component act on each other (*e.g.*, water self-arrange and orient around the lipid polar head group inducing in turn re-orientation the substituent of the amino acids; protein adapts its conformation according to the surrounding lipids, which, in turn, adapt their arrangement around the protein).

The last step is the production run, which is performed under NPT conditions. A sufficient production time is mandatory to enable a sufficient sampling (Maxwell-Boltzmann statistics) of the biological process under investigation (here insertion of a molecule from bulk water to the lipid bilayer). To check convergence depends on the process under investigation. In this example, the system is converged when the molecule reaches its equilibrium position and stays long in it. The insertion of xenobiotics in lipid bilayers is usually about the hundreds of nanoseconds timescale. In the case of rare events, *i.e.*, requiring much longer time scale (from several microseconds to seconds), biased MD simulations (see section III) are required, as molecule flip-flop from one to another leaflet, passive permeation or large conformational changes in membrane proteins.

## II.3. Metadynamics

### II.3.1. Introduction

Although MD simulations can apply on very large molecular systems made of hundreds of thousands of atoms at equilibrium, rare events can hardly be sufficiently sampled with the current computational facilities. In other words, complex biological processes requiring large movements of atoms and high energetic barriers to be crossed (*e.g.*, protein folding, protein domain swapping, diffusion processes, membrane crossing) are not accessible by regular MD simulations. The energy landscapes of such processes are complex hypersurfaces made of many local minima with physical relevance or not (*i.e.*, more or less populated under physiological conditions). Some minima are separated by energy barriers of a few $k_BT$ that can be easily crossed during regular MD simulations, but some others are separated by high energy barriers, namely requiring the formation of highly instable transition states. Various enhanced-sampling methods have been developed to theoretically study such biological macromolecular systems, including umbrella sampling, steered MD, adaptive force bias, adiabatic MD, metadynamics, which all allow non-equilibrium driven MD simulations [47–54]. In this chapter, only the basic concepts of metadynamics are developed.

### II.3.2. Basic principle

As other enhanced-sampling MD-based methods, metadynamics reduces the phase space with a focus given to a limited number of dimensions defined by CVs (collective variables) that are representative of the process of interest. Several CVs can be treated simultaneously, however which evolve in their own dimension. The exploration of each potential energy curve is achieved by the construction of a history-dependent potential which is the sum of Gaussian functions added to the global Hamiltonian at regular intervals.

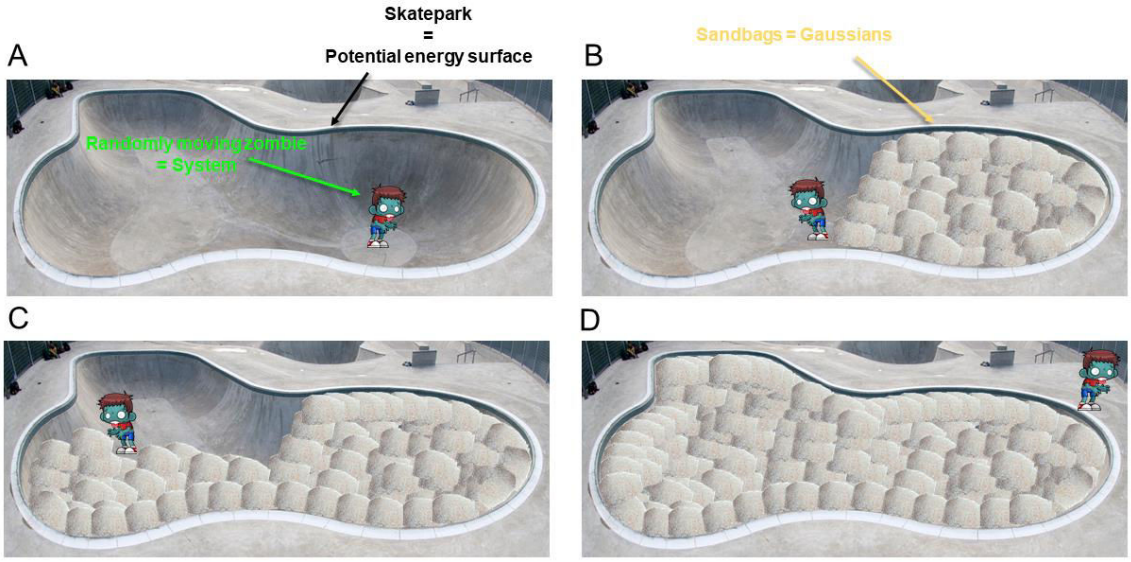



Figure 8: Schematic representation of the metadynamics process. (A) Beginning of the simulation on one CV pictured as a zombie (randomly-walking person) stuck in a well representative of an energetic minimum. (B) The energetic landscape is filled with sand bags representative of Gaussians. (C) and (D) When wells are filled by enough sand bags, the zombie escapes the wells and explores the entire empty skatepark (whole free energy landscape).

In metadynamics, the sum of Gaussians is exploited to reconstruct iteratively an estimator of the free energy profile in a one-dimensional way but for several representative reaction coordinates (the CVs). The procedure can be pictured by a zombie (randomly) walking in an empty pool copying skatepark made of several wells, or basins (Figure 8). The wells of the pool represent the energetic minima and the bumps represent the saddle points. At each step of the random walk of the zombie, a sandbag (a Gaussian) is deposited in its surrounding to fill the well. Due to its random motion, the zombie is stuck in a well until there are enough bags to escape the well (Figure 8). When a well is sufficiently filled, the zombie explores another well, and eventually the entire pool surface (Figure 8C&D), *i.e.*, exploration of the whole free energy curve of the corresponding CV. The zombie's memory about his way allows reconstructing the surface of the pool (the free energy profile).

In metadynamics, it is assumed that the time-dependent potential defined as the sum of Gaussians, deposited up to time $t$, provides an unbiased estimate of the free energy potential curve in the region explored during the corresponding MD simulation. As for other biased methods, the choice of the different CVs is key in metadynamics. The exploration of an appropriate set of CVs highly indeed accelerates the exploration of rare events; conversely, using wrong CVs requires a big computational effort with no benefit.

### II.3.3. Equations[55,56]

Let us consider a finite number of CVs named $\mathrm{S}_\alpha(\mathrm{x})$, with $\alpha = 1 \ldots \mathrm{d}$, where $\mathrm{d}$ is a small number, CVs being function of atomic coordinates $\mathrm{x}$ (see section II.3.4.1 for more details). The probability to observe value $\mathrm{s}$ for the CVs is as follows:

$$\mathrm{P(s)} = \frac{\exp\left(-\frac{\mathrm{F(s)}}{\mathrm{k_B T}}\right)}{\int \mathrm{ds} \exp\left(-\frac{\mathrm{F(s)}}{\mathrm{k_B T}}\right)} \qquad (\mathrm{Eq.}\ 36),$$

where $F(s)$ is the Helmholtz free energy.

As the CVs are supposed to comprehensively describe the phase space, the exploration of the different CVs is supposed to provide a series of microstates (molecular conformations), which constitute a canonical ensemble representative of the molecular system evolution. Thus, the free energy can be calculated from the histogram of all $N(s)$ microstates, as follows:

$$F(s) = -k_B T \cdot \log N(s) \qquad (Eq.\ 37).$$

In metadynamics, the bias potential $V(s,t)$ is added, which acts at time $t$ only on the CVs and which is given by the sum of Gaussian functions, written as follows:

$$V(s,t) = \sum_{i=1}^{t/\tau_G} \omega_G \exp\left[\sum_{\alpha=1}^{d} \frac{(s_\alpha(t) - s_\alpha(i\tau_G))^2}{2\sigma_\alpha^2}\right] \qquad (Eq.\ 38),$$

where $s_\alpha(t)$ is the value taken by the CV $S_\alpha(x)$ at time $t$, $\omega_G$ is the height of the added Gaussian, $\sigma_\alpha$ is the Gaussian width, $\tau_G$ is the frequency of deposition of the Gaussians, and $d$ is the number of CVs, which corresponds to the number of key dimensions of the system evolution.

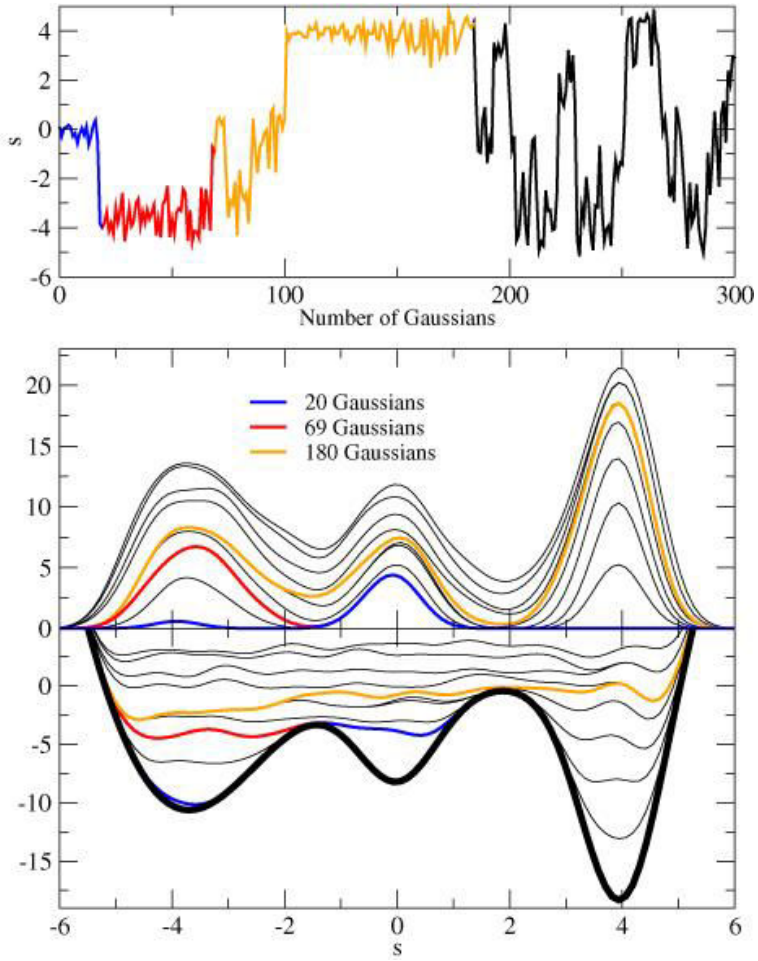



Figure 9: This figure was reproduced from[56] and it provides a typical example of a converged metadynamics simulation. The upper panel plots the CV value ($s_\alpha(t)$) *vs.* the number of Gaussians added. The middle panel represents the evolution of the bias potential $V(s,t)$ when increasing the number of Gaussians. The lower panel is the evolution of the sum of $V(s,t)$ and $\tilde{F}(s)$, estimate of the free energy which is depicted by the thick black line.

The value of the CV can be followed according to the number of added Gaussians (see the upper panel of Figure 9), as well as the $V(s,t)$ values (middle panel of Figure 9). Each jump in the CV value allows escaping from an energetic minimum. This allows an iterative and converging construction of the free energy profile along one CV (Figure 9 lower panel). The metadynamics simulation is stopped when the system becomes diffusive, meaning that the region of the energetic landscape is entirely explored. The negative of $V(s,t)$ is an estimate value the underlying free energy, named $\tilde{F}(s)$. After an infinite simulation time, namely, in practice, a sufficiently long time, one can obtain the exact free energy, $F(s)$ within a constant $C$, as follows:

$$\lim_{t\to\infty} V(s,t) = -F(s) + C \qquad (\text{Eq. } 39).$$

This equation states that $F(s)$ (equilibrium quantity) can be estimated by a non-equilibrium dynamics. The efficiency of the method scales exponentially with the number of dimensions, $d$, required for a correct description of the phase space. Metadynamics works properly only if $d$ is small. The quality of the reconstructed free energy is strongly influenced by the $\omega_G$ and $\sigma_\alpha$ , and $\tau_G$ parameters. These parameters must be carefully chosen to find the best balance between accuracy and sampling efficiency (see next section).

### II.3.4. Metadynamics set-up

#### II.3.4.1. Collective Variables

The choice of the CVs is a complex step that should be consider with care prior to run metadynamics simulations. As general rules, CVs should: i) take values that are clearly distinguishable between the initial, final and intermediate states; and ii) describe all slow events that occur during the studied process. A relatively small number is likely to limit the computational cost and to ensure the accuracy of the results.

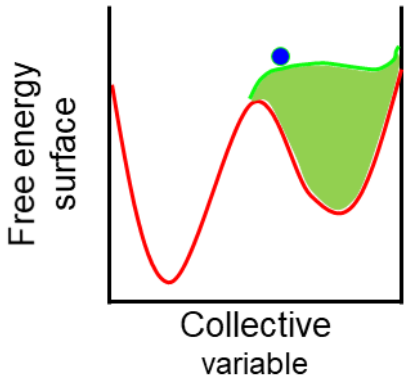



Figure 10: Representation of the potential energy surface along a key CV which is ignored in the metadynamics process. The basin where the CV value was first defined is overfilled, whereas the other basin is not explored at all.

The choice of CVs should be exhaustive, *i.e.*, ignoring a key CV may miss exploration of a key region of the energy landscape. In other words, the exploration of the transition between two potential wells (*i.e.*, two basins) connected through this ignored CV may take too long (Figure 10). The usual way to choose CVs is by trials-and-errors; it means that different combinations of CVs can be tested from expected structural differences or driving forces of the process under study. A broad range of CV can be used. The simplest CVs are the geometry-related

variables, *i.e.*, distances, torsions and angles. Nevertheless, more sophisticated and dedicated CV can also be defined, *e.g.*, linear combination of geometric variables; number of $C_\gamma$ contacts in peptide chains to follow protein folding, unfolding and aggregation[57]; helicity of protein or DNA backbone [58]; movement of a ligand along a protein channel access defined by the combination of distances to different milestones in the channel[50]; number of hydrophobic contacts; difference between experimental and calculated NMR chemical shifts [59]. The set of CVs can be thought along the metadynamics process, according to the evolution of the simulations, *e.g.*, non-convergence.

#### II.3.4.2. The $\omega_G$, $\sigma_\alpha$ and $\tau_G$ metadynamics parameters

Prior to metadynamics simulations, a preliminary unbiased MD run is usually performed, which allows estimate the shape and the width of the initial free-energy wells; this allows a guess of both $\omega_G$ and $\sigma_\alpha$ for a given CV. A general rule to choose $\sigma_\alpha$ is to use a fraction (*i.e.* $\frac{1}{2}$ or $\frac{1}{3}$) of the standard deviation of the CV fluctuation obtained from the preliminary unbiased MD run. Lower $\sigma_\alpha$ values can be used, however increasing computational cost. The $\sigma_\alpha$ values can be re-adapted along the metadynamics simulations. The Gaussian height $\omega_G$ should be chosen to avoid too abrupt changes when a new Gaussian is added. Usually the $\omega_G$ is a fraction of the $k_BT$ (*e.g.*, 0.1 kcal/mol). Nonetheless, it may be adapted according to the molecular process under study (see Chapter IV). The deposition time $\tau_G$ should be sufficiently large to accelerate exploration of the free energy landscape and sufficiently small to correctly explore each potential well of interest. It can also be determined from preliminary short free MD simulations. Indeed, $\tau_G$ is linked to the autocorrelation time of the CV. Another parameter to consider is the deposition rate $w$ of the Gaussians:

$$w = \frac{\omega_G}{\tau_G} \qquad (\text{Eq. } 40).$$

This parameter is crucial in metadynamics simulations by affecting the error and filling speed of the different basins. A balance should be found to fill the free energy landscape: i) a sufficiently large deposition rate to limit computational time; but ii) small enough to reduce the error and allow resampling many times the same zone in the CV phase space to produce a good estimate of the free energy, namely: $\tilde{F}(s) \cong F(s)$.

### II.3.5. Enhanced metadynamics methods

There are multiple extensions to metadynamics including multiple-walkers metadynamics[60] which consists of several simulations (walkers) performed in parallel which are then combined again to enhance sampling; replica-exchange metadynamics which also performs independent parallel metadynamics runs applied to the same number of replicas, which are attempting to swap by pairs at regular intervals; parallel tempering metadynamics [61], which use the replica-exchange methodology to exchange coordinates of two replicas with adjacent temperatures; or bias-exchange metadynamics (BE-MTD)[62]. Here we detail better the last method which was used in Chapter IV.

BE-MTD performs $d$ metadynamics simulation runs on $d$ replicas, at the same temperature in parallel, $d$ being the number of CVs. Each replica is biased by a metadynamics potential $V(s,t)$. At fixed time intervals $\tau$, an exchange of the bias potentials between pairs of replicas is attempted. Let us consider two CVs for which the bias potentials $V^i(s_i(x_1,t))$ and $V^j\left(s_j(x_2,t)\right)$ is acting on replicas 1 and 2 ($x_1$ and $x_2$ sets of coordinates), respectively. In BE-MTD, when the exchange is accepted, the bias potential acting on both replicas jump, as follows:

$V^i(s_i(x_1,t)) \rightarrow V^i(s_i(x_2,t))$ and *vice et versa* $V^j\left(s_j(x_2,t)\right) \rightarrow V^j\left(s_j(x_1,t)\right)$. The exchange is accepted, or not, according to the Metropolis algorithm, *i.e.*, following the probability, $p_{ij}$, to exchange bias $V^i$ with bias $V^j$:

$$p_{ij} = \min\left(1, e^{\left[V^i(s_i(x_k,t))+V^j\left(s_j(x_l,t)\right)-V^i(s_i(x_l,t))-V^j(s_j(x_k,t))\right]/k_BT}\right) \qquad (\text{Eq. } 41),$$

where $V^i(s_i(x_k,t))$ and $V^j\left(s_j(x_l,t)\right)$ are the bias potential $i$ applied on replica $k$ and the bias potential $j$ applied on replica $l$, respectively; and $V^i(s_i(x_l,t))$ and $V^j(s_j(x_k,t))$ are the bias potential i applied on replica l and the bias potential j applied on replica k, respectively (Figure 11).

The efficiency of the method is enhanced if the exchanges are not attempted too frequently (*e.g.*, every 10 ps in classical MD simulations), as replicas need sufficient time to relax under the added bias. As in metadynamics, in BE-MTD, the result of the simulation is not a free energy as a function of $d$ CVs, but $d$ low-dimensional projections of the free energy. BE-MTD does not solve the need to use a proper set of key CVs, however, it allows using more CVs at the same time, to increase the chance to use the right CVs. For instance, BE-MTD successfully rationalized the reversible folding of the tryptophane-cage[62] with eight replicas including the number of backbone-backbone hydrogen bonds.

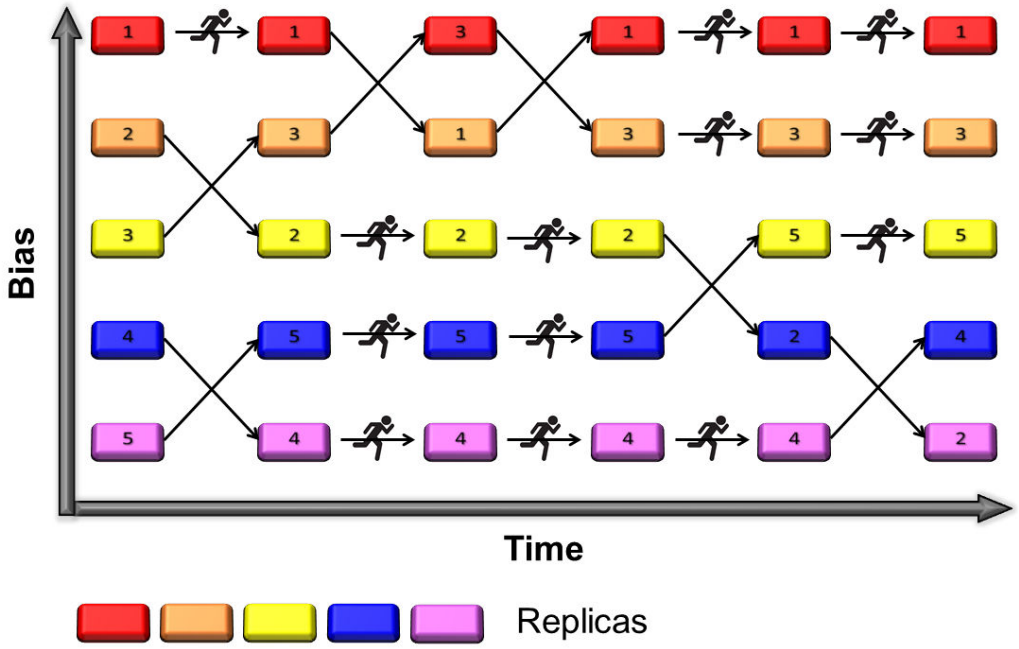



Figure 11: Scheme of the bias-exchange.

## II.4. Protein construction

### II.4.1. Principle of homology modeling (HM)

In drug discovery, High-Throughput Screening (HTS) techniques have been intensively used[63]. *In silico* methods have raised (*e.g.*, Structure-Based Drug Design - SBDD) which support HTS data[64]. Joining the efforts of these techniques has allowed gaining into structure-activity relationship and the search for novel drugs. A particular attention was paid to drug-protein interactions, *i.e.* identification of binding pocket[65,66] and development of

libraries of compounds for a three-dimensional (3D) model using the docking approach[67–72]. There exist several data bases which collect all elucidated protein structures, with or without ligands including the extensively used Protein Data Bank (PDB). However, many proteins have not been structurally resolved experimentally (*e.g.*, due to difficulties in crystallization) so far. In this case, there exist databases including SWISS-MODEL[73], Protein Model Portal[74] or Modbase[75], which predict 3D models of these unknown protein structures. These automated models are prone to inaccuracies, limiting their real use in drug discovery[76,77]. Homology Modeling (HM), also known as comparative modeling or template-based modeling, is a solid alternative to construct a 3D model of these unknown proteins by using homologous protein structures that have been resolved. HM uses all 3D structural parameters of one or several known 3D structure(s) and it aligns the sequence of the protein having no 3D structure onto these parameters [78]. Before describing HM, it is worth mentioning that confusion is often made with protein threading. Protein threading is a method aiming the same as HM, but it is used when no homologous protein structures are available, thus the method to construct a 3D model differ from HM. The Critical Assessment of protein Structure Prediction (CASP) community supports the choice of the method or software for HM [79].

### II.4.2. Homology modeling steps

HM procedure can be decomposed into several steps (Figure 12): i) identification, by sequence alignment of all sequences of 3D known structures (templates) which are homologous of the studied protein, for which no 3D structure exists; ii) generation, refinement of the protein model; iii) validation of the model; and then iii) use the constructed model in drug discovery.

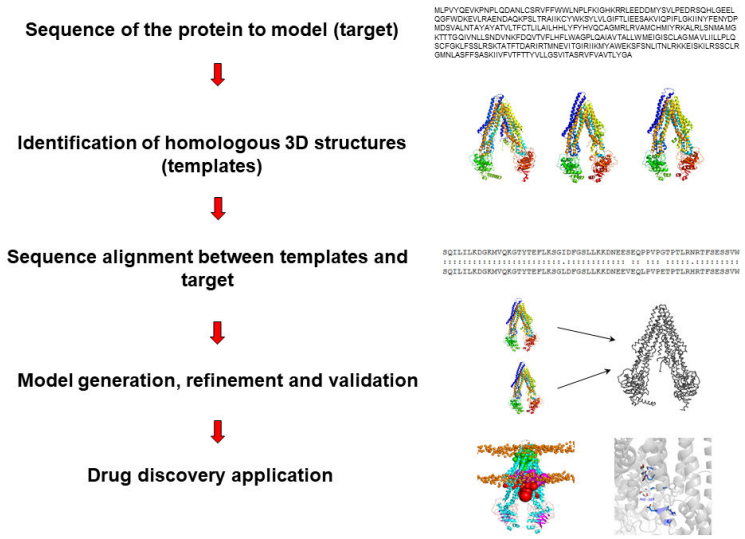



Figure 12: Scheme of homology modeling steps to build a 3D model from a protein sequence.

### II.4.3. Selection and alignment of template/target proteins

The selection of the templates will determine the quality of the final model. Indeed, the quality depends on the degree of sequence identity, so the accuracy of the method used to proceed this alignment[80]. Models with more than 50% of identities are usually appropriate for drug discovery applications; between 25 and 50%, the models can be used, however with care, to identify relevant therapeutic drug target and to design mutagenesis experiments (*e.g.*, directed mutagenesis); below 25%, the conclusions are highly questionable, although some trends can

be found[64]. Proteins can exhibit low sequence identity on specific domains but exhibit high identity on other domains. Conversely, a structure with a high sequence identity can hindered important conserved parts that are more flexible and can lead to inaccuracies in HM[81].

### II.4.4. Alignment algorithms

There are several pairwise alignment tools for optimal search between pairs of sequences (*e.g.*, FASTA[82], BLAST[83], PSI-BLAST[84]). The alignment can be complicated if there are gaps, as for ABC transporters, which contain highly flexible domains [85]. The alignment tools use either global or local alignments (Figure 13); two well-known algorithms have been developed, namely the Needleman-Wunsch[86] and Smith-Waterman[87] algorithms for global and local alignments[88], respectively. As the former algorithm takes the end-to-end sequence into consideration, problems may arise from gaps, or important conserved regions. The latter algorithm aligns small domains and allows more flexibility in the alignment process. Better sensitivity can be achieved by multiple sequence alignments, *e.g.*, as made by the PSI-BLAST[84,89] tool.

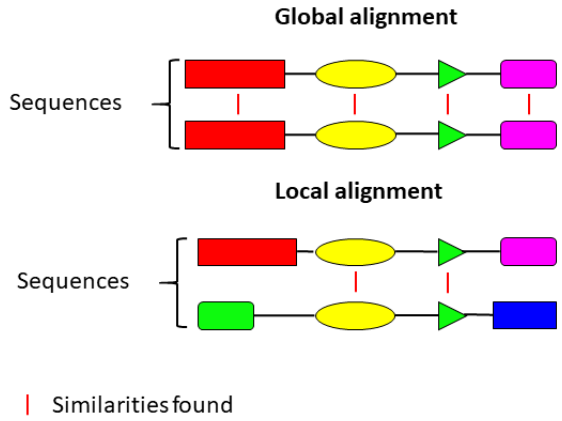



Figure 13: Scheme of the differences between the global and local alignment.

### II.4.5. Scoring matrices

To discriminate the sequences and to select the best homologues after the alignments, scoring matrices can be used. There are multiple scoring matrices that should be carefully chosen according to what closely-homolog means for the desired accuracy. The most common matrices types are the PAM (Point Accepted Mutation) and BLOSUM (BLOcks SUbstitution Matrix) matrices.

In PAM matrices, the columns and rows are constituted by the twenty indispensable amino acids. Each entry in the matrix indicates the probability that one amino acid has been replaced by another one (*i.e.*, mutation). There exist different PAM matrices (*e.g.*, PAM250[90]), which differ from the evolutionary interval (*e.g.*, from bacteria to human, or from fish to human) chosen to evaluate these probabilities. PAM are based on comparison with highly related sequences, *i.e.*, more than 85%. BLOSUM matrices can be estimated with lower identity, and they focus on specific sequence motifs, *e.g.*, conserved regions along the evolution. The different BLOSUM matrices differ from the percentage of identity between sequences. For instance, in BLOSUM62 the sequences to build the matrix exhibit less than 62% identity[91].

There are other tools including BLAST[92] which provides various scores to confirm how good are the homologous proteins found by the alignment process. A key score in BLAST is the E-

value, which calculates the probability to obtain identity between two portions of sequences (excluding gaps) only by chance. In other words, the lower the E-value (the closer to zero), the more relevant the homologous proteins. The E-value should nonetheless be carefully considered in the presence of large gaps (or extra loops) which may multiply the number of short identical sequences, thus artificially decreasing the E-value (good homology).

### II.4.6. Model generation and refinement

There are two main types of methods to construct a model[93–95], namely those based on rigid-body superimposition and those based on geometric restraints with respect to the resolved 3D structures (*e.g.*, X-ray, NMR, Cryo-EM). The aim of these two methods is mainly to fit the conserved domains.

In the rigid-body superimposition, there are two ways to construct the model, either by fragment assembly or by segment matching[96]. The former way relies on the assembly of conserved fragments identified from the previous sequence-alignment step, *e.g.*, in ABC transporters, the NBDs are relatively well-conserved regions. Once the well-conserved and well-defined regions are constructed, the unresolved region are considered by the algorithm. The different fragment-assembly methods differ one from each other by the way to treat the unresolved regions, which are usually constructed from fragment libraries[97,98]. In the latter construction way, the target protein is divided into segments that match structures from the Protein Data Bank. The alignment is then achieved on each segment to assess the similarity with the template structures (*e.g.*, comparison of the α-carbon coordinates).

The geometric-restraint methods (*e.g.*, used in the MODELLER software[75,99]) are based on the NMR method to generate a 3D structure. It uses target/template alignments (*i.e.*, regions that are common between the template and the target) to generate spatial restraints that are applied to the internal coordinates of the proteins (*i.e.*, backbone coordinates, dihedral angles). The spatial restraints are then used for a global optimization procedure to minimize the heavy atoms in the model. However, the fitting between the template and the target is not always accurate, thus an iterative procedure is usually used to refine the alignment[100].

A critical point in HM is to model the unresolved regions, mainly loops. This step constitutes the main source of errors in HM. The major conformations of loops are identified and fitted based on their beginning and ending parts. The possible conformations are predicted either from *ab initio* methods or from database-searching techniques. The former (*ab initio* based) methods refine the conformations of loops by a score, *i.e.*, an energy obtained by space sampling and global energy minimization, or Monte Carlo sampling as in the I-TASSER software[101,102]. The latter methods search chains in databases to fit to the loop regions, with homology lower than 25% but elucidated with resolution higher than 2.5 Å. Following these different steps, all side-chain conformations are predicted to complete the model.

### II.4.7. Model validation

The errors in HM mainly come from: i) the packing of the side-chains due to sequence divergence (*i.e.*, divergence between target/template); ii) misalignments or shifts in the aligned residues; iii) the gaps in the aligned regions, often located in loops; iv) use of incorrect templates (*i.e.*, identity lower than 25%).

The quality of the model is assessed by different ways. Webservers are available to check the stereochemistry as with PROCHECK[103] (analysis of the stereochemistry residue by residue geometry and overall geometry), AQUA[104] or WHATCHECK[105]. The quality of the

secondary structure can be assessed by the Ramachandran plot[106,107], which can be further refined by optimizing and relaxing the protein to remove steric clashes and correct local organization. Other analyses are carried out to evaluate the accuracy in the protein fold. A simple method to ensure the validity of the model is to calculate the Root-Mean-Square Deviation (RMSD) between the template and the final model, which is expected to be lower than $2$ Å. This approach remains insufficient, especially in the presence of many loops. The 3D identity between the templates and the target protein can also be calculated, *e.g.*, by using PDBeFold[108].

A further step in evaluating accuracy of HM models is to calculate 3D scores. The folding can be evaluated by calculating the Z-score either on the whole protein or on domains (*e.g.*, implemented in VERIFY3D[109], PROSAII[110] or ANOLEA[111]). It is calculated as follows:

$$Z = \frac{(x - \mu)}{\sigma} \qquad (\text{Eq. } 42),$$

where $x$ and $\mu$ are the raw scores (*e.g.*, energy of the structure obtained from physical-based energy functions, QMEAN and QMEAN Z-score [112]) of the constructed model and an average of all the templates used for the HM procedure, respectively, and $\sigma$ is the corresponding standard deviation. The TM-score (Template Modeling score) is also important to assess the similarity between the templates and the 3D model:

$$TM - score = max\left[\frac{1}{L_{aligned}} \sum_{i}^{L_{aligned}} \frac{1}{1 + \left(\frac{d_i}{d_0(L_{\text{3D model}})}\right)^2}\right] \qquad (\text{Eq. } 43),$$

where $L_{\text{3D model}}$ and $L_{aligned}$ are the 3D model sequence length and the length of the aligned sequence region of the template, respectively; $d_i$ is the distance between the $ith$ residue of the 3D model and the $ith$ residue of the template; and $d_0$ is a distance normalization term. Softwares have their own score based on the method used to build the 3D model, *e.g.*, the C-score in like I-TASSER[102] which ranges from -5 to 2, with the higher the score the better the confidence in the 3D model. The quality of the model can also be assessed by using experimental data as obtained by directed mutagenesis.

To follow specific distances (*e.g.*, hydrogen bonds, salt bridges or disulfide bonds) or angles and dihedrals can be achieved by simple analyses. This allows a better evaluation of the secondary structure, *e.g.*, local distortion of α-helices (*e.g.*, DSSP program). Evolution of the Ramachandran plot along MD simulation time is likely to document on the protein secondary structure. The hydrogen bond network can be followed to characterize the evolution of the secondary structure, but also specific interactions, *e.g.*, interactions between α-helices, interactions in hydrophobic cavities of proteins[113], interactions between proteins and surrounding environment, mainly with water[114]. Occurrence of hydrogen bonds and their distribution (or their number) are evaluated by geometrical criteria possibly user-defined, default values being *e.g.*, in the GROMACS package:

$$r \leq r_{HB} = 0.35\ nm$$

$$\alpha \leq \alpha_{HB} = 30°$$

where $r_{HB} = 0.35\ nm$ is the first minimum of the radial distribution function (RDF) of SPC water.

## II.5. Molecular docking and related analyses

Molecular docking is a method used to predict the spatial positioning and arrangement of ligands in binding site of receptors; protein-protein interaction can also be studied however with more attention paid to the methodology and accuracy. Drug design and clinical research have gained much progress thanks to high-throughput protein identification and the structural and protein-ligand information obtained by X-ray, NMR or cryo-EM experiments. In this perspective, computational methods including molecular docking have appeared as a promising support[115,116], allowing development of virtual screening methods[117–120]. The advantages of molecular docking are its speed to guess drug-protein interaction, and the related free energy of interaction (score, see below).

### II.5.1. Search algorithms

The first step of molecular docking procedures is to define a docking space, namely a portion of the protein, which is the binding site if it is known, where the best positions (poses) of the ligand are randomly searched. The second step is to find a series of ligand-protein arrangements, defining the position of the ligand in the binding site, thus the conformation of the ligand-protein complex; each position is named a pose. The algorithms to evaluate the poses can be either of rigid-body type or of flexible-ligand type.

The former type of algorithms is the simplest implementation, in which the ligand and the receptor are rigid (*i.e.* no movement allowed in torsions, angles or bonds). Only the six molecular degrees of freedom (translation and rotation) are explored for the ligand within the docking space. Although this type of method is fast, the restriction of degrees of freedom is often too constraining, leading to miss important poses or conformational adaptation of the ligand. An alternative (latter type of algorithms) is to allow the ligand to be flexible, according to user-defined rules. Some residues of the docking space can also be flexible (with however a limited number of degrees of freedom in this space). The ligand flexibility is ensured either by systematic or by stochastic algorithms.

The systematic algorithms explore all degrees of freedom of a ligand by: i) a systematic conformational search, in which all rotatable bonds (defining a torsion angle) are scanned from 0° to 360° with a fixed increment; ii) ligand structure fragmentation as in place-and-join method, in which individual docking are made for key fragments of the ligand, providing fragment-poses linked back to determine a pose for the ligand; iii) incremental methods, in which the ligand is docked as a rigid core, and flexibility of key fragments is explored afterwards to refine the pose (*e.g.*, in DOCK[121] or flexX[122] programs); and iv) using libraries of pre-generated conformations (*e.g.*, in FLOG[123]).

The stochastic search algorithms are sampling the space through random changes in the ligand conformation, either by: i) Monte-Carlo that consider a Boltzmann probability function to define the acceptance of a new pose (with a random selection with a Metropolis condition); ii) the genetic algorithm which is inspired by natural selection, with the poses described as “genes” and the best poses “mate” to generate offspring (this method converges faster than Monte-Carlo methods); or iii) simulated annealing, in which the heating allows to cross energetic barriers and the cooling allows to reach minima.

### II.5.2. Scoring function

The different poses are discriminated by the scoring functions, which are expressed in terms of Gibbs energy of binding ($\Delta G_{bind}$) including various energetic terms related to the solvent

effect, the conformational changes that occur in the protein and the ligand when flexibility is allowed, the protein-ligand interaction itself, the change in internal rotation, and the change in vibrational modes[124].

The different terms can be obtained on the different poses by various ways: i) by defining a grid on which the MM (force field) potential energy is provided at each crossing point and, *e.g.*, protein-ligand interaction energies are obtained by a simple multiplication; ii) by QM calculations, mainly semi-empirical methods to avoid huge computational effort; iii) by empirical scoring functions, which pre-define energetic contributions for different types of contacts, *e.g.*, hydrophobic contacts, H-bonding, $\pi - \pi$ stacking; and iv) by knowledge-based scoring function, *i.e.*, by correlation of the poses with known structural data about ligand-protein interactions.

### II.5.3. Docking validation

The quality and thus significance of docking results can be assessed either by comparison with respect to experimental data[125,126], or by re-docking and calculating the RMSD differences between the different docking runs for a given pose, which should be lower than 2 Å. Another validation analysis consists in docking negative controls (decoys), *i.e.*, ligands which are known not to bind[127]. Docking validation is usually evaluated with a Receiver Operating Characteristic (ROC) curve. The significance of docking can also be evaluated with respect to related protein-ligand X-ray crystal structures[128] or from libraries of decoys[127].

## References


[1] R.S. Paton, J.M. Goodman, Hydrogen bonding and pi-stacking: how reliable are force fields? A critical evaluation of force field descriptions of nonbonded interactions, J Chem Inf Model. 49 (2009) 944–955. doi:10.1021/ci900009f.

[2] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, The Journal of Chemical Physics. 98 (1993) 10089–10092. doi:10.1063/1.464397.

[3] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins, J. Phys. Chem. B. 102 (1998) 3586–3616. doi:10.1021/jp973084f.

[4] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, J. Chem. Theory Comput. 11 (2015) 3696–3713. doi:10.1021/acs.jctc.5b00255.

[5] J. Huang, A.D. MacKerell, CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data, J Comput Chem. 34 (2013) 2135–2145. doi:10.1002/jcc.23354.

[6] W. Huang, Z. Lin, W.F. van Gunsteren, Validation of the GROMOS 54A7 Force Field with Respect to β-Peptide Folding, J Chem Theory Comput. 7 (2011) 1237–1243. doi:10.1021/ct100747y.

[7] J.P.M. Jämbeck, A.P. Lyubartsev, Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids, J. Phys. Chem. B. 116 (2012) 3164–3179. doi:10.1021/jp212503e.

[8] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids, J. Am. Chem. Soc. 118 (1996) 11225–11236. doi:10.1021/ja9621760.

[9] M. Möllhoff, U. Sternberg, Molecular mechanics with fluctuating atomic charges – a new force field with a semi-empirical charge calculation, J Mol Model. 7 (2001) 90–102. doi:10.1007/s008940100008.

[10] S.J. Marrink, H.J. Risselada, S. Yefimov, D.P. Tieleman, A.H. de Vries, The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations, The Journal of Physical Chemistry B. 111 (2007) 7812–7824. doi:10.1021/jp071097f.

[11] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J.W. Ponder, P. Ren, Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins, J. Chem. Theory Comput. 9 (2013) 4046–4063. doi:10.1021/ct4003702.

[12] S. Patel, A.D. Mackerell, C.L. Brooks, CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model, J Comput Chem. 25 (2004) 1504–1514. doi:10.1002/jcc.20077.

[13] C.M. Baker, Polarizable force fields for molecular dynamics simulations of biomolecules, Wiley Interdisciplinary Reviews: Computational Molecular Science. 5 (2015) 241–254. doi:10.1002/wcms.1215.

[14] J. Shearer, S. Khalid, Communication between the leaflets of asymmetric membranes revealed from coarse-grain molecular dynamics simulations, Scientific Reports. 8 (2018) 1805. doi:10.1038/s41598-018-20227-1.

[15] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197. doi:10.1021/ja00124a002.

[16] C. Oostenbrink, A. Villa, A.E. Mark, W.F. van Gunsteren, A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6, J Comput Chem. 25 (2004) 1656–1676. doi:10.1002/jcc.20090.

[17] J.P.M. Jämbeck, A.P. Lyubartsev, An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes, J. Chem. Theory Comput. 8 (2012) 2938–2948. doi:10.1021/ct300342n.

[18] J.B. Klauda, R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A.D. MacKerell, R.W. Pastor, Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types, J Phys Chem B. 114 (2010) 7830–7843. doi:10.1021/jp101759q.

[19] C.I. Bayly, P. Cieplak, W. Cornell, P.A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, J. Phys. Chem. 97 (1993) 10269–10280. doi:10.1021/j100142a004.

[20] I. Ermilova, A.P. Lyubartsev, Extension of the Slipids Force Field to Polyunsaturated Lipids, J. Phys. Chem. B. 120 (2016) 12826–12842. doi:10.1021/acs.jpcb.6b05422.

[21] I. Chandrasekhar, M. Kastenholz, R.D. Lins, C. Oostenbrink, L.D. Schuler, D.P. Tieleman, W.F. van Gunsteren, A consistent potential energy parameter set for lipids: dipalmitoylphosphatidylcholine as a benchmark of the GROMOS96 45A3 force field, Eur. Biophys. J. 32 (2003) 67–77. doi:10.1007/s00249-002-0269-4.

[22] R. Tjörnhammar, O. Edholm, Reparameterized United Atom Model for Molecular Dynamics Simulations of Gel and Fluid Phosphatidylcholine Bilayers, J Chem Theory Comput. 10 (2014) 5706–5715. doi:10.1021/ct500589z.

[23] S.-W. Chiu, S.A. Pandit, H.L. Scott, E. Jakobsson, An Improved United Atom Force Field for Simulation of Mixed Lipid Bilayers, J. Phys. Chem. B. 113 (2009) 2748–2763. doi:10.1021/jp807056c.

[24] C.J. Dickson, B.D. Madej, Å.A. Skjevik, R.M. Betz, K. Teigen, I.R. Gould, R.C. Walker, Lipid14: The Amber Lipid Force Field, J. Chem. Theory Comput. 10 (2014) 865–879. doi:10.1021/ct4010307.

[25] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (2004) 1157–1174. doi:10.1002/jcc.20035.

[26] O. Berger, O. Edholm, F. Jähnig, Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature, Biophysical Journal. 72 (1997) 2002–2013. doi:10.1016/S0006-3495(97)78845-3.

[27] K. Pluhackova, S.A. Kirsch, J. Han, L. Sun, Z. Jiang, T. Unruh, R.A. Böckmann, A Critical Comparison of Biomembrane Force Fields: Structure and Dynamics of Model DMPC, POPC, and POPE Bilayers, J. Phys. Chem. B. 120 (2016) 3888–3903. doi:10.1021/acs.jpcb.6b01870.

[28] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, The Journal of Chemical Physics. 126 (2007) 014101. doi:10.1063/1.2408420.

[29] H.J.C. Berendsen, Transport Properties Computed by Linear Response through Weak Coupling to a Bath, in: Computer Simulation in Materials Science, Springer, Dordrecht, 1991: pp. 139–155. doi:10.1007/978-94-011-3546-7_7.

[30] W.G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, Phys. Rev. A. 31 (1985) 1695–1697. doi:10.1103/PhysRevA.31.1695.

[31] D.J. Coughtrie, D.P. Tew, The Nosé–Hoover looped chain thermostat for low temperature thawed Gaussian wave-packet dynamics, The Journal of Chemical Physics. 140 (2014) 194106. doi:10.1063/1.4875517.

[32] R.J. Loncharich, B.R. Brooks, R.W. Pastor, Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide, Biopolymers. 32 (1992) 523–535. doi:10.1002/bip.360320508.

[33] H.C. Andersen, Molecular dynamics simulations at constant pressure and/or temperature, The Journal of Chemical Physics. 72 (1980) 2384–2393. doi:10.1063/1.439486.

[34] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, The Journal of Chemical Physics. 81 (1984) 3684–3690. doi:10.1063/1.448118.

[35] M. Parrinello, A. Rahman, Crystal Structure and Pair Potentials: A Molecular-Dynamics Study, Phys. Rev. Lett. 45 (1980) 1196–1199. doi:10.1103/PhysRevLett.45.1196.

[36] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, Journal of Applied Physics. 52 (1981) 7182–7190. doi:10.1063/1.328693.

[37] M. Parrinello, A. Rahman, Strain fluctuations and elastic constants, The Journal of Chemical Physics. 76 (1982) 2662–2666. doi:10.1063/1.443248.

[38] P. Koehl, Electrostatics calculations: latest methodological advances, Current Opinion in Structural Biology. 16 (2006) 142–151. doi:10.1016/j.sbi.2006.03.001.

[39] B. Guillot, A reappraisal of what we have learnt during three decades of computer simulations on water, Journal of Molecular Liquids. 101 (2002) 219–260. doi:10.1016/S0167-7322(02)00094-6.

[40] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, Interaction Models for Water in Relation to Protein Hydration, in: B. Pullman (Ed.), Intermolecular Forces, Springer Netherlands, 1981: pp. 331–342. doi:10.1007/978-94-015-7658-1_21.

[41] H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, The missing term in effective pair potentials, J. Phys. Chem. 91 (1987) 6269–6271. doi:10.1021/j100308a038.

[42] W.L. Jorgensen, J.D. Madura, Quantum and statistical mechanical studies of liquids. 25. Solvation and conformation of methanol in water, J. Am. Chem. Soc. 105 (1983) 1407–1413. doi:10.1021/ja00344a001.

[43] M.W. Mahoney, W.L. Jorgensen, A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions, The Journal of Chemical Physics. 112 (2000) 8910–8922. doi:10.1063/1.481505.

[44] P. Mark, L. Nilsson, Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K, J. Phys. Chem. A. 105 (2001) 9954–9960. doi:10.1021/jp003020w.

[45] H. Nada, J.P.J.M. van der Eerden, An intermolecular potential model for the simulation of ice and water near the melting point: A six-site model of H2O, The Journal of Chemical Physics. 118 (2003) 7401–7413. doi:10.1063/1.1562610.

[46] M.S. Lee, F.R. Salsbury, M.A. Olson, An efficient hybrid explicit/implicit solvent method for biomolecular simulations, J Comput Chem. 25 (2004) 1967–1978. doi:10.1002/jcc.20119.

[47] M. Moradi, E. Tajkhorshid, Mechanistic picture for conformational transition of a membrane transporter at atomic resolution, PNAS. 110 (2013) 18916–18921. doi:10.1073/pnas.1313202110.

[48] J. Li, E. Tajkhorshid, A gate-free pathway for substrate release from the inward-facing state of the Na+-galactose transporter, Biochimica et Biophysica Acta (BBA) - Biomembranes. 1818 (2012) 263–271. doi:10.1016/j.bbamem.2011.09.011.

[49] I. Bisha, A. Rodriguez, A. Laio, A. Magistrato, Metadynamics simulations reveal a Na+ independent exiting path of galactose for the inward-facing conformation of vSGLT, PLoS Comput. Biol. 10 (2014) e1004017. doi:10.1371/journal.pcbi.1004017.

[50] M. Paloncýová, V. Navrátilová, K. Berka, A. Laio, M. Otyepka, Role of Enzyme Flexibility in Ligand Access and Egress to Active Site: Bias-Exchange Metadynamics Study of 1,3,7-Trimethyluric Acid in Cytochrome P450 3A4, J. Chem. Theory Comput. 12 (2016) 2101–2109. doi:10.1021/acs.jctc.6b00075.

[51] M. Moradi, E. Tajkhorshid, Computational Recipe for Efficient Description of Large-Scale Conformational Changes in Biomolecular Systems, J. Chem. Theory Comput. 10 (2014) 2866–2880. doi:10.1021/ct5002285.

[52] G. Bussi, F.L. Gervasio, A. Laio, M. Parrinello, Free-Energy Landscape for β Hairpin Folding from Combined Parallel Tempering and Metadynamics, J. Am. Chem. Soc. 128 (2006) 13435–13441. doi:10.1021/ja062463w.

[53] Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, Chemical Physics Letters. 314 (1999) 141–151. doi:10.1016/S0009-2614(99)01123-9.

[54] A. Noy, A. Pérez, C.A. Laughton, M. Orozco, Theoretical study of large conformational transitions in DNA: the B↔A conformational change in water and ethanol/water, Nucleic Acids Res. 35 (2007) 3330–3338. doi:10.1093/nar/gkl1135.

[55] A. Laio, M. Parrinello, Escaping free-energy minima, PNAS. 99 (2002) 12562–12566. doi:10.1073/pnas.202427399.

[56] A. Laio, F.L. Gervasio, Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science, Rep. Prog. Phys. 71 (2008) 126601. doi:10.1088/0034-4885/71/12/126601.

[57] C. Chiu, S. Singh, J.J. de Pablo, Effect of Proline Mutations on the Monomer Conformations of Amylin, Biophysical Journal. 105 (2013) 1227–1235. doi:10.1016/j.bpj.2013.07.029.

[58] L. Jin, F. Ye, D. Zhao, S. Chen, K. Zhu, M. Zheng, R.-W. Jiang, H. Jiang, C. Luo, Metadynamics Simulation Study on the Conformational Transformation of HhaI Methyltransferase: An Induced-Fit Base-Flipping Hypothesis, Biomed Res Int. 2014 (2014). doi:10.1155/2014/304563.

[59] D. Granata, C. Camilloni, M. Vendruscolo, A. Laio, Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics, PNAS. 110 (2013) 6817–6822. doi:10.1073/pnas.1218350110.

[60] P. Raiteri, A. Laio, F.L. Gervasio, C. Micheletti, M. Parrinello, Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics, J. Phys. Chem. B. 110 (2006) 3533–3539. doi:10.1021/jp054359r.

[61] G. Bussi, F.L. Gervasio, A. Laio, M. Parrinello, Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics, J. Am. Chem. Soc. 128 (2006) 13435–13441. doi:10.1021/ja062463w.

[62] S. Piana, A. Laio, A Bias-Exchange Approach to Protein Folding, J. Phys. Chem. B. 111 (2007) 4553–4559. doi:10.1021/jp067873l.

[63] M. Congreve, C.W. Murray, T.L. Blundell, Structural biology and drug discovery, Drug Discov. Today. 10 (2005) 895–907. doi:10.1016/S1359-6446(05)03484-7.

