Ph.D. Thesis

Université de Limoges

University of Limoges

ED 615 - Sciences Biologiques et Santé (SBS)

Pharmacology & Transplantation Inserm U1248

A thesis submitted to University of Limoges
in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

Computational Chemistry

Presented and defenced by

Angelika Janaszkiewicz

On December 9, 2022

In silico Models of Pharmacologically Relevant Membrane Transporters: Focus on the Major Facilitator Superfamily Fold

Thesis supervisor: Florent Di Meo, Ph.D., HDR

JURY:

President of jury
Prof. Marie Essig, Ph.D. HDR, Full Prof. CESP, Inserm U1018, Université Paris-Saclay, France

Reporters
Dr. Lucie Delemotte, Ph.D. Associate Prof., KTH Royal Institute of Technology
Prof. Catherine Etchebest, Ph.D. HDR, FullProf., Inserm U1134, DSIMB, Université Paris Diderot, Sorbonne Paris Cité

Examiners
Prof. Marie Essig, Ph.D. HDR, Full Prof. CESP, Inserm U1018, Université Paris-Saclay, France
Dr. Florent Di Meo, Ph.D. HDR, Inserm U1248, P&T, Ω-Health Institute, Université de Limoges, France

Z dedykacją dla babci, u której zawsze odnajdywałam wsparcie.

“I am among those who think that science has great beauty.”

Maria Skłodowska-Curie

Acknowledgements

This doctoral dissertation would not have been possible without the support of many people.

I am deeply grateful to my mentor, **Florent Di Meo**, for giving me the opportunity to learn under his supervision. His expertise, support, understanding, and kindness were beyond my expectations. I am grateful for many insightful discussions, advice, and the wonderful work and encouragement that enabled this work to be completed. I feel privileged to be your student.

I would like to thank the thesis committee for their time and effort they put into reviewing this manuscript. I am pleased that **Marie Essig** has agreed be the committee's president. She is a brilliant nephrology researcher who has inspired and advise the projects presenting the clinical perspective. **Cathrine Etchebest's** work has been an inspiration to me, so I am overjoyed to have her on the committee. I am delighted to have **Lucie Delemotte** on the committee. During my internship in Lucie Delemotte's laboratory, I had the pleasure of meeting her in person, witnessing her brilliant mind and experiencing her kindness. She has been a role model for me, and her work has influenced my research. I am beyond grateful that **Florent Di Meo**, my mentor, supervisor, and scientific inspiration, is a member of the committee. I feel privileged to be able to share my research with such esteemed scientists. Thank you for your time and effort.

I would like to thank the Pharmacology & Transplantation unit members, especially **Pierre Marquet**, for guidance and participation in my journey.

The computational group, Pharmacology & Transplantation, Limoges, 2022

I have had the honour of working with incredible members of the computational group. I am grateful to my colleagues, **Ágota Tóth, Veronica Crespi, Mehdi Benmameri, Benjamin Chantemargue,** and **Gabin Fabre**, who offered great support. The exchange of ideas boosted the passion and scientific curiosity.

Special thanks to **Ágota Tóth**, with whom I had the privilege of working and developing a beautiful friendship that brought an amazing value to this journey.

I am grateful to **Dominik Gront**, my friend and scientific advisor, for his time, many excellent discussions, and words of wisdom.

I am incredibly grateful to my love, **David**, that had been for me in the hardest and happiest moments. You have showed me kindness, warmth and support that made it all easier. You have been consoling and encouraging me to celebrate little victories. Thank you for standing by my side.

I am grateful to my dear friends for their encouragement and challenges, as well as their unquestioning support. Thanks to **Sandra, Klaudia, Kinga, Ania, Kasia, Nadia, Ada, Agnieszka** and **Clarisse**, who have always been there for me.

Dziękuję mojej rodzinie, mojej mamie **Ani**, moim braciom **Radosławowi** i **Piotrowi**, mojej babci **Helenie**, mojemu wujkowi **Ryszardowi**, cioci **Danucie** i innym którzy z ciekawością i optymizmem przyglądali się mojej podróży wspierając mnie z całych sił.

Many thanks to **Université de Limoges, Agence Nationale de la Recherche** for financial support that enabled this work. Many thanks for the administrative support to **Emmanuelle Izorche, Fanny Escure** and **Karen Poole**. Thanks also to **Xavier Montagutelii** for technical support.

Rights

Table of Contents

List of Figures

List of Tables

Abbreviations

ABC	ATP-Binding Cassette
ACE	Angiotensin-Converting Enzyme
ADME	Absorption, Distribution, Metabolism, and Elimination
AF2	AlphaFold2
AFED	Adiabatic Free Energy Dynamics
AFV	Adefovir
AI	Artificial Intelligence
aKG	a-Ketoglutarate
AMBD	Adaptive Biassing Molecular Dynamics
ANP	Acyclic Nucleoside Phosphonate
ATP	Adenosine Triphosphate
BAR	Bennett's Acceptance Ratio
cAMP	Cyclic Adenosine-3′, 5′-MonoPhosphate
CASP	Critical Assessment of Protein Structure Prediction
CDF	Cidofovir
CFTR	Cystic Fibrosis Transmembrane Conductance
CG	Coarse-graine
Chol	Cholesterol
COX-2	CycloOXygenase 2
Cryo-EM	Cryogenic Electron Microscopy
CV	Collective Variables
DAPC	2,3-diarachidonyl-d-glycero-1-phosphatidylcholine
DAPE	1,2-diarachidonyl-phosphatidylethanolamine
DDIs	Drug-Drug Interactions
DDoPE	1-docosahexaenoyl-2-docosahexaenoyl-phosphatidylethanolamine
DDoPS	1-docosahexaenoyl-2-docosahexaenoyl-phosphatidylserine
DOPC	2,3-dioleoyl-d-glycero-1-phosphatidylcholine
DOPE	2,3-dioleoyl-d-glycero-1-phosphatidylethanolamine
ECL	ExtraCellular Loop
EMA	European Medicine Agency
FDA	Food & Drug Administration
FEP	Free Energy Perturbation
GI	GastroIntestinal
GlpT	Glycerol-3-phosphate Transporter
GLUT	Glucose Transporter
GPACMAN	grapefruit, protease inhibitors, azole antifungals, cimetidine, macrolides, amiodarone, and nor
GPCR	G-Protein Coupled Receptor
GPLs	GlyceroPhosphoLipids
GTP	Guanosine TriphosPhate
H-bond	Hydrogen bond
IC	IntraCellular
ICH	IntraCellular Helix
ICLs	IntraCellular Loops
IF	Inward-Facing
IF^{occ}	Inward-Facing occluded
ITC	International Transporter Consortium
K_M	Binding affinity

LacY	Lactose permease
LeuT	Leucine Transporter
LJ	Lennard-Jones
MATE	Multi-Antimicrobial Extrusion
MBAR	Multistate Bennett Acceptance Ratio
MC	Monte Carlo
MD	Molecular Dynamics
MFS	Major Facilitator Superfamily
ML	Machine Learning
ML-FF	Machine-Learning based Force Fields
MM	Molecular Mechanics
MSA	Multiple sequence alignment
MTX	Methotrexate
NaDC3	Na^+/Dicarboxylate Cotransporter
NBDs	Nucleotide Binding Domains
NKT	New Kidney Transporter
NMR	Nuclear Magnetic Resonance
OATs	Organic Anion Transporters
OCTs	Organic Cation Transporters
OF	Outward-Facing
OF^{occ}	Outward-Facing occluded
OTA	Ochratoxin A
PAH	P-AminoHippurate
PAPA	1-palmitoyl-2-arachidonyl-phosphotidic acid
PAPC	1-palmitoyl-2-arachidonyl-phosphatidylcholine
PAPE	1-palmitoyl-2-arachidonyl-phosphatidylethanolamine
PAPS	1-palmitoyl-2-arachidonyl-phosphatidylserine
PBC	Periodic Boundary Condition
PC	PhosphatidylCholine
PCA	Principal Component Analysis
PCRABS	phenytoin, carbamazepine, rifampin, alcohol, barbiturates, and St. John's Wort
PD	PharmacoDynamics
PDB	Protein Data Bank
PDoPC	3-palmitoyl-2-docosahexaenoyl-d-glycero-1-phosphatidylcholine
PDoPE	1-palmitoyl-2-docosahexaenoyl-phosphatidylethanolamine
PE	PhosphatidylEthanolamine
PGx	Pharmacogenetics
PK	PharmacoKinetics
PK-PD	Pharmacokinetic-pharmacodynamic
PKTD	Proximal Kidney Tubular Dysfunction
PME	Particle Mesh Ewald
POPA	3-palmitoyl-2-oleoyl-d-glycero-1-phosphatidic acid
POPC	3-palmitoyl-2-oleoyl-d-glycero-1-phosphatidylcholine
POPE	3-palmitoyl-2-oleoyl-d-glycero-1-phosphatidylethanolamine
POPS	3-palmitoyl-2-oleoyl-d-glycero-1-phosphatidylserine
PTC	Proximal Tubular Cells
PUFA	polyunsaturated
QM	Quantum Mechanics
RMSD	Root Mean Square Deviation

RSST	Remote Sensing Signalling Theory
SCOP	Structural Classification of Proteins
SLC	SoLute Carrier
SNP	Single Nucleotide Polymorphisms
TAMD	Temperature accelerated molecular dynamics
TI	Thermodynamic Integration
TMH	TransMembrane Helix
TmrAB	Thermus thermophilus
TNF	Tenofovir
UR	Unlocked-Return
XylE	Xylose transporter

Introduction

Unique characteristics within individuals resulting from molecular, physiological, environmental, and psychosocial factors have become the focus of personalised medicine over the past decade. Adapting treatments that account for all these parameters is of particular importance to increase the ratio between benefits and adverse effects. In the context of pharmacology, the individual factors that modulate drug exposure cocktails are known to affect patient response. This is even more important when drug cocktails are considered in severe pathologies since used xenobiotics can also pharmacologically interact with each other. The basis of personalised pharmacological treatment has emerged in light of technological advancements including DNA sequencing, proteomics, wireless health monitoring devices and theoretical methods. Besides, the better understanding and knowledge of pharmacological events such as drug journey (pharmacokinetics) has emphasised inter-individual variations as well as longitudinal intra-individual variations over time in the context of (life-)long treatments [1,2]. However, despite the extensive studies, the inter-individual variability in some drug responses remains challenging and poorly understood.

The field of pharmacokinetics/pharmacodynamics (PK/PD) relationships has been widely used over the past decades to estimate and predict major determinants of overall drug response, whether linked to therapeutic or adverse effects. In addition to assessing a drug dose in the so-called therapeutic window, it is important to thoroughly determine the response and the drug's mechanism of action. On a daily basis, clinicians use PK/PD relationships from a systemic point of view, e.g., considering blood xenobiotic concentrations. However, there is still a gap between systemic and local PK, i.e., drug concentration close to its (off-)targets. Globally, both local and systemic PK and, in turn, pharmacodynamics are particularly impacted by membrane crossing events since such events drive overall drug absorption, distribution, metabolism, and elimination. Therefore, the field of pharmacological research would benefit from a fine and deep understanding of such events at the atomic scale to better decipher the link between local and systemic PK/PD [3].

Drug membrane transporters are major determinants of pharmacokinetics and drug effects. Furthermore, due to the well-known wide range of substrate specificity of drug membrane transporters, they can also mediate drug-drug interactions (DDIs). All these membrane-mediated events affecting drug journey have been stressed by the International Transporter Consortium, which has defined in 2012 and 2018 lists of "transporters of emerging clinical importance" [4,5]. Special attention was oriented toward transporters expressed in the liver and kidneys, where the majority of drugs are metabolised and eliminated. Among the proteins of "emerging importance" are Solute Carriers (SLC) as well as ATP-Binding Cassette (ABC), highlighting the roles of e.g., ABCB1 (P-gp), ABCC2 (MRP2), ABCC4 (MRP4), SLC22A6 (OAT1), SLC22A8 (OAT3), and SLCO1B1 (OATP1B1) that actively participate in processes and have been suggested to mediate PK and DDIs [4,5]. ITC recommendations have led federal agencies (e.g., FDA, EMA) to define transporter lists that must be tested in the field of drug discovery and development prior to their release on the market. Therefore, recent efforts have been made toward the understanding of drug protein interactions to decipher transporters' roles in drug safety and efficacy. For instance, genetic polymorphisms were carried out to depict loss-of-function genetic variants in transporters that may contribute to modulating drug PK [6].

Despite the fact that the transporter studies have emphasised their pharmacokinetic, pharmacological, and physiological relevance, their molecular understanding remains fragmented. Knowledge of drug delivery and disposition, as well as molecular mechanisms of drug interaction accounting for interindividual/interspecies differences, can be supported by the scientific study of transporter structure, dynamics, and mechanism [7].

Due to experimental limitations, drug-protein interactions and drug-drug interactions, and therefore their underlying mechanisms, remain poorly understood at the molecular level. Computational methods can support experimental observations to decipher transporter-mediated effects on drug disposition and efficacy. For instance, by means of all-atom Molecular Dynamics (MD) simulation, one can gain a better understanding of the proteins' structures in different conformational states (inward-/outward-facing) at the atomic scale. Such approaches can investigate the role of the environment, namely the surrounding lipid bilayer, which in turn provides insights on the mechanism of substrate binding, inhibition, translocation, as well as kinetics. The computational tools are also expected to shed light on the impact of polymorphisms on the transporter structures, which may be associated with function modulation. Modern drug development actually incorporates in silico studies that predict the drug's effects and offer understanding at the atomic scale [7].

The present study investigates solute carriers, a superfamily of membrane transporters. Focus was paid to the Major Facilitator Superfamily (MFS) of the SLCs. The manuscript includes the structural description of a clinically relevant human transporter, namely Organic Anion Transporter 1 (SLC22A6/OAT1), but also the well-described glucose transporters 1 (SLC2A1/GLUT1), and glucose transporter 3 (SLC2A3/GLUT3), given that MFS folds were shown to share structural patterns in spite of large differences in terms of sequence. The research extends to various conformational states, substrate binding modes, the mode of action for inhibitors and the impact of the composition of the lipid bilayer on the transporters.

The present manuscript is organised into five chapters. The chapter I presents the context, focusing on drug disposition, including concepts of pharmacokinetics (section I.1) and pharmacodynamics (section I.2). An overview of the role of drug membrane transporters is provided in section I.3, including the importance of lipid membrane bilayers. The chapter II presents the fundamentals of molecular dynamics describing molecular mechanics basis and force fields (section II.1-II.3), classical molecular dynamics simulations (section II.4), free energy calculations (section II.5) and more specifically protein modelling techniques (section II.6).

The following chapters are dedicated to *hOAT1,* including its broad structural description (Chapter III) and the description of binding modes and allosteric effects of the lipid membrane bilayer (section IV). Chapter V is dedicated to investigation of a few MFS transporters, including *h*OAT1 in terms of structural dynamics, substrate, and inhibitor binding, emphasising the role of the lipid/protein interplay.

Chapter I. Pharmacological significance of drug disposition

I.1. Pharmacokinetics: How a drug is processed by the organism

Drug disposition describes the drug's journey, from its entry to its exit out from a body. This is usually described by four primary processes, namely drug absorption, distribution, metabolism, and elimination (ADME). In some models, toxicity may be taken into account as well, and all together is referred to as ADMET or ADME-Tox. These processes impact the drug concentrations in body compartments and the kinetics of drug exposure to certain tissues, which are associated with the drug's efficacy. That stands for pharmacokinetics (PK), a fundamental concept in pharmacology. In other words, pharmacokinetics can be defined by the study of how a drug is processed by the organism.

A drug may be administered via oral, parenteral (injection), sublingual (under the tongue), transdermal (directly on skin), rectal (suppository), pulmonary (inhalation), topical (applied to the skin), or ocular (through the eyes) routes. Some of the alternatives provide a drug in its active form, while others, such as the oral route, may use a prodrug that must be chemically modified in order to be effective. It depends on physicochemical properties and the purpose for which a drug should be administered conveniently. The drug is subsequently absorbed into the bloodstream and transported to its site of action. If a drug is sufficiently polar (and thus hydrophilic), it can be eliminated directly; otherwise, it must undergo biotransformation during metabolism before being eliminated. Understanding the processes is essential for estimating drug PK during the process of drug discovery or in pharmacological research [8–10].

I.1.1. Absorption

Absorption refers to the journey of a drug from its administration until it reaches the biological main fluid (i.e., the blood stream). In the field of pharmacology, particular attention must be paid to the concentration of effective drug form that reaches the systemic circulation, so-called bioavailability [11]. The bioavailability strongly depends on the administration route. The most common way for drug intake is oral administration, which is thought to be the most convenient. However, gastro-intestinal absorption depends on pH of the gastrointestinal system, psychochemical properties of the drug, interaction with food, as well as the patient's health and individual ability of the organism. Therefore, some drugs may be poorly absorbed through oral administration. On the contrary, an injected drug (parenteral injection) does not consider the absorption event, and drug bioavailability easily reaches 100%. Unless the drug is administrated by injection, the bioavailability is always decreased during absorption (Fig. 1). Bioavailability is essential for calculating drug dosages. Absorption affects bioavailability as well as physiological and pathological factors, which must be considered in order to determine the most convenient method of administering a drug [9].

The rate of absorption of orally administered drugs is affected by the rate of dissolution in gastrointestinal (GI) fluids. For controlled-release products, for example, prolonged absorption may be achieved using devices that reduce the frequency of drug release and thus its intake while maintaining its therapeutic effect [9].

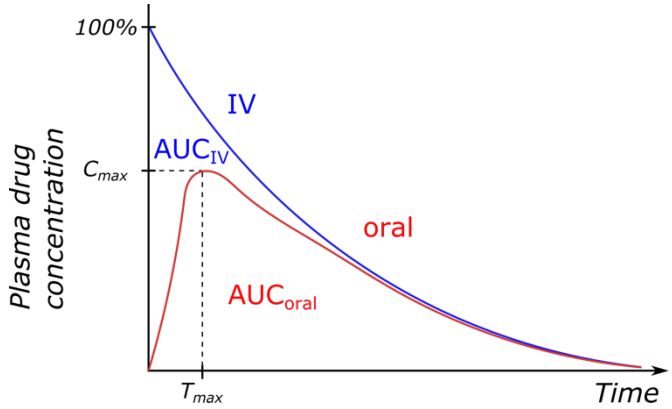


Figure 1. Bioavailability as a function of plasma drug concentration in time of a drug taken by intravenous (IV, blue curve) and oral route (red curve). The peak concentration is presented as C_{max} at given time (T_{max}). The difference between drug intake and T_{max} stands for absorption.

I.1.2. Distribution

Distribution depends on (i) drug lipophilicity, the drug with high the lipophilicity will diffuse the membrane easier than hydrophilic compounds; (ii) blood flow, for instance, brain receives more blood flow than the skin, therefore a drug will accumulate faster in the brain than skin, (iii) capillary permeability, presence of tight junctions in the intracellular gaps limits distribution; (iv) binding to plasma proteins (albumins) slows down the distribution. A pharmacokinetic factor *volume of distribution* (V_d) can be used to predict whether a drug accumulates more in the blood or tissues. That is crucial when estimating a drug's dose.

$$V_d = \frac{amount\ of\ drug\ in\ the\ body}{drug\ concentration\ in\ blood\ plasma}$$

Drugs with high molecular weight binding to plasma proteins will display low value of V_d, while lipophilic drugs with low molecular weight will be characterized by high V_d [9,10].

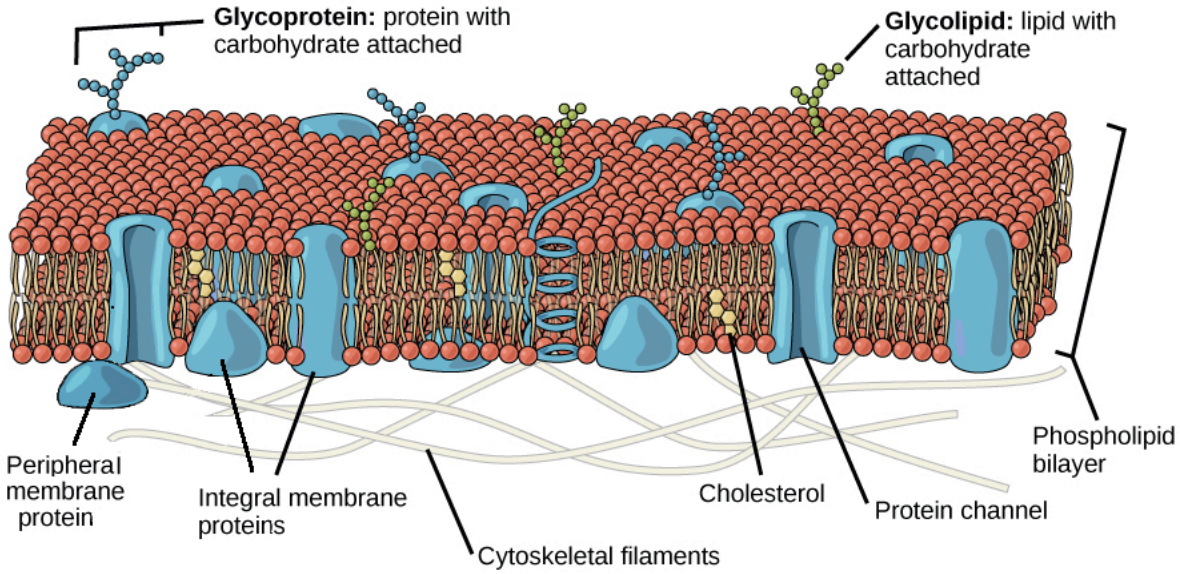


Figure 2. Schematic picture of plasma membrane. Figure adapted from *OpenStax Biology* [12].

Since a drug frequently crosses multiple cell membranes before reaching targets, its physicochemical properties are critical to understanding its disposition in an organism. The plasma membrane of a cell is formed of a lipid bilayer with a polar surface and a highly hydrophobic interior. This property is a consequence of the amphiphilicity of the lipids, which are composed of lipophilic acyl chains oriented inwards and a polar head-oriented outwards from the centre of the bilayer. Cell membranes also contain proteins, glycolipids, glycoproteins, and carbohydrates, which create the complexity of a membrane (Fig. 2). The composition of a membrane varies among cells and tissue types, with different types of lipids determining its polarity, fluidity, flexibility, and permeability.

Certain physicochemical features, in terms of shape, size, charge, and polarity, of a drug determine its distribution route [9]. Drug transport into the inside of a cell occurs in various ways, including passive and active transport mechanisms. The passive transport transfers molecules downhill, the concentration gradient of the substrate and the electrochemical potential without using energy. Where active transport transfers the molecules uphill, the gradient concentration is driven by either a co-substrate, an ion, or energy released from the hydrolysis of adenosine triphosphate (ATP) [8]. This is true when the molecules are transported through the membrane. However, molecules may be transported through paracellular gaps, which are usually limited only by blood flow. Unless the paracellular gaps are equipped with so-called "tight junctions," size- and charge-selective intercellular adhesion complexes that are selectively permeable. They are abundant in epithelial cells, largely in brain capillary endothelial cells [13].

I.1.3. Elimination

Drugs are eliminated via renal excretion, liver bile, sweat, lungs, and breast milk. Besides xenobiotic elimination, the kidneys regulate homeostasis of the body’s fluids, minerals, nutrition, and acidity. Renal-mediated urine-blood exchange events involve three distinct mechanisms: (i) glomerular filtration, (ii) tubular secretion, and (iii) tubular reabsorption. Each event has its own role and limitations.

The glomerular filtration eliminates mainly low-molecular weight organic compounds with positive or no charge as well as inorganic ions. Its effectiveness is determined by the blood flow and rate of glomerular filtration [14]. Tubular secretion of drugs involves transcellular transport from plasma to the urine (Fig. 3). It involves membrane proteins or (membrane transporters) that translocate substrates across the membrane at the apical and basolateral of kidney cells. Solute carriers (SLC) and ATP-binding cassette transporters (ABC) play a major role in drug secretion. Drug uptake from the blood is mediated by e.g., Organic Anion Transporters (OATs) and Organic Cation Transporters (OCTs), whereas drug excretion is mediated primarily by multi-antimicrobial extrusion protein (MATE) and ABC proteins (e.g., MRP4) [15,16]. Since elimination considers two steps for drug translocation, namely from blood into the proximal tubule cells and from the proximal tubule cells into the urine, efflux transporter impairment can contribute to nephrotoxicity associated with intracellular drug accumulation [17]. Transporters may be saturated when a drug concentration reaches a given limit. Given the diversity of transporter substrates, which may result in drug-drug interactions (DDIs) in the case of competition [4]. Fortunately, transporters, on the other hand, exhibit substrate overlaps, which is important because if one transporter is blocked, the other can take over. This is the case of, e.g., OAT1 and OAT3. However, their substrate preferences differ primarily in terms of molecular size [18].

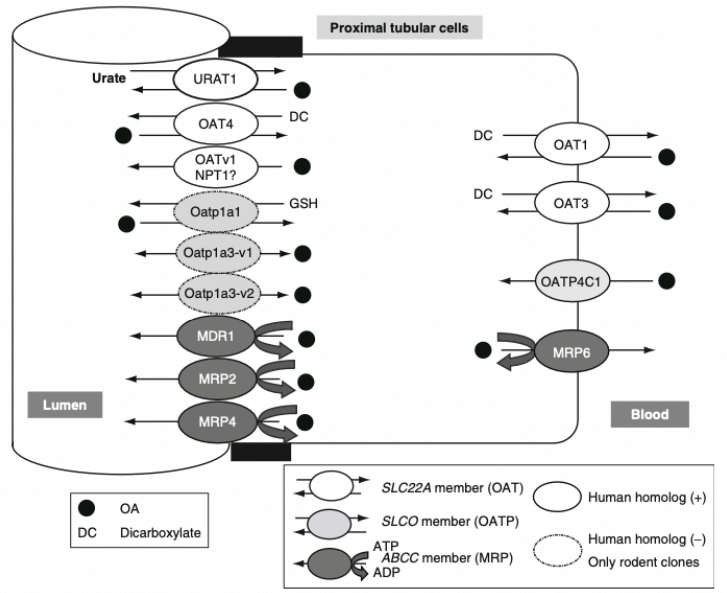


Figure 3. The uptake from the blood into the proximal tubule cell and excretion to the lumen in the elimination process. Figure adapted from *Pharmacology. Principles and Practice* [16].

In contrast to filtration and secretion, tubular reabsorption aims to uptake substances from the urine into the bloodstream instead of the other way around (Fig. 4). Tubular reabsorption may also involve membrane proteins even tough passive permeation was shown to play an important role. Therefore, small, nonionized, lipophilic molecules are easily reabsorbed, while in the case of weak acids and bases, the perturbation of pH impacts the reabsorption [14].

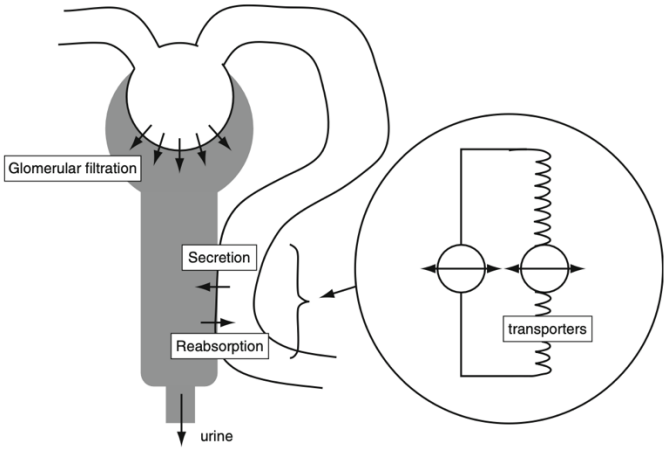


Figure 4. Tubular reabsorption. Figure adapted from *Pharmacology. Principles and Practice* [16].

Drug elimination can be monitored by clearance, or rate of elimination in the urine, which is relative to the plasma drug concentration as a function of time. Considering the whole body, total clearance of a given drug can be calculated by summing renal, hepatic, and other clearances [9,14,16].

$$Cl = Cl_{renal} + Cl_{hepatic} + Cl_{other}$$

Most drugs follow first-order kinetics at the systemic level, so the rate of elimination of a drug is proportional to its concentration in the body. However, some drugs, such as aspirin, are eliminated by zero-order kinetics [19], where the amount of an eliminated drug is independent of its concentration, resulting in a constant rate of elimination.

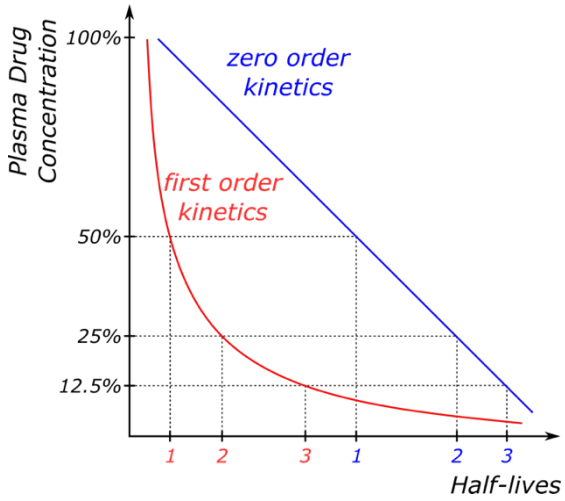


Figure 5. The zero (blue) and first (red) order kinetics in drug elimination.

Given the exponential behaviour of first-order kinetics, plasma drug concentration can be assessed by using its half-life, i.e., the time required for 50% elimination (Fig. 5). Both clearance and half-life are important to predict the steady state of a drug, and in turn, if blood drug concentration ranges in the therapeutic window over time [9,14,16].

I.1.4. Metabolism

While hydrophilic drugs are easily excreted from the body, lipophilic and amphiphilic ones may need to be biotransformed to increase their hydrophilicity. Most of these chemical reactions occur in the liver. However, many enzymes are also found in the epithelial cells of the upper intestine, kidneys, lungs, placenta, and brain. Metabolic reactions increase hydrophilicity by conjugation synthetic reaction. The metabolism mostly consists of Phase I and Phase II reactions. The irreversible reactions of Phase I include oxidation, hydrolysis, and reduction. While conjugations (i.e., covalent binding to a given fragment) occur as Phase II reactions, such as glutathione conjugation, acetylation, sulfonation, and glucuronidation. It is important to note that xenobiotic metabolism can lead to the activation or inactivation of the parent xenobiotic. Even though an inactive drug metabolite is typically produced, biologically active and potentially toxic metabolites can also be generated. In this context, drugs are proposed in the form of so-called prodrugs, which require activation by using biotransformation reactions [9,10,20].

The enzymes mediating the biotransformation include cytochromes P450, alcohol dehydrogenase, peroxidases, monoamine oxidases, or xanthine oxidase. However, the cytochromes P450 participate in majority of the reactions, where CYP 3A4/5, CYP 2D6, CYP 2C8/9, and CYP IA2 are the most common. Some substances may promote or inhibit CYP enzymes, suggesting PK-related DDIs. Typical inducers include phenytoin, carbamazepine, rifampin, alcohol, barbiturates, and St. John's Wort (PCRABS), while inhibitors include grapefruit, protease inhibitors, azole antifungals, cimetidine, macrolides, amiodarone, and non-DHP CCBs (GPACMAN) [21].

I.1.5. Membrane crossing processes as key events in ADME

Apart from paracellular transport, every single ADME-step is governed by membrane crossing events. For instance, absorption and elimination require transcellular transport of xenobiotics or metabolites across two membranes. Metabolism mainly occurs in hepatocytes in which

xenobiotics and metabolites must enter and exit, respectively. Several molecular mechanisms were suggested for these events, which are briefly presented below [10].

I.1.5.1. Passive permeation transport

Permeation diffusion driven by the concentration gradient enables molecules to pass through lipid bilayers. A so-called permeant must display appropriate lipid solubility and ionisation to efficiently cross the membrane by passive permeation. The permeation of a molecule includes (i) membrane partitioning, (ii) diffusion across the lipid bilayer (or flip-flop), and (iii) membrane exiting (repartition, Fig. 6B) [22]. The permeation of a drug is governed by its electrochemical gradient, which depends on the charge of the molecule (z), membrane voltage (E_m), and concentration of the drug outside (C_o) and inside (C_i) of the plasma membrane:

$$\Delta\mu = zE_mF + RTln\left(\frac{C_i}{C_o}\right),$$

where F is faraday constant, R is the gas constant and T stands for temperature. The structure-kinetic relationship of the crossing the membrane event is described by the permeability constant (P) which positively correlates with lipophilicity[1] and inversely correlates with the root square of molecular weight. The movement through the membrane can be modelled using the 1th Fick's law of diffusion, where the permeability P is multiplied by the concentration difference [9].

$$J = -P\frac{zE_mF}{RT}\left[\frac{C_i - C_o\exp\left(E_mF/RT\right)}{1 - \exp\left(E_mF/RT\right)}\right]$$

I.1.5.2. Protein-mediated transport

Drug influx and efflux are mediated by membrane carriers through polarised membranes at a rate that is significantly faster than passive diffusion for given molecules. Membrane proteins can transfer molecules either against their electrochemical gradient using energy or by simply facilitating and regulating substrate translocation, where molecules are transported selectively.

An important role in drug transport is played by ATP-binding cassette transporters (Fig. 6D), which usually efflux molecules out of the cell using energy stored in ATP molecules. The importance and function of pharmacologically relevant transporters are described in section 1.3.3. The energy stored in an electrochemical gradient is known to be used in secondary active transport. For instance, a membrane transporter Na^+/K^+-ATPase transports Na^+ and K^+ against its gradient concentration using energy released from ATP hydrolysis. The intracellular gradient of Na+ is then used by the Na^+/Ca^{2+} exchange protein to export Ca^{2+}, maintaining its low intracellular concentration [23,24]. Some membrane carriers represent facilitated transport, in which the molecule transfer occurs down the electrochemical gradient without energy input. This is, for instance, the case of glucose uptake by GLUT1 (Fig. 6C).

In the pharmacological context, it is worth mentioning that protein-mediated transport can be inhibited or saturated, which will significantly decrease the kinetics of transport. This has been shown to strongly affect drug PK, such as in probenecid/G-penicillin drug-drug interactions [4,20,25,26].