[64] A. Hillisch, L.F. Pineda, R. Hilgenfeld, Utility of homology models in the drug discovery process, Drug Discov. Today. 9 (2004) 659–669. doi:10.1016/S1359-6446(04)03196-4.

[65] T.L. Blundell, B.L. Sibanda, R.W. Montalvão, S. Brewerton, V. Chelliah, C.L. Worth, N.J. Harmer, O. Davies, D. Burke, Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery, Philos. Trans. R. Soc. Lond., B, Biol. Sci. 361 (2006) 413–423. doi:10.1098/rstb.2005.1800.

[66] T.A. Halgren, Identifying and Characterizing Binding Sites and Assessing Druggability, J. Chem. Inf. Model. 49 (2009) 377–389. doi:10.1021/ci800324m.

[67] X.-Y. Meng, H.-X. Zhang, M. Mezei, M. Cui, Molecular Docking: A powerful approach for structure-based drug discovery, Curr Comput Aided Drug Des. 7 (2011) 146–157. doi:10.2174/157340911795677602.

[68] G. Wang, W. Zhu, Molecular docking for drug discovery and development: a widely used approach but far from perfect, Future Med Chem. 8 (2016) 1707–1710. doi:10.4155/fmc-2016-0143.

[69] R.E. Amaro, J. Baudry, J. Chodera, Ö. Demir, J.A. McCammon, Y. Miao, J.C. Smith, Ensemble Docking in Drug Discovery, Biophysical Journal. 114 (2018) 2271–2278. doi:10.1016/j.bpj.2018.02.038.

[70] A.J.W. Orry, R.A. Abagyan, C.N. Cavasotto, Structure-based development of target-specific compound libraries, Drug Discov. Today. 11 (2006) 261–266. doi:10.1016/S1359-6446(05)03717-7.

[71] A.C. Anderson, The process of structure-based drug design, Chem. Biol. 10 (2003) 787–797. doi:10.1016/j.chembiol.2003.09.002.

[72] J. Weigelt, L.D. McBroom-Cerajewski, M. Schapira, Y. Zhao, C.H. Arrowmsmith, Structural genomics and drug discovery: all in the family, Current Opinion in Chemical Biology. 12 (2008) 32–39. doi:10.1016/j.cbpa.2008.01.045.

[73] S. Bienert, A. Waterhouse, T.A.P. de Beer, G. Tauriello, G. Studer, L. Bordoli, T. Schwede, The SWISS-MODEL Repository—new features and functionality, Nucleic Acids Res. 45 (2017) D313–D319. doi:10.1093/nar/gkw1132.

[74] K. Arnold, F. Kiefer, J. Kopp, J.N.D. Battey, M. Podvinec, J.D. Westbrook, H.M. Berman, L. Bordoli, T. Schwede, The Protein Model Portal, J. Struct. Funct. Genomics. 10 (2009) 1–8. doi:10.1007/s10969-008-9048-5.

[75] U. Pieper, N. Eswar, H. Braberg, M.S. Madhusudhan, F.P. Davis, A.C. Stuart, N. Mirkovic, A. Rossi, M.A. Marti-Renom, A. Fiser, B. Webb, D. Greenblatt, C.C. Huang, T.E. Ferrin, A. Sali, MODBASE, a database of annotated comparative protein structure models, and associated resources, Nucleic Acids Res. 32 (2004) D217-222. doi:10.1093/nar/gkh095.

[76] J.A.R. Dalton, R.M. Jackson, An evaluation of automated homology modelling methods at low target template sequence similarity, Bioinformatics. 23 (2007) 1901–1908. doi:10.1093/bioinformatics/btm262.

[77] D. Fischer, Servers for protein structure prediction, Curr. Opin. Struct. Biol. 16 (2006) 178–182. doi:10.1016/j.sbi.2006.03.004.

[78] C.N. Cavasotto, S.S. Phatak, Homology modeling in drug discovery: current trends and applications, Drug Discovery Today. 14 (2009) 676–683. doi:10.1016/j.drudis.2009.04.006.

[79] J. Moult, J.T. Pedersen, R. Judson, K. Fidelis, A large-scale experiment to assess protein structure prediction methods, Proteins: Structure, Function, and Bioinformatics. 23 (1995) ii–iv. doi:10.1002/prot.340230303.

[80] K. Ginalski, Comparative modeling for protein structure prediction, Current Opinion in Structural Biology. 16 (2006) 172–177. doi:10.1016/j.sbi.2006.02.003.

[81] P. Larsson, B. Wallner, E. Lindahl, A. Elofsson, Using multiple templates to improve quality of homology models in automated homology modeling, Protein Sci. 17 (2008) 990–1002. doi:10.1110/ps.073344908.

[82] D.J. Lipman, W.R. Pearson, Rapid and sensitive protein similarity searches, Science. 227 (1985) 1435–1441. doi:10.1126/science.2983426.

[83] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410. doi:10.1016/S0022-2836(05)80360-2.

[84] M. Bhagwat, L. Aravind, PSI-BLAST Tutorial, Methods Mol. Biol. 395 (2007) 177-186. doi:10.1007/978-1-59745-514-5_10.

[85] Z. Zhang, J. Chen, Atomic Structure of the Cystic Fibrosis Transmembrane Conductance Regulator, Cell. 167 (2016) 1586-1597.e9. doi:10.1016/j.cell.2016.11.014.

[86] S.B. Needleman, C.D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins, Journal of Molecular Biology. 48 (1970) 443–453. doi:10.1016/0022-2836(70)90057-4.

[87] T.F. Smith, M.S. Waterman, Comparison of biosequences, Advances in Applied Mathematics. 2 (1981) 482–489. doi:10.1016/0196-8858(81)90046-4.

[88] V.O. Polyanovsky, M.A. Roytberg, V.G. Tumanyan, Comparative analysis of the quality of a global algorithm and a local algorithm for alignment of two sequences, Algorithms Mol Biol. 6 (2011) 25. doi:10.1186/1748-7188-6-25.

[89] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs., Nucleic Acids Res. 25 (1997) 3389–3402. doi:10.1093/nar/25.17.3389.

[90] M.O. Dayhoff, R.M. Schwartz, R.M. and B.C. Orcutt, A model of evolutionary change in proteins, in: In Atlas of Protein Sequence and Structure, M.O. Edition, Atlas of Protein Sequence and Structure, Natl. Biomed. Res. Found., 5(3) (1978) 345-352.

[91] S. Henikoff, J.G. Henikoff, Amino acid substitution matrices from protein blocks., Proc Natl Acad Sci U S A. 89 (1992) 10915–10919. doi:10.1073/pnas.89.22.10915.

[92] T. Madden, The BLAST Sequence Analysis Tool, (2013). The NCBI Handbook. 2nd edition. Bethesda (MD): National Center for Biotechnology Information (US); 2013-.

[93] D. Baker, A. Sali, Protein structure prediction and structural genomics, Science. 294 (2001) 93–96. doi:10.1126/science.1065659.

[94] Y. Zhang, Progress and challenges in protein structure prediction, Curr. Opin. Struct. Biol. 18 (2008) 342–348. doi:10.1016/j.sbi.2008.02.004.

[95] A. Fiser, Template-Based Protein Structure Modeling, Methods Mol Biol. 673 (2010) 73–94. doi:10.1007/978-1-60761-842-3_6.

[96] M. Levitt, Accurate modeling of protein conformation by automatic segment matching, Journal of Molecular Biology. 226 (1992) 507–533. doi:10.1016/0022-2836(92)90964-L.

[97] B. Wallner, A. Elofsson, All are not equal: a benchmark of different homology modeling programs, Protein Sci. 14 (2005) 1315–1327. doi:10.1110/ps.041253405.

[98] T. Bohnuud, L. Luo, S.J. Wodak, A.M.J.J. Bonvin, Z. Weng, S. Vajda, O. Schueler-Furman, D. Kozakov, A benchmark testing ground for integrating homology modeling and protein docking, Proteins. 85 (2017) 10–16. doi:10.1002/prot.25063.

[99] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815. doi:10.1006/jmbi.1993.1626.

[100] B. John, A. Sali, Comparative protein structure modeling by iterative alignment, model building and model assessment, Nucleic Acids Res. 31 (2003) 3982–3992. doi:10.1093/ar/gkg460.

[101] Y. Wang, J. Virtanen, Z. Xue, Y. Zhang, I-TASSER-MR: automated molecular replacement for distant-homology proteins using iterative fragment assembly and

progressive sequence truncation, Nucleic Acids Res. 45 (2017) W429–W434. doi:10.1093/nar/gkx349.

[102] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER Suite: protein structure and function prediction, Nat Meth. 12 (2015) 7–8. doi:10.1038/nmeth.3213.

[103] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, Journal of Applied Crystallography. 26 (1993) 283–291. doi:10.1107/S0021889892009944.

[104] R.A. Laskowski, J.A. Rullmannn, M.W. MacArthur, R. Kaptein, J.M. Thornton, AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, J. Biomol. NMR. 8 (1996) 477–486. doi:10.1007/BF00228148.

[105] R.W.W. Hooft, G. Vriend, C. Sander, E.E. Abola, Errors in protein structures, Nature. 381 (1996) 272–272. doi:10.1038/381272a0.

[106] G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, Stereochemistry of polypeptide chain configurations, Journal of Molecular Biology. 7 (1963) 95–99. doi:10.1016/S0022-2836(63)80023-6.

[107] S.A. Hollingsworth, P.A. Karplus, A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins, Biomol Concepts. 1 (2010) 271–283. doi:10.1515/BMC.2010.022.

[108] E. Krissinel, K. Henrick, Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, Acta Crystallogr. D Biol. Crystallogr. 60 (2004) 2256–2268. doi:10.1107/S0907444904026460.

[109] D. Eisenberg, R. Lüthy, J.U. Bowie, VERIFY3D: assessment of protein models with three-dimensional profiles, Meth. Enzymol. 277 (1997) 396–404. doi:10.1016/S0076-6879(97)77022-8.

[110] M.J. Sippl, Recognition of errors in three-dimensional structures of proteins, Proteins. 17 (1993) 355–362. doi:10.1002/prot.340170404.

[111] F. Melo, D. Devos, E. Depiereux, E. Feytmans, ANOLEA: a www server to assess protein structures, Proc Int Conf Intell Syst Mol Biol. 5 (1997) 187–190.

[112] P. Benkert, M. Biasini, T. Schwede, Toward the estimation of the absolute quality of individual protein structure models, Bioinformatics. 27 (2011) 343–350. doi:10.1093/bioinformatics/btq662.

[113] C. Pan, J. Weng, W. Wang, Conformational Dynamics and Protein-Substrate Interaction of ABC Transporter BtuCD at the Occluded State Revealed by Molecular Dynamics Simulations, Biochemistry. 55 (2016) 6897–6907. doi:10.1021/acs.biochem.6b00386.

[114] N. Bhattacharjee, P. Biswas, Structure of hydration water in proteins: a comparison of molecular dynamics simulations and database analysis, Biophys. Chem. 158 (2011) 73–80. doi:10.1016/j.bpc.2011.05.009.

[115] H. Gohlke, G. Klebe, Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors, Angew. Chem. Int. Ed. Engl. 41

(2002) 2644–2676. doi:10.1002/1521-3773(20020802)41:15<2644::AID-ANIE2644>3.0.CO;2-O.

[116] W.L. Jorgensen, The Many Roles of Computation in Drug Discovery, Science. 303 (2004) 1813–1818. doi:10.1126/science.1096361.

[117] W.P. Walters, M.T. Stahl, M.A. Murcko, Virtual screening—an overview, Drug Discovery Today. 3 (1998) 160–178. doi:10.1016/S1359-6446(97)01163-X.

[118] T. Langer, R.D. Hoffmann, Virtual screening: an effective tool for lead structure discovery?, Curr. Pharm. Des. 7 (2001) 509–527. doi:10.2174/1381612013397861.

[119] J. Bajorath, Integration of virtual and high-throughput screening, Nat Rev Drug Discov. 1 (2002) 882–894. doi:10.1038/nrd941.

[120] D.B. Kitchen, H. Decornez, J.R. Furr, J. Bajorath, Docking and scoring in virtual screening for drug discovery: methods and applications, Nat Rev Drug Discov. 3 (2004) 935–949. doi:10.1038/nrd1549.

[121] I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin, A geometric approach to macromolecule-ligand interactions, Journal of Molecular Biology. 161 (1982) 269–288. doi:10.1016/0022-2836(82)90153-X.

[122] S.F. Sousa, P.A. Fernandes, M.J. Ramos, Protein-ligand docking: current status and future challenges, Proteins. 65 (2006) 15–26. doi:10.1002/prot.21082.

[123] M.D. Miller, S.K. Kearsley, D.J. Underwood, R.P. Sheridan, FLOG: a system to select “quasi-flexible” ligands complementary to a receptor of known three-dimensional structure, J. Comput. Aided Mol. Des. 8 (1994) 153–174. doi:10.1007/BF00119865.

[124] Ajay, M.A. Murcko, Computational methods to predict binding free energy in ligand-receptor complexes, J. Med. Chem. 38 (1995) 4953–4967. doi:10.1021/jm00026a001.

[125] F. Ballante, G.R. Marshall, An Automated Strategy for Binding-Pose Selection and Docking Assessment in Structure-Based Drug Design, J Chem Inf Model. 56 (2016) 54–72. doi:10.1021/acs.jcim.5b00603.

[126] F. Ballante, Protein-Ligand Docking in Drug Design: Performance Assessment and Binding-Pose Selection, in: Rational Drug Design, Humana Press, New York, NY, 2018: pp. 67–88. doi:10.1007/978-1-4939-8630-9_5.

[127] N. Huang, B.K. Shoichet, J.J. Irwin, Benchmarking sets for molecular docking, J. Med. Chem. 49 (2006) 6789–6801. doi:10.1021/jm0608356.

[128] M.J. Hartshorn, M.L. Verdonk, G. Chessari, S.C. Brewerton, W.T.M. Mooij, P.N. Mortenson, C.W. Murray, Diverse, high-quality test set for the validation of protein-ligand docking performance, J. Med. Chem. 50 (2007) 726–741. doi:10.1021/jm061277y.

# Chapitre III. Drug penetration in lipid bilayers

## III.1. Introduction

Xenobiotics (drugs, nutrients including antioxidants, endogenous compounds, toxics) cross several lipid bilayers to reach their biological targets, then to play their role in the organism. Some xenobiotics accumulate in membranes, where they can act (*e.g.*, as lipid peroxidation inhibitors for antioxidants) or they can be taken by membrane proteins through entries which are facing the surrounding lipid bilayers. There is a broad range of experimental methods able to tackle xenobiotics-membrane interactions (*e.g.,* advanced fluorescence spectroscopy, solid-state nuclear magnetic resonance, surface plasmon resonance, and neutron scattering). These methods are generally applied on liposomes of different sizes or self-supported mono/bi-layers. However, these experimental techniques can hardly be applied for screening on large drug banks or data bases, *e.g.*, of natural compounds. Moreover, these techniques provide only fragmented information and can hardly document on the exact position and orientation of the xenobiotics.

The complete rationalization of xenobiotic-membrane interactions can be evaluated *in silico*, by using molecular dynamics (MD). If well parameterized, MD simulations provide an accurate atomic description of these interactions, elucidating the driving forces responsible for xenobiotic insertion/positioning/orientation. The current MD methods also allow tackling binary or ternary lipid mixtures or intermolecular interactions between xenobiotics. They have shown success at reproducing experimental data when available. In this chapter, four studies about penetration, position and orientation of xenobiotics (curcumin; a novel flavonolignan derivative, beta-lapachone and the nitrone-trolox conjugate) will be described. These four studies correspond to four publications which have either been submitted or which have already been accepted. To facilitate reading, only the theoretical results or the experimental data required for understanding are presented here. The supplementary informations in this chapter and chapter IV are available at the following link: https://my.pcloud.com/publink/show?code=kZpyR97ZG1RgQjgSU5ze6zwPPccO8QwVTOX7
.

## III.2. Local low dose curcumin treatment improves functional recovery and remyelination in a rat model of sciatic nerve crush through inhibition of oxidative stress

Martial Caillaud[a], Benjamin Chantemargue[c,d], Laurence Richard[a,b], Laetitia Vignaud[a], Frédéric Favreau[a,f], Pierre-Antoine Faye[a,f], Philippe Vignoles[e], Franck Sturtz[a,f], Patrick Trouillas[c,d], Jean-Michel Vallat[b], Alexis Desmoulière[a], Fabrice Billet[a*].

*Published, DOI: 10.1016/j.neuropharm.2018.07.001*

### Foreword

*Antioxidants play a major role in the prevention of lipid peroxidation. Using MD simulations, we have predicted the position and orientation of curcumin in relation with their antioxidant properties. This supported the relationship, suggested experimentally, between neuroprotective properties of curcumin and its lipid peroxidation inhibition capacity.*

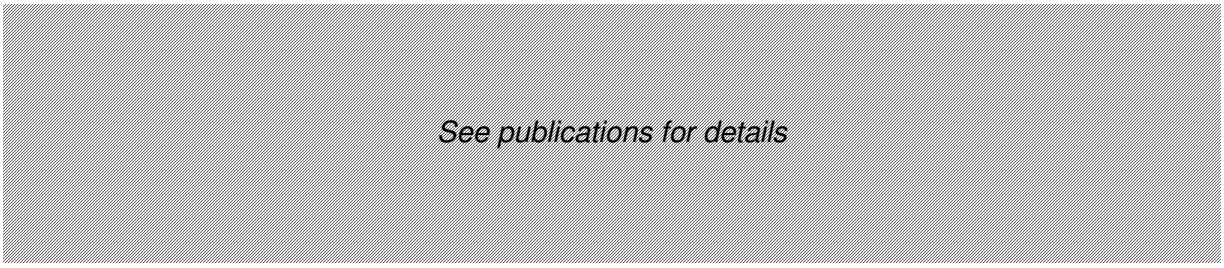
#### III.2.1. Abstract

Traumatic injuries to peripheral nerves are frequent, however, specific pharmacological treatments are currently lacking. Curcumin has antioxidant, anti-inflammatory and neuroprotective properties but high oral doses are required for therapeutic use, particularly due to its low bioavailability. The aim of the present study was to investigate the effects of local and continuous treatment using low curcumin doses on functional recovery and nerve regeneration after rat sciatic nerve crush (SNC). Curcumin was administered by osmotic pumps with a catheter delivering the drug at the injury site (0.2 mg/day for 4 weeks). Functionally, early improvements in mechanical sensitivity, finger spacing of the injured paw, skilful walking and grip strength were observed in curcumin-treated animals. The curcumin treatment increased expression of compact myelin proteins (MPZ and PMP22), myelin sheath thickness and, correspondingly, increased motor and sensitive nerve conduction velocity. Microscopic analysis of gastrocnemius muscle indicated a curcumin-induced decrease in neurogenic lesions. Curcumin treatment reduced the production of reactive oxygen species (ROS) (which were notably produced by macrophages), lipid peroxidation and increased expression of transcription factor Nrf2. *In silico* analyses indicated that curcumin combines all the characteristics required to be an efficient lipid peroxidation inhibitor at the heart of biological membranes, hence protecting their degradation due to ROS. This antioxidant capacity is likely to contribute to the beneficial effects of curcumin after SNC injury. These results demonstrate that, when administrated locally, low doses of curcumin represent a promising therapy for peripheral nerve regeneration.

### Keywords

Sciatic nerve crush; Curcumin; Oxidative stress; Myelin.

### III.2.2. *Material and methods*




*In silico methodology*

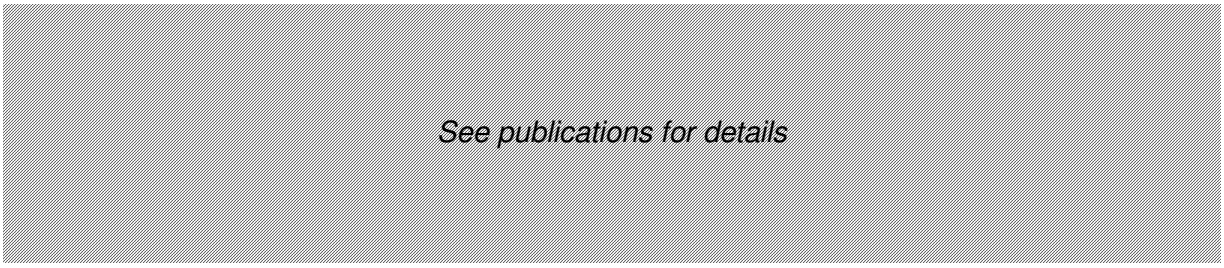
MD simulations were carried out using a lipid bilayer made of 128 lipids, either containing 100% POPC (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine) molecules or a 25% cholesterol and 75% POPC mixture (*i.e.*, 1:3 cholesterol:POPC). The surrounding hydration system was made using 42 water molecules per lipid, the water molecules being described by the explicit TIP3P model[1]. Physiological conditions were mimicked by 0.15 M NaCl distributed in the bulk water. Each membrane model was first equilibrated in the absence of curcumin. The topology of curcumin in its enol form was obtained from the PRODRG webserver[2]. Atomic charges were derived from the Restrained fit of Electrostatic Potential (RESP) approach as implemented in R.E.D. software[3], calculated at the DFT B3LYP/aug-cc-pVTZ level of theory with Gaussian09[4]. MD simulations were carried out using the GAFF[5] and Slipids[6] force fields for curcumin and lipid bilayers, respectively. Particular attention was paid to the atom types along the curcumin chain to correctly account for π-conjugation.

Two types of simulations were performed: one mimicking low concentrations, for which, only 1 curcumin was placed per leaflet, the 2 curcumins in the two leaflets having no interaction along the simulation; another simulation was carried out mimicking high local curcumin concentration, for which 6 curcumin molecules per leaflet were placed.

MD simulations were performed with a 2 fs time step. The pressure was maintained constant at 1atm by the Parinello-Rahman barostat[7] a semi-isotropic pressure coupling. The temperature was kept constant at 310 K using the Nosé-Hoover thermostat[8]. The electrostatic interactions were treated by the Particle Mesh Ewald scheme[9] with a 1.4 nm cutoff. The cutoff for the van der Waals interactions was set to 1.5 nm, with a switching function from 1.4 nm. The NPT (constant mole number, pressure and temperature) ensemble was used for all simulations. Periodic boundary conditions were used in all dimensions. All MD simulations were performed using the GROMACS package version 5.1.4[10]. Each MD simulation was 300 ns long, sufficient to reach a converged location.

For analyses, the z-axis was defined as the vector normal to the membrane surface. All analyses (position and orientation distributions as well as membrane properties) were assessed after MD simulations had reached convergence, *i.e.*, during the last 100 ns.

### III.2.3. Results




*Positioning and orientation of curcumin in lipid bilayers*

In MD simulations using one curcumin per leaflet (*i.e.*, low curcumin concentration), the insertion from bulk water to the lipid bilayer was rapid (*i.e.*, within the first 10 ns of the MD simulation when the molecule started from the water phase). This is in agreement with previous results obtained with various other polyphenols, showing rapid insertion into membranes[11]. During the rest of the MD simulation (300 ns), curcumin was anchored in between the lipid unsaturation and the polar head group region, curcumin's COM (center of mass) being at *ca*. 1.2 nm from the membrane centre (Fig. 8a). The OH groups at both edges of the chemical structure of curcumin drive anchoring to the polar head group by electrostatic and hydrogen bonding interactions, whereas the centre of the aromatic rings lie deeper in the membrane (Fig. 8b). Curcumin lies parallel to the membrane surface (see Fig. 8b), as confirmed by the angles between the z-axis (perpendicular to the membrane surface) and vectors 1–3 (defined in Supplementary data 1), which were *ca*. 90° (Supplementary data 1). Due to its extended $\pi$–conjugated system, the curcumin conformation remained planar throughout the simulation.

a

| Membrane | %[Cholesterol] | Number of curcumin molecule per leaflet | Group | Position (nm) |
|---|---|---|---|---|
| POPC | 0 | 1 | Curcumin COM | 1.2 ± 0.2 |
| | | | POPC C=C double bond | 0.8 ± 0.0 |
| | | | Phosphates | 1.9 ± 0.0 |
| POPC | 0 | 6 | Curcumin COM | 0.9 ± 0.6 |
| | | | POPC C=C double bond | 0.8 ± 0.1 |
| | | | Phosphates | 1.9 ± 0.2 |
| POPC | 25 | 1 | Curcumin COM | 1.4 ± 0.2 |
| | | | POPC C=C double bond | 0.9 ± 0.0 |
| | | | Phosphates | 2.1 ± 0.0 |
| | | | Cholesterol OH moiety | 1.4 ± 0.9 |

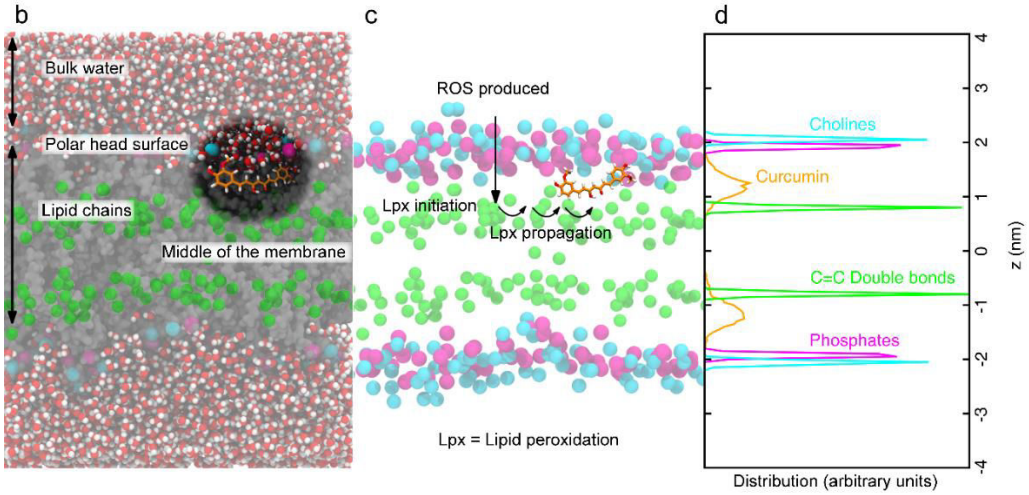



Fig. 8. Analysis of curcumin insertion in an *in silico* membrane model. (**a**) Averaged location of the center of mass (COM) of curcumin, lipid unsaturation (C=C double bond) and phosphorus atoms (picturing membrane surface) for systems with 1 curcumin per leaflet in POPC and 1:3 cholesterol:POPC bilayers, as well as systems with 6 curcumin molecules per leaflet in POPC bilayer; (**b**) Characteristic snapshot of curcumin in POPC bilayer lying parallel to the membrane surface. (**c**) Schematic description of lipid peroxidation. (**d**) Distribution of COM of curcumin, phosphates, cholines and C=C along the z-axis perpendicular to the membrane surface.

In Schwann cells (SCs), cholesterol concentration is high[12]. Therefore, to better model physiological conditions, MD simulations were also performed in a cholesterol enriched (25%) POPC bilayer. In this case, curcumin molecules insert deeper in the membrane. This can be attributed to a competition between cholesterol and curcumin, which are both prone to interact with the polar head group region of the membrane. Therefore, in the presence of cholesterol, curcumin is in closer contact with the unsaturated lipid region (Fig. 8a).

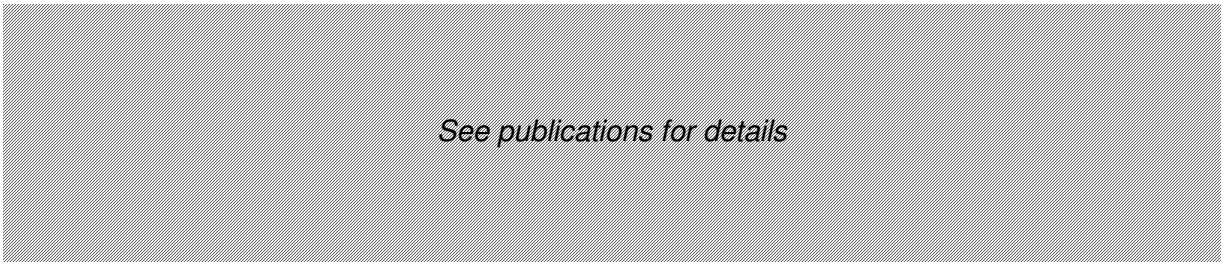
We also mimicked the situation corresponding to curcumin accumulation in the bilayer, which may induce a high local concentration of curcumin, by using 6 molecules per leaflet. In this case, the averaged location was slightly deeper in the bilayer (Fig. 8a), *i.e.*, again driving curcumin into closer contact with the unsaturated lipid region. It is worth mentioning that under this high-concentration condition, noncovalent self-associations between the different curcumin molecules were observed, which were driven by both $\pi$-$\pi$ stacking or by hydrogen bonding. Binary, tertiary and, to a lesser extent, higher-degree noncovalent complexes were

observed during the simulation (Supplementary data 2) in a dynamic way (namely, the different noncovalent assemblies stacked and unstacked several times during the simulation, see Supplementary data 2). Larger assemblies were also observed, in which molecules of curcumin from both leaflets were interacting (Supplementary data 2). Such transient noncovalent low-size molecular assemblies are likely to be formed, have previously been described between polyphenols[13] and have partially rationalized the possible accumulation of these compounds in tissues[14,15].

*Impact of curcumin on lipid bilayer structure*

The presence of curcumin in the lipid bilayer only slightly modified the organization of the bilayer, as seen from the slight increase of order parameters of the two lipid tails (Supplementary data 3). This, however, did not significantly modify bilayer thickness and APL (area per lipid), even at high curcumin concentration (Supplementary data 4). APL was not modified as curcumin partitioned sufficiently deep in the bilayer. As extensively described in the literature[16], the presence of cholesterol dramatically increases lipid bilayer organization as shown by an increase of *ca*. 30% of lipid order values (Supplementary data 3). This results in a decrease of APL and increase of bilayer thickness, respectively. In the presence of curcumin in a 1:3 cholesterol:POPC bilayer, this effect was slightly reduced, *i.e.*, slightly disorganizing bilayers (*i.e.*, decrease of lipid order parameter values, see Supplementary data 4). Again, this is attributed to the competing location close to the polar head group surface of both cholesterol and curcumin.

### III.2.4. Discussion




Free radical produced carbon-centred radicals on the lipid chains ($L^{\bullet}$), which after oxygen addition, form peroxy radicals ($LOO^{\bullet}$). These peroxy radicals then propagate from chain to chain until termination (*i.e.*, formation of oxidative products)[17]. The most efficient antioxidants inhibit the propagation stage; while inhibitors of the initiation stage may instead only control lipid peroxidation but cannot efficiently inhibit the process, as they only absorb the burst of free radicals entering membranes to a small extent. To represent a candidate propagation-stage inhibitor, the required properties are: i) efficient free radical scavenging capacity and ii) sufficiently deep insertion into the lipid bilayer to be in close contact with the unsaturated region where the propagation occurs. Concerning the former property, curcumin has repeatedly been shown to be an efficient free radical scavenger due to its efficient hydrogen atom donor capacity from the OH groups (at both edges of the molecule) or its efficient electron donor capacity from the extended $\pi$–conjugated system[18]. Concerning the latter property, here we show by MD simulations that curcumin is prone to quickly insert and partition into the lipid bilayers of Schwann Cells (SCs). Due to its amphiphilic character, curcumin is positioned in contact with both the polar head group region and unsaturated region of the polyunsatured fatty acids. This shows that curcumin is a relevant candidate to penetrate lipid bilayers of SCs, inhibiting lipid peroxidation and protecting the membrane from degradation due to ROS

produced by macrophages at 5 weeks after Sciatic Nerve Crush (SNC) and likely from other sources at the early post-injury stages. This mechanism of action is likely to be involved in the improvement of the remyelination process that is induced by curcumin.


Taken together, our results show that local infusion of low-dose curcumin improves functional recovery, nerve conduction velocity and remyelination after SNC in rats. Moreover, the present study indicates that curcumin combines all of the characteristics required to be an efficient antioxidant notably in SCs membranes, therefore protecting the myelin sheath. These results demonstrate that the local and continuous delivery of low curcumin dose represents a promising therapeutic modality in peripheral nerve repair. Thus, curcumin may be considered as a good candidate for therapeutic approaches based on controlled and continuous delivery system such as hydrogels, subcutaneous implants, functionalised tubing or nanoparticles.

## References


[1] D.J. Price, C.L. Brooks, A modified TIP3P water potential for simulation with Ewald summation, The Journal of Chemical Physics. 121 (2004) 10096–10103. doi:10.1063/1.1808117.

[2] A.W. Schüttelkopf, D.M.F. van Aalten, PRODRG: a tool for high-throughput crystallography of protein-ligand complexes, Acta Crystallogr. D Biol. Crystallogr. 60 (2004) 1355–1363. doi:10.1107/S0907444904011679.

[3] F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, P. Cieplak, The R.E.D. Tools: Advances in RESP and ESP charge derivation and force field library building, Phys Chem Chem Phys. 12 (2010) 7821–7839. doi:10.1039/c0cp00111b.

[4] M.J. Frish, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssel, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, Gaussian, Inc., Wallingford CT, 2016.

[5] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (2004) 1157–1174. doi:10.1002/jcc.20035.

[6] J.P.M. Jämbeck, A.P. Lyubartsev, An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes, J. Chem. Theory Comput. 8 (2012) 2938–2948. doi:10.1021/ct300342n.

[7] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, Journal of Applied Physics. 52 (1981) 7182–7190. doi:10.1063/1.328693.

[8] W.G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, Phys. Rev. A. 31 (1985) 1695–1697. doi:10.1103/PhysRevA.31.1695.

[9] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, The Journal of Chemical Physics. 98 (1993) 10089–10092. doi:10.1063/1.464397.

[10] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX. 1–2 (2015) 19–25. doi:10.1016/j.softx.2015.06.001.

[11] P. Košinová, K. Berka, M. Wykes, M. Otyepka, P. Trouillas, Positioning of Antioxidant Quercetin and Its Metabolites in Lipid Bilayer Membranes: Implication for Their Lipid-Peroxidation Inhibition, J. Phys. Chem. B. 116 (2012) 1309–1318. doi:10.1021/jp208731g.

[12] G. Saher, M. Simons, Cholesterol and Myelin Biogenesis, in: Cholesterol Binding and Cholesterol Transport Proteins:, Springer, Dordrecht, 2010: pp. 489–508. doi:10.1007/978-90-481-8622-8_18.

[13] P. Trouillas, J.C. Sancho-García, V. De Freitas, J. Gierschner, M. Otyepka, O. Dangles, Stabilizing and Modulating Color by Copigmentation: Insights from Theory and Experiment, Chem. Rev. 116 (2016) 4937–4982. doi:10.1021/acs.chemrev.5b00507.

[14] Y. Okamoto, D. Pehlivan, W. Wiszniewski, C.R. Beck, G.J. Snipes, J.R. Lupski, M. Khajavi, Curcumin facilitates a transitory cellular stress response in Trembler-J mice, Hum. Mol. Genet. 22 (2013) 4698–4705. doi:10.1093/hmg/ddt318.

[15] G. Fabre, I. Bayach, K. Berka, M. Paloncýová, M. Starok, C. Rossi, J.-L. Duroux, M. Otyepka, P. Trouillas, Synergism of antioxidant action of vitamins E, C and quercetin is related to formation of molecular associations in biomembranes, Chem. Commun. 51 (2015) 7713–7716. doi:10.1039/C5CC00636H.

[16] T. Róg, M. Pasenkiewicz-Gierula, I. Vattulainen, M. Karttunen, Ordering effects of cholesterol and its analogues, Biochimica et Biophysica Acta (BBA) - Biomembranes. 1788 (2009) 97–121. doi:10.1016/j.bbamem.2008.08.022.

[17] M. Valko, D. Leibfritz, J. Moncol, M.T.D. Cronin, M. Mazur, J. Telser, Free radicals and antioxidants in normal physiological functions and human disease, Int. J. Biochem. Cell Biol. 39 (2007) 44–84. doi:10.1016/j.biocel.2006.07.001.

[18] A. Barzegar, A.A. Moosavi-Movahedi, Intracellular ROS Protection Efficiency and Free Radical-Scavenging Activity of Curcumin, PLOS ONE. 6 (2011) e26012. doi:10.1371/journal.pone.0026012.

## III.3. Novel flavonolignan hybrid antioxidants: From enzymatic preparation to molecular rationalization

Eva Vavříková[a],Vladimír Křen[a], Lubica Jezova-Kalachova[a], Michal Biler[bc], Benjamin Chantemargue[bd], Michaela Pyszková[e], Sergio Riva[f], Marek Kuzma[a], Kateřina Valentová[a], Jitka Ulrichová[e], Jiří Vrba[e], PatrickTrouillas[bd], Jan Vacek[e]*

*Published, DOI: 10.1016/j.ejmech.2016.12.051*

### Foreword

*As for the previous study, the MD simulations have supported understanding of the antioxidant properties of novel hybrid flavonolignans, e.g., combination of silybin and ascorbic acid. This combination has enhanced lipid peroxidation inhibition capacity, which was mainly attributed to the positioning of the derivatives in membranes.*

#### III.3.1. Abstract

A series of antioxidants was designed and synthesized based on conjugation of the hepatoprotective flavonolignan silybin with l-ascorbic acid, trolox alcohol or tyrosol via a C12 aliphatic linker. These hybrid molecules were prepared from 12-vinyl dodecanedioate-23-O-silybin using the enzymatic regioselective acylation procedure with Novozym 435 (lipase B) or with lipase PS. Voltammetric analyses showed that the silybin-ascorbic acid conjugate exhibited excellent electron donating ability, in comparison to the other conjugates. Free radical scavenging, antioxidant activities and cytoprotective action were evaluated. The silybin-ascorbic acid hybrid exhibited the best activities (IC50 = 30.2 μM) in terms of lipid peroxidation inhibition. The promising protective action of the conjugate against lipid peroxidation can be attributed to modulated electron transfer abilities of both the silybin and ascorbate moieties, but also to the hydrophobic C12 linker facilitating membrane insertion. This was supported experimentally and theoretically by density functional theory (DFT) and molecular dynamics (MD) calculations. The results presented here can be used in the further development of novel multipotent antioxidants and cytoprotective agents, in particular for substances acting at an aqueous/lipid interface.

### Keywords

Sylibin; Vitamin C; Enzymatic coupling; Multipotent antioxidant; Lipid peroxidation; Cytoprotection.

### III.3.2. Material and methods

*See publications for details*

*Quantum calculations*

The geometries were optimized using DFT (density functional theory) formalism with the ωB97XD/6-31G(d) method. The ωB97XD functional is well-adapted to describe dispersion forces[53]. The frequency analysis was performed at 298 K and 1 atm, confirming the absence of any negative frequency (imaginary mode of vibration). The electronic energies were further computed with both ωB97XD/6-31 + G(d,p) and B3P86/6-31 + G(d,p) levels. The former method enables consistency to be maintained with the functional used for optimization, whereas the latter has demonstrated effectiveness at evaluating energetic values related to the redox properties of compounds investigated[12], [41]. Only the values obtained with the latter functional are reported here, noting that both functionals gave the same trends. For the different phenolic compounds (ArO•), the O-H bond-dissociation enthalpies (BDE), the ionization potential (IP) and the electron transfer energy (ETE) values were calculated as follows:

1) BDE = $H°_{(298K)}$ (ArO•) + $H°_{(298K)}$ (H•) – $H°_{(298K)}$ (ArOH), where ArOH is the neutral form of the investigated molecules (**1**, **2**, **3** and **6**) and ArO• the phenoxyl radical formed after hydrogen atom abstraction from ArOH.

2) IP = $H°_{(298K)}$ ($ArO^{•+}$) – $H°_{(298K)}$ (ArOH), where $ArO^{•+}$ is the radical cation obtained after electron transfer from the neutral form of the compounds.

3) ETE = $H°_{(298K)}$ (ArO•) – $H°_{(298K)}$ ($ArO^-$), where $ArO^-$ is the anionic form; ETE represents the energy required to transfer an electron from the deprotonated form of the compounds.

Solvent effects were taken into account implicitly using the polarizable continuum model (PCM). In PCM models, the substrate is embedded into a shape-adapted cavity surrounded by a dielectric continuum characterized by its dielectric constant. Calculations were performed in gas and water (ε = 78.35). All calculations were performed with Gaussian[54].

*Force field and membrane model*

All MD simulations were carried out using the package GROMACS version 5.0.4[55]. A 43A1-S3 force field was used[56]. The bilayer model consisted of 128 molecules of 1,2-dioleoyl-*sn*-glycero-3-phosphatidylcholine (DOPC)[57] surrounded by approximatively 4900 water molecules (SPC model)[58]. Phosphatidylcholines are the main type of phospholipids that constitute human membranes, and DOPC bilayers are a relevant model to get a global picture of drug membrane penetration. $Na^+$ and $Cl^-$ ions were added using the program Genion, according to their physiological concentrations (0.9% = 0.154 M).

*Molecule parameters*

For phenolic derivatives, the topologies were obtained from the PRODRG webserver[59]. The Restrained fit of Electrostatic Potential (RESP)[60,61] partial charges were alsoused. The ESP

charges were obtained from B3LYP/aug-cc-pVTZ[62] calculations on geometries optimized at the same level using the software Gaussian 09[54].

*Free MD simulations*

Several free MD simulations were carried out for each molecule in different protonated states, with different starting points (in water, inside bilayer, parallel or not to membrane surface). An energy minimization was done using the steepest-descent algorithm before the MD free simulations were performed over hundreds of ns. The Leap-frog Verlet integrator was used with a 2-fs time step. The cut-off for both electrostatic and vdW interactions was set to 14 Å. Long-range electrostatic interactions were calculated with the particle mesh Ewald (PME) algorithm[63]. Temperature and pressure were set to mimic physiological conditions, with 310 K maintained by a Nosé-Hoover thermostat[64,65] every 0.1 ps and the pressure was regulated anisotropically (each axis independently) at 1 atm by Parinello-Rahman barostat[66] with a time constant of 5 ps and a $4.5 \times 10^{-5}$ $bar^{-1}$ compressibility. All bonds were constrained by the LINCS algorithm (linear constraint solver)[67]. Periodic boundary conditions were used along the three axes. The simulations were carried out up to 2 µs, allowing averaging over more than 1 µs, which appears to be sufficiently long for all of the processes followed here. Diastereoisomer A of the silybin moiety was chosen in the initial setup for all calculations.

*Analysis of orientation*

Orientation was followed using 3 angles, named $\alpha_1$, $\alpha_2$, $\alpha_3$, describing the orientation of the ascorbic acid, silybin and linker (**9**) moieties, respectively. Specifically, the 3 angles were measured between the vectors aligned along the longest axis of these moieties and the *z*-axis perpendicular to the membrane surface (represented with phosphorus atoms), see Fig. 6.

### III.3.3. Results and discussion

*See publications for details*

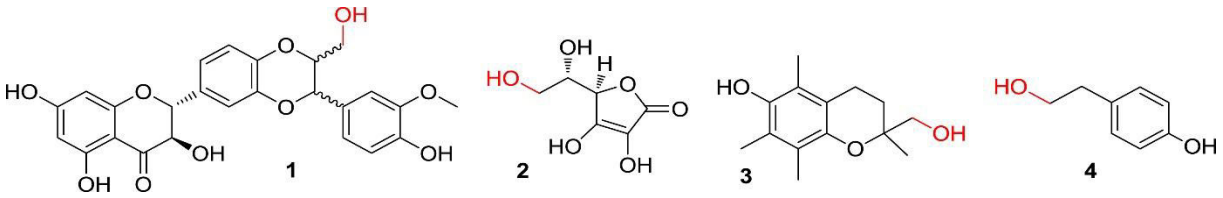
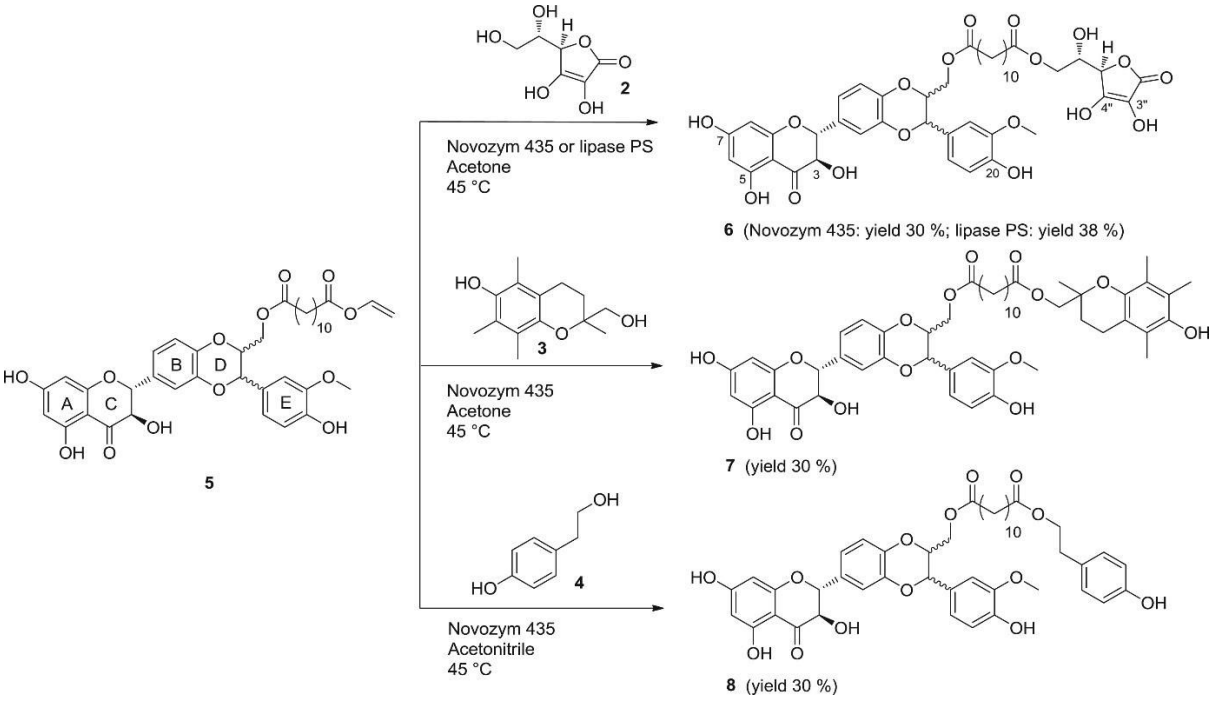



Fig. 1. Structures of silybin **(1)**, l-ascorbic acid **(2)**, trolox alcohol **(3)**, and tyrosol **(4)**. The hydroxyl groups involved in hybrid conjugate formation are marked in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)




Scheme 1. Enzymatic synthesis of compounds **6**, **7** and **8**



*Structure and folding of conjugate 6*

Prior to any redox property prediction, a careful conformational analysis was performed. Whereas **2** exhibits no specific conformational features and **1** was largely described in the literature[12,35,41], the conjugates, such as **6**, exhibited many more degrees of freedom. Due to the fact that **1** exists in the form of two diastereoisomers, diastereoisomer A was chosen in the initial setup for all calculations. With **1**, the type of diastereoisomers was shown to not significantly influence the overall redox properties or electron-donating capacity[12,35]. Sufficiently long molecular dynamics (MD) simulations were performed in water, which enabled the identification of several folded conformers (see Fig. S70 in Supporting Information). These different conformers were further optimized with a DFT-D (density functional theory including dispersion forces) method, which properly describes any existing dispersion interactions. The most stable conformer was folded, with a stabilizing energy of 78 kcal/mol compared to the linear form (Fig. 5). This conformer was mainly stabilized by π-π stacking interaction putting the B-ring of the silybin moiety and the moiety of ascorbic acid in close contact, but also by hydrogen bonding interactions (Fig. 5). Interestingly, no folding was observed in DMSO (even after 500-ns long MD simulations) and instead only linear conformers were obtained (Fig. S71 in Supporting Information).

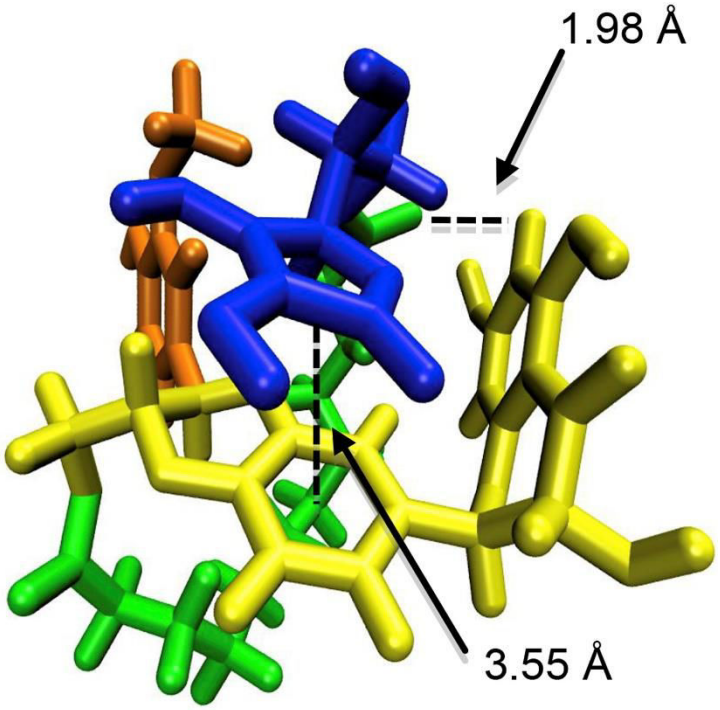



Fig. 5. The most stable conformer of compound **6**, highlighting $\pi$–$\pi$ stacking (between the ascorbic acid moiety and B-ring of the silybin moiety) and H-bonding (between C-7 OH and the keto group of the $C_{12}$ linker) interactions. The ascorbic acid (blue), $C_{12}$ linker (green), A-B-C-D rings (yellow) and the E-ring (orange) of silybin moieties are displayed in the respective colors. Diastereoisomer A of the silybin moiety was chosen in the initial setup for all calculations.