[1] The lipophilicity can be described by $logP$, the octanol-water partition coefficient, with lower values indicating hydrophilic molecules and higher numbers indicating lipophilic molecules.

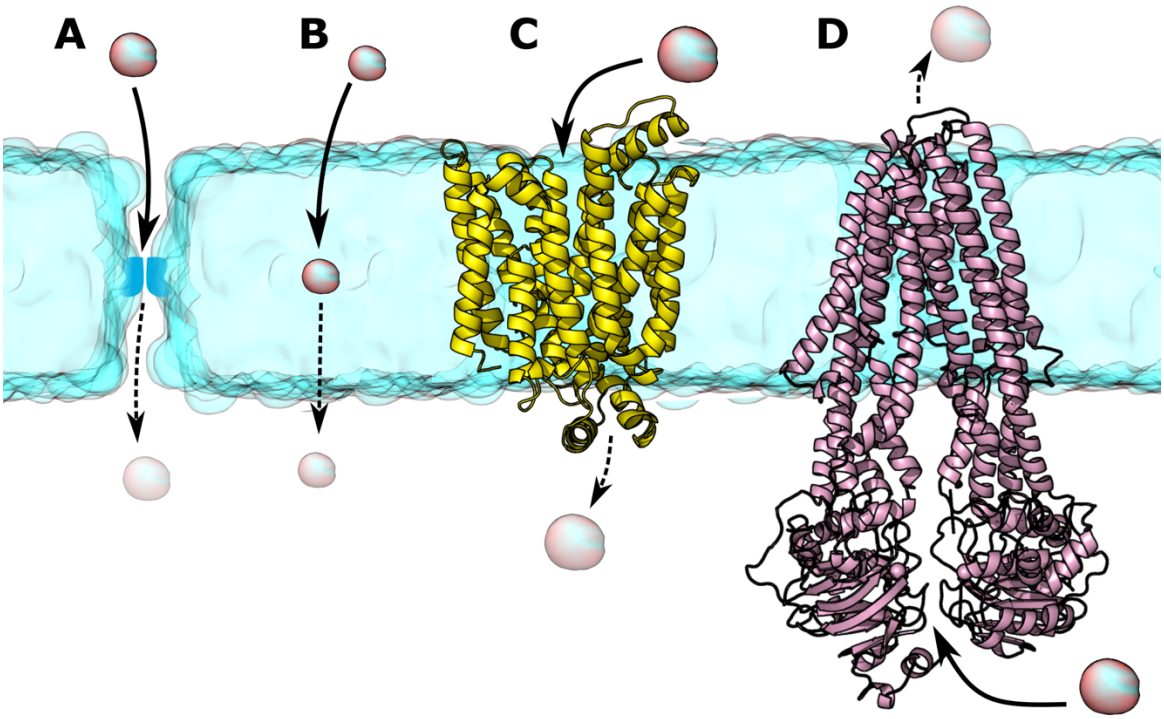


Figure 6. Molecule permeation through the lipid bilayer (cyan sphere). (A) Movement through intracellular gaps (tight junctions). (B) Passive permeation of molecules through lipid bilayer using the concentration gradient. (C) Facilitated diffusion. The molecules are transported by Solute Carriers (SLC) proteins. (D) Active transport. The transport depends on energy derived from ATP (adenosine triphosphate) via ATP-binding cassette (ABC) transporter.

It is worth mentioning that large molecules can enter into (or exit out of) cells via "endocytosis". A molecule enters the cell by forming a vesicle from the cell membrane to be released into the cytosol. However, molecular mechanisms remain under debate given the large amount of energy required by the cell membrane deformation [10].

I.2. Pharmacodynamics - Drug's effect on the body

Pharmacodynamics (PD) focuses on the biochemical and physiological effects of a drug. The effective use of a therapeutic agent as well as the development of new drugs mostly depends on the understanding of the mechanism of action. Organism response to the active substance is often associated with interactions with targets. A drug targets may be a protein, including receptors, enzymes, carriers, and ion channels, but also DNA. A drug target is often referred to as a receptor in a general manner [8,9].

Protein receptors often bind xenobiotics in a non-covalent manner, including hydrogen bonds, van der Waals interactions, hydrophobic or ionic interactions. The binding affinity of a drug is correlated to the structures of both xenobiotics and receptors. The interactions between the binding pocket and the drug can be very specific, and structural changes in one or both can significantly modulate receptor-xenobiotic binding affinity. Such understanding is key to understanding the rationales behind drug design [3,27]. Interestingly, some drugs may covalently bind to receptor, such as acetylsalicylic acid to COX-2 (cyclooxygenase 2) leading to a "permanent" effect [28].

While a drug ideally displays selectivity toward tissues, it has been shown that receptors may selectively bind individual chemical classes of drugs. In this context, owing to selective nature of the interaction between a receptor and a drug, drug effect is observed in tissue where

receptor is highly expressed. Therefore, the main drug's site of action is defined by the location of receptor expression. However, adverse effects can be observed when xenobiotic interacts with targeted receptors (or proteins from the same family) in other tissues. Besides, it is important to note that receptor interactions are saturable; overexposure favouring thus interaction with off-targets [9,10].

The drug's effect may also be limited by the inherent structure modulation of structure of the target receptor. For instance, at the molecular level, xenobiotic-target interactions may be increased or decreased by a receptor mutation or the presence of a competitor or an allosteric partner (i.e., a molecule binding a distant site of the receptor) [29].

I.2.1. Molecular mechanism of xenobiotic-receptor interactions

Expected activation or inhibition of a target receptor is often associated with its conformational state. Biological responses may be enhanced or reduced by binding small molecules that can activate or inactivate a receptor by promoting a given conformation.

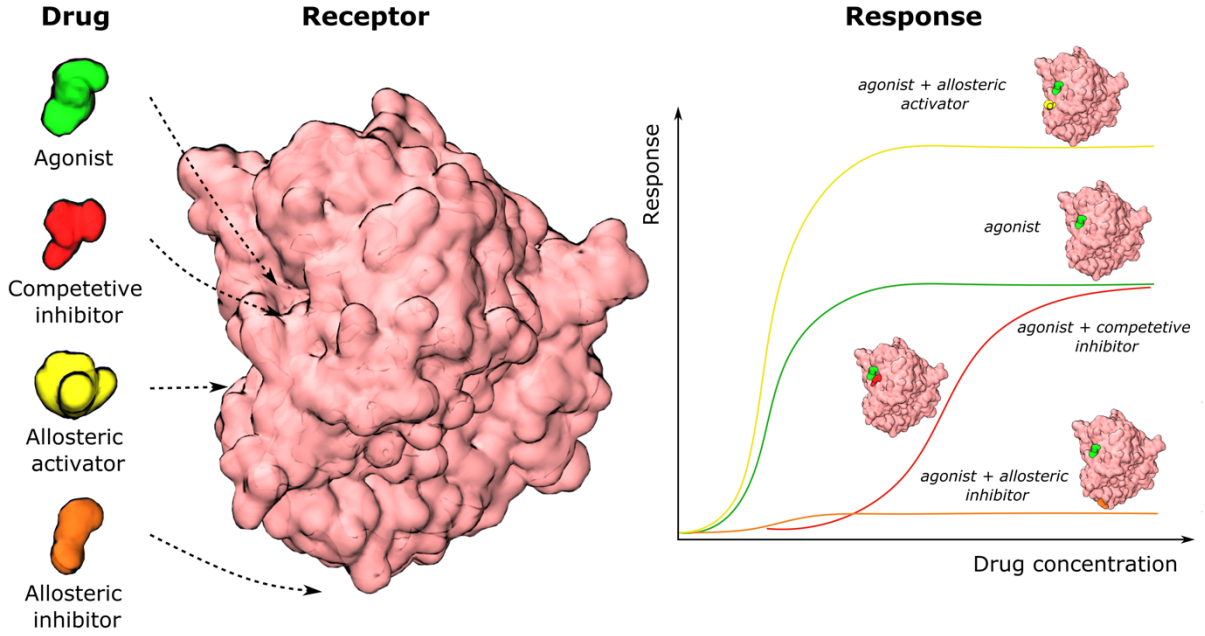


Figure 7. Binding of an agonist (green) to a receptor induce a biological response which can be (i) enhanced by an activator (yellow) that binds to the allosteric binding site, or (ii) significantly reduced by an allosteric inhibitor (orange) that binds independently from the agonist. Otherwise, the inhibitor (red) can have similar binding affinity as the agonist, displaying competitive inhibition, which decays with increasing agonist drug dose.

A drug is an agonist if the binding event favours the active state of a receptor and initiates its biological response by related cell pathway activation. Depending on the level of biological response, an agonist may be full, partial, or inverse. A full agonist at a high concentration achieves a maximal response, while a partial agonist, regardless of the concentration, never reaches the maximal biological response due to the incomplete stabilisation of the active conformation. However, the invert agonist displays higher affinity toward the inactive conformation that leads to the stabilisation of the inactive state, and thus its inhibition. An antagonist has high affinity for the inactive state of a receptor, blocking the action of a drug. A competitive antagonist may compete with the endogenous agonist over the same binding pocket. This may lead to the increase of blood concentration of the endogenous substrate. In

another case, a xenobiotic can bind to a distant site as compared to the known active binding site. The so-called allosteric effect may favour or preclude the endogenous substrate binding [30].

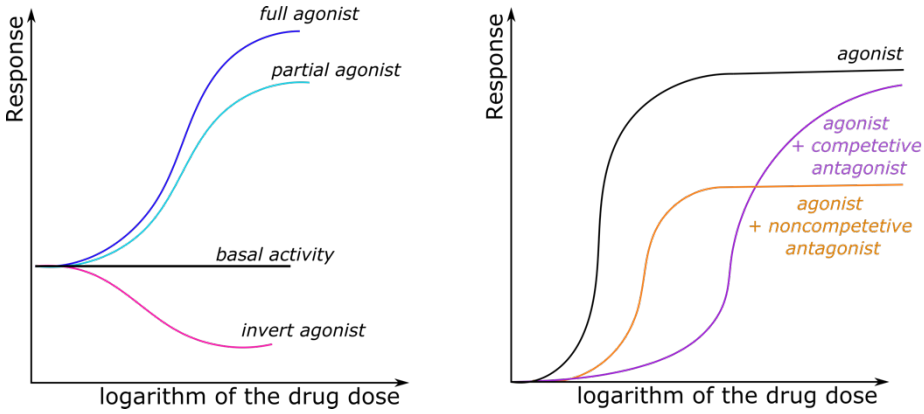


Figure 8. The biological response to an increased dose of an agonist in the case of a full, partial, or inverted agonist (left) or in the presence of a competitive or non-competitive inhibitor (right).

The PK-PD relationship models the relation between a drug dose (PK) and the biological response (PD). Often, drug effect is defined as the concentration-dependent ratio between a drug maximal response (E_{max}) and the drug concertation at which 50% of the maximal response is produced (EC_{50}).

$$E = \frac{E_{max} \times C}{C + EC_{50}}$$

A drug can be evaluated based on its potency, which is indicated by EC_{50}. The term refers to the drug concentration (or dose) at which EC_{50} is achieved. The lower the EC_{50}, the more potent the xenobiotic [9,10,30].

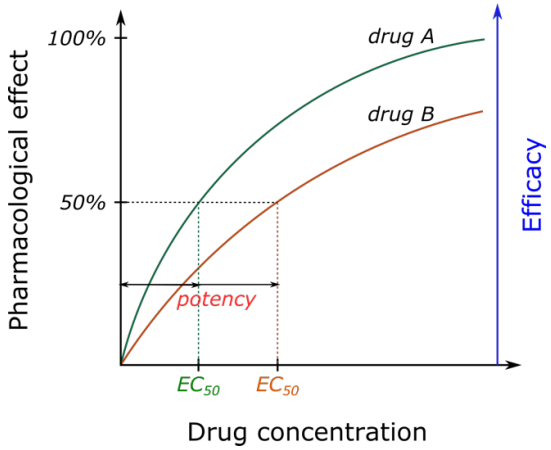


Figure 9. The potency of a drug is evaluated by the drug concentration needed to reach 50% of its pharmacological effect. In the plot, drug A is more potent than drug B and reaches higher maximal efficacy.

The *clinical effectiveness* describes a drug's ability to reach its site of action, accounting the xenobiotic journey, i.e., its ADME model. For instance, in Figure 9, drug B does not achieve the same pharmacological effect as drug A due to its lower maximal efficacy, even though the

molecule B can be more potent at the molecular level. This highlights the gap existing between the molecular and systemic understanding of PK/PD relationships for which no robust model has been developed so far [31,32].

Due to PK and PD external factors, defining an “ideal” drug dose is still challenging. It can be an individual value for a pathological condition or even to a person also considering its genetic background in the field of pharmacogenetics (PGx) research [33]. A measure of therapeutic index is important to determine the safest drug dose range for which drug responses is in the therapeutic window (Fig. 10). The therapeutic index is indicated by the ratio of TD_{50} and ED_{50}, which describes the drug dose that generates toxicity and effective response, respectively, in 50% of the individuals of the studied population [9,10].

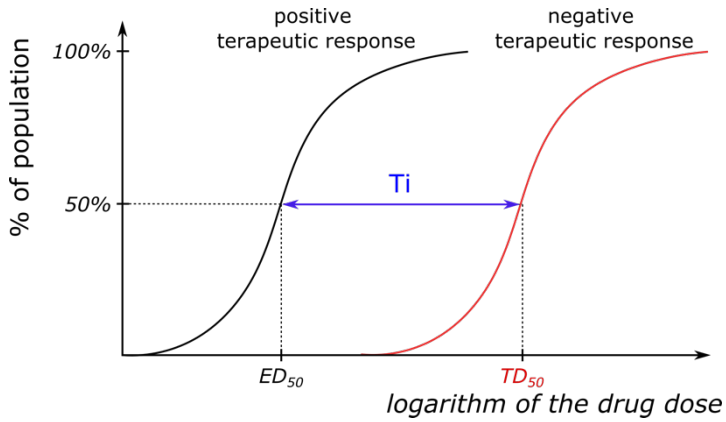


Figure 10. Therapeutic index.

I.2.2. Example of drug receptors

When a drug binds to a receptor, it often triggers a cascade of events via an effector. The effector may be a component of the receptor or an independent molecule, such as an ion or another protein. The process can occur through distinct pathways with distinct biochemical mechanisms. Receptors within superfamilies share structural patterns and mechanisms of action. The mechanisms of signalling pathways are important for understanding cell signalling that translates the received signals into a precise signal of action. There are several well-known mechanisms concerning (i) intracellular receptors; (ii) ligand-gated ion channels; (iii) enzyme-like receptors; and (iv) G-protein coupled receptors [8,9].

I.2.2.1. Intracellular receptor

Xenobiotics can bind to an intracellular receptor (Fig. 11A). For instance, steroids and hormone-like compounds bind intracellular receptors that are translocated to the nucleus to modulate gene expression. The therapeutic consequence of the mechanism is the lag period, which can take up to several hours. However, the therapeutic effect can last for up to days [34].

I.2.2.2. Protein kinases

Protein kinases are enzyme receptors that intracellularly phosphorylate effector proteins (Fig. 11C). Phosphorylation is often associated with activation/inhibition of cellular pathway. Membrane protein kinase structures often exhibits an extracellular domain that binds an agonist, an intracellular phosphorylating domain. Both domains are distantly connected through a transmembrane domain which propagate the signal from substrate binding across

the membrane. Kinases typically contain additional larger domains that may diversify their properties. Substrate or agonist binding to the extracellular domain may trigger conformational changes that e.g., favour receptor dimerization which in turn activates the phosphorylating domain. The active receptor can then phosphorylate other proteins activating cell pathways for further biological functions. It is important to note that there almost systematically exist cell control systems which aims at (up)downregulating kinases if (under)overactivity is detected. For instance, downregulation can be achieved by receptor endocytosis which translocate substrate-receptor complex into the cytosol to be then degraded. Upregulation can occur by promoting gene expression leading to increase receptor concentration at the membrane [8,35].

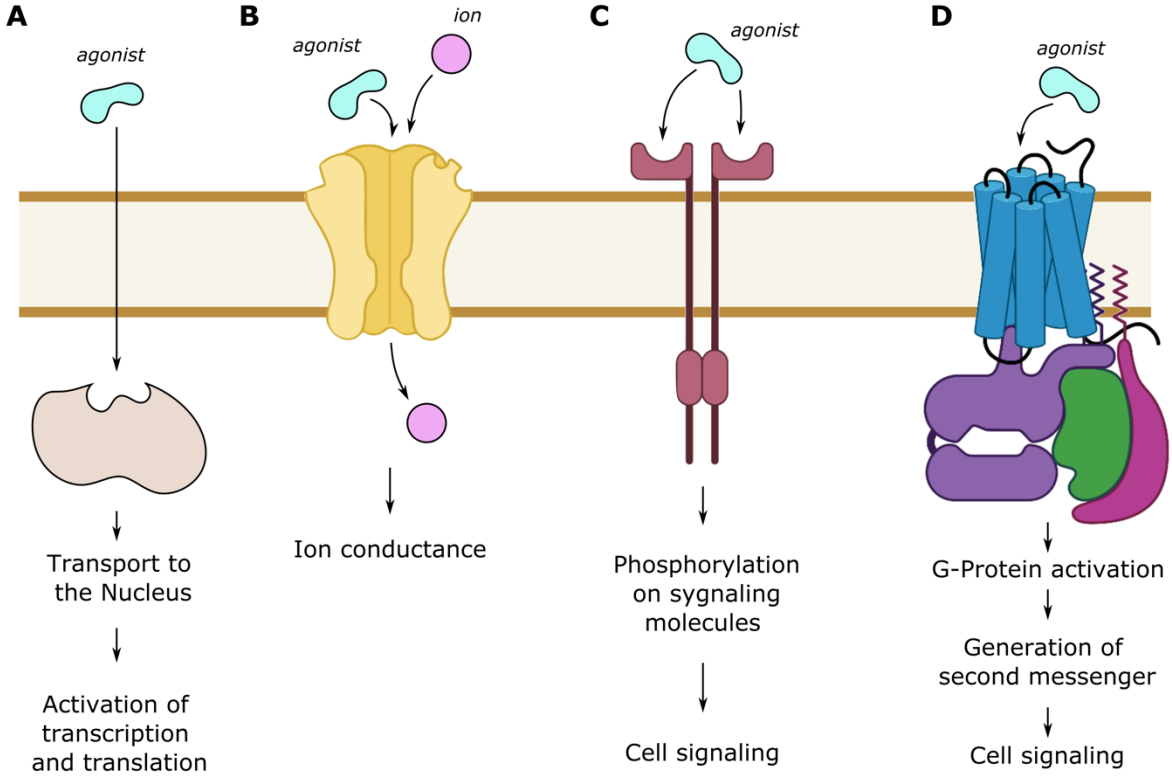


Figure 11. Examples of drug receptors include: (A) an intracellular receptor; (B) an ion channel; (C) a protein kinase; and (D) a G protein coupled receptor.

I.2.2.3. Ligand and voltage-gated ion channels

Ion conduction between compartments is governed by ligand- and voltage-gated channels (Fig. 11B). The transport occurs down the electrochemical gradient. The transmembrane domain adapts to a cylindrical shape, forming a channel that enables selective ion transport. In response to binding a messenger molecule, the ion channel can be activated toward opening or inhibited by closing. For instance, endogenous acetylcholine promotes Na^+ transport into the cell by binding to the nAChR receptor and inducing the channel opening. Voltage-gated receptors, on the other hand, are activated by membrane potential. Transmembrane ion conductance causes the receptor to transmit the signal via a change in membrane polarisation. Their response is nearly instantaneous and can be regulated by phosphorylation or degradation (endocytosis) [9,36].

I.2.2.4. G proteins

Drugs can impact the concentration of substances known as second messengers, such as Ca^{2+}, the phosphoinositides, or cyclic adenosine-3′, 5′-monophosphate (cAMP), to induce biological responses without entering the cell. On the cell surface, the drug binds to a receptor, which immediately activates the intracellular coupled G-protein that governs the activity of an effector (Fig. 11D). An effector, such as an enzyme or ion channel, directly regulates the concentration of second messengers. For example, the effector adenylyl cyclase increases the intracellular concentration of cAMP. G-protein coupled receptor (GPCR) superfamily is a widely known target for drugs. They are made of seven transmembrane domains and an intracellular G-protein component. The receptor is responsible for selective drug binding, while the G-protein is responsible for the signal transmission. G-protein activation is dependent on the binding and hydrolysis of guanosine triphosphate (GTP). A drug bound on the extracellular side of a receptor drives its conformational change that may promotes the binding of GTP instead of GDP to G proteins. As long as GTP is bound, activation is promoted, whereas GTP hydrolysis terminates activity [37,38].

I.3. Membrane Transporters

I.3.1. Overview of the roles and functions of membrane transporters

I.3.1.1. Function

Membrane transporters are proteins that facilitate uptake or excretion of exogenous and endogenous substances (see section 1.1.5.2 above). Transporters are essential for cell and thus, whole-body homeostasis because they translocate a broad range of substrates, including nutrients, metabolites, drugs, and toxins. Given that membrane transporters mediate influx and efflux of rate-limiting metabolites, signalling molecules, and other endogenous substances, they actively participate in the cell signalling process [39].

Interestingly, transporters have been suggested to be involved in distant communication between body fluid compartments, cells, organs, or even organisms. This has been originally proposed within the frameworks of the remote sensing and signalling theory (RSST) [23]. The physiological function and expression of transporters are regulated according to *sensing* through by intracellular and blood substance concentrations. It aims to regulate the concentration of a myriad of endogenous compounds in distant cells. Typically, concentration of parent endogenous compounds and their metabolic forms depends on the transporters' activity. In other words, the increased concentration of a given molecule may reflect the abnormal function of organs which may be sensed by distant organs. In response, distant organs can overexpress other transporters, receptors in order to overcome abnormal activities. The remote sensing and signalling system was also suggested to be coordinated with other regulatory systems, such as hormone regulation or growth factors, that together balance out homeostatic perturbations [23].

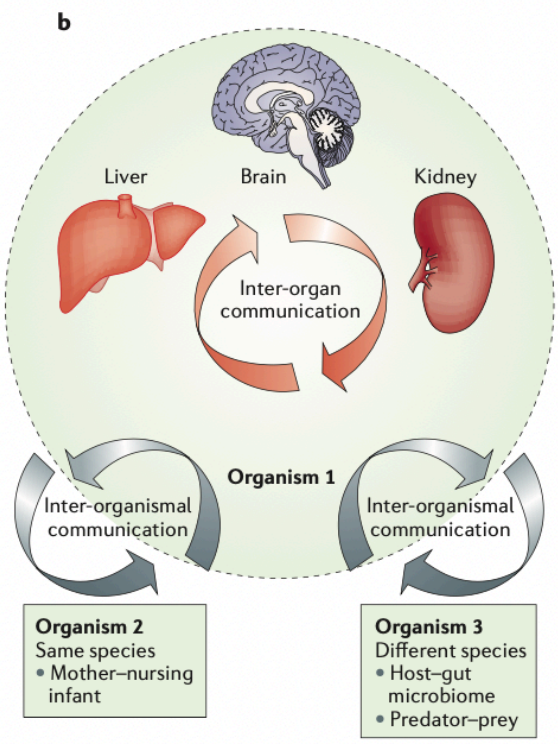


Figure 12. The remote sensing and signalling theory. Figure adapted from *What drug transporters really do?* [23].

Membrane transporters actively participate in drug absorption, distribution, metabolism, and elimination; their function is crucial for pharmacokinetics and pharmacodynamics. Xenobiotics can compete for transporters and modulate their activity, for example by inhibition. The inhibition of Organic Anion Transporters (OATs) by probenecid can decreases the renal clearance and increase the half-life of substances like penicillin, ACE (angiotensin-converting enzyme) inhibitors, and antiviral drugs. However, the consequences of DDIs can extend to cholestasis, nephrotoxicity, and hepatotoxicity by inhibiting important transporters, such as MDR1, BSEP, and OAT1 [40]. Therefore, the understanding of xenobiotic-transporter interactions is helpful to decipher modulations of drug disposition, bioavailability. Drug membrane transporters should also be considered in aspects for possible single nucleotide polymorphisms (SNPs) that result in clinical phenotypes [41]. Their presence affects transport activity toward particular drugs or causes overall dysfunction of a transporter. In some cases, the dysfunction of a transporter can be partially compensated for by another transporter. Drug transporters are not only multi-specific but also display a partial substrate overlap, especially among the families. For instance, OAT1 and OAT3, but also OCT2 and MATE2 [23] display at least partial substrate overlaps. Therefore, if one of the transporter pair is dysfunctional, the other can partially take over its function. Even though drugs display substrate overlap, they display different binding affinities for a given drug. Otherwise, the dysfunction of a transporter can be related to a disease [42].

I.3.1.2. Alternating access mechanism

The majority of drug membrane transporters adopt multiple conformations throughout the different stages of the transport cycle. According to alternating access, the protein undergoes large-scale conformational changes in order to alternately open the extracellular or intracellular gate for substrate uptake and release. Therefore, two major conformations, namely inward-

facing (IF) and outward-facing (OF), are typically highlighted, while the transition between them occurs via metastable transition states.

I.3.1.3. Drug transporter classification

Drug membrane transporters can be classified into two main superfamilies, namely ATP-binding cassette (ABC) proteins and solute carriers (SLC). They significantly differ in their structures and mechanisms. In ABC transporters, molecules are transported via active transport, where the substrates are translocated against the molecular gradient. The transport cycle is driven by the energy acquired from ATP-hydrolysis. For SLCs, several mechanisms were observed or suggested, mostly either facilitated transport or secondary/tertiary active mechanisms (see section 1.3.4.1). Another important aspect that differs between both superfamilies relies on the transport direction. In humans, while ABC transporters almost exclusively mediate out-of-the-cell transport (efflux), SLC proteins can mediate substrate translocation into and out-of-the-cell (influx and efflux) substrate translocation, the former being more often observed [43,44].

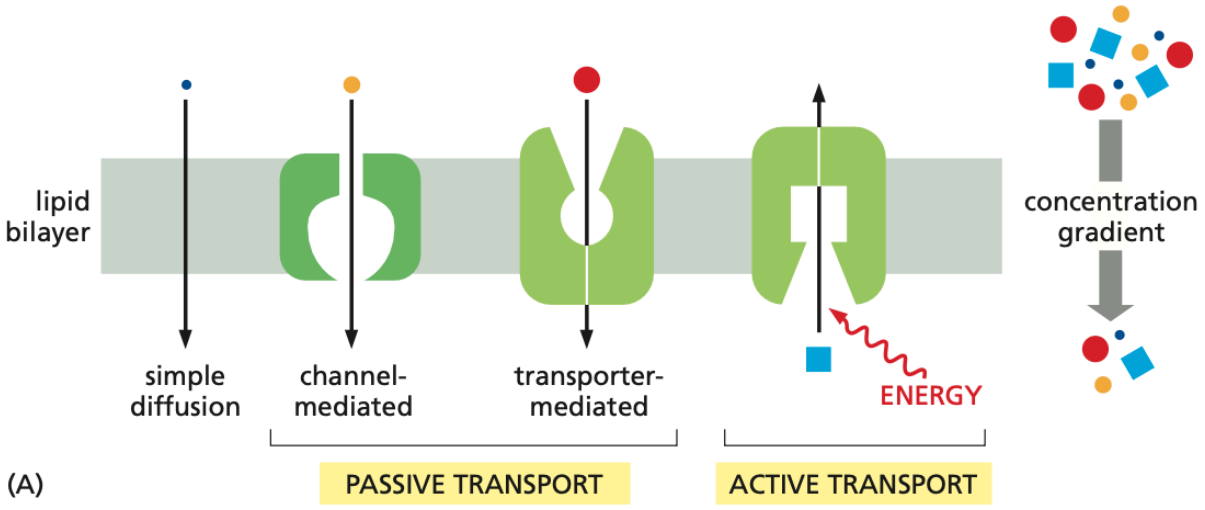


Figure 13. Active and passive transport across the membrane. Figure adapted from *The Cell* [45].

I.3.1.4. Tissues, cells, and membrane transporters

Drug membrane transporters are mostly expressed in epithelial cells, which form a physical barrier between compartments. Cell types tend to express particular sets of transporters. For instance, epithelial cells of the intestine and blood barrier express abundant efflux transporters P-gp (ABCB1) and BCRP (ABCG2) that are considered crucial for drug absorption. Solute carriers are expressed in liver and kidney cells and are of the utmost importance in the context of pharmacology. For instance, organic anion/cation transporters (OATs, OCTs) are considered fundamental for the drug elimination process. Given the wide substrate range, transporters are exposed to PK-DDIs, especially those expressed in the liver, kidney, blood-brain barrier, and intestine. Independently, FDA (2020), EMA (2013) and PMDA (2017) have recommended evaluation of transporters regarding drug-drug interactions (DDIs), including P-gp, BCRP, OATP1B1, OATP1B3, OAT1, OAT3, OCT2, MATE1, MATE2-K [5,46].

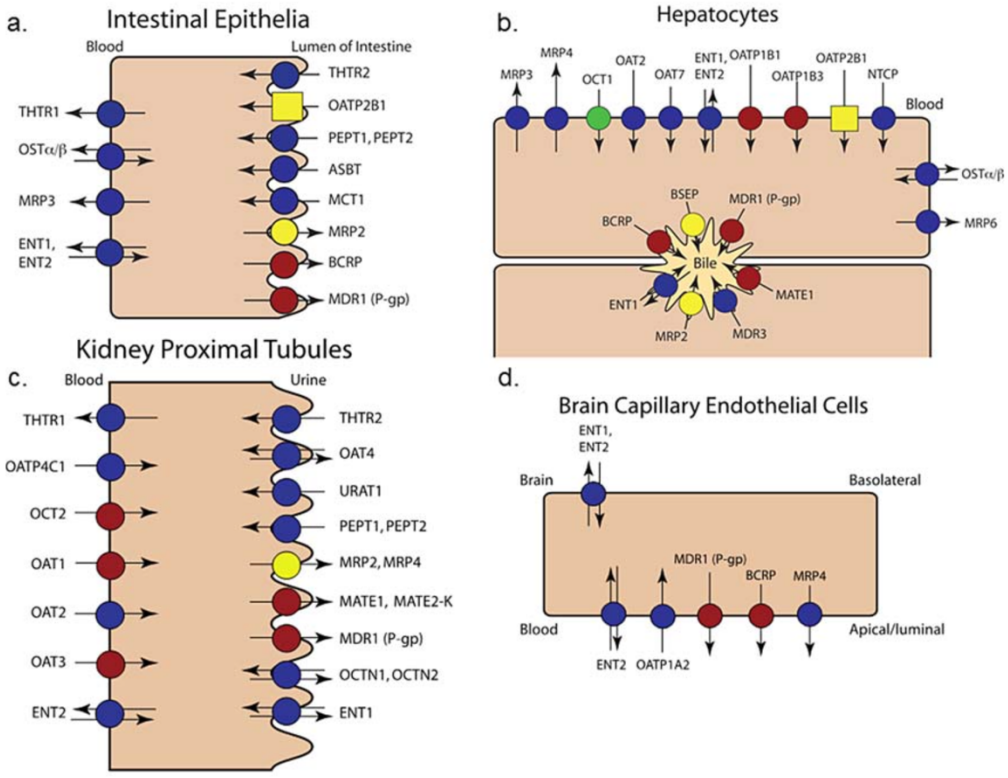


Figure 14. Transporter expression in epithelial cells. The transporters marked with red and yellow circles were recommended for mechanistic study at the first ITC meeting. Where green circle and yellow square marked transporters are called for DDIs studies. The blue-marked transporters do not have any study-specific recommendations. Figure adapted from *Transporters in Drug Development: 2018 ITC Recommendations for Transporters of Emerging Clinical Importance* [5].

Epithelial cells are polarized. This is true for kidney proximal tubule cells, which have a basolateral side exposed to blood circulation and an apical side facing the lumen. Since transporters mediate substrate efflux or influx, their expression is restricted to one side. Drug transcellular transport requires influx and efflux. As a result, a pair of transporters is frequently involved in mediating drug transport from the blood circulation into, for example, urine, bile, or the gut lumen. Likewise, the absorption process requires a similar transporter setup to mediate substances from the gastrointestinal tract into the bloodstream. For instance, in kidney proximal cells, the uptake of anionic xenobiotics is governed by Organic Anion Transporters (OAT) 1 and 3 at the basolateral side of the cell. They may be translocated out of the cell by ABCC2 and ABCC4 (MRP2, MRP4, respectively) [15,17].

I.3.1.5. Determining of transporters' structure

At the molecular level, experimental methods may struggle to obtain both high-resolution and dynamic pictures due to their high conformational flexibility and their low abundance on the cell surface. It is still challenging to experimentally resolve membrane transporter structures. Consequently, numerous transporters lack experimentally determined structures. In this field, the molecular understanding of membrane transporter functions and their interactions with xenobiotics or substrates relies on the joint use of multiple experimental and computational techniques. Indeed, all techniques only provide fragmented pictures when taken individually. However, structural information about different conformations of membrane transport proteins is fundamental for the understanding of their dynamics and function [7].

I.3.2. ABC transporters

I.3.2.1. Overview and clinical significance

ABC transporters have been found to export substances out of mammalians eukaryotic cells. Even though they are classified into seven subfamilies, most of them shared the so-called type IV and V folding according to a recent structural classification [47]. Most of ABC proteins are membrane transporters including drug-membrane transporters.

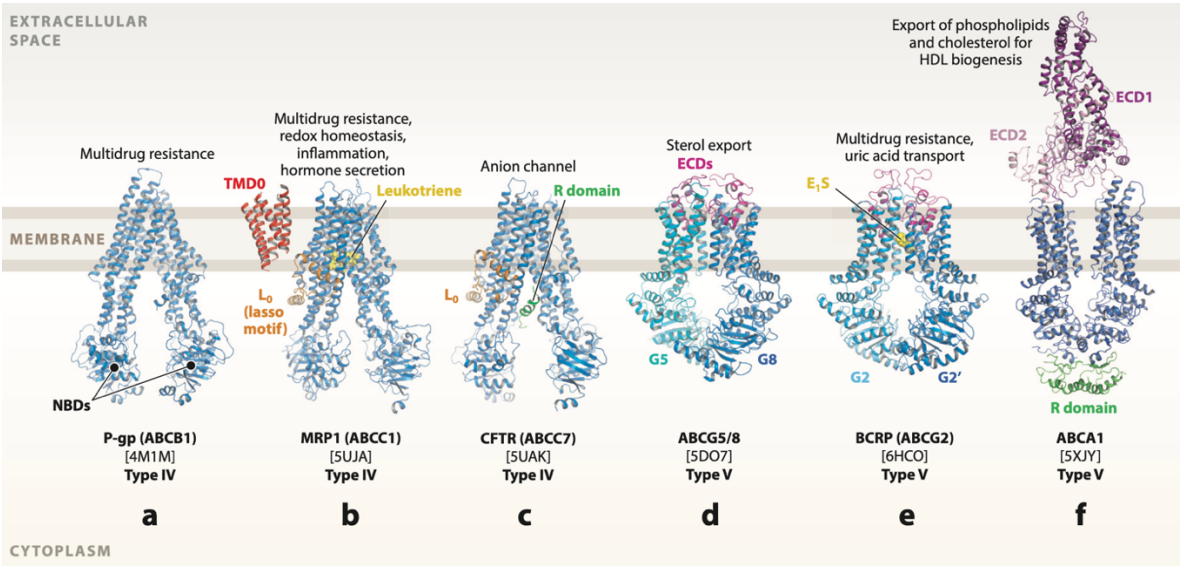


Figure 15. Type IV and V ABC transporters found in mammalians. Figure adapted from *Structural and Mechanistic Principles of ABC Transporters [*47].

One of the most studied human ABC members is the so-called P-glycoprotein (P-gp, ABCB1, MDR1). A growing interest has emerged in cancer chemotherapy research owing to its known overexpression in cancer cells. A wide selection of studies has provided a comprehensive overview of its functional properties, ranging from the atomic scale to clinical studies. In general, P-gp is responsible for drug extrusion and detoxification. The P-gp overexpression in cancer cells were pointed out to increase the rate of chemotherapy agents' efflux out of the cell, leading to a significantly lower effectiveness. Similarly, multidrug resistance-associated protein 1 (MRP1, ABCC1), which is involved in the efflux of a wide variety of xenobiotics (e.g., antidepressants, anticancer agents), has been linked to drug resistance in chemotherapy. However, this role should be streamlined with respect to its cousin P-gp [45]. The structures of P-gp [48–51] and MRP1 [52–54] have been resolved in several conformations, which have initiated extensive studies, providing solid insights into (i) the ABC transport cycle and (ii) the ATP-hydrolysis catalytic cycle. Investigations on P-gp provided the basis to study other ABC transporters of importance. For example, breast cancer resistance protein (BCRP/ABCG2) has received a lot of attention. BCRP adopts type IV-folding and transports a broad range of endogenous substances. Its dysfunction has been associated with several disorders. For instance, kidney diseases, hypertension, and gout are associated with BCRP's failure because of its role in the transport of uric acid [55]. Another well-known member of the ABC family is cystic fibrosis transmembrane conductance regulator protein (CFTR). However, it is worth mentioning that CFTR is out of scope since it is a channel regulating ion conduction in and out of the cell. Interestingly, CFTR does not utilise ATP hydrolysis to drive the conformational changes. Instead, the protein works like an ATP-gated ion channel that uses the ion gradient to move substrates down its electrochemical gradient. Several mutations and Single

Nucleotide Polymorphisms (SNPs) lead to channel dysfunctions, which are directly associated with cystic fibrosis [56]. The disorder can be extremely dangerous for health or even life since it limits lung function and favours severe infections.

I.3.2.2. Transport mechanism

ABC transporters use the energy derived from the hydrolysis of ATP to ADP to drive conformational changes leading to the transport cycle. Briefly, the ABC transport cycle is expected to start with either substrate binding and then ATP molecules or vice versa. The cytosolic nucleotide binding domains (NBDs), which contain conserved ATPase domains, undergo dimerization thanks to the presence of ATP and Mg^{2+} ions. This leads to the opening of the extracellular gate, favouring substrate release from the extracellular side. The open, outward-facing conformation has been suggested to be a transient state since spontaneous closure occurs in the absence of substrate to preclude substrate re-entry. NBDs are expected to finally split after ATP hydrolysis and phosphate release. The transporter opens the intracellular gate in order to get back to the initial IF conformation [47,57]. It is worth mentioning that such transport cycle is still under debate, according to ABC subfamily.

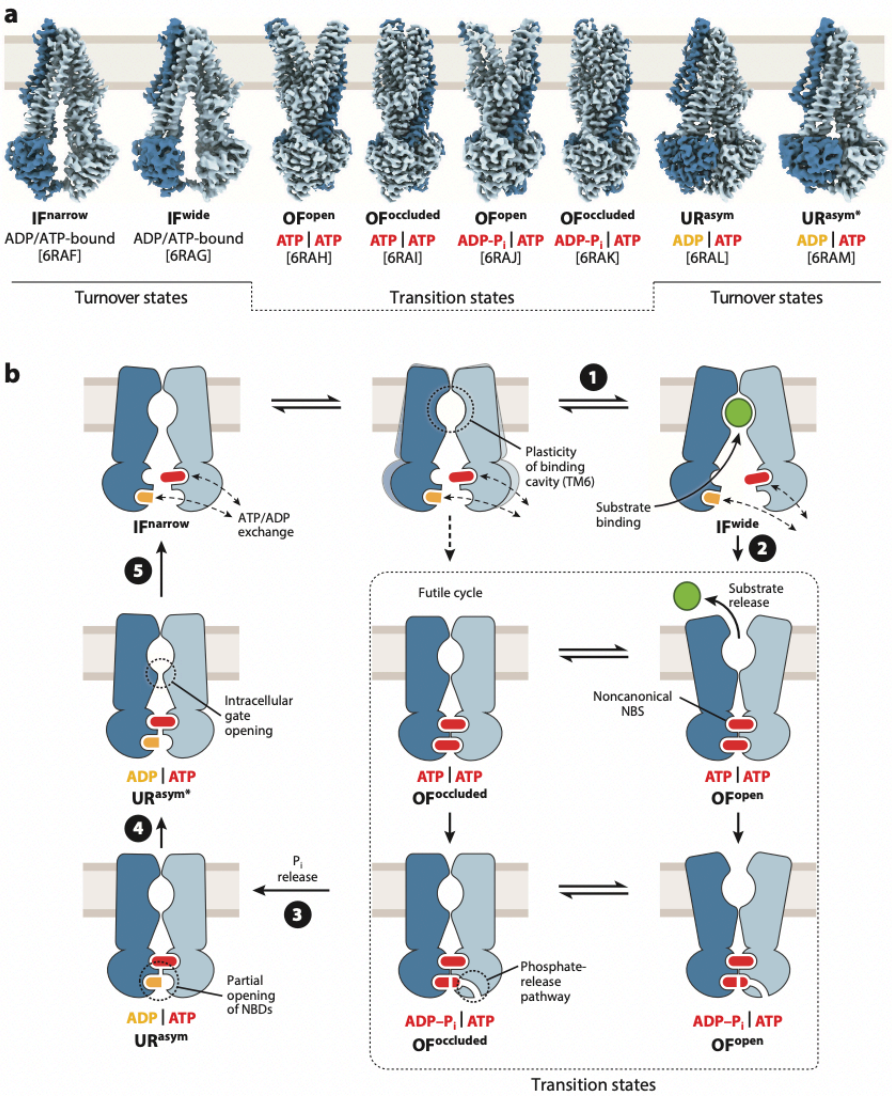


Figure 16. The structures (A) and mechanistic details of transport cycle (B) of TmrAB (*Thermus thermophilus)* multidrug-resistance proteins A and B). The unlocked-return (UR) are the post-hydrolysis conformations. Figure adapted from *Structural and Mechanistic Principles of ABC Transporters* [47].

The ABC transporters share structural patterns that consist of (i) NBDs and (ii) transmembrane domains that serve as translocating substrates. In NBDs, several motifs are highly conserved, namely the A-loop, walker A (P-loop) and walker B, the signature motif (C-loop), the A-, D-, and Q-loops, and the H-switch. The P-loops bind the phosphates of the ATP and Mg^{2+}, while the A-loop stabilises the ATP aromatic moiety by means of π–π interactions. A catalytic reaction occurs via the catalytic glutamate located in the walker B motif, which strongly polarises water molecules during the ATP hydrolysis. Two ATP molecules are sandwiched between the two NBDs. The dissociation of NBD dimers has been suggested to carry a signal to the transmembrane domain through coupling helices that structurally conserve elements along the ABC transporters [47,58].

I.3.3. Solute Carriers

I.3.3.1. Overview and clinical significance

Solute carrier (SLC) transporters play a significant role in numerous cellular and whole-body homeostasis-related processes. Since SLCs transport metabolites, xenobiotics, and ions across membranes, they are physiologically and pharmacologically relevant. Their function is tightly related to maintaining the electrochemical gradients, transitions of chemical and electrical signals, and equilibrium of the cell volume. There are more than 420 SLCs divided into 65 protein families. Numerous members were shown to be associated with numerous diseases, including cancer, diabetes, and disorders of the central nervous system [27,43,59]. For instance, numerous SLCs have been associated with monogenic diseases [59] that are caused by specific gene mutations that may cause loss-of-function transporters. Several transporters that translocate amino acids, glucose, metals, vitamins, neurotransmitters, or hormones have been associated with Mendelian diseases. For example, inflammatory bowel disease has been associated with dysfunction of both SLC22A4 and SLC22A5. Similarly, dysfunctions of SLCO1B1 and SLCO1B3 were associated with rotor syndrome, which is characterised by a relatively high concentration of bilirubins in the blood that causes yellowing of the skin and whites of the eyes [59]. However, only a few families have been depicted as pharmacologically significant drug targets.

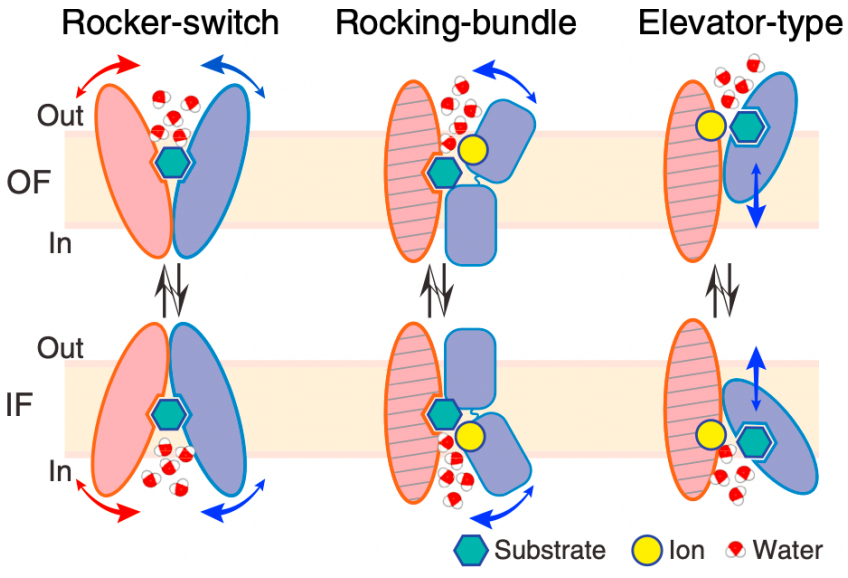


Figure 17. The mechanisms of alternating access model. Figure adapted from *Computational Dissection of Membrane Transport at a Microscopic Level* [44].

I.3.3.2. Transport mechanism

At the molecular level, different mechanisms have incorporated the basic ideas of alternating access for SLCs. Surprisingly, it was proposed that SLCs exist in closed states in which neither gate is open, and the substrate is trapped inside the cavity. *The rocker-switch*, *rocking-bundle,* and *elevator-type* are depicted in Figure 17 [44].

Rocker-switch, *rocking-bundle,* and *elevator-type* are the main mechanism variances of the alternating access model for SLCs. Rocker-switch and rocking-bundle mechanisms involve rearrangement domains for the access of the binding cavity from either intra- or extracellular side. In the *rocker-switch* model, the domains act symmetrically, which is depicted by the resolved structures of glucose transporter (GLUT) 5 in OF and IF conformation (PDB 4YBQ, 4YB9 [60]).

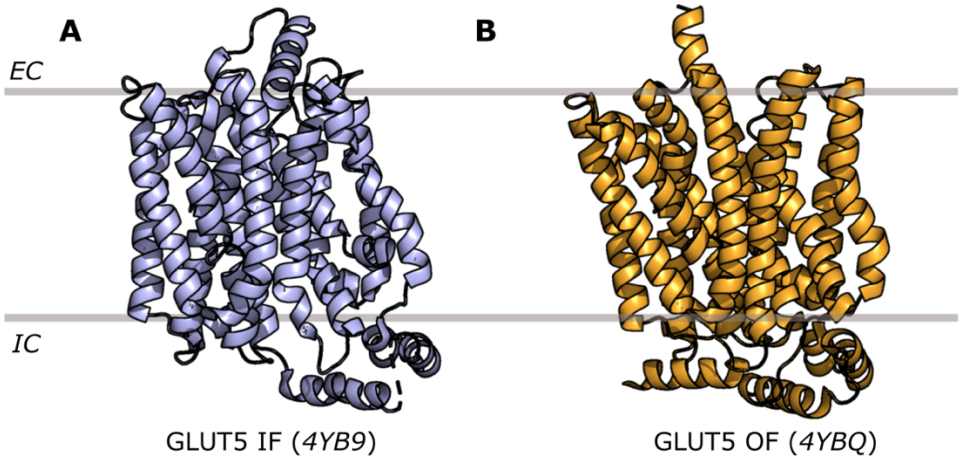


Figure 18. The inward- (A) and outward-facing (B) conformation of GLUT5.

In the rocking-bundle model, the domains are different from each other, where usually one domain is more flexible along the transport cycle. This mechanism has been suggested for LeuT transport [61], where the N bundle is stable while the C bundle undergoes large conformational changes. In the elevator mechanism, only the C-bundle transports molecules, while the N-bundle is fixed to oligomers (Fig. 18) [62].

I.3.4. Major Facilitator Superfamily

I.3.4.1. On the physiological and pharmacological importance of MFS

Major Facilitator Superfamily comprises one of the largest branches of SLC transporters. They are widely spread throughout tissues, translocating xenobiotics and endogenous substances. Their members have been found in yeast, bacteria, plants, and animals. MFS proteins comprise 16 different families. However, a particular interest has been paid to SLC2 family that includes sugar porters (SP) and SLC22, comprising of organic anion/cation/zwitterion transporters (OAT, OCT), since they are also involved in xenobiotic translocation. Their importance has been emphasized by the International Transporter Consortium (ITC) as well as medicine agencies (e.g., Food & Drug Administration – FDA – or European Medicine Agency – EMA) which call for investigating drug-transporter interactions in the drug development process [5]. MFS transporter dysfunction may be associated with metabolic disorders as well as cancer. For instance, the SLC2 family constitutes one of the main glucose

transporters maintaining the homeostasis of sugars and urate. Their dysfunctions are associated with diabetes type II (GLUT2/SLC2A2 and GLUT4/SLC2A4), hyperuricaemia, and gout (SLC2A9). GLUT1 (SLC2A1) is associated with cancer [63,64]. The SLC22 family is associated with hyperuricaemia (SLC22A11 and SLC22A12), prostate cancer (SLC22A3), or coronary heart disease (SLC22A3), to name a few [65,66].

The prototypical proteins of MFS are glucose transporters (GLUTs) that play a fundamental role in regulating sugar in the body. Even though the substrate range can be nearly identical, different isoforms bind substrates with varying affinities. The well-defined class I of glucose transporters is comprised of GLUT1-4. The expression of different GLUT isoforms is specific to preferential tissues. GLUT1 is expressed primarily in the brain, red blood cells, and foetal tissues, GLUT2 in the liver, GLUT3 in neurons, and GLUT4 in muscle cells. Moreover, GLUT1 is often upregulated in tumour tissues [67]. Therefore, the elucidation of their mechanistic insights into the transport function is of utmost importance. The GLUT family includes one of the few proteins that has been resolved in several conformational states [68–72]. In other words, given the MFS structural pattern (see below), GLUTs can be used in structural investigations as prototypes for MFS transporters.

Besides, Organic Anion Transporters (OATs) have received a lot of attention in pharmacology. OAT1 is the prototypical representative of the SLC22 family and, together with OAT3, they are essential for drug elimination [17,65]. OAT1 transports metabolites, toxins, and drugs, including antivirals, antibiotics, diuretics, statins, chemotherapeutic agents, antihypertensive, and anti-inflammatory agents [73]. Studies using OAT1-knocked out mice have demonstrated that the transport of the above-mentioned drugs was significantly impaired [74]. OAT1 and OAT3 partially share a common substrate range, so if one is dysfunctional, usually the other can partially compensate. Even though their dysfunctions have physiological impacts, the whole transporter balance prevents the existence of OAT-related pathological conditions. From the pharmacological point of view, xenobiotics' toxicities have been associated with transport dysfunctions from OAT1 and OAT3 SNPs. This is the case of e.g., methotrexate, non-steroidal anti-inflammatory drugs [ref] or mercury [ref]. Therefore, OATs have attracted a great deal of attention, mainly owing to the relationships between PGx and DDI [5,75].

I.3.4.2. Topology

Even though MFS transporters display rather low sequence conservation within the same family, they share structural patterns known as the MFS canonical fold. MFS protein structure consists of 12 transmembrane helices (TMH), which are organised into two N- and C-bundles. Additionally, MFS proteins have extra- and intracellular loops, which can usually be glycosylated or phosphorylated. The transmembrane helices are pseudo-symmetrically arranged to favour inter-TMH interactions that are particular for (i) given conformation or (ii) substrate binding. This non-coincidental arrangement forms 3 sets of functional helices. A-helices are known for their role in drug binding. B-helices are responsible for the rocking movement of the bundles along the transport cycle, and C-helices were suggested to maintain transporter integrity by interacting with the surrounding lipid bilayer [62,76]. However, there exist MFS proteins that slightly differ from the canonical fold, for instance, having 14 or 11 TMH or lacking an extracellular loop.

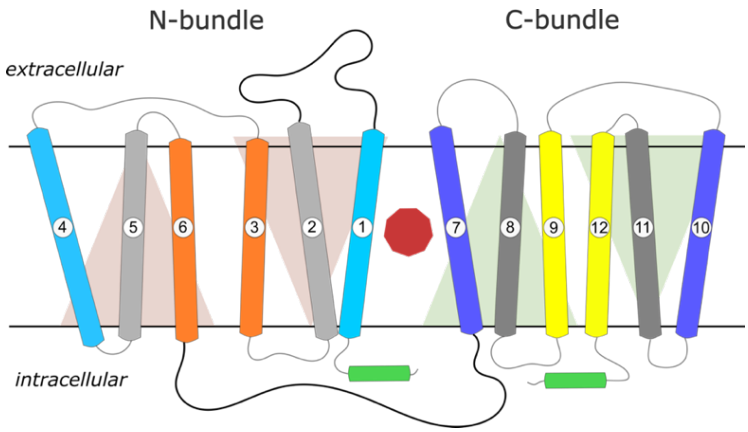


Figure 19. The canonical MFS fold is organised into N- and C-bundles as well as extra and intracellular loops. Each bundle is arranged into 3+3 inverted TMH repeats that complementarily interact with the respective helix. A-helices (blueish), B-helices (grey) and C-helices (yellow and orange) constitute functional helices. Figure adapted from *Insights into the structure and function of the human organic anion transporter 1 in lipid bilayer membranes* [77].

I.3.4.3. Transport

Regardless of large-scale conformational changes, it is important to note that some MFS can be involved in the translocation of another substrate during the transport cycles, defined by antiporter and symporter depending on if the co-substrate is transported in the opposite or same direction as the substrate, respectively [78].

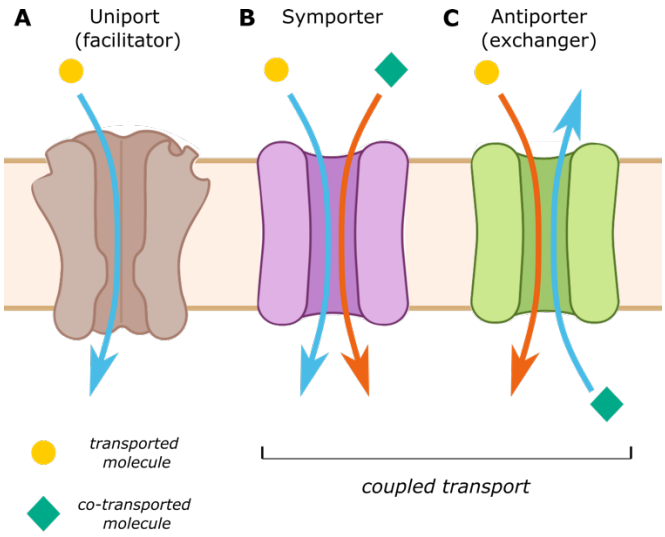


Figure 20. Members of the Major Facilitator Superfamily do rely directly on ATP hydrolysis; however, the produced electrochemical gradient drives the transport. As secondary active transporters, they utilize concentration gradient (A, facilitator) or energy released from downhill transport of one substrate to drive translocation of another substrate in the same (symporter) or opposite (antiporter) direction.

The large conformational changes during rocker-switch occur in a symmetric manner. The bundles in the rocker-switch mechanism move symmetrically around the binding spot throughout the transport cycle, exposing the cavity to either the intracellular or extracellular side, in contrast to the rocking-switch mechanism, in which only one bundle is significantly flexible. The opening of the cavity leans on the A-helices (Fig. 19), TMH 1, 7 on the extracellular side and TMH 4 and 10 on the intracellular side [62]. Non-covalent interactions are involved in

the *gating event* upon substrate binding. The concept of gating in MFS transporters refers to residues or helices that maintain non-covalent interactions between N and C bundles, precluding substrate release. For instance, non-covalent interactions between aforementioned the A-helices can be formed, leading to the so-called occluded conformation where neither gate is open (or at least partially open). The large conformational changes that lead from the IF to OF state are maintained by B-helices, (TMH 5 and 8, and TMH 2 and 11; Figure 19, greyish) that stand directly on the N and C bundle interface [76].

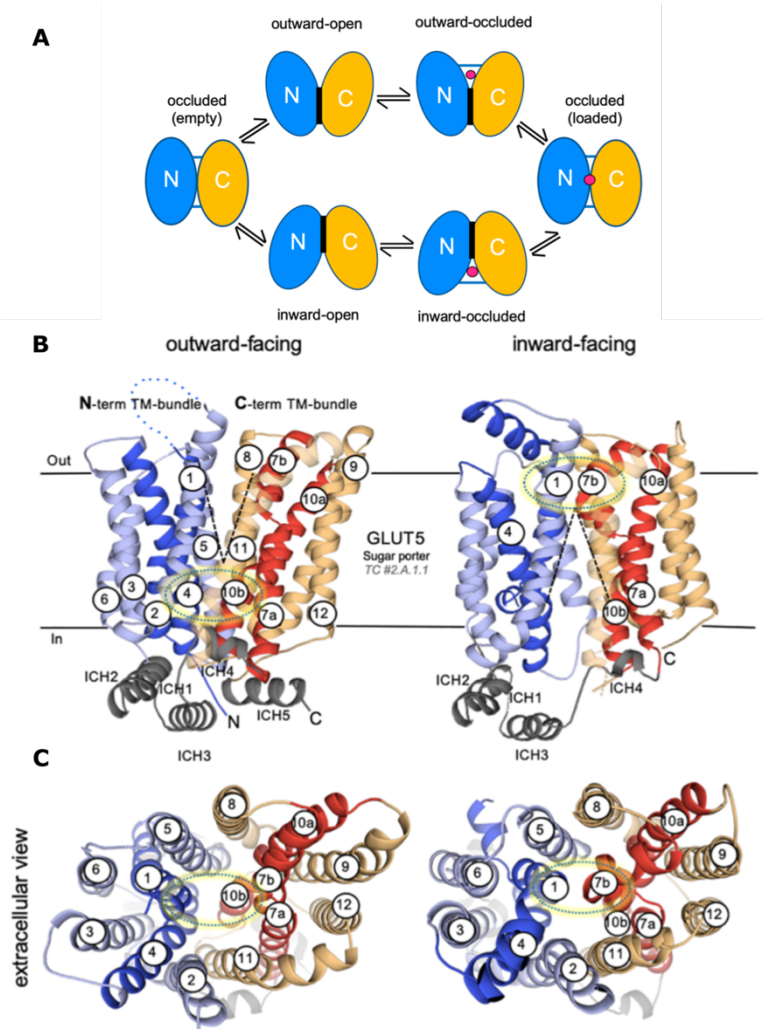


Figure 21. The schematic visualisation of conformations appeared during the transport cycle (A). GLUT5 is presented in OF (left) and IF (right) conformation (B). The gating event that leads to occlusion is maintained by TMH1 and 7 on the extracellular side and TMH4 and 10 on the intracellular side (C). The figure has been adapted from *Structures and General Transport Mechanisms by the Major Facilitator Superfamily (MFS)* [62].

I.3.4.4. On the importance of intracellular motifs in MFS

Not only does the transmembrane domain play a central role in the MFS transport cycle. Intracellular domains are typically made up of α-helices (ICHs) and loops that form a tightly interacting network to maintain a specific conformation throughout the transport cycle. Since the intracellular domains of MFS proteins contain highly conserved motifs, the pattern of interactions between MFS proteins is highly conserved. These motifs are distributed pseudo-symmetrically in the N and C bundles and are made of charged and polar residues favouring salt-bridge and H-bond interactions. The A-motif is placed on the intracellular side between

TMH2 and 3 and TMH8 and 9 in the N and C bundles, respectively. The charged residues of the A-motif interact with the so-called PETL (which may differ from one MFS member to another) and E[X_6]R motifs in each bundle [79,80]. Since salt-bridges are specific to a given conformation, it is presumed that the network break during the transport cycle can tune the energy barrier associated with conformational changes [72].

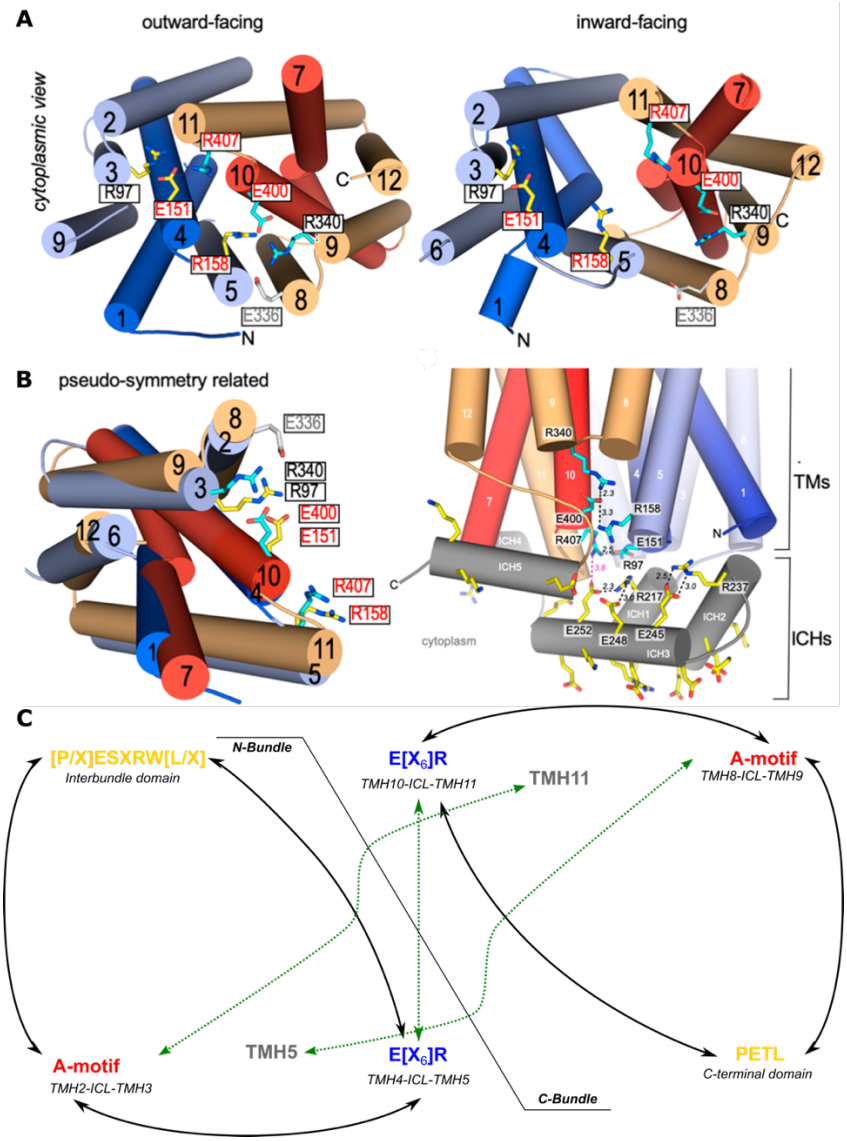


Figure 22. The intracellular interactions of motifs in MFS proteins are highly conserved. (A) The interactions are specific to the OF (left) and IF (right) conformations. In addition, the residues and their interactions along the N and C bundles exhibit pseudo-symmetry, which becomes clear when the bundles are aligned (B). A map of interactions (C) depicts typical interactions in the OF state, while the green lines indicate the missing interactions in the IF state. Adapted from *Structure and mechanism of the mammalian fructose transporter GLUT5* [60] and *Insights into the structure and function of the human organic anion transporter 1 in lipid bilayer membranes* [77].

I.4. Lipid membrane bilayers

Biological membranes constitute the boundaries of cells that separate the interior from the extracellular compartments of cells. Membranes also enclose organelles within the cell, such as mitochondria and the nucleus. The presence of membranes is also responsible for ion and small molecule gradients between the inside and outside of membranes, which drive a variety of physiological processes [12,45].

I.4.1. Membrane composition

Membranes are heterogeneous, containing a variety of lipids that vary in charge, saturation degree, and size. Cell membranes also contain proteins, including transporters or receptors, as well as structural proteins that maintain cell and membrane integrities. The composition of a membrane varies across different tissues, cell types, and species. Since the plasma membrane is dynamic and fluid, lipids move freely in the plane of the membrane (so-called lateral diffusion). The types of membrane components drive membrane permeability to small molecules, membrane fluidity, and interactions between lipids and proteins. The membrane is mostly made of amphiphilic lipids (e.g., phospholipids) with hydrophilic polar heads and hydrophobic acyl chains, arranged in a lipid bilayer in which lipid tails face the interior of the membrane while polar heads are exposed to either intracellular or extracellular compartments [45].

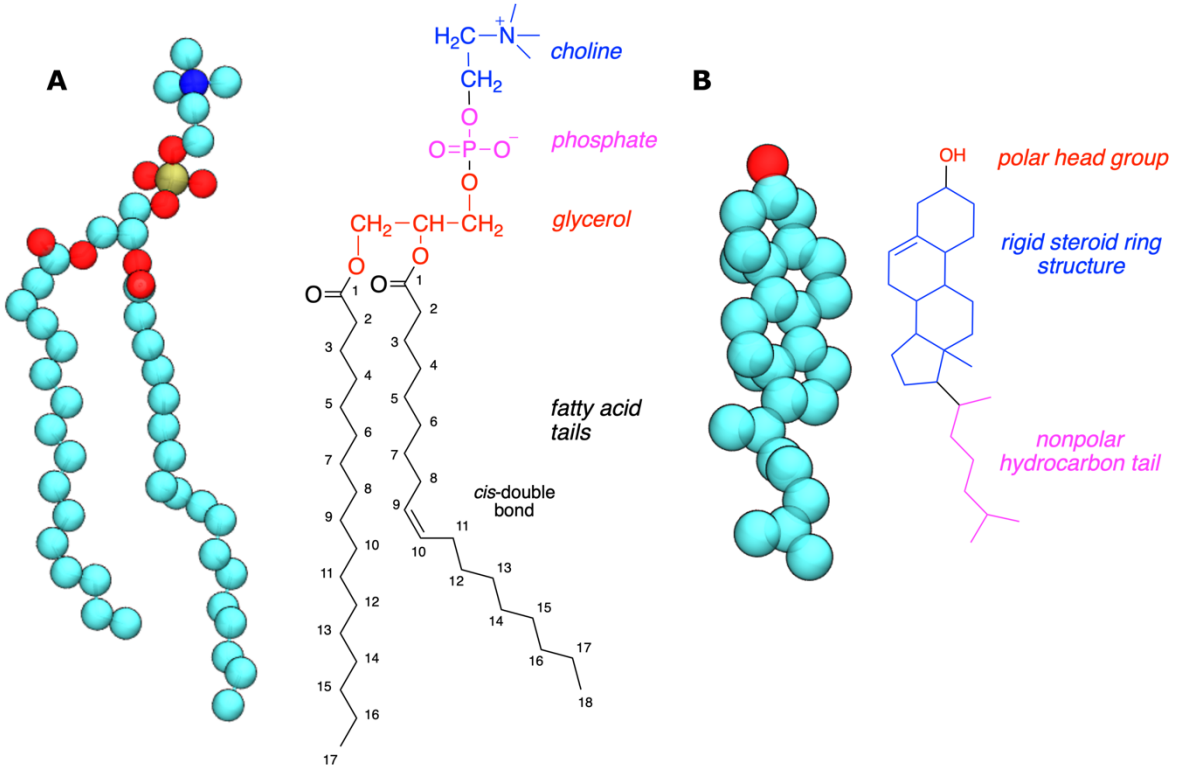


Figure 23. Common components of lipid bilayers, phospholipid (A) and sterol (B). Phosphoglycerides (A) are built from the polar head that includes choline, phosphate, and glycerol, to which are connected fatty acid chains. Sterols (B) have a hydroxyl group that represents the polar head and a hydrophobic part made of steroid ring structure and a hydrocarbon tail.

Lipids such as phospholipids, sterols, and sphingolipids are the main components of lipid bilayers, however, the ratio of components depends on the types of cells and tissues. Phospholipids contain a phosphate group as the polar head and two hydrocarbon tails that are

adjacent (namely sn-1 and sn-2 chains). Typically, hydrocarbon tails consist of fatty acids with varying lengths and saturation levels. Unsaturated fatty acids typically contain *cis* double bonds that kink the acyl chain. The length and saturation of acyl chains influence the packing of lipids and thus the membrane's fluidity [45,81,82].