*Rationalization of compound 6 redox behavior*

To rationalize the redox properties of compound **6** with respect to its constitutive components **1** and **2**, hydrogen atom and electron transfer capacities were calculated. The O-H BDE (bond dissociation enthalpy) reflects the former transfer, whereas IP (ionization potential) and ETE (electron transfer energy) reflect the latter. The IP corresponds to an electron transfer from the protonated form, *i.e.* at pH = 5. The ETE reflects electron transfer from the deprotonated form, *i.e.* at physiological pH. No significant differences were observed in terms of BDE between **6** and both components, *i.e.* the lowest BDE value of **6** was obtained for C-3″ OH, which exhibited very similar BDE values in **2** of about 76 kcal/mol (Table 2). These calculations are in very good agreement with the fact that **2** is more easily oxidized than **1**. In addition to this, a thorough analysis of frontier molecular orbitals (MOs) was also performed. The only difference was that in **6**, the highest occupied molecular orbital (HOMO) and HOMO-1 appear to be almost degenerated, and their distribution was the combination of the two separated HOMO and HOMO-1 in **1** (Fig. S72 in Supporting Information).

Table 2. O-H BDE (bond dissociation enthalpies), IP (ionization potential) and ETE (electron transfer energy) as calculated in water (PCM solvent) with B3P86/6-31 + G(d,p)//ωB97XD/6-31G(d). All energetic values are given in kcal/mol. For atom numbering, see Scheme 1.

| | Compounds | **1** | **2** | **3** | **6 folded** | **6 linear** |
|---|---|---|---|---|---|---|
| IP | | 150.1 | 162.1 | 133.7 | 152.0 | 161.9 |
| | C-7 OH | 131.7 | | | 136.6 | 134.0 |
| | C-20 OH | 112.1 | | | 115.0 | 117.2 |
| ETE | C-3″ OH | | 124.5 | | 123.8 | 126.3 |
| | C-4″ OH | | 116.8 | | 116.7 | 116.3 |
| | OH | | | 98.4 | | |
| | C-7 OH | 91.4 | | | 92.9 | 93.3 |
| | C-20 OH | 82.5 | | | 83.5 | 84.7 |
| BDE | C-3″ OH | | 76.9 | | 76.5 | 77.8 |
| | C-4″ OH | | 79.4 | | 79.4 | 79.0 |
| | OH | | | 72.7 | | |

The absence of any enhancement of the redox properties of **6** with respect to its constitutive components is in good agreement with its DPPH scavenging capacity. In this case, the most active compound is **2**, due to its low BDE value, whereas **1** is much less active, which is attributed to a much higher BDE value (*i.e.* much lower capacity for hydrogen atom transfer), see Table 2. The combination of the two components in **6** decreases the DPPH scavenging of **2**, most likely due to hydrophobic effects attributed to the $C_{12}$ linker (Table 1), and conjugate **7** exhibits the same DPPH scavenging capacity as trolox (**3**, $IC_{50}$ 3.58 µM). This clearly shows a chemical difference between **6** and **7** in terms of their amphiphilic properties, which will be reported and discussed elsewhere.

In terms of ABTS scavenging, **1** is more active than **2** due to a lower ETE value (Table 2). This means that, as expected in this assay performed at physiological pH, the mechanism of action is electron transfer from the deprotonated forms to the free radical. When combining both moieties in **6**, their ABTS scavenging activities are enhanced in an additive way ($TEAC_{additive}$ = 0.57 + 0.78 = 1.35, very close to the TEAC of **6**, *i.e.*, 1.32; see Table 1). Therefore, no specific synergistic effects are observed, *e.g.,* no self-regeneration by electron transfer between the two moieties.

*Lpx inhibition and interactions of **6** with a model membrane*

This Lpx inhibition cannot be easily rationalized purely from the capacity to scavenge free radicals, which indeed exhibits the reverse trend, for instance when comparing **6** to **2**. To efficiently inhibit Lpx, not only sheer free radical scavenging capacity but also a specific inhibition propagation stage is required. This requires that the compound could be sufficiently embedded in the lipid bilayer membrane to be in contact with the lipid chains where oxidation occurs[43]. This is affected by the presence of hydrophobic moieties such as the $C_{12}$ linker (**9**). To resolve this issue we performed MD simulations, which have proved to be highly effective at providing accurate locations of bioactive compounds in lipid bilayers of a given composition, as well as accurate energetics of membrane penetration and crossing[42].

Compound **6** penetrated the 1,2-dioleoyl-*sn*-glycero-3-phosphatidylcholine (DOPC) bilayer and reached a converged location in a relatively short time, within ca. 400 ns. The center of mass (COM) of **6** was positioned 1.2 nm from the membrane center (Table 3 and Fig. 6). The

COM of the silybin moiety appeared to be deeper in the bilayer than the COM of the ascorbic acid moiety. Moreover, the silybin moiety appeared not to be perpendicular to the membrane surface (α2-angle of 93°, see Table S4 in Supporting Information). In contrast, the A-ring was oriented towards the polar head group region and the E-ring containing the most active free radical scavenger group of the silybin moiety (C-20 OH)[11,12] was deeper in the lipid bilayer (Table 3). This results in efficient inhibition of the propagation stage of Lpx. When starting the MD simulations, whether with the linear conformer or with the folded one, compound **6** quickly folded when approaching or inside the lipid bilayer. The only case in which **6** remained linear was with the transmembrane guess arrangement (Fig. 6); in this case, even after long-term MD simulations compound **6** remained linear and transmembrane. This compound is indeed an ideal transmembrane candidate; however, the limiting step to reaching the transmembrane arrangement is the flip-flop of one moiety, *e.g.* silybin. Indeed, compound **6** most likely inserts the membrane in its folded conformer, and then stabilizes below the polar head group region. From this region, the silybin moiety may cross the bilayer, thus leading to the transmembrane arrangement. However, evaluating the energy required for the flip-flop of the silybin moiety is a particularly complex issue with both experimental and theoretical techniques.

Table 3. Center of mass [nm] for protonated and deprotonated states of compounds **6** (folded), **6** (linear), **1** and **2**. COM values correspond to the center of mass for the entire molecule. (**2**) moiety, (**1**) moiety, A-ring of (**1**) moiety and E-ring of (**1**) moiety values correspond to the center of mass for the ascorbic acid moiety without the dihydroxyethyl chain, silybin moiety, A-ring of silybin moiety and E-ring of silybin moiety of the given compound, respectively.

| | COM | (**2**) moiety | (**1**) moiety | A-ring of (**1**) moiety | E-ring of (**1**) moiety |
|---|---|---|---|---|---|
| **Protonated state** | | | | | |
| **6** (folded) | 1.2 ± 0.1 | 1.6 ± 0.1 | 1.2 ± 0.2 | 1.4 ± 0.2 | 1.1 ± 0.2 |
| **6** (linear) | 1.1 ± 0.1 | 1.3 ± 0.2 | 1.0 ± 0.1 | 1.2 ± 0.2 | 0.4 ± 0.2 |
| **1** | 1.4 ± 0.2 | – | – | 1.4 ± 0.2 | 1.6 ± 0.2 |
| **2** | 1.6 ± 0.2 | 1.7 ± 0.2 | – | – | – |
| **Deprotonation at (2) moiety** | | | | | |
| **6** (folded) | 1.5 ± 0.2 | 2.0 ± 0.3 | 1.4 ± 0.3 | 1.6 ± 0.1 | 1.4 ± 0.2 |
| **6** (linear) | 1.6 ± 0.1 | 1.7 ± 0.2 | 0.6 ± 0.1 | 1.0 ± 0.1 | 1.5 ± 0.2 |
| **2** | 2.4 ± 0.2 | 2.4 ± 0.2 | – | – | – |
| **Deprotonation at (1) moiety** | | | | | |
| **6** (folded) | 2.0 ± 0.2 | 1.8 ± 0.3 | 2.2 ± 0.2 | 2.1 ± 0.2 | 2.2 ± 0.3 |
| **6** (linear) | 0.3 ± 0.1 | 1.4 ± 0.1 | 1.0 ± 0.1 | 1.4 ± 0.1 | 0.9 ± 0.2 |
| **Double deprotonation at (1) and (2) moieties** | | | | | |
| **6** (folded) | 2.5 ± 0.1 | 2.7 ± 0.2 | 2.8 ± 0.2 | 2.4 ± 0.2 | 2.1 ± 0.2 |
| **6** (linear) | 0.8 ± 0.1 | 1.7 ± 0.2 | 0.7 ± 0.1 | 1.4 ± 0.2 | 0.7 ± 0.2 |

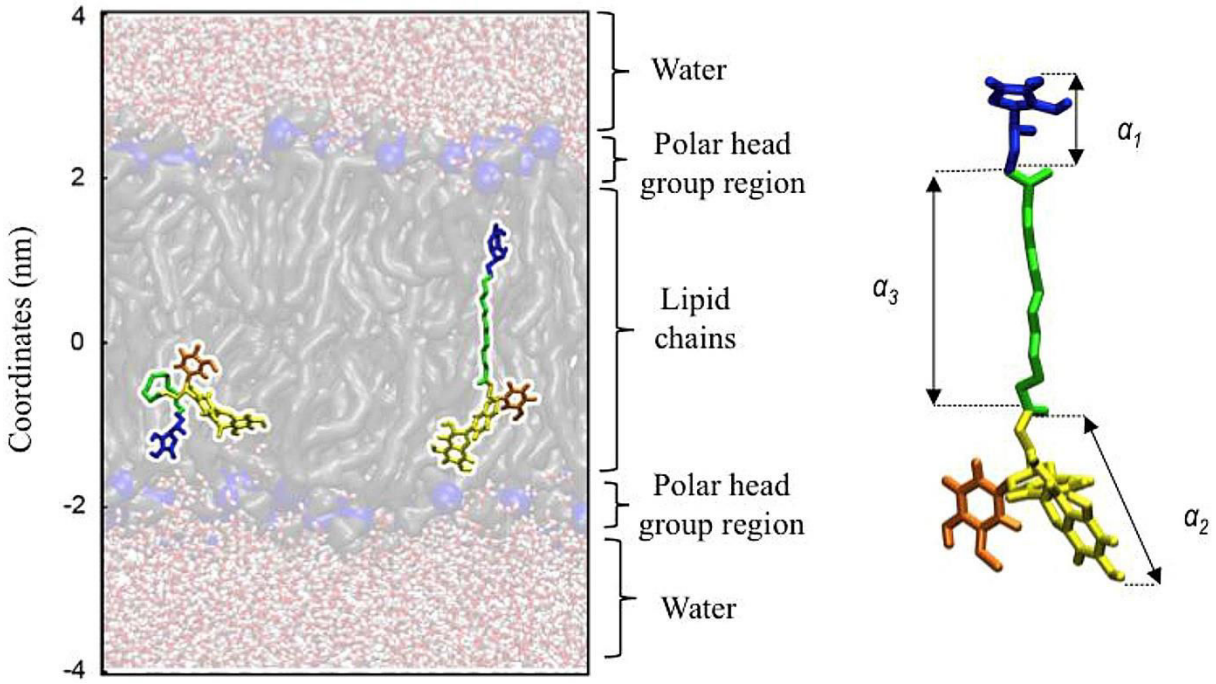



Fig. 6. Equilibrated location of compound **6** in DOPC lipid bilayers as averaged over the last 100 ns of 500-ns MD simulations. Two cases are shown, namely a folded and a linear conformer on the left and right hand side, respectively. The arrows show the vectors defining the $\alpha_1$, $\alpha_2$, $\alpha_3$ angles with respect to the z-axis of the membrane, which provide the orientation of the ascorbic acid, $C_{12}$ linker and silybin moieties, respectively. For color highlights of **6**, see Fig. 5.

In any event, the antioxidant active moiety (E-ring) of **6** (folded form) is pulled deeper into the bilayer than in **1**, namely the E-ring of **6** and **1** is at 1.1 nm and 1.6 nm, respectively (Table 3). This effect is less pronounced with the ascorbic acid (**2**) moiety, showing that this part of the molecule is still strongly anchored in the polar head group region. The global pulling of **6** deeper into lipid bilayers confirms its better capacity to inhibit the propagation stage of Lpx, compared to **1** and even more compared to **2**, which is not active. Moreover, in the folded conformer of **6**, both antioxidant moieties are in relatively close contact, which can favor the regeneration processes of, *e.g.* the polyphenol moiety by ascorbic acid[42]. We must note that even the linear transmembrane conformer pulls the antioxidant moieties deeper into the membrane and may partly rationalize the effective Lpx inhibition of **6**, which definitely spends relatively long periods deeply buried in the membrane.

As both **1** and **2** components of **6** have p$K_a$ values favoring the deprotonated states under physiological conditions, we also performed MD simulations with the deprotonated form at C-7 OH (on the silybin moiety), C-3″ OH (on the ascorbic acid moiety) and both. In both mono-deprotonated forms, the charge moiety pulls **6** in the polar head group region into close contact with water (COM at 2.0 nm and 1.5 nm, respectively as seen in Table 3). This is mainly attributed to strong electrostatic interactions with the choline groups of DOPC. The double deprotonated form of **6** did not penetrate the membrane at all, staying in close contact with both the choline moieties and water, within a µs time scale (COM at 2.5 nm, see Table 3). This clearly suggests that the proton transfer processes should occur at the membrane surface (most likely due to shifted acid-base balances) prior to penetration.

### III.3.4. Conclusion

The hepatoprotectant silybin **1** was enzymatically coupled to three potential antioxidants, namely ascorbate **2**, trolox alcohol **3** or tyrosol **4** *via* an aliphatic $C_{12}$ linker. The developed semi-synthetic protocol enables the preparation of pure conjugates in just two steps, avoiding cumbersome protective and deprotective reactions. The conjugation of **1** and **2** resulted in excellent electron-donor, antiradical, antioxidant as well as cytoprotective abilities of the conjugate **6**, 23-*O*-silybin-6-*O*′-[(+)-l-ascorbic acid] dodecanedioate. Compound **6** thus represents a novel type of hybrid antioxidant that is particularly effective at inhibiting lipid peroxidation, which was confirmed and rationalized by experimental as well as computational approaches.

## References


[12] P. Trouillas, P. Marsal, A. Svobodová, J. Vostálová, R. Gažák, J. Hrbáč, P. Sedmera, V. Křen, R. Lazzaroni, J.-L. Duroux, D. Walterová, Mechanism of the Antioxidant Action of Silybin and 2,3-Dehydrosilybin Flavonolignans: A Joint Experimental and Theoretical Study, J. Phys. Chem. A. 112 (2008) 1054–1063. doi:10.1021/jp075814h.

[35] M. Zatloukalová, V. Křen, R. Gažák, M. Kubala, P. Trouillas, J. Ulrichová, J. Vacek, Electrochemical investigation of flavonolignans and study of their interactions with DNA in the presence of Cu(II), Bioelectrochemistry. 82 (2011) 117–124. doi:10.1016/j.bioelechem.2011.06.005.

[41] M. Pyszková, M. Biler, D. Biedermann, K. Valentová, M. Kuzma, J. Vrba, J. Ulrichová, R. Sokolová, M. Mojović, A. Popović-Bijelić, M. Kubala, P. Trouillas, V. Křen, J. Vacek, Flavonolignan 2,3-dehydroderivatives: Preparation, antiradical and cytoprotective activity, Free Radic. Biol. Med. 90 (2016) 114–125. doi:10.1016/j.freeradbiomed.2015.11.014.

[42] G. Fabre, I. Bayach, K. Berka, M. Paloncýová, M. Starok, C. Rossi, J.-L. Duroux, M. Otyepka, P. Trouillas, Synergism of antioxidant action of vitamins E, C and quercetin is related to formation of molecular associations in biomembranes, Chem. Commun. 51 (2015) 7713–7716. doi:10.1039/C5CC00636H.

[53] J.-D. Chai, M. Head-Gordon, Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections, Physical Chemistry Chemical Physics. 10 (2008) 6615–6620. doi:10.1039/B810189B.

[54] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssel, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, Gaussian 09, Version A02, Gaussian, Inc., Wallingford CT, 2016.

[55] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX. 1–2 (2015) 19–25. doi:10.1016/j.softx.2015.06.001.

[56] S.-W. Chiu, S.A. Pandit, H.L. Scott, E. Jakobsson, An Improved United Atom Force Field for Simulation of Mixed Lipid Bilayers, J. Phys. Chem. B. 113 (2009) 2748–2763. doi:10.1021/jp807056c.

[57] D. Poger, A.E. Mark, On the Validation of Molecular Dynamics Simulations of Saturated and cis-Monounsaturated Phosphatidylcholine Lipid Bilayers: A Comparison with Experiment, J. Chem. Theory Comput. 6 (2010) 325–336. doi:10.1021/ct900487a.

[58] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, Interaction Models for Water in Relation to Protein Hydration, in: B. Pullman (Ed.), Intermolecular Forces, Springer Netherlands, 1981: pp. 331–342. doi:10.1007/978-94-015-7658-1_21.

[59] A.W. Schüttelkopf, D.M.F. van Aalten, PRODRG: a tool for high-throughput crystallography of protein-ligand complexes, Acta Crystallogr. D Biol. Crystallogr. 60 (2004) 1355–1363. doi:10.1107/S0907444904011679.

[60] C.I. Bayly, P. Cieplak, W. Cornell, P.A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, J. Phys. Chem. 97 (1993) 10269–10280. doi:10.1021/j100142a004.

[61] W.D. Cornell, P. Cieplak, C.I. Bayly, P.A. Kollmann, Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation, J. Am. Chem. Soc. 115 (1993) 9620–9631. doi:10.1021/ja00074a030.

[62] R.A. Kendall, T.H.D. Jr, R.J. Harrison, Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions, The Journal of Chemical Physics. 96 (1992) 6796–6806. doi:10.1063/1.462569.

[63] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, The Journal of Chemical Physics. 98 (1993) 10089–10092. doi:10.1063/1.464397.

[64] W.G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, Phys. Rev. A. 31 (1985) 1695–1697. doi:10.1103/PhysRevA.31.1695.

[65] S. Nosé, A molecular dynamics method for simulations in the canonical ensemble, Molecular Physics. 52 (1984) 255–268. doi:10.1080/00268978400101201.

[66] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, Journal of Applied Physics. 52 (1981) 7182–7190. doi:10.1063/1.328693.

[67] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: A linear constraint solver for molecular simulations, J. Comput. Chem. 18 (1997) 1463–1472. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

## III.4. Experimental and theoretical approaches for deciphering the peculiar behavior of beta-lapachone in lipid monolayers and bilayers

Xiao Wu[1], Benjamin Chantemargue[2,3], Florent Di Meo[2], Claudie Bourgaux[1], Christopher Thorne[1], W. Chollet[1], David Chapron[1], Patrick Trouillas[2,3], Veronique Rosilio[1*]

*Submitted*

### Foreword

*This study deals with the impact of beta-lapachone (b-lap) on the structural properties of membranes. MD simulations allow supporting experimental measurements to provide the position and orientation of b-lap in lipid bilayers with different lipid compositions, with or without 30% of cholesterol.*

#### III.4.1. Abstract

Beta-lapachone (b-lap) is a promising anticancer drug whose applications are limited by its poor solubility and stability. Its encapsulation in liposomes has been proposed to overcome these issues. However, in this study we show that b-lap exhibits an atypical interfacial behavior when mixed with lipids. Although the drug does not seem retained at the interface as deduced from the unchanged characteristics of lipid $\pi-$A isotherms, b-lap modifies the surface potential, compressibility moduli and compression cycles of lipid monolayers. At low concentration, the drug effect depends on the presence of double bonds in phospholipid chains. At high concentration, b-lap rigidifies the mixed monolayers, especially those made of saturated phospholipids. At low and high concentration, it disorganizes all phospholipid-cholesterol monolayers. The effect of b-lap on the thermal behavior and structure of DPPC and POPC 2 bilayers confirms its location in the upper part of the bilayers, close to the polar head groups, however with relatively high diffusibility. Molecular dynamic simulations allow a comprehensive description of its position and orientation in DOPC and POPC bilayers, in the presence or absence of cholesterol.

### Keywords

Surface pressure; differential scanning calorimetry; X-ray scattering (SAXS); molecular dynamics; phospholipid; cholesterol; stearylamine.

### III.4.2. Material and methods

*See publications for details*

*Molecular dynamic (MD) simulation*

MD simulations were carried out using four lipid bilayer compositions, namely POPC, DOPC, CHOL:POPC (1:3) and CHOL:DOPC (1:3). Each membrane contained 128 molecules (*i.e.*, 38 molecules of CHOL in the mixtures). An 8x8x8 nm box was created which included the bilayer, explicit water molecules and ions ($Na^+$ and $Cl^-$) at physiological concentration (0.9%). The periodic boundary conditions were applied in all directions. The z-axis was defined by the normal vector to membrane surface.

The Gromos43A1-s3 force field was used for the lipids and b-lap[1]. Atomic charges of b-lap were derived from the Restrained fit of Electrostatic Potential (RESP) calculated at the DFT B3LYP/aug-cc-pVTZ level of theory[2]. The explicit SPC (single point charge) water model was used for water molecules surrounding the membrane[3]. All MD simulations were performed using the GROMACS package, version 5.0.4[4]. The DFT calculations were performed by the Gaussian09 software[5]. Each membrane model was prior equilibrated without b-lap. B-lap was then positioned inside or outside the lipid bilayers. For each membrane, 1, 2, 3 and 4 molecules of b-lap per leaflet were successively included in the membrane models, corresponding to 1.5, 3, 4.7, 6.3 mol%, respectively. The MD simulations were 600 ns long each. They were carried out in the NPT ensemble. The temperature was regulated by the Nose-Hoover thermostat at 293 K (0.5 ps coupling constant)[6,7], the pressure was kept constant semi-isotropically at 1 atm by the Parinello-Rahman barostat (5.0 ps coupling constant and 4.5 x $10^{-5}$ compressibility)[8]. The cutoff for short-range van der Waals and Coulomb interactions was set to 1.0 nm. Long range electrostatic interactions were calculated with the particle mesh Ewald (PME) algorithm[9]. All analyses were performed over the last 400 ns of each simulation to ensure sampling over the well-equilibrated system.

### III.4.3. Results

*See publications for details*

*In silico evaluation of b-lap – membrane interaction*

*Positioning and orientation of b-lap in lipid bilayers*

The position of b-lap in the different lipid bilayers was given in terms of z-value with respect to the center of the lipid bilayer (Table 3) and of distributions along the z-axis (Figure 5). B-lap inserts into the lipid bilayers and it preferentially lies in between the lipid unsaturation and the

polar head group, with relatively high diffusibility. To exemplify this diffusibility, relatively high standard deviation was noticed, and it is worth noting that two b-lap flip-flop events occurred within the 600 ns of the MD simulations. Although not tightly anchored in a preferred position, keto groups are, in average, oriented towards membrane surface to maximize H16 bonding with the polar head groups. The analysis of the molecule orientation showed that the longest axis of b-lap was perpendicular to the z-axis, with greater fluctuations in POPC than in DOPC membrane (Figure 6). The depth of penetration was almost not affected by the number of b-lap molecules; however, the orientation was more constrained (less fluctuations) with 4 molecules per leaflet instead of 1. No stacking, e.g. as driven by $\pi-\pi$ stacking interactions, was observed during the simulation time. The results also show that the presence of CHOL pulls b-lap towards the polar head groups (Table 3).

Table 3: Distance $z$ (center of mass of b-lap), $z_{C=O}$ (carbonyl group of b-lap), $z_{Phosphate}$ (phosphate groups of DOPC or POPC), $z_{C=C}$ (unsaturation of DOPC or POPC) and $z_{OH\text{-}CHOL}$ (OH group of CHOL) from membrane center, in DOPC, POPC, CHOL:POPC (1:3) and CHOL:DOPC (1:3).

| Lipid composition | Number of b-lap molecules per leaflet | $z$ (nm) | $z_{C=O}$ (nm) | $z_{Phosphate}$ (nm) | $z_{C=C}$ (nm) | $z_{OH\text{-}CHOL}$ (nm) |
|---|---|---|---|---|---|---|
| DOPC | 1 | 1.5 ± 0.2 | 1.6 ± 0.2 | 2.0 ± 0.0 | 0.8 ± 0.0 | - |
| | 2 | 1.4 ± 0.2 | 1.6 ± 0.2 | 2.0 ± 0.0 | 0.8 ± 0.0 | - |
| | 3 | 1.3 ± 0.2 | 1.5 ± 0.2 | 2.0 ± 0.0 | 0.8 ± 0.0 | - |
| | 4 | 1.4 ± 0.2 | 1.6 ± 0.2 | 2.0 ± 0.0 | 0.8 ± 0.0 | - |
| CHO:DOPC 1:3 | 1 | 1.6 ± 0.2 | 1.8 ± 0.2 | 2.2 ± 0.0 | 0.9 ± 0.0 | 1.7 ± 0.0 |
| | 2 | 1.6 ± 0.2 | 1.8 ± 0.2 | 2.2 ± 0.0 | 0.9 ± 0.0 | 1.6 ± 0.0 |
| | 3 | 1.5 ± 0.2 | 1.7 ± 0.3 | 2.2 ± 0.0 | 0.9 ± 0.0 | 1.6 ± 0.0 |
| | 4 | 1.6 ± 0.2 | 1.8 ± 0.2 | 2.2 ± 0.0 | 0.9 ± 0.0 | 1.6 ± 0.0 |
| POPC | 1 | 1.3 ± 0.2 | 1.4 ± 0.2 | 1.9 ± 0.0 | 0.8 ± 0.0 | - |
| | 2 | 1.2 ± 0.2 | 1.5 ± 0.2 | 1.9 ± 0.0 | 0.8 ± 0.0 | - |
| | 3 | 1.2 ± 0.2 | 1.5 ± 0.2 | 1.9 ± 0.0 | 0.8 ± 0.0 | - |
| | 4 | 1.2 ± 0.2 | 1.6 ± 0.2 | 1.9 ± 0.0 | 0.8 ± 0.0 | - |
| CHOL:POPC 1:3 | 1 | 1.4 ± 0.2 | 1.7 ± 0.1 | 2.1 ± 0.0 | 0.9 ± 0.0 | 1.5 ± 0.1 |
| | 2 | 1.5 ± 0.2 | 1.7 ± 0.1 | 2.1 ± 0.0 | 0.9 ± 0.0 | 1.5 ± 0.1 |
| | 3 | 1.6 ± 0.2 | 1.8 ± 0.2 | 2.1 ± 0.0 | 0.9 ± 0.0 | 1.5 ± 0.1 |
| | 4 | 1.4 ± 0.2 | 1.6 ± 0.2 | 2.1 ± 0.0 | 0.9 ± 0.0 | 1.4 ± 0.1 |

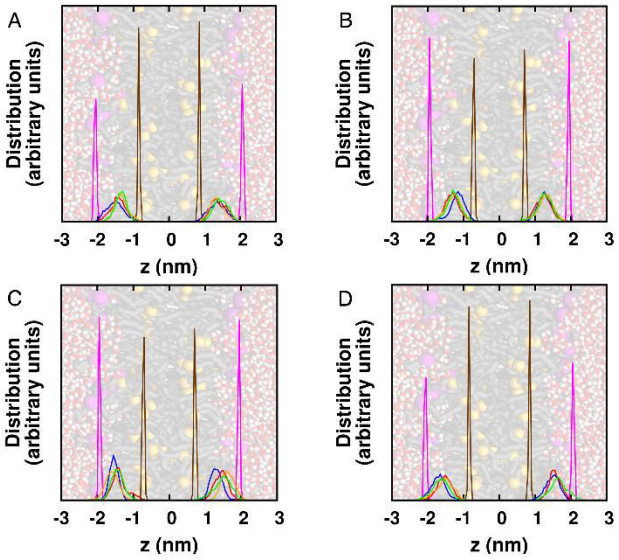



Figure 5: Distribution of b-lap along the z-axis in A) DOPC, B) POPC, C) 1:3 CHOL:POPC (1:3) and D) CHOL:DOPC (1:3). The different b-lap profiles (blue, red, orange and green) correspond to the situations of 1, 2, 3 and 4 molecules per leaflet, respectively. Double-bonds and phosphates are colored in dark-yellow and magenta, respectively.

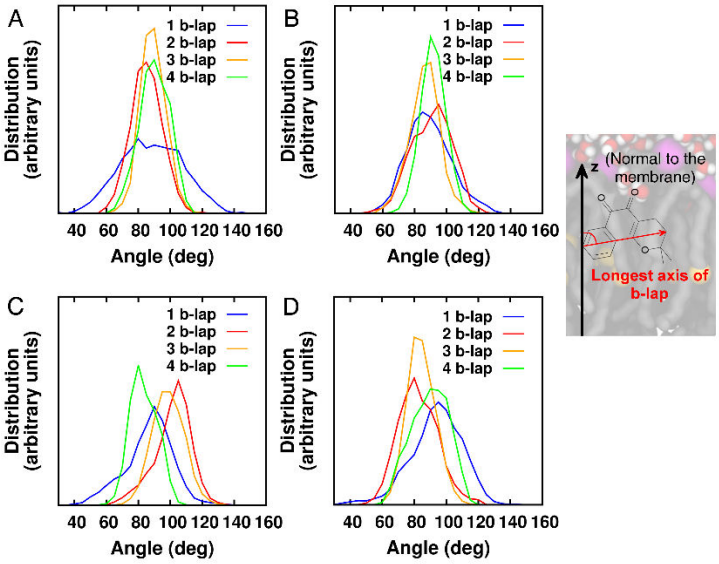



Figure 6: Orientation of b-lap in A) DOPC, B) POPC, C) CHOL:POPC (1:3) and D) CHOL:DOPC (1:3), as seen by the distribution of the angle between the vector colinear with the longest axis of b-lap and the normal vector to membrane surface colinear with the z-axis (see definition in the box).

*Structural impact of b-lap on lipid bilayers*

As expected, CHOL decreases area per lipid and conversely increases thickness of the different membranes (Table 4). B-lap (using up to four molecules per leaflet) generally tends to exhibit either the reverse effect - namely slight increase of area per lipid and decrease of thickness – or has no effect (Table 4). If any, the effect remains very weak due to the small molar fraction of b-lap. Using higher molar fraction could have induced greater effects nonetheless it would have not mimicked the experimental conditions. CHOL increases lipid

order parameters for the two acyl chains of both POPC and DOPC-based membranes (Figure 7), which is consistent with its structuring effect. In the absence of CHOL, no significant effects attributed to b-lap are observed on the order parameter. Conversely, in CHOL-containing membranes, b-lap decreases the order parameter values. In DOPC, its effect is greater at the second half of the lipid chain (*i.e.*, below unsaturation, see Figure 7A). In POPC, its effect is clearer on the sn1 chain, whereas almost no effect is observed on sn2 (Figure 7B).

Table 4: Thickness and area per lipid in DOPC, POPC, CHOL:POPC (1:3) and CHOL:DOPC (1:3).

| Lipid composition | Number of b-lap molecules per leaflet | Thickness (nm) | Area per lipid (Å$^2$) |
|---|---|---|---|
| DOPC | 0 | 4.14 ± 0.10 | 62.3 ± 1.0 |
| | 1 | 4.13 ± 0.10 | 61.8 ± 1.0 |
| | 2 | 4.06 ± 0.10 | 62.7 ± 1.0 |
| | 3 | 4.09 ± 0.10 | 62.8 ± 1.0 |
| | 4 | 4.07 ± 0.10 | 63.3 ± 1.0 |
| CHO:DOPC 1:3 | 0 | 4.36 ± 0.10 | 50.2 ± 0.5 |
| | 1 | 4.32 ± 0.10 | 51.0 ± 1.0 |
| | 2 | 4.31 ± 0.10 | 51.3 ± 1.0 |
| | 3 | 4.28 ± 0.10 | 52.0 ± 1.0 |
| | 4 | 4.30 ± 0.10 | 52.0 ± 1.0 |
| POPC | 0 | 3.92 ± 0.10 | 62.8 ± 1.0 |
| | 1 | 3.92 ± 0.10 | 63.4 ± 1.0 |
| | 2 | 3.91 ± 0.10 | 63.9 ± 1.0 |
| | 3 | 3.95 ± 0.10 | 63.2 ± 1.0 |
| | 4 | 3.91 ± 0.10 | 64.4 ± 1.0 |
| CHOL:POPC 1:3 | 0 | 4.27 ± 0.10 | 50.8 ± 1.0 |
| | 1 | 4.26 ± 0.10 | 50.8 ± 1.0 |
| | 2 | 4.22 ± 0.10 | 51.5 ± 1.0 |
| | 3 | 4.18 ± 0.10 | 51.8 ± 1.0 |
| | 4 | 4.18 ± 0.10 | 52.5 ± 1.0 |

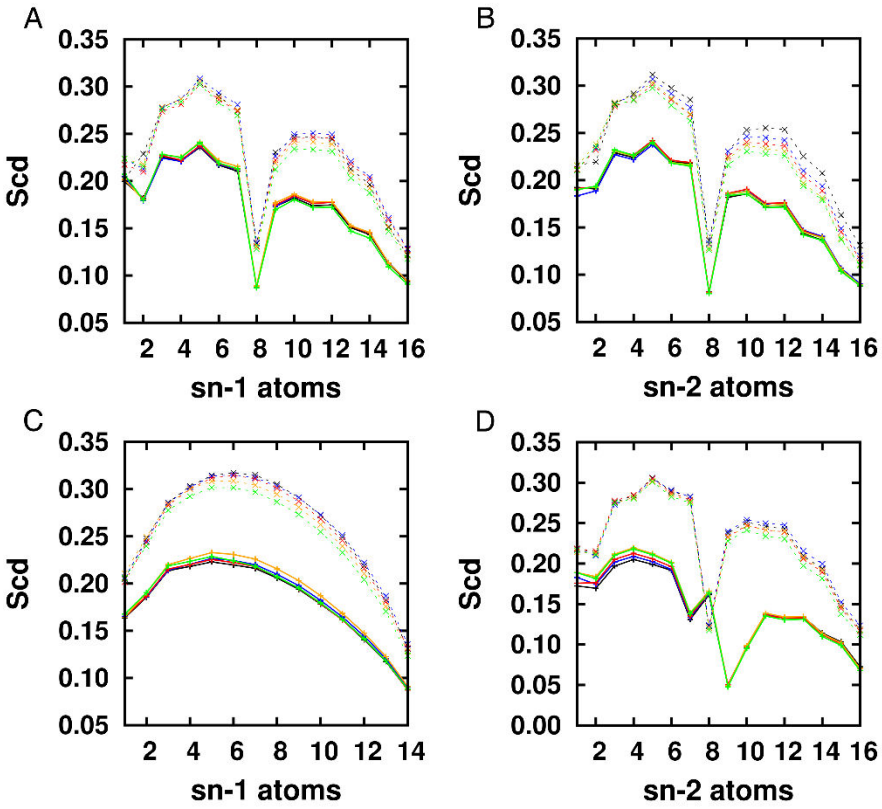



Figure 7: Lipid order parameters of sn1 and sn2 lipid chains in A) DOPC and B) POPC-based membranes. The profiles along the atoms of both chains are in black, blue, red, orange and green according to the number of b-lap per leaflet (0, 1, 2, 3 and 4, respectively). The dotted lines represent the membranes with CHOL.

### III.4.4. Discussion

*See publications for details*

The X-ray patterns of the two phospholipid bilayers were unchanged in the presence of 3.5 mol% of b-lap. The scattering profiles were characteristic of the Lβ' gel phase of DPPC and the Lα fluid phase of POPC, respectively. The d-spacing of the bilayers were 6.34 and 6.45 nm, respectively, in agreement with values reported in the literature for the pure phospholipids[10-13]. The d-spacing corresponds to the sum of the bilayer and the water thicknesses and it is therefore very sensitive to any change in membrane structure and interactions between bilayers. The constant d-spacing means that the drug was able to interact with the phospholipids in a gel phase (DPPC) as well as in a fluid lipid phase (POPC), without perturbing their lamellar organization. These results perfectly agree with MD simulations showing that the thickness of DOPC and POPC bilayers remained constant upon addition of b-lap.

The exact location and orientation of b-lap in the lipid bilayers was further documented by MD simulations. MD analyses showed that b-lap locates slightly deeper in DOPC than in POPC membrane, which agrees with the suggested key role of lipid unsaturation, POPC lipids exhibiting a chain without unsaturation. This key role is supported by the observation, made at low b-lap molar fraction, of its lesser impact on the compressibility modulus of POPC compared to DOPC and DLOPC monolayers.

Huynh *et al.* [14,15] showed evidence of dynamic clusters of acyl chains in POPC monolayers. The existence of these nano-clusters of palmitoyl (sn-1) or oleoyl (sn-2) chains was confirmed in the present simulation. Interestingly, the radial distribution function (RDF), also denoted g(r), clearly showed that b-lap preferentially partitioned in the sn-1 chain clusters in POPC bilayers and to a less but still significant extent in POPC-CHOL bilayers (Figure 8A). This can also be seen on one snapshot depicting the density mapping of these nano-clusters and the location of b-lap (Figure 8B-E). In other words, in bilayers made of a hybrid phospholipid, b-lap appeared prone to interact more with the saturated than with the unsaturated chains, inducing a rigidifying effect on partially or fully saturated phospholipid monolayers.

The transfer of b-lap towards the polar head group region in the presence of CHOL is consistent with the structuring effect of CHOL in membranes. Namely, the better packing of CHOL with lipid chains reduces free volume, which drives b-lap in closer contact to the polar head groups with an increase of H-bonding interactions with the keto groups of b-lap. In CHOL-containing membranes, b-lap decreases the order parameter values. In POPC bilayers, its effect was clearer on the sn-1 chain, whereas almost no effect was observed on sn-2. This noticeable observation reflects a destructuring effect assigned to b-lap, which counterbalances the structuring effect of CHOL, as seen in surface pressure measurements. This agrees with the presence of b-lap at the interface in contact with the polar head group region however with the capacity to diffuse towards the middle of the lipid chains. This dynamical presence of b-lap between the polar head groups and the lipid tails counteracts CHOL-lipid tail ordering and rationalize the biophysical observations.

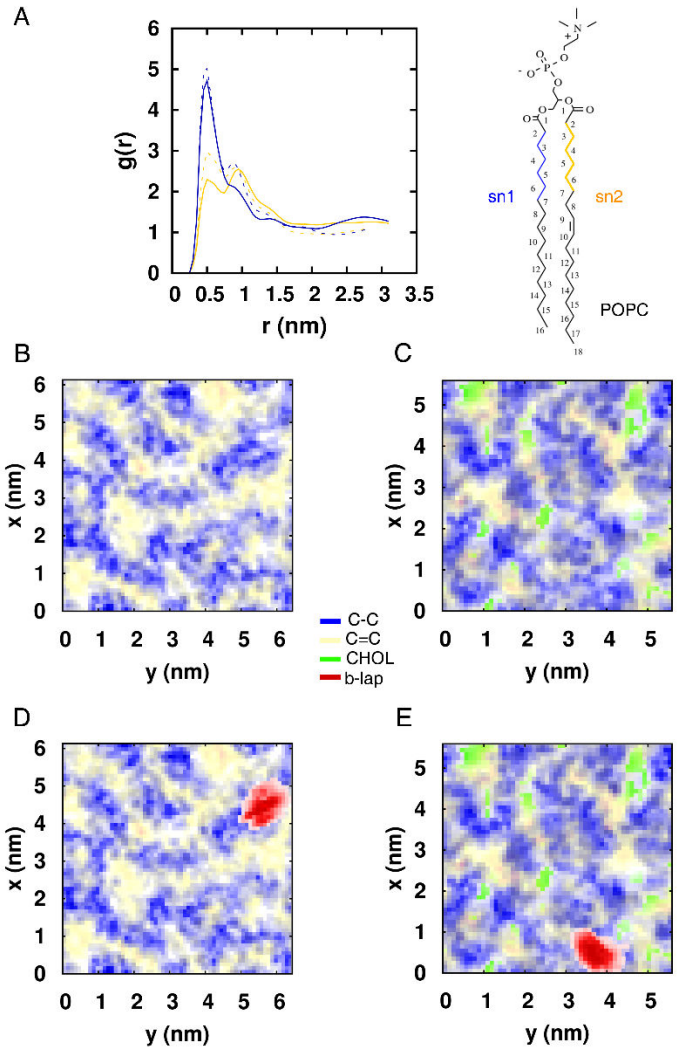



Figure 8: (A) Radial distribution function (RDF), also noted g(r), of atoms C2-C7 from sn-1 chain (blue) and sn-2 chain (yellow) with respect to the center of mass of b-lap in POPC in the presence (full line) or absence (dashed line) of 25 % CHOL. Snapshots of densities of C2-C7 from sn-1 (blue) and sn-2 chain (yellow) in the absence (B) or presence (C) of CHOL, obtained by projection onto the x,y plan, perpendicular to the z-axis; CHOL is depicted in green. The densities show clustering of sn-1 chains (blue) and sn-2 chains (yellow). (D) and (E) represent the same snapshots on which the density of b-lap (red) is additionally depicted, localized in close contact with the sn-1 chain.

### III.4.5. Conclusion

*See publications for details*

The position of b-lap within bilayers is similar to that of CHOL, which may lead to competition between the two molecules. This is supported by MD simulations revealing that the drug, lying between the lipid chain unsaturations and the phosphate groups, is driven towards the polar

head groups in the presence of CHOL. In hybrid phospholipid (POPC) monolayers, it preferentially partitions in nano-clusters made of the saturated chains.


Beside the information gained about the behavior of b-lap in phospholipid mono- and bilayers, this study highlights the complementarity of physical-chemical techniques and MD simulations for the better understanding of drug-membrane interactions.

## References


[1] S.-W. Chiu, S.A. Pandit, H.L. Scott, E. Jakobsson, An Improved United Atom Force Field for Simulation of Mixed Lipid Bilayers, J. Phys. Chem. B. 113 (2009) 2748–2763. doi:10.1021/jp807056c.

[2] F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, P. Cieplak, The R.E.D. Tools: Advances in RESP and ESP charge derivation and force field library building, Phys Chem Chem Phys. 12 (2010) 7821–7839. doi:10.1039/c0cp00111b.

[3] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, Interaction Models for Water in Relation to Protein Hydration, in: Intermolecular Forces, Springer, Dordrecht, 1981: pp. 331–342. doi:10.1007/978-94-015-7658-1_21.

[4] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX. 1–2 (2015) 19–25. doi:10.1016/j.softx.2015.06.001.

[5] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssel, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, Gaussian 09, Version A02, Gaussian, Inc., Wallingford CT, 2016.

[6] S. Nosé, A molecular dynamics method for simulations in the canonical ensemble, Molecular Physics. 52 (1984) 255–268. doi:10.1080/00268978400101201.

[7] W.G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, Phys. Rev. A. 31 (1985) 1695–1697. doi:10.1103/PhysRevA.31.1695.

[8] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, Journal of Applied Physics. 52 (1981) 7182–7190. doi:10.1063/1.328693.

[9] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, The Journal of Chemical Physics. 98 (1993) 10089–10092. doi:10.1063/1.464397.

[10] K. Lohner, A. Schuster, G. Degovics, K. Müller, P. Laggner, Thermal phase behaviour and structure of hydrated mixtures between dipalmitoyl- and dihexadecylphosphatidylcholine, Chemistry and Physics of Lipids. 44 (1987) 61–70. doi:10.1016/0009-3084(87)90005-3.

[11] B. Pili, C. Bourgaux, H. Amenitsch, G. Keller, S. Lepêtre-Mouelhi, D. Desmaële, P. Couvreur, M. Ollivon, Interaction of a new anticancer prodrug, gemcitabine–squalene, with

a model membrane: Coupled DSC and XRD study, Biochimica et Biophysica Acta (BBA) - Biomembranes. 1798 (2010) 1522–1532. doi:10.1016/j.bbamem.2010.04.011.

[12] C. Pereira-Leite, C. Carneiro, J.X. Soares, C. Afonso, C. Nunes, M. Lúcio, S. Reis, Biophysical characterization of the drug-membrane interactions: the case of propranolol and acebutolol, Eur J Pharm Biopharm. 84 (2013) 183–191. doi:10.1016/j.ejpb.2012.12.005.

[13] B. Drasler, D. Drobne, A. Sadeghpour, M. Rappolt, Fullerene up-take alters bilayer structure and elasticity: A small angle X-ray study, Chem. Phys. Lipids. 188 (2015) 46–53. doi:10.1016/j.chemphyslip.2015.04.001.

[14] L. Huynh, N. Perrot, V. Beswick, V. Rosilio, P.A. Curmi, A. Sanson, N. Jamin, Structural properties of POPC monolayers under lateral compression: computer simulations analysis, Langmuir. 30 (2014) 564–573. doi:10.1021/la4043809.

[15] L. Huynh, N. Perrot, V. Beswick, V. Rosilio, P.A. Curmi, A. Sanson, N. Jamin, Reply to “Comment on ‘Structural properties of POPC monolayers under lateral compression: computer simulations analysis,’” Langmuir. 31 (2015) 888–889. doi:10.1021/la504104e.

## III.5. Nitrone-Trolox as an inhibitor of lipid oxidation: Towards synergetic antioxidant effects

Larissa Socrier[a], Marie Rosselin[b], Ana Milena Gomez Giraldo[a], Benjamin Chantemargue[c, d], Florent Di Meo[c], Patrick Trouillas[c, d], Grégory Durand[b] and Sandrine Morandat[a]

*Submitted*

### Foreword

*This study is in the line with the second publication about bifunctional antioxidants. MD simulations allow rationalizing the enhanced lipid peroxidation inhibition when combining a nitrone, namely α-phenyl-N-tert-butylnitrone (PBN), and Trolox, the hydro-soluble vitamin E.*

### III.5.1. Abstract

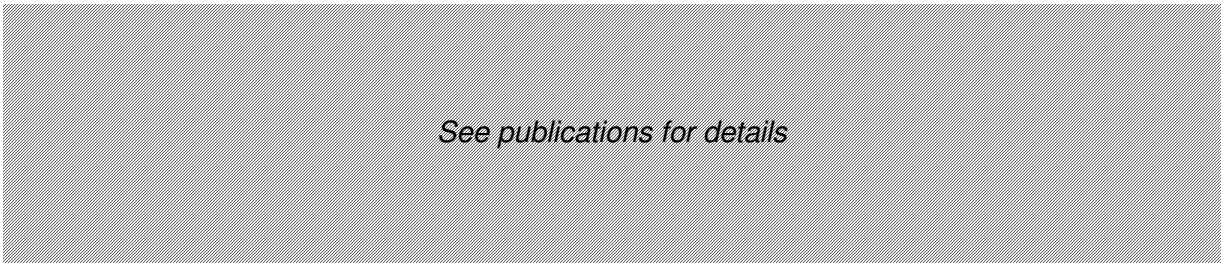
Free radical scavengers like α-phenyl-*N-tert*-butylnitrone (PBN) and 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Trolox) have been widely used as protective agents in various biomimetic and biological models. A series of three amphiphilic Trolox and PBN derivatives have been designed by adding to the parent molecules both a perfluorinated chain and a sugar group in order to render them amphiphilic. In this work, we have studied the interactions between these derivatives and lipid membranes and how they influence antioxidant properties.

The three derivatives form monolayers at the air/water interface. We next investigated the ability of each derivative to interact with 1,2-dilinoleoyl-sn-glycero-3-phosphocholine (DLPC) as well as their efficiency to inhibit the AAPH-induced oxidation of DLPC liposomes. The location of these derivatives in the membrane is a key parameter to rationalize their antioxidant efficiency. The derivative bearing both the PBN and the Trolox antioxidant moieties on the same fluorinated carrier exhibited a synergistic antioxidant effect by delaying the oxidation process. Molecular dynamics (MD) simulations supported the understanding of the mechanism of action, highlighting various key physical-chemical descriptors.

### Keywords

Lipid oxidation, amphiphilic nitrone, synergism.

### III.5.2. Materials and methods




*Theoretical methodology*

Two DLPC membranes made of 128 and 72 lipids each, were created using the membrane bilayer builder from the CHARMM-GUI server(32). The membranes were solvated with a hydration number of 35 water molecules per lipid. $Na^+$ and $Cl^-$ ions were added at a 0.154 M concentration, ensuring the neutrality of the system. The DLPC lipids were described using the lipid 11 force field(33). The force field parameters for the different compounds (Trolox, anionic Trolox, PBN, FATx, FAPBN, FATxPBN) were derived from the Generalized Amber Force Field version 2 (GAFF2)(34–36) for all moieties but the sugars, which were described by the GLYCAM force field (37), using the antechamber package(38). FATx, FAPBN and FATxPBN were built using a building block approach (*i.e.*, considering the β-D-galactopyranosyl, 2,6-diaminohexanoyl, 1-amino-dodecafluorooctanyl, Trolox and PBN residues). Atomic charges were derived from RESP (Restrained fit of Electro Static Potential) based on the calculations achieved within the Density Functional Theory (DFT) formalism with the IEFPCM-B3LYP/cc-pVDZ method in diethylether(39), accounting for the chemical environment in the building block approach. The DFT calculations and the atomic charge fitting were performed with the Gaussian 09, RevA(40) and R.E.D. III(41) softwares, respectively. The three-point TIP3P water model(42) was used to describe water molecules. The anionic Trolox was used because, regarding the low pKa of the carboxylic moiety, Trolox exists in its deprotonated form in the bulk water, and both (neutral and anionic) forms may co-exist in the bilayer. However, regarding insertion in membrane, no significant differences were observed between the two forms, therefore the results have not been detailed.

MD simulations were carried out using CPU-Particle-Mesh Ewald (PME)(43) MD codes available in Amber16(44, 45), and according to the following procedure: minimization of water molecules prior to minimization of the entire system to prevent from steric clash; slow thermalization of the water molecules up to 100 K in the (N,V,T) ensemble for 200 ps; thermalization of the whole system to the final temperature (310 K) for 500 ps in the (N,P,T) ensemble; equilibration of the system for 5 ns (N,P,T) MD simulations. Productions of 600 ns (N,P,T) MD simulations were then achieved. PME MD simulations were carried out using the SHAKE algorithm(46) on H-bonds and a 10 Å non covalent interaction cut-off for both Coulomb and van der Waals interactions. The temperature was maintained using the Langevin dynamics(47) with a collision frequency of 1 $ps^{-1}$. Semi-isotropic pressure scaling with constant surface tension in the xy-plane perpendicular to the membrane normal (z-axis) was used with Berendsen barostat(48), in which the pressure relaxation time was set to 1 ps.