Phosphoglycerides (also referred to as glycerophospholipids, or GPLs) are the most common phospholipids and represent nearly 75% of the biological membrane components. They are derived from glycerol in which one O-atom is bound to a phosphate group (polar head) and the other two connect the acyl tails via ester bonds. Polar heads differ in size and charge. *Phosphatidylethanolamine*, *phosphatidylserine*, and *phosphatidylcholine* are examples of common glycerophospholipids. Sphingolipids, which are also abundant in the lipid membrane, derive from sphingosine with an amine and two hydroxyl groups. Finally, cholesterol is the second most abundant lipid in membranes after phospholipids. Its structure is drastically different as compared to other lipids since it shares steroid structural patterns made of four fused rings bearing a hydroxyl group on one side and a short hydrophobic tail on the other side [45,81].

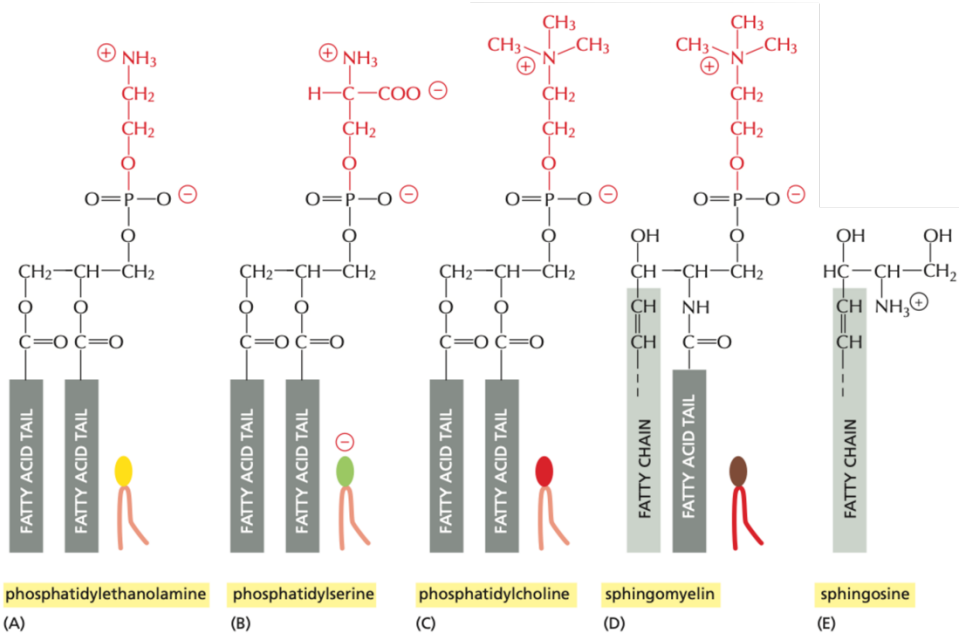


Figure 24. Polar heads (red) of phospholipids abundantly present found in lipid bilayers. (A-D) represent glycerophospholipids, while (D) derive from sphingosine and (E) is sphingolipid. Adapted from *Molecular Biology of the cell* [45].

The hydrophobic part of lipids is constituted by fatty acids that can be saturated or polyunsaturated (PUFA) and usually have from 12 to 24 carbons. Saturated fatty acids favour higher lipid order, leading to more rigid structure. In contrast, unsaturated fatty acids are more flexible, favouring more fluid lipid bilayer membranes. Nearly all natural unsaturated fatty acids adopt a cis conformation that generates a kink in the structure. However, since the cis conformation is less stable than the trans conformation, a fatty acid can undergo isomerisation

under certain conditions, for instance a chemical agent [83]. Having a variety of lipid chains and polar heads, the phospholipids display a spectrum of naturally occurring lipids.

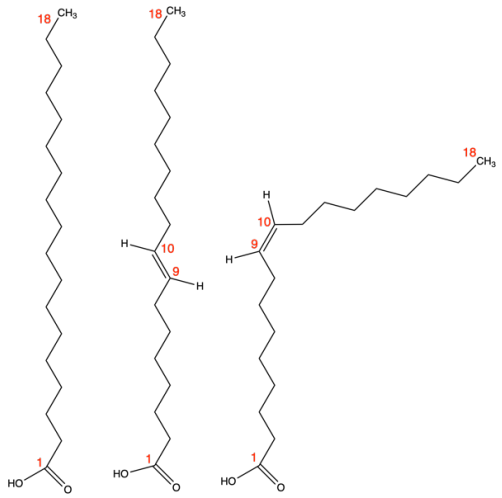


Figure 25. Examples of fatty acid chains of 18 carbons in form (from left) saturated (stearic acid) and unsaturated in trans (elaidic acid) and cis (oleic acid) conformation.

I.4.2. Membrane structure

Biophysical experiments are of utmost importance in order to decipher the role of lipid structure in membrane structure. Basically, in a pure bilayer membrane, lipid shape drives the self-assembly structure of the lipid bilayer.

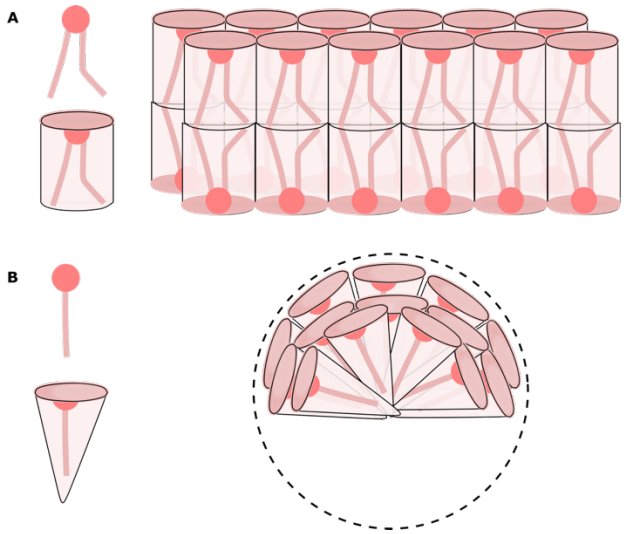


Figure 26. Lipids with a cylindrical shape are more likely to form a lipid bilayer (A), while lipids with a conical shape lipids form micelles (B).

From these experiments, trends can be defined. (Un)saturation of lipid tails participate in driving membrane properties such as thickness and fluidity. Unsaturated lipid tails favour a thinner and more fluid membrane. Membranes are known to exist in several states, such as liquid ordered, disordered, or gel state, depending on lipid tail length and saturation as well as temperature. *cis*-Unsaturated lipids introduce a kink that creates packing defects (voids) in the membrane, resulting in increased flexibility and fluidity of the membrane. For instance, even at low temperatures, if the lipid chains are short and unsaturated, the membrane can remain fluid. The role of cholesterol is of utmost importance in the native cell membrane. Cholesterol

regulates membrane fluidity in a cholesterol: phospholipid mixture, regardless of the nature of lipid tails. For example, cholesterol interlocks slightly below the phospholipid polar head region where the steroid structure is in contact with carbons 1 to 9 of the lipid tails. Since the double bonds occur frequently in that region, cholesterol can sterically "block" motions of unsaturated lipid chains, decreasing lipid bilayer fluidity of pure unsaturated lipid bilayer membrane. In other words, cholesterol can either increase or decrease the rigidity of lipid bilayer membranes: the gel phase lipid bilayer becomes more fluid in the presence of cholesterol, whereas the fluid lipid bilayer becomes more ordered. If lipids are densely packed, cholesterol prevents membrane crystallisation and thus gel phase [45,81].

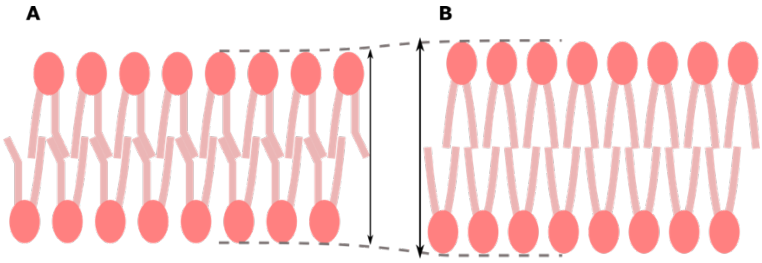


Figure 27. The thickness of a membrane containing unsaturated (A) and saturated (B) lipid chains.

Among other parameters, membrane fluidity may be assessed by the lipid order parameter, which measures the angle between carbon and adjacent hydrogens in $-CH_2-$ or $-CH_3$ moieties of hydrocarbon tails with respect to the membrane normal. The larger the lipid order parameter (S) the more ordered the membrane:

$$S = \frac{1}{2}(3 < cos^2\theta > -1)$$

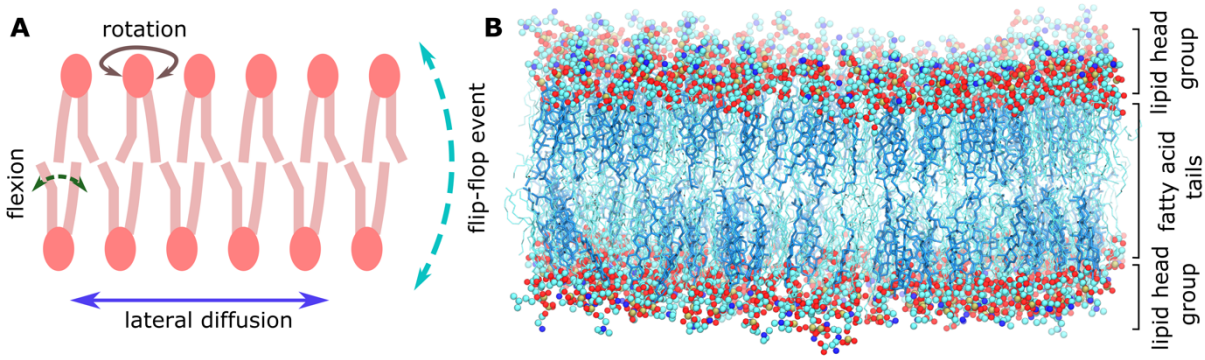


Figure 28. Membrane lipid bilayer and its dynamics.

Lipids in membranes are dynamic and migrate along the plane of the membrane. Typically, their motions occur only within the leaflet plane via so-called lateral diffusion, which can be calculated by computational methods or assessed experimentally [84,85]. In addition, the lipid dynamics include lipid rotation around their vertical axis. To maintain the equilibrium between the leaflets, rare instances of "flip-flop" migration of lipids from one leaflet to another may occur, especially for cholesterol, which undergoes the event spontaneously. Regarding phospholipids translocation occurs via proteins known as flippase, floppase, or scramblase [86]. It is worth mentioning that some ABC transporters are known to be involved in such processes. It is indeed essential to recognise that the membrane lipid bilayer exhibits a natural asymmetry between leaflets. Biologically speaking, flip-flop events may be of importance since cell membrane composition and lipid distribution in leaflets can be associated with cell events. For

example, the presence of phosphoserine lipids in the outer leaflets may be a fingerprint for cell apoptosis [45,87].

I.4.3. Lipid – protein interaction

The composition and asymmetry of membranes appear to be of particular importance for lipid-protein interactions. By playing a role in protein functions, annular lipids modulate membrane properties such as fluidity and packing around the protein. In another type of interaction, allosteric effect of lipids is involved in transporter function modulation. Membrane components can mediate oligomerisation given that some proteins function as oligomers. Since certain proteins mediate lipid transport, lipids may be substrates (e.g., phosphatidylcholine, ABCB4/MDR3 [88,89]) or promote inhibition of proteins (enhanced inhibition effect in ABCB1/P-gp [90]).

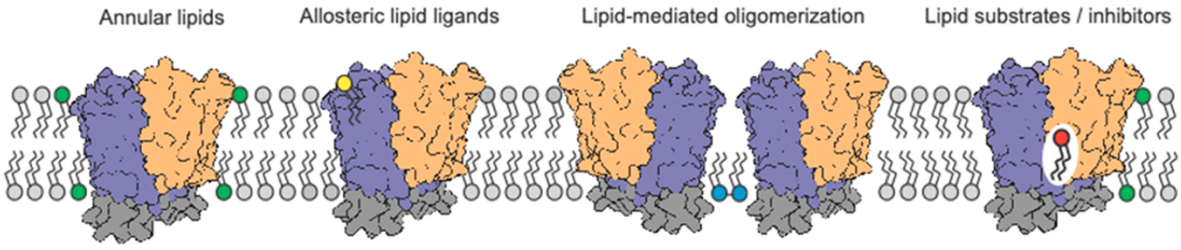


Figure 29. Types of lipid-protein interactions. Figure adapted from *Structures and General Transport Mechanisms by the Major Facilitator Superfamily (MFS)* [62].

In many MFS transporters, the presence of phosphatidylethanolamine (PE) has been highlighted as crucial for the proper transporter function [62,91–93]. For example, PE components are crucial for proper folding of LacY transporter in *E. coli* [94–96]. Moreover, it is suggested that PE components facilitate the transport cycle of MFS proteins by stabilising OF conformation and then destabilising interactions that promote IF conformation [93]. In the case of ABC proteins, however, several transporters have been resolved with cholesterol bonds, indicating a possible active role in transporter function [52,54,97].

I.5. References

[1] L.H. Goetz, N.J. Schork, Personalized medicine: motivation, challenges, and progress, Fertility and Sterility. 109 (2018) 952–963. https://doi.org/10.1016/j.fertnstert.2018.05.006.

[2] L. Hood, L. Rowen, The human genome project: big science transforms biology and medicine, Genome Med. 5 (2013) 79. https://doi.org/10.1186/gm483.

[3] 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. Wah Yee, M.J. Zamek-Gliszczynski, L. Zhang, The International Transporter Consortium, Membrane transporters in drug development, Nature Reviews Drug Discovery. 9 (2010) 215–236. https://doi.org/10.1038/nrd3028.

[4] M.J. Zamek-Gliszczynski, C.A. Lee, A. Poirier, J. Bentz, X. Chu, H. Ellens, T. Ishikawa, M. Jamei, J.C. Kalvass, S. Nagar, K.S. Pang, K. Korzekwa, P.W. Swaan, M.E. Taub, P. Zhao, A. Galetin, ITC Recommendations for Transporter Kinetic Parameter Estimation and Translational Modeling of Transport-Mediated PK and DDIs in Humans, Clin Pharmacol Ther. 94 (2013) 64–79. https://doi.org/10.1038/clpt.2013.45.

[5] M.J. Zamek-Gliszczynski, M.E. Taub, P.P. Chothe, X. Chu, K.M. Giacomini, R.B. Kim, A.S. Ray, S.L. Stocker, J.D. Unadkat, M.B. Wittwer, C. Xia, S.-W. Yee, L. Zhang, Y. Zhang, International Transporter Consortium, Transporters in Drug Development: 2018 ITC Recommendations for Transporters of Emerging Clinical Importance, Clin. Pharmacol. Ther. 104 (2018) 890–899. https://doi.org/10.1002/cpt.1112.

[6] K. Maeda, Y. Sugiyama, Impact of Genetic Polymorphisms of Transporters on the Pharmacokinetic, Pharmacodynamic and Toxicological Properties of Anionic Drugs, Drug Metabolism and Pharmacokinetics. 23 (2008) 223–235. https://doi.org/10.2133/dmpk.23.223.

[7] A. Schlessinger, M.A. Welch, H. van Vlijmen, K. Korzekwa, P.W. Swaan, P. Matsson, Molecular Modeling of Drug-Transporter Interactions-An International Transporter Consortium Perspective, Clin. Pharmacol. Ther. 104 (2018) 818–835. https://doi.org/10.1002/cpt.1174.

[8] B.G. Katzung, S.B. Masters, A.J. Trevor, Basic & clinical pharmacology, 12th ed, McGraw-Hill Medical ; McGraw-Hill [distributor], New York, London, 2012.

[9] L.S. Goodman, A. Gilman, L.L. Brunton, J.S. Lazo, K.L. Parker, eds., Goodman & Gilman's the pharmacological basis of therapeutics, 11th ed, McGraw-Hill, New York, 2006.

[10] H.P. Rang, M.M. Dale, eds., Rang and Dale's pharmacology, 8. ed, Elsevier, Churchill Livingstone, Amsterdam Heidelberg, 2016.

[11] G.M. Currie, Pharmacology, Part 2: Introduction to Pharmacokinetics, J. Nucl. Med. Technol. 46 (2018) 221–230. https://doi.org/10.2967/jnmt.117.199638.

[12] 3.4 The Cell Membrane - Concepts of Biology | OpenStax, (n.d.). https://openstax.org/books/concepts-biology/pages/3-4-the-cell-membrane.

[13] C. Zihni, C. Mills, K. Matter, M.S. Balda, Tight junctions: from simple barriers to multifunctional molecular gates, Nat Rev Mol Cell Biol. 17 (2016) 564–580. https://doi.org/10.1038/nrm.2016.80.

[14] S.K. Bardal, J.E. Waechter, D.S. Martin, Pharmacokinetics, in: Applied Pharmacology, Elsevier, 2011: pp. 17–34. https://doi.org/10.1016/B978-1-4377-0310-8.00002-6.

[15] K. Wang, B. Kestenbaum, Proximal Tubular Secretory Clearance: A Neglected Partner of Kidney Function, CJASN. 13 (2018) 1291–1296. https://doi.org/10.2215/CJN.12001017.

[16] D.R. Taft, Drug Excretion, in: Pharmacology, Elsevier, 2009: pp. 175–199. https://doi.org/10.1016/B978-0-12-369521-5.00009-9.

[17] S.K. Nigam, W. Wu, K.T. Bush, M.P. Hoenig, R.C. Blantz, V. Bhatnagar, Handling of Drugs, Metabolites, and Uremic Toxins by Kidney Proximal Tubule Drug Transporters, CJASN. 10 (2015) 2039–2049. https://doi.org/10.2215/CJN.02440314.

[18] A.K. Nigam, J.G. Li, K. Lall, D. Shi, K.T. Bush, V. Bhatnagar, R. Abagyan, S.K. Nigam, Unique metabolite preferences of the drug transporters OAT1 and OAT3 analyzed by machine learning, J. Biol. Chem. 295 (2020) 1829–1842. https://doi.org/10.1074/jbc.RA119.010729.

[19] G. Levy, Pharmacokinetics of Salicylate Elimination in Man, Journal of Pharmaceutical Sciences. 54 (1965) 959–967. https://doi.org/10.1002/jps.2600540703.

[20] K. Bachmann, Drug Metabolism, in: Pharmacology, Elsevier, 2009: pp. 131–173. https://doi.org/10.1016/B978-0-12-369521-5.00008-7.

[21] T. Lynch, A. Price, The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects, Am Fam Physician. 76 (2007) 391–396.

[22] C.J. Dickson, V. Hornak, R.A. Pearlstein, J.S. Duca, Structure–Kinetic Relationships of Passive Membrane Permeation from Multiscale Modeling, J. Am. Chem. Soc. 139 (2017) 442–452. https://doi.org/10.1021/jacs.6b11215.

[23] S.K. Nigam, What do drug transporters really do?, Nat Rev Drug Discov. 14 (2015) 29–44. https://doi.org/10.1038/nrd4461.

[24] The International Transporter Consortium, Membrane transporters in drug development, Nat Rev Drug Discov. 9 (2010) 215–236. https://doi.org/10.1038/nrd3028.

[25] A. Ivanyuk, F. Livio, J. Biollaz, T. Buclin, Renal Drug Transporters and Drug Interactions, Clin Pharmacokinet. 56 (2017) 825–892. https://doi.org/10.1007/s40262-017-0506-8.

[26] R. Ho, R. Kim, Transporters and drug therapy: Implications for drug disposition and disease, Clinical Pharmacology & Therapeutics. 78 (2005) 260–277. https://doi.org/10.1016/j.clpt.2005.05.011.

[27] R.-A.A. Garibsingh, A. Schlessinger, Advances and Challenges in Rational Drug Design for SLCs, Trends in Pharmacological Sciences. 40 (2019) 790–800. https://doi.org/10.1016/j.tips.2019.08.006.

[28] J.R. Vane, R.M. Botting, The mechanism of action of aspirin, Thrombosis Research. 110 (2003) 255–258. https://doi.org/10.1016/S0049-3848(03)00379-7.

[29] Z. Cournia, A. Chatzigoulas, Allostery in membrane proteins, Current Opinion in Structural Biology. 62 (2020) 197–204. https://doi.org/10.1016/j.sbi.2020.03.006.

[30] M. Marino, Z. Jamal, P.M. Zito, Pharmacodynamics, in: StatPearls, StatPearls Publishing, Treasure Island (FL), 2022. http://www.ncbi.nlm.nih.gov/books/NBK507791/ (accessed October 7, 2022).

[31] H. Zou, P. Banerjee, S.S.Y. Leung, X. Yan, Application of Pharmacokinetic-Pharmacodynamic Modeling in Drug Delivery: Development and Challenges, Front. Pharmacol. 11 (2020) 997. https://doi.org/10.3389/fphar.2020.00997.

[32] B. Franck, C. Monchaud, F. Saint-Marcoux, J.-P. Rérolle, J. Allard, V. Allot, P. Marquet, M. Essig, J.-B. Woillard, Population pharmacokinetics of gentamicin in haemodialysis patients: modelling, simulations and recommendations, European Journal of Clinical Pharmacology. 76 (2020) 947–955. https://doi.org/10.1007/s00228-020-02867-3.

[33] J.T. Brown, J.R. Bishop, M.E. Schneiderhan, Using pharmacogenomics and therapeutic drug monitoring to guide drug selection and dosing in outpatient mental health

comprehensive medication management, Mental Health Clinician. 10 (2020) 254–258. https://doi.org/10.9740/mhc.2020.07.254.

[34] J.-P. Gies, Y. Landry, DRUG TARGETS: MOLECULAR MECHANISMS OF DRUG ACTION, in: The Practice of Medicinal Chemistry, Elsevier, 2003: pp. 51–65. https://doi.org/10.1016/B978-012744481-9/50008-8.

[35] P. Cohen, Protein kinases — the major drug targets of the twenty-first century?, Nat Rev Drug Discov. 1 (2002) 309–315. https://doi.org/10.1038/nrd773.

[36] M. Griguoli, R. Scuri, D. Ragozzino, E. Cherubini, Activation of nicotinic acetylcholine receptors enhances a slow calcium-dependent potassium conductance and reduces the firing of stratum oriens interneurons, European Journal of Neuroscience. 30 (2009) 1011–1022. https://doi.org/10.1111/j.1460-9568.2009.06914.x.

[37] J. Li, Y. Ning, W. Hedley, B. Saunders, Y. Chen, N. Tindill, T. Hannay, S. Subramaniam, The Molecule Pages database, Nature. 420 (2002) 716–717. https://doi.org/10.1038/nature01307.

[38] O. Fleetwood, P. Matricon, J. Carlsson, L. Delemotte, Energy Landscapes Reveal Agonist Control of G Protein-Coupled Receptor Activation via Microswitches, Biochemistry. 59 (2020) 880–891. https://doi.org/10.1021/acs.biochem.9b00842.

[39] S.K. Nigam, K.T. Bush, V. Bhatnagar, S.M. Poloyac, J.D. Momper, The Systems Biology of Drug Metabolizing Enzymes and Transporters: Relevance to Quantitative Systems Pharmacology, Clin. Pharmacol. Ther. 108 (2020) 40–53. https://doi.org/10.1002/cpt.1818.

[40] R. Ho, R. Kim, Transporters and drug therapy: Implications for drug disposition and disease, Clinical Pharmacology & Therapeutics. 78 (2005) 260–277. https://doi.org/10.1016/j.clpt.2005.05.011.

[41] R.A. Wilke, M.E. Dolan, Genetics and Variable Drug Response, JAMA. 306 (2011). https://doi.org/10.1001/jama.2011.998.

[42] S.W. Yee, D.J. Brackman, E.A. Ennis, Y. Sugiyama, L.K. Kamdem, R. Blanchard, A. Galetin, L. Zhang, K.M. Giacomini, Influence of Transporter Polymorphisms on Drug Disposition and Response: A Perspective From the International Transporter Consortium, Clin. Pharmacol. Ther. 104 (2018) 803–817. https://doi.org/10.1002/cpt.1098.

[43] A. Schlessinger, S.W. Yee, A. Sali, K.M. Giacomini, SLC Classification: An Update, Clin Pharmacol Ther. 94 (2013) 19–23. https://doi.org/10.1038/clpt.2013.73.

[44] T. Jiang, P.-C. Wen, N. Trebesch, Z. Zhao, S. Pant, K. Kapoor, M. Shekhar, E. Tajkhorshid, Computational Dissection of Membrane Transport at a Microscopic Level, Trends in Biochemical Sciences. 45 (2020) 202–216. https://doi.org/10.1016/j.tibs.2019.09.001.

[45] B. Alberts, Molecular biology of the cell, Sixth edition, Garland Science, Taylor and Francis Group, New York, NY, 2015.

[46] L. Zhang, S.-M. Huang, L.J. Lesko, Transporter-Mediated Drug–Drug Interactions, Clin Pharmacol Ther. 89 (2011) 481–484. https://doi.org/10.1038/clpt.2010.359.

[47] C. Thomas, R. Tampé, Structural and Mechanistic Principles of ABC Transporters, Annu. Rev. Biochem. 89 (2020) 605–636. https://doi.org/10.1146/annurev-biochem-011520-105201.

[48] A. Alam, R. Küng, J. Kowal, R.A. McLeod, N. Tremp, E.V. Broude, I.B. Roninson, H. Stahlberg, K.P. Locher, Structure of a zosuquidar and UIC2-bound human-mouse chimeric ABCB1, Proc. Natl. Acad. Sci. U.S.A. 115 (2018). https://doi.org/10.1073/pnas.1717044115.

[49] K. Nosol, K. Romane, R.N. Irobalieva, A. Alam, J. Kowal, N. Fujita, K.P. Locher, Cryo-EM structures reveal distinct mechanisms of inhibition of the human multidrug transporter

ABCB1, Proc. Natl. Acad. Sci. U.S.A. 117 (2020) 26245–26253. https://doi.org/10.1073/pnas.2010264117.

[50] A. Alam, J. Kowal, E. Broude, I. Roninson, K.P. Locher, Structural insight into substrate and inhibitor discrimination by human P-glycoprotein, Science. 363 (2019) 753–756. https://doi.org/10.1126/science.aav7102.

[51] S. Urgaonkar, K. Nosol, A.M. Said, N.N. Nasief, Y. Bu, K.P. Locher, J.Y.N. Lau, M.P. Smolinski, Discovery and Characterization of Potent Dual P-Glycoprotein and CYP3A4 Inhibitors: Design, Synthesis, Cryo-EM Analysis, and Biological Evaluations, J. Med. Chem. 65 (2022) 191–216. https://doi.org/10.1021/acs.jmedchem.1c01272.

[52] L. Wang, Z.L. Johnson, M.R. Wasserman, J. Levring, J. Chen, S. Liu, Characterization of the kinetic cycle of an ABC transporter by single-molecule and cryo-EM analyses, ELife. 9 (2020) e56451. https://doi.org/10.7554/eLife.56451.

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

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

[55] R.W. Robey, K.K.K. To, O. Polgar, M. Dohse, P. Fetsch, M. Dean, S.E. Bates, ABCG2: A perspective, Advanced Drug Delivery Reviews. 61 (2009) 3–13. https://doi.org/10.1016/j.addr.2008.11.003.

[56] M. Lopes-Pacheco, CFTR Modulators: The Changing Face of Cystic Fibrosis in the Era of Precision Medicine, Front. Pharmacol. 10 (2020) 1662. https://doi.org/10.3389/fphar.2019.01662.

[57] V.V. Rybenkov, H.I. Zgurskaya, C. Ganguly, I.V. Leus, Z. Zhang, M. Moniruzzaman, The Whole Is Bigger than the Sum of Its Parts: Drug Transport in the Context of Two Membranes with Active Efflux, Chem. Rev. 121 (2021) 5597–5631. https://doi.org/10.1021/acs.chemrev.0c01137.

[58] S. Hofmann, D. Januliene, A.R. Mehdipour, C. Thomas, E. Stefan, S. Brüchert, B.T. Kuhn, E.R. Geertsma, G. Hummer, R. Tampé, A. Moeller, Conformation space of a heterodimeric ABC exporter under turnover conditions, Nature. 571 (2019) 580–583. https://doi.org/10.1038/s41586-019-1391-0.

[59] L. Lin, S.W. Yee, R.B. Kim, K.M. Giacomini, SLC transporters as therapeutic targets: emerging opportunities, Nat Rev Drug Discov. 14 (2015) 543–560. https://doi.org/10.1038/nrd4626.

[60] N. Nomura, G. Verdon, H.J. Kang, T. Shimamura, Y. Nomura, Y. Sonoda, S.A. Hussien, A.A. Qureshi, M. Coincon, Y. Sato, H. Abe, Y. Nakada-Nakura, T. Hino, T. Arakawa, O. Kusano-Arai, H. Iwanari, T. Murata, T. Kobayashi, T. Hamakubo, M. Kasahara, S. Iwata, D. Drew, Structure and mechanism of the mammalian fructose transporter GLUT5, Nature. 526 (2015) 397–401. https://doi.org/10.1038/nature14909.

[61] L.R. Forrest, Y.-W. Zhang, M.T. Jacobs, J. Gesmonde, L. Xie, B.H. Honig, G. Rudnick, Mechanism for alternating access in neurotransmitter transporters, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 10338–10343. https://doi.org/10.1073/pnas.0804659105.

[62] D. Drew, R.A. North, K. Nagarathinam, M. Tanabe, Structures and General Transport Mechanisms by the Major Facilitator Superfamily (MFS), Chem. Rev. 121 (2021) 5289–5335. https://doi.org/10.1021/acs.chemrev.0c00983.

[63] R. Augustin, The protein family of glucose transport facilitators: It's not only about glucose after all, IUBMB Life. (2010) NA-NA. https://doi.org/10.1002/iub.315.

[64] A.J. Cura, A. Carruthers, Role of Monosaccharide Transport Proteins in Carbohydrate Assimilation, Distribution, Metabolism, and Homeostasis, in: R. Terjung (Ed.), Comprehensive Physiology, 1st ed., Wiley, 2012: pp. 863–914. https://doi.org/10.1002/cphy.c110024.

[65] M. Roth, A. Obaidat, B. Hagenbuch, OATPs, OATs and OCTs: the organic anion and cation transporters of the SLCO and SLC22A gene superfamilies: OATPs, OATs and OCTs, British Journal of Pharmacology. 165 (2012) 1260–1287. https://doi.org/10.1111/j.1476-5381.2011.01724.x.

[66] H. Koepsell, The SLC22 family with transporters of organic cations, anions and zwitterions, Molecular Aspects of Medicine. 34 (2013) 413–435. https://doi.org/10.1016/j.mam.2012.10.010.

[67] N. Yan, A Glimpse of Membrane Transport through Structures—Advances in the Structural Biology of the GLUT Glucose Transporters, Journal of Molecular Biology. 429 (2017) 2710–2725. https://doi.org/10.1016/j.jmb.2017.07.009.

[68] 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. https://doi.org/10.1038/nature14655.

[69] 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. https://doi.org/10.1038/nature13306.

[70] N. Wang, S. Zhang, Y. Yuan, H. Xu, E. Defossa, H. Matter, M. Besenius, V. Derdau, M. Dreyer, N. Halland, K.H. He, S. Petry, M. Podeschwa, N. Tennagels, X. Jiang, N. Yan, Molecular basis for inhibiting human glucose transporters by exofacial inhibitors, Nat Commun. 13 (2022) 2632. https://doi.org/10.1038/s41467-022-30326-3.

[71] Y. Yuan, F. Kong, H. Xu, A. Zhu, N. Yan, C. Yan, Cryo-EM structure of human glucose transporter GLUT4, Nat Commun. 13 (2022) 2671. https://doi.org/10.1038/s41467-022-30235-5.

[72] T.F. Custódio, P.A. Paulsen, K.M. Frain, B.P. Pedersen, Structural comparison of GLUT1 to GLUT3 reveal transport regulation mechanism in sugar porter family, Life Sci. Alliance. 4 (2021) e202000858. https://doi.org/10.26508/lsa.202000858.

[73] S.K. Nigam, K.T. Bush, G. Martovetsky, S.-Y. Ahn, H.C. Liu, E. Richard, V. Bhatnagar, W. Wu, The Organic Anion Transporter (OAT) Family: A Systems Biology Perspective, Physiological Reviews. 95 (2015) 83–123. https://doi.org/10.1152/physrev.00025.2013.

[74] W. Wu, K.T. Bush, S.K. Nigam, Key Role for the Organic Anion Transporters, OAT1 and OAT3, in the in vivo Handling of Uremic Toxins and Solutes, Sci Rep. 7 (2017) 4939. https://doi.org/10.1038/s41598-017-04949-2.

[75] C.-Y. Sun, M.-S. Wu, C.-C. Lee, S.-H. Chen, K.-C. Lo, Y.-H. Chen, A novel SNP in the 5' regulatory region of organic anion transporter 1 is associated with chronic kidney disease, Sci Rep. 8 (2018) 8085. https://doi.org/10.1038/s41598-018-26460-y.

[76] 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. https://doi.org/10.1038/nrm.2015.25.

[77] A. Janaszkiewicz, Á. Tóth, Q. Faucher, M. Martin, B. Chantemargue, C. Barin-Le Guellec, P. Marquet, F. Di Meo, Insights into the structure and function of the human organic anion transporter 1 in lipid bilayer membranes, Sci Rep. 12 (2022) 7057. https://doi.org/10.1038/s41598-022-10755-2.

[78] N. Yan, Structural Biology of the Major Facilitator Superfamily Transporters, Annu. Rev. Biophys. 44 (2015) 257–283. https://doi.org/10.1146/annurev-biophys-060414-033901.

[79] X.C. Zhang, Y. Zhao, J. Heng, D. Jiang, Energy coupling mechanisms of MFS transporters: Energy Coupling Mechanisms of MFS Transporters, Protein Science. 24 (2015) 1560–1579. https://doi.org/10.1002/pro.2759.

[80] E.M. Quistgaard, C. Löw, P. Moberg, L. Trésaugues, P. Nordlund, Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters, Nature Structural & Molecular Biology. 20 (2013) 766–768. https://doi.org/10.1038/nsmb.2569.