The compounds were inserted in the bulk water of the equilibrated membrane systems, preventing from steric clash with water molecules. MD simulations of 600 ns were then carried out with all the compounds; the total MD simulation time for the six derivatives was 3.6 µs. The analyses were performed along the last 300 ns, which allowed a robust sampling of structural properties, *i.e.*, after the equilibrium was reached. They were carried out using the cpptraj

software(49). The positions of the center of mass (COM) or key moieties of the different compounds in membranes are given with respect to the middle of the bilayer, (*i.e.,* z-component of the vector originated at the COM of the lipid bilayer and pointing towards the COM of the compound). The orientation of the compounds in the lipid bilayer was assessed by the α-angle between the z-axis and the normal vector(s) defined for each compound (see supplemental data).

### III.5.3. Results and discussion

*See publications for details*

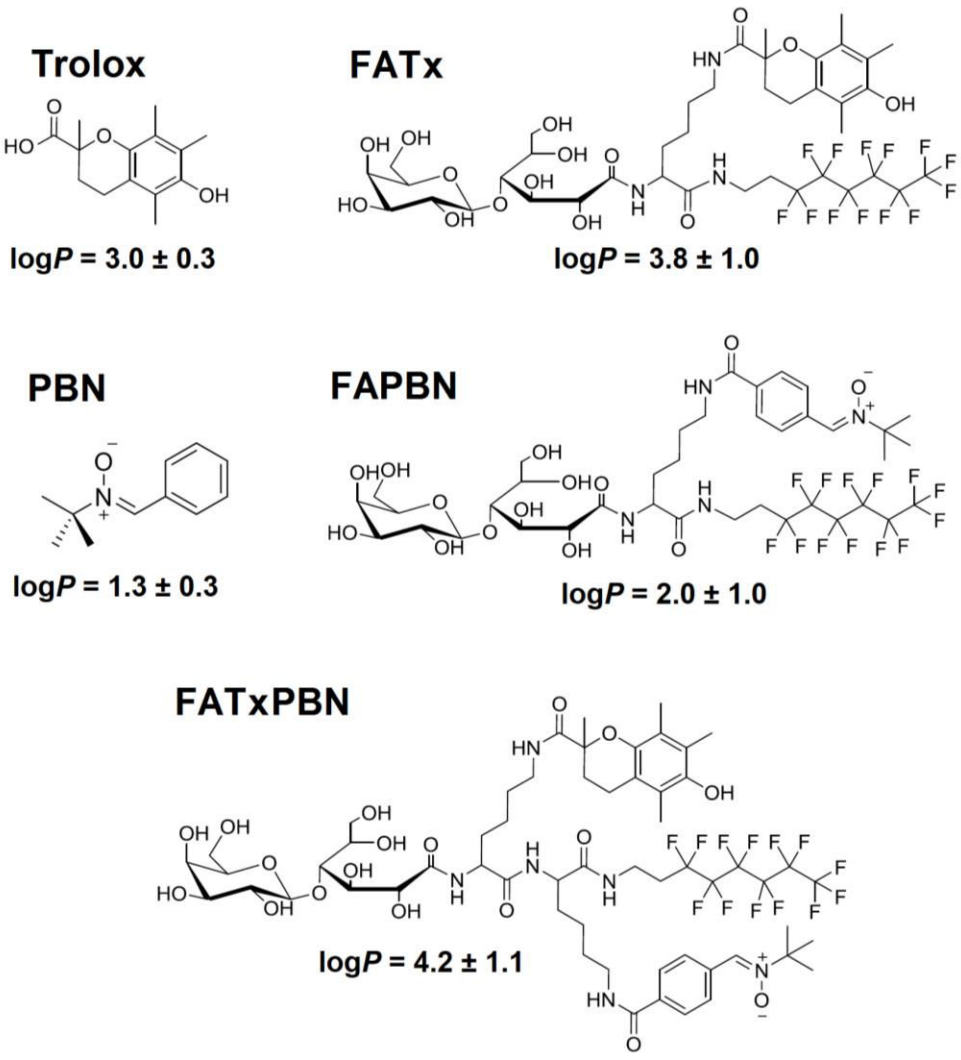



Fig. 1. Chemical structures and log*P* values of the studied molecules.


*MD simulations of derivative positioning in DLPC membrane*

All derivatives inserted into the DLPC lipid bilayers when starting from the bulk water. The preferred position of the center of mass (COM) of Trolox and PBN were just below the polar head group region in close contact with the phosphates and to a less extent with the choline moieties (14.0 and 13.6 Å, respectively from the middle of the bilayer, see Table 2 and Fig. S6 for the radial distribution function). The fluorinated antioxidants also partition just below the polar head group region, the COM being at 14.3, 14.7 and 13.7 Å from the middle of the bilayer for FATx, FAPBN and FATxPBN, respectively (Table 2). The different moieties drive the preferred location of these amphiphilic derivatives. Namely, from one side, both $\beta$-D-Galactose and gluconic acid moieties strongly anchor the derivatives to the choline and the phosphate groups (*ca*. 16.5 and 20.0 Å, respectively, from the middle of the bilayer) by electrostatic and H-bonding interactions (Fig. 6). From the other side, the fluorinated chain lines up along the DLPC lipid chains (COM of the fluorinated chain ranging from 8.2 to 10.4 Å, see Table 2), anchoring the derivatives into the membrane (Fig. 6). In the case of fluorinated antioxidants, the Trolox and PBN moieties fluctuate below the polar head group region (ranging from 13.8 to 15.7 Å from the middle of the bilayer, see Table 2) with a preferred orientation parallel to the surface (Fig. S7) in close contact with the phosphates, as for the parent Trolox and PBN derivatives. Interestingly, in the case of FAPBN, a minor conformer was observed inside the membrane, which drives the PBN moiety deeper in the membrane (9.0 Å from the middle of the bilayer, see Table 2), in close contact with the fluorinated chain (Fig. 6D). This minor conformer corresponds to a 3-dimensional molecular arrangement which has been observed during short-lasting period (compared to the major conformer) along the MD simulation time. In this conformer, PBN is lining up along the fluorinated chain (Fig. 6D), whereas in the major conformer (seen during long-lasting period of the MD simulations), both PBN and Trolox moieties are far apart from the fluorinated chain. Within this molecular arrangement, inhibition of the propagation stage is likely to occur. A similar conformational rearrangement is also expected in FATxPBN, *i.e.*, lining up the PBN moiety along the fluorinated chains. Conversely, similar alignment of the Trolox moiety and the fluorinated chain (in FATx and FATxPBN) has appeared unlikely, even though we have conducted a MD simulation in which these two moieties were in closer contact. It is worth noting that no intramolecular interactions between the Trolox and PBN moieties were observed in FATxPBN throughout the whole MD simulation.

Table 2. Positions given as distances from the middle of the bilayer (in Å) of key moieties of DLPC (*i.e.,* choline, phosphates, unsaturations) and the different derivatives Trolox, Anionic Trolox, PBN, FATx, FAPBN, minor conformer of FAPBN (FAPBN*) and FATxPBN (*i.e.,* Trolox moiety, Trolox OH-phenolic, PBN, PBN amine moiety, 1LB = β-D-Galactose, GA4 = gluconic acid; DAH = 2,6-diaminohexanoic acid, DFA = 3,3,4,4,5,5,6,6,7,7,8,8,8-dodecafluorooctan-1-amine), as well as center-of-mass (COM) of the three fluorinated antioxidants FATx, FAPBN and FATxPBN. The positions were averaged over the second half of the MD simulations.

| | **Trolox** | **Anionic Trolox** | **PBN** | **FATx** | **FAPBN** | **FAPBN*** | **FATxPBN** |
|---|---|---|---|---|---|---|---|
| **Cholines** | 20.4 ± 0.4 | 20.3 ± 0.4 | 20.5 ± 0.4 | 20.5 ± 0.3 | 20.5 ± 0.3 | 20.5 ± 0.3 | 20.5 ± 0.3 |
| **Phosphates** | 19.1 ± 0.4 | 19.0 ± 0.4 | 19.2 ± 0.3 | 19.2 ± 0.3 | 19.2 ± 0.2 | 19.2 ± 0.3 | 19.2 ± 0.3 |
| **1st lipid unsaturation** | 7.2 ± 0.3 | 7.2 ± 0.3 | 7.2 ± 0.3 | 7.2 ± 0.2 | 7.3 ± 0.2 | 7.3 ± 0.2 | 7.2 ± 0.2 |
| **2nd lipid unsaturation** | 5.6 ± 0.2 | 5.6 ± 0.2 | 5.7 ± 0.2 | 5.7 ± 0.2 | 5.7 ± 0.1 | 5.7 ± 0.2 | 5.7 ± 0.2 |
| **Trolox** | 14.0 ± 1.7 | 15.3 ± 1.8 | | 14.0 ± 2.3 | | | 13.9 ± 1.7 |
| **Trolox phenolic OH group** | 15.1 ± 2.2 | 14.2 ± 2.2 | | 14.7 ± 2.6 | | | 14.0 ± 2.3 |
| **PBN** | | | 13.6 ± 2.7 | | 15.7 ± 2.5 | 9.0 ± 2.1 | 13.8 ± 1.5 |
| **PBN amine** | | | 13.8 ± 2.8 | | 15.2 ± 2.9 | 7.4 ± 2.0 | 13.9 ± 1.8 |
| **1LB** | | | | 16.7 ± 1.7 | 16.3 ± 2.0 | 17.6 ± 1.7 | 16.8 ± 1.9 |
| **GA4** | | | | 20.4 ± 1.5 | 20.0 ± 1.9 | 20.9 ± 1.8 | 19.6 ± 1.7 |
| **DAH** | | | | 16.3 ± 1.9 | 16.2 ± 1.8 | 16.1 ± 2.1 | 15.2 ± 1.2 |
| **DFA** | | | | 10.0 ± 1.8 | 10.4 ± 1.9 | 11.1 ± 1.9 | 8.2 ± 1.6 |
| **FATx COM** | | | | 14.3 ± 1.7 | | | |
| **FAPBN COM** | | | | | 14.7 ± 1.7 | 13.9 ± 1.8 | |
| **FATxPBN COM** | | | | | | | 13.7 ± 1.1 |

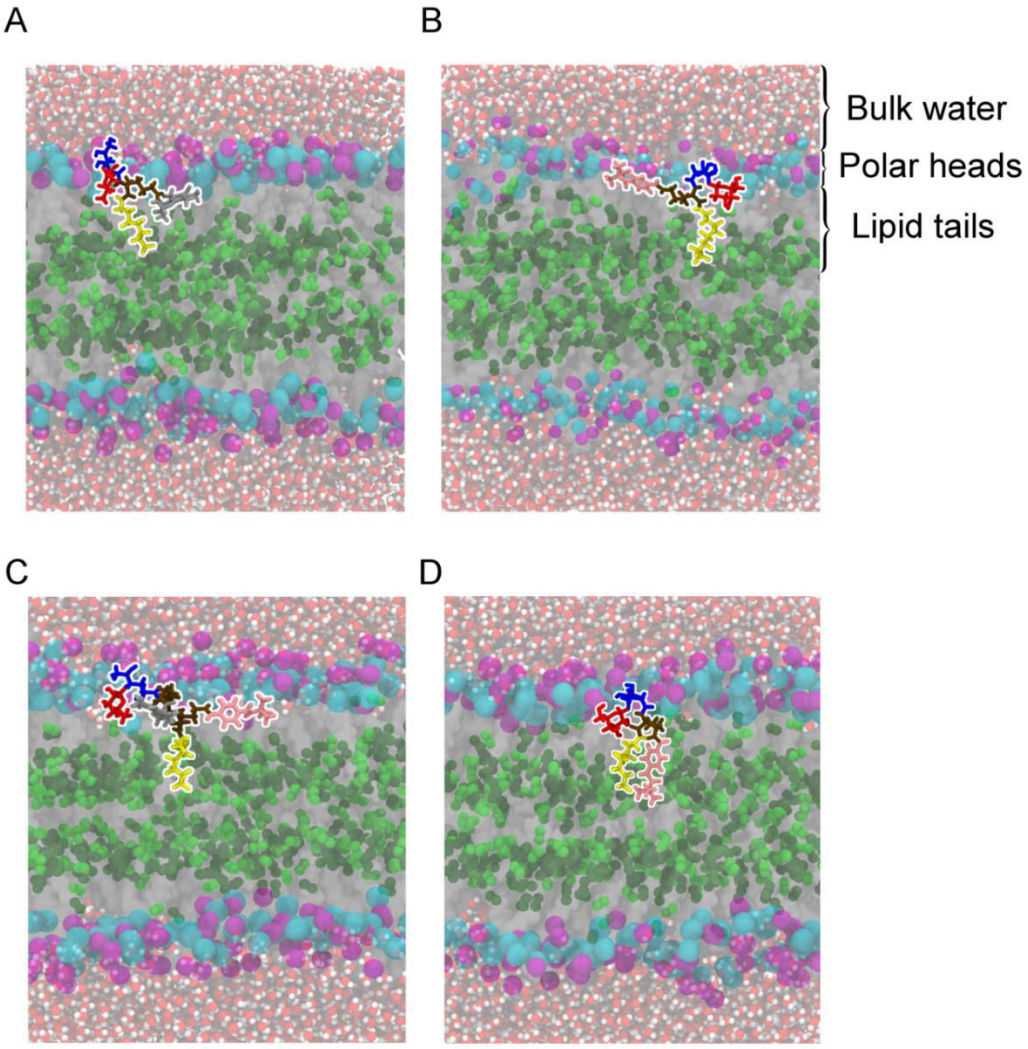



Fig. 6. Representative snapshots of the preferred location of FATx (A), FAPBN (B) and FATxPBN (C). Fig. 6D depicts the minor conformation of FAPBN, in which the PBN moiety is lined up along the fluorinated chain. The 1LB (β-D-Galactose), GA4 (gluconic acid), DAH (2,6-diaminohexanoic acid) and DFA (3,3,4,4,5,5,6,6,7,7,8,8,8-dodecafluorooctan-1-amine) moieties are depicted in red, blue, brown and yellow, respectively. Trolox and PBN are depicted in grey and pink, respectively. The cholines are depicted in cyan, the phosphates are depicted in magenta, the 1st and 2nd unsaturations are depicted in green and dark-green, respectively.

*Mechanism of lipid peroxidation inhibition of FATx, FAPBN and FATxPBN*

*See publications for details*

The differences observed between FAPBN and the other two fluorinated antioxidants can be rationalized by a lower affinity to lipid (slightly lower hydrophobicity, see log*P* on Fig. 1) of the former parent compound (*i.e.,* PBN) and by the absence of the Trolox moiety in this derivative. The scavenging mechanisms of both Trolox and PBN are different (most probably hydrogen atom abstraction and adduct formation – spin trap –, respectively), which makes the former

derivative a much better free radical scavenger. Concomitantly, the amphiphilic feature of the derivatives finely tunes the capacity for lipid peroxidation inhibition, by ensuring a positioning in the membrane below the polar head groups, in close contact with the oxidation process (*i.e.*, as close as possible from unsaturation). At this preferred location, the derivatives are likely to inhibit both the initiation stage or the propagation stage for lipid chains which can adopt a transient snorkel-like shape(55, 56) of lipid peroxidation. Also, it is well-known that location of the antioxidants in close contact to the unsaturations of the phospholipid acyl chains favors a better inhibition of lipid peroxidation(57).

*See publications for details*

Although both PBN itself and the PBN moiety of FAPBN lie at almost the same preferred position, one can expect an anchoring effect attributed to the fluorinated chain which better constrain (anchor) the PBN moiety positioning in the membrane, which can more efficiently act as a spin trap. In its minor FAPBN conformation (Fig. 6D), the PBN moiety may even spend some time deeper in the bilayer to inhibit lipid peroxidation more efficiently than PBN alone, however not as efficient as both FATxPBN and FATx due to a much less efficient free radical scavenging capacity.

FATx is 33 times more active than FAPBN, which is mainly attributed to the nature of the antioxidant moiety and not to positioning. Indeed, Trolox is a much more efficient free radical scavenger than PBN, while both moieties are positioned in the same membrane region, in close contact with the polar head group region.

*See publications for details*

In the case of FATxPBN, no intramolecular interactions between both antioxidant (Trolox and PBN) moieties were observed by MD simulations, which preclude regeneration effect within the same molecule. However, regarding to the similar depth of insertion of both Trolox and PBN moieties, intermolecular interactions are likely to occur, which could favor a regeneration process.

### III.5.4. Conclusion

The present work is in line with the strategy aiming at grafting two antioxidant moieties on the same carrier to enhance the global antioxidant action with a nanoscale fine tuning. The key role of the amphiphilic feature of this series of fluorinated antioxidants bearing Trolox and/or PBN moieties was studied in terms of interaction with membranes. The presence of the perfluorinated chain increases hydrophobicity, allowing these derivatives to form monolayers at the air-water interface. Because of higher hydrophobicity, they are anchored to the DLPC

membranes, with a greater antioxidant activity against AAPH-induced lipid oxidation. Interestingly, although PBN is inefficient, its combination with a Trolox moiety on the same carrier creates a synergistic effect that improves the global antioxidant efficiency. Thanks to a dual mode of action, such hybrid antioxidants are very promising molecules for a wide range of applications (cosmetic, pharmaceutical and food) by lowering the concentrations required to efficiently fight against lipid oxidation.

## References


[32] E.L. Wu, X. Cheng, S. Jo, H. Rui, K.C. Song, E.M. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R.M. Venable, J.B. Klauda, W. Im, CHARMM-GUI Membrane Builder toward realistic biological membrane simulations, J. Comput. Chem. 35 (2014) 1997–2004. doi:10.1002/jcc.23702.

[33] Å.A. Skjevik, B.D. Madej, R.C. Walker, K. Teigen, LIPID11: A Modular Framework for Lipid Simulations Using Amber, J. Phys. Chem. B. 116 (2012) 11124–11136. doi:10.1021/jp3059992.

[34] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (2004) 1157–1174. doi:10.1002/jcc.20035.

[35] J. Wang, T. Hou, Application of Molecular Dynamics Simulations in Molecular Property Prediction. 1. Density and Heat of Vaporization, J. Chem. Theory Comput. 7 (2011) 2151–2165. doi:10.1021/ct200142z.

[36] D.L. Mobley, C.I. Bayly, M.D. Cooper, M.R. Shirts, K.A. Dill, Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations, J. Chem. Theory Comput. 5 (2009) 350–358. doi:10.1021/ct800409d.

[37] K.N. KIRSCHNER, A.B. YONGYE, S.M. TSCHAMPEL, J. GONZÁLEZ-OUTEIRIÑO, C.R. DANIELS, B.L. FOLEY, R.J. WOODS, GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates, J Comput Chem. 29 (2008) 622–655. doi:10.1002/jcc.20820.

[38] J. Wang, W. Wang, P.A. Kollman, D.A. Case, Automatic atom type and bond type perception in molecular mechanical calculations, Journal of Molecular Graphics and Modelling. 25 (2006) 247–260. doi:10.1016/j.jmgm.2005.12.005.

[39] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, J. Comput. Chem. 24 (2003) 1999–2012. doi:10.1002/jcc.10349.

[40] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssel, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, Gaussian 09, Version A02, Gaussian, Inc., Wallingford CT, 2016.

[41] F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, P. Cieplak, The R.E.D. Tools: Advances in RESP and ESP charge derivation

and force field library building, Phys Chem Chem Phys. 12 (2010) 7821–7839. doi:10.1039/c0cp00111b.

[42] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, The Journal of Chemical Physics. 79 (1983) 926–935. doi:10.1063/1.445869.

[43] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, The Journal of Chemical Physics. 98 (1993) 10089–10092. doi:10.1063/1.464397.

[44] Salomon-Ferrer Romelia, Case David A., Walker Ross C., An overview of the Amber biomolecular simulation package, Wiley Interdisciplinary Reviews: Computational Molecular Science. 3 (2013) 198–210. doi:10.1002/wcms.1121.

[45] D.A. Case, D.S. Cerutti, T.E.. Cheatham, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, AMBER 2017;University of California: San Francisco, CA, USA, 2017.

[46] J.-P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, Journal of Computational Physics. 23 (1977) 327–341. doi:10.1016/0021-9991(77)90098-5.

[47] R.J. Loncharich, B.R. Brooks, R.W. Pastor, Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide, Biopolymers. 32 (1992) 523–535. doi:10.1002/bip.360320508.

[48] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, The Journal of Chemical Physics. 81 (1984) 3684–3690. doi:10.1063/1.448118.

[49] D.R. Roe, T.E. Cheatham, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, J. Chem. Theory Comput. 9 (2013) 3084–3095. doi:10.1021/ct400341p.

[55] D. Marquardt, J.A. Williams, N. Kučerka, J. Atkinson, S.R. Wassall, J. Katsaras, T.A. Harroun, Tocopherol Activity Correlates with Its Location in a Membrane: A New Perspective on the Antioxidant Vitamin E, J. Am. Chem. Soc. 135 (2013) 7523–7533. doi:10.1021/ja312665r.

[56] J. Garrec, A. Monari, X. Assfeld, L.M. Mir, M. Tarek, Lipid Peroxidation in Membranes: The Peroxyl Radical Does Not "Float," J. Phys. Chem. Lett. 5 (2014) 1653–1658. doi:10.1021/jz500502q.

[57] P. Košinová, K. Berka, M. Wykes, M. Otyepka, P. Trouillas, Positioning of Antioxidant Quercetin and Its Metabolites in Lipid Bilayer Membranes: Implication for Their Lipid-Peroxidation Inhibition, J. Phys. Chem. B. 116 (2012) 1309–1318. doi:10.1021/jp208731g.

## Chapitre IV. Membrane proteins

### IV.1. Structural patterns of the human ABCC4/MRP4 exported in lipid bilayers rationalize clinically observed polymorphisms

B. Chantemargue[a,b,†], F. Di Meo[a,†,*], K. Berka[b], N. Picard[a,c], H. Arnion[a], M. Essig,[a] P. Marquet[a,c], M. Otyepka[b], P. Trouillas[a,b]

*Published, DOI*: *10.1016/j.phrs.2018.02.029*

#### Foreword

*This study aims at better understanding clinically observed toxicities related to the ABCC4/MRP4 protein. To this end, a molecular model of the ABCC4/MRP4 exporter was constructed. MD simulations allowed to report the surrounding lipid bilayer, especially cholesterol on this protein. We addressed several structural and energetic issues about the different domains involved in xenobiotics export, including the so far poorly resolved $L_1$, $L_0$ and zipper domains. The dramatic structural impact of the p.Gly187Trp mutation was rationalized.*

#### IV.1.1. Abstract

The ABCC4/MRP4 exporter has a clinical impact on membrane transport of a broad range of xenobiotics. It is expressed at key locations for drug disposition or effects such as in the liver, the kidney and blood cells. Several polymorphisms and mutations (*e.g.*, p.Gly187Trp) leading to MRP4 dysfunction are associated with an increased risk of toxicity of some drugs. So far, no human MRP4 structure has been elucidated, precluding rationalization of these dysfunctions at a molecular level. We constructed atomistic model of the wild type (WT) MRP4 and the p.Gly187Trp mutant embedded in different lipid bilayers and relaxed them for hundreds of nanoseconds by molecular dynamics simulations. The WT MRP4 molecular structure confirmed and ameliorated the general knowledge about the transmembrane helices and the two nucleotide binding domains. Moreover, our model elucidated positions of three generally unresolved domains: $L_1$ (linker between the two halves of the exporter); $L_0$ (N-terminal domain); and the zipper helices (between the two NBDs). Each domain was thoroughly described in view of its function. The p.Gly187Trp mutation induced a huge structural impact on MRP4, mainly affecting NBD 1 structure and flexibility. The structure of transporter enabled rationalization of known dysfunctions associated with polymorphism of MRP4. This model is available to the pharmacology community to decipher the impact of any other clinically observed polymorphism and mutation on drug transport, giving rise to in silico predictive pharmacogenetics.

#### Keywords

ABCC4/MRP4 exporter; Molecular dynamics; *In silico* pharmacogenomics; Lipid-protein interactions; Polymorphism and mutation; Atomistic elucidation.

### IV.1.2. Introduction

Drug crossing through cell and organelle membranes affects drug disposition, activity and toxicity *in vivo*. Membrane crossing can proceed either by (unassisted) passive permeation or through membrane protein transporters. The International Transporter Consortium (ITC) has described a series of human membrane transporters as being of "emerging clinical importance"[1]. This includes ABCB1 (known as P-glycoprotein, P-gp), a highly described member of the superfamily of ATP-binding cassette (ABC) transporters (exporters in mammals). ABC dysfunction is often associated with multidrug resistance and drug-drug interactions in the form of competitive affinity for these transporters, with a clinical impact in many therapeutic domains.

Less attention has been paid to ABCC4/MRP4 (multidrug resistance protein type 4), although it is expressed at key locations for drug disposition/action such as in the liver, the kidney and peripheral blood cells, and it has a broad range of xenobiotics as substrates[2]. Dysfunction of this exporter can modify drug pharmacokinetics or pharmacodynamics. Among the mutations associated with MRP4 dysfunction, the p.Gly187Trp polymorphism[3] is the most frequent in Caucasians and it was described as having the greatest impact on the membrane transport of the two drugs PMEA and AZT (9-(2-phosphonylmethoxyethyl)adenine - adefovir - and azidothymidine, respectively)[4]. This polymorphism was also associated with decreased efflux of ganciclovir (GCV) *in vitro*[5], which is associated with an increased risk of toxicity due to the intracellular accumulation of the drug in neutrophils. The impact of other rare mutations on MRP4 function has been poorly documented so far, which does not allow for a systematic prediction of the overall MRP4 (dys)function and its use in clinical decisions (*e.g.*, in the form of biomarkers or next-generation sequencing data).

Despite highly variable amino acid sequences due to the large number of variants that appeared with phylogeny, all ABC transporters exhibit surprising structural similarities. In their functioning form, they are constituted of two transmembrane domains (TMDs) and two nucleotide binding domains (NBDs) with very similar secondary and tertiary structures. Their catalytic sites, where ATP molecules are hydrolyzed, are in two well-conserved and well-defined regions of NBDs[6]. To date, neither the crystallographic nor the cryo-electronic microscopic (cryo-EM) structures of any MRP4 protein have been elucidated. Nonetheless, it is known to bear: twelve transmembrane helices (TMH 1-12); two NBDs (NBD 1 and 2); the N-terminal domain, which is constituted of the $L_0$ motif; and the linker $L_1$ making a covalent linkage between the two halves of the exporter (Fig. 1A). The supramolecular packing between TMHs is of crucial importance for drug accessibility to the dynamic transmembrane pore, allowing transport from the inner (cytoplasmic) to the outer compartments[7]. The TMHs are arranged in bundles A, B, C and D, which result from an entanglement of the different TMHs of the two halves of MRP4 (Fig. 1). ECL and ICL are the extra- and intra-cytoplasmic loops connecting the TMHs with each other, and ICD 1 and 2 are two intra-cytoplasmic domains (covalently) connecting TMDs and NBDs.

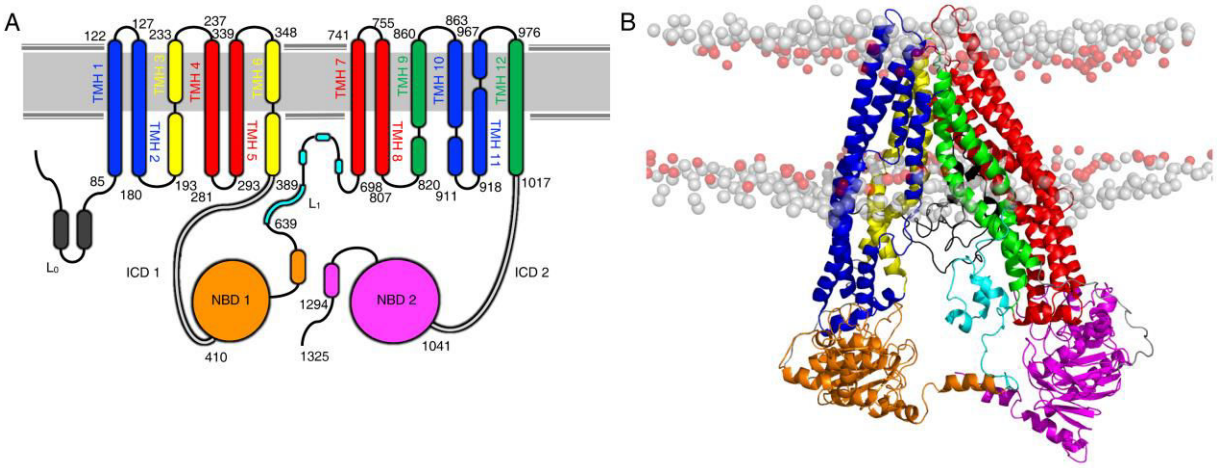



Fig. 1. A) Topology of the human MRP4 exporter, highlighting: the two TMDs made of twelve TMHs (TMH 1-12); the two NBDs (NBD 1 and 2); ICD 1-2; the linker $L_1$; the zipper helices; the $L_0$ motif. If any, the irregularities in helices are depicted. B) Fully relaxed 3D-structure of the molecular WT MRP4 model.

Although pure homology molecular models can document on the different domains[8-13], they do not include the surrounding lipid bilayer and a sufficient (dynamic) conformational sampling. In the current study, we used homology modeling to construct the human wild-type (WT) MRP4 protein in its inward-facing (IF) conformer (*i.e.*, drug chamber open towards inner compartment) and in its ligand-free (apo) state. To become predictive and descriptive, this structure was embedded in different lipid bilayers. The models were relaxed by 300 ns molecular dynamic (MD) simulations. This was mandatory to relax any bias associated with homology modeling construction, thus ensuring correction of both secondary and tertiary structures, especially for the most flexible and the less known domains. The topology of the models was thoroughly and critically analyzed with respect to known experimental data. MD simulations also give the considerable advantage to track conformational motions in lipid bilayer. The p.Gly187Trp mutant molecular model was also constructed, which gave clues about how this single nucleotide polymorphism affects the tertiary structure of MRP4, at an atomic resolution, rationalizing its subsequent transport dysfunction.

### IV.1.3. Methods

*Homology modeling*

The human MRP4 amino acid sequence was retrieved from the UniProt database[14] with the primary accession number NM_005845.4. Isoform 1 was chosen as being the canonical form. The construction of the IF model of the human wild-type (WT) MRP4 protein was achieved using the I-TASSER webserver[15]. In the first step of this construction, three different relevant templates were identified, namely P-gp from *C. elegans* (PDB ID: 4F4C, X-ray crystallographic structure obtained at 3.4 Å resolution)[16], and from *Mus musculus* (PDB ID: 3G5U and 4M1M, X-ray crystallographic structures, both obtained at 3.8 Å resolution)[17,18]. The robustness of this molecular WT MRP4 model was assessed by a series of 'sequence and structure' similarity scores (see Table S1).

Protonation states of all histidine residues were systematically checked according to the neighboring chemical environment to maximize hydrogen bond networking. Histidines 55, 152, 172, 466, 472, 592, 710, 831, 903, 1037, and 1233 were defined as ε-protonated; Histidines 35, 153, 158, 213, 411, 572, 583, 930, 934, and 1111 were defined as δ-protonated; and Histidines 798, 1060, 1141, and 1225 were defined as double-protonated.

After the construction of this molecular WT MRP4 model in its IF form, the end of the sequence Lys1294 - Leu1325 (PDZ binding domain including the ETAL motif at the end of the sequence) was removed from the structure, as this flexible part (loop plus α-helix) is expected to be far from the NBD domains and likely involved in protein-protein interactions[19]. The molecular p.Gly187Trp mutant model was built from the optimized WT MRP4 model by simply replacing the amino acid residue.

*Molecular dynamics simulations*

Prior to MD simulations, the WT MRP4 model was embedded into three different POPC (1-palmitoyl,2-oleoyl-*sn*-glycero-3-phosphocholine)-based lipid bilayers including 0, 25 and 50 mol. % cholesterol (Chol), *i.e.*, (1:0), (3:1) and (1:1) POPC:Chol ratios, using the CHARMM-GUI membrane builder[20]. The p.Gly187Trp mutant model was embedded into the (3:1) POPC:Chol lipid bilayer. All systems were solvated by the TIP3P explicit water model[21] and neutralized with 154 mM NaCl to imitate physiological conditions. The final molecular systems were made of *ca.* 350,000 atoms. Membrane contents and box sizes are reported in Table S2. The Amber FF14SB[22] and Lipid14[23] force fields were used to model the protein and lipids (including cholesterol), respectively.

The minimization and MD simulations were performed with AMBER 16[24] using the CPU and GPU PMEMD versions, respectively[25]. Water molecules were minimized and then shortly thermalized to 100 K during 200 ps under (*N, V, T*) ensemble conditions. Afterwards, additional 500 ps thermalization steps were performed to reach 310 K under (*N, P, T*) ensemble conditions, with a pressure relaxation time of 2.0 ps. The whole systems were then equilibrated during 5 ns. Temperature was maintained using Langevin dynamics[26] with a damping coefficient of $\gamma = 1.0\ ps^{-1}$. Constant pressure (1 atm) was maintained using Berendsen barostat [27], in which pressure relaxation time was set up to 1 ps, showing stable trajectories. Non-bonded interactions were explicitly calculated using cutoff distances of 10 Å for both Coulomb and van der Waals interactions, in agreement with Lipid14 parameterization[23]; the particle mesh Ewald method[28] was applied. The integration time step was set to 2 fs and trajectory snapshots were saved every 10 ps. The MD runs produced 300 ns (*N, P, T*) trajectories, while the analyses were carried out on the last 100 ns of the trajectories to ensure stabilized geometrical features. The total simulation time for the five systems (1- WT MRP4 in 3:1 POPC:Chol bilayer; 2- WT MRP4 in 1:0 POPC:Chol bilayer; 3- WT MRP4 in 1:1 POPC:Chol bilayer; 4- WT MRP4 in 3:1 POPC:Chol bilayer - $L_0$ split from the transporter, see section 3.1 for more details; 5- p.Gly187Trp in 3:1 POPC:Chol bilayer) was *ca.* 1 µs corresponding to *ca.* 4,800 h GPU on NVIDIA K20 and K40m cards. The robustness of the WT MRP4 model was significantly increased after the 300 ns MD simulations (see Supplementary Data for all methodological and analysis details).

Conversely to simple static homology model, the 300 ns MD simulations allowed a correct structural organization of any single residue and its direct chemical environment. This was analyzed by assessing local residue-residue contact and H-bonding over the last 100 ns (see supplementary information), which allowed to anticipate how a given mutation may affect the structure of the protein.

### IV.1.4. Results and discussion

*Topology description of WT MRP4*

*TMD pattern* - The packing between TMHs was analyzed in terms of energy of interaction between bundles and TMHs with a qualitative perspective. As expected, the energy of interaction between TMHs belonging to a given bundle was high with respect to the total energy (Table S4). Using only one bundle as an example: the formation of bundle A (made of TMHs 1, 2, 10 and 11) is mainly attributed to strong attractive interactions between i) TMH 2 and 11 (-300 kcal.mol$^{-1}$); ii) TMH 10 and 11 (-170 kcal.mol$^{-1}$); iii) TMH 1 and 2 (-105 kcal.mol$^{-1}$).

Bundles A and C from one side and bundles B and D from the other side exhibited large energies of attraction of -460 and -403 kcal.mol$^{-1}$, respectively (Table S5). This confirms the interlacing of bundles A&C and bundles B&D in the IF-apo-state of MRP4. Interestingly, intra-bundle interactions are more likely driven by van der Waals interactions, whereas inter-bundle interactions are mainly driven by Coulomb interactions. This is consistent with general knowledge about the role of the substrate in IF-to-OF (outward-facing) large-scale conformational transitions. Indeed, substrate binding depends on charge distribution in the protein chamber[8,9,12]. The noncovalent electrostatic interactions between substrates and amino acids of this chamber drive a very dynamic process, allowing for bundle interchanges, following a complex energetic landscape to avoid huge energetic barriers[29]. This is fully supported by recent findings showing that substrate binding is simultaneous to ATP-binding and hydrolysis[30].

The ECLs and ICLs are connecting the different TMDs between them. Interestingly, the short ECLs (ECL 2 and 5) are not exactly extra-cytoplasmic, because they are pulled inside the lipid bilayer beneath the polar head group region. ECL 3 and 6 are lying at the membrane surface, at the bottom of the curvature (Fig. S3). ICLs are in close contact with NBDs and will be described in the next section.

*NBD pattern* - This new molecular MRP4 model confirms previous knowledge on NBD geometrical features and it ameliorates the description of inter-domain interactions typical of MRP4. The NBDs are mainly made of highly conserved regions constituted of different motifs, namely: A-loop, walker A (or P-loop), Q-loop, ABC signature (C-motif, or LSGGQ motif), walker B, D-loop, H-loop (switch motif) (see sequences in Fig. S4). This explains why most ABC transporters exhibit similar transport mechanisms, which involve ATP binding followed by its hydrolysis.

The two NBDs of MRP4 are asymmetric, *i.e.*, NBD 2 exhibits a longer chain than the consensus NBD 1 domain. Structurally, this induces slight conformational changes between the two NBDs, mainly in the signature motif, where the end of the helical arrangement is disrupted in NBD 2 (Fig. S4), possibly decreasing ATP affinity to this site.

The two NBDs are connected to TMDs either covalently through the ICDs, one in each half of the transporter, or non-covalently through the ICLs. The ICDs are unstructured external loops, which are in close contact with the two NBDs thanks to Coulomb interactions (Fig. S5 and Table S6). Due to these strong noncovalent interactions, the ICDs are poorly flexible. Interestingly, much stronger interactions were observed between NBD 2 and ICD 2 (-300 kcal.mol$^{-1}$) than between NBD 1 and ICD 1 (-170 kcal.mol$^{-1}$), see Table S6. This strengthens the assumption of the asymmetric feature of NBDs in MRP4.

The ICLs connect the different TMHs following the amino acid sequence. They are made of two very short loops and one short α-helix (*ca.* 10 amino acids long) (Fig. S5), making elbow-shaped connections between TMHs. The four ICLs have been described to make a ball-and-socket assembly with the two NBDs[31], forming noncovalent NBD/TMD joints that can efficiently transfer the ATP-hydrolysis energy, decreasing energetic barriers to favor the large conformational changes that occur in the TMDs to transport drugs. Again, energies of noncovalent interaction between ICLs and NBDs underline the NBD asymmetry feature of MRP4. NBD 1 exhibited more favorable energies of interaction with ICL 1/4 than NBD 2 with ICL 2/3 (-200 and -150 kcal.mol$^{-1}$, respectively, see Table S7). This is mostly attributed to a lower Coulomb contribution between NBD 2 and ICL 2/3, associated with a weaker hydrogen bond network.

*$L_1$ linker* - This domain is likely to play a regulatory role and it is sometimes called R-domain. To the best of our knowledge, no 3D structure of $L_1$ has been resolved by X-ray crystallography so far, suggesting a highly unstructured region. MD simulations are suitable to analyze the structure and conformational dynamics of this motif, having in mind that the initial homology models are biased by the template structures.

$L_1$ locates in the protein chamber (Fig. 1B). It consists in a pair of two structurally-different motifs, namely Glu639-Val659 and Trp660-Phe698. The former motif is highly flexible. Along the 300 ns, its structure quickly folds and unfolds (Fig. 2A). Its length fluctuates substantially from 17 Å to 45 Å (Fig. 2A). It may act as a rubber band, smoothly limiting too large opening in the IF-state. The latter motif is more structured, as it includes three short helices connected by flexible loops. Despite the presence of these loops, this motif is strongly packed into a ball-shaped nest, due to specific inter-residue hydrogen bonds (Fig. 2A). Its structure is stabilized within the first 80 ns of the MD simulation (Fig. S6). The robustness of this $L_1$ model relies on the full agreement of its structural feature with the recent cryo-EM density observed for this linker in ABCC7/CFTR (cystic fibrosis transmembrane conductance regulator)[32]. This study showed spread over EM density for the former (flexible) region, reflecting large conformational flexibility, whereas narrower EM density was observed for the latter region identifying it as more structured.

*Zipper helices* - Recently, the two C-terminal helices of a bacterial heterodimeric ABC transporter were shown to modulate the substrate transport function[33]. They were suggested first to weakly bind to each other, stabilizing NBD dimerization upon substrate binding, then to unzip and switch their positions, locking NBDs and triggering ATP hydrolysis.

The corresponding helices of MRP4 (residues 627-639 and 1282-1293 for NBD 1 and 2, respectively) were found to weakly interact in an antiparallel fashion (Fig. 2B). During MD simulations, this conformational arrangement was maintained by i) van der Waals interactions, *i.e.*, between leucine and valine and between Gly628 and Phe1293, forming CH-π interactions; ii) electrostatics, *i.e.*, polar contacts involving lysine, glutamate and aspartate residues; and iii) H-bonding, *i.e.*, involving threonine, glutamine, serine and valine residues. Both van der Waals and Coulomb interactions contributed (59 and 41%, respectively) to the total energy of interaction between these two helices. These sufficiently flexible interactions are consistent with the need of zipping and unzipping events along transport cycle[33]. This finding stress again that drug efflux through ABC exporters results from a strongly cooperative set of events.

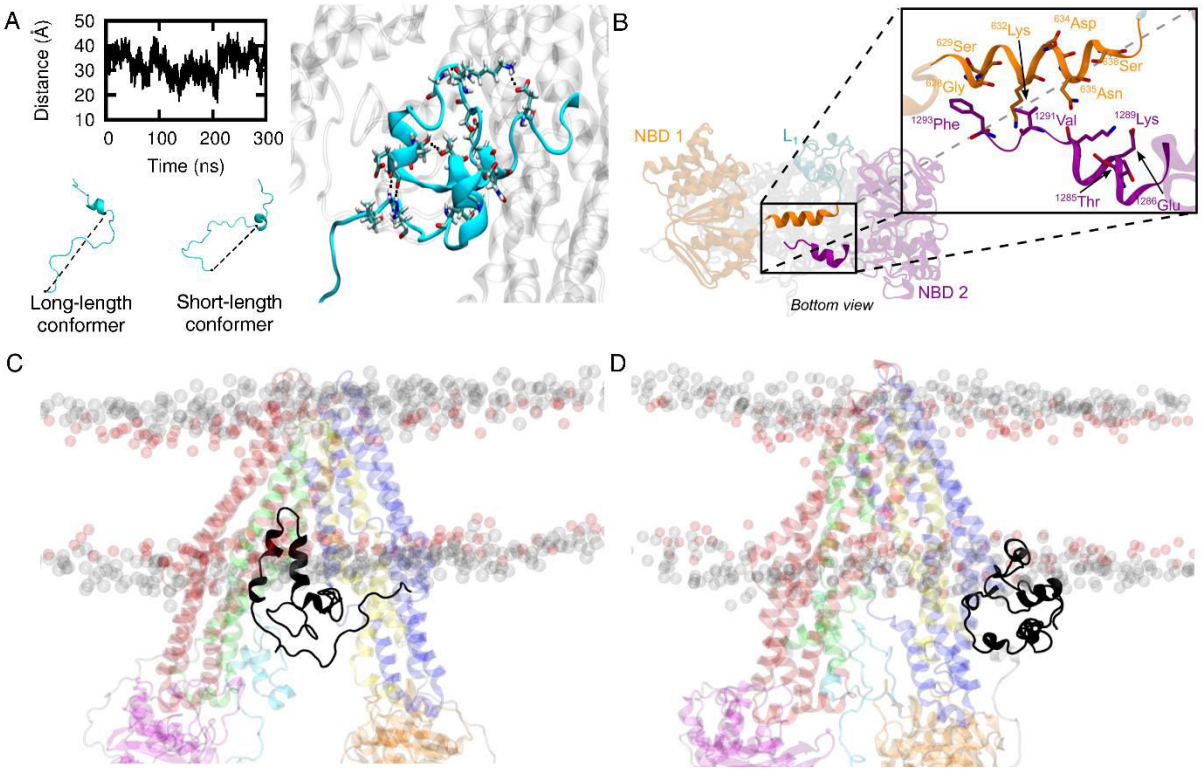



Fig. 2. Representative 3D-structures of: A) $L_1$, which is made of a flexible domain (with a fluctuating length) and a more structured domain arranged within a H-bond network; B) the zipper helices and the related interaction network; C) $L_0$ in contact with the exporter; D) $L_0$ in contact with the polar head groups of the exporter.

*$L_0$ motif* - Contrary to MRP4, in some ABC transporters, $L_0$ is the link between TMDs and the so-called TMD0 domain, which is involved in protein trafficking and protein-protein interactions [34]. In some cases, $L_0$ was described as crucial for transport, as it participates in the correct folding[32,34,35]. Possible contacts with the exporter and with $L_1$ were suggested[36,37]. The $L_0$ motif was only recently and partially elucidated structurally in CFTR by cryo-EM[32]; it was described as an interfacial motif in close contact with the transporter, in the vicinity (however without contact) with $L_1$. In the presence of TMD0, a recent cryo-EM study with a MRP1 suggested that $L_0$ is more split from the exporter[38]. As no consensus could be reached so far, we constructed two different initial geometries, namely one with $L_0$ in close contact with the exporter (as obtained the 4F4C template), the other with $L_0$ split from the exporter (with similar location than in [38]. Along the 300 ns MD simulation, the former geometry drove $L_0$ in between bundles B and D (Fig. 2C) at *ca*. 5 Å distance from these two bundles. In this position, $L_0$ appeared to be in contact with the protein pore and with $L_1$. In the latter geometry, $L_0$ was lying at the membrane interface in contact with the choline moieties of the polar head group region (Fig. 2D). Due to the hydrophilicity of $L_0$, electrostatic interactions drove its location (either embedded in between TMHs or at membrane interface, respectively), precluding any possible positioning in between the lipid tails. Interestingly, in the former case ($L_0$ in close contact with the exporter) increased inter-NBD distances by *ca.* 15 Å (Fig. S7) as compared to the latter case. The former geometry is more likely in the absence of TMD0 (*e.g.*, in MRP4), whereas the latter geometry is likely in the presence of TMD0 in other ABC transporters[34,38].

*Role of the surrounding lipid bilayer*

As mentioned above, the WT MRP4 model was embedded in (3:1), (1:0) and (1:1) POPC:Chol bilayers. The first bilayer contained a physiologically relevant amount of Chol with respect to phospholipids (taking here POPC as a representative phospholipid of biological membranes), the other two POPC:Chol ratios were mimicking two frontier cases. Although these three membrane models do not comprehensively account for the whole membrane complexity, they are expected to faithfully represent the influence of the surrounding membrane on MRP4.

*Protein and cholesterol positioning in the bilayer* - Along the 300 ns, lateral diffusion of the MRP4 protein was observed (Fig. S 8A-B). Along the vertical direction, the protein positioning was driven by tryptophan residues, which distributed at the polar head group region (Fig. S9). Salt bridges also strongly anchored the protein into the lipid bilayer, mainly between the numerous arginine residues and the polar head groups of POPC (Table S8). To a much less extent, serine, threonine, asparagine, glutamine, histidine, tyrosine, tryptophan, glutamic acid and lysine residues also formed hydrogen bonds with POPC.

The highest Chol density regions were in close contact with TMHs (Fig. 3A for 3:1 POPC:Chol model). This effect was enhanced in the (1:1) POPC:Chol model (Fig. 3B). Several flip-flop events of Chol were observed driving Chol molecules in specific contacts with TMHs' residues (Fig. 3C-E).

*Membrane-Protein effects* - MRP4 induced strong reorganization of the bilayer in the relatively close neighboring of the protein (Table S9), mainly impeding too close contacts between lipid tails, as seen by the repulsive contribution of Coulomb interactions, which are only partially compensated by attractive van der Waals interactions. Conversely, the surrounding lipid bilayer slightly but significantly affected the protein tertiary structure and flexibility, although most inter-TMHs and inter-bundle distances were similar with the three different membrane models (Fig. S10 and 11). When Chol concentration was increased up to 50% Chol, a dramatic decrease of the inter-NBD distance was observed, from 62 to 43 Å (distance between the two centers of mass). Interestingly, this drove the model structure closer to an IF-occluded conformer (Fig. S12). Also, upon increase of Chol concentration, membrane thickness increased by *ca.* 10% (Fig. S13).

These three models showed the dramatic impact of membrane composition, which in turn may affect all driving forces ruling ABC function. This agrees with Jerabek *et al.*, who showed the role of the membrane in ABC-mediated drug transport [39]. Regarding the large diversity of membranes in the organism, their contribution should be systematically considered when tackling ABC-mediated active transport.

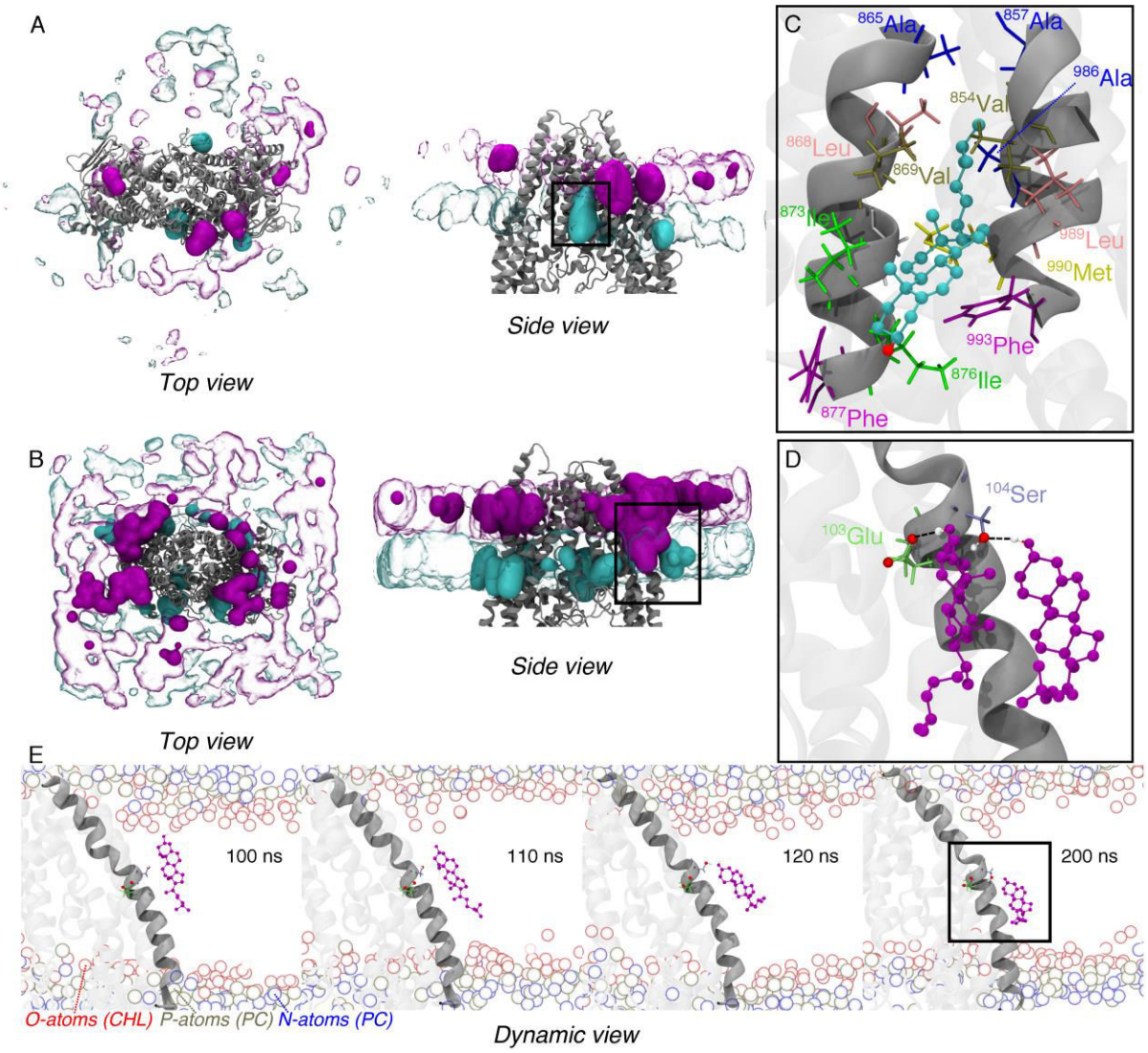



Fig. 3. Top and side views of cholesterol distributions around the exporter embedded in: A) (3:1) POPC:Chol, and B) (1:1) POPC:Chol in bilayer mixtures. The respective specific TMH-Chol interactions are zoomed in panels C) and D). E) Dynamic view of hydrogen bond-driven Chol flip-flop. Upper and lower leaflets are depicted in purple and cyan, respectively.

### *Molecular description of MRP4 polymorphisms*

*Substrate-water pocket* - Although the structural features of the MRP4 substrate-binding chamber still require clarification, distinct structural subdomains have been identified[40], with possible allosteric domains[11]. Various residues are involved in substrate binding, namely Phe368, Phe369, Glu374, Arg375, Glu378, Trp995 and Arg998 (Fig. 4A)[9,11,12]. These amino acids are clearly located at the surface of the intra-membrane water pore, in between TMHs (Fig. 4B). As already suggested[9], we confirm that this region is rich in charged amino acids (Fig. S14). We can infer that this region is also rich in aromatic amino acids, mostly phenylalanine, that can selectively bind aromatic substrates by π- π stacking (*e.g.*, ganciclovir, adefovir, furosemide, indometacin). Interestingly, one model also suggests that $L_0$ can be involved in substrate binding when it is in contact with the exporter, the Lys36-Gln61 subdomain being rich in charged amino acids (*i.e.*, Arg37, Arg38, Glu40, Glu41, Asp42, Asp43, Asp51, Arg52).

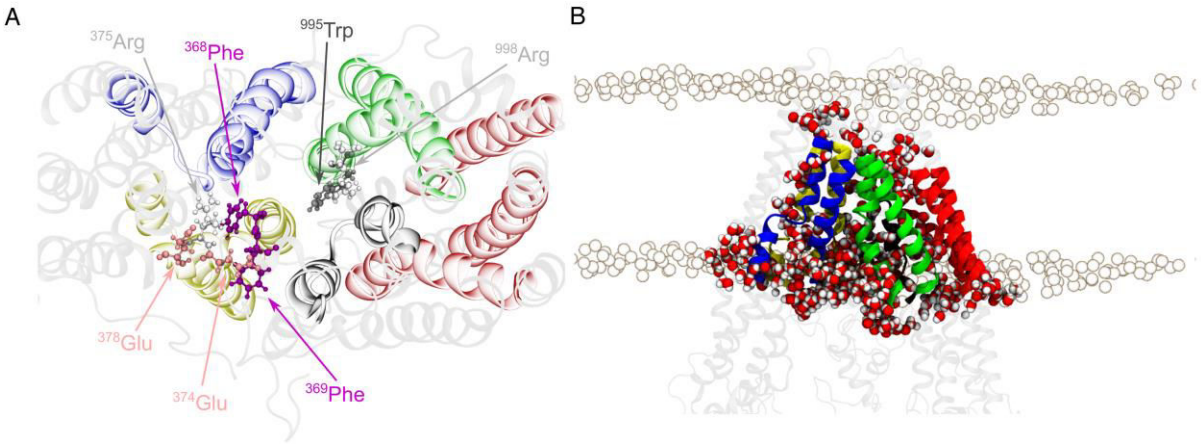



Fig. 4. A) 3D location of residues defined in the literature as playing a key role in substrate binding event. and B) representation of the substrate chamber, filled by water molecules

*Mapping of known MRP4 polymorphisms* - There exist several (naturally-occurring or induced) MRP4 polymorphisms (Table 1 and Table S10), which are spread all over the exporter (Fig. S15). Many of them affect the transporter function by affecting membrane trafficking, thus modifying or preventing protein location in the plasma membrane. This molecular MRP4 model can be used to rationalize the effects of mutations, but only for those which allow expression and insertion of the protein in the plasma membrane. In other words, here we only focus on the variants for which membrane location was confirmed (Table 1 and Fig. 5). Consequently, eleven mutations were selected, which can be classified according to their location on the molecular MRP4 model, namely in the: i) outer TMDs; ii) ATP binding sites; or iii) TMD-NBD interacting regions (Fig. 5). Interestingly, all of them but p.Glu757Lys are located in highly-conserved regions in several species (*e.g.*, *Danio rerio*, *Equus caballus*, *Oryctolagus cuniculus*, see Fig. S16).

Table 1. MRP4 variants, for which plasma membrane location have been confirmed, their effect on substrate-dependent MRP4 transport function, and their structural location in the molecular MRP4 model.

| Variant | Substrates | | | | | | Location in 3D model | |
|---|---|---|---|---|---|---|---|---|
| | $E_2$17βG[a] | $PGE_2$[a] | AZT[a] | PMEA[a] | 6-MP[a] | GCV[a] | Topology | Region/Domains |
| p.Cys171Gly [40] | ↓ | ↓ | nd[b] | nd[b] | nd[b] | nd[b] | TMH 2 | Bundles A&C/ICL 1 |
| p.Gly187Trp [4] | nd[b] | nd[b] | ↓ | ↓ | ↓ | ↓ | ICL 1 | Bundle A/ICL 1/NBD 1 |
| p.Lys304Asn [4,42] | ↓ | = | = | = | = | nd[b] | TMH 5 | Bundle B |
| p.Gly487Glu [4,42] | nd[b] | ↓ | ↓ | ↓ | ↓ | nd[b] | NBD 1 | NBD 1/ICL 1/Bundle A |
| p.Gly538Asp[c] [4] | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | NBD 1 | ABC signature |
| p.Tyr556Cys [42] | = | ↑ | nd[b] | ↑ | = | nd[b] | NBD 1 | Walker B |
| p.Glu757Lys[d] [41,42] | = | = | nd[b] | nd[b] | ↓ | nd[b] | TMH 8 | Bundle B |
| p.Arg820Ile [42] | nd[b] | nd[b] | nd[b] | ↓ | ↓ | nd[b] | ICL 3 | NBD 2/Bundle D/$L_1$ |
| p.Val854Phe [42] | nd[b] | nd[b] | nd[b] | = | = | nd[b] | TMH 9 | Bundle A&D |
| p.Ile866Val [42] | nd[b] | nd[b] | nd[b] | = | = | nd[b] | TMH 10 | Bundle A&D |
| p.Lys1081Asn[c] [42] | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | NBD 2 | Walker A |

[a] *Only organic endogenous ($E_2$17βG and $PGE_2$) and exogenous substrates (AZT, PMEA, 6-MP and ganciclovir) are discussed, omitting results about the MRP4 transport of arsenic derivatives that often exhibit a specific behaviour.*

[b] *nd: Not determined*

[c] *p.Gly538Glu and p.Lys1081Asn variants were experimentally designed and are used as negative control for MRP4 transporter function assays*

[d] *p.Glu757Lys variant has been shown to either not modify*[40] *or dramatically decrease*[41] *transport function.*

The three mutations located in the outer TMDs (p.Glu757Lys, p.Val854Phe and p.Ile866Val) were not described to significantly alter the MRP4 transporter function[40]. Here we confirm that they are not expected to dramatically modify the local physical-chemical environment and the conformational changes over the transport cycle, except p.Glu757Lys. This mutation transforms a negatively charged residue into a positively charged one. This most probably modifies the noncovalent interactions (mainly salt bridge) with TMH 7&8, *i.e.*, interaction of residue 757 with Asp760 (in the variant) instead of Lys741 (in the WT, see Fig. 5). In turn, a significant local modification of the secondary structure is expected. Because of its outer location, this is not expected to modify the function of the exporter, however this could affect membrane trafficking, as already suggested[41].

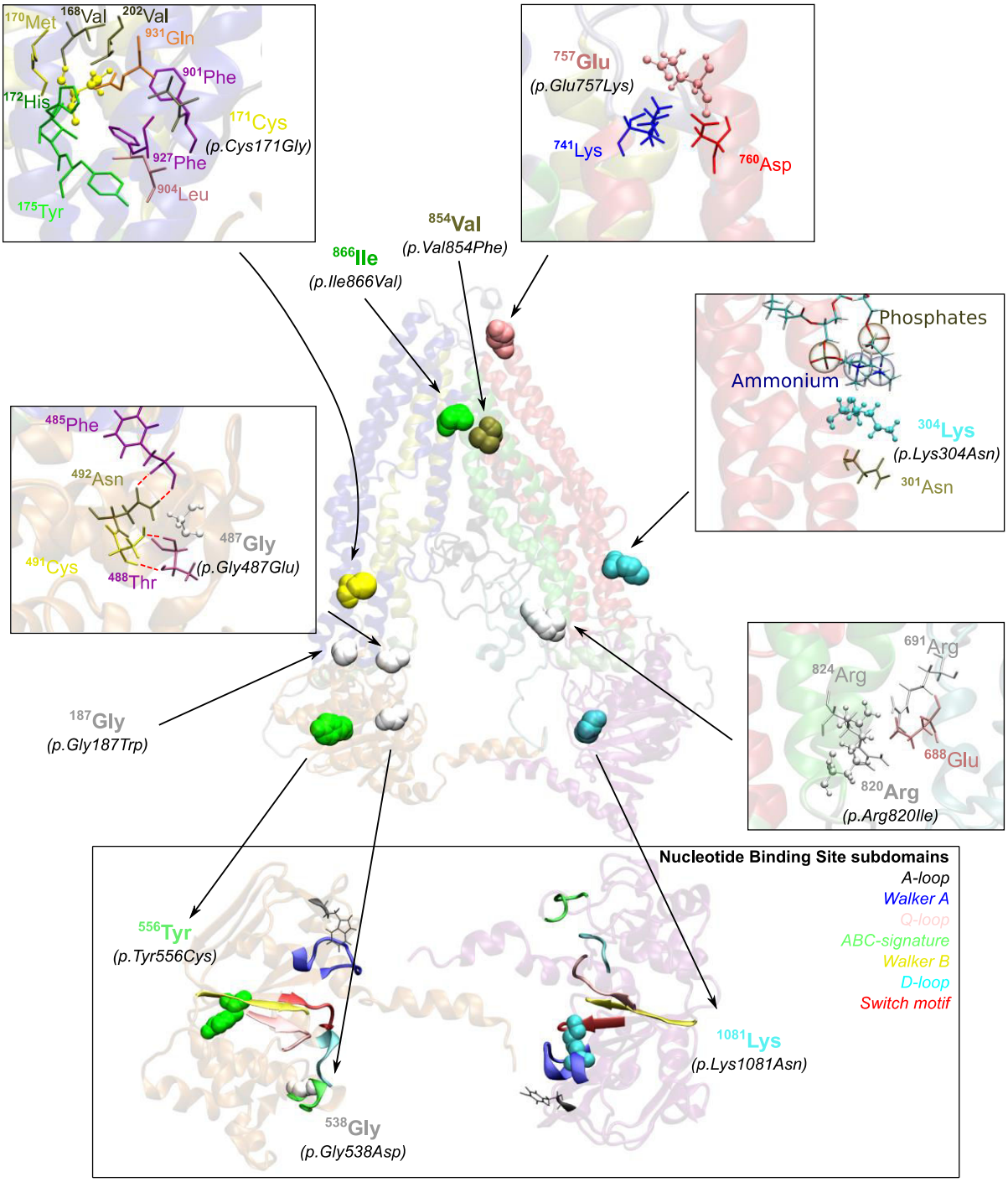



Fig. 5. Mapping of eleven mutations for which plasma membrane location of the corresponding variants have been confirmed. The different boxes highlight the local supramolecular organization and the possible inter-residue contacts of importance.

As expected, three mutations have been described in the ATP-binding sites, which strongly affect MRP4 transporter function. They are located in both NBD 1 (p.Gly538Asp and p.Tyr556Cys) and NBD 2 (p.Lys1081Asn). Both the p.Gly538Asp and p.Lys1081Asn variants were experimentally designed to be used as a negative control of MRP4 transport function [4,42].

- p.Gly538Asp is located in the ABC signature (Fig. 5), which normally bound the □-phosphate moiety of ATP [43]. The mutation to the anionic aspartate is likely to dramatically decrease ATP-binding by increasing repulsive Coulomb interactions.

- p.Lys1081Asn is located in the Walker A of NBD 2 (Fig. 5). This residue is expected to play a key role in ATP-binding together with the ABC signature of NBD 1, within an antiparallel arrangement of the two NBDs[7]. The cationic feature of Lys1081 is expected to maximize attractive Coulomb interactions with the negatively charged ATP and Walker A. Its mutation into the electronegative asparagine is likely to lower the affinity of ATP.

- p.Tyr556Cys is located in the Walker B (Fig. 5). The transport function of the corresponding variant is either unaltered or augmented however in a substrate-dependent manner[4,40,42]. Long-range coupling with the substrate binding pocket has been suggested to rationalize its role in MRP1[40,44,45]; a molecular understanding would require MD simulations of this variant made from the WT MRP4 model, upon substrate binding, to entirely sample the conformational reorganization.

The third class of mutations stems from the key role of the NBD-TMD interacting region in the ATP-hydrolysis energy transduction to the TMDs[30]. Most of these mutations decrease the transport function, possibly in a substrate-dependent way[40]. Two roles should be considered: either the close vicinity from the ATP-binding region or the substrate binding pocket are distorted by allosteric effects, or the energy transduction is affected; both options can occur simultaneously.

- p.Gly487Glu is located in NBD 1. This residue is in close contact with a H-bonding network involving Phe485, Asn492, Thr488 and Cys491 (Fig. 5), however not directly involved in the H-bond interaction. Its mutation to the polar anionic glutamate may affect the organization of this H-bonding network, directly located at the NBD-TMD interface where energy transduction is excepted.

- p.Arg820Ile is located in ICL 3. This residue is involved in a H-bond network with Glu688, Arg824 and Arg691 (Fig. 5). Its mutation may affect energy transduction from NBD 2 but also the interaction with $L_1$ domain, which is in close contact (Fig. S17).

- p.Cys171Gly is located in TMH 2. This residue is mostly surrounded by apolar and aromatic residue (Fig. 5); it exhibits no contact to the neighboring helices, except with Gly931 of TMH 10.

- p.Lys304Asn is located in TMH 5. This mutation was described to decrease the MRP4 function only for $E_2$17$\beta$G as a substrate. This residue can interact either with Asn301 or the polar head groups of the surrounding lipid bilayer. Its mutation to the neutral asparagine may affect protein-lipid interactions.

A particular attention is now paid to the p.Gly187Trp variant, for which MD simulations were performed after having mutated the residue.

*p.Gly187Trp polymorphism* - The Gly187 residue is in ICL1, in contact with NBD 1. Compared to the WT model, the p.Gly187Trp variant showed a dramatically modified structure (Fig. 6A). NBD 1 was turned and twisted as seen by the difference in the α-angle of *ca.* 10°, respectively (see Fig. 6C for angle definition and picture of differences in tertiary structures). This distortion made the two NBDs in closer contact with each other, with an inter-NBD distance of 43 instead of 62 Å (Fig. 6B). Moreover, there were many more native contacts for residue 187 in the mutant (Trp187) than in the WT model (Gly187), mainly within NBD 1 (Table S11). As tryptophan is significantly bigger than glycine, this dramatically increased steric hindrance (Fig. 6A). The conformational rearrangements led to stronger hydrogen bonding and salt bridge networks (mainly involving Glu461) between ICLs and NBD 1 (Table S12), which in turn decreased NBD flexibility, as shown by the overall NBD 1 B-factor decrease by *ca.* 115 $Å^2$ in p.Gly187Trp (Fig. S14). Significantly more powerful attractive interaction was observed between NBD and ICLs ($E_{int}$ = -200 as compared to -270 $kcal.mol^{-1}$, Table S7). Only the Coulomb contribution was increased, confirming the strengthening of hydrogen bond / salt bridge network in the mutant.

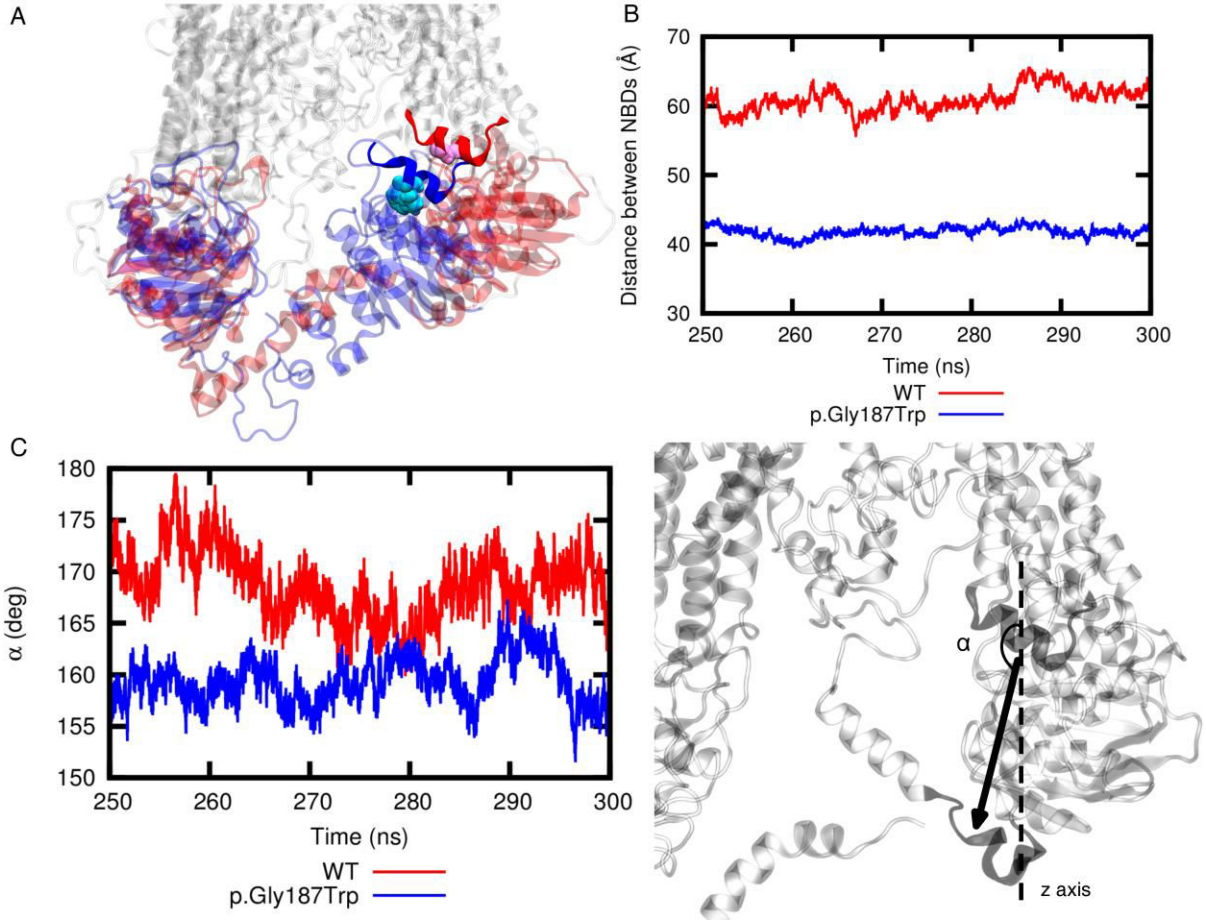



Fig. 6. Differences observed between p.Gly187Trp (blue) and WT MRP4 (red), as seen as: A) superimposition of NBDs; B) comparison of inter-NDB distance along the last 50 ns of MD simulation; C) description and evolution of α-angle along the last 50 ns of MD simulation.

Besides the conformational modifications in the direct surrounding of the p.Gly187Trp mutation, TMD packing was affected, in agreement with the communication role of the ICLs. The full structural description is given as Supplementary Data, but briefly bundle A conformation exhibited slight deformations, *i.e. ca.* 10° larger angles between i) TMH 1 and 2;

ii) TMH 1 and 10; and iii) TMH 1 and 12. This intra-bundle disorganization affected the interaction between bundles A and B, which was slightly increased (Table S13) making the IF-open to IF-occluded transformation more energy-demanding.

### IV.1.5. Conclusion

We propose a novel full molecular model of the human MRP4. The TMDs and NBDs structures are highly coherent with the literature. Attention has been paid to the structural features of the domains poorly defined so far by X-ray crystallography and cryo-EM, namely $L_1$, $L_0$ and the zipper helices. $L_1$ converged into a coherent conformation, located into the protein chamber. The role of $L_0$ and $L_1$ in the substrate binding chamber during drug transport should be carefully considered in drug efflux by MRP4. Each known polymorphism or rare mutation can be implemented in turn in this model, to elucidate its role at an atomic resolution. Simply by replacing the amino acid concerned and achieving MD relaxation in a reasonable timescale (tenths of ns), the conformational modifications can be elucidated. This should help to rationalize the impact of clinically observed polymorphisms and mutations on MRP4 exporter, as well as to predict the transport of new drugs (thus action and toxicity) related to a given genotype, giving rise to *in silico* predictive pharmacogenetics.

## References


[1] K.M. Hillgren, D. Keppler, A.A. Zur, K.M. Giacomini, B. Stieger, C.E. Cass, L. Zhang, International Transporter Consortium, Emerging transporters of clinical importance: an update from the International Transporter Consortium, Clin. Pharmacol. Ther. 94 (2013) 52–63. doi:10.1038/clpt.2013.74.

[2] International Transporter Consortium, K.M. Giacomini, S.-M. Huang, D.J. Tweedie, L.Z. Benet, K.L.R. Brouwer, X. Chu, A. Dahlin, R. Evers, V. Fischer, K.M. Hillgren, K.A. Hoffmaster, T. Ishikawa, D. Keppler, R.B. Kim, C.A. Lee, M. Niemi, J.W. Polli, Y. Sugiyama, P.W. Swaan, J.A. Ware, S.H. Wright, S.W. Yee, M.J. Zamek-Gliszczynski, L. Zhang, Membrane transporters in drug development, Nat. Rev. Drug Discov. 9 (2010) 215–236. doi:10.1038/nrd3028.

[3] K.M. Morrissey, C.C. Wen, S.J. Johns, L. Zhang, S.-M. Huang, K.M. Giacomini, The UCSF-FDA TransPortal: A Public Drug Transporter Database, Clin. Pharmacol. Ther. 92 (2012) 545–546. doi:10.1038/clpt.2012.44.

[4] N. Abla, L.W. Chinn, T. Nakamura, L. Liu, C.C. Huang, S.J. Johns, M. Kawamoto, D. Stryke, T.R. Taylor, T.E. Ferrin, K.M. Giacomini, D.L. Kroetz, The Human Multidrug Resistance Protein 4 (MRP4, ABCC4): Functional Analysis of a Highly Polymorphic Gene, J. Pharmacol. Exp. Ther. 325 (2008) 859–868. doi:10.1124/jpet.108.136523.

[5] P.-A. Billat, T. Ossman, F. Saint-Marcoux, M. Essig, J.-P. Rerolle, N. Kamar, L. Rostaing, H. Kaminski, G. Fabre, M. Otyepka, J.-B. Woillard, P. Marquet, P. Trouillas, N. Picard, Multidrug resistance-associated protein 4 (MRP4) controls ganciclovir intracellular accumulation and contributes to ganciclovir-induced neutropenia in renal transplant patients, Pharmacol. Res. 111 (2016) 501–508. doi:10.1016/j.phrs.2016.07.012.

[6] J. ter Beek, A. Guskov, D.J. Slotboom, Structural diversity of ABC transporters, J. Gen. Physiol. 143 (2014) 419–435. doi:10.1085/jgp.201411164.

[7] S. Wilkens, Structure and mechanism of ABC transporters, F1000Prime Rep 2015. 7 (2015). doi:10.12703/P7-14.

[8] A.W. Ravna, I. Sylte, G. Sager, Binding site of ABC transporter homology models confirmed by ABCB1 crystal structure, Theor. Biol. Med. Model. 6 (2009) 20. doi:10.1186/1742-4682-6-20.

[9] A.A.K. El-Sheikh, J.J.M.W. van den Heuvel, E. Krieger, F.G.M. Russel, J.B. Koenderink, Functional Role of Arginine 375 in Transmembrane Helix 6 of Multidrug Resistance Protein 4 (MRP4/ABCC4), Mol. Pharmacol. 74 (2008) 964–971. doi:10.1124/mol.107.043661.

[10] A.W. Ravna, G. Sager, Molecular model of the outward facing state of the human multidrug resistance protein 4 (MRP4/ABCC4), Bioorg. Med. Chem. Lett. 18 (2008) 3481–3483. doi:10.1016/j.bmcl.2008.05.047.

[11] F.G.M. Russel, J.B. Koenderink, R. Masereeuw, Multidrug resistance protein 4 (MRP4/ABCC4): a versatile efflux transporter for drugs and signalling molecules, Trends Pharmacol. Sci. 29 (2008) 200–207. doi:10.1016/j.tips.2008.01.006.

[12] H.G.M. Wittgen, J.J.M.W. van den Heuvel, E. Krieger, G. Schaftenaar, F.G.M. Russel, J.B. Koenderink, Phenylalanine 368 of multidrug resistance-associated protein 4

(MRP4/ABCC4) plays a crucial role in substrate-specific transport activity, Biochem. Pharmacol. 84 (2012) 366–373. doi:10.1016/j.bcp.2012.04.012.

[13] Y. Chen, H. Jin, L. Zhang, Z. Liu, Molecular models of different states of the human multidrug resistance protein 4 (MRP4/ABCC4), J. Chin. Pharm. Sci. 25 (2016) 428–437. doi:10.5246/jcps.2016.06.048.

[14] The UniProt Consortium, UniProt: a hub for protein information, Nucleic Acids Res. 43 (2015) D204–D212. doi:10.1093/nar/gku989.

[15] Y. Zhang, I-TASSER server for protein 3D structure prediction, BMC Bioinformatics. 9 (2008) 40. doi:10.1186/1471-2105-9-40.

[16] M.S. Jin, M.L. Oldham, Q. Zhang, J. Chen, Crystal structure of the multidrug transporter P-glycoprotein from C. elegans, Nature. 490 (2012) 566–569. doi:10.1038/nature11448.

[17] S.G. Aller, J. Yu, A. Ward, Y. Weng, S. Chittaboina, R. Zhuo, P.M. Harrell, Y.T. Trinh, Q. Zhang, I.L. Urbatsch, G. Chang, Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding, Science. 323 (2009) 1718–1722. doi:10.1126/science.1168750.

[18] J. Li, K.F. Jaimes, S.G. Aller, Refined structures of mouse P-glycoprotein, Protein Sci. Publ. Protein Soc. 23 (2014) 34–46. doi:10.1002/pro.2387.

[19] H. Hayashi, S. Naoi, T. Nakagawa, T. Nishikawa, H. Fukuda, S. Imajoh-Ohmi, A. Kondo, K. Kubo, T. Yabuki, A. Hattori, M. Hirouchi, Y. Sugiyama, Sorting nexin 27 interacts with multidrug resistance-associated protein 4 (MRP4) and mediates internalization of MRP4, J. Biol. Chem. 287 (2012) 15054–15065. doi:10.1074/jbc.M111.337931.

[20] E.L. Wu, X. Cheng, S. Jo, H. Rui, K.C. Song, E.M. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R.M. Venable, J.B. Klauda, W. Im, CHARMM-GUI Membrane Builder toward realistic biological membrane simulations, J. Comput. Chem. 35 (2014) 1997–2004. doi:10.1002/jcc.23702.

[21] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935. doi:10.1063/1.445869.

[22] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, J. Chem. Theory Comput. 11 (2015) 3696–3713. doi:10.1021/acs.jctc.5b00255.

[23] C.J. Dickson, B.D. Madej, Å.A. Skjevik, R.M. Betz, K. Teigen, I.R. Gould, R.C. Walker, Lipid14: The Amber Lipid Force Field, J. Chem. Theory Comput. 10 (2014) 865–879. doi:10.1021/ct4010307.

[24] D.A. Case, D.S. Cerutti, T.E. Cheatham, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, D.M. York, P.A. Kollman, (2017), AMBER 2017, University of California, San Francisco. http://ambermd.org.

[25] R. Salomon-Ferrer, A.W. Götz, D. Poole, S. Le Grand, R.C. Walker, Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald, J. Chem. Theory Comput. 9 (2013) 3878–3888. doi:10.1021/ct400314y.

[26] R.J. Loncharich, B.R. Brooks, R.W. Pastor, Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide, Biopolymers. 32 (1992) 523–535. doi:10.1002/bip.360320508.

[27] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690. doi:10.1063/1.448118.

[28] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092. doi:10.1063/1.464397.

[29] M. Moradi, E. Tajkhorshid, Mechanistic picture for conformational transition of a membrane transporter at atomic resolution, Proc. Natl. Acad. Sci. 110 (2013) 18916–18921. doi:10.1073/pnas.1313202110.

[30] B. Verhalen, R. Dastvan, S. Thangapandian, Y. Peskova, H.A. Koteiche, R.K. Nakamoto, E. Tajkhorshid, H.S. Mchaourab, Energy transduction and alternating access of the mammalian ABC transporter P-glycoprotein, Nature. 543 (2017) 738–741. doi:10.1038/nature21414.

[31] K. Hollenstein, D.C. Frei, K.P. Locher, Structure of an ABC transporter in complex with its binding protein, Nature. 446 (2007) 213–216. doi:10.1038/nature05626.

[32] Z. Zhang, J. Chen, Atomic Structure of the Cystic Fibrosis Transmembrane Conductance Regulator, Cell. 167 (2016) 1586–1597.e9. doi:10.1016/j.cell.2016.11.014.

[33] A. Nöll, C. Thomas, V. Herbring, T. Zollmann, K. Barth, A.R. Mehdipour, T.M. Tomasiak, S. Brüchert, B. Joseph, R. Abele, V. Oliéric, M. Wang, K. Diederichs, G. Hummer, R.M. Stroud, K.M. Pos, R. Tampé, Crystal structure and mechanistic basis of a functional homolog of the antigen transporter TAP, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) E438–E447. doi:10.1073/pnas.1620009114.

[34] G.E. Tusnády, B. Sarkadi, I. Simon, A. Váradi, Membrane topology of human ABC proteins, FEBS Lett. 580 (2006) 1017–1022. doi:10.1016/j.febslet.2005.11.040.

[35] E. Biemans-Oldehinkel, M.K. Doeven, B. Poolman, ABC transporter architecture and regulatory roles of accessory domains, FEBS Lett. 580 (2006) 1023–1035. doi:10.1016/j.febslet.2005.11.079.

[36] A.P. Naren, M.W. Quick, J.F. Collawn, D.J. Nelson, K.L. Kirk, Syntaxin 1A inhibits CFTR chloride channels by means of domain-specific protein-protein interactions, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 10972–10977.

[37] A.P. Naren, E. Cormet-Boyaka, J. Fu, M. Villain, J.E. Blalock, M.W. Quick, K.L. Kirk, CFTR Chloride Channel Regulation by an Interdomain Interaction, Science. 286 (1999) 544–548. doi:10.1126/science.286.5439.544.

[38] Z.L. Johnson, J. Chen, Structural Basis of Substrate Recognition by the Multidrug Resistance Protein MRP1, Cell. 168 (2017) 1075–1085.e9. doi:10.1016/j.cell.2017.01.041.

[39] H. Jerabek, G. Pabst, M. Rappolt, T. Stockner, Membrane-mediated effect on ion channels induced by the anesthetic drug ketamine, J. Am. Chem. Soc. 132 (2010) 7990–7997. doi:10.1021/ja910843d.

[40] M. Banerjee, V. Marensi, G. Conseil, X.C. Le, S.P.C. Cole, E.M. Leslie, Polymorphic variants of MRP4/ABCC4 differentially modulate the transport of methylated arsenic metabolites and physiological organic anions, Biochem. Pharmacol. 120 (2016) 72–82. doi:10.1016/j.bcp.2016.09.016.

[41] P. Krishnamurthy, M. Schwab, K. Takenaka, D. Nachagari, J. Morgan, M. Leslie, W. Du, K. Boyd, M. Cheok, H. Nakauchi, C. Marzolini, R.B. Kim, B. Poonkuzhali, E. Schuetz, W. Evans, M. Relling, J.D. Schuetz, Transporter-mediated protection against thiopurine-induced hematopoietic toxicity, Cancer Res. 68 (2008) 4983–4989. doi:10.1158/0008-5472.CAN-07-6790.

[42] D. Janke, S. Mehralivand, D. Strand, U. Gödtel-Armbrust, A. Habermeier, U. Gradhand, C. Fischer, M.R. Toliat, P. Fritz, U.M. Zanger, M. Schwab, M.F. Fromm, P. Nürnberg, L. Wojnowski, E.I. Closs, T. Lang, 6-mercaptopurine and 9-(2-phosphonyl-methoxyethyl) adenine (PMEA) transport altered by two missense mutations in the drug transporter gene ABCC4, Hum. Mutat. 29 (2008) 659–669. doi:10.1002/humu.20694.

[43] H. Nikaido, How are the ABC transporters energized?, Proc. Natl. Acad. Sci. 99 (2002) 9609–9610. doi:10.1073/pnas.162375699.

[44] K. Koike, G. Conseil, E.M. Leslie, R.G. Deeley, S.P.C. Cole, Identification of proline residues in the core cytoplasmic and transmembrane regions of multidrug resistance protein 1 (MRP1/ABCC1) important for transport function, substrate specificity, and nucleotide interactions, J. Biol. Chem. 279 (2004) 12325–12336. doi:10.1074/jbc.M311435200.

[45] I.J. Létourneau, A. Nakajima, R.G. Deeley, S.P.C. Cole, Role of proline 1150 in functional interactions between the membrane spanning domains and nucleotide binding domains of the MRP1 (ABCC1) transporter, Biochem. Pharmacol. 75 (2008) 1659–1669. doi:10.1016/j.bcp.2008.01.009.

## IV.2. Exploration of the human MRP4 protein chamber by thorough molecular docking calculations

*Note to the reader: This section is written and edited as a publication; which however has not been submitted yet.*

### IV.2.1. Introduction

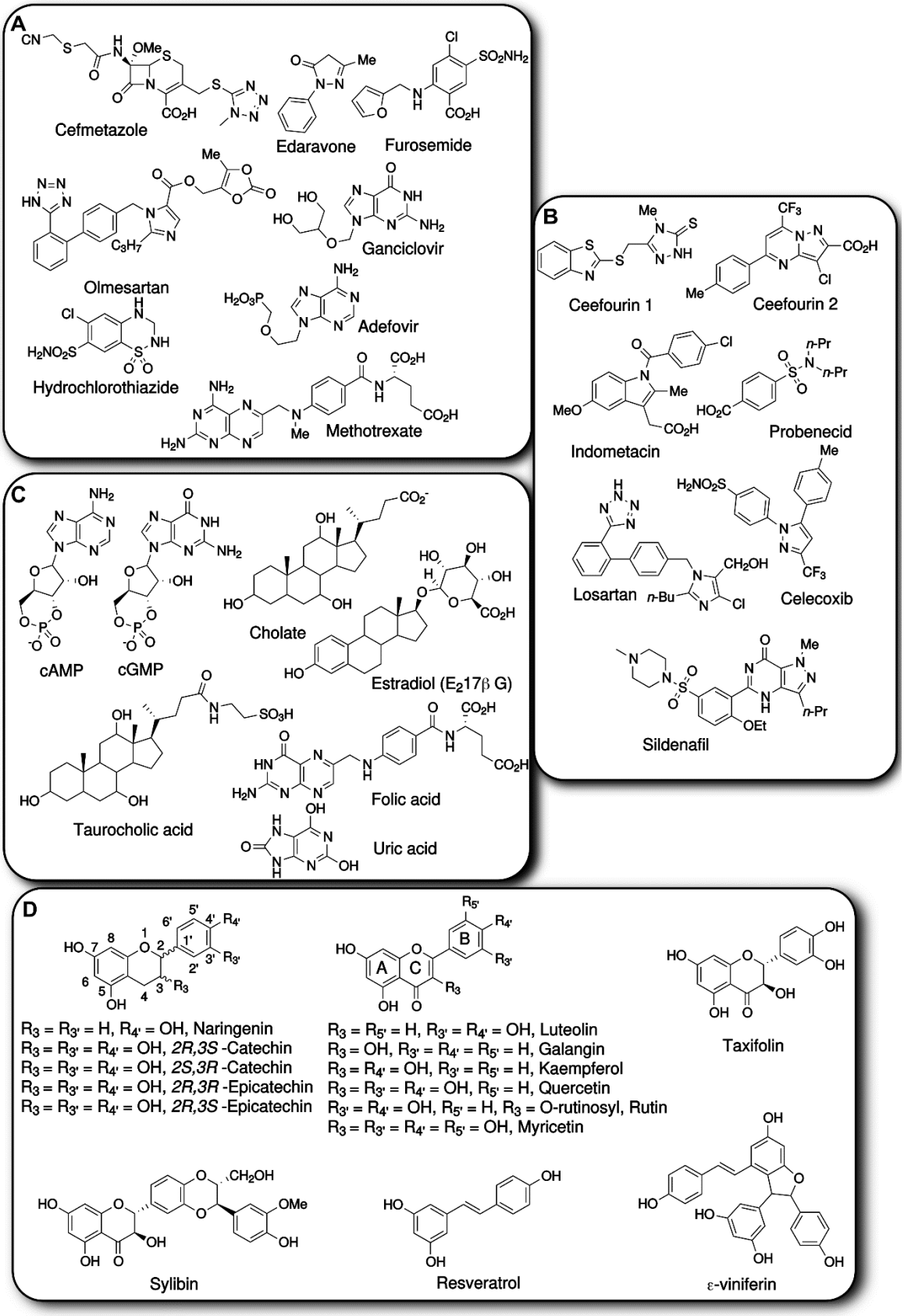
The human ABCC4/MRP4 (multidrug resistance-associated protein 4) exporter belongs to the ATP-binding cassette (ABC) transporter superfamily[1], which is involved in the active transport of a broad range of substrates, namely endogenous compounds (*e.g.*, cAMP/GMP, leukotrienes)[2] or xenobiotics (*e.g.*, antiviral, antibiotics and nutrients including polyphenols). Drug transport through MRP4 plays a key role in drug pharmacokinetics (PK) and pharmacodynamics (PD). Regarding the broad range of chemical structures, one might expect various roles of this exporter. For instance, its expression at the brain capillary endothelial cells suggests that it protects the brain from certain xenobiotics, whereas its key location at the basolateral membranes of hepatocytes and apical membranes of kidney proximal tubule cells are attributed to a drug detoxification role[3,4]. Even though no disease has been associated to alter MRP4 activity[5], the transport of endogenous substrate by this exporter has been linked to physiological events such as dendritic cell migration in T cell-mediated response[6] as well as the liver detoxification of patients with cholestasis[3,7].

MRP4 was listed by the international transporter consortium (ITC) as of “emerging clinical importance in drug absorption and disposition[3,4]. To study the relationship between MRP4 and PK/PD drugs is of utmost relevance, with special attention paid to both intracellular drug bioavailability[8] and PK drug-drug interactions (PK DDIs). Dysfunctions of ABC exporters, including MRP4, has been be linked to intracellular accumulation of drugs and/or endogenous substrates that in turn may provoke harmful effects[9–11]. A growing attention has been paid to the relationship between PK DDIs and ABC/MRPs; indeed, drug transport activity can be modulated by other substrates (inhibitors or enhancers)[12]. A better understanding of this relationship requires a thorough characterization of the MRP4 protein chamber and channel, where the substrates dynamically bind.

MRP4 is a highly polymorphic protein[5] for which the canonical form is made of 1325 amino acids. It adopts the generic structure of mammalian monomeric ABC transporters made of twelve transmembrane helices (TMHs) 1-12, two nucleotide binding domains (NBDs) 1 and 2, a N-terminal Lasso motif ($L_0$) and a linker $L_1$ between NBD1 and TMH7. So far, the structure of MRP4 has not been resolved neither by X-ray nor by cryo-EM. In agreement with recent findings on the transport cycle of ABC transporters[13,14], we recently constructed a three-dimensional model of MRP4 from a protein threading procedure and post-MD (molecular dynamics) treatment to allow sufficient relaxation of the secondary structure. Moreover, to enhance realistic environment description, the protein was embedded in a lipid bilayer. The model was in its inward-facing conformer. The TMHs are arranged in four bundles, named A (TMHs 1, 2, 10 and 11), B (TMHs 4, 5, 7 and 8), C (TMHs 3 and 6) and D (TMHs 9 and 12)[15]. Even though there is a lack of structural insights about the MRP4 substrate binding pocket, recent experimental studies pointed out the importance of a series of residues, namely Phe368, Phe369, Glu374, Arg375, Glu378, Trp995 and Arg998[16,17]. A docking study linked to experimental data suggested Glu103, Ser328, Gly359, Arg362, Val726 and Leu987 to be involved in substrate binding[18], however, this was made on a homology model of MRP4 in its outward-facing conformer. The thorough identification of residues constituting the protein

chamber of other ABC transporters (*e.g.*, ABCB1/P-gp - P-glycoprotein - or ABCC1/MRP1) enables documenting about the driving forces responsible for substrate binding. For instance, MRPs were suggested to bind substrates mainly through hydrogen bonding and $\pi$–$\pi$ stacking interactions and have multiple allosteric binding sites in the protein chamber [2,19]. Another issue that is partially related to the structural feature of the protein chamber, but also to its direct environment (the lipid bilayer), is the xenobiotic membrane access. This key event is likely transporter-dependent since *e.g.*, in P-gp, xenobiotics can enter the chamber directly from the lipid bilayer while this access is completely clogged in MRP1. Further investigations are required for MRP4 since there is a limited knowledge about the substrate translocation pathway, as there are only a few studies which have dealt with. These studies were based on docking procedures performed in a static way, *i.e.,* not accounting for the highly dynamic character of the translocation path[16–18,20].

By extending our previous work about the structural pattern of MRP4, we here propose an *in silico* investigation of protein-ligand interactions using a broad series of ligands (Scheme 1) by means of molecular docking performed on the lipid-bilayer embedded relaxed MRP4 models. It is important to stress that the $L_0$-domain of MRP4 remains relatively flexible as it is not link to the transmembrane $TMD_0$ domain contrary to other MRPs. Therefore, this study is made on two possible MRP4 conformers, named $L_0$-in or $L_0$-out, corresponding to $L_0$ in contact with the protein or $L_0$ embedded in the lipid bilayer, respectively[15]. As consensual conclusions could not be reached about the $L_0$ location, both conformers should be considered as it might strongly impact on the structural feature of the protein chamber and in turn on binding. A set of 41 ligands was chosen including known substrates that are either drugs (Scheme 1A), endogenous substrates (Scheme 1B) as well as known inhibitors (Scheme 1C), several natural polyphenols were also included regarding their potential interaction with MRP4 (Scheme 1D) [ 21–24]. Extra polyphenols were also selected, as this family is a textbook case allowing investigation of structure-property relationship mainly on $\pi$–$\pi$ stacking and H-bonding capacity. It is worth mentioning that a few substrates with sugar moieties were also included (namely, edaravone glucuronide, 17-$\beta$-estradiol glucuronide $E_2$17$\beta$G and rutin) as MRP4 is known to carry glucuronide drug metabolites.




Scheme 1. Chemical structures of the studied A) drugs, B) inhibitors, C) endogenous substrates and D) natural polyphenols. The edaravone glucuronide, the quercetin-3-*O*'-glucuronide and quercetin-3-*O*'-sulfate are not included for the sake of clarity.

### IV.2.2. Methods

*MRP4 structures*

The human MRP4 structures used for molecular docking calculations were obtained from the MD-refined protein-threading models embedded in a (3:1) POPC:Cholesterol lipid bilayer membrane[15]. The two $L_0$-in and $L_0$-out conformers were considered (see [15] for more details). For both $L_0$-in/out conformations, five different snapshots were extracted from the last 100 ns of previous MD simulations performed at 310 K to account for local conformational flexibility of amino acids side chains in the protein chamber. It is important to note that the whole protein was considered as static for molecular docking calculations except for the sidechains of Phe368, Phe369, Glu374, Arg375, Glu378, Trp995 and Arg998 that were experimentally shown to be involved in substrate bindings[16,17].

*Ligands*

The 41 different ligands were therefore divided into four categories (Scheme 1 and Table S1), namely drugs, endogenous substrates, inhibitors and natural polyphenols. All ligand structures were built and minimized at the molecular mechanic level (MMFF94s forcefield) using the Avogadro software[22]. For most of ligands, the non-aromatic dihedral angles were considered flexible for molecular docking calculations except those known to exhibit large rotational energy barrier (*e.g.*, H-bonding between 5-OH and keto group of flavonols, see Fig. S1 for a list of rigid dihedral angles).

*Molecular Docking*

Molecular docking calculations were carried out using the Autodock Vina software[23]. The PDBQT ligand and MRP4 files were generated with AutoDockTools[24]. First, “blind”-docking calculations were performed with all ligands in which the whole cavity of the protein chamber was considered as the volume search space (*i.e.*, with x-, y-, z-dimensions of 40.0 Å x 65.0 Å x 55.0 Å or 143 000 $Å^3$, see Table S2). All molecular poses obtained from this blind docking exhibited clusters located at different spatial regions. A refined molecular docking procedure was then performed in smaller space search volumes of each identified cluster. It is important to note that the maximum search volume used was *ca*. 27000 $Å^3$ (*i.e.*, 29.0 Å x 29.0 Å x 29.0 Å) to ensure reliability of results (Table S2).

The molecular docking calculations were performed on twenty replicas for each ligand-MRP4 system to ensure correct sampling as suggested earlier[25]. Owing to the large space search volume, the “exhaustiveness” parameter was set at 40 to increase the computational effort used during molecular pose search[26] (*i.e.,* thoroughness of search which is roughly proportional to time). The maximum number of selected binding poses was limited at 100 per replica and the calculation affinity cut-off was set at 15 kcal/mol with respect to the top pose of a given replica.

*Structural analysis*

Molecular docking calculations provided roughly more than 70 000 different molecular poses by ligand-protein pair (*i.e.*, more than 140 000 for both $L_0$-in and $L_0$-out conformers), which correspond to roughly 20 poses docking calculation on 20 replicas, in 4 regions and in the 5 MRP4 snapshots. Molecular docking results and poses were thoroughly analyzed using a homemade python script available on request. For a given ligand-MRP4 system, all molecular poses obtained with the five snapshots were ranked according to their affinities. Only the poses

lying within a cut-off of 1 kcal/mol, with respect to the top-ranked pose of the twenty replicas, were selected. We assume in our calculations a chemical accuracy of 1 kcal/mol, the standard error of Autodock Vina being between 2.75 and 2.85 kcal/mol[23]. Protein Data Bank format (PDB) trajectories gathering selected molecular poses and MRP4 coordinates were automatically generated including flexible residues but neglecting non-polar H-atoms. The generation of these PDB trajectories enabled the use of the CPPTRAJ software[27] for a systematic investigation of H-bond interactions (using a cut-off XH--Y distance and angle set up at 3.5 Å and 120°), of the protein-ligand contact lying within a 4.5 Å distance, as well as of the clustering of molecular poses. Distributions of H-bonding were performed by summing the atomic H-bond fractions for a given residue over all selected molecular poses. A systematic investigation of $\pi$–$\pi$ stacking interactions was also carried out; namely, the interactions between two aromatic residues was considered as $\pi$-stacked when (i) the distance between the two rings' centers-of-mass was below 5.0 Å and (ii) angle between the two normal vectors of the aromatic rings ('corrplane' keyword in CPPTRAJ) was below 10°. These analyses are believed to provide statistically relevant insights about plausible molecular poses, accounting for the flexibility of protein-ligand interaction regions in MRP4.

### IV.2.3. Results and Discussion

#### IV.2.3.1. Identification of drug-binding pockets

*Three regions to bind them all*

Since the MRP4 protein is expected to have different substrate binding regions[28], the blind docking was first performed in a volume centered in the middle of the protein chamber (Fig. S2). This step highlighted three binding regions, named regions 1, 2 and 3 (see Fig. 1). Region 1 is mostly constituted of solvent-accessible parts of TMHs 3, 6, 10, 11 (bundles A and C) and to a lower extend TMH 12 (bundle D). Despite a significant overlap with region 1, region 2 is defined as a distinct binding site since it is located more buried in the protein chamber, in contact with the upper part of region 1. It is therefore less solvent-accessible, mainly in contact with TMHs 1, 3, 6, 12. Region 3 is an allosteric binding pocket defined between bundles B and D and located in the vicinity of the polar head group surface. It is worth mentioning that the $L_0$-motif is in close contact with region 1 in the case of the MRP4 $L_0$-in conformer.