[81] A. Blanco, G. Blanco, Lipids, in: Medical Biochemistry, Elsevier, 2017: pp. 99–119. https://doi.org/10.1016/B978-0-12-803550-4.00005-7.

[82] R. Dawaliby, C. Trubbia, C. Delporte, C. Noyon, J.-M. Ruysschaert, P. Van Antwerpen, C. Govaerts, Phosphatidylethanolamine Is a Key Regulator of Membrane Fluidity in Eukaryotic Cells, Journal of Biological Chemistry. 291 (2016) 3658–3667. https://doi.org/10.1074/jbc.M115.706523.

[83] A.C. Rustan, C.A. Drevon, Fatty Acids: Structures and Properties, in: John Wiley & Sons, Ltd (Ed.), ELS, 1st ed., Wiley, 2005. https://doi.org/10.1038/npg.els.0003894.

[84] Y. Chen, B.C. Lagerholm, B. Yang, K. Jacobson, Methods to measure the lateral diffusion of membrane lipids and proteins, Methods. 39 (2006) 147–153. https://doi.org/10.1016/j.ymeth.2006.05.008.

[85] M. Jan Akhunzada, F. D'Autilia, B. Chandramouli, N. Bhattacharjee, A. Catte, R. Di Rienzo, F. Cardarelli, G. Brancato, Interplay between lipid lateral diffusion, dye concentration and membrane permeability unveiled by a combined spectroscopic and computational study of a model lipid bilayer, Sci Rep. 9 (2019) 1508. https://doi.org/10.1038/s41598-018-37814-x.

[86] H.M. Hankins, R.D. Baldridge, P. Xu, T.R. Graham, Role of Flippases, Scramblases and Transfer Proteins in Phosphatidylserine Subcellular Distribution, Traffic. 16 (2015) 35–47. https://doi.org/10.1111/tra.12233.

[87] J. Razzokov, M. Yusupov, S. Vanuytsel, E.C. Neyts, A. Bogaerts, Phosphatidylserine flip-flop induced by oxidation of the plasma membrane: a better insight by atomic scale modeling, Plasma Process Polym. 14 (2017) 1700013. https://doi.org/10.1002/ppap.201700013.

[88] K. Nosol, R. Bang-Sørensen, R.N. Irobalieva, S.K. Erramilli, B. Stieger, A.A. Kossiakoff, K.P. Locher, Structures of ABCB4 provide insight into phosphatidylcholine translocation, Proc. Natl. Acad. Sci. U.S.A. 118 (2021) e2106702118. https://doi.org/10.1073/pnas.2106702118.

[89] A. Ben Saad, V. Vauthier, Á. Tóth, A. Janaszkiewicz, A. Durand-Schneider, A. Bruneau, J. Delaunay, M. Lapalus, E. Mareux, I. Garcin, E. Gonzales, C. Housset, T. Aït-Slimane, E. Jacquemin, F. Di Meo, T. Falguières, Effect of CFTR correctors on the traffic and the function of intracellularly retained ABCB4 variants, Liver Int. 41 (2021) 1344–1357. https://doi.org/10.1111/liv.14839.

[90] K. Kapoor, S. Pant, E. Tajkhorshid, Active participation of membrane lipids in inhibition of multidrug transporter P-glycoprotein, Chem. Sci. 12 (2021) 6293–6306. https://doi.org/10.1039/D0SC06288J.

[91] E.E. Pohl, O. Jovanovic, The Role of Phosphatidylethanolamine Adducts in Modification of the Activity of Membrane Proteins under Oxidative Stress, Molecules. 24 (2019) 4545. https://doi.org/10.3390/molecules24244545.

[92] C. Martens, R.A. Stein, M. Masureel, A. Roth, S. Mishra, R. Dawaliby, A. Konijnenberg, F. Sobott, C. Govaerts, H.S. Mchaourab, Lipids modulate the conformational dynamics of a secondary multidrug transporter, Nat Struct Mol Biol. 23 (2016) 744–751. https://doi.org/10.1038/nsmb.3262.

[93] C. Martens, M. Shekhar, A.J. Borysik, A.M. Lau, E. Reading, E. Tajkhorshid, P.J. Booth, A. Politis, Direct protein-lipid interactions shape the conformational landscape of secondary transporters, Nat Commun. 9 (2018) 4151. https://doi.org/10.1038/s41467-018-06704-1.

[94] H. Vitrac, V.K.P.S. Mallampalli, M. Bogdanov, W. Dowhan, The lipid-dependent structure and function of LacY can be recapitulated and analyzed in phospholipid-containing detergent micelles, Sci Rep. 9 (2019) 11338. https://doi.org/10.1038/s41598-019-47824-y.

[95] M. Bogdanov, W. Dowhan, Phosphatidylethanolamine Is Required for in Vivo Function of the Membrane-associated Lactose Permease of Escherichia coli, Journal of Biological Chemistry. 270 (1995) 732–739. https://doi.org/10.1074/jbc.270.2.732.

[96] M. Bogdanov, W. Dowhan, Phospholipid-assisted protein folding: phosphatidylethanolamine is required at a late step of the conformational maturation of the polytopic membrane protein lactose permease, EMBO J. 17 (1998) 5255–5264. https://doi.org/10.1093/emboj/17.18.5255.

[97] Á. Tóth, A. Janaszkiewicz, V. Crespi, F. Di Meo, On the interplay between lipids and asymmetric dynamics of an NBS degenerate ABC transporter, Biophysics, 2022. https://doi.org/10.1101/2022.05.16.492073.

Chapter II. Molecular Dynamics

II.1. Fundamentals of molecular mechanics

Molecular modelling uses the equations of quantum and classical physics to simulate the behaviour of molecular systems. Historically, there have been two families of methods: quantum mechanical (QM) methods and molecular mechanics (MM), which focus on electron- or nucleus-oriented properties, respectively. The advantage of MM calculations over quantum mechanics (QM) calculations is that larger systems, such as proteins, membranes, and DNA, can be simulated. Within the framework of the Born-Oppenheimer approximation, electrons are not considered as individual particles, thus their motion is neglected. Therefore, molecular systems are treated as the "ball-and-string" model, in which atoms' motions are treated by classical mechanics, i.e., by applying Newton's second law to explore system dynamics in Molecular Dynamics (MD) simulations [1].

II.1.1. Force field

The simulation uses a *force field* (FF), an empirical approximation that defines molecules and calculates the potential energy of a system. A force field is a collection of parameters and functions that describe the interactions between atoms and molecules. The empirical expression of potential energy used in MM comprises bonded and non-bonded interactions.

$$V_{total} = V_{bonded} + V_{non-bonded} \qquad (1)$$

The energy of the system is calculated as a function of nuclear position only. Atoms within molecules are represented as balls connected to each other by bonds, modelled as springs (Fig. 30). The bonded term includes bond stretching, angle bending, and bond torsion terms, while the non-bonding term includes electrostatic and Van der Waals interactions.

$$V_{bonded} = V_{bond} + V_{angle} + V_{torsion}$$
$$V_{non-bonded} = V_{Van\ der\ Waals} + V_{electrostatic} \qquad (2)$$

The functional forms in force fields are described as follows [1].

$$\begin{aligned} V(r^N) = & \sum_{bonds} \frac{k_i}{2}\left(l_i - l_{i,0}\right)^2 \\ & + \sum_{angles} \frac{k_i}{2}\left(\theta_i - \theta_{i,0}\right)^2 \\ & + \sum_{torsion} \frac{V_n}{2}\left(1 + \cos(n\omega - \gamma)\right) \\ & + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right) \end{aligned} \qquad (3)$$

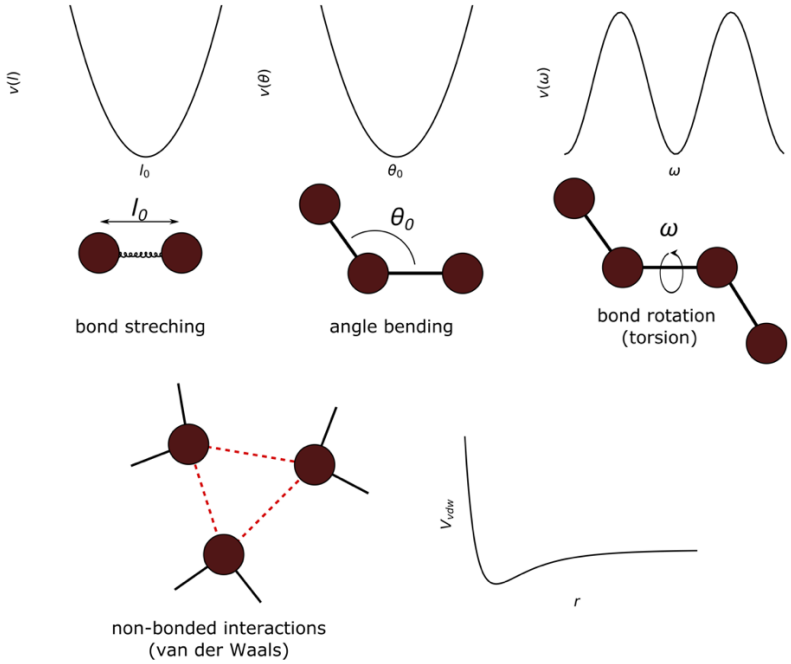


Figure 30. Basic components of molecular mechanics potentials.

II.1.1.1. Bond stretching

The bond lengths display deviation due to oscillation. The "stiffness" of a bond is restricted by a force constant k adequate to a bond type. Bond lengths in equilibrium (l_0) and force constant k are assigned either by experimental data or QM-based calculations. Most force fields describe the bond stretching potential within the harmonic approximation, applying Hooke's law:

$$v(l) = \frac{k}{2}(l - l_0)^2 \qquad (4)$$

Nevertheless, at extreme bond stretching, harmonic potential is a rather poor approximation. For large l, the Morse potential is a better approximation. However, its complexity, coming from additional terms and expansion functions, leads to a significantly higher computational cost:

$$v(l) = D_e\{1 - \exp[-a(l - l_0)]\}^2 \qquad (5)$$

where D_e is the depths of the potential energy minimum and $a = \omega\sqrt{\mu/2D_e}$ with μ being reduced mass and ω representing the frequency of bond vibrations (Fig. 31).

For most biomolecular MD simulations, the harmonic potential is considered as an elementary approach. Altogether with implicit electronic treatment, this basically precludes the sampling of out-of-equilibrium events such as reactivity (bond breaking/formation) by conventional FFs[1] and MD simulations [3,4].

[1] The ReaxFF [2] force field takes into account bord orders, which allow for the formation and breaking of bonds during MD simulations.

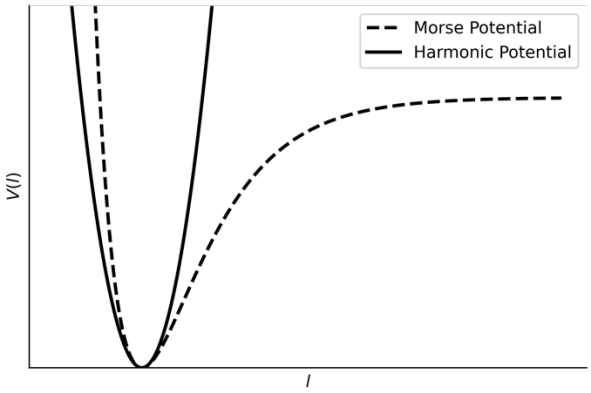


Figure 31. Harmonic potential compared to Morse potential.

II.1.1.2. Angle Bending

The potential energy of deviation of angle bending from the reference value (θ_0) is also expressed as an approximation of harmonic potential given by:

$$v(\theta) = \frac{k}{2}(\theta - \theta_0)^2, \qquad (6)$$

where contribution is defined by force constant k and the angle equilibrium value (θ_0). The force constants are significantly smaller than those for bond stretching, given that angle distortion is less energy-costing. Angle bending term may be improved, aiming at higher accuracy by the incorporation of higher-order terms for extreme cases such as highly strained molecules [1,5].

$$v(\theta) = \frac{k}{2}(\theta - \theta_0)^2[1 - k'(\theta - \theta_0) - k'(\theta - \theta_0) - k''(\theta - \theta_0)^2 \ldots] \qquad (7)$$

II.1.1.3. Torsion Terms

The potential energy of torsion term is expressed as cosine series expansion of the dihedral angle between A B C and D atoms as pictures in Figure 32A:

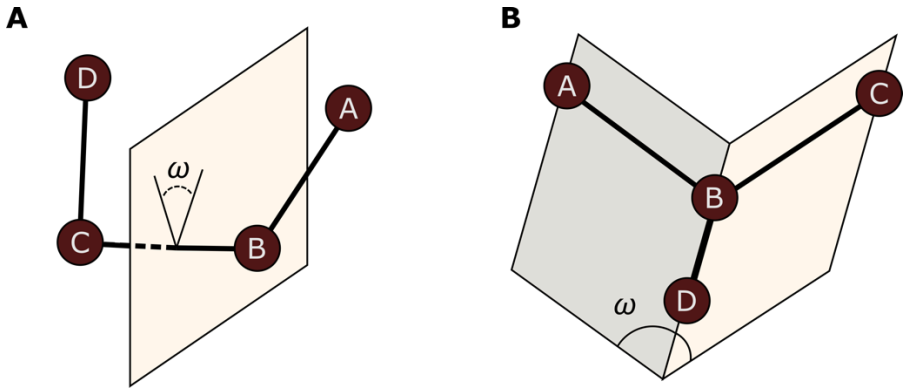


Figure 32. Torsion angles. A) proper torsion angle, B) improper torsion angle.

Torsional potential involves V_n, which indicates the relative energy barrier to rotation, while γ describes the minimum energy torsion angle, whereas n denotes the multiple energy minima in the function accounting for the periodicity. Compared to angles and bonds, the torsion term is more flexible regarding degrees of freedom.

$$v(\omega) = \sum_{n=0}^{N} \frac{V_n}{2}[1 + \cos(n\omega - \gamma)] \qquad (8)$$

Nevertheless, the energetic profile along torsion angle varies for some molecules. Depending on the atom types there might be added a torsional contribution which will correct the torsional term to provide more accurate value of $v_{(\omega)}$.

$$v_{(\omega)} = \frac{V_1}{2}(1 + cos\omega) + \frac{V_2}{2}(1 + cos2\omega) + \frac{V_3}{2}(1 + cos3\omega) \qquad (9)$$

Since the torsion itself is periodic so the torsional potential energy changes also periodically with the bond rotation (Fig. 33). Major conformational and energetic changes may occur due to bond rotation.

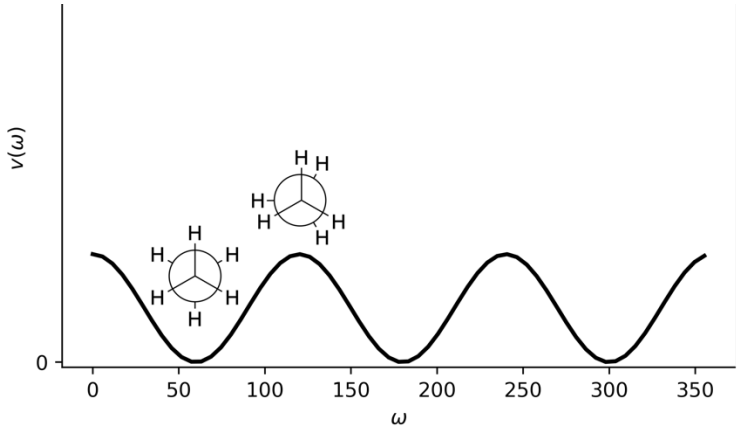


Figure 33. The potential energy of ethane along rotation of carbon-carbon bond.

Besides, an improper torsion angle is not characterized by four atoms bound sequentially, but a central atom (B) bonded to three atoms (A, C, D) as it is illustrated on Figure 32A. Torsion is defined as the angle between two planes. The so-called *out-of-plane* bending term can be represented:

$$v_{(\omega)} = k(1 - cos2\omega) \qquad (10)$$

However, the out-of-plane (improper torsion) terms are not always necessary. Moreover, the out-of-the-plane bending can simply adjust owing to non-covalent interactions of adjacent atoms, therefore some force fields do not use improper torsion potential terms [1,5].

II.1.1.4. Cross Terms

Cross terms are applied in force fields to describe the coupling of the aforementioned internal motion. Various couplings of those components can be included in FF potential, such as stretch-stretch (eq. 11), stretch-bend (eq. 12), stretch-torsion (eq. 13). For instance, by decreasing an angle in a molecule, as shown in Figure 34, the interaction of terminal atoms

increases. Thus, a cross term can be designed in order to account for bond length changes while the compression of the angle. Coupling between internal coordinates may appear in several forms, however, only a few are usually considered necessary for describing the structure and interactions of a molecule.

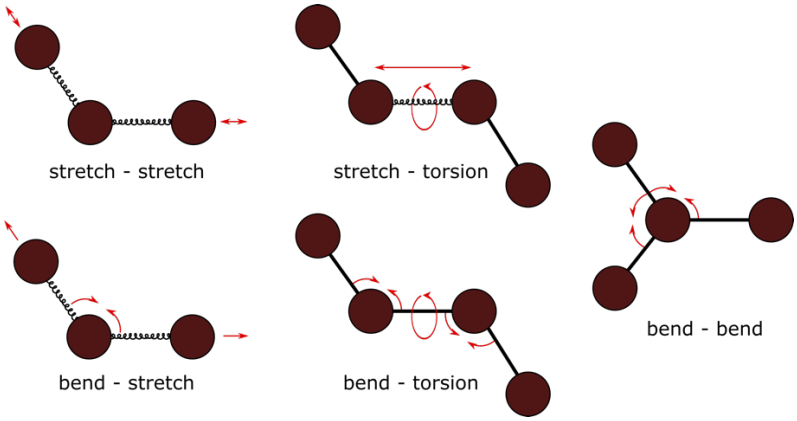


Figure 34. Cross terms.

Among several various possibilities to describe such interactions, the stretch-stretch can be expressed as:

$$v(l_1, l_2) = \frac{k_{l_1,l_2}}{2}[(l_1 - l_{1,0})(l_2 - l_{2,0})] \quad (11)$$

Particularly essential is the stretch-bend term when it comes to predicting the vibrational frequencies. One way to express such term is by using non-bonded interaction term of terminal (1 and 3) atoms known as Urey-Bradley potential approximated by a harmonic function.

$$v(r_{1,3}) = \frac{k_{r_{1,3}}}{2}(r_{1,3} - r_{1,3}^0)^2 \quad (12)$$

The stretch-torsion coupling is valuable especially in case of strained systems.

$$v(l, \omega) = k(l - l_0)cosn\omega$$
$$v(l, \omega) = k(l - l_0)[1 + cosn\omega] \quad (13)$$

Moreover, cross terms involving more than two internal coordinates are also possible. For instance, stretching of two bonds and an angle can be expressed as follows:

$$v(l_1, l_2, \theta) = \frac{k_{l_1,l_2,\theta}}{2}[(l_1 - l_{1,0}) + (l_2 - l_{2,0})](\theta - \theta_0) \quad (14)$$

Nonetheless, the force fields for biomolecular systems do not employ cross-terms, but rather force fields of class II and III (described in section II.3) utilize such terms [1].

II.1.1.5. Non-bonded interactions

There exist several different components of non-covalent interactions including e.g., electrostatic interactions, π-effect, van der Waals forces and hydrophobic effects. Nevertheless, usually force fields comprise as groups of electrostatic and van der Waals interactions. They

are considered as 'through-space' interactions modelled as function of force of the distance. The interactions are considered between all pairs of atoms [1].

II.1.1.5.1. Electrostatic interactions

The same or opposing (partial) charges on adjacent molecules cause repulsive or attractive forces, which are the basis for electrostatic interactions. They are considered to occur between atoms of non-zero electric moments. Numerous biological events rely on non-covalent interactions, such as the secondary and tertiary structures of proteins and nucleic acids, and the binding and unbinding of ligands.

Molecules display an unequal distribution of partial atomic charges. This is associated with the electrostatic properties of a molecule. First-order electrostatic between two molecules (or even within one) is often described as a sum of interactions between pairs of point charges by assuming first-order Coulomb's law:

$$v = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad (15)$$

where N_A and N_B represent the number of point charges in two molecules. Alternatively, the electrostatics can be calculated by considering a molecule as a single entity. The sufficient intermolecular electrostatics is based upon electric moments such as charge (q) and higher terms (e.g., dipole μ, quadrupole Θ, octupole Φ). Molecules are represented as distribution of charges, so dipole has two opposite charges on distant edges, whereas quadrupole has four charges and octupole has eight distributed charges. Usually, the lowest non-zero electric moment is considered; therefore, many uncharged molecules are treated as dipoles. Nevertheless, some molecules require more detail description of charge distribution thus higher-order terms can be considered. For instance, the electrostatic potential at point *P* between charge q_1 and q_2, like pictures the Figure 35, would have a following form including various electric moments using the cosine rule.

$$\phi(r) = \frac{1}{4\pi\varepsilon_0}\left(\frac{q}{r} + \frac{\mu cos\theta}{r^2} + \frac{\Theta(3cos^2\theta - 1)}{2r^3} + \cdots\right) \qquad (16)$$

Where μ illustrate dipole in form of $q_2 z_z - q_1 z_1$ and Θ is quadrupole in form $q_1 z_1^2 + q_2 z_2^2$.

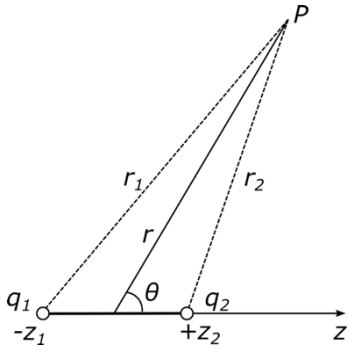


Figure 35. Electrostatic potential of two charges.

Complete electrostatic treatment should also consider the so-called induced electric moments (μ_{ind}) between molecules. Polarization is an event which results in an induced electric moment

caused by the environment external electric field, often neighbouring molecules. Such terms are often included in the so-called polarisable FFs.

$$\mu_{ind} = \alpha \mathrm{E} \qquad (17)$$

The induced dipole is proportional to the external electric field E with the constant of proportionality (α), polarizability [1,5,6].

II.1.1.5.2. Lennard-Jones potential

Even though van der Waals forces are considered weak interactions, in nature they are often associated with key events such as the formulation of water droplets (cohesion) or the gecko's ability to walk on transverse flat surfaces (adhesion).

Van der Waals forces lead to attractive interactions that can be strong enough to bind molecules in a non-covalent manner. Those forces are often described as the interaction between polar or non-polar molecules that exhibit an instantaneous electronic dipole moment. As two atoms approach each other, their electron densities impact each other. The fluctuations of electron densities create a slightly uneven distribution, giving rise to instantaneous electrical dipole moments. The developed electric moment will induce a slight charge polarisation in a neighbouring molecule. The simultaneously developed instantaneous electrical moments may be affected by long-range attractive forces over short distances in the case of a lack of bonding opportunities. The attracting potential is the effect of the so-called dispersive forces [4,7].

It is important to note that at a short distance, two particles are strongly repulsive owing to steric hindrance. Therefore, it is possible in FF to describe both repulsive and attractive non-electrostatic interactions by the so-called Lennard-Jones (LJ) potential. It is near to 0 for large distances, while for very short interatomic distances it becomes strongly repulsive. However, at intermediate distances, the attractive inductive effect plays a role, and its energy depends inversely on the distance between molecules to the sixth power.

$$E_{vdw}(r^{AB}) = E_{repulsion}(r^{AB}) - \frac{C^{AB}}{(R^{AB})^6} \qquad (18)$$

A widely used approach applies the *LJ 12-6 functional* where the repulsive interactions are given by interatomic distance to the negative twelfth power (r^{-12}).'

$$v_{LJ}(r^{AB}) = \frac{C_1}{(r^{AB})^{12}} - \frac{C_2}{(r^{AB})^6} \qquad (19)$$

The adjustable parameters C_1 and C_2 may be described by a collision parameter σ that determines r for which the energy is zero, and the well depth ε as shown on Figure 36.

Depending on FFs, the equation can also take the following form if $r_0 = 2^{1/6}\sigma$.

$$v_{LJ}(r) = \varepsilon\left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6}\right] \qquad (20)$$

Given that Lennard-Jones potential is specified by two parameters σ and ε, it provides simplicity to apply the equation to many different atoms, molecules.

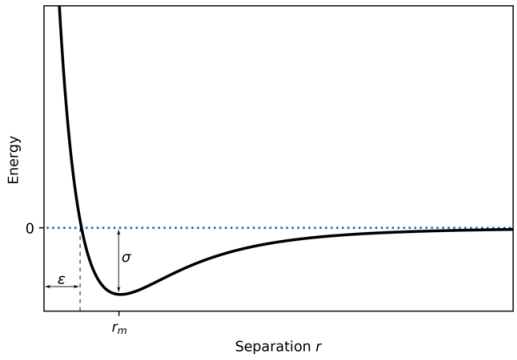


Figure 36. Lennard-Jones potential.

The Lennard Jones potential initiated the development of other potentials such as *Hill* or *Buckingham potential* that are based on the original LJ 12-6 with some adjustments to better reflect van der Waals interactions in particular systems. However, those adjustments often increase the computational cost of calculations [1].

II.1.1.5.3. Aromatic – aromatic interactions

The interactions of aromatic compounds have a great impact on structure stability. The aromatic-aromatic interactions are not treated uniquely, rather their interactions are encompass by general consideration of electrostatics and Van der Waals interactions. Among them, stand out conformation as face-to-face, t-shape and edge-on. For instance, in t-shape interactions the main role plays the quadrupole moments adapting the most convenient energetically orientation of molecules [8].

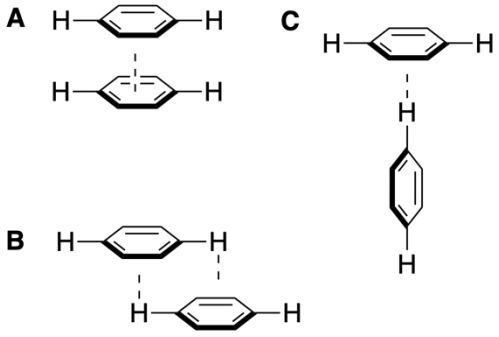


Figure 37. The aromatic-aromatic interactions exemplified on benzenes molecules: (A) face-to-face, (B) edge-on-face, (C) T-shape.

In general, main aromatic interactions can be classified as:

$\pi - \pi$ repulsion dominates in a face-to-face geometry,

$\pi - \delta$ attraction dominates in edge-on geometry,

$\pi - \delta$ attraction dominates in t-shape geometry.

II.1.1.5.4. Hydrogen bonding

Hydrogen bonds are known to maintain the secondary structure of proteins and nucleic acids. They are also the primary interactions between water molecules, shaping solvent-solute interactions in an aqueous environment. Therefore, that is especially important for describing the explicit solvent model. From the electronic point of view, hydrogen bonds occur between

hydrogen atoms (acceptors) bound to an electronegative element and another electronegative element (donor) that bears a lone pair (e.g., oxygen O, nitrogen N or fluoride F atoms). Usually, hydrogen bonding is described by the Lennard-Jones 6–12 term. However, some FFs include an explicit term, using a 10–12 Lennard-Jones-like potential as follows:

$$v(r) = \frac{A}{r^{12}} - \frac{C}{r^{10}} \qquad (21)$$

Given the focus on nuclei, it is worth mentioning that the terms “donor” and “acceptor” are inverted in MM with respect to the electronic view. Indeed, MD packages often deal with H-atom donors and acceptors [4].

II.2. Parametrization of force fields

Force fields attempt to mimic the behaviour of molecules and calculate the potential energy of a system. The function requires a set of optimized empirical parameters that are based on experimental observations or QM-based calculations. A suitable force field must satisfy the best compromise between accuracy and efficiency, providing accurate results within a reasonable time scale.

Once the complexity of the functional description is established, an issue arises when considering assigning values to parameters. The first important matter to bring up is the atom type, which varies from one to another not only by an element but also considering its geometry. In other words, hybridization but also the chemical environment like a moiety and bond a given atom type makes. The number of atom types strongly varies between force fields.

Subsequently, considering a force field comprises of *n* atom types, each entry must have assigned parameters such as:

- Stretching force constant (k^{AB}) and bond length (l_0^{AB}),
- Bending force constant (k^{ABC}) and angle (θ_0^{ABC}),
- relative energy rotation barriers (V_1^{ABCD}, V_2^{ABCD}, V_3^{ABCD})
- electric moments (q, μ,Θ,Φ)
- collision parameter (σ) and well depth (ε) for Van der Waals term

The number of cross terms in a force field usually ranges from zero to a dozen or so. The value assignment to a parameter requires a sufficient number of independent sets of data. It accounts for a hassle coming from rather limited experimental data. Moreover, since experiments provide data considering a molecule in its equilibrium state, it is a great challenge to obtain some values, such as the torsional profile. Therefore, it has become common to employ QM methods to obtain parameter values. Electronic structure calculations might be useful to gain knowledge about the bonded parameters, including stretching, bending, and torsional [1,5,6].

II.2.1. Parametrization of missing parameters

Despite the extensiveness of force fields, the common problem of lacking parameters frequently occurs. However, for a given molecule of interest, the issue of insufficient parameters can be solved in at least two ways. Taking advantage of already existing force field parameters for a similar system, the missing elements can be carefully assessed by

comparison. It is often applied in the case of missing torsion angles where the atoms are sufficiently similar. However, a potential energy function for one molecule may be extended to a larger range of molecules with related chemical groups, which is the basis for the concept of transferability. Yet, the typical technique is to use external data, such as experimental data or the electrical structure of a system, in the scenario of a system for which there are none or almost no parameters. The parametrization procedure can also combine the two methods. Regardless of the process, the parameters need to be carefully checked, validated, and compared to other systems. It is important to keep in mind that the parameters cannot be treated separately, therefore they cannot be transferred across distinct force fields, since their values depend on the remaining terms of energy functions and how they relate to the other terms [1,4].

II.2.2. Parameter reduction

To prevent excessive complexity, force fields assume structural transferability between different molecules. In other words, force fields encompass the majority of common chemical compounds adapting to a given generality, so-called "generic" parameters. Another way to reduce parameters in force fields can be by minimising the torsional parameters. The torsional parameter would depend only on the central atoms. For instance, regardless of the first and last atoms, every single carbon–carbon bond (X-C-C-X) would use the same parameter for different torsions. Again, such an approach must be carefully monitored along with simulations [1,5].

II.3. Types of force fields

Force field families vary in terms of (i) functional forms of energy terms, (ii) the number and type of included cross-terms, and thus (iii) the precision of parameters. Considering the accuracy level, force fields can be categorised into three main classes. If a force field is intended for large systems such as proteins or nucleic acids, its functional forms are kept quite simple, and no cross-term contribution is involved. They represent the *Class I* force fields (AMBER [9], CHARMM [10], and GROMOS [11]). The bond stretching and angle bending are represented by the harmonic potential, while repulsive and van der Waals interactions are described by the Lenard-Jones 12-6 potential. Since the harmonic description for bond stretching and angle bending is true only close to equilibrium, to study systems that require high accuracy, the anharmonic description is necessary. The *Class II* is considered as high degree accuracy force fields (MM2 [12], MM3 [13]), but it is limited to reproduce only small and medium-size molecules. The high accuracy is achieved by employment of several cross terms and use of higher-order function expansions for stretching and bending, such as anharmonic cubic and quadric terms to the potential energy. Nowadays, further improvements have been made in force field accuracy by including e.g., hyperconjugation or electronic polarisation as denoted in *Class III FF*s (e.g., AMOEBA [14,15], DRUDE [16]) [5].

Table 1. Comparison of force fields forms (adapted from *Introduction to Computational Chemistry* [5])

Force Field	Number of atom types	E_{str}	E_{bend}	E_{tors}*	E_{vdw}	E_{el}	E_{cross}	Molecules

AMBER [9]	41	P2	P2	imp.	12-6 12-10	charge	none	proteins, nucleic acids, carbohydrates
CHARMM [10]	29	P2	P2	imp.	12-6	charge	none	proteins
GROMOS [11]		P2	P2	P2(imp.)	12-6	charge	none	proteins, nucleic acids, carbohydrates
MM2 [12]	71	P3	P2+P6	P2	Exp-6	dipole	sb	general
MM3 [13,17]	153	P4	P6	P2	Exp-6	dipole or charge	sb, bb, st	general (all elements)
PFF [18]		P2	P2	Imp.	12-6	polar	none	proteins

Abbreviations: ***Pn***: Polynomial of order n; ***Exp–6***: exponential + R−6; ***n − m***: R−n + R−m Lennard-Jones type potential; ***imp.***: improper torsional angle; ***polar***: polarizable; ***bb***: bend–bend: ***sb***: stretch–bend; ***st***: stretch–torsional;

Most biomolecular systems simulations utilise the *Class I* force fields since they allow for the simulation of relatively large systems on a reasonable timescale with satisfying accuracy. The computational cost of simulations is tightly correlated with the number of atoms. The most time-consuming term in simulations is the pair-wise calculation of non-bonded interactions. Lowering the number of non-bonded interactions by reducing the number of explicit atoms is one method to lower the computing cost of simulations without diminishing the system. **The united atom** model simplifies molecules by combining groups of atoms and treating them as a single entity. Usually, it joins hydrogens to the atoms to which they are bonded. Typically, non-polar hydrogen atoms bound to carbon atoms are joined in united atom models. For instance, a single group (i.e., united atom) concept is applied to each of the moieties in lipids for GROMOS 53a6, 53a5 and 43a1p FFs [11]. One disadvantage is considering the inverse of stereochemistry that can occur in the united atom model during simulations. A way to prevent this event is to use improper torsions. One of the first force fields, such as the initial Amber [9], CHARMM [10] and OPLS [19] used the united atom models [5]. Given the recent advances in GPU-based MD simulations, united atoms have been less used over the past years [20].

However, in order to simulate large-scale systems and extend the time scale of simulations, greater approximations must be adopted. The "coarse-grained" approach represents molecules in reduced details, leading to greater efficiency with significantly longer time-scale simulations. Coarse-grained models represent a group of "heavy" atoms and their associated hydrogen atoms as a bean. Since there are fewer interactions to assess, the efficiency of the simulations rises several orders of magnitude compared to all-atom models. It is usually utilised for simulating large-scale protein folding, aggregation of membrane-bound proteins, polymer configuration, and more often to discover a protein's conformational space [4,21]. Nonetheless, several issues with CG simulations may arise: (i) lower accuracy; (ii) because some of the interactions were already assumed and fixed, some important events may be omitted; (iii) or an event may be incorrectly explained [4]. CG simulations can be used when an atomistic description is not required.

The point charge approximation has its own limitations, which come from the lack of responsiveness of electronic distribution in a molecule to the environment. Incorporating the

polarization into a force field significantly rises relevance of a study, since the electrostatic description is essential for the recognition and stability of biomolecular systems. The explicit description of electrostatic interactions sets apart polarizable force fields, such as AMOEBA [14,15] from the classical fixed-charge force fields, providing a more realistic treatment of electrostatics [4]. The difference between non-polarizable and polarizable force fields is most pronounced for hydrogen bonding, the dynamics of water around proteins [22], and sequence-specific effect in DNA [23]. Given that the polarisation effect is associated with supramolecular arrangement accounting for solute-solvent interactions, it shows special importance in studies involving interactions of nucleic acids with divalent ions [23], protein folding and stabilisation [24], or ligand-protein interactions [25]. Nevertheless, the increased accuracy comes with higher computational cost since not only non-bonded terms are improved but also the bonded terms use high-order potentials that better describe the system not only in equilibrium [26].

The variety of force fields allows it to investigate various biological processes like protein folding, structure, dynamics, and interactions with other molecules. It is currently possible to simulate complex systems, such as membrane proteins with lipids, ions, and water owing the GPU accelerated simulation and the variety of force fields. In order to describe specific types of molecules, such as proteins, nucleic acids, lipids, solvents, or inorganic molecules, dedicated force fields have been developed [27]. Focus given on smaller range of molecules provide highly optimized parameters favouring accuracy.

Table 2. Examples of popular force fields are listed in table that distinguish non-polarizable and polarizable force fields.

Types of molecules	Non-polarizable	Polarizable
Nucleic acids	AMBER [9], CHARMM [28], GROMOS [29]	AMBER ff02pol [30] CHARMM Drude [16] AMOEBA [31,32]
Proteins	AMBER ff19SB [33], CHARMM [28], GROMOS [29]	AMBER ff02pol [30] CHARMM Drude [16] AMOEBA [31,32]
Lipids	CHARMM [34,35] Lipid14 [36] GROMOS [37]	CHARMM Drude [16] AMOEBA [31,32]
Sugars	GLYCAM06 [38]	CHARMM Drude [16]

Therefore, the choice of a method depends not only on the system but also on the desired accuracy level, available computational resources, and accessible software. Extra attention needs to be paid if a simulated system requires employment of different force fields. It is not recommended to combine force fields that have been developed inconsistently, meaning that the forms of the potential energy functions must be the same.

Moreover, as a living perspective, it is worth noting that the branch of Machine-Learning based Force Fields (ML-FF) has been dynamically developing. The aim is to give up the compromise

between accuracy and computational efficiency. The principle of the algorithm is to learn from statistical relations of potential energy and chemical structure. Where the learning data set includes forces and energies obtained from *ab initio* calculations, As a result, ML methods learn an appropriate energy function for the system, which overcomes one of the problems in classical FFs [39].

II.4. Classical Molecular Dynamics Simulations

Molecular Dynamics aims to recreate the motion of atoms, while force field employs molecular mechanics function to calculate the potential energies $V(r)$ of atoms and determine forces acting on them. By providing the initial coordinates and velocities of a system, the function of potential energy (force field) calculates the potential energy. The potential energy function evaluates the molecular interactions by allowing to calculate forces:

$$\vec{F} = -\nabla U(\vec{r})$$

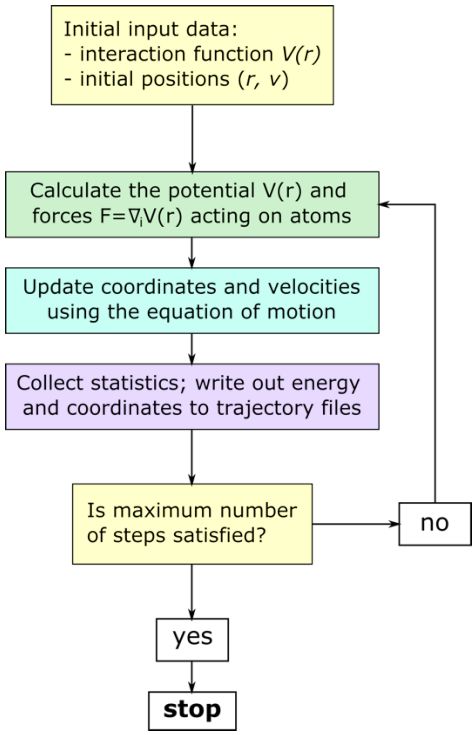


Figure 38. Flowchart for typical molecular dynamic simulations

However, the Newton's classical equation of motion allows to calculate the forces acting on a dynamic object in progressing time, accounting particle mass (m) and its acceleration (a).

$$F = ma \qquad (22)$$

Given that the force is a function of time and acceleration is the derivative of velocity (v) with respect to time, the equation of motion can be rewritten as follows:

$$F(t) = m\frac{\Delta v}{\Delta t} \qquad (23)$$

The equation of motion is calculated using an integration method. A so-called *integrator* is a mathematical function that combines the differential equations to compute atomic positions,

velocities, and forces. An integrator is a fundamental part of the typical workflow for molecular dynamic simulations. Such a mathematical function must meet certain expectations, including:

- conservation of energy and momentum,
- good accuracy – numerical solution should be close to the original solution,
- good stability for large time steps – to reduce computational cost,
- time-reversable – able to reproduce the preceding step,
- area-preserving (symplectic) – preserve the volume and total energy of a system between steps occurring one by another.

Algorithms such as the Verlet [40], velocity-Verlet [41], and leapfrog [42] meet expectations while displaying satisfactory accuracy. Moreover, they are adaptable to different equations of motion. Once the new positions of atoms and their velocities are established, the energy and coordinates are written out to trajectory files. The process is iterative until the number of steps is satisfied [1,4,6].

II.4.1. Integration algorithms

The integration should be divided into many minor changes with a relatively small timestep (Δt). At each t, the determined force provides information about acceleration, velocities and positions of particles that is used to calculate new positions and velocities at $t + \Delta t$ step. The algorithms that integrate the equation of motion adopt Taylor series expansion to approximate positions (r), velocities (v), acceleration (a) [1].

$$r(t+dt) = r(t) + dtv(t) + \frac{1}{2}dt^2a(t) + \frac{1}{6}dt^3b(t) + \frac{1}{24}dt^4c(t) + \cdots$$

$$v(t+dt) = v(t) + dta(t) + \frac{1}{2}dt^2b(t) + \frac{1}{6}dt^3c(t) + \cdots$$

$$a(t+dt) = a(t) + dtb(t) + \frac{1}{2}dt^2c(t) + \cdots$$

$$b(t+dt) = b(t) + dtc(t) + \cdots \qquad (24)$$

Given that velocity is the second derivative, while the acceleration is the third derivative, then b and c are the fourth and fifth derivatives of the position with respect to time, respectively. By using the position and acceleration of current (t) and previous ($t - dt$) step, the *Verlet* [40] algorithm calculates the new position as follows:

$$r(t+dt) = 2r(t) - r(t-dt) + dt^2a(t) \qquad (25)$$

Nevertheless, that algorithm does not include explicitly the velocities which require an indirect solution where the velocities at t time can be calculated from the half-step ($t + \frac{1}{2}dt$).

$$v(t) = \frac{r(t+dt) - r(t-dt)}{2dt} \qquad (26)$$

A variation of that approach is the *leap-from* [42] algorithm in which the velocities at time ($t + \frac{1}{2}dt$) are calculated from previous step velocities ($t - \frac{1}{2}dt$) and acceleration at time t.

$$v\left(t+\frac{1}{2}dt\right)=v\left(t-\frac{1}{2}dt\right)+\frac{dt}{m}F(t)=v\left(t-\frac{1}{2}dt\right)+a(t)dt$$

$$r(t+dt)=r(t)+v(t)dt\left(t+\frac{1}{2}dt\right) \qquad (27)$$

The *velocity Verlet* [41] algorithm is a well-adopted alteration of Verlet algorithm, which enables to calculate position and velocity:

$$v(t+dt)=v(t)+[a(t)+a(t+dt)]\frac{dt}{2}$$

$$r(t+dt)=r(t)+v(t)dt+\frac{1}{2}a(t)dt^2 \qquad (28)$$

The timestep Δt is an adjustable parameter and its value is important for effective integration. The value should be small and simultaneously large enough to enable large time-scale simulation. Timestep should be smaller than the vibrations of fastest entities. Usually, the timestep is in order of 2 fs, which means that, bonds involving hydrogen must be maintained by applying external constraints such as SHAKE [43] or LINCS [44].

The above-mentioned methods are illustrated in the Figure 39 showing comparison the order of computing positions (r), velocities (v) and acceleration (a).

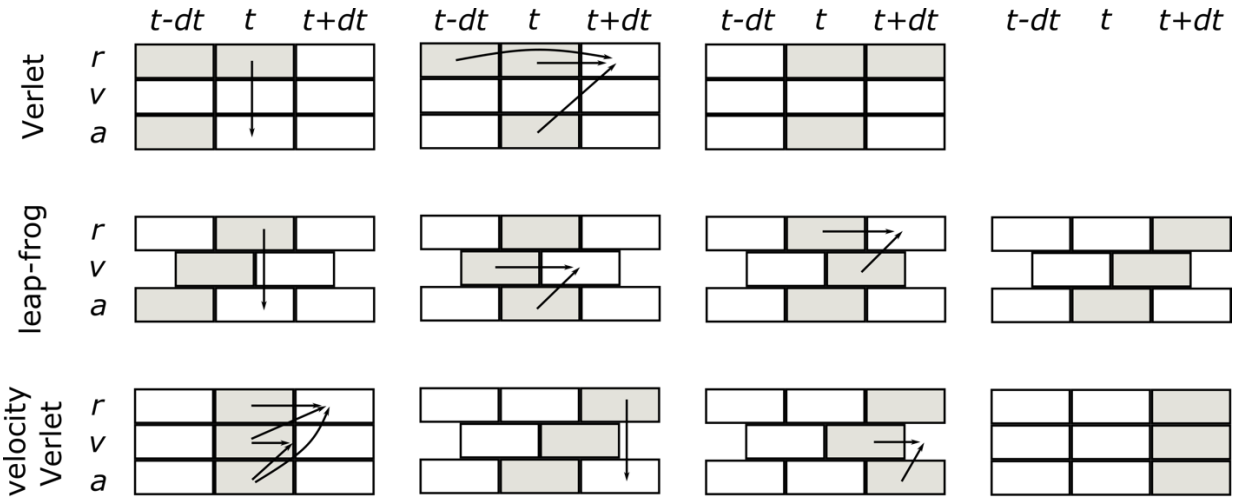


Figure 39. Integration algorithms Adapted from *Computer simulation of liquids* [45].

II.4.2. Ensembles

Thermodynamic ensembles are probability distributions of system particles assuming certain constant properties. The microcanonical ensemble arises directly from the integration of Newton's equation of motion and is characterized by a constant number of particles (*N*), volume (*V*), and energy (*E*). Microcanonical ensemble describes an isolated system that does not exchange energy with its surroundings. Since the potential energy of the system should decrease due to its relaxation, the kinetic energy in the form of the temperature must increase to maintain the constant total energy.

The canonical ensemble maintains a constant number of particles (*N*), volume (*V*), and temperature (*T*). As a result, the system's overall energy might alter while maintaining a constant temperature. It models a closed system, i.e., the system can exchange energy with

its environment (here, the thermostat “bath”) but not particles. Finally, MD simulations are often performed under the constant temperature, constant pressure, and constant number of particles (N, P, T) ensemble, also known as the isothermal-isobaric ensemble. The volume (box size) can change while the pressure remains constant.

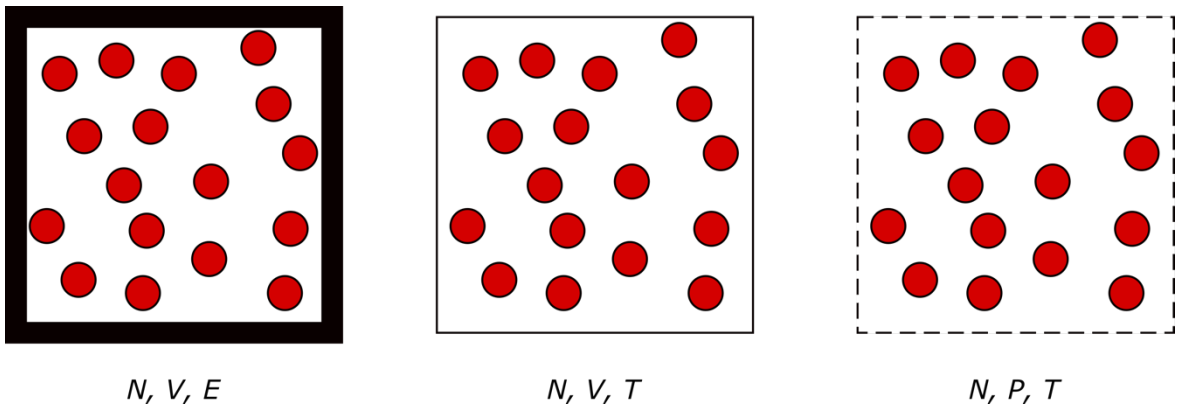


Figure 40. The schematic visualisation of ensembles.

Initial velocities of atoms may be randomly assigned by assuming the Maxwell-Boltzmann distribution at a given temperature. It provides the probability (P) of velocity v_{ix} and direction x of an atom with given mass m at a certain temperature T.

$$P(v_{ix}) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} exp\left[-\frac{1}{2}\frac{m_i v_{ix}^2}{k_B T}\right] \qquad (29)$$

Temperature and/or pressure control must be applied in order to be consistent with the biological or physical conditions. Canonical or isothermal-isobaric ensembles are frequently used [1,4,45].

II.4.3. Thermostat

A variety of thermostats are available to modulate the temperature of a system. However, the role of thermostats is not to strictly maintain a constant temperature but rather to ensure that the average temperature of a system matches the targeted value. Thermostats instead scale the velocities of atoms since temperature and the average particle kinetic energy fluctuations are related as follows:

$$\langle E_{kin} \rangle = \sum_{i=1}^{N} \frac{|p_i|^2}{2m_i} = \frac{3}{2} N k_B T \qquad (30)$$

where N is the total number of degrees of freedom, k_B the Boltzmann constant.

Thermostats may generally be divided into global (e.g., Nosé-Hoover [46,47], and Berendsen weak-coupling [48]); and local families (e.g., Langevin [49], Andersen [50,51]). The latter acts on specific particles almost randomly, whereas the former instantaneously affects all particles with the same strength. Berendsen weak-coupling and Langevin are popularly used thermostats, therefore the focus will be given mainly to these two methods.

An efficient temperature scaling method is Berendsen weak coupling. Furthermore, Berendsen formalism can also be extended to pressure control for (N,P,T) MD simulations. Berendsen

method is based on weak coupling between the system and (in case of thermostat) heat bath at given T_d. In case if the system deviates from T_d, it is instantly corrected by:

$$\frac{dT}{dt} = \frac{T_d - T}{\tau}, \qquad (31)$$

where τ determinates the strength of the bath and system coupling. It is expected that a high parameter τ would lead to weak coupling, whereas a small value would result in strong coupling. The basic velocity scaling approach, however, produces results when τ equals the timestep dt. Otherwise, the technique applies a scaling factor χ to rescale velocities, defined by:

$$\chi = \left[1 + \frac{\Delta t}{\tau}\left(\frac{T_d}{T} - 1\right)\right]^{1/2} \qquad (32)$$

The kinetic energy variations are eliminated by rescaling the velocities, and therefore it was shown that the method does not reflect the canonical ensemble [52]. Rescaling the velocities suppresses the kinetic energy fluctuations. Vibrations and internal kinetic energy decrease during the course of the simulation in favour of translational, external kinetic energy. The kinetic energy of low-frequency motion is growing while the kinetic energy of high-frequency motion is dropping, and this is the weakness of the Berendsen thermostat. It results in the alleged "flying ice cube" occurrence [52]. Despite that, for larger systems, built of thousands of atoms, the thermostat yields approximately correct results, while the results disappoint for small systems. The exhibited method is a poor representation of the canonical ensemble [52,53,53].

The Nosé-Hoover thermostat is a preferable choice, if it comes to computing structural properties, as the resulting trajectories are in line with the canonical ensemble. This method foundation is an expanded Lagrangian with additional artificial coordinate (s) that represents additional degree of freedom standing in for a thermal reservoir. The reservoir has the potential energy given by:

$$U_s = gk_BTlns \qquad (33)$$

as well as the kinetic energy

$$K_s = \frac{Qs^2}{2}. \qquad (34)$$

The parameter g is equivalent to the number of degrees of freedom in the system, while Q stands for the effective mass associated with s. The fluctuation of temperature is determined by Q which demonstrates the coupling between the real system and reservoir s. Therefore, the Lagrangian is defined:

$$\mathcal{L}_s = K + K_s - U(\{\vec{r}_i\}) - U_s, \qquad (35)$$

given that the kinetic energy $K = \frac{1}{2}\sum_i m_i \vec{r}_i^{\,2}$.

The Langevin thermostat on the other hand, describes the motion of particles by adding a friction (Γ) and noise (ξ) term to Newton's equation of motion

$$m_i\vec{r_i} = \vec{F_i} - m_i\Gamma_i r_i + \vec{\xi}_i(t) \quad (36)$$

Therefore, the force acting on an atom ($\vec{F}$) is reduced by the friction (Γ) which can be understood as the rate of temperature relaxation and added the noise (ξ) which corresponds to particle's random collision with solvent atoms modelled as the Brownian motion at given temperature [4].

II.4.4. Pressure control

Since many experiments are conducted at constant temperature and pressure, simulations performed in isothermal-isobaric ensemble are most relevant to the experimental data. The volume of the box is modified to maintain the pressure deviations, not the pressure itself, in a manner similar to how the temperature is controlled. The volume fluctuation is negatively inversely related to isothermal compressibility (κ).

$$\kappa = -\frac{1}{V}\left(\frac{dV}{dP}\right)_T \quad (37)$$

Due to relatively easily compressible materials, a high value of is associated to adequate volume deviation (dV). Similar to how the temperature deviation in the canonical ensemble was corrected via thermal energy exchange, the isothermal-isobaric ensemble makes advantage of energy fluctuation [4,7]. The algorithms for controlling pressure are comparable to those used for maintaining temperature. The approach proposed by Berendsen [48] uses an extra term P_d which is the desired pressure, also referred to as the pressure of the 'bath', while P is the actual pressure. The parameter τ_p, in this context determines how a system is coupled to the pressure bath.

$$\frac{dP}{dt} = \frac{P_d - P}{\tau_p} \quad (38)$$

The box volume is rescaled at each time step (dt) by the scaling factor χ.

$$\chi = 1 - \kappa\frac{dt}{\tau_p}(P_d - P) \quad (39)$$

As a consequence, new atomic positions (r_i') are rescaled by $\chi^{1/3}$:

$$r_i' = \chi^{1/3} r_i \quad (40)$$

The scaling of the pressure may cause its oscillations, which is a defect in the Berendsen barostat. Thus, the Berendsen barostat is commonly used for pre-equilibrating the system rather than MD production.

The Andersen [51] barostat adds an additional degree of freedom by introducing an external variable, V, which corresponds to the box's volume. In an actual system, it produces the effect of a piston. The kinetic energy associated with V takes form:

$$K_V = \frac{1}{2}QV^2, \quad (41)$$

where Q stands for the piston's mass. The potential energy of the piston is given by $U_V = P_d V$. Rapid oscillation is caused by a piston with a small mass, whereas the converse is true for a piston with a large mass, whereas the infinite mass piston exhibits the typical molecular dynamics behaviour. The average box volume is given by the balance of external and internal pressure of a system. The coordinates and velocities are scaled by $\vec{r} = \frac{\vec{r}}{V^{1/3}}$ and $\vec{v} = \frac{\vec{v}}{V^{1/3}}$, respectively.

II.4.5. Periodic Boundary Conditions

Simulation systems in MD must be enclosed in a box with suitable so-called boundary conditions. Closed borders are inappropriate for biological systems, and it should not be limited by physical walls. Periodic Boundary Condition (PBC) systems are models in which the original box is imaged in each dimension. Therefore, particles can cross boundaries without consequences as a particle disappearing from one box edge reappears on the other side (Fig. 41), maintaining a constant number of particles. The short- and long-range interactions are calculated using the minimum image convention approach, i.e., taking into account not only the atoms in the original box but also the interacting atoms from adherent boxes (Fig. 41B) [1,4].

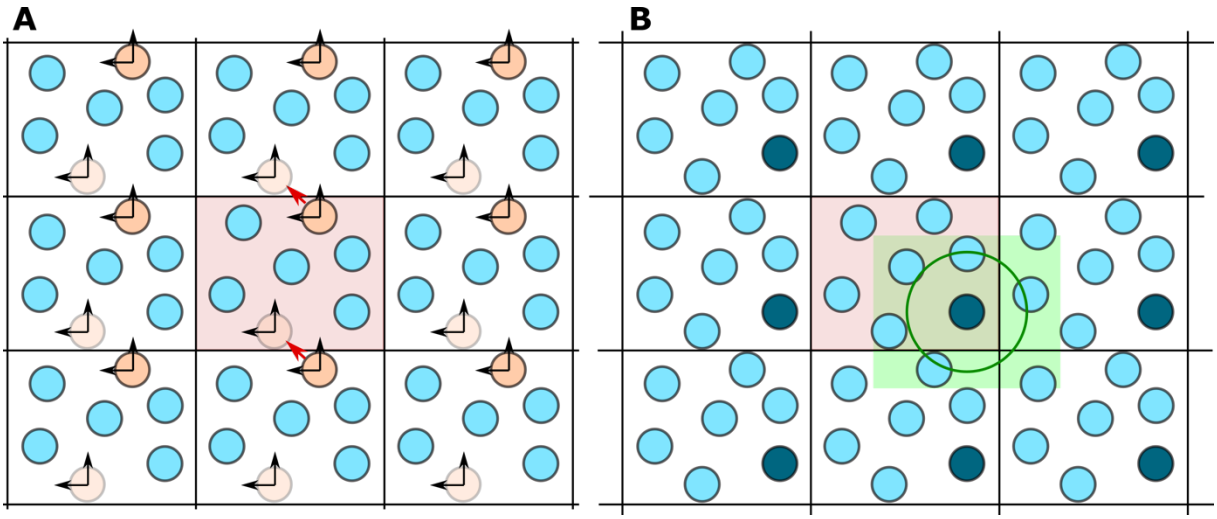


Figure 41. The shaded box is called the central box in which the simulation is performed

II.4.6. Cut-off

Non-bonded interactions are the most computationally expensive aspect of MD simulations. As a result, the *minimum image convention* and *cut-off* are typically used to address non-bonded interactions.

The minimum image convention, which limits given atom interactions to one image of each atom, is used to compute the energy of non-bonded interactions exclusively inside the cut-off distance (Fig. 41B). For atom pairs extending the cut-off, the energy of non-bonded interactions is set to 0. When utilising periodic boundary conditions, the cut-off value should be carefully adjusted since too large values will lead to self-interactions, usually ranging between 8 and 14 Å depending on the box size and the FFs. Since the Lennard-Jones interactions decay very abruptly, the cut-off method works well. On the contrary, the Coulomb interactions decay much slower than LJ $(1/r)$, therefore using cut-off can lead to artefacts. As a solution, it is common practise to use Particle Mesh Ewald (PME) summation [54], which splits the interactions into

short-range (before the cut-off) and long-range (after the cut-off) interactions. Assigning charges to a grid in a periodic system enables the use of Fourier transform to solve out of the cut-off long-range electrostatic interactions [1,4].

II.4.7. Constraint MD

Since the non-bonded interactions are updated at each computational step, larger timestep values might be a solution to sample more of the conformational space during MD simulations. However, the time step must be one order of magnitude smaller than the highest frequency of the system. Covalent bonds involving hydrogen atoms have the highest frequency vibrations, reaching 10 fs^{-1} in the case of the $C-H$ bond, for example. Therefore, the minimal timestep in bioorganic systems cannot be greater than 1 fs. Fortunately, H-atom containing bond stretching motions have little impact on the system. Hence, removing those degrees of freedom, by constraining the hydrogen/heavy-atom bonds to their equilibrium length, enables to extend the time step and thus increase the efficiency of simulations. Algorithms such as SHAKE [43] and LINCS [44] introduce bond constraints preventing from bond vibrations thus allowing to increase the time step from 1 fs to 2 fs. In the Verlet integration method, the SHAKE algorithm freezes the bond between i and j as follows:

$$\left|r_{ij}\right|^2 - d_{ij}^2 = 0 \qquad (42)$$

Given that r_{ij} is the instantaneous interatomic distance vector and d_{ij} is the distance constraint. SHAKE iteratively adjust the coordinates to satisfy the constraints. First, the atoms are allowed to move according to forces and subsequently the atoms are forced to obey the constraints via iterating through variables [1,4].

II.4.8. Solvent model

Water molecules may be represented with different levels of detail. Implicit solvent considers only electrostatic contributions between solute and solvent, leading to a substantial lower computational cost. However, such models ignore specific interactions, such as solvent-solute H-bonding, which have been demonstrated to be important in biological events. Therefore, explicit solvent models provide a more detailed description of water molecules, being thus more computationally demanding. According to the level of accuracy, they can be divided into categories. The simple solvent models represent water molecules as rigid molecules with point charges (from 3 to 5 points in water) for non-bonded interactions. A certain improvement in accuracy provides the flexible models with leading water molecule conformation flexibility. Another improvement considers many-body effects and polarisation [5,55].

The most popular force fields include TIP3P, TIP4P and SPC [56,57] that differ within parameters regarding geometry and point charges that represent *interaction sites.* TIP4P stands out with additional interacting site that charge is placed below the oxygen toward hydrogens.

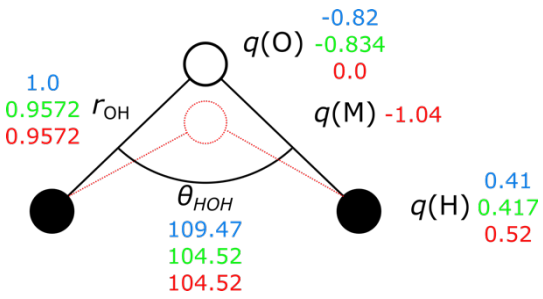


Figure 42. The differences between SPC (blue), TIP3P (green) and TIP4P (red) water models.

II.5. Modelling protein-ligand interactions by means of molecular docking

Molecular docking is a key approach in structural biology for predicting binding the modes of a molecule to a protein with a known three-dimensional structure. Biological processes are better understood when molecular docking provides initial information about ligand-protein interactions at the atomic level. The insights from the behaviour of small molecules in binding sites help model appropriate drugs as well as drug targets [58]. Although protein-protein docking is progressively employed, in the majority, small molecules are often docked into macromolecules (including proteins and nucleic acids). Combining protein-ligand docking is a commonly used technique in pharmaceutical research, where together with virtual screening and MD simulations, it has become a standard pipeline for computer-aided drug discovery [6,59]. To briefly introduce the concept of the *in silico* pipeline for drug development, the procedure includes several generic stages, where each step results in narrowing the range of molecules until the desired molecule(s) are derived. A typical strategy involves (i) virtual screening from an appropriate database, (ii) molecular docking calculations to assess potentially relevant compounds, and (iii) MD simulations to estimate reasonably accurate free energy binding, leading to a small number of suitable molecules (Fig. 43).

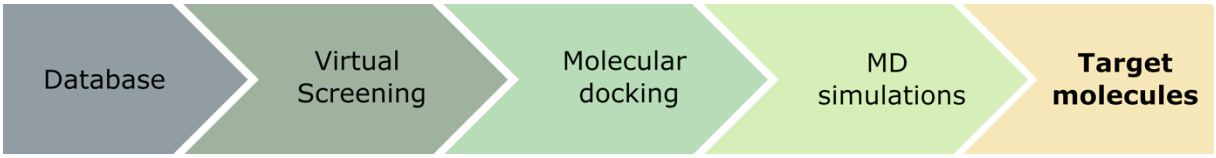


Figure 43. A generic scheme for a range of studies related to drug development in *in silico* manner.

The origins of molecular docking can be traced back to Fischer's idea of "lock-and-key" [60], in which the ligand fits into the protein binding pocket like a lock and key as rigid bodies. The concept was then refined by Koshland, who introduced the "induce-fit" theory [61], which complements the previous consideration by including the local arrangement of binding pocket residues interacting with a ligand. As a result, the incorporation of flexible molecules into docking calculations became important [58].

Essentially, docking attempts to predict a ligand-receptor complex by emphasising the molecule's conformation, position, and the so-called binding affinity. A typical approach consists of two docking phases:

(i) The first phase is *blind docking*, which involves searching throughout the whole protein without prior knowledge of the binding pockets. Typically, this stage involves several runs in which binding affinities are estimated, revealing favourable binding regions.

(ii) The next docking search is restricted to binding regions defined in the first step. This stage often enables ligand flexibility and partial receptor flexibility, with the binding region residues free to rotate. However, the rules of flexible docking are determined by the method used.

There are numerous docking calculation methods, although comparing them is complicated as they use different sampling algorithms and scoring functions. Additionally, some approaches work better with specific groups of molecules. Consequently, one method may be better suited for a certain system than another [59].

In general, searching algorithms can be divided into *systematic* and *stochastic* methods. While systematic methods are deterministic and the result depends on the resolution of the searching area, stochastic methods randomly make changes to a state variable iteratively until the criteria are satisfied [59]. For instance, systematic search implemented in methods such as DOCK [62] or FLOG [63] employs matching algorithms [64] which are based on molecular shape and chemical information. Stochastic methods, like the used Monte Carlo (MC) formalism [65], generate conformations by allowing for bond rotation. Then, the conformers are evaluated by an energy-based score function. The implementation of iteration uses the best poses for further conformer generation. That allows to overcome large energetic barriers. For example, the MC algorithm can be used in Autodock [66].

It is essential to implement a scoring system in order to rank the binding modes of molecules and extract the relevant poses. The scoring function estimates binding affinity based on either force field, knowledge, or empiric parameters.

The FF-based scoring function assesses the energy binding by summing the non-bonded interactions (electrostatics and van der Waals). However, to diminish computational cost, the non-bonded interactions are restricted by a cut-off, which decreases the accuracy of long-range interactions. For instance, Autodock, for example, uses an AMBER force field that takes into account solvation and entropy contributions. The empirical scoring function takes into account the energy of hydrogen bonds, ionic interaction, “hydrophobic effects” and binding entropy that are appropriately weighted and summed up.

Knowledge-based scoring [67] is a statistical approach that uses information from databases of resolved structures of ligand-protein complexes.

The scoring system often suffered from inaccuracies in binding affinities caused by, e.g., lack of or poor solvation effect representation. However, if the solvation effect is extremely important for a given system, one can utilize a physics-based score like MM-PB/SA that employs Poisson-Boltzmann potential for solvent effect.

Since molecules display a variety of conformations due to their high flexibility, the initial docking in which the molecules were considered to be rigid bodies is a limitation; despite the fact that conformational adaptation of a protein to a ligand is regarded as advantageous for enhancing binding affinity. Even though the majority of current methods allow for the complete or at least partial flexibility of molecules, the flexibility of macromolecules remains challenging [1].

II.6. Assessing Free Energy surfaces by means of advanced MD techniques

II.6.1. Background

Depending on the circumstances, macromolecules can exist in a variety of conformational states. A transition between different states is possible, but it also relies on the energy barriers

and thus on the transition timescale. Assessing energy barriers and pathways between given states is essential to thoroughly understanding the mechanism of a given biological event at an atomic level. Since the conventional MD simulations are limited by their length, the biological events are almost impossible to observe. Assessing free energy barriers using advanced MD methods can overcome this issue by describing plausible pathways (s) between states and conformations on an affordable time scale.

Free energy is a thermodynamic equilibrium measure associated with stability that indicates the probability of a system adopting a given state. The free energy calculations may be used exclusively to estimate energy barriers, as the free energy difference between states, or to sample the path between states, as the free energy surface along a coordinate. The calculation of free energy essentially consists of three parts: (i) the functional form of energy; (ii) a sampling method that produces a representative ensemble of configurations; and (iii) a method to estimate the free energy difference; with the sampling step usually requiring the largest time and computing power [68]. As a result, the next section will concentrate on computing the free energy difference before introducing the idea of enhanced sampling techniques.

II.6.2. Definition

There are two ensembles to consider in terms of free energy. The free energy of a system with a constant number of particles, volume, and temperature (NVT) is expressed as the Helmholtz function, whereas the free energy of a system with a constant number of particles, pressure, and temperature (NPT) is expressed as the Gibbs free energy function [1,4]. The Helmholtz free energy formula is provided below, while Gibbs free energy is its NPT ensemble equivalent.

$$A = -\frac{1}{\beta} ln Q_{NVT}, \qquad (43)$$

Where $\beta = \frac{1}{k_B T}$ with k_b being the Boltzmann constant and T stands for the temperature. Q_{NVT} stands for partition function defined:

$$Q_{NVT} = \frac{1}{h^{3N} N!} \iint \exp\left[-\beta \mathcal{H}(x, p_x)\right] dx\, dp_x \qquad (44)$$

where $\mathcal{H}$ is Hamiltonian representing the potential energy, x stands for atomic coordinates and p_x is the momentum. Basically, the partition function can be perceived as fundamental for the statistical mechanical description of particles. The partition function gives rise to the concept of the probability distribution (PDF, $\mathcal{P}(x, p_x)$), which is a function of atomic positions (x) and momenta (p_x), that calculates the likelihood of various state of existence. The probability distribution predicts that low energy states will be the ones that are sampled the most in respect to their Boltzmann weight [68].

$$\mathcal{P}(x, p_x) = \frac{1}{h^{3N} N!} \frac{1}{Q_{NVT}} \exp\left[-\beta \mathcal{H}(x, p_x)\right] \qquad (45)$$

Thus, the free energy difference between two states is associated with the ratio of probabilities of those states through the ratio of partition functions as follows [69]:

$$\Delta A_{0\to1} = A_1 - A_0 = -\frac{1}{\beta} ln \frac{Q_1}{Q_0} \qquad (46)$$

The calculations of free energy are meant to represent the difference between states rather than absolute free energy. Even the so-called “absolute” binding free energies used in protein-ligand interactions are calculated as the difference in energy between a ligand bonded to a binding pocket and free a ligand/target. The evaluation of the free energy difference between two states can reveal details about, for example, the difference in binding energies of various molecules; the dissociation energy of a small molecule from a receptor; the stability of a mutant protein; the partition coefficient of a molecule between two phases, and thus its concentrations in each phase [1,4,68].