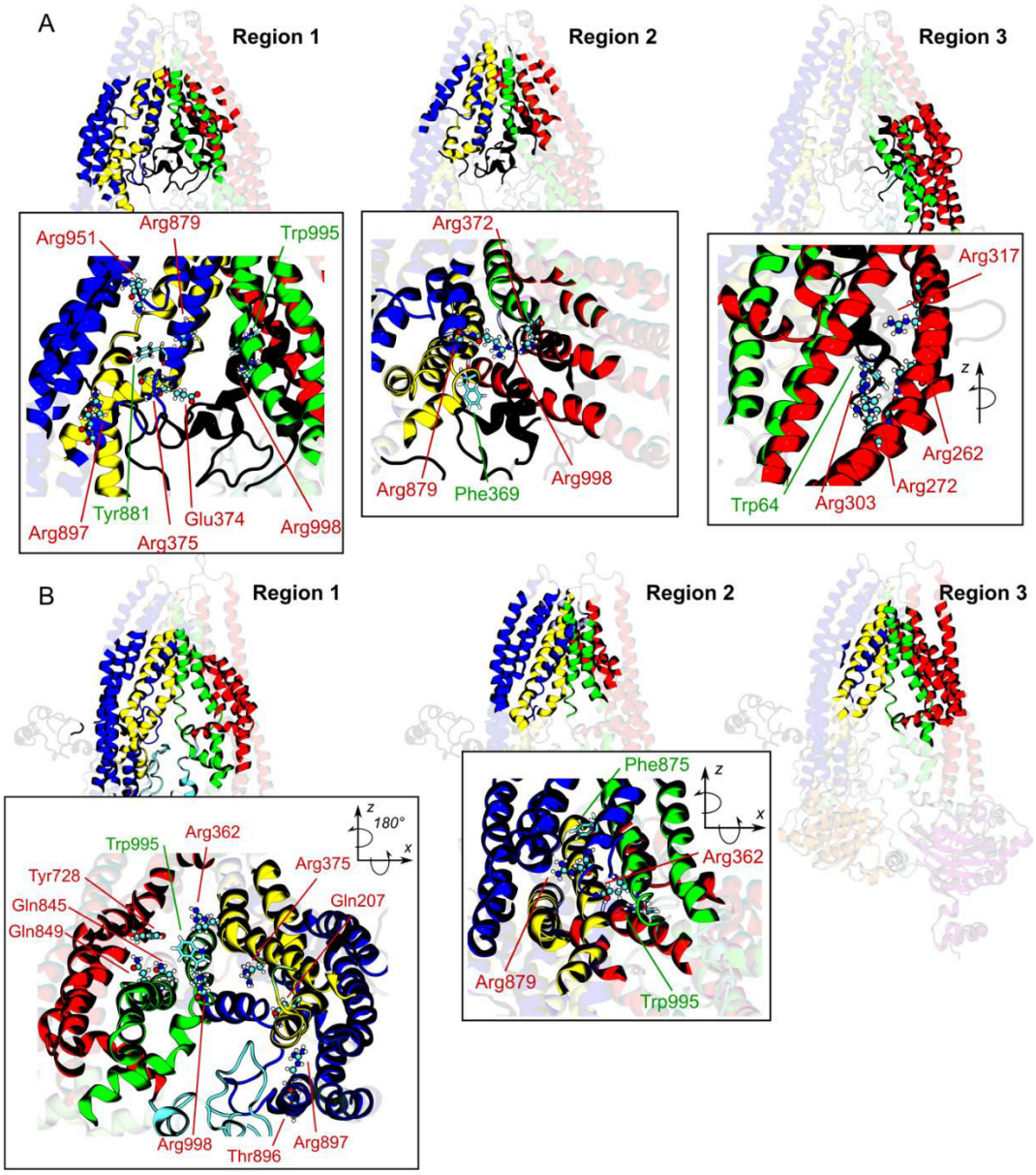



Fig. 1. Binding regions and residues of importance for the modes of binding modes to MRP4 in A) $L_0$-in and B) $L_0$-out conformations. Protein secondary structures are colored according to the topology (see Fig. S3 for a detailed definition). Residues of importance involved in H-bonding are depicted in ball and sticks and the corresponding labels are in red. Residues of importance engaged in $\pi$-$\pi$ stacking interactions are in licorice and labels are in green. Residues of importance of region 3 in MRP4 $L_0$-out are not showed owing to a too high overlap with region 1.

A refined docking was performed by splitting region 1 into smaller search volumes to match the recommended maximum volume (*i.e.*, 27000 Å$^3$). Namely, four/five search zones were defined in MRP4 $L_0$-in/out (Fig. S4 and Table S2). Given the high structural diversity of known substrates, a multi-mode of binding is expected, a thorough analysis of non-covalent interactions was performed in a very exhaustive way, *i.e.*, all poses ranging within 1 kcal/mol (see section 2). A set of 107839 poses were thus analysed for both $L_0$-in and $L_0$-out conformations. This analysis (counting H-bonds as relevant when present in a fraction greater than 10%) highlighted a series of amino acids involved in the binding modes.

*Key residues to ligand binding modes*

*Region 1.* Region 1 is defined as the canonical binding region, crucial for substrate binding in the first step of in the translocation process. Regardless to the MRP4 conformation, Arg375, Arg998, Thr896 and Arg897 have been highlighted to play a key role in substrate-transporter, mainly involved in the formation of H-bond networks. For instance, the average H-bond fractions over all substrate poses in region 1 are 0.69 (0.50), 0.37 (1.60), 0.26 (0.19) and 0.19 (0.15), for Arg375, Arg998, Thr896 and Arg897, respectively in the $L_0$-in ($L_0$-out) conformers (Table 1). The fact that (cationic) arginine residues are key in substrate binding is in the line with experimental evidences suggesting that MRP4 preferentially efflux anionic[16]. This is particularly true for the $L_0$-in conformer, for which five of the key residues are arginine, namely Arg375, Arg879, Arg897, Arg951 and Arg998 (Table 1). It is worth mentioning that Arg897 and Arg951 are nearer to the border of region 1 with respect to other key residues, which may suggest an indirect role. For instance, Arg951 could play a key role for substrate access from the lipid bilayer.

Table 1. Average H-bond fractions in A) $L_0$-in and B) $L_0$-out conformations. The fractions are presented for key residues, for all the dataset, and for each defined group (i) purine-like, (ii) sugars, (iii) steroids and (iv) flavonoids.

(A)

| Region | Residues | Dataset | Purine-like | Sugars | Steroids | Flavonoids |
|---|---|---|---|---|---|---|
| Region 1 | Arg375 | 0.69 | 0.65 | 1.10 | 0.99 | 0.58 |
| | Arg879 | 0.38 | 0.47 | - | 0.31 | 0.37 |
| | Arg998 | 0.37 | 0.22 | 0.59 | 0.16 | 0.57 |
| | Glu374 | 0.31 | 0.33 | 0.55 | 0.66 | 0.22 |
| | Arg951 | 0.29 | 0.23 | 0.15 | 0.38 | 0.28 |
| | Thr896 | 0.26 | 0.31 | 0.44 | 0.23 | 0.24 |
| | Arg897 | 0.19 | 0.20 | 0.48 | 0.15 | 0.21 |
| | Arg887 | - | - | - | 0.19 | - |
| | Ile373 | - | - | - | 0.16 | - |
| | Arg21 | - | - | - | 0.15 | 0.17 |
| | Ser894 | - | - | 0.26 | - | - |
| | Gln54 | - | 0.18 | 0.24 | - | - |
| | Trp995 | - | 0.19 | 0.20 | - | - |
| | Gln994 | - | - | - | - | 0.18 |
| Region 2 | Arg362 | 0.19 | - | 0.49 | 0.40 | 0.20 |
| | Arg998 | 0.19 | 0.16 | 0.39 | 0.16 | 0.28 |
| | Arg879 | 0.16 | - | - | - | 0.22 |
| | Glu374 | - | - | 0.16 | - | - |
| | Trp995 | - | 0.18 | - | - | - |
| Region 3 | Arg317 | 0.77 | 0.80 | 1.18 | 0.37 | 0.88 |
| | Arg303 | 0.39 | 0.35 | 0.63 | 1.13 | 0.29 |
| | Arg272 | 0.20 | 0.17 | 0.28 | 0.17 | 0.16 |
| | Arg262 | - | - | - | 0.28 | - |

(B)

| Region | Residues | Dataset | Purine-like | Sugars | Steroids | Flavonoids |
|---|---|---|---|---|---|---|
| Region1 | Arg998 | 1.60 | 1.43 | 2.50 | 2.01 | 1.76 |
| | Arg375 | 0.50 | 0.26 | 0.70 | 0.43 | 0.77 |
| | Gln845 | 0.39 | 0.54 | 0.57 | 0.43 | 0.30 |
| | Tyr728 | 0.33 | 0.34 | 0.51 | 0.22 | 0.39 |
| | Gln849 | 0.33 | 0.38 | 0.57 | 0.15 | 0.41 |
| | Thr896 | 0.19 | - | - | - | 0.17 |
| | Arg362 | 0.18 | 0.32 | 0.23 | 0.19 | - |
| | Gln207 | 0.18 | 0.22 | 0.27 | - | 0.19 |
| | Arg897 | 0.15 | 0.19 | 0.16 | - | 0.20 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Arg880 | - | - | 0.16 | 0.56 | - |
| | Phe882 | - | - | 0.21 | 0.54 | - |
| | Leu883 | - | - | 0.17 | 0.51 | - |
| | Arg317 | - | - | - | 0.26 | - |
| | Arg891 | - | - | 0.16 | 0.24 | - |
| | Ser886 | - | - | - | 0.19 | - |
| | Arg782 | - | - | - | 0.17 | - |
| | Ser668 | - | 0.25 | 0.28 | - | - |
| | Lys890 | - | - | 0.27 | - | - |
| | Lys670 | - | - | 0.25 | - | - |
| | Ser894 | - | 0.16 | 0.15 | - | - |
| | Ala673 | - | - | - | - | 0.24 |
| | Gly672 | - | - | - | - | 0.23 |
| Region2 | Arg879 | 0.41 | 0.50 | 0.46 | - | 0.30 |
| | Arg362 | 0.17 | 0.30 | 0.25 | - | 0.16 |
| | Gln221 | - | - | - | 0.18 | - |
| | Tyr155 | - | - | - | 0.17 | - |
| | Lys106 | - | 0.17 | - | - | - |
| Region3 | Arg998 | 0.30 | 0.31 | 0.50 | 0.21 | 0.32 |
| | Tyr728 | 0.17 | 0.17 | 0.27 | 0.15 | 0.20 |
| | Arg375 | 0.15 | - | 0.22 | 0.16 | 0.18 |
| | Arg362 | - | 0.20 | 0.15 | 0.19 | - |
| | Gln845 | - | 0.18 | 0.19 | 0.16 | - |
| | Gln849 | - | - | 0.20 | - | 0.19 |

Many poses in region 1 adopt $\pi$-$\pi$ stacked conformations, mainly with two aromatic residues, namely Trp995 and Tyr881 (the latter being highlighted in the $L_0$-in conformation). The key role of Trp995 was suggested from experimental data[17], underlining the robustness of the MRP4 molecular models and the docking procedure and analysis.

*Region 2.* The docking in region 2 highlighted less key residues than in region 1, namely Arg362 and Arg879, with averaged H-bond fractions of 0.19 (0.41) and 0.19 (0.17), respectively in $L_0$-in ($L_0$-out) (Table 1). The presence of Arg879 in both regions 1 and 2 may be explained by the spatial overlap between these two regions. It suggests a dynamic contributions of Arg362 to ease the translocation from region 1 to region 2, in agreement with previous studies[20]. Trp995 was highlighted as a key residue of region 2 capable of forming $\pi$-$\pi$ stacking interactions with various substrates. In the $L_0$-in conformation only, Phe369 was also identified as capable of such interactions, in agreement with experimental evidences [16,17]. Regarding its location in the upper part of MRP4, region 2 is expected to correspond to the second step of substrate binding, *i.e.*, when MRP4 is in its inward-facing close state, in other words, when the substrate has no access back to the inner compartment. At this stage, the molecular docking in region 2 cannot be considered as comprehensive, as it is performed in an inward-facing open conformer whereas in the inward-facing close conformation the protein chamber accessible volume is likely to be modified.

*Region 3.* The poses in region 3 suggest that this region is either a transient or an allosteric binding site. In the latter case, region 3 could be responsible for transport inhibition or modulation. Indeed, strong binding in this region may prevent large-scale conformational transitions occurring along the transport cycle by modifying interactions between bundles B and D, while the latter should swap with bundle C to reach inward-facing close then the outward-facing conformation[13]. It is worth mentioning that a large overlap between region 1 and 3 is observed in the $L_0$-out conformation. Therefore, key residues will only be discussed for the $L_0$-in conformation. Region 3 is also rich in arginine residues (namely Arg317, Arg303 and Arg272) that can easily bind anionic substrates. These residues are often involved in H-

bonding to the substrates, averaged H-bond fraction being 0.77, 0.39 and 0.20 for Arg317, Arg303 and Arg272, respectively. Interestingly, in the $L_0$-in conformation, docking suggests that the $L_0$-motif is involved in substrate binding in region 3, mostly thanks to Trp64 that was observed to bind aromatic substrates by $\pi-\pi$ stacking interactions.

### IV.2.3.2. Binding modes of model substrates and comparison with experiments

To further validate the robustness of these results, they were compared to previous studies focusing on specific transporter-ligand interactions for well-identified, so prototypical, MRP4 substrates, namely cGMP, folic acid, methotrexate and $E_2\beta_{17}$ glucuronide. To date, joint site-directed mutagenesis and molecular docking studies have shown that a few amino acids might play a key role either in substrate binding or in funneling of the substrate out of the binding pocket[16,17]. These studies suggested the role of Phe368, Phe369, Glu374, Arg375, Glu378, Trp995 and Arg998 even though no structural evidence has been directly provided so far. Here, we suggest that $\pi$-$\pi$ stacking interactions may significantly contribute to the mode of substrate binding. In the present section, attention is paid to the energetic of the best-ranked poses, which are expected to be the most predominant ones.

Table 2. Top-ranked affinities for cGMP, folic acid, methotrexate and $E_2$17$\beta$ glucuronide in region 1, 2 and 3 for L0-in (-out) conformation.

| Substrate | $L_0$-in conformation | | | $L_0$-out conformation | | |
|---|---|---|---|---|---|---|
| | Region 1 | Region 2 | Region 3 | Region 1 | Region 2 | Region 3 |
| cGMP | -9.6 | -8.6 | -8.5 | -9.4 | -8.3 | -9.3 |
| Folic acid | -10.4 | -9.7 | -10.0 | -9.9 | -9.4 | -9.6 |
| Methotrexate | -10.3 | -9.3 | -9.6 | -9.5 | -9.3 | -9.3 |
| $E_2$17$\beta$ glucuronide | -12.0 | -10.4 | -11.2 | -11.4 | -9.8 | -10.9 |

*Binding modes of cGMP.* Regardless $L_0$ location ($L_0$-in or $L_0$-out conformations), *cGMP* preferentially binds to MRP4 at region 1, *i.e.*, in the protein chamber (the top-ranked pose affinities being -9.6, -8.6, and -8.5 kcal/mol for regions 1, 2 and 3 of $L_0$-in conformation, respectively, see Fig. 3A). The mapping of contacts below 4.0 Å suggests that cGMP interacts preferentially with TMHs 12 and 6 (Fig. S5-S6). This clearly underlines the driving contribution of H-bonding to the cGMP binding mode to MRP4, mainly with (cationic) arginine residues. cGMP can interact by H-bonding (Fig. 2A) thanks to H-atom donors and acceptors located over the whole molecule (purine, sugar or intracyclic phosphate moieties, see Fig. 1B). H-bonding between cGMP and both Arg375 and Arg998 is observed at a maximum occurrence of 0.83 and 0.46, respectively in region 1 of the $L_0$-in ($L_0$-out) conformation, which support previous findings suggesting the direct interaction between Arg375/998 and cGMP[16]. Flexible docking calculations over several snapshots also suggested H-bond interactions between cGMP and Arg879, Glu374 and Trp995, averaged H-bond fraction of 0.20, 0.19 and 0.17, respectively, in the $L_0$-in conformation. Concerning Glu374 and Trp995, their presence as key residues in the binding pocket was suggested previously [16]. In the case of the $L_0$-in conformation, Arg21 was also expected to interact with cGMP owing to the presence of the $L_0$-motif in the protein chamber (H-bond fraction of *ca.*, 0.17, see Table 2). Interestingly, the docking in the $L_0$-out conformation also suggested the formation of a H-bond network with Arg998 and Arg375 (H-bond fractions of *ca.* 0.76 and 0.15, respectively).

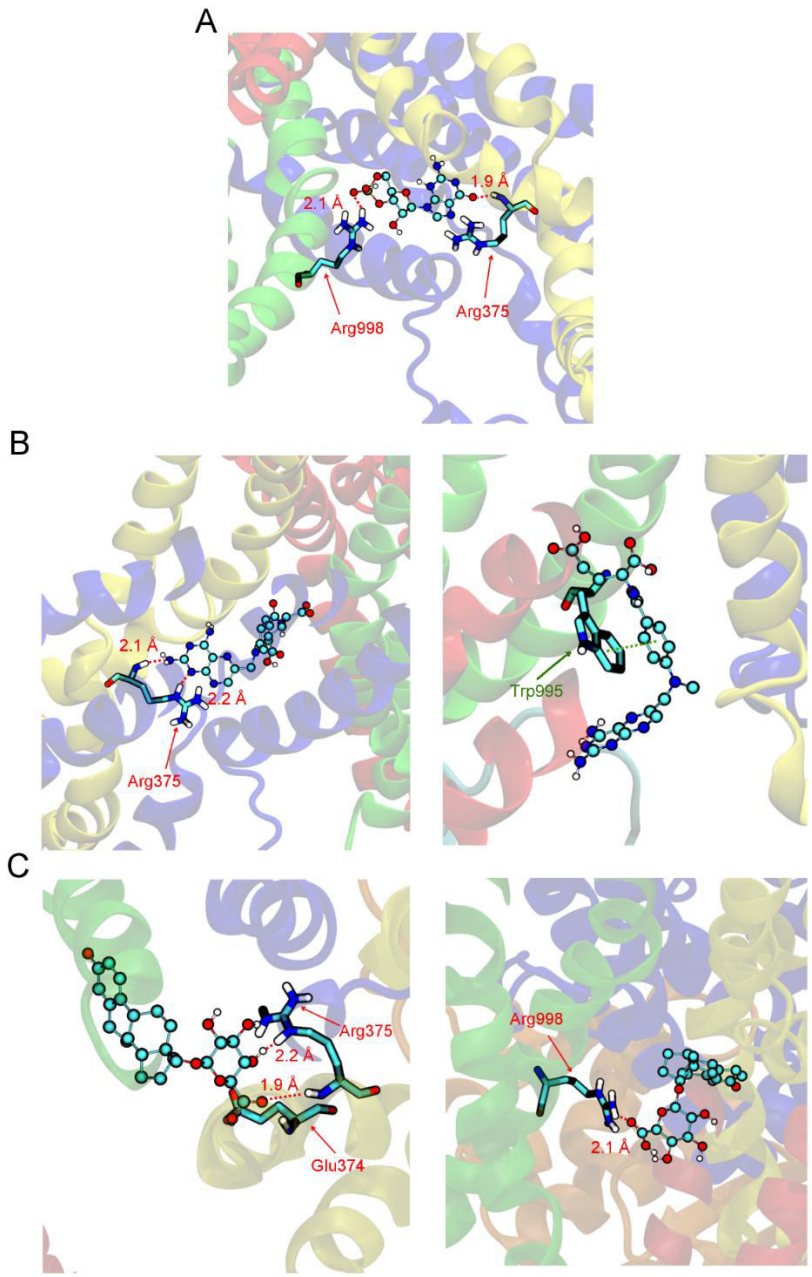



Fig. 2. Representative poses of A) cGMP, B) methotrexate and C) $E_2$17*β* glucuronide with key residues experimentally described.

*Binding modes of methotrexate and folic acid.* Folic acid and methotrexate share the same chemical backbone, namely pteridinyl and *p*-aminobenzamidyl moieties. They both exhibit similar preferential poses in region 1 in close contact with TMH 12, for both the $L_0$-in and $L_0$-out conformations. The calculated affinities in the $L_0$-in conformation were -10.4 and -10.3 kcal/mol for the two compounds, respectively (Table 2). As a comparison, these affinities were -9.7 and -9.3 kcal/mol, respectively, in region 2. Slightly less region-selectivity was seen in the $L_0$-out conformation with affinities of -9.9 (-9.5) *vs.* -9.4 (-9.3) kcal/mol for folic acid (methotrexate) in region 1 *vs.* region 2 (Table 2), suggesting a greater competition between the two binding regions. H-bonding strongly contributes to the modes of binding to both folic acid and methotrexate. Interestingly, Arg375 strongly binds both compounds (H-bond fractions being 0.57 and 0.62, respectively, see Fig. 2B). To a lower extent, Arg998 and Glu374 also contribute to the H-bond network with both substrates (Table S3). As for cGMP, cationic

Arg379 is seen to strongly participate in protein-ligand H-bond network (H-bond fractions being 0.87 and 0.79 for folic acid and methotrexate, respectively). The binding of Trp995 to both folic acid and methotrexate (Fig. 2B) suggests a mode of binding mainly through $\pi$-$\pi$ stacking involving the pteridinyl and tryptophan indolyl moieties, and to a lower extent H-bonding with the tryptophan N-heteroatom. These findings agree with previous studies suggesting the key role of Trp995, Arg375, Arg998 and Glu374. Interestingly, the suggested contributions of Phe368 and Phe369 to protein-ligand interactions appear negligible in region 1. Conversely, $\pi$-$\pi$ stacking between Phe368 and both folic acid and methotrexate were observed in region 2. This suggests a role of Phe368 and Phe369 in the intermediate steps of the substrate translocation process leading to the inward-facing close and outward-facing conformations [13]**.**

*Binding modes of $E_2 17\beta$ glucuronide*. The docking suggested preferential binding of $E_2\beta_{17}$ glucuronide in region 1. Despite the aromatic A-ring of $E_2 17\beta$ glucuronide, the driving contribution is mainly H-bonding thanks to the OH groups of the glucuronide moiety. The key role of Arg998, Arg375, Glu374 is again brought to the light, with maximum H-bond fractions of 0.33, 1.32, and 1.12, respectively[4] (Fig. 2C).

#### IV.2.3.3. Ligand-based structure-property relationship and preferential binding modes in the MRP4 exporter

*Overall preferential binding regions*

To better document preferential binding modes in MRP4 over the 41-ligand dataset, the difference in averaged top-ranked affinity between region 1 and region j ($<\Delta\Delta G_{1,j}>$ with j = 2 or 3, see Table 3). Regardless MRP4 conformation, region 1 is globally the most favorable binding region, $<\Delta\Delta G_{1,2}>$ and $<\Delta\Delta G_{1,3}>$ being -1.0 ± 0.5 (-1.0 ± 0.6) and -0.9 ± 0.5 (-0.2 ± 0.3) kcal/mol, respectively for the $L_0$-in ($L_0$-out) conformation. As expected, region 2 is the least favourable binding region owing to the smallest binding pocket volume and the absence of flexibility due to molecular docking calculations. Due to large overlapping space between regions 1 and 3 in $L_0$-out, both regions were almost equally favourable; conversely, in $L_0$-in a significant difference in affinity was observed (Table 3). This result points out the limitations of the single-pose approach specially to tackle binding in the protein chamber of transporters, for which using the affinity only of the top-ranked pose is highly questionable. To consider multi-pose appears mandatory, given the high flexibility of binding regions; in other word, the greater, the sampling, the better. Thus, from a broad analysis of docking results, Boltzmann distributions weighted by affinities have been evaluated, which may provide a better qualitative overview of binding region competitions (Table 3). Region 1 is the preferred binding site for most of compounds, given the Boltzmann distribution of 0.57 ± 0.09 (0.78 ± 0.07), 0.15 ± 0.08 (0.13 ± 0.06) and 0.28 ± 0.10 (0.10 ± 0.06) in region 1, 2 and 3, respectively, in $L_0$-in ($L_0$-out). This confirms the canonical feature of region 1 for binding to MRP4.

[4] Fraction higher than 100% indicates that more than one H-bond occurs between residue and ligand.

Table 3. Average difference in top-ranked affinity $<\Delta\Delta G_{1,j}>$ (j= 2 or 3 for region 2 or 3, respectively) over the whole 41-ligand dataset, and $D_{reg}i$ are the Boltzmann distribution of regions i for each ligand in the dataset.

| Substrate | MRP4 $L_0$-in conformation | | | | | MRP4 $L_0$-out conformation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\Delta\Delta G_{1,2}$ | $\Delta\Delta G_{1,3}$ | $D_{reg1}$ | $D_{reg2}$ | $D_{reg3}$ | $\Delta\Delta G_{1,2}$ | $\Delta\Delta G_{1,3}$ | $D_{reg1}$ | $D_{reg2}$ | $D_{reg3}$ |
| 2R, 2R-epicatechin | -0.9 | -0.8 | 0.46 | 0.27 | 0.27 | -1.2 | -0.1 | 0.77 | 0.17 | 0.06 |
| 2R, 3S-catechin | -0.9 | -1.4 | 0.52 | 0.19 | 0.29 | -1.4 | 0.2 | 0.83 | 0.11 | 0.06 |
| 2S, 3R-catechin | -1.1 | -1.0 | 0.49 | 0.17 | 0.33 | -0.9 | -0.1 | 0.84 | 0.05 | 0.12 |
| 2S,3S-epicatechin | -0.7 | -0.8 | 0.56 | 0.20 | 0.24 | -1.4 | -0.2 | 0.70 | 0.16 | 0.15 |
| Adefovir | -0.6 | -1.0 | 0.58 | 0.18 | 0.24 | -0.9 | -0.6 | 0.77 | 0.13 | 0.10 |
| cAMP | -1.0 | -0.6 | 0.62 | 0.16 | 0.22 | -1.4 | 0.2 | 0.89 | 0.07 | 0.05 |
| Ceefourin1 | -0.4 | -1.0 | 0.51 | 0.10 | 0.39 | -0.8 | -0.2 | 0.81 | 0.18 | 0.01 |
| Ceefourin2 | -0.8 | -0.7 | 0.78 | 0.10 | 0.12 | -1.2 | -0.2 | 0.77 | 0.17 | 0.06 |
| Ceftmetazole | -0.7 | -0.6 | 0.42 | 0.07 | 0.52 | -1.1 | -0.5 | 0.75 | 0.14 | 0.11 |
| Celecoxib | -1.4 | -1.3 | 0.64 | 0.16 | 0.20 | 0.2 | 0.3 | 0.75 | 0.15 | 0.11 |
| cGMP | -1.0 | -1.1 | 0.61 | 0.13 | 0.26 | -1.1 | -0.1 | 0.85 | 0.10 | 0.06 |
| Cholate | -1.9 | 0.0 | 0.63 | 0.09 | 0.28 | -1.1 | -0.2 | 0.84 | 0.12 | 0.04 |
| Edaravone | 0.1 | -0.1 | 0.48 | 0.31 | 0.21 | -0.9 | -0.6 | 0.74 | 0.19 | 0.06 |
| Edaravone glucuronide | -0.9 | -0.6 | 0.63 | 0.21 | 0.15 | -0.8 | -0.3 | 0.71 | 0.19 | 0.11 |
| $E_2$17$\beta$ glucuronide | -1.6 | -0.8 | 0.77 | 0.05 | 0.18 | -1.6 | -0.5 | 0.61 | 0.07 | 0.33 |
| Folic acid | -0.7 | -0.4 | 0.69 | 0.07 | 0.24 | -0.5 | -0.3 | 0.66 | 0.23 | 0.11 |
| Furosemide | -0.8 | -0.1 | 0.64 | 0.18 | 0.19 | -0.7 | 0.0 | 0.73 | 0.17 | 0.10 |
| Galangin | -0.8 | -0.9 | 0.57 | 0.22 | 0.22 | -1.0 | -0.1 | 0.79 | 0.11 | 0.10 |
| Ganciclovir | -0.3 | -0.5 | 0.59 | 0.29 | 0.12 | -1.5 | -0.5 | 0.75 | 0.19 | 0.06 |
| Hydrochlorothiazide | -1.0 | -1.5 | 0.52 | 0.16 | 0.32 | -2.2 | -1.2 | 0.84 | 0.14 | 0.03 |
| Indometacin | -0.5 | -0.9 | 0.49 | 0.18 | 0.34 | -0.3 | 0.1 | 0.73 | 0.15 | 0.12 |
| Kaempferol | -1.1 | -1.4 | 0.54 | 0.19 | 0.27 | -1.3 | -0.2 | 0.87 | 0.07 | 0.07 |
| Losartan | -0.9 | -0.4 | 0.41 | 0.11 | 0.49 | -0.4 | -0.3 | 0.76 | 0.16 | 0.08 |
| Luteolin | -1.2 | -1.2 | 0.49 | 0.24 | 0.27 | -1.6 | -0.2 | 0.77 | 0.17 | 0.07 |
| Methotrexate | -1.0 | -0.7 | 0.56 | 0.03 | 0.41 | -0.2 | -0.2 | 0.71 | 0.23 | 0.06 |
| Myricetin | -1.1 | -1.9 | 0.59 | 0.15 | 0.26 | -1.4 | -0.3 | 0.82 | 0.12 | 0.06 |
| Naringenin | -1.7 | -1.5 | 0.51 | 0.24 | 0.25 | -1.5 | 0.0 | 0.77 | 0.12 | 0.11 |
| Olmesartan | -1.1 | -0.9 | 0.69 | 0.05 | 0.26 | -1.0 | -0.3 | 0.79 | 0.08 | 0.13 |
| Probenecid | -0.2 | -0.7 | 0.53 | 0.20 | 0.28 | -0.3 | -0.2 | 0.69 | 0.24 | 0.07 |
| Quercetin | -1.5 | -1.7 | 0.52 | 0.18 | 0.30 | -1.2 | -0.1 | 0.86 | 0.10 | 0.04 |
| 3'-*O*-glucuronide | -1.9 | -1.3 | 0.63 | 0.03 | 0.33 | -1.7 | 0.0 | 0.85 | 0.09 | 0.07 |
| 3'-*O*-sulfate | -1.1 | -2.1 | 0.56 | 0.17 | 0.27 | -1.2 | 0.1 | 0.73 | 0.09 | 0.18 |
| Resveratrol | -0.3 | -0.2 | 0.65 | 0.23 | 0.13 | -0.9 | -0.4 | 0.76 | 0.12 | 0.12 |
| Rutin | -1.0 | -0.5 | 0.73 | 0.00 | 0.27 | -1.2 | -0.2 | 0.91 | 0.02 | 0.07 |
| Sildenafil | -1.5 | -1.1 | 0.47 | 0.06 | 0.48 | -0.9 | 0.0 | 0.74 | 0.12 | 0.14 |
| Silybin | -1.6 | -1.2 | 0.70 | 0.03 | 0.27 | -1.8 | -0.2 | 0.82 | 0.01 | 0.16 |
| Taurocholic acid | -1.6 | -0.5 | 0.63 | 0.04 | 0.33 | 0.2 | -0.2 | 0.61 | 0.20 | 0.19 |
| Taxifolin | -1.1 | -1.3 | 0.45 | 0.14 | 0.41 | -0.3 | 0.2 | 0.74 | 0.05 | 0.21 |
| Urate | -0.8 | -1.1 | 0.54 | 0.30 | 0.16 | 0.2 | -0.3 | 0.92 | 0.04 | 0.04 |
| ε-viniferin | -0.9 | -0.7 | 0.53 | 0.08 | 0.39 | -1.0 | 0.1 | 0.77 | 0.03 | 0.20 |

*Ligand-based structure-property relationship of MRP4 protein-ligand interactions*

The use of multi-replica molecular docking enables thorough analyses to establish a ligand-based structure property relationship of binding to the chamber of MRP4. The 41 substrates included in the dataset were grouped into four subsets defined by the presence of either (i) a purine-like ring, (ii) a steroid moiety, (iii) a flavonoid skeleton or (iv) a sugar moiety (Scheme S1 and Table S4). Some substrates bearing two of these moieties were in the corresponding two subsets.

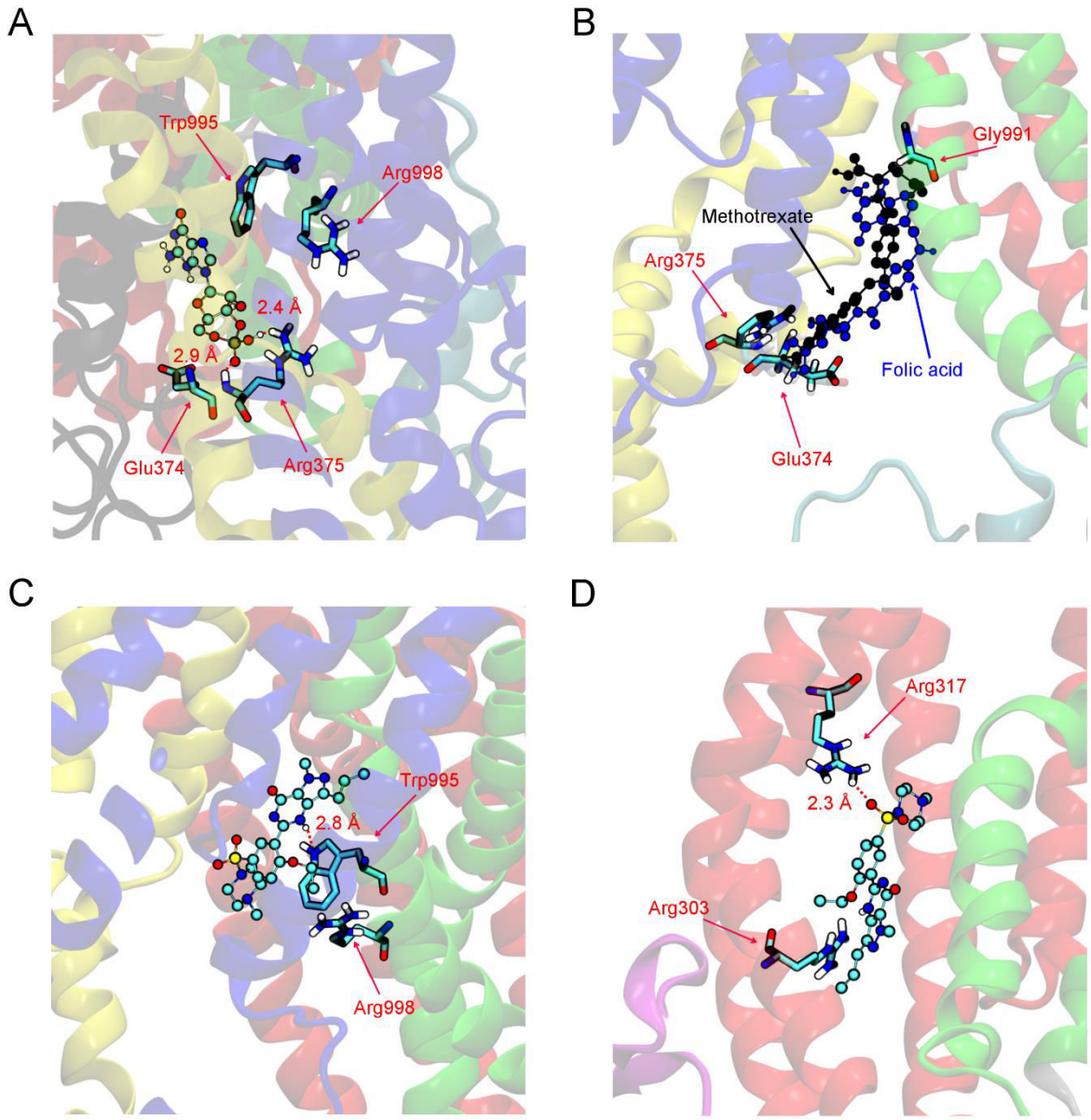



Fig. 3. Top ranked poses and surrounding key residues (within a 4 Å sphere) of A) cGMP in region 1; B) methotrexate and folic acid in region 1; and C) sildenafil in regions 1 and 3.

Table 4. Key residues and corresponding occurrence of $\pi-\pi$ stacked poses in the A) $L_0$-in and B) $L_0$-out conformations.

(A)

| Region | Residues | Dataset | Purine-like | Sugars | Steroids | Flavonoids |
|---|---|---|---|---|---|---|
| Region1 | Tyr881 | 0.53 | 0.53 | 0.54 | 0.54 | 0.53 |
| | Trp995 | 0.43 | 0.44 | 0.42 | 0.43 | 0.43 |
| Region2 | Phe369 | 0.91 | 0.92 | 0.92 | 0.91 | 0.92 |
| Region3 | Trp64 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |

(B)

| Region | Residues | Dataset | Purine-like | Sugars | Steroids | Flavonoids |
|---|---|---|---|---|---|---|
| Region1 | Trp995 | 0.97 | 0.96 | 0.96 | 0.97 | 0.96 |
| Region2 | Trp995 | 0.81 | 0.86 | 0.85 | 0.82 | 0.79 |
| | Phe875 | 0.18 | 0.14 | 0.15 | 0.17 | 0.21 |
| Region3 | Trp995 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |

*Purine-like groups.* This subset contains substrates bearing: (i) a purine moiety (cAMP, cGMP, urate, adefovir, ganciclovir); (ii) a pteridine moiety (folate, methotrexate); or (iii) a 5/6-fused N-aromatic ring moiety (ceefourin1, ceefourin2, indometacine, sildenafil). These substrates bind MRP4 in the same pocket of region 1 mostly either with Arg375, Arg879, Arg998 and Glu374 by H-bonding, or with Trp995 by $\pi-\pi$ stacking interactions (Table 4). Trp995 can also interact with purine-like moiety by H-bonding in $L_0$-in (average H-bond fraction of 0.19). The orientation of the substrates in this pocket appears very flexible, as the driving force is the formation of a H-bond network with the different key residues mentioned above. For example, the top-ranked poses of methotrexate and folic acid exhibit two opposite orientations, however involving 4 Å-distance contact with the same residues (namely Glu374, Arg375 and Gly991) despite high structural similarity (Fig. 3B). This agrees with and it rationalizes the high specificity but low selectivity of ABCC/MRP transporters because of high flexibility in binding site regions.

Table 5. Top-rank affinities for all 41ligands in each region of $L_0$-in and $L_0$-out conformations.

| Substrate | $L_0$-in | | | $L_0$-out | | |
|---|---|---|---|---|---|---|
| | Region 1 affinity (kcal/mol) | Region 2 affinity (kcal/mol) | Region 3 affinity (kcal/mol) | Region 1 affinity (kcal/mol) | Region 2 affinity (kcal/mol) | Region 3 affinity (kcal/mol) |
| 2R, 2R-epicatechin | -9.7 | -8.8 | -8.9 | -9.7 | -8.5 | -9.6 |
| 2R, 3S-catechin | -9.9 | -9.0 | -8.5 | -9.6 | -8.2 | -9.8 |
| 2S, 3R-catechin | -9.9 | -8.8 | -8.9 | -9.7 | -8.8 | -9.6 |
| 2S,3S-epicatechin | -9.7 | -9.0 | -8.9 | -9.7 | -8.3 | -9.5 |
| Adefovir | -7.9 | -7.3 | -6.9 | -7.6 | -6.7 | -7.0 |
| cAMP | -9.2 | -8.2 | -8.6 | -9.2 | -7.8 | -9.4 |
| Ceefourin1 | -8.3 | -7.9 | -7.3 | -7.9 | -7.1 | -7.7 |
| Ceefourin2 | -10.2 | -9.4 | -9.5 | -10.1 | -8.9 | -9.9 |
| Ceftmetazole | -9.0 | -8.3 | -8.4 | -8.7 | -7.6 | -8.2 |
| Celecoxib | -10.9 | -9.5 | -9.6 | -9.6 | -9.8 | -9.9 |
| cGMP | -9.6 | -8.6 | -8.5 | -9.4 | -8.3 | -9.3 |
| Cholate | -9.9 | -8.0 | -9.9 | -9.9 | -8.8 | -9.7 |
| Edaravone | -6.9 | -7.0 | -6.8 | -7.2 | -6.3 | -6.6 |
| Edaravone glucuronide | -9.6 | -8.7 | -9.0 | -9.3 | -8.5 | -9.0 |
| $E_2 17\beta$ glucuronide | -12.0 | -10.4 | -11.2 | -11.4 | -9.8 | -10.9 |
| Folic acid | -10.4 | -9.7 | -10.0 | -9.9 | -9.4 | -9.6 |
| Furosemide | -8.8 | -8.0 | -8.7 | -8.4 | -7.7 | -8.4 |
| Galangin | -9.8 | -9.0 | -8.9 | -9.3 | -8.3 | -9.2 |
| Ganciclovir | -7.6 | -7.3 | -7.1 | -7.9 | -6.4 | -7.4 |
| Hydrochlorothiazide | -8.6 | -7.6 | -7.1 | -9.3 | -7.1 | -8.1 |
| Indometacin | -9.4 | -8.9 | -8.5 | -8.8 | -8.5 | -8.9 |
| Kaempferol | -9.9 | -8.8 | -8.5 | -9.8 | -8.5 | -9.6 |
| Losartan | -9.9 | -9.0 | -9.5 | -9.4 | -9.0 | -9.1 |
| Luteolin | -10.1 | -8.9 | -8.9 | -10.0 | -8.4 | -9.8 |
| Methotrexate | -10.3 | -9.3 | -9.6 | -9.5 | -9.3 | -9.3 |
| Myricetin | -10.3 | -9.2 | -8.4 | -9.6 | -8.2 | -9.3 |
| Naringenin | -10.3 | -8.6 | -8.8 | -9.7 | -8.2 | -9.7 |
| Olmesartan | -11.2 | -10.1 | -10.3 | -10.3 | -9.3 | -10.0 |
| Probenecid | -7.8 | -7.6 | -7.1 | -7.1 | -6.8 | -6.9 |
| Quercetin | -10.4 | -8.9 | -8.7 | -9.9 | -8.7 | -9.8 |
| 3'-*O*-glucuronide | -10.9 | -9 | -9.6 | -10.9 | -9.2 | -10.9 |
| 3'-*O*-sulfate | -10.6 | -9.5 | -8.5 | -9.5 | -8.3 | -9.6 |
| Resveratrol | -8.2 | -7.9 | -8.0 | -8.6 | -7.7 | -8.2 |
| Rutin | -11.7 | -10.7 | -11.2 | -10.8 | -9.6 | -10.6 |
| Sildenafil | -10.4 | -8.9 | -9.3 | -9.9 | -9.0 | -9.9 |
| Silybin | -12.4 | -10.8 | -11.2 | -11.6 | -9.8 | -11.4 |
| Taurocholic acid | -10.8 | -9.2 | -10.3 | -9.7 | -9.9 | -9.5 |
| Taxifolin | -10 | -8.9 | -8.7 | -9.2 | -8.9 | -9.4 |
| Urate | -7.6 | -6.8 | -6.5 | -6.8 | -7.0 | -6.5 |
| ε-viniferin | -11.3 | -10.4 | -10.6 | -10.9 | -9.9 | -11.0 |

A particular attention has been paid to cGMP and sildenafil, since sildenafil is known to inhibit cGMP hydrolysis in many tissues by binding allosteric site of phosphodiesterase 5[32]. Based on top-ranked affinity, sildenafil might preferentially bind MRP4 prior to cGMP, suggesting a competitive inhibition (Table 5). Moreover, sildenafil exhibits a significantly similar or higher

occurrence in region 3 than in region 1 of $L_0$-in, Boltzmann distributions being *e.g.*, 0.47 and 0.48 for region 1 and 3, respectively (Table 3). This reinforces the hypothesis that region 3 is an allosteric binding site involved inhibition mechanism, hence drug-drug interactions. Interestingly, ceefourin1 which is a specific MRP4 inhibitor[24] exhibits a similar behavior in the $L_0$-in conformation, Boltzmann distributions of ceefourin1 being 0.51 and 0.39 for region 1 and 3, respectively (Table 3). This finding should be confirmed by site-directed mutagenesis, targeting this allosteric binding site, namely Arg317 and Arg303.

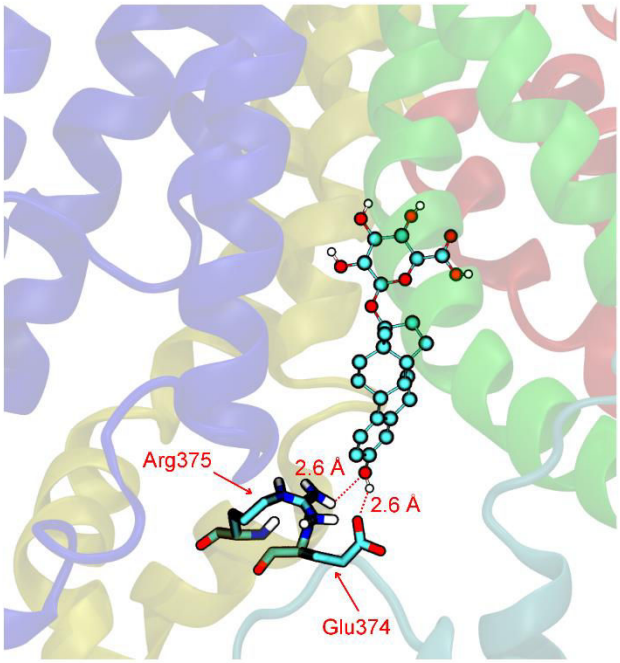



Fig. 4. Representative pose of $E_2 17\beta$ glucuronide in region 1.

*Steroids.* Three endogenous steroids known as substrates of MRP4 were docked. The three of them preferentially bind the same binding pocket as purine-like compounds, *i.e.*, region 1, here again confirmed as the canonical binding site of MRP4 (Fig. 4). H-bonding interactions are key in the binding mode in region 1, owing to the presence of the OH-groups in the steroid skeleton. $E_2 17\beta$ glucuronide is the only substrate of this subset for which $\pi$–$\pi$ stacking interaction were expected, however, only a few poses exhibited this arrangement with Trp995 and Tyr881 (21 poses over 911 selected poses for $L_0$-in). A stronger occurrence of H-bonding with both Glu374 and Arg375 was observed in this subset, highlighting again the central role of these two residues. For instance, the fraction of H-bonding with Arg375 and Glu374 were 0.99 and 0.69, respectively, for the steroid subset, whereas they were 0.66 and 0.31, respectively, for the whole dataset. The docking in $L_0$-in also suggested a possibility for steroids to bind the (allosteric) region 3.

*Flavonoids.* Quercetin and naringenin have been shown to be inhibitors of MRP4[21], the latter substrate being capable of regulating efflux of cyclic nucleotides such as cAMP and cGMP[33–35]. Several flavonoids were thus docked to provide structural trends responsible for inhibition. They preferentially bind the canonical binding site, even though a significant part may bind the (allosteric) region 3 in $L_0$-in. The docking also showed that most of flavonoids can bind region 2 as well (Tables 1, 3 and 5). This behaviour (efficient binding in the different regions of the MRP4 protein chamber) may be attributed to the relatively low molecular weight of flavonoids, as well as their amphiphilic character, namely the $\pi$–conjugated flavonoid skeleton to favour

$\pi$–$\pi$ interactions and hydrophobic effects on one side, and several OH-groups favouring H-bond interactions on the other side. Flavonoids preferentially bind Arg375 and Arg998 by H-bond interactions, regardless to the MRP4 conformation (Table 1). Flavonoids can be divided into subgroups according to the aromaticity degree of their central ring (Scheme 1). The so-called C-ring is aromatic in flavonol and flavones, whereas it is not in flavanols such as naringenin. The molecular docking suggested the central role of the C-ring, as for the top-ranked poses in region 1, naringenin and quercetin systematically adopted $\pi$-$\pi$ stacking arrangement, mostly involving their B-ring. No $\pi$-$\pi$ stacked poses were observed for catechins and epicatechins (Fig. 5A), regardless their configurations. One can expect that the joint presence of the 3-OH group and absence of aromaticity in C-ring preclude $\pi$-$\pi$ stacked orientation. More surprisingly, $\pi$-$\pi$ stacked poses were observed with myricetin among the top-ranked poses. This might be assigned to the predominant role of H-bond interactions owing to the presence of three OH-groups on myricetin B-ring (Fig. 5B).

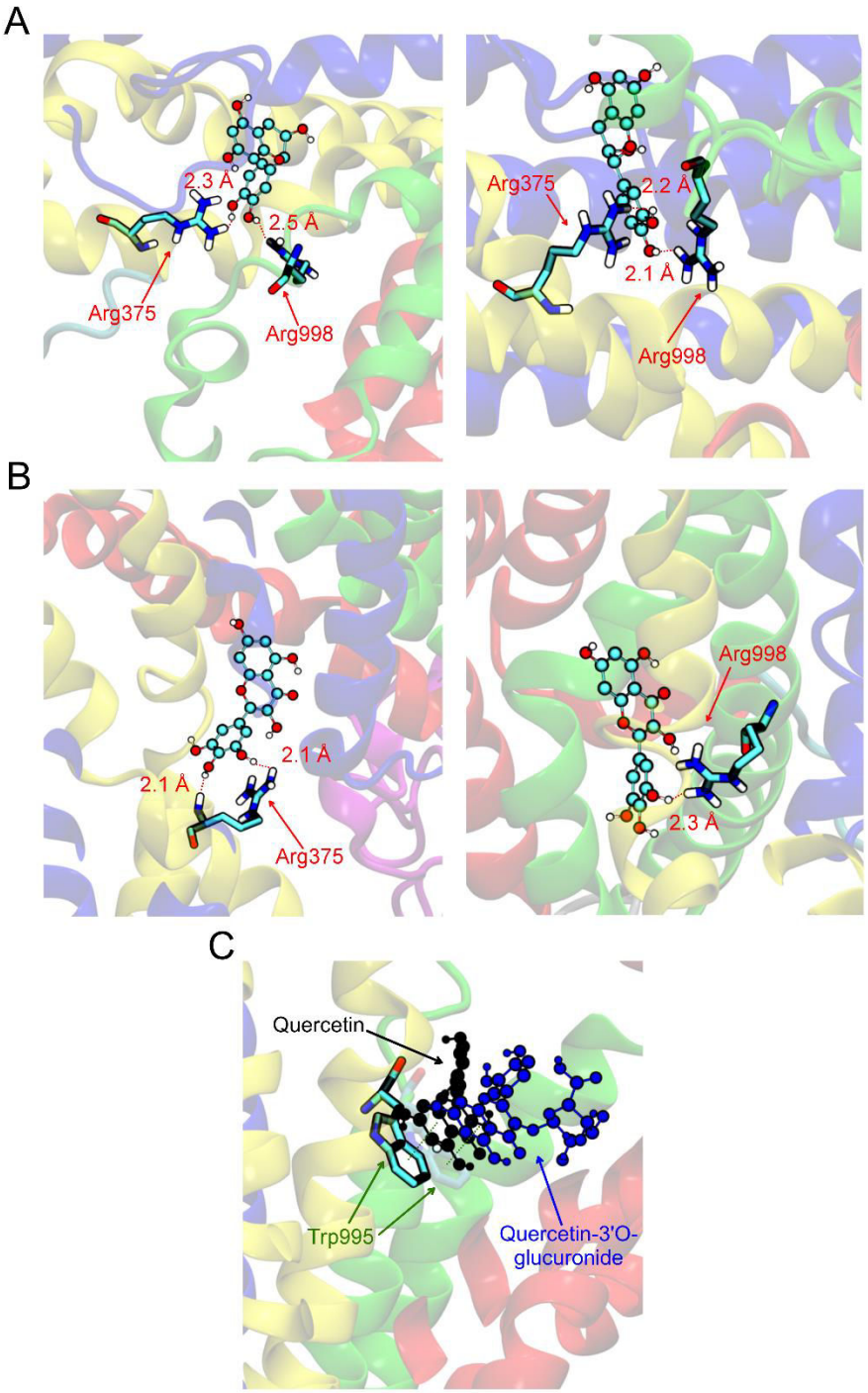



Fig. 5. Representative poses for A) 2S, 3R-catechin and 2R, 2R-epicatechin, B) myricetin and C) quercetin and quercetin-3'O-glucuronide. The H-bonding is highlighted in red and the $\pi$–$\pi$ stacking is highlighted in green.