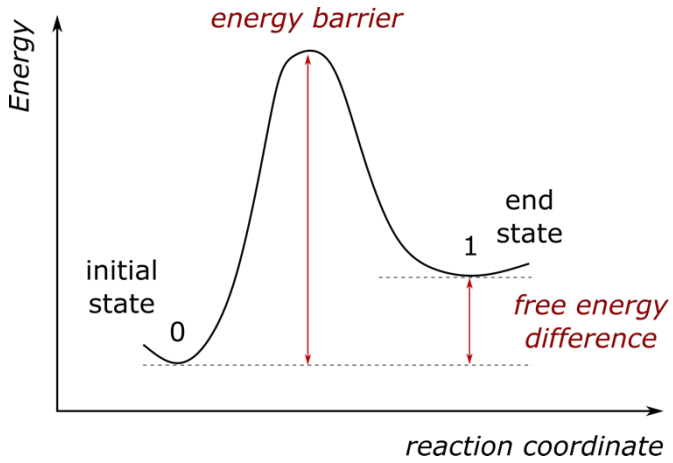


Figure 44. The schematic representation of a reaction between an initial state (0) and an end state (1) along a reaction coordinate. To get from state 0 to state 1 a system must overcome an energy barrier, while the free energy difference is calculated between the states, regardless of the energy barrier.

As a result, multiple approaches can be applied to calculate the free energy difference, including, for instance, (i) methods based on histogram (WHAM), (ii) non-equilibrium simulations (Jarzynski equation), (iii) perturbation theory techniques (Free Energy Perturbation; FEP), or (iv) derivate and integrate force method (thermodynamic integration; TI) [68,69].

II.6.3. Alchemical binding free energy

The calculations of *alchemical transformations* correspond to a perturbation only in the region of a system where two states differ. The term *alchemical* refers to system transformations such as converting moieties like an alcohol to an amine or hydroxyl group to a chloride, but also atom disappearance. Molecules can be computationally modified via alchemical transformation described in the initial and end state using either a single or dual topological paradigm. When using a dual topology paradigm, the initial and end state share the topology, where the states co-exist and the transformation occurs within a single simulation. However, the transformed atoms are not able to "see" each other. While in the single topology paradigm, the initial and end state are described separately, therefore two separate simulations are performed using "dummy" atoms. In contrast to the dual topology paradigm, which scales the interaction energy with the environment, the single topology paradigm scales only the non-bonded parameters so that their values range between 0 and 1. The free energy contribution is known in the case of the single-topology approach, since electrostatic and van der Waals interactions are decoupled, yet the method calls for two distinct simulations. Unlike a single

simulation used in the dual-topology approach, however, it does not provide the free energy contributions [68]. Although there are a number of ways to determine free energy difference, the following sections will focus on thermodynamic perturbation and thermodynamic integration in terms of the alchemical free energy binding [68,69].

II.6.3.1. Free energy perturbation (FEP) technique

Considering that the initial and end state are distinct, the transition from the initial to the end state can be accomplished by the linear scaling of the potential energy function or its parameters using a scaling parameter λ whose values range between 0 and 1. Where λ_0 denotes the initial state and λ_1 stands for the end state. Although the λ parameter may be employed in a variety of processes, including calculating the potential mean force, it is here presented in the context of free energy difference.

$$\Delta A_{0\to1} = -\frac{1}{\beta} ln\langle \exp\{-\beta[\mathcal{H}(x,p_x;\lambda_1) - \mathcal{H}(x,p_x;\lambda_0)]\}\rangle_{\lambda_0} \qquad (47)$$

The transition may be identified as an average over configurations by using the partition function and the identity of FEP to describe the free energy difference. Where the definition of FEP using the identity is:

$$\iint \exp[+\beta\mathcal{H}(x,p_x)]\exp[-\beta\mathcal{H}(x,p_x)]\,dx\,dp_x = h^{3N}N! \qquad (48)$$

If $\langle \dots \rangle_{\lambda_0}$ in eq. 47 represents an ensemble average over representative configurations of the initial λ_0 state, then the same calculation may be carried out backwards. Such approach to carry a simulation from $0 \to 1$ and $1 \to 0$ can improve accuracy.

$$\Delta A_{1\to0} = \frac{1}{\beta} ln\langle \exp\{-\beta[\mathcal{H}(x,p_x;\lambda_0) - \mathcal{H}(x,p_x;\lambda_1)]\}\rangle_{\lambda_1} \qquad (49)$$

The configuration ensembles must overlap in order to converge and achieve an accurate free energy difference. The transformation can be carried out by minor perturbation changes via physically meaningless intermediate states. Free energy is a state function and does not depend on the path but only on the end states. To ensure ensemble overlap and connect the end states, the path from state 0 to state 1, is divided into a number of intermediate states. The previously mentioned coupling parameter λ corresponds to the interval of intermediate states, often referred to as “window”. The coupling parameter describes a relationship between the initial, end and intermediate states. Along with λ changes, the force field terms can be expressed as a linear combination of λ that scales the electrostatic and Lenard-Jones potentials. Therefore, for each λ a separate simulation is performed with appropriate force field parameters. The potential is a function of λ and coordinates x. For N intermediate states, the free energy of transformation from state 0 to state 1 can be calculated as follows:

$$\Delta A_{0\to1} = -\frac{1}{\beta}\sum_{k=1}^{N-1} ln\langle \exp\{-\beta[\mathcal{H}(x,p_x;\lambda_{k+1}) - \mathcal{H}(x,p_x;\lambda_k)]\}\rangle_{\lambda_k} \qquad (50)$$

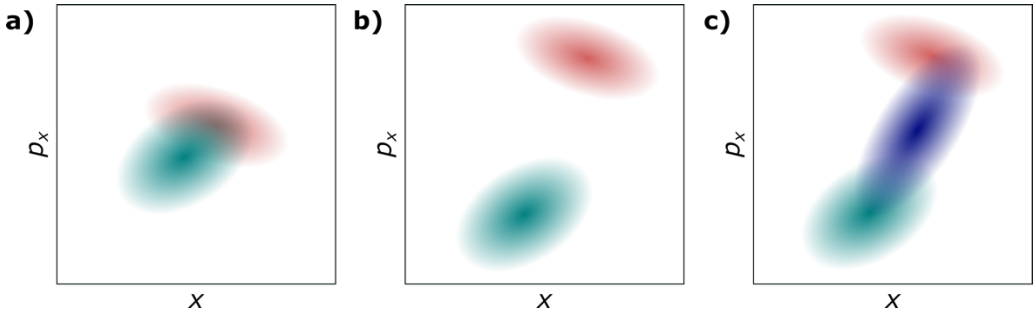


Figure 45. The overlap of the phase space of two ensembles determines the convergence of FEP calculations. If two ensemble representative states (red and green) are overlapping (**a**), the calculations are expected to converge, giving an accurate free energy result. If the states are too distinct from each other (**b**), the calculation will not converge. However, the calculation can reach convergence if mutually overlapping intermediate states (blue) are introduced (**c**) [68].

The number of states (N) and the values of λ should be determined according to a system, although ideally, the space between states should be constant to facilitate further calculations [6,68,69].

II.6.3.2. Thermodynamic integration

The free energy difference between state 0 and state 1 can be expressed as finite integrate using thermodynamic integration (TI) method:

$$\Delta A_{0\to1} = A(\lambda_1) - A(\lambda_0) = \int_{\lambda_0}^{\lambda_1} \frac{dA(\lambda)}{d\lambda} d\lambda \qquad (51)$$

For each simulation corresponding to discrete value of λ, the ensemble is averaged. The free energy corresponds to the integrate along λ values.

$$\Delta A_{0\to1} = \int_{\lambda_0}^{\lambda_1} \langle \frac{\partial \mathcal{H}(x, p_x; \lambda)}{\partial \lambda} \rangle_\lambda \, d\lambda \qquad (52)$$

The main difference between TI and FEP is the criterion of convergence. While in FEP the ensemble representative states must overlap, in TI simulations the convergence occur a by smooth transition of $A(\lambda)$ [1,6,68].

II.6.3.3. Thermodynamic cycle

Essentially, all sorts of free energy calculations can be constructed as thermodynamic cycles in which appropriate end states are identified. Because free energy is a state function, in theory, the reaction between end states does not depend on the path. However, in practice, determining a meaningful pathway is essential [70]. The idea of the thermodynamic cycle relies on the first law of thermodynamics, i.e., there is no energy gain or loss inside the cycle, hence the value cycle should be 0.

$$\Delta G_2 - \Delta G_{binding} = \Delta G_1 - \Delta G_3 \qquad (53)$$

The thermodynamic cycle can be used to compute relative or absolute binding free energy by establishing appropriate end states and conducting two separate simulations. Consider the

free energy binding of a molecule to a protein as an example of a thermodynamic cycle (Fig. 46A).

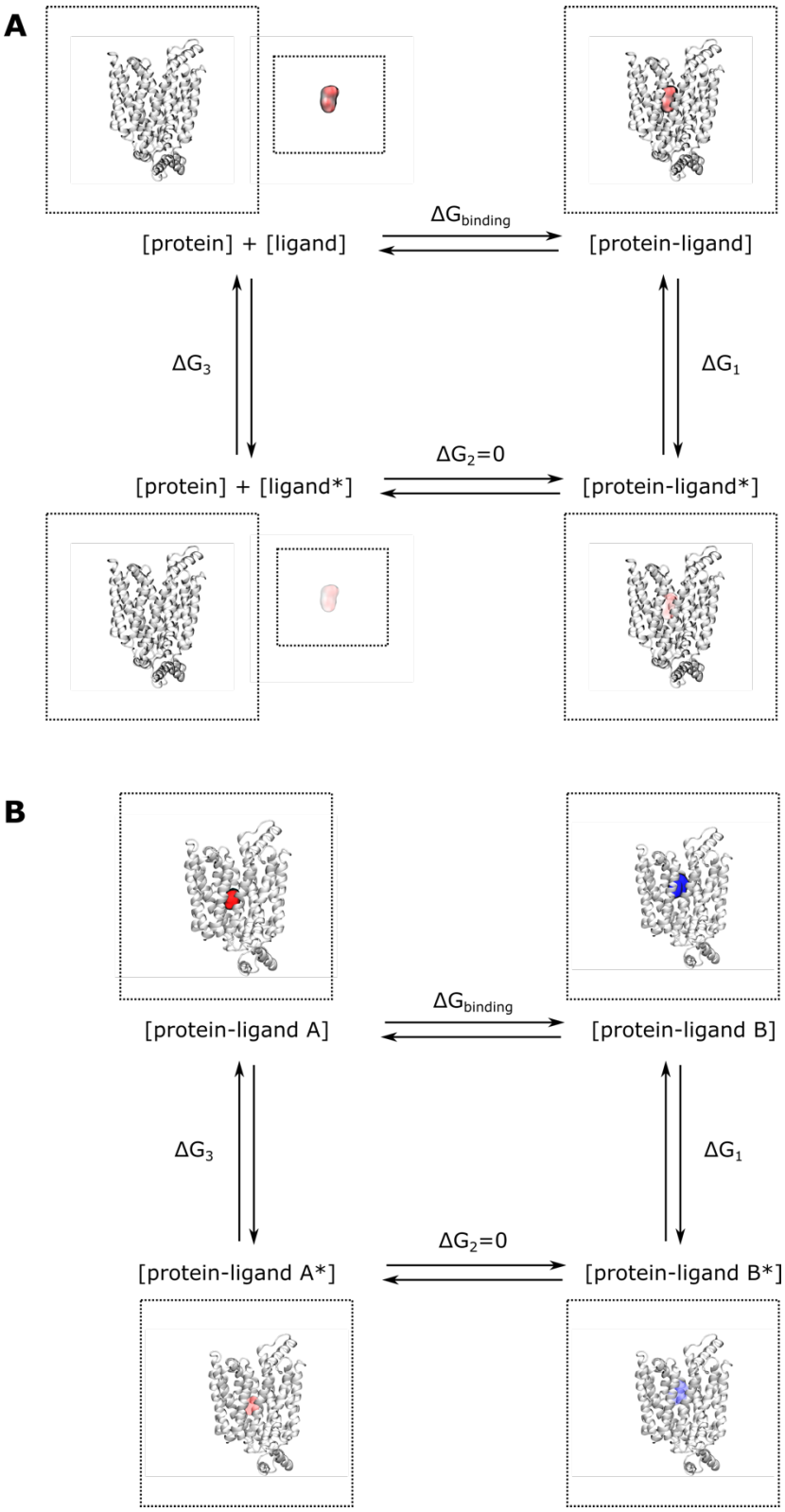


Figure 46. The thermodynamic cycle of free binding **A**) Absolute binding free energy takes into account ligand-protein bound systems, unbound protein systems, and free ligand systems. **B**) Relative binding free energy calculates the energy difference between protein systems bound to different ligands.

The alchemical transformation considers non-bonding forces between ligand and protein, and since the system involves isothermal-isobaric conditions, the energy is expressed as Gibbs free energy $\Delta G = -RTlnK$. As a conclusion, using the thermodynamic cycle depicted in Figure 46 A and B, the binding free energy can be calculated by performing simulations corresponding to ΔG_1 and ΔG_3:

$$\Delta G_{binding} = \Delta G_3 - \Delta G_1 \qquad (54)$$

Molecular states are derived from molecular dynamic simulations and then analysed statistically as part of the conventional free energy computation methodology. There is always a necessity for error analysis since the outcomes are never exact but rather estimated statistically by sampling thermodynamic probability distributions [1].

In general, proper free energy calculations are difficult to perform. To begin with, the simulation setup can be sensitive to certain parameter selections, requiring a great deal of attention and manual modifications while creating input files. Several challenges relating to the nature of free energy calculations, such as *end-point disaster*, *particle collapse problem*[1], or *large gradient-jump problem*[2], may arise [70]. Furthermore, there is a wide range of procedures and protocols accessible for application; nevertheless, a methodology that is not appropriate for a given system and improperly selected/adjusted parameters might yield a high number of potential errors. Even though a lot of issues have been addressed, finding appropriate solutions. For instance, the end-point catastrophe arises from a sharp divergence of free energy contribution in the initial $\lambda \cong 0$ and the end $\lambda \cong 1$ points. However, that can be prevented by using soft core potential, which smooths or "softens" the interactions to decouple by introducing adjustable parameters [69,70].

II.6.4. Enhanced Molecular Dynamics sampling

Due to inadequate sampling of high energy areas, traditional simulations cannot properly sample conformational space and thus provide free energy differences. Low free energy states characterise a system in equilibrium, displaying ideally the global minimum or, at the very least, a local minimum on the energy surface. Therefore, sampling of those states is usually abundant by conventional MD. However, the sampling of high energy states is not as trivial.

Many enhanced sampling methods have been developed in order to boost sampling efficiency to "force" or "carry" the system from one minimum to another. In this framework, MD simulations can be extended by sampling a broader region of the conformational hyperspaces. However, the choice of methods is not less difficult depending on the system. The choice of a method is strongly dependent on the nature of the issue; therefore, the method must be chosen carefully, since an inappropriate choice can provide unexpected results and, worse yet, lead to incorrect conclusions. Regardless of the methods, the sampling algorithm should be able to sample all important configurations for the biological event of interest [4].

Besides, enhanced sampling techniques can generally be classified into those that rely on collective variables (CVs) and those that do not. A collective variable can be understood as a descriptor of a molecular system state and the transition of interest. The CVs represent a broad

[1] Particle collapse problem arises when new spurious minima appear in the intermediate state as a result of an imbalance between Coulomb attraction and exchange repulsion.

[2] Large gradient-jump issues are caused by spurious jumps in the free energy close to the thermodynamic end points.

range of functions depending on atomic coordinates that allow one to calculate a force acting on atoms from the force acting on a CV. Typically, slow degrees of freedom are sought for CVs in order to drive a transition to unpopulated conformations. Given that the representation of collective variables is usually rather complicated, machine-learning (ML) techniques have become the gold standard for generating relevant CVs for a given system [71,72]. However, the methods can be categorised in various ways, indicating that there is more than one correct classification. A recent categorization of available methods has been made by Hénin et. al., [73] in which the approaches were divided based on the main methodological questions. That way, the methods are classified based on a system characteristic to help make an appropriate decision regarding a suitable method.

It is important to introduce the concept of free energy estimators before moving on to the introduction of method classes. In general, free energy estimators are mathematical formulas that are used to determine the free energy, whether it is free energy surface or free energy difference. The estimator of free energy is unique to the selected method, or at least shares the same concept as its derivative method.

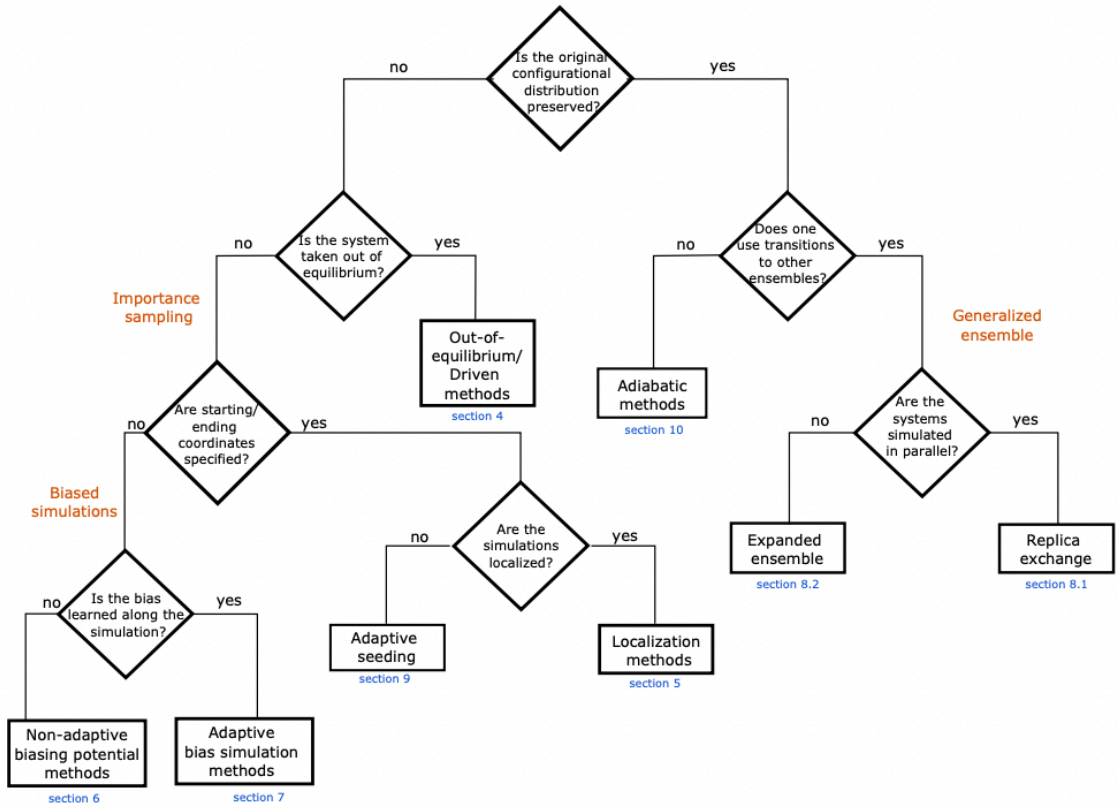


Figure 47. Figure adopted from *Enhanced sampling methods for molecular dynamics simulations* [73].

For example, estimators of the free energy difference:

- the free energy difference may rely on a ratio of occurred configuration while visiting two distinct states,
- the free energy difference may be estimated by integrating the average ensembles along the coupling parameter using a trapezoidal rule as in TI,

- in case of a significant overlap between thermodynamic states, the free energy difference may be estimated by exponential averaging the potential energy of both states,
- Bennett's Acceptance Ratio (BAR) relies on the energy sampled in both states using the thermodynamic identity.

Estimators of free energy:

- the estimation of free energy may rely upon the count or probability of a given transition occurrence along a given CV,
- Multistate Bennett Acceptance Ratio (MBAR) estimates free energies by summing all the energy samples collected from a state,
- The methods based on histograms sum the corrected energies over histogram bins.

Basically, the enhanced sampling methods have been categorised into eight classes of methods. However, considering different definitions, a given method can be associated with more than one class. The classes are introduced below in brief, without a full discussion of concepts and techniques [73].

Exploration of configuration space occurs in **out-of-equilibrium methods** by forcing the system to follow a specific direction of transition indicated by CV(s) or alchemical parameter. Targeted MD [74] is a classic example in which the initial state is forced to achieve a specific value of RMSD (root mean square deviation) to adopt the expected end (targeted) state. However, RMSD is rarely a suitable CV since, even though its value may meet the expectations, the observed conformation may not entirely correspond to the target structure. SMD, or steered MD, also corresponds to that description. Applying an external force on atoms (for example, pulling them) to cause a conformation change is the basis of SMD. The Jarzynski identity may then be used to calculate the free energy of intermediate states:

$$e^{-\beta(F_\lambda - F_0)} = \langle e^{-\beta W_\lambda} \rangle, \qquad (55)$$

which uses weights according to λ parameter [75].

The **localization methods** rely on sampling a narrow, well-defined area of configuration space, known as "stratum" or a window. The simulation can be constrained either by the sampling area (free energy surface) or the potential to remain in the sampling area. A popular approach found in that class is Umbrella Sampling (US) [76], where the CV that drives the transition is assumed to be known. Then the reaction path is broken down into small conformational areas (or windows) where the system is restricted by harmonic potential [4,73].

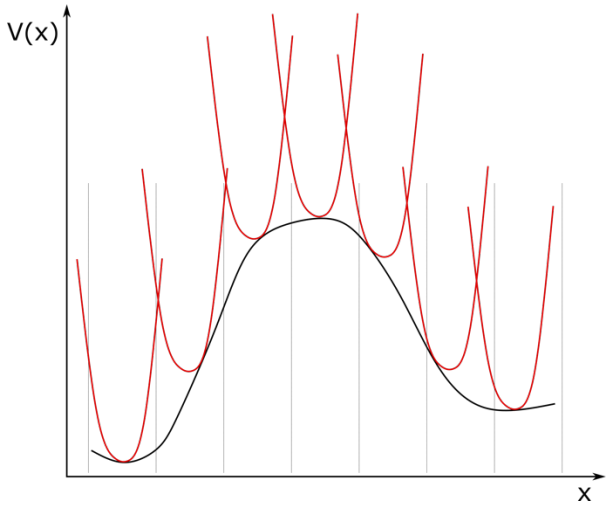


Figure 48. The simulation in umbrella sampling uses stratums (or windows) to break down the reaction path into several sampling configuration spaces. Each of the configurations is constrained by harmonic potential (red “umbrellas”).

Accelerated MD [77] and Gaussian-accelerated MD [78,79] approaches fall into the category of **Non-adaptive Biasing Potential methods,** where the parameters are manipulated in a way to flatten the energy landscape. It often concerns decreasing energy barriers of transitions along a given parameter (for example, dihedral angle). Then, the unbiased statistics can be reconstructed by reweighting [73].

The concept behind **Adaptive Biasing Potential Methods** is to provide an external potential that biases a system's dynamics through a certain CV space. The free energy is balanced by the external bias potential that aims to flatten sampling in that CV space. The external bias is adjusted (*adapted*) in an iterative manner, since the free energy surface is unknown. As for every single CV-oriented enhanced sampling method, the selection of CVs must be made with special caution. Metadynamics [80–82] is a typical example of a method in this category. In this approach, sampling is done by adding a time-dependent gaussian-shaped bias along CVs that correspond to slow degrees of freedom. By using CVs, the FES may be calculated from the biassed potential, assuming convergence at the end of the simulations. Over the years, metadynamics has developed several variants of the method, including, for example, well-tempered metadynamics [81]. Numerous techniques may be recognised, including adaptive biassing MD (ABMD) [83] and Gaussian-mixture umbrella sampling [84].

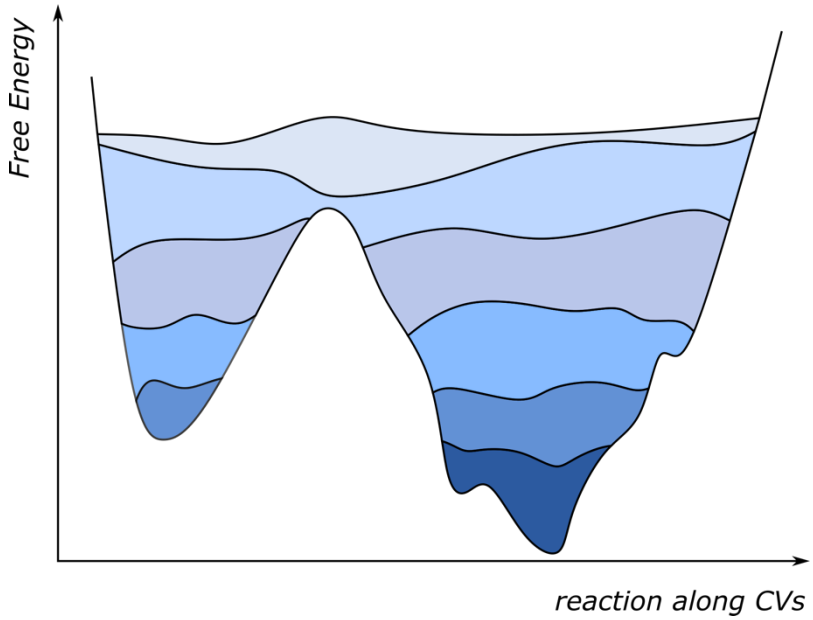


Figure 49. In the concept of metadynamics, the free energy is fulfilled with the biassed potential along a given CV, in order to flatten the energy surface and enable one to visit all the possible thermodynamic states.

Generalized ensemble encompass a wide variety of methodologies including the range of replica exchange methods [85], simulated tempering [86] and simulated scaling [87]. The simulation framework allows to sample all the thermodynamic states by visiting them in a stochastic manner. The principle of replica exchange method is to perform several simulations of a system in different temperatures or Hamiltonians parallelly, which periodically exchange thermodynamic states between the simulations. The exchange of thermodynamic states is evaluated based on Metropolis criterion [4,73].

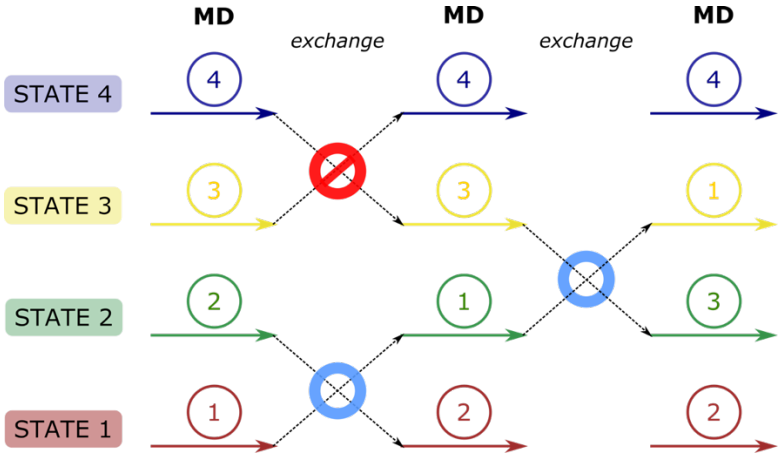


Figure 50. The concept of replica exchange simulations. Multiple copies of a system in different thermodynamic states undergo parallel MD simulations that exchange the states periodically. The exchange may be accepted (blue circles) or rejected (red crossed circles) based on the Metropolis criterion.

The adaptive seeding methods aim to obtain an accurate description of an ensemble by focusing on sampling particular conformational space regions, to then join the trajectories and create a solid model containing all the relevant thermodynamic states. That is achieved using Markov State modelling that takes into account the probability distribution of microstates as well as the rates of transitions between them at a given time lag [73,88].

The free energy calculation using **adiabatic methods** aims to obtain Markovian-like dynamics along a CV due to the fixed separation of timescales (it may be fixed by alchemical parameter). The concept is implemented in methods such as the adiabatic free energy dynamics (AFED) [89] or Temperature accelerated molecular dynamics (TAMD) [90].

II.7. Modelling Proteins

II.7.1. Protein sequence alignment

Prediction and modelling of protein structures have always been challenging tasks in the field of computational chemistry. The aim of sequence alignment is to arrange the amino acid sequences of at least two proteins with a high identity/similarity score so that the residues may share structural, functional, or evolutionary relationships. The origin of alignment reaches algorithms developed by Needleman and Wunsh [91] and Smith and Waterman [64], which consider global and local alignment, respectively. While the global algorithm searches for similarities end-to-end, the local alignment seeks for local similar regions. The Needleman-Wunsh approach is based on dynamic programming, which is thought to produce rather good alignment. Nevertheless, the exhaustive search is computationally intensive, making it unsuitable for large sequences. This is particularly true in the situation of multiple sequence alignment (MSA), in which more than two structures are aligned. Heuristic approaches, on the other hand, are faster and less demanding because the alignment occurs in smaller search spaces, but they do not guarantee the optimal alignment [1,92].

The alignment is composed of homologous residues and deleted/inserted entities (gaps) in order to produce the same length of all sequences, simultaneously having residues that share common features at the same positions in different sequences. Considering MSA, a matrix of size N x M is produced, where N is the number of sequences and M is the number of columns containing residues and gaps. It is assumed that within a column, the residues are homologous. As a result, the evolutionary relationship can be estimated. Many applications can be derived from the evolutionary relationship, including structure prediction, sequence conservation, and single nucleotide variations. The methods of MSA use a heuristic approach where the computational effort is reduced. Some of the popular methods, including ClustalW [93] and T-Coffee [94], use progressive techniques that build the final MSA by aligning the most-alike pairs of sequences and progressively aligning the least similar sequences. In the initial step, the algorithm builds a guide tree that clusters sequences according to their similarity to progressively align and add entities to the MSA [1,95]. However, if the alignment is poor at the early stage, the error propagates to the next sequences.

The iterative methods are based on the same principle as progressive methods, with the improvement of iteratively realigning sequences until the error is minimal. The iterative method, MUSCLE [96], achieves excellent accuracy with a large number of sequences in a computationally affordable amount of time. The algorithm first creates a rough alignment, and then using the *k-mer distance*[1], it clusters the obtained matrix to create the initial guide tree. The second stage is almost identical to the first, except a *Kimura* distance is used in place of *k-mer* to produce a more precise matrix and refine the guide tree. The MSA is then improved iteratively by realigning sequences until the threshold SP-score[2] is met [95,96].

[1] K-mer distance is employed to estimate evolutionary distance using multiple alignments.

[2] SP-score is essentially the sum of all pairwise sequence scores.

II.7.2. Predicting protein structure

II.7.2.1. Background

Even though a myriad of protein sequences has been made available, only a small fraction of structures have been experimentally resolved in the Protein Data Bank [97] (PDB; http://www.rcsb.org/). Protein structure elucidation is usually carried out by means of X-ray crystallography, cryogenic electron microscopy (cryo-EM) and, to a lesser extent, nuclear magnetic resonance (NMR). However, difficulties in purification, production, and access to techniques continue to make experimental protein structure elucidation challenging.

Biological functions are often associated with protein structure. Therefore, investigating three-dimensional protein structures and dynamics is of biological, pharmacological, and medical relevance. Over the past decades, computational approaches to predict and/or support experimental observations, if available, have become necessary to rationalise protein structure-function relationships.

Protein structures were shown to structurally adapt into secondary structures (e.g., α-helices, β-sheets and loops) which may be structurally arranged into several domains that are of particular importance for their function. Protein structure can display patterns that are common to a particular class of proteins despite lacking regularity and perfect symmetry [1,95]. Proteins can be distinguished by considering their structural and evolutionary relationships. The primary classification yields *families*, *superfamilies*, and below them, *fold* classes according to Structural Classification of Proteins (SCOP) database [98]. Proteins within *families* exhibit close evolutionary relationships due to shared functions, structural patterns (domains), and typically high sequence identity and/or similarity. Even though proteins may display relatively low sequence identity when assigned to a *superfamily*, they often share structural and functional properties. These are further subdivided into structural folds, which exhibit specific domain topology organised in a consistent pattern [98,99]. Computational methods take advantage of structural evolutionary relationships, providing methods for protein structure prediction when no experimental structure is available.

II.7.2.2. Comparative modelling

Comparative ("homology") modelling attempts to predict the protein structure based on its alignment with one or more related proteins of a known structure, so-called *templates*. Essentially, this approach assigns the fold of the target sequence based on identified similarities with the template(s). The target sequence is then aligned to the template(s), which serves as a foundation for building a model. Finally, the relevance of the model must be evaluated (Fig. 51).

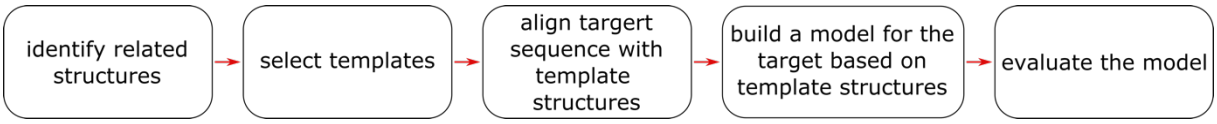


Figure 51. Step diagram for comparative modelling.

The procedure depends on the templates used. Consequently, particular care should be taken when selecting them. It is important to remember that database search tools like BLAST and FASTA are especially useful for finding suitable templates. The sequence identity between a target and a template should be greater than 30%. Following the selection of the template(s), the sequences should be aligned, for instance using the methods for sequence alignment

mentioned above. There are available computer programs (e.g., Modeller [100], Rosetta [101]) as well as online servers (e.g., i-TASSER [102], Robetta [101], SWISS-MODEL [103]) that automatize the comparative modelling. However, due to uncertainty of loops and side-chain conformation, the obtained model may display low confidence. The loops could be enhanced by using high-resolution fragment libraries like SCOP or CATH [104]. It is recommended to refine the model with molecular dynamics [105].