*Sugar moieties.* MRP4 is known to carry glucuronide substrates[36]. The dataset includes a few glucosides (rutin), glucuronides ($E_2 17\beta$ glucuronide, quercetin and edaravone glucuronides) and cAMP and cGMP bearing a ribose moiety. Sugar moieties are expected to play a strong anchor to the protein by providing both H-bond acceptor and donor groups. For instance, the affinity of edaravone to MRP4 was significantly increased in the presence of the glucuronide moiety (top-ranked poses $\Delta G$ of -6.9 and -9.6 kcal/mol for the standalone edaravone and its glucuronide, respectively). However, quercetin 3-*O*-glucuronide did not

exhibit the same behavior. Concerning both region 1 and 3, affinities and Boltzmann distributions of standalone quercetin and its glucuronide were similar (Table 3). This might be explained by the predominant role of $\pi$-$\pi$ stacking interactions in this case, both compounds adopting the same conformation in the binding sites (Fig. 5C). The main difference between both compounds is observed in region 2, for which the presence of the sugar moiety leads to steric clashes, preventing favorable anchoring in this region.

## References


[1] K. Lee, M.G. Belinsky, D.W. Bell, J.R. Testa, G.D. Kruh, Isolation of MOAT-B, a widely expressed multidrug resistance-associated protein/canalicular multispecific organic anion transporter-related transporter, Cancer Res. 58 (1998) 2741–2747.

[2] F.G.M. Russel, J.B. Koenderink, R. Masereeuw, Multidrug resistance protein 4 (MRP4/ABCC4): a versatile efflux transporter for drugs and signalling molecules, Trends Pharmacol. Sci. 29 (2008) 200–207. doi:10.1016/j.tips.2008.01.006.

[3] K.M. Hillgren, D. Keppler, A.A. Zur, K.M. Giacomini, B. Stieger, C.E. Cass, L. Zhang, International Transporter Consortium, Emerging transporters of clinical importance: an update from the International Transporter Consortium, Clin. Pharmacol. Ther. 94 (2013) 52–63. doi:10.1038/clpt.2013.74.

[4] International Transporter Consortium, K.M. Giacomini, S.-M. Huang, D.J. Tweedie, L.Z. Benet, K.L.R. Brouwer, X. Chu, A. Dahlin, R. Evers, V. Fischer, K.M. Hillgren, K.A. Hoffmaster, T. Ishikawa, D. Keppler, R.B. Kim, C.A. Lee, M. Niemi, J.W. Polli, Y. Sugiyama, P.W. Swaan, J.A. Ware, S.H. Wright, S.W. Yee, M.J. Zamek-Gliszczynski, L. Zhang, Membrane transporters in drug development, Nat Rev Drug Discov. 9 (2010) 215–236. doi:10.1038/nrd3028.

[5] N. Abla, L.W. Chinn, T. Nakamura, L. Liu, C.C. Huang, S.J. Johns, M. Kawamoto, D. Stryke, T.R. Taylor, T.E. Ferrin, K.M. Giacomini, D.L. Kroetz, The Human Multidrug Resistance Protein 4 (MRP4, ABCC4): Functional Analysis of a Highly Polymorphic Gene, J Pharmacol Exp Ther. 325 (2008) 859–868. doi:10.1124/jpet.108.136523.

[6] R. van de Ven, G.L. Scheffer, A.W. Reurs, J.J. Lindenberg, R. Oerlemans, G. Jansen, J.-P. Gillet, J.N. Glasgow, A. Pereboev, D.T. Curiel, R.J. Scheper, T.D. de Gruijl, A role for multidrug resistance protein 4 (MRP4; ABCC4) in human dendritic cell migration, Blood. 112 (2008) 2353–2359. doi:10.1182/blood-2008-03-147850.

[7] U. Gradhand, T. Lang, E. Schaeffeler, H. Glaeser, H. Tegude, K. Klein, P. Fritz, G. Jedlitschky, H.K. Kroemer, I. Bachmakov, B. Anwald, R. Kerb, U.M. Zanger, M. Eichelbaum, M. Schwab, M.F. Fromm, Variability in human hepatic MRP4 expression: influence of cholestasis and genotype, Pharmacogenomics J. 8 (2008) 42–52. doi:10.1038/sj.tpj.6500451.

[8] H. Nikaido, How are the ABC transporters energized?, PNAS. 99 (2002) 9609–9610. doi:10.1073/pnas.162375699.

[9] P.-A. Billat, T. Ossman, F. Saint-Marcoux, M. Essig, J.-P. Rerolle, N. Kamar, L. Rostaing, H. Kaminski, G. Fabre, M. Otyepka, J.-B. Woillard, P. Marquet, P. Trouillas, N. Picard, Multidrug resistance-associated protein 4 (MRP4) controls ganciclovir intracellular accumulation and contributes to ganciclovir-induced neutropenia in renal transplant patients, Pharmacological Research. 111 (2016) 501–508. doi:10.1016/j.phrs.2016.07.012.

[10] A.K. Nanayakkara, C.A. Follit, G. Chen, N.S. Williams, P.D. Vogel, J.G. Wise, Targeted inhibitors of P-glycoprotein increase chemotherapeutic-induced mortality of multidrug resistant tumor cells, Scientific Reports. 8 (2018) 967. doi:10.1038/s41598-018-19325-x.

[11] J. Renes, E.E. de Vries, G.J. Hooiveld, I. Krikken, P.L. Jansen, M. Müller, Multidrug resistance protein MRP1 protects against the toxicity of the major lipid peroxidation product 4-hydroxynonenal., Biochem J. 350 (2000) 555–561.

[12] B. Marquez, F. Van Bambeke, ABC multidrug transporters: target for modulation of drug pharmacokinetics and drug-drug interactions, Curr Drug Targets. 12 (2011) 600–620.

[13] M. Moradi, E. Tajkhorshid, Mechanistic picture for conformational transition of a membrane transporter at atomic resolution, PNAS. 110 (2013) 18916–18921. doi:10.1073/pnas.1313202110.

[14] B. Verhalen, R. Dastvan, S. Thangapandian, Y. Peskova, H.A. Koteiche, R.K. Nakamoto, E. Tajkhorshid, H.S. Mchaourab, Energy transduction and alternating access of the mammalian ABC transporter P-glycoprotein, Nature. 543 (2017) 738–741. doi:10.1038/nature21414.

[15] B. Chantemargue, F. Di Meo, K. Berka, N. Picard, H. Arnion, M. Essig, P. Marquet, M. Otyepka, P. Trouillas, Structural patterns of the human ABCC4/MRP4 exporter in lipid bilayers rationalize clinically observed polymorphisms, Pharmacol. Res. 133 (2018) 318–327. doi:10.1016/j.phrs.2018.02.029.

[16] A.A.K. El-Sheikh, J.J.M.W. van den Heuvel, E. Krieger, F.G.M. Russel, J.B. Koenderink, Functional Role of Arginine 375 in Transmembrane Helix 6 of Multidrug Resistance Protein 4 (MRP4/ABCC4), Mol Pharmacol. 74 (2008) 964–971. doi:10.1124/mol.107.043661.

[17] H.G.M. Wittgen, J.J.M.W. van den Heuvel, E. Krieger, G. Schaftenaar, F.G.M. Russel, J.B. Koenderink, Phenylalanine 368 of multidrug resistance-associated protein 4 (MRP4/ABCC4) plays a crucial role in substrate-specific transport activity, Biochemical Pharmacology. 84 (2012) 366–373. doi:10.1016/j.bcp.2012.04.012.

[18] A.W. Ravna, G. Sager, Molecular model of the outward facing state of the human multidrug resistance protein 4 (MRP4/ABCC4), Bioorg. Med. Chem. Lett. 18 (2008) 3481–3483. doi:10.1016/j.bmcl.2008.05.047.

[19] M.S. Jin, M.L. Oldham, Q. Zhang, J. Chen, Crystal structure of the multidrug transporter P-glycoprotein from C. elegans, Nature. 490 (2012) 566–569. doi:10.1038/nature11448.

[20] A.W. Ravna, I. Sylte, G. Sager, Binding site of ABC transporter homology models confirmed by ABCB1 crystal structure, Theor Biol Med Model. 6 (2009) 20. doi:10.1186/1742-4682-6-20.

[21] C.-P. Wu, A.M. Calcagno, S.B. Hladky, S.V. Ambudkar, M.A. Barrand, Modulatory effects of plant phenols on human multidrug-resistance proteins 1, 4 and 5 (ABCC1, 4 and 5), FEBS J. 272 (2005) 4725–4740. doi:10.1111/j.1742-4658.2005.04888.x.

[22] M.D. Hanwell, D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek, G.R. Hutchison, Avogadro: an advanced semantic chemical editor, visualization, and analysis platform, J Cheminform. 4 (2012) 17. doi:10.1186/1758-2946-4-17.

[23] O. Trott, A.J. Olson, AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, Journal of Computational Chemistry. 31 (2010) 455–461. doi:10.1002/jcc.21334.

[24] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility, J Comput Chem. 30 (2009) 2785–2791. doi:10.1002/jcc.21256.

[25] P.H. Palestro, L. Gavernet, G.L. Estiu, B. Blanch, L. E, Docking Applied to the Prediction of the Affinity of Compounds to P-Glycoprotein, BioMed Research International. (2014). doi:10.1155/2014/358425.

[26] S. Forli, R. Huey, M.E. Pique, M.F. Sanner, D.S. Goodsell, A.J. Olson, Computational protein-ligand docking and virtual drug screening with the AutoDock suite, Nat Protoc. 11 (2016) 905–919. doi:10.1038/nprot.2016.051.

[27] D.R. Roe, T.E. Cheatham, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, J. Chem. Theory Comput. 9 (2013) 3084–3095. doi:10.1021/ct400341p.

[28] R.A.M.H. Van Aubel, P.H.E. Smeets, J.J.M.W. van den Heuvel, F.G.M. Russel, Human organic anion transporter MRP4 (ABCC4) is an efflux pump for the purine end metabolite urate with multiple allosteric substrate binding sites, Am. J. Physiol. Renal Physiol. 288 (2005) F327-333. doi:10.1152/ajprenal.00133.2004.

[29] I.V. Turko, S.A. Ballard, S.H. Francis, J.D. Corbin, Inhibition of Cyclic GMP-Binding Cyclic GMP-Specific Phosphodiesterase (Type 5) by Sildenafil and Related Compounds, Mol Pharmacol. 56 (1999) 124–130. doi:10.1124/mol.56.1.124.

[30] L. Cheung, C.L. Flemming, F. Watt, N. Masada, D.M.T. Yu, T. Huynh, G. Conseil, A. Tivnan, A. Polinsky, A.V. Gudkov, M.A. Munoz, A. Vishvanath, D.M.F. Cooper, M.J. Henderson, S.P.C. Cole, J.I. Fletcher, M. Haber, M.D. Norris, High-throughput screening identifies Ceefourin 1 and Ceefourin 2 as highly selective inhibitors of multidrug resistance protein 4 (MRP4), Biochem. Pharmacol. 91 (2014) 97–108. doi:10.1016/j.bcp.2014.05.023.

[31] P.P. Pattabiraman, P.E. Pecen, P.V. Rao, MRP4-mediated regulation of intracellular cAMP and cGMP levels in trabecular meshwork cells and homeostasis of intraocular pressure, Invest. Ophthalmol. Vis. Sci. 54 (2013) 1636–1649. doi:10.1167/iovs.12-11107.

[32] Y. Hara, Y. Sassi, C. Guibert, N. Gambaryan, P. Dorfmüller, S. Eddahibi, A.-M. Lompré, M. Humbert, J.-S. Hulot, Inhibition of MRP4 prevents and reverses pulmonary hypertension in mice, J. Clin. Invest. 121 (2011) 2888–2897. doi:10.1172/JCI45023.

[33] F. Hofmann, R. Feil, T. Kleppisch, J. Schlossmann, Function of cGMP-dependent protein kinases as revealed by gene deletion, Physiol. Rev. 86 (2006) 1–23. doi:10.1152/physrev.00015.2005.

[34] J. Bai, L. Lai, H.C. Yeo, B.C. Goh, T.M.C. Tan, Multidrug resistance protein 4 (MRP4/ABCC4) mediates efflux of bimane-glutathione, The International Journal of Biochemistry & Cell Biology. 36 (2004) 247–257. doi:10.1016/S1357-2725(03)00236-X.

## IV.3. An attempt to exploring the Inward-Facing to Outward-Facing path of the SAV1866 ABC transporter by metadynamics

*Note to the reader: This section is written and edited as a publication; which however has not been submitted yet.*

### IV.3.1. Introduction

There exist many clinical conditions in which complex drug cocktails (including immunosuppressants, antibiotics, antivirals, antidiuretics, or analgesics) are required, which are often life-long treatments. Such drug cocktails may affect many biological functions, including functioning of membrane protein transporters[1,2], which are involved in drug influx and efflux. In such situation, drug-drug interactions (DDI) involving membrane transporters have been related to unwanted adverse effects, leading to either interruption of the treatment or lowering of the drug dose, and potentially graft rejection and patient morbidity[3]. Two superfamilies of drug transporters have been involved in drug pharmacokinetics and pharmacodynamics (PK/PD)[1], namely the solute carrier[1,4] (SLC) and ATP-binding cassette[5] (ABC) transporter families. The former is involved in both drug influx and efflux whereas the latter is exclusively dedicated to drug efflux in human cells. The International Transporter Consortium (ITC)[6,7] has described a series of human transporters from both families as being of "emerging clinical importance", *e.g.*, ABCB1 (known as P-glycoprotein, P-gp), or ABCG2 (known as Breast Cancer Resistance Protein, BCRP)[1,8]. Other ABC transporters have gained interest, *e.g.*, ABCC1, 2 and 4 / MRP1, 2 and 4 (multidrug resistance protein types 1, 2 and 4), which are expressed at key locations for drug disposition/action such as in the liver, the kidney and peripheral blood cells, and have a broad range of xenobiotics as substrates[1,9,10]. Many polymorphisms in these transporters have been highlighted in the (dys)-functions of drug transport, however the molecular rationalization is highly fragmented and has deserved much attention in the recent years.

Drug transport through ABC membrane proteins is a very complex mechanism due to the complexicity of their structural features, which are characterized by highly interlinked intra/extra-cellular and transmembrane domains. Most ABC transporters are made of twelve (*i.e.*, 2 times 6) TransMembrane alpha-helices (TM), which constitutes four different bundles, namely bundles A-D (Figure 1), and two nucleotide binding domains (NBDs, see Figure 1), which are highly conserved regions, and which contain the catalytic sites where ATP molecules can bind and can be hydrolyzed. Although most of three-dimensional structures of ABC transporters have been elucidated in their inward-facing open (IF-o) conformer, two other conformers have also been identified (Figure 1), mainly by recent cryo-EM studies[11]. It has been shown that the apo-state (conformer bound neither with ATP nor with substrate) of the human P-gp co-exists as an IF-o and an IF-close (IF-c) state[12]. Outward-facing (OF) conformers have also been elucidated, *e.g.*, for the bovine MRP1 transporter[13]. Based on these structural data and the experimental conditions used to stabilize these different conformers, various mechanisms for drug translocation through ABC transporter have been suggested. The beginning of the story can be seen from the IF-o, which is considered as the resting state, in which the two NBDs are far from each other. Then the transport requires large conformational changes. ATP binding and its hydrolysis into ADP are key steps in this process. Although the two NBDs can approach to mimic the IF-c conformer by thermal fluctuation, stabilization of this conformer is mandatory to pursue the process. This stabilization requires ATP binding into NBD 1 and 2 and a head-to-tail lock (NBD dimerization). Large conformational

changes occur in the transmembrane domains, in bundles are interchanged (Figure 1). There exists no experimental way to fully tackle the dynamics of ABC transporters, even though cryo-EM can provide some intermediates states[12]. In principle, only *in silico* methods can elucidate this highly dynamic process, however regarding the size of the molecular system, it is a very challenging task of modern theoretical chemistry. Recently, a few studies based on biased molecular dynamics (MD), have got new insights into these mechanisms[14,15]. The energetic landscape of the IF-o ↔ OF process was studied on the apo-state of the bacterial MsbA transporter by non-equilibrium MD simulations, following key collective variables (CV)[15]. Although the distance between the two NDBs is an obvious CV, the twisting between the two NBDs was here shown as a key CV to trigger the large conformational changes occurring in the IF-o ↔ OF process. This highlighted the importance of the NBD-NBD interface (predominantly of electrostatic nature and highly modified during twisting) and the highly cooperative NBD-TMD interactions. By elucidating the minimum free-energy path, this MD-based study has highlighted several intermediates states, which were connected to confirm an alternative-access process, as also shown by cryo-EM[11].

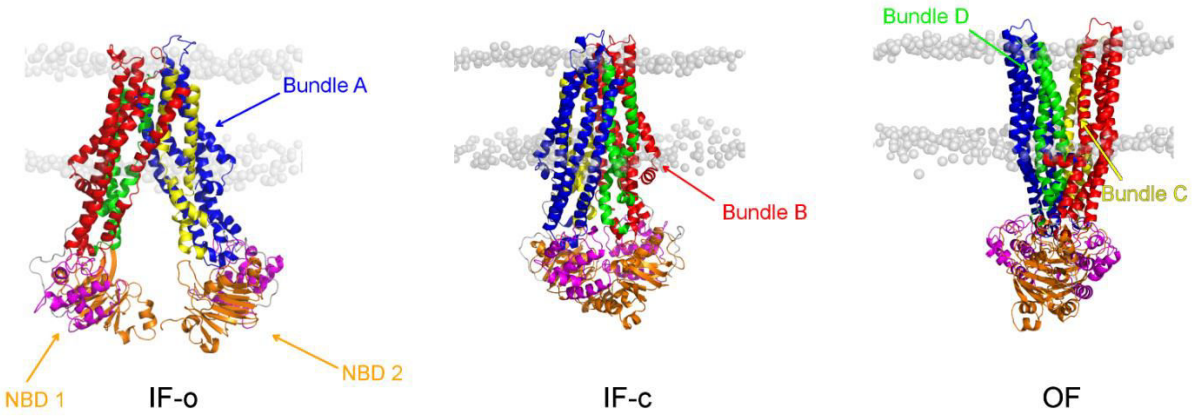



Figure 1: Three different states known for ABC transporters inward-facing open (IF-o), inward-facing close (IF-c) and outward-facing (OF) conformers. Bundles A, B, C and D are depicted in blue, red, yellow and green, respectively. NBD1 and 2 are depicted in yellow and purple, respectively.

Metadynamics (MTD) is a biased MD method which has appeared promising and elegant is the sense that it is based on a smooth populating process of the different basins which are representative of a given conformation of a molecular system. It is based on the addition of Gaussian functions in the phase space of a given CV value. This methodology reduces the phase space of a very complex highly multidimensional mechanism (say the ABC transporter IF ↔ OF process) into a low dimensional space. By choosing a set of CVs ($N_{CV}$ = number of CVs) which are representative of the process, the real phase space is focused on the CVs' space, which has $N_{CV}$ as a dimension. The addition of the bias potential (here Gaussian functions) at a CV value and at regular time steps allows building a history-dependent bias-potential ($V(s,t)$), where $s$ is the value of the CV at time $t$. In between each interval, a MD simulation is achieved to relax the system around the initial value of the CV under the constrain of the new Gaussian-like potential which is added. All along the MTD simulation, the total bias potential, $V(s,t)$, equals the sum of the (repulsive) Gaussian functions in the CV space:

$$V(s,t) = \sum_{i=1}^{t/\tau_G} \omega_G exp\left[\sum_{\alpha=1}^{N_{CV}} \frac{\left(s_\alpha - s_\alpha(i\tau_G)\right)^2}{2\sigma_\alpha^2}\right] \qquad (\text{Eq. 1}),$$

where $\omega_G$ is the height of the added Gaussian, $\sigma_\alpha$ is the width of the Gaussian, $\tau_G$ is the deposition time interval (*i.e.*, regular time interval at which a new Gaussian is added to the potential), $\frac{\omega_G}{\tau_G} = \omega$ is named deposition rate, $s_\alpha$ is the value of the CV at time t of the simulation when focusing on CV n°$\alpha$, and $s_\alpha(i\tau_G)$ is the value of the CV at the time when the Gaussian is added.

At small $\tau_G$ values, the expression of $V(s,t)$ can be written as an integral:

$$V(s,t) = \omega \int_0^t dt'.exp\left[\sum_{\alpha=1}^{N_{CV}} \frac{\left(s_\alpha - s_\alpha(t')\right)^2}{2\sigma_\alpha^2}\right] \qquad \text{(Eq. 2)}.$$

This history-dependent bias-potential allows populating the different basins representative of the IF ↔ OF process, *i.e.*, all intermediate conformers, but also transition states which are milestones of the process. In other word, if the MTD simulation is converged it means that the sampling of the conformational space is sufficient, and the potential energy surface of the CV space is given as the negative of $V(s,t)$.

The choice of the $\omega_G$, $\sigma_\alpha$ and $\tau_G$ is a critical step in MTD. If the sampling is too slow (*e.g.*, too small $\omega_G$ and $\sigma_\alpha$ values), the MTD process does not converge in a reasonable timescale. Conversely, if too high $\omega_G$ values are set, the free energy surface is error prone, as the system can be driven in states which are physically irrelevant. A fraction of the thermal energy $k_{BT}$ can be taken as a first guess of $\omega_G$ values, however this should be, *a posteriori*, tuned again according to the system as it could be insufficient to escape wells surrounded by high barriers within a reasonable timescale. $\tau_G$ values should be large enough to allow sufficient relaxation of the molecular system between two successive deposits, hence to avoid placing a new Gaussian on top of the previous deposited one. However, too high $\tau_G$ values may induce discontinuities in the construction of the free energy surface.

An unsuccessful convergence in the MTD process can also be due to the choice in the CVs. Indeed, CVs should be representative of the process, *i.e.*, allowing to guide the systems from the initial to the final conformer through the different basins (the path). If a key CV is missing, the sampling will barely reach convergence and the system could likely be stuck in a CV space that could not allow to reconstruct the canonical ensemble (global phase space). In MTD, only a few CVs can be selected. A solution to accelerate the sampling without using too high $\omega_G$, $\sigma_\alpha$ and $\tau_G$ values, is to combine the process to replica-exchange or bias-exchange techniques[16–18]. In the latter case (bias-exchange), several replicas (initial geometries) are subjected to MTD simulations performed in parallel and the bias is regularly exchanged from one to another replica. The exchange is proposed every $\tau$ time-unit and it is accepted, or not, following the Metropolis algorithm, *i.e.*, following the probability to exchange bias $V_G^i$ with bias $V_G^j$:

$$p_{ij} = min\left(1, e^{\left[V_G^i(s_i)+V_G^j(s_j)-V_G^i(s_j)-V_G^j(s_i)\right]/k_BT}\right) \qquad \text{(Eq. 3)},$$

where $V_G^i(s_i)$ and $V_G^j(s_j)$ are the bias potential i applied on replica i and the bias potential j applied on replica j, respectively, and $V_G^i(s_j)$ and $V_G^j(s_i)$ are the bias potential i applied on replica j and the bias potential j applied on replica i, respectively.

In bias-exchange metadynamics (BE-MTD), a super set of CVs (as independent as possible) can be selected, which increases the chance to tackle the problem with the appropriate CVs, *i.e.*, representative of the process. The present manuscript is an attempt to use BE-MTD to study the IF ↔ OF process of Sav1866. Sav1866 is a bacterial ABC transporter; it was the first elucidated ABC structure isolated from *Staphylococcus aureus*[19]. It is homo-dimer made of the twelve transmembrane domains arranged in bundles A-D, and two NBDs. Due to the complexity of the system, the first section is dedicated to the methodological strategy, providing technical details and highlighting the different limitations faced by this study. From this knowledge, the second section deals with the process in terms of conformational changing, which provides a dynamic view of the ABC transporter and the xenobiotic channel.

### IV.3.2. Computational details and strategy

*Protein models*

The Sav1866 model in OF conformation was prepared using the X-ray structure of SAV1866 from *Staphylococcus aureus* (2HYD[19], 3.0 Å). It was embedded in a POPC (palmitoyl-2-oleoylphosphatidylcholine) lipid bilayer and it was relaxed according the MD procedure described below. The IF-o and IF-c structures were obtained by homology modeling using the I-TASSER[20,21] webserver. The construction of IF-o was achieved sequentially for the two monomers. Namely one monomer was constructed by using, as templates: the structure of the NBDs of MsbA from *Acinetobacter baumannii* (5IDV_A, 1.45 Å resolution); the P-gp structure in IF conformer of the TM287/288 transporter from *Thermotoga* maritima (3QF4_B[22], 2.9 Å resolution); and the structure of NBDs from ABCB6 from *Homo Sapiens (*3NH6_A[23], 2.0 Å resolution). The second monomer was simply duplicated from the first one and both monomers were assembled to construct the Sav1866 homodimer in IF-o. The IF-c model was constructed following the same procedure as for IF-o with the following templates: the P-gp structure from *Caenorhabditis elegans* (4F4C[24], 3.4 Å resolution) and the P-gp structure from *Mus musculus* (4Q9H[25], 3.4 Å resolution).

The robustness of the IF models was assessed by various scores. The C-score, or confidence score, for which the rule is the higher C-score value, the higher the confidence in the predicted model, within the [-5, 2] interval. For IF-o and IF-c models, the C-score was -0.2 and 2, respectively[21]. The template modeling (TM)-score also allows measuring structural similarity between template and modeled structures. A TM-score higher than 0.5 indicates a similar fold between the model and proteins for which the topology has already been robustly elucidated[26], the maximum value being 1. For both IF-o and IF-c, the TM-score was 0.69 ± 0.12 and 0.99 ± 0.04, respectively, confirming a correct topology of both models. The 70 ns unbiased MD simulations of the models embedded in POPC lipid bilayers (see below for simulation details) allow relaxation of the system by exploring the near-neighboring conformational space of the homology models. As a post-(MD)relaxation analysis, the Errat score estimated the quality of these models based on pairwise (non-covalently bonded) interactions; a score above 90% indicating a good model, *i.e.*, correct overall secondary structure as compared to high-resolution known 3D protein structures. It was 94.4% and 93.9% for IF-o and IF-c, respectively. Additionally, the Ramachandran plots (as obtained using RAMPAGE webserver[27]) for IF-o conformer showed that 85.6 and 11.5% of residues were

in favored and allowed regions, respectively, while 3.0% were in outlier regions (Figure S1). Concerning the IF-c conformer, 94.9 and 4.6% of residues were in favored and allowed regions, respectively, while 0.5% were in outlier regions (Figure S1). The (ProSA) Z-score[28] also validated the robustness of both models, which were in the range of X-ray and NMR-derived structural parameters (Figure S2). Regarding the different scores obtained after reconstruction/relaxation of the IF-c geometry was strongly robust to be considered for the biased MD simulation. Although, the IF-o geometry was slightly less robust, the scores were high enough to consider the geometry as a relevant replica. Moreover, in the IF-o state, the inter domain interactions are less critical than in IF-c, attenuating the impact of the slightly less good score with IF-o compared to IF-c.

*Molecular dynamics settings*

The three states of the Sav1866 ABC transporter (IF-o, IF-c and OF) were embedded in lipid bilayers made of POPC, using the g_membed tool[29]. The GROMOS54a7 force field[30] was used to describe both the lipids and the proteins. The systems were hydrated with an explicit solvent, using the SPC/E[31] model to describe water molecules. Na+ and Cl- ions were added at a physiological concentration of 0.154 mol%. MD simulations were performed with a 2 fs integration time step. Periodic boundary conditions were applied in all three dimensions. Above a cut-off set at 1.0 nm, long-range electrostatic interactions were treated by the particle-mesh Ewald (PME) method[32]. The van der Waals interactions were smoothly switched to 0 from 0.9 to 1.0 nm. The temperature was maintained at 310 K with the Nosé-Hoover thermostat[33,34] using a 0.5 ps integration time step. The pressure was maintained at 1.013 bar with the Parrinello-Rahman barostat[35] using a 10 ps integration time step, under semi-isotropic conditions. Constraints were applied on bonds with hydrogens using the SHAKE algorithm. The water molecule system was preliminary minimized, keeping the membrane and the protein frozen, and then shortly thermalized to 310 K during 500 ps under (N,V,T) ensemble conditions. The whole system was then thermalized to 310 K during 500 ps under (N,V,T) ensemble conditions. Equilibration of the whole molecular system at 310 K was ensured by a 1 ns MD simulation under (N,P,T) ensemble conditions. For all three Sav1866 states (IF-o, IF-c and OF) a preliminary 70 ns unbiased MD simulation was performed. All MD simulations were performed using the GROMACS 4.5.5 package[36].

*Preliminary biased-exchange metadynamics (BE-MTD) parameterization*

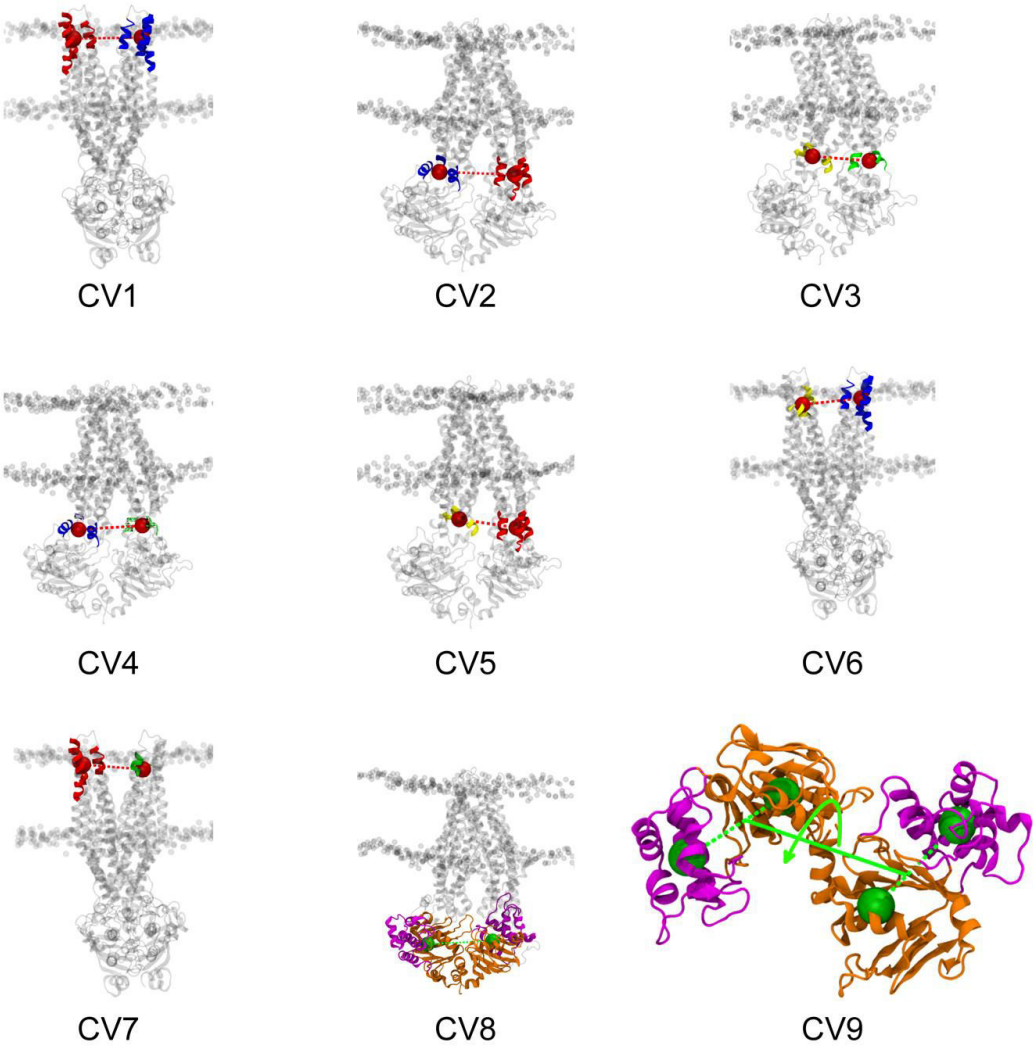
The functioning of ABC transporters, which are embedded in membranes, is a very complex process. It means that from an energetic point of view, the system moves on a complex multidimensional landscape. Due to this complexity, the estimation of the MTD parameters, especially $\omega_G$, cannot be simply achieved from preliminary unbiased MD simulations, as it is often recommended. We therefore performed a preliminary BE-MTD simulation with a set of parameters classical considered as reasonable, namely 0.6 kJ/mol and 0.05 CV-unit for $\omega_G$ and $\sigma_\alpha$, respectively [37–39]. Similarly, a first set of ten CVs was chosen according to the geometrical differences between the OF, IF-c and IF-o states. Inter-domain distances, angles between transmembrane bundles, torsion between transmembrane domains and a torsion between the two NBD domains constituted this set of ten CVs (see Figure S3). The inter-domain distances appeared as the most intuitive geometrical descriptors of the ABC functioning from one to another state. The angles and the torsion between transmembrane domains were added to *a priori* better mimic the flexibility of the system. The torsion between NBDs was inspired from the parameter described by Moradi *et al.*[15]. For this preliminary BE-

MTD simulation, a 10 ps value was chosen for $\tau_G$, as for the exchange rate which was set to 20 ps. At this preliminary stage, the biased-exchange was performed only with the OF conformer. The data were analyzed after 100 ns simulation for each CV, making a total exploration time of 1 µs. The potential energy curves for all CVs allowed a coarse-grained estimation of the potential barriers to escape from the wells. As expected, the energetic barriers to exchange between the different states were relatively high, *ca.* several tens of kJ/mol (Figure S4), which agree with the energy release by ATP hydrolysis, which is known to participate in catalyzing the OF ↔ IF process. In homo-dimers, the energy release by ATP hydrolysis can be doubled. Additionally, the surrounding lipid bilayer is expected to act (mainly entropically) on the ABC transporter which could add to the global energy required to proceed to the exchange from the IF to the OF state. These different considerations are coherent with high activation free energies required to go from one to another basin. All BE-MTD simulations were achieved by Plumed 2.1[40].

*Definition of the different replicas and of nine CVs for the second biased-exchange metadynamics (BE-MTD) simulation*

Convergence in potential energy curve *vs.* each CV was not reached. Therefore, at this stage, all parameters for the BE-MTD simulations were revisited to accelerate the exploration of the phase space. First, different geometries were used to define the different replicas, namely, the OF, IF-c and IF-o states, which obviously greatly enhanced the sampling of the entire process by forcing the system to exchange and to reach the OF, IF-c and IF-o basins, although the energetic barriers could not be undertaken, at least at the beginning of the BE-MTD simulations.

All CVs were more well-thought than with the preliminary BE-MTD, from a thorough analysis of the structural differences between the different replicas. The angles were not considered anymore as appearing too much correlated with the inter-domain distances. Nine CVs (Scheme 1) were chosen as being representative of the OF ↔ IF conformational changes, including bundle swapping. CV1 and 2 were defined as the distances between the top (extracellular) and the bottom (intracellular), respectively of bundles A and B. CV3 was defined as the distance between the bottom of bundles C and D. CV4 was defined as the distance between the bottom of bundles A and D. CV5 was defined as the distance between the bottom of bundles B and C. CV6 was defined as the distance between the top of bundles A and C. CV7 was defined as the distance between the top of bundles B and D. C1-7 were chosen to explicitly describe bundle swapping, *i.e.*, in IF state, bundle A is in close contact with bundle C and bundle B is in close contact with bundle D, whereas in the OF state, bundle A is in close contact with bundle D and bundles B is in close contact with bundle C (see Figure 1). CV8 was defined as the distance between the center of mass of the two NDBs, this CV appearing mandatory to follow the conformational modification between IF-o and IF-c. CV9 was kept from the previous simulation, *i.e.*, twist between NBDs, as inspired from the study of Moradi *et al.*[15], which highlighted this geometrical parameter as a key CV in the conformational transition occurring in the MsbA transporter.




Scheme 1: Definition of CV1-9 for the BE-MTD simulations. Bundles A, B, C and D are depicted in blue, red, yellow and green, respectively. NBD1 and 2 are depicted in yellow and purple, respectively. For CV1-7, the top and bottom of the bundles are precisely defined by specific residues which are quoted in Table S1. The red or green spheres represent the centers of mass of these selected residues to define the CVs.

*Definition of $\omega_G$, $\sigma_\alpha$ and $\tau_G$*

The $\omega_G$ values of the different CVs were chosen as being roughly tenth of the energetic barrier estimated by the preliminary BE-MTD (Table 1). The $\sigma_\alpha$ values were chosen to smoothly decompose the motion in the different basins (Table 1); they were set at the same order of magnitude than the standard deviation of the CVs observed from the unbiased MD simulations. To define a correct $\tau_G$ values is a critical issue which requires special attention. Too high $\tau_G$ values would require too long to explore the energetic landscape of the process. Too low values would not allow a sufficient relaxation of the system after the bias was added (relaxation between two successive Gaussian deposits); subsequently the Gaussians would be added at the same CV value and would not allow exploration of the potential wells. A correct guide to choose $\tau_G$ is the autocorrelation time of the CVs of the system. It is usually of the ps timescale

but it is system-dependent; the greater the complexity and interconnection, the higher the autocorrelation time values. We estimated the autocorrelation time from the 70 ns unbiased MD simulations. The autocorrelation function $C_f(t)$ was calculated as follows:

$$C_f(t) = \frac{1}{N-j} \sum_{i=0}^{N-1-j} f(i\Delta t) f\big((i+j)\Delta t\big) \qquad \text{(Eq. 4)},$$

where $f(t)$ is the property, $N$ is the number of available time frames and $\Delta t$ is the time interval.

As seen in Figure 2 for three CVs, it was around 100 ps. Such high values could be attributed to the highly-interconnected domains of the ABC transporters, which means that any motion of a given domain can impact allosterically on another domain (say NBD1 onto NBD2 through the transmembrane helices, which in turn may affect the motion of NBD1 itself).

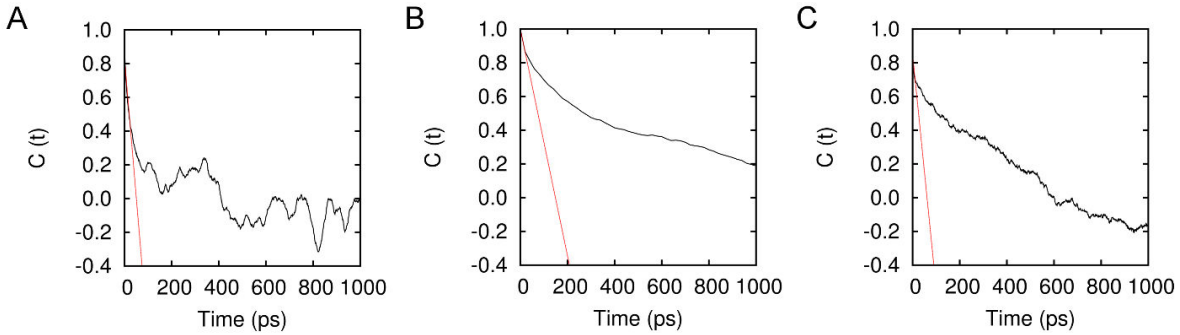



Figure 2: Autocorrelation function *vs.* time for A) CV1, B) CV8 and C) CV9, as obtained from the 70 ns unbiased MD simulations on the OF conformer. The red line is the tangent of the autocorrelation function curve at time 0.

| CV | $\tau_G$ (ps) | $\omega_G$ (kJ/mol) | $\sigma_\alpha$ (unit of the CV) | Interval of CV evolution (unit of the CV) |
|---|---|---|---|---|
| CV1 | 100 | 8 | 0.08 nm | [1.8 nm; 5.5 nm] |
| CV2 | 100 | 16 | 0.06 nm | [2.0 nm; 5.0 nm] |
| CV3 | 100 | 8 | 0.06 nm | [2.2 nm; 5.0 nm] |
| CV4 | 100 | 8 | 0.08 nm | [2.0 nm; 4.5 nm] |
| CV5 | 100 | 8 | 0.09 nm | [2.0 nm; 4.5 nm] |
| CV6 | 100 | 8 | 0.08 nm | [1.8 nm; 3.5 nm] |
| CV7 | 100 | 8 | 0.1 nm | [1.8 nm; 3.5 nm] |
| CV8 | 100 | 16 | 0.04 nm | [2.2 nm; 5.5 nm] |
| CV9 | 100 | 4 | 0.06 rad | [-2.8 rad; -1.9 rad] [-160°; -109°] |

Table 1: Set of parameters selected for the BE-MTD simulations for each CV.

*Exploration of the CV space*

To check whether the parameterization is adapted to sufficiently sample the conformational space, thus to access the canonical ensemble of the Sav1866 transporter dynamics, is a critical issue. Within the context of biased MD simulation, this can only be achieved on a small part of the phase space, namely the low-dimensional projection on the phase space defined by the different CVs. Within the BE-MTD formalism, a first indicator of sampling is the rate of exchange between the different replicas. According to the three topologies used to define the different replicas, a sufficient exchange rate would mean at least a sampling of the basins around the different intermediate states of the IF ↔ OF process, and ideally free jumps between the different basins. The exchange was proposed every 500 ps and it was accepted following the Metropolis algorithm, see (Eq. 3). The exchange rate was regular along the simulation time; whatever the time range used to calculate this rate, it was ranging from 19 to 28% (Figures 3 and 4). This highlighted an extensive sampling of the OF, IF-c and IF-o basins along the different CVs as well as interconnections between the basins (Figure 5). A value above 25% highlights a reasonable exchange rate between replicas, however in our case, it is fluctuating around this value and it does not converge to a threshold value as it would be expected in case free diffusion of the ABC (IF ↔ OF) dynamics.

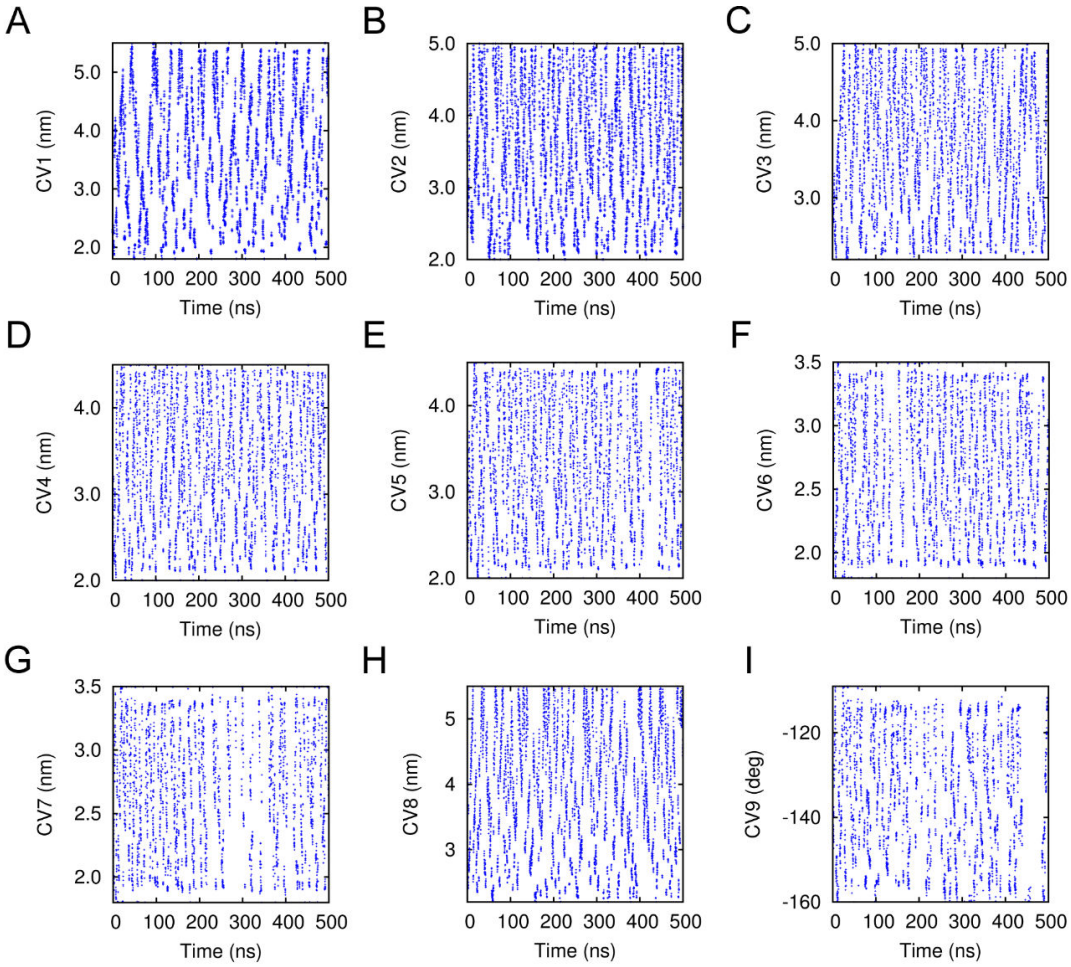



Figure 3: Evolution of the CV values in the corresponding 1D space for all CVs (A to I for CV1 to CV9, successively) along the 500 ns simulation time.

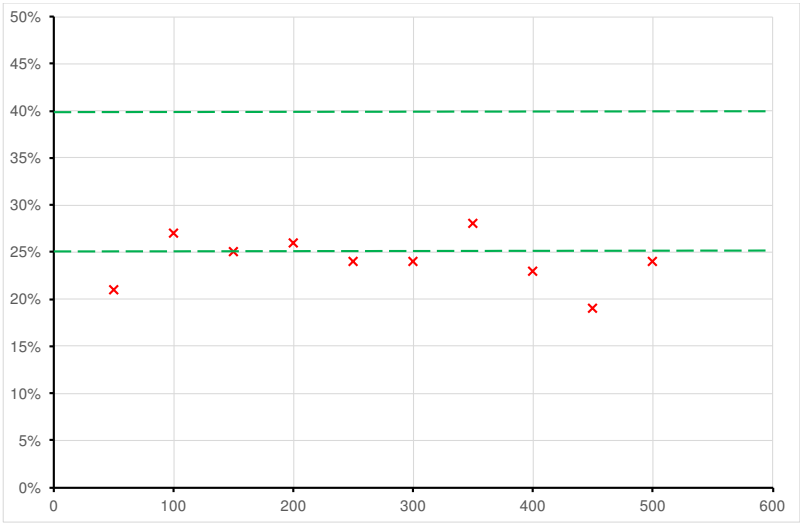



Figure 4: Exchange rate *vs.* simulation time steps. The green zone represents the confidence interval.

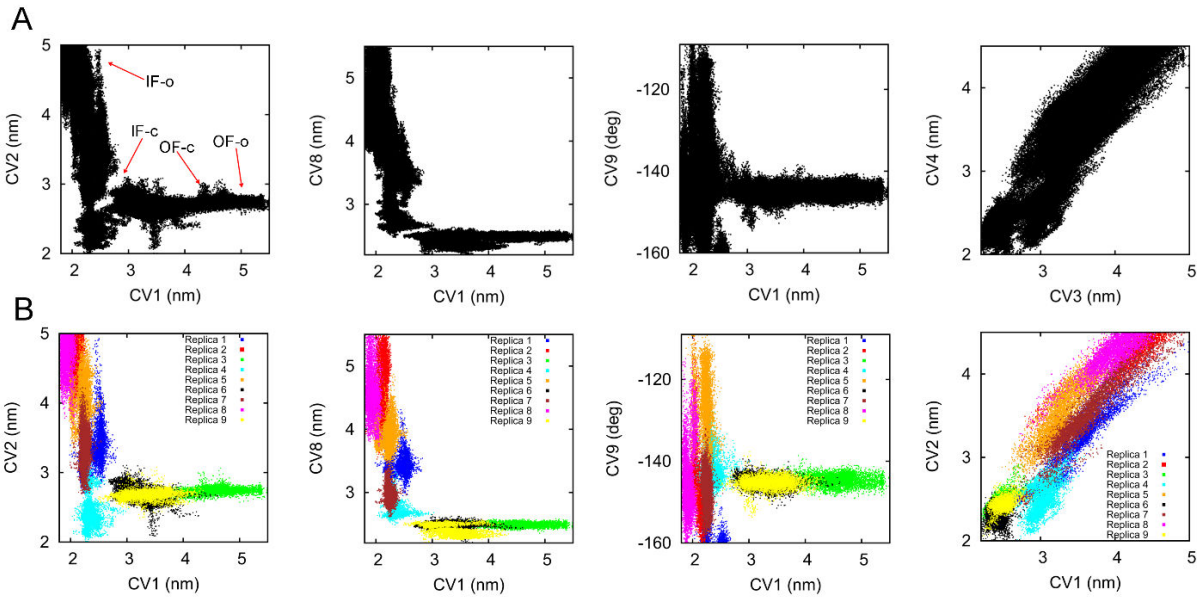



Figure 5: Exploration of the CV space depicted as the distribution of values of CV1-2, CV1-8, CV1-9 and CV3-4 pairs, showing correlation between all these pairs of CVs. A) Exploration of CV pairs between all replicas, B) exploration of CV pairs following each replica. Coarse locations of the OF-c, OF-o, IF-c and IF-o related basins are indicated by red arrows.

Figure 5 also highlighted that some CVs were correlated (linear-shape curve) and were not mandatory to describe the transfer between the different basins. A focus was therefore given only to CV1, 2, 8 and 9, namely distances between the top (extracellular) of bundles A and B, the bottom (intracellular) of bundles A and B, inter-NBD distance and twist between NBDs, respectively. For these four CVs, we followed the biased potential *vs.* time which is supposed to converge when the BE-MTD simulation is completed. Figure 6 plots the negative of the potential energy curve *vs.* each CV, in which the reference free energy at each step was taken at the minimum of each potential energy curve. A series of potential wells were observed which reflect different basins representative of the OF, IF-c and IF-o conformers. However, no convergence was observed for any of these wells. Instead, a continuous exploration of the space was observed. This agrees with the non-convergence observed with the exchange rate,

and it reflects that either the exploration was not completed at 500 ns simulation time per CV, or the exploration was rapidly achieved by a too quick filling of the wells (too high $\omega_G$), and a constant but unnecessary adding of extra Gaussian above 100 ns. The latter option was precluded by plotting the potential energy curves on shorter time intervals, showing that the local wells were correctly dug by the BE-MTD process (Figure 7). This reflects that sampling is on the way but not completed, and extraction from local wells would require much longer time scale and/or possibly the use of other CVs representative of the process, yet not included.

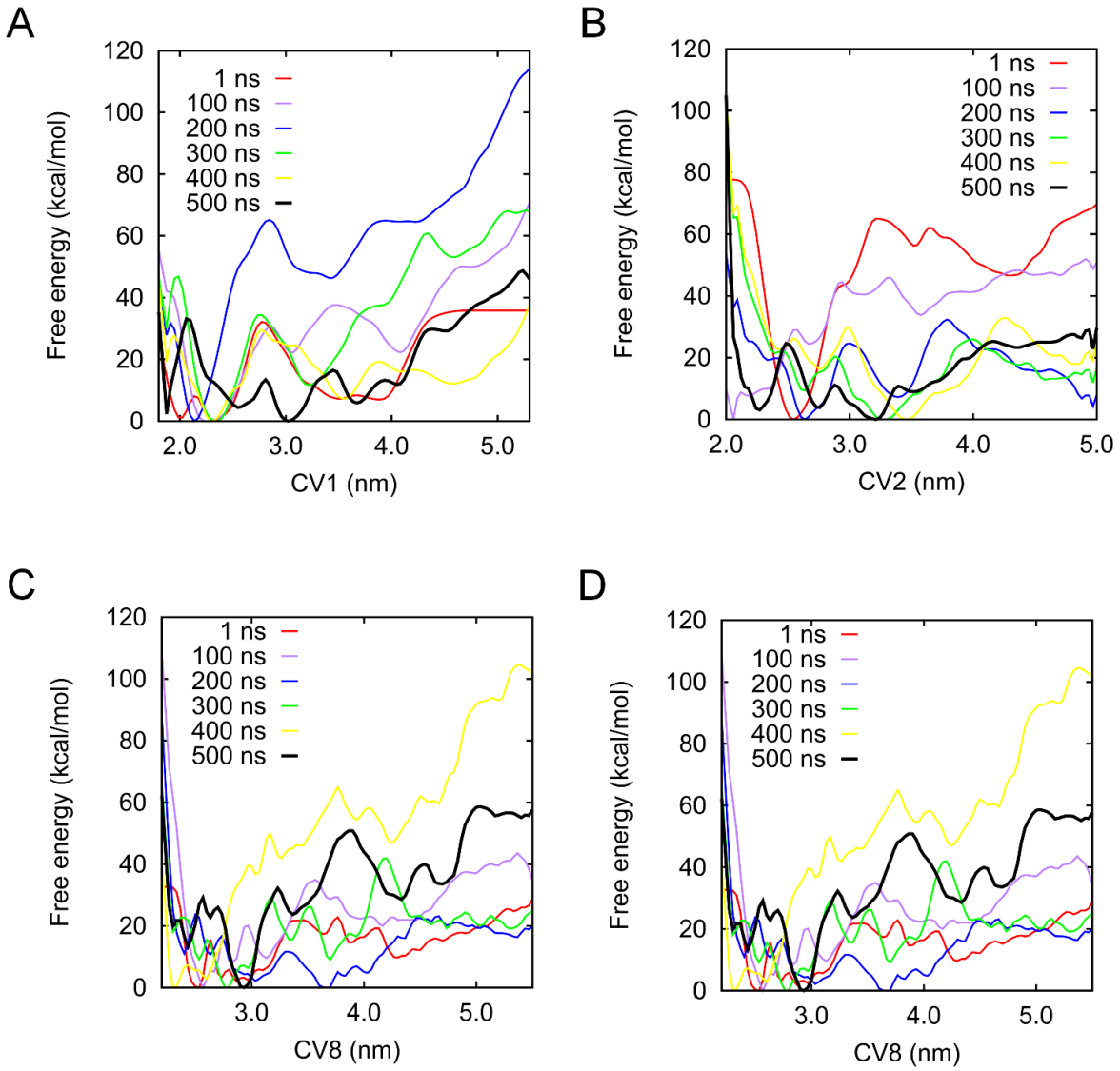



Figure 6: Negative of the biased potential -V(s) *vs.* A) CV1, B) CV2, C) CV8 and D) CV9 along the 500 ns simulation time for each CV. The minimum of the energy for each CV was set at zero. The color palette is red, purple, blue, green, yellow for the different simulation step and finally black for the last profile; 1000 Gaussians were added every 100 ns.

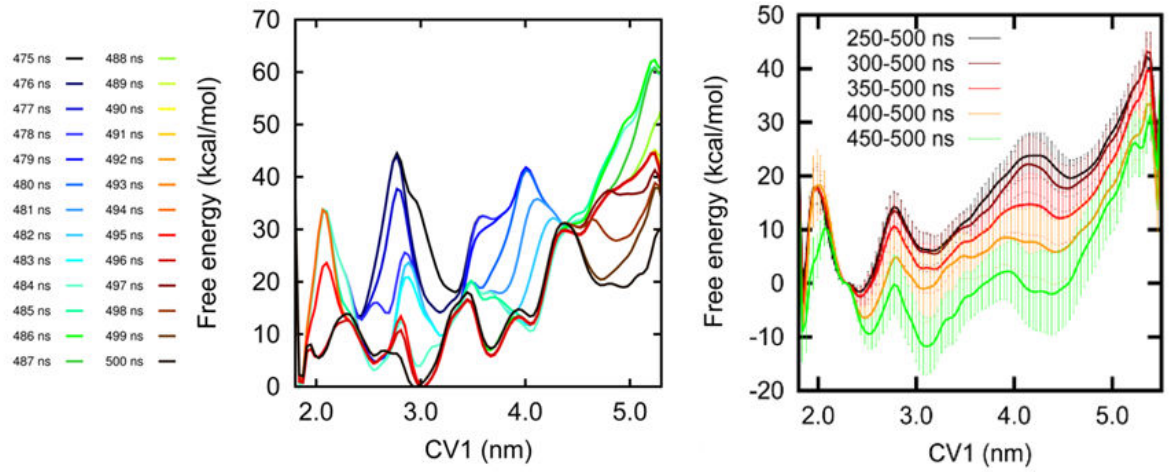



Figure 7: Left panel: Negative of the biased potential -V(s) *vs.* CV1 in the 475-500 ns range of the simulation time. The minimum of the energy for each CV was set at zero. 1000 Gaussians were added every 100 ns. Right panel: History dependent potential as a function of CV1 obtained by averaging over various time periods from the middle to the end of the simulation. The error bars were estimated by block analysis after each of these times.

*Analysis of the conformational landscape*

To recast CV phase space into the global phase space can only be achieved by cluster analysis, *i.e.*, regrouping all frames of all BE-MTD trajectories exhibiting similar structural characteristics (*i.e.*, in this case, similar CV1, CV2, CV8 and CV9 values) into bins, which define a hypercube (having the number of CVs, $N_{CV}$, as a dimension). The different bins correspond to different microstates, which should be characteristic of the representative basins of the IF ↔ OF process. In principle, the population of each bin, $N(s)$, can be related to the relative free energy, $F(s)$, of each corresponding microstate, which is given, considering a Boltzmann distribution, by:

$$F(s) = -k_B T . log N(s) \text{ (Eq. 5).}$$

However, as far as the converge is not fully ensured, these energies should be taken with care and can only be considered as a relative value, showing how much populated were the different microstates explored at a given BE-MTD simulation time. The cluster analysis was performed using METAGUI[41] in combination with VMD. The grid spacing values were set at *ca.* 0.2 nm for CV1, CV2 and CV8, and 0.09 rad (*ca.* 5°) for CV9, respectively. All relevant microstates (*i.e.*, sufficiently populated) appeared correct structurally, *i.e.*, correct distribution in the Ramachandran plots, no significant helix distortion, or no specific changes in the secondary structure. This was another indication that although description of the canonical ensemble describing the IF ↔ OF process was not completed, sampling around the OF, IF-c and IF-o states was correct, *i.e.*, providing correct intermediates. All the microstates were ranked according to the evolution of the CV values. Following a given replica with a given geometry (*i.e.*, OF, IF-c or IF-o), no jump in CV values was observed between the three geometries (Figure 5B). However, the exploration was sufficiently extended to envisage inter-connection between the three different states (L-shape distributions in Figure 5). This precludes a discussion about the energy barriers required to be crossed in the IF ↔ OF process. Nonetheless the conformational path leading from one to another state is worth discussing from these preliminary data to rationalize the complexity of the IF ↔ OF process in the Sav1866 ABC transporter.

*A structural view of the IF ↔ OF process*

There exists no experimental technique capable of providing a complete view of the conformational evolution of membrane transporters during drug translocation. Even though cryo-EM has opened fantastic perspectives in this respect, yet this technique can only capture only a few intermediate states and for a limited number of transporters, *e.g.*, for Sav1866, one of the oldest elucidated 3D ABC structure, only one state has been elucidated so far. Biased MD simulations currently appear as the only possibility to tackle the dynamic of the IF ↔ OF process. The BE-MTD simulations have allowed constructing a coherent path following mainly four independent CVs describing slow events. As seen in Figure 8A, CV1, 2 and 8 are particularly adapted to capture the structural evolution from the IF-o to the OF conformer. Namely, CV8 (distance between NBDs) first decreases from *ca*. 5.1 nm to *ca*. 2.6 nm. Then, the system reaches the complex region of the phase space in which all domains are in close contact and which has been described as IF-c conformer. Here we believe form our simulation that this region is a cluster of conformers smoothly evolving from the IF-c conformers to OF-c (OF-close) conformers, from which extra-cellular opening is likely to occur for drug release. The OF-c → OF-o (OF-open) transition constitutes the final stage, depicted in Figure 8A by a regular increase of CV1 (top distance between transmembrane bundles A and B) from *ca*. 1.9 nm to *ca*. 5.2 nm. Figure 8B and C shows that CV9 is a highly fluctuating CV. The value around -130° is possible in IF-o conformers. The two NBDs are likely to twist freely in this conformer until the two NBD are in close contact all along the rest of path (*i.e.*, IF-c, OF and other intermediate states). At this stage, although CV9 still fluctuates, the fluctuation range is much narrower. It must stress that in our study, ATP is absent from the binding site, it means that the head-to-tail lock is only ensured by the supramolecular conformational arrangement. In other words, the lock between the two NBDs is not optimum from an enthalpy (intermolecular interaction) point of view. Therefore, the fluctuation of CV9 is likely to be even more reduced in the presence of ATP. This process agrees with the $(\beta, \gamma) \rightarrow \alpha$ process previously described for the OF → IF process occurring in the apo-MsbA transporter, where $\alpha, \beta, \gamma$ are the angle between transmembrane bundles, the angle between NBDs, and the twist between NBDs, respectively, which are related to CV8, CV1 and CV9, respectively [15].

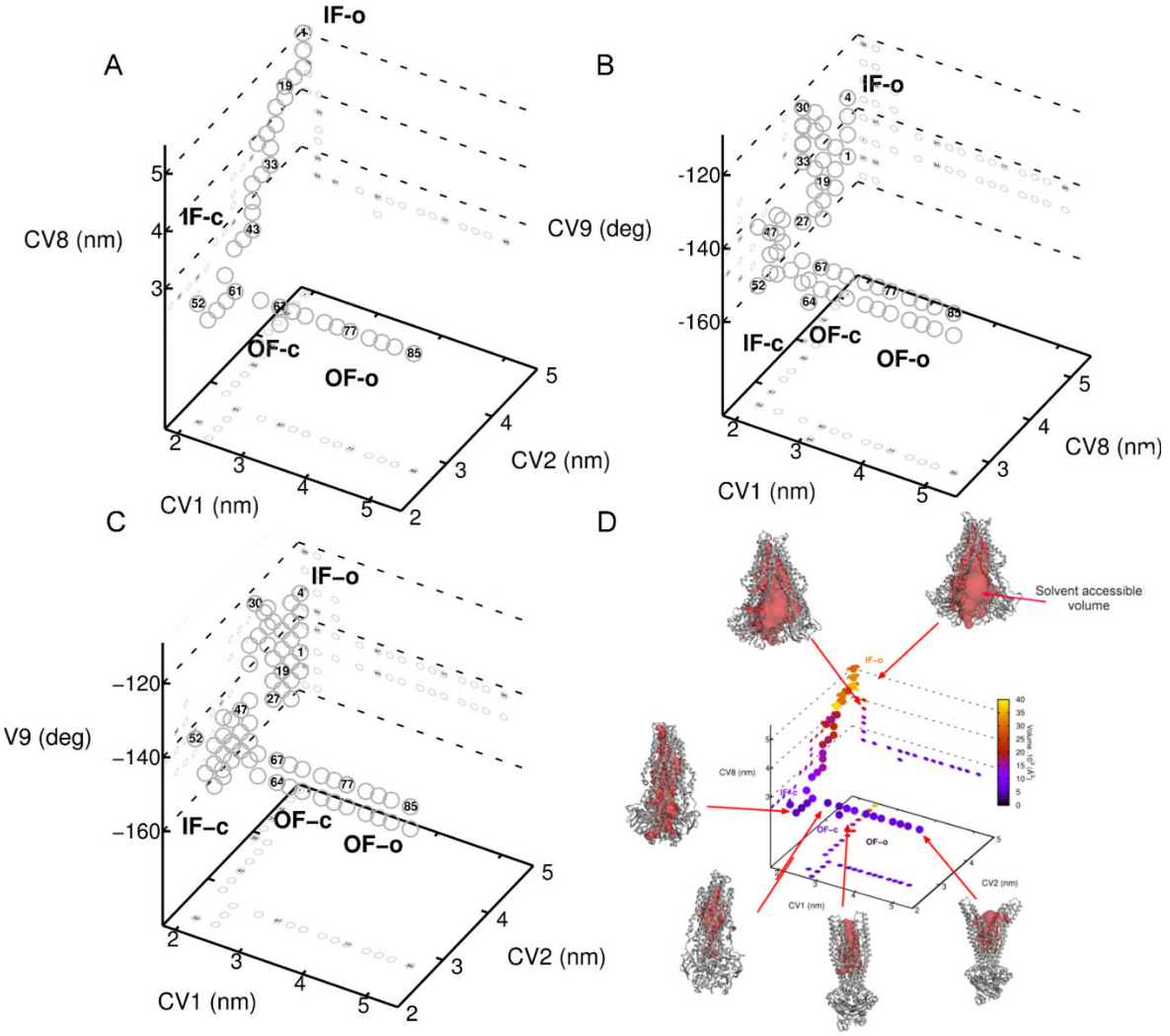



Figure 8: Evolution of A) CV1, CV2, CV8, B) CV1, CV8, CV9, C) CV1, CV2, CV9 along the IF ↔ OF path. D) evolution of CV1, CV2, CV8 along the evolution of the chamber volume.

From all data, a chronology of events was constructed following various structural parameters, which are characteristic of the IF ↔ OF process. The protein chamber is key in xenobiotic translocation through ABC. It is a dynamic channel in the sense that it is never open from one to another face of the membrane at the same time (Figure 9A). As expected, the IF-o conformer exhibits a widely IF opened chamber allowing access of a broad range of xenobiotics (microstates 1-39). Substrate binding is then mainly driven by specific non-covalent interactions at different sites, which are ABC transporter dependent [9,24,42–44]. The system can fluctuate until the two NBDs are in close contact, in the first IF-c microstates (microstates 40-50). At this stage, the chamber volume starts decreasing (Figure 8D) while the TMDs highly re-organize to create a volume which can encapsulate the xenobiotic (microstates 51-60). These microstates constitute a set of walnut-like shape conformers (here simply named walnut), which are intermediates between IF-c and OF-c, and which are neither IF, nor OF. From this stage, the embedded volume is shifted towards the outer compartment within the formation of OF-c intermediates (microstates 61-85). Eventually, the OF-o state is created, opening a volume corresponding to the xenobiotic release. As expected for ABC transporters, Sav1866 is never open simultaneously towards the inner and outer compartments. This scheme fully agrees with the vectorial transport of the xenobiotics occurring in ABC transporters.

The chronology of events and chamber volume arrangement can also be followed with a focus given on TMDs (Figure 9B). In IF-o conformer, bundles A and C from one side and bundles B and D from the other side are in close contact likely due to attractive electrostatic interactions. The two pairs of bundles form a V-shape allowing a widely open chamber towards the inner compartment. In the IF-c conformer, the four transmembrane bundles are packed. This region of the phase space is key and it is the crossroad where intense conformational fluctuations occur. It is also a region of the phase space in which there exists the highest energetic barriers, which was not overcome through this 4.5 µs BE-MTD simulations (Figure 5B). In OF-c, the bundles initiate the greatest large-scale conformational changes, *i.e.*, swapping between bundles C and B from one side and bundles D and A from the other side. Once this swapping is completed the ABC transporter reopens towards the outer compartment. This evolution supports the alternating-access mechanism proposed for drug translocation through ABC transporters [11,45].

Also, following the chronology of events, the structural behavior of the two C-terminal helices requires special attention (Figure 9C). They are known to adopt a highly specific organization, *i.e.*, parallel to each other, in OF conformers [46]. The domain has been described as a zipper-like domain, which impacts on the activity of the heterodimeric *Thermus thermophilus* multidrug resistance proteins A and B (TmrAB); mutation or deletion of these zipper helices was shown to decrease the transporter activity [46]. A zipper-like role was suggested to facilitate the large conformational changes in the IF ↔ OF process. The two helices were suggested to weakly interact with each other in the IF-o state stabilizing the conformer and limiting opening of the transporter towards the inner compartment. Unzipping and twisting the helices to zip them again in the OF conformer was then suggested. Here, by constructing a IF ↔ OF path for Sav1866, we observed a similar mechanism, which strongly agrees with this scheme (Figure 7C of [46]). Namely, in the IF-o microstates, the two helices are relatively weakly non-covalently linked. Although weak, the link exists through contacts between residues His571, Leu572, Tyr573, Gln576, Asn577, Leu578 and Gln1144, Glu1148, Tyr1151, Asn1155, Leu1156. The inter-helix distance is however relatively prone to fluctuate (Figure 9C). The two helices line up as soon as the system reaches the IF-c state and it only slightly fluctuates all along the further events (Figure 9C). This path fully confirms the key role played by the zipper helices in the large-conformation modifications which occur during the xenobiotic translocation. This zipper-like role of the C-terminal helices is likely to occur in many other ABC transporters, however, as mentioned in Nöll *et al.*[46], with an impact that still need to be elucidated from a global point of view within the different ABC families. It should systematically be considered with care, and with a dynamic perspective, when studying structural features of ABC transporters to tackle the active transport of xenobiotics.

As expected, the two NBDs approach in a head-to-tail way with great flexibility (Figure 9D). When reaching the IF-c microstates, the two NBDs are embedded with each other in a very specific way, *i.e.*, the Walker A and A-loop domains of one NBD facing the Signature domain of the other (Figure 9D). It must be stressed that the all simulations were performed in the absence of ATP or ADP. Upon ATP/ADP binding, the A-loop, adjacent to the N-terminus of the NBDs, containing an aromatic residue (*i.e.*, Trp, Tyr, or Phe), allows attractive $\pi - \pi$ stacking interactions with the adenosine moiety of ATP [47]. Walker A has been shown to be in contact with the α and β-phosphates of ATP, while ABC signature and H-loop are in contact with γ-phosphate ATP moiety [48–51]. It is important to note that NBD dimerization is mandatory to properly bind ATP molecule since it is sandwiched between one Walker A of one and the ABC

signature of the other NBD [11,52]. In turn, ATP binding is described to stabilize this dimerization. Walker B is of particular importance since it provides the Glu residue that orients water molecule in the catalytic site and thus allow ATP-hydrolysis by nucleophilic attack. The path obtained by metadynamics shows that although ATP binding is mandatory it is not the only reason for the head-to-tail arrangement of the two NBDs in a very specific orientation[5]. In other words, when correctly oriented, the interface between the two NBDs allow attractive electrostatic and van der Waals interactions, which favor the lock of the two NBDs, then to allow all conformational changes responsible for the xenobiotic translocation.

Regarding NBD dimerization, two key steps have been highlighted in this chronology of events. The first one appears when moving from IF-o to IF-c; as expected in this step there are large directional movement to lead the two NBD in close contact (see PCA - principal component analysis - in the last row in Figure 9E). For IF-c to the walnut-conformers, no such directional movements were observed, as from OF-c to OF-o. However, from the walnut-conformers to OF-c, large directional movements were observed, which reflects the twist between NBDs. This step appears crucial to trigger the xenobiotic release.

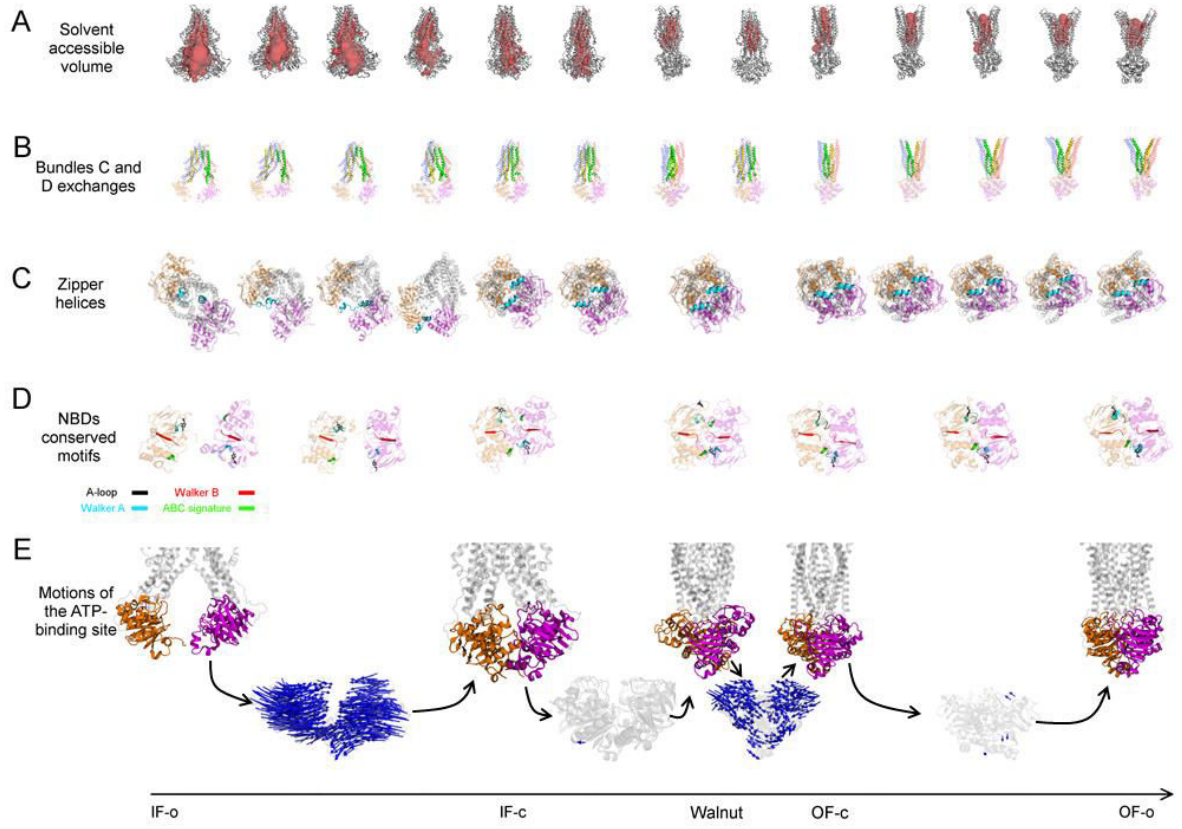



Figure 9: Chronology of events of the IF ↔ OF path re-constructed from the BE-MTD simulations, following: A) the chamber volume; B) bundle swapping; C) zipper-like C-terminal helices; D) head-to-tail NBD lock; E) NBD dimerization and corresponding PCA analysis between key intermediates (*i.e.*, IF-o → IF-c, IFc → walnut, walnut → OF-c, and OF-c → OF-o).

[5]Although biased for some CVs, the simulations have been performed to allow all possible conformational relaxation, in particular in this region. If the previous statement was wrong, the IF-c → OF-o microstates would have shown much more distortion from this NBD packing.

### IV.3.3. Conclusion

Although the recent developments in cryo-EM have allowed significant progress in visualization of key intermediate states, the investment is still a limiting step and these techniques only provide a few intermediate ABC conformers. Biased MD simulation is the only way to tackle the IF ↔ OF path, of a given ABC transporter, in its entirety. Metadynamics is a promising alternative to escape from the "brute-force" techniques, which in principle allows a more natural exploration of the phase space, *i.e.*, without pushing the system into unrealistic conformers. Nonetheless, metadynamics has still been preferred for smaller systems than membrane ABC transporters. By following nine CVs, then focusing on four uncorrelated and key CVs, we have studied the IF ↔ OF path of the prototypical, Sav1866 ABC transporter. Although 4.5 µs simulations were performed, the system appeared not sufficiently converged to fully trust the exact path in terms of free energy. It appeared that the phase space was not fully filled by Gaussian functions in the region where the energy barriers were the highest, *i.e.*, from the IF-o to the IF-c basins and from the IF-c to the OF-o basins. The sampling was most probably insufficient, which could be due to: i) insufficient computational time; ii) too many key geometrical variables to be considered; iii) or sub-optimum metadynamics parametrization. The height of the Gaussians could not be increased further to accelerate the exploration as it was already set at 4-16 kJ/mol, which is expected to be *ca*. tenth of the highest barriers of the IF ↔ OF process. This also stands true for the deposition rate which was set at 100 ps and which can hardly be increased. Among all reasons responsible for not reaching the complete exploration of the phase space likely lies in the choice of the set of key CVs. If only one CV was sub-optimally defined or was missing, the system may be stuck in a region of the phase space from which it can hardly escape. From our data, a final revisit of this set is recommended to reach the optimum metadynamics parameterization. However, it must be stressed that regarding the size of the system, any benchmarking is likely to take months of simulation, therefore a step-by-step strategy is still mandatory.

Although the energy landscape was not elucidated from this study, at this stage of the simulations, a relevant series of connected intermediates, smoothly describing the large-scale conformational changes from IF-o to OF- was obtained. The analysis of this path provides very relevant details in the path, which confirms some hypotheses or mechanisms found and suggested in the literature. A series of key intermediates was highlighted, defining two new regions of the phase space, namely defining OF-c and walnut conformers. In the near future, we believe that biased MD simulations should pave the way toward a better understanding of this key pharmacological event. A dynamic picture of the xenobiotic transport through ABCs is indeed mandatory and the time has come to go beyond the static or semi-static (fluctuation around one energetic minimum) models, which stands particularly true for drug membrane transporters.

## References


[1] International Transporter Consortium, K.M. Giacomini, S.-M. Huang, D.J. Tweedie, L.Z. Benet, K.L.R. Brouwer, X. Chu, A. Dahlin, R. Evers, V. Fischer, K.M. Hillgren, K.A. Hoffmaster, T. Ishikawa, D. Keppler, R.B. Kim, C.A. Lee, M. Niemi, J.W. Polli, Y. Sugiyama, P.W. Swaan, J.A. Ware, S.H. Wright, S.W. Yee, M.J. Zamek-Gliszczynski, L. Zhang, Membrane transporters in drug development, Nat Rev Drug Discov. 9 (2010) 215–236. doi:10.1038/nrd3028.

[2] T. Ebner, N. Ishiguro, M.E. Taub, The Use of Transporter Probe Drug Cocktails for the Assessment of Transporter-Based Drug–Drug Interactions in a Clinical Setting—Proposal of a Four Component Transporter Cocktail, Journal of Pharmaceutical Sciences. 104 (2015) 3220–3228. doi:10.1002/jps.24489.

[3] J. König, F. Müller, M.F. Fromm, Transporters and Drug-Drug Interactions: Important Determinants of Drug Disposition and Effects, Pharmacol Rev. 65 (2013) 944–966. doi:10.1124/pr.113.007518.

[4] K.M. Giacomini, S.-M. Huang, Transporters in drug development and clinical pharmacology, Clin. Pharmacol. Ther. 94 (2013) 3–9. doi:10.1038/clpt.2013.86.

[5] Y.H. Choi, A.-M. Yu, ABC transporters in multidrug resistance and pharmacokinetics, and strategies for drug development, Curr. Pharm. Des. 20 (2014) 793–807. doi:10.2174/13816128200514021465212.

[6] J.A. DiMasi, L. Feldman, A. Seckler, A. Wilson, Trends in risks associated with new drug development: success rates for investigational drugs, Clin. Pharmacol. Ther. 87 (2010) 272–277. doi:10.1038/clpt.2009.295.

[7] M.J. Zamek-Gliszczynski, K.A. Hoffmaster, D.J. Tweedie, K.M. Giacomini, K.M. Hillgren, Highlights from the International Transporter Consortium second workshop, Clin. Pharmacol. Ther. 92 (2012) 553–556. doi:10.1038/clpt.2012.126.

[8] K.M. Hillgren, D. Keppler, A.A. Zur, K.M. Giacomini, B. Stieger, C.E. Cass, L. Zhang, International Transporter Consortium, Emerging transporters of clinical importance: an update from the International Transporter Consortium, Clin. Pharmacol. Ther. 94 (2013) 52–63. doi:10.1038/clpt.2013.74.

[9] A.A.K. El-Sheikh, J.J.M.W. van den Heuvel, E. Krieger, F.G.M. Russel, J.B. Koenderink, Functional Role of Arginine 375 in Transmembrane Helix 6 of Multidrug Resistance Protein 4 (MRP4/ABCC4), Mol Pharmacol. 74 (2008) 964–971. doi:10.1124/mol.107.043661.

[10] F.G.M. Russel, J.B. Koenderink, R. Masereeuw, Multidrug resistance protein 4 (MRP4/ABCC4): a versatile efflux transporter for drugs and signalling molecules, Trends Pharmacol. Sci. 29 (2008) 200–207. doi:10.1016/j.tips.2008.01.006.

[11] B. Verhalen, R. Dastvan, S. Thangapandian, Y. Peskova, H.A. Koteiche, R.K. Nakamoto, E. Tajkhorshid, H.S. Mchaourab, Energy transduction and alternating access of the mammalian ABC transporter P-glycoprotein, Nature. 543 (2017) 738–741. doi:10.1038/nature21414.

[12] G.A. Frank, S. Shukla, P. Rao, M.J. Borgnia, A. Bartesaghi, A. Merk, A. Mobin, L. Esser, L.A. Earl, M.M. Gottesman, D. Xia, S.V. Ambudkar, S. Subramaniam, Cryo-EM Analysis

of the Conformational Landscape of Human P-glycoprotein (ABCB1) During its Catalytic Cycle, Mol Pharmacol. 90 (2016) 35–41. doi:10.1124/mol.116.104190.

[13] Z.L. Johnson, J. Chen, ATP Binding Enables Substrate Release from Multidrug Resistance Protein 1, Cell. 172 (2018) 81-89.e10. doi:10.1016/j.cell.2017.12.005.

[14] P.-C. Wen, E. Tajkhorshid, Conformational Coupling of the Nucleotide-Binding and the Transmembrane Domains in ABC Transporters, Biophys J. 101 (2011) 680–690. doi:10.1016/j.bpj.2011.06.031.

[15] M. Moradi, E. Tajkhorshid, Mechanistic picture for conformational transition of a membrane transporter at atomic resolution, PNAS. 110 (2013) 18916–18921. doi:10.1073/pnas.1313202110.

[16] A. Laio, F.L. Gervasio, Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science, Rep. Prog. Phys. 71 (2008) 126601. doi:10.1088/0034-4885/71/12/126601.

[17] S. Piana, A. Laio, A Bias-Exchange Approach to Protein Folding, J. Phys. Chem. B. 111 (2007) 4553–4559. doi:10.1021/jp067873l.

[18] F. Marinelli, F. Pietrucci, A. Laio, S. Piana, A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations, PLOS Computational Biology. 5 (2009) e1000452. doi:10.1371/journal.pcbi.1000452.

[19] R.J.P. Dawson, K.P. Locher, Structure of a bacterial multidrug ABC transporter, Nature. 443 (2006) 180–185. doi:10.1038/nature05155.

[20] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, Nat Protoc. 5 (2010) 725–738. doi:10.1038/nprot.2010.5.

[21] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER Suite: protein structure and function prediction, Nat Meth. 12 (2015) 7–8. doi:10.1038/nmeth.3213.

[22] M. Hohl, C. Briand, M.G. Grütter, M.A. Seeger, Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation, Nat Struct Mol Biol. 19 (2012) 395–402. doi:10.1038/nsmb.2267.

[23] M. Haffke, A. Menzel, Y. Carius, D. Jahn, D.W. Heinz, Structures of the nucleotide-binding domain of the human ABCB6 transporter and its complexes with nucleotides, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 979–987. doi:10.1107/S0907444910028593.

[24] M.S. Jin, M.L. Oldham, Q. Zhang, J. Chen, Crystal structure of the multidrug transporter P-glycoprotein from C. elegans, Nature. 490 (2012) 566–569. doi:10.1038/nature11448.

[25] P. Szewczyk, H. Tao, A.P. McGrath, M. Villaluz, S.D. Rees, S.C. Lee, R. Doshi, I.L. Urbatsch, Q. Zhang, G. Chang, Snapshots of ligand entry, malleable binding and induced helical movement in P-glycoprotein, Acta Crystallogr. D Biol. Crystallogr. 71 (2015) 732–741. doi:10.1107/S1399004715000978.

[26] Y. Zhang, J. Skolnick, TM-align: a protein structure alignment algorithm based on the TM-score, Nucleic Acids Res. 33 (2005) 2302–2309. doi:10.1093/nar/gki524.

[27] S.C. Lovell, I.W. Davis, W.B. Arendall, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson, Structure validation by Calpha geometry: phi,psi and Cbeta deviation, Proteins. 50 (2003) 437–450. doi:10.1002/prot.10286.

[28] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Res. 35 (2007) W407-410. doi:10.1093/nar/gkm290.

[29] M.G. Wolf, M. Hoefling, C. Aponte-Santamaría, H. Grubmüller, G. Groenhof, g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation, J Comput Chem. 31 (2010) 2169–2174. doi:10.1002/jcc.21507.

[30] N. Schmid, A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, W.F. van Gunsteren, Definition and testing of the GROMOS force-field versions 54A7 and 54B7, Eur. Biophys. J. 40 (2011) 843–856. doi:10.1007/s00249-011-0700-9.

[31] H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, The missing term in effective pair potentials, J. Phys. Chem. 91 (1987) 6269–6271. doi:10.1021/j100308a038.

[32] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, The Journal of Chemical Physics. 98 (1993) 10089–10092. doi:10.1063/1.464397.

[33] S. Nosé, A molecular dynamics method for simulations in the canonical ensemble, Molecular Physics. 52 (1984) 255–268. doi:10.1080/00268978400101201.

[34] W.G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, Phys. Rev. A. 31 (1985) 1695–1697. doi:10.1103/PhysRevA.31.1695. doi:10.1103/PhysRevA.31.1695.

[35] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, Journal of Applied Physics. 52 (1981) 7182–7190. doi:10.1063/1.328693.

[36] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: fast, flexible, and free, J Comput Chem. 26 (2005) 1701–1718. doi:10.1002/jcc.20291.

[37] G. Bussi, D. Branduardi, Free-Energy Calculations with Metadynamics: Theory and Practice, in: Reviews in Computational Chemistry Volume 28, Wiley-Blackwell, 2015: pp. 1–49. doi:10.1002/9781118889886.ch1.

[38] D. Branduardi, G. Bussi, M. Parrinello, Metadynamics with Adaptive Gaussians, J. Chem. Theory Comput. 8 (2012) 2247–2254. doi:10.1021/ct3002464.

[39] F. Marinelli, Following Easy Slope Paths on a Free Energy Landscape: The Case Study of the Trp-Cage Folding Mechanism, Biophysical Journal. 105 (2013) 1236–1247. doi:10.1016/j.bpj.2013.07.046.

[40] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R.A. Broglia, M. Parrinello, PLUMED: a portable plugin for free-energy calculations with molecular dynamics, Computer Physics Communications. 180 (2009) 1961–1972. doi:10.1016/j.cpc.2009.05.011.

[41] T. Giorgino, A. Laio, A. Rodriguez, METAGUI 3: A graphical user interface for choosing the collective variables in molecular dynamics simulations, Computer Physics Communications. 217 (2017) 204–209. doi:10.1016/j.cpc.2017.04.009.

[42] S. Wilkens, Structure and mechanism of ABC transporters, F1000Prime Rep 2015. 7 (2015). doi:10.12703/P7-14.

[43] T.W. Loo, M.C. Bartlett, D.M. Clarke, Simultaneous binding of two different drugs in the binding pocket of the human multidrug resistance P-glycoprotein, J. Biol. Chem. 278 (2003) 39706–39710. doi:10.1074/jbc.M308559200.

[44] K. Beis, Structural basis for the mechanism of ABC transporters, Biochem. Soc. Trans. 43 (2015) 889–893. doi:10.1042/BST20150047.

[45] A. Ward, C.L. Reyes, J. Yu, C.B. Roth, G. Chang, Flexibility in the ABC transporter MsbA: Alternating access with a twist, PNAS. 104 (2007) 19005–19010. doi:10.1073/pnas.0709388104.

[46] A. Nöll, C. Thomas, V. Herbring, T. Zollmann, K. Barth, A.R. Mehdipour, T.M. Tomasiak, S. Brüchert, B. Joseph, R. Abele, V. Oliéric, M. Wang, K. Diederichs, G. Hummer, R.M. Stroud, K.M. Pos, R. Tampé, Crystal structure and mechanistic basis of a functional homolog of the antigen transporter TAP, Proc. Natl. Acad. Sci. U.S.A. 114 (2017) E438–E447. doi:10.1073/pnas.1620009114.

[47] S.V. Ambudkar, I.-W. Kim, D. Xia, Z.E. Sauna, The A-loop, a novel conserved aromatic acid subdomain upstream of the Walker A motif in ABC transporters, is critical for ATP binding, FEBS Lett. 580 (2006) 1049–1055. doi:10.1016/j.febslet.2005.12.051.

[48] E. Procko, I. Ferrin-O'Connell, S.-L. Ng, R. Gaudet, Distinct structural and functional properties of the ATPase sites in an asymmetric ABC transporter, Mol. Cell. 24 (2006) 51–62. doi:10.1016/j.molcel.2006.07.034.

[49] P.C. Smith, N. Karpowich, L. Millen, J.E. Moody, J. Rosen, P.J. Thomas, J.F. Hunt, ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer, Mol. Cell. 10 (2002) 139–149. doi:10.1016/S1097-2765(02)00576-2.

[50] K.P. Hopfner, A. Karcher, D.S. Shin, L. Craig, L.M. Arthur, J.P. Carney, J.A. Tainer, Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily, Cell. 101 (2000) 789–800. doi:10.1016/S0092-8674(00)80890-9.

[51] J.E. Walker, M. Saraste, M.J. Runswick, N.J. Gay, Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold, EMBO J. 1 (1982) 945–951. doi:10.1002/j.1460-2075.1982.tb01276.x.

[52] I.B. Holland, M. A. Blight, ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans, Journal of Molecular Biology. 293 (1999) 381–399. doi:10.1006/jmbi.1999.2993.

## Conclusion

Interactions between xenobiotics and lipid bilayers are crucial to rationalize their activity, toxicity and bioavailability. Drug crossing events are essential for most xenobiotics to reach their biological targets. In some cases, the behavior of some therapeutic compounds in lipid bilayers can be directly or indirectly related to their mechanism of action. The molecular rationalization of xenobiotic-membrane interactions may support the design of more available and less toxic drugs. Also, structure-activity relationship can be established. The quality of the *in silico* lipid bilayer models as well as their increasing diversity appear crucial. Lipid bilayer mixtures, especially including cholesterol significantly affect membrane dynamics and modulate the interaction of xenobiotics with membranes. The interactions of xenobiotics with membrane proteins are also a crucial issue related to drug influx or efflux. This understanding can support rationalization of the impact of polymorphisms on xenobiotics transport through membrane proteins. Despite experimental advances in the resolution of 3D structures of these proteins, the knowledge about human ABC transporter functioning is still challenging, mainly regarding their transport cycle, their interactions with lipid bilayers, the selectivity of their substrates and the impact of physiological and clinical conditions. MD simulations and biased methods have appeared interesting alternative to study these issues.

Supported by MD simulations, Chapter III of this PhD work described the positioning of various xenobiotics in membranes, *i.e.*, antioxidants or a potential anticancer drug, related either to their mechanisms of action or to their transport. We also highlighted the importance of cholesterol in biological membranes. Chapter IV focused on membrane proteins, for which various domains (*e.g.*, $L_0$, $L_1$ linkers, zipper helices) and specific residues are shed light by membrane protein model construction. A particular attention was paid to the significant structural impact of the substitution of a single residue as well as the importance of specific regions and residues for substrate binding (Sections IV.1 and 2). Proteins can be pictured as tightrope walkers for which all domains, all residues and molecular arrangements are key for their functioning in the organism. The investigation of the transport cycle of ABC transport has shown us the cooperativity and the complexity of the large-scale conformational changes (Section IV.3). The transmembrane bundles, NBDs, zipper helices highly cooperate to allow the pump motion responsible for transport cycle through ABCs.

From this PhD work arise few points that are of crucial importance for modeling xenobiotic insertion. A particular attention should be paid to the choice of force fields. Lipid force field must be wisely chosen accounting the compatibility with xenobiotic force field. It is nowadays well-admitted that force fields must be picked within a given family (*e.g.,* AMBER or CHARMM). It is worth mentioning that xenobiotic force field must be carefully parametrize. A particular attention should be paid to non-covalent terms (*e.g.*, atomic partial charges) since they are likely to be driving force of xenobiotic insertion.

Likewise, several validation steps are required to ensure the robustness of membrane protein models obtained by either homology modeling or protein threading. A careful comparison with homologs, similar proteins is mandatory as well as with indirect structural clues obtained at the experimental level (*e.g.*, function loss owing to site-directed mutagenesis). Prior to running MD simulations, protonation states of titrable residues must be carefully consider accounting the impact of lipid bilayer membrane electrostatic environment. However, it is well admitted in the community that it is almost impossible to model the “exact” *in situ* protonation state. In biased MD simulations, Collective Variables must to be wisely defined for cost-effective and accurate

description of the phenomenon of interest. In metadynamics, the choice of the Gaussian function parameters is key to correctly fill potential energy wells in an efficient manner.

A further step for the xenobiotics and membrane interaction is to simulate even more complex mixtures of lipid bilayers *e.g.*, with different charged lipids to get even closer to the biological and experimental conditions. Regarding the human ABCC4/MRP4 model, the following crucial methodological step is to refine the model to help at deciphering the role and position of both $L_1$ and $L_0$ linkers. Besides, from our metadynamic MD simulations, ABC transporter transport cycle can be further investigated not only by refining the choice of the Collective Variables, but also by assessing conformational changes upon ATP/ADP and substrate binding to NBDs and protein chamber, respectively.

Over the past few years, composition of the lipid bilayer has gained strong interest in the field of membrane protein. However, the role of lipid composition to large-scale conformational transitions remains unclear and should be investigated as a dedicated Collective Variable. Additionally to the refinement of the biased MD procedure, the investigation of other ABC transporters could help decipher transporter family function relationship. The construction of additional unresolved human ABC transporters (*e.g.,* MRP2) could help decipher the clinical and pharmacological impact of these transporters.

The ultimate goal of molecular pharmacology is to provide thorough pictures of the interactions between xenobiotics, lipid bilayers and membrane proteins, as well as to predict the impact of xenobiotics and polymorphisms on active transport. The aim is to go towards an *in silico* personalized predictive tool. This PhD work and the increasing literature in this research field allow us to believe that *in silico* methods will decipher clinical issues related to membrane protein transporters with an atomic scale perspective.

***In silico* interakcí xenobiotik s lipidovými dvojvrstvami a membránovými ABC transportery, na příkladu ABCC4/MRP4**

K porozumění biologických mechanismů chování membránových proteinů je třeba hlubšího pochopení interakcí xenobiotik a těchto proteinů s lipidovými membránami. Experimentální metody jsou často metodologicky náročné a jen velmi přibližně popisují interakce proteinů s membránami. *In silico* molekulární modelování poskytujealternativu k uchopení těchto problémů. Molekulární dynamika (MD) a dynamické simulace s externím potenciálem otevírají nové možnosti k atomistickému popisu těchto intermolekulárních interakcí. S využitím MD simulací jsme vytvořili model lidského ABC ABC4/MRP4 transportéru. Studovali jsme také vliv cholesterolu na chování tohoto proteinu a dopad polymorfizmu zodpovědného za ztrátu transportních schopností proteinu. Pro hlubší porozumění chování proteinu jsme dále studovali interakce xenobiotik s modelem lidského transportéru spolu s transportním cyklem ABC transportéru.
Interakce xenobiotik a membránových lipidů byly studovány z hlediska jejich schopnosti inkorporovat se do membránových struktur. U směsných membrán s obsahem cholesterolu byl prokázán značný dopad jak na chování lidského ABC4/MTP4, tak xenobiotik. Byla zde demonstrována důležitost jednotlivých regionů, domén, jakož i jednotlivých residuí připraveného modelu ABC4/MRP4. Dále jsme pozorovali strukturní změny, ke kterým docházelo během tranzitních stavů transportního cyklu.

Klíčová slova: Molekulární dynamika, lipidová dvojvrstva, membrány, xenobiotika, metadynamika

***Etude in silico des interactions des xénobiotiques avec les bicouches lipidiques et les transporteurs membranaires ABC, le cas d'ABCC4/MRP4***

L'appréhension des mécanismes d'action biologiques des protéines membranaires nécessite de comprendre les interactions des xénobiotiques avec ces protéines et avec les membranes lipidiques. Les méthodes expérimentales sont parfois coûteuses et ne permettent d'obtenir que des informations partielles sur les interactions xénobiotiques-membrane-protéine. La modélisation moléculaire est une sérieuse alternative. Les simulations de dynamique moléculaire et de dynamique biaisées ont ouvert de nombreuses perspectives en permettant de décrire ces interactions moléculaires à l'échelle atomique. Grâce à des simulations de dynamique moléculaire, nous avons été capables de construire un modèle de transporteur humain ABC : ABCC4/MRP4. Cette protéine a été choisie pour sa présence dans le rein, notamment, et son importance clinique. Nous avons évalué l'influence du cholestérol sur cette protéine. L'étude de domaines spécifiques et l'impact d'un polymorphisme a été reliée à l'activité de transport de cette protéine. Nous avons également étudié l'interaction de xénobiotiques avec ce transporteur humain. Le cycle de transport des transporteurs ABC a été examiné afin de comprendre leur fonctionnement. L'incorporation de cholestérol a montré un impact significatif sur la protéine humaine ABCC4/MRP4 et sur les xénobiotiques étudiés. L'importance de domaines constituant la protéine ABCC4/MRP4 ainsi que l'importance de résidus individuels a clairement été prouvée. Nous avons également pu observer des intermédiaires du cycle de transport d'un transporteur ABC conjointement avec des changements structuraux.

Mots-clés: dynamique moléculaire, transporteurs ABC, xénobiotiques, membranes lipidiques, métadynamique

***In silico* investigation of xenobiotic interactions with lipid bilayers and ABC membrane transporters, the case of ABCC4/MRP4;**

Understanding the biological mechanisms of action of membrane proteins requires the comprehension of the interactions of xenobiotics with these proteins and with lipid membranes. Experimental methods are often demanding and only partially respond to xenobiotic-membrane-protein interactions. *In silico* molecular modeling is a serious alternative to tackle these issues. Molecular dynamics (MD) and biased dynamics simulations have opened many perspectives by providing an atomistic description of these intermolecular interactions. Using MD simulations, we built a model of the human ABC ABCC4/MRP4 transporter. We explored the influence of cholesterol on this protein as well as the impact of a polymorphism known to shut down the transport activity of this protein. We also studied the interaction of xenobiotics with this human transporter. The transport cycle of the ABC transporters was investigated in an attempt to better understand how it works.
Interactions between lipid membranes and xenobiotics were explored by examining their ability to incorporate lipid membranes. Lipid mixtures with cholesterol showed a significant impact on the human protein ABCC4/MRP4 and on the xenobiotics studied. The importance of regions, domains constituting the ABCC4/MRP4 protein as well as the importance of specific residues has been clearly demonstrated. We also observed intermediates in the transport cycle of an ABC transporter in conjunction with structural changes occurring during this cycle.

Keywords: molecular dynamics, lipid bilayer membranes, xenobiotics, metadynamics