II.7.2.3. Model quality

There are several parameters which must be monitored to assess the model quality. For instance, a model of a native protein should:

(i) exhibit distributions of phi/psi (Φ/Ψ) torsion angles (Ramachandran plot) in the "allowed" area;

(ii) minimize contacts of hydrophobic residues with water and polar environment;

(iii) display essential interactions including covalent bonds, salt bridges, disulphide bridges, and hydrogen bonds [105].

The model can be evaluated by PROTCHECK [106] or MolProbity [107], which assess whether the structure is plausible by considering Ramachandran plot, steric clashes, hydrogen bonds, and non-equilibrium bond angles. The model may also be evaluated by ensuring that the protein fold lies in the energetically lowest state. A popular model quality assessment method, DOPE [108], uses statistical potential dependent on interatomic distances. Likewise, other assessment methods exist, such as SOAP [109]. It includes solvent accessibility, interatomic contact areas, angles, and distances.

More recently, a model can be assessed by machine-learning methods using data derived from MSA, structural features, and physico-chemical properties (e.g., ProQ3 [110], DeepQA [111], and MQAPRank [112]).

II.7.2.4. Machine-Learning based structure prediction

Every second year since 1994, the Critical Assessment of Protein Structure Prediction (CASP) competition has been carried out to evaluate structure prediction algorithms. In 2020, the DeepMind team introduced AlphaFold2 (AF2) [113], an Artificial Intelligence (AI) algorithm that predicts protein structures with accuracy assumed to be comparable to experimental methods, winning the CASP14 competition [114]. The high accuracy is achieved by employing a neural network architecture that has been trained on experimentally resolved structures from the Protein Data Bank [97] to learn about physical and geometric properties as well as, evolutionary conservation information from MSA.

To briefly summarise the approach, AF2 searches databases using the amino acid sequence as input to produce a multiple sequence alignment (MSA). Simultaneously, the algorithm identifies complex relationships of protein residues and creates a *pair representation*, indicating residues likely to be in contact, accounting for an initial structure. The MSA and the pair representation are then used in an *evoformer*[1], which processes and identifies the essential information to iteratively modify both the MSA and the pair representation. The

[1] Evoformer can be understood as neural network that transforms the evolutionary information.

refined matrices are then utilised to generate the final model that goes through the *evoformer* again to refine the prediction.

It is worth undersigning the central role of both MSA and pair representation, which stand for the idea of co-evolving residues. It indicates that if two residues are in close contact, it is very likely that if one of them mutates, the other residue will adapt in order to preserve the structure. For instance, this corresponds to the preservation of electrostatic interaction within the protein that maintains the tertiary structure. In other words, the correlated mutation contains structural information. More than 65,000 and 214,000,000 reported structures are now available in both the PDB database and AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/), respectively. It is worth noting that the predicted structures contain per-residue confidence score estimates that correspond to predicted local-distance differences of C_α ($pLDDT$-C_α).

Although the AlphaFold2 structure prediction is ground-breaking and can be used in different ways, yet there is still potential for improvement. Despite the fact that hundreds of thousands of protein structures have been recorded, the database only records one conformational state, which is a major drawback. Since a protein may adopt different conformations, it is of interest to explore alternative conformations. Fortunately, an open-source AF2 code enables one to predict a protein of choice in a desired state by choosing appropriate templates. Furthermore, identifying the impact of missense mutations on protein structure is essential when determining the biological effect. AF2 has been proven to reliably predict wild-type structures, however it is unsuccessful at predicting the effects of missense mutations on a structure [115]. Moreover, the protein's whole biological function cannot be determined only from its structure alone. Since metal ions, post-translational modifications, cofactors, or interactions with other proteins or nucleic acids complement the function of a protein, it is important to know their interactions and position in the structure. AF2 does not predict any of those features so far. Nonetheless, this does not exclude further ligand screening or docking calculations on AF2 structures [116].

The revolution of AlphaFold has paved the way toward the development of machine-learning methods that attempt to fill the missing elements of the protein structure prediction field. Therefore, it is essential to bring attention to RoseTTAFold, a different structure prediction machine-learning algorithm recently released by Baker lab [117]. The performance of RoseTTAFold is considered nearly as good as AF2. RoseTTAFold's indisputable advantage over AlphaFold2 is its capacity to predict protein-protein complex models. Likewise, RoseTTAFold, a very recently introduced I-TASSER-MTD [118], presents a method to fill a niche in protein structure prediction. I-TASSER-MTD is publicly available platform that offers a fully automated pipeline for modelling large multi-domain protein structures and annotation of structure function based only on the sequence.

II.8. References

[1] A.R. Leach, Molecular modelling: principles and applications, 2nd ed, Prentice Hall, Harlow, England ; New York, 2001.

[2] T.P. Senftle, S. Hong, M.M. Islam, S.B. Kylasa, Y. Zheng, Y.K. Shin, C. Junkermeier, R. Engel-Herbert, M.J. Janik, H.M. Aktulga, T. Verstraelen, A. Grama, A.C.T. van Duin, The ReaxFF reactive force-field: development, applications and future directions, Npj Comput Mater. 2 (2016) 15011. https://doi.org/10.1038/npjcompumats.2015.11.

[3] A.R. Leach, Molecular modelling: principles and applications, Repr, Longman, Harlow, 1997.

[4] L. Monticelli, E. Salonen, eds., Biomolecular Simulations, Humana Press, Totowa, NJ, 2013. https://doi.org/10.1007/978-1-62703-017-5.

[5] F. Jensen, Introduction to computational chemistry, 2nd ed, John Wiley & Sons, Chichester, England ; Hoboken, NJ, 2007.

[6] A. Kukol, ed., Molecular Modeling of Proteins, Humana Press, Totowa, NJ, 2008. https://doi.org/10.1007/978-1-59745-177-2.

[7] P.W. Atkins, J. De Paula, Physical chemistry: thermodynamics, structure, and change, Tenth edition, W.H. Freeman, New York, 2014.

[8] L.M. Salonen, M. Ellermann, F. Diederich, Aromatic Rings in Chemical and Biological Recognition: Energetics and Structures, Angew. Chem. Int. Ed. 50 (2011) 4808–4842. https://doi.org/10.1002/anie.201007560.

[9] 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. https://doi.org/10.1021/ja00124a002.

[10] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217. https://doi.org/10.1002/jcc.540040211.

[11] W.R.P. Scott, P.H. Hünenberger, I.G. Tironi, A.E. Mark, S.R. Billeter, J. Fennen, A.E. Torda, T. Huber, P. Krüger, W.F. van Gunsteren, The GROMOS Biomolecular Simulation Program Package, J. Phys. Chem. A. 103 (1999) 3596–3607. https://doi.org/10.1021/jp984217f.

[12] N.L. Allinger, Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms, J. Am. Chem. Soc. 99 (1977) 8127–8134. https://doi.org/10.1021/ja00467a001.

[13] N.L. Allinger, Y.H. Yuh, J.H. Lii, Molecular mechanics. The MM3 force field for hydrocarbons. 1, J. Am. Chem. Soc. 111 (1989) 8551–8566. https://doi.org/10.1021/ja00205a001.

[14] P. Ren, J.W. Ponder, Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation, J. Phys. Chem. B. 107 (2003) 5933–5947. https://doi.org/10.1021/jp027815+.

[15] P. Ren, J.W. Ponder, Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations, J. Comput. Chem. 23 (2002) 1497–1506. https://doi.org/10.1002/jcc.10127.

[16] J.A. Lemkul, J. Huang, B. Roux, A.D. MacKerell, An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications, Chem. Rev. 116 (2016) 4983–5013. https://doi.org/10.1021/acs.chemrev.5b00505.

[17] J.H. Lii, N.L. Allinger, Molecular mechanics. The MM3 force field for hydrocarbons. 2. Vibrational frequencies and thermodynamics, J. Am. Chem. Soc. 111 (1989) 8566–8575. https://doi.org/10.1021/ja00205a002.

[18] G.A. Kaminski, H.A. Stern, B.J. Berne, R.A. Friesner, Y.X. Cao, R.B. Murphy, R. Zhou, T.A. Halgren, Development of a polarizable force field for proteins via ab initio quantum chemistry: First generation model and gas phase tests, J. Comput. Chem. 23 (2002) 1515–1531. https://doi.org/10.1002/jcc.10125.

[19] W. Damm, A. Frontera, J. Tirado-Rives, W.L. Jorgensen, OPLS all-atom force field for carbohydrates, J. Comput. Chem. 18 (1997) 1955–1970. https://doi.org/10.1002/(SICI)1096-987X(199712)18:16<1955::AID-JCC1>3.0.CO;2-L.

[20] 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. https://doi.org/10.1021/ct400314y.

[21] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A.E. Dawid, A. Kolinski, Coarse-Grained Protein Models and Their Applications, Chem. Rev. 116 (2016) 7898–7936. https://doi.org/10.1021/acs.chemrev.6b00163.

[22] B. Kim, T. Young, E. Harder, R.A. Friesner, B.J. Berne, Structure and Dynamics of the Solvation of Bovine Pancreatic Trypsin Inhibitor in Explicit Water: A Comparative Study of the Effects of Solvent and Protein Polarizability, J. Phys. Chem. B. 109 (2005) 16529–16538. https://doi.org/10.1021/jp051569v.

[23] V. Babin, J. Baucom, T.A. Darden, C. Sagui, Molecular Dynamics Simulations of DNA with Polarizable Force Fields: Convergence of an Ideal B-DNA Structure to the Crystallographic Structure, J. Phys. Chem. B. 110 (2006) 11571–11581. https://doi.org/10.1021/jp061421r.

[24] Y. Tong, C.G. Ji, Y. Mei, J.Z.H. Zhang, Simulation of NMR Data Reveals That Proteins' Local Structures Are Stabilized by Electronic Polarization, J. Am. Chem. Soc. 131 (2009) 8636–8641. https://doi.org/10.1021/ja901650r.

[25] D. Jiao, P.A. Golubkov, T.A. Darden, P. Ren, Calculation of protein–ligand binding free energy by using a polarizable potential, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 6290–6295. https://doi.org/10.1073/pnas.0711686105.

[26] Z. Jing, C. Liu, S.Y. Cheng, R. Qi, B.D. Walker, J.-P. Piquemal, P. Ren, Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications, Annu. Rev. Biophys. 48 (2019) 371–394. https://doi.org/10.1146/annurev-biophys-070317-033349.

[27] M. Karplus, J.A. McCammon, Molecular dynamics simulations of biomolecules, Nat. Struct Biol. 9 (2002) 646–652. https://doi.org/10.1038/nsb0902-646.

[28] A.D. MacKerell, J. Wiorkiewicz-Kuczera, M. Karplus, An all-atom empirical energy function for the simulation of nucleic acids, J. Am. Chem. Soc. 117 (1995) 11946–11975. https://doi.org/10.1021/ja00153a017.

[29] T.A. Soares, P.H. Hünenberger, M.A. Kastenholz, V. Kräutler, T. Lenz, R.D. Lins, C. Oostenbrink, W.F. van Gunsteren, An improved nucleic acid parameter set for the GROMOS force field: Improved Nucleic Acid Parameter Set, J. Comput. Chem. 26 (2005) 725–737. https://doi.org/10.1002/jcc.20193.

[30] P. Cieplak, J. Caldwell, P. Kollman, Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases, J. Comput. Chem. 22 (2001) 1048–1057. https://doi.org/10.1002/jcc.1065.

[31] X. Peng, Y. Zhang, H. Chu, Y. Li, D. Zhang, L. Cao, G. Li, Accurate Evaluation of Ion Conductivity of the Gramicidin A Channel Using a Polarizable Force Field without Any

Corrections, J. Chem. Theory Comput. 12 (2016) 2973–2982. https://doi.org/10.1021/acs.jctc.6b00128.

[32] J.W. Ponder, C. Wu, P. Ren, V.S. Pande, J.D. Chodera, M.J. Schnieders, I. Haque, D.L. Mobley, D.S. Lambrecht, R.A. DiStasio, M. Head-Gordon, G.N.I. Clark, M.E. Johnson, T. Head-Gordon, Current Status of the AMOEBA Polarizable Force Field, J. Phys. Chem. B. 114 (2010) 2549–2564. https://doi.org/10.1021/jp910674d.

[33] C. Tian, K. Kasavajhala, K.A.A. Belfon, L. Raguette, H. Huang, A.N. Migues, J. Bickel, Y. Wang, J. Pincay, Q. Wu, C. Simmerling, ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution, J. Chem. Theory Comput. 16 (2020) 528–552. https://doi.org/10.1021/acs.jctc.9b00591.

[34] 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. https://doi.org/10.1021/jp101759q.

[35] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. de Groot, H. Grubmüller, A.D. MacKerell, CHARMM36m: an improved force field for folded and intrinsically disordered proteins, Nat Methods. 14 (2017) 71–73. https://doi.org/10.1038/nmeth.4067.

[36] 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. https://doi.org/10.1021/ct4010307.

[37] I. Marzuoli, C. Margreitter, F. Fraternali, Lipid Head Group Parameterization for GROMOS 54A8: A Consistent Approach with Protein Force Field Description, J. Chem. Theory Comput. 15 (2019) 5175–5193. https://doi.org/10.1021/acs.jctc.9b00509.

[38] 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: GLYCAM06, J. Comput. Chem. 29 (2008) 622–655. https://doi.org/10.1002/jcc.20820.

[39] O.T. Unke, S. Chmiela, H.E. Sauceda, M. Gastegger, I. Poltavsky, K.T. Schütt, A. Tkatchenko, K.-R. Müller, Machine Learning Force Fields, Chem. Rev. 121 (2021) 10142–10186. https://doi.org/10.1021/acs.chemrev.0c01111.

[40] L. Verlet, Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules, Phys. Rev. 159 (1967) 98–103. https://doi.org/10.1103/PhysRev.159.98.

[41] W.C. Swope, H.C. Andersen, P.H. Berens, K.R. Wilson, A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters, The Journal of Chemical Physics. 76 (1982) 637–649. https://doi.org/10.1063/1.442716.

[42] R.W. Hockney, S.P. Goel, J.W. Eastwood, Quiet high-resolution computer models of a plasma, Journal of Computational Physics. 14 (1974) 148–158. https://doi.org/10.1016/0021-9991(74)90010-2.

[43] V. Kr□utler, W.F. van Gunsteren, P.H. H□nenberger, A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations, J. Comput. Chem. 22 (2001) 501–508. https://doi.org/10.1002/1096-987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V.

[44] M. Elstner, K.J. Jalkanen, M. Knapp-Mohammady, T. Frauenheim, S. Suhai, Energetics and structure of glycine and alanine based model peptides: Approximate SCC-DFTB, AM1 and PM3 methods in comparison with DFT, HF and MP2 calculations, Chemical Physics. 263 (2001) 203–219. https://doi.org/10.1016/S0301-0104(00)00375-X.

[45] M.P. Allen, D.J. Tildesley, Computer simulation of liquids, Second edition, Oxford University Press, Oxford, United Kingdom, 2017.

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

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

[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. https://doi.org/10.1063/1.448118.

[49] T. Schneider, E. Stoll, Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions, Phys. Rev. B. 17 (1978) 1302–1322. https://doi.org/10.1103/PhysRevB.17.1302.

[50] C.P. Lowe, An alternative approach to dissipative particle dynamics, Europhysics Letters (EPL). 47 (1999) 145–151. https://doi.org/10.1209/epl/i1999-00365-x.

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

[52] S.C. Harvey, R.K.-Z. Tan, T.E. Cheatham, The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition, J. Comput. Chem. 19 (1998) 726–740. https://doi.org/10.1002/(SICI)1096-987X(199805)19:7<726::AID-JCC4>3.0.CO;2-S.

[53] T. Morishita, Fluctuation formulas in molecular-dynamics simulations with the weak coupling heat bath, The Journal of Chemical Physics. 113 (2000) 2976–2982. https://doi.org/10.1063/1.1287333.

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

[55] C.J. Cramer, Essentials of computational chemistry: theories and models, 2nd ed, Wiley, Chichester, West Sussex, England ; Hoboken, NJ, 2004.

[56] 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. https://doi.org/10.1063/1.445869.

[57] 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, Dordrecht, 1981: pp. 331–342. https://doi.org/10.1007/978-94-015-7658-1_21.

[58] X.-Y. Meng, H.-X. Zhang, M. Mezei, M. Cui, Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery, CAD. 7 (2011) 146–157. https://doi.org/10.2174/157340911795677602.

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

[60] E. Fischer, Einfluss der Configuration auf die Wirkung der Enzyme, Ber. Dtsch. Chem. Ges. 27 (1894) 2985–2993. https://doi.org/10.1002/cber.18940270364.

[61] D.E. Koshland, Correlation of Structure and Function in Enzyme Action: Theoretical and experimental tools are leading to correlations between enzyme structure and function., Science. 142 (1963) 1533–1541. https://doi.org/10.1126/science.142.3599.1533.

[62] 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. https://doi.org/10.1016/0022-2836(82)90153-X.

[63] R.L. DesJarlais, R.P. Sheridan, J.S. Dixon, I.D. Kuntz, R. Venkataraghavan, Docking flexible ligands to macromolecular receptors by molecular shape, J. Med. Chem. 29 (1986) 2149–2153. https://doi.org/10.1021/jm00161a004.

[64] A.T. Brint, P. Willett, Algorithms for the identification of three-dimensional maximal common substructures, J. Chem. Inf. Comput. Sci. 27 (1987) 152–158. https://doi.org/10.1021/ci00056a002.

[65] D.S. Goodsell, H. Lauble, C.D. Stout, A.J. Olson, Automated docking in crystallography: Analysis of the substrates of aconitase, Proteins. 17 (1993) 1–10. https://doi.org/10.1002/prot.340170104.

[66] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, J. Comput. Chem. 19 (1998) 1639–1662. https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.

[67] I. Muegge, Y.C. Martin, A General and Fast Scoring Function for Protein−Ligand Interactions: A Simplified Potential Approach, J. Med. Chem. 42 (1999) 791–804. https://doi.org/10.1021/jm980536j.

[68] C. Chipot, Free Energy Calculations in Biological Systems. How Useful Are They in Practice?, in: B. Leimkuhler, C. Chipot, R. Elber, A. Laaksonen, A. Mark, T. Schlick, C. Schütte, R. Skeel (Eds.), New Algorithms for Macromolecular Simulation, Springer-Verlag, Berlin/Heidelberg, 2006: pp. 185–211. https://doi.org/10.1007/3-540-31618-3_12.

[69] A. Pohorille, C. Jarzynski, C. Chipot, Good Practices in Free-Energy Calculations, J. Phys. Chem. B. 114 (2010) 10235–10253. https://doi.org/10.1021/jp102971x.

[70] T.-S. Lee, B.K. Allen, T.J. Giese, Z. Guo, P. Li, C. Lin, T.D. McGee, D.A. Pearlman, B.K. Radak, Y. Tao, H.-C. Tsai, H. Xu, W. Sherman, D.M. York, Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery, J. Chem. Inf. Model. 60 (2020) 5595–5623. https://doi.org/10.1021/acs.jcim.0c00613.

[71] M.M. Sultan, H.K. Wayment-Steele, V.S. Pande, Transferable Neural Networks for Enhanced Sampling of Protein Dynamics, J. Chem. Theory Comput. 14 (2018) 1887–1894. https://doi.org/10.1021/acs.jctc.8b00025.

[72] O. Fleetwood, M.A. Kasimova, A.M. Westerlund, L. Delemotte, Molecular Insights from Conformational Ensembles via Machine Learning, Biophysical Journal. 118 (2020) 765–780. https://doi.org/10.1016/j.bpj.2019.12.016.

[73] J. Hénin, T. Lelièvre, M.R. Shirts, O. Valsson, L. Delemotte, Enhanced sampling methods for molecular dynamics simulations, (2022). https://doi.org/10.48550/ARXIV.2202.04164.

[74] J. Schlitter, M. Engels, P. Krüger, E. Jacoby, A. Wollmer, Targeted Molecular Dynamics Simulation of Conformational Change-Application to the T ↔ R Transition in Insulin, Molecular Simulation. 10 (1993) 291–308. https://doi.org/10.1080/08927029308022170.

[75] C. Jarzynski, Nonequilibrium Equality for Free Energy Differences, Phys. Rev. Lett. 78 (1997) 2690–2693. https://doi.org/10.1103/PhysRevLett.78.2690.

[76] G.M. Torrie, J.P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, Journal of Computational Physics. 23 (1977) 187–199. https://doi.org/10.1016/0021-9991(77)90121-8.

[77] D. Hamelberg, J. Mongan, J.A. McCammon, Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules, The Journal of Chemical Physics. 120 (2004) 11919–11929. https://doi.org/10.1063/1.1755656.

[78] Y. Miao, J.A. McCammon, Gaussian Accelerated Molecular Dynamics: Theory, Implementation, and Applications, in: Annual Reports in Computational Chemistry, Elsevier, 2017: pp. 231–278. https://doi.org/10.1016/bs.arcc.2017.06.005.

[79] J. Wang, P.R. Arantes, A. Bhattarai, R.V. Hsu, S. Pawnikar, Y.M. Huang, G. Palermo, Y. Miao, Gaussian accelerated molecular dynamics: Principles and applications, WIREs Comput Mol Sci. 11 (2021). https://doi.org/10.1002/wcms.1521.

[80] G. Bussi, A. Laio, Using metadynamics to explore complex free-energy landscapes, Nat Rev Phys. 2 (2020) 200–212. https://doi.org/10.1038/s42254-020-0153-0.

[81] A. Barducci, G. Bussi, M. Parrinello, Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method, Phys. Rev. Lett. 100 (2008) 020603. https://doi.org/10.1103/PhysRevLett.100.020603.

[82] A. Laio, M. Parrinello, Escaping free-energy minima, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 12562–12566. https://doi.org/10.1073/pnas.202427399.

[83] V. Babin, C. Roland, C. Sagui, Adaptively biased molecular dynamics for free energy calculations, The Journal of Chemical Physics. 128 (2008) 134101. https://doi.org/10.1063/1.2844595.

[84] P. Maragakis, A. van der Vaart, M. Karplus, Gaussian-Mixture Umbrella Sampling, J. Phys. Chem. B. 113 (2009) 4664–4673. https://doi.org/10.1021/jp808381s.

[85] A. Mitsutake, Y. Sugita, Y. Okamoto, Generalized-ensemble algorithms for molecular simulations of biopolymers, Biopolymers. 60 (2001) 96–123. https://doi.org/10.1002/1097-0282(2001)60:2<96::AID-BIP1007>3.0.CO;2-F.

[86] E. Marinari, G. Parisi, Simulated Tempering: A New Monte Carlo Scheme, Europhys. Lett. 19 (1992) 451–458. https://doi.org/10.1209/0295-5075/19/6/002.

[87] H. Li, M. Fajer, W. Yang, Simulated scaling method for localized enhanced sampling and simultaneous “alchemical” free energy simulations: A general method for molecular mechanical, quantum mechanical, and quantum mechanical/molecular mechanical simulations, The Journal of Chemical Physics. 126 (2007) 024106. https://doi.org/10.1063/1.2424700.

[88] M.R. Shirts, V.S. Pande, Mathematical Analysis of Coupled Parallel Simulations, Phys. Rev. Lett. 86 (2001) 4983–4987. https://doi.org/10.1103/PhysRevLett.86.4983.

[89] L. Rosso, P. Mináry, Z. Zhu, M.E. Tuckerman, On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles, The Journal of Chemical Physics. 116 (2002) 4389–4402. https://doi.org/10.1063/1.1448491.

[90] L. Maragliano, E. Vanden-Eijnden, A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations, Chemical Physics Letters. 426 (2006) 168–175. https://doi.org/10.1016/j.cplett.2006.05.062.

[91] 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. https://doi.org/10.1016/0022-2836(70)90057-4.

[92] J. Chao, F. Tang, L. Xu, Developments in Algorithms for Sequence Alignment: A Review, Biomolecules. 12 (2022) 546. https://doi.org/10.3390/biom12040546.

[93] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific

gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680. https://doi.org/10.1093/nar/22.22.4673.

[94] C. Notredame, D.G. Higgins, J. Heringa, T-Coffee: A novel method for fast and accurate multiple sequence alignment, J Mol Biol. 302 (2000) 205–217. https://doi.org/10.1006/jmbi.2000.4042.

[95] D.W. Mount, Bioinformatics: sequence and genome analysis, 2nd ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y, 2004.

[96] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797. https://doi.org/10.1093/nar/gkh340.

[97] H. Berman, K. Henrick, H. Nakamura, Announcing the worldwide Protein Data Bank, Nat Struct Mol Biol. 10 (2003) 980–980. https://doi.org/10.1038/nsb1203-980.

[98] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, SCOP: a structural classification of proteins database for the investigation of sequences and structures, J Mol Biol. 247 (1995) 536–540. https://doi.org/10.1006/jmbi.1995.0159.

[99] S. Kaczanowski, P. Zielenkiewicz, Why similar protein sequences encode similar three-dimensional structures?, Theor Chem Acc. 125 (2010) 643–650. https://doi.org/10.1007/s00214-009-0656-3.

[100] A. Šali, T.L. Blundell, Comparative Protein Modelling by Satisfaction of Spatial Restraints, Journal of Molecular Biology. 234 (1993) 779–815. https://doi.org/10.1006/jmbi.1993.1626.

[101] Y. Song, F. DiMaio, R.Y.-R. Wang, D. Kim, C. Miles, T. Brunette, J. Thompson, D. Baker, High-Resolution Comparative Modeling with RosettaCM, Structure. 21 (2013) 1735–1742. https://doi.org/10.1016/j.str.2013.08.005.

[102] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, Nat Protoc. 5 (2010) 725–738. https://doi.org/10.1038/nprot.2010.5.

[103] T. Schwede, SWISS-MODEL: an automated protein homology-modeling server, Nucleic Acids Research. 31 (2003) 3381–3385. https://doi.org/10.1093/nar/gkg520.

[104] J. Abbass, J.-C. Nebel, Customised fragments libraries for protein structure prediction based on structural class annotations, BMC Bioinformatics. 16 (2015) 136. https://doi.org/10.1186/s12859-015-0576-2.

[105] S.D. Lam, S. Das, I. Sillitoe, C. Orengo, An overview of comparative modelling and resources dedicated to large-scale modelling of genome sequences, Acta Crystallogr D Struct Biol. 73 (2017) 628–640. https://doi.org/10.1107/S2059798317008920.

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

[107] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, *MolProbity* : all-atom structure validation for macromolecular crystallography, Acta Crystallogr D Biol Crystallogr. 66 (2010) 12–21. https://doi.org/10.1107/S0907444909042073.

[108] M. Shen, A. Sali, Statistical potential for assessment and prediction of protein structures, Protein Sci. 15 (2006) 2507–2524. https://doi.org/10.1110/ps.062416606.

[109] G.Q. Dong, H. Fan, D. Schneidman-Duhovny, B. Webb, A. Sali, Optimized atomic statistical potentials: assessment of protein interfaces and loops, Bioinformatics. 29 (2013) 3158–3166. https://doi.org/10.1093/bioinformatics/btt560.

[110] A. Leaver-Fay, M. Tyka, S.M. Lewis, O.F. Lange, J. Thompson, R. Jacak, K.W. Kaufman, P.D. Renfrew, C.A. Smith, W. Sheffler, I.W. Davis, S. Cooper, A. Treuille, D.J. Mandell, F. Richter, Y.-E.A. Ban, S.J. Fleishman, J.E. Corn, D.E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J.J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J.J. Gray, B. Kuhlman, D. Baker, P. Bradley, Rosetta3, in: Methods in Enzymology, Elsevier, 2011: pp. 545–574. https://doi.org/10.1016/B978-0-12-381270-4.00019-6.

[111] R. Cao, D. Bhattacharya, J. Hou, J. Cheng, DeepQA: improving the estimation of single protein model quality with deep belief networks, BMC Bioinformatics. 17 (2016) 495. https://doi.org/10.1186/s12859-016-1405-y.

[112] X. Jing, K. Wang, R. Lu, Q. Dong, Sorting protein decoys by machine-learning-to-rank, Sci Rep. 6 (2016) 31571. https://doi.org/10.1038/srep31571.

[113] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature. 596 (2021) 583–589. https://doi.org/10.1038/s41586-021-03819-2.

[114] J. Pereira, A.J. Simpkin, M.D. Hartmann, D.J. Rigden, R.M. Keegan, A.N. Lupas, High-accuracy protein structure prediction in CASP14, Proteins. 89 (2021) 1687–1699. https://doi.org/10.1002/prot.26171.

[115] G.R. Buel, K.J. Walters, Can AlphaFold2 predict the impact of missense mutations on structure?, Nat Struct Mol Biol. 29 (2022) 1–2. https://doi.org/10.1038/s41594-021-00714-2.

[116] J. Skolnick, M. Gao, H. Zhou, S. Singh, AlphaFold 2: Why It Works and Its Implications for Understanding the Relationships of Protein Sequence, Structure, and Function, J. Chem. Inf. Model. 61 (2021) 4827–4831. https://doi.org/10.1021/acs.jcim.1c01114.

[117] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G.R. Lee, J. Wang, Q. Cong, L.N. Kinch, R.D. Schaeffer, C. Millán, H. Park, C. Adams, C.R. Glassman, A. DeGiovanni, J.H. Pereira, A.V. Rodrigues, A.A. van Dijk, A.C. Ebrecht, D.J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M.K. Rathinaswamy, U. Dalwadi, C.K. Yip, J.E. Burke, K.C. Garcia, N.V. Grishin, P.D. Adams, R.J. Read, D. Baker, Accurate prediction of protein structures and interactions using a three-track neural network, Science. 373 (2021) 871–876. https://doi.org/10.1126/science.abj8754.

[118] X. Zhou, W. Zheng, Y. Li, R. Pearce, C. Zhang, E.W. Bell, G. Zhang, Y. Zhang, I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction, Nat Protoc. (2022). https://doi.org/10.1038/s41596-022-00728-0.

Chapter III. Insights into the structure and function of the human organic anion transporter 1 in lipid bilayer membranes

Angelika Janaszkiewicz,[1] Ágota Tóth,[1] Quentin Faucher,[1] Marving Martin,[1,2] Benjamin Chantemargue,[2] Chantal Barin-Le Guellec,[1,3] Pierre Marquet,[1,4] Florent Di Meo,[1,*]

Published, doi: https://doi.org/10.1038/s41598-022-10755-2

Supplementary Information are available here

Abstract

The human SLC22A6/OAT1 plays an important role in the elimination of a broad range of endogenous substances and xenobiotics thus attracting attention from the pharmacological community. Furthermore, OAT1 is also involved in key physiological events such as the remote inter-organ communication. Despite its significance, the knowledge about *h*OAT1 structure and the transport mechanism at the atomic level remains fragmented owing to the lack of resolved structures. By means of protein-threading modelling refined by μs-scaled Molecular Dynamics simulations, the present study provides the first robust model of *h*OAT1 in outward-facing conformation. Taking advantage of the AlphaFold 2 predicted structure of *h*OAT1 in inward-facing conformation, we here provide the essential structural and functional features comparing both states. The intracellular motifs conserved among Major Facilitator Superfamily members create a so-called “charge-relay system” that works as molecular switches modulating the conformation. The principal element of the event points at interactions of charged residues that appear crucial for the transporter dynamics and function. Moreover, *h*OAT1 model was embedded in different lipid bilayer membranes highlighting the crucial structural dependence on lipid-protein interactions. MD simulations supported the pivotal role of phosphatidylethanolamine components to the protein conformation stability. The present model is made available to decipher the impact of any observed polymorphism and mutation on drug transport as well as to understand substrate binding modes.

Keywords: Membrane Transporters; Structural Pharmacology; Molecular Dynamics; Protein-lipid interactions; Major Facilitator Superfamily; Organic Anion Transporter 1

[1] *INSERM U1248 Pharmacology & Transplantation, Univ. Limoges, CBRS, 2 rue du prof. Descottes, F-87000 Limoges, France*

[2] *InSiliBio, 1 avenue d’Ester, Ester Technopôle, F-87000 Limoges, France*

[3] *CHU de Tours, 2 Boulevard Tonnellé, F-37044 Tours, France*

[4] *Department of Pharmacology and Toxicology, CHU Limoges, F-87000 Limoges, France*

[*] Corresponding Author: Florent Di Meo – florent.di-meo@inserm.fr; INSERM U1248 P&T, CBRS, Faculté de Médecine et Pharmacie, Univ. Limoges, 2 rue du Prof. Descottes, 87000 Limoges, France

III.1. Introduction

Major Facilitator Superfamily (MFS) proteins belong to the SoLute Carrier (SLC) superfamily, one of the most important classes of membrane transporters. They can translocate a broad range of endogenous compounds and xenobiotics across cell membranes and play important pharmacological and physiological roles [1,2]. MFS transporters can affect drug pharmacokinetics by modulating absorption, distribution, and elimination [3] since they are involved in cellular influx or efflux. Understanding MFS transporter functions and kinetics is of particular importance to decipher how do they modulate the local pharmacokinetics *i.e.*, local drug concentration at the target sites, whether linked with xenobiotic journey and/or therapeutic/adverse effects. This is particularly relevant since, over the past years, growing interest has been paid to local xenobiotic bioavailability (*i.e.,* at the intracellular scale) [4] which can help fulfil the gap between systemic and cellular-scaled pharmacological investigations [5].

From the physiological point of view, MFS transporters also play an essential role in maintaining homeostasis at the systemic and cellular scales. MFS transporters are involved in cellular nutrient disposition [1,2] (*e.g.*, sugar porters including Glucose transporters – GLUTs – family) as well as in detoxification processes [6,7] (*e.g.*, Organic Anion Transporter family). By modulating body fluid and tissue concentrations of a broad range of specific endo/exogenous molecules, MFS transporters might even drive hormone-independent remote inter-organ communications [8]. The so-called “remote sensing signaling theory” [2] is key to rationalize the remote modulation of transporter expressions or functions in distant organs as already suggested for SLC [8,9] and ATP-Binding Cassette (ABC) transporters, in physiology but also in pharmacology [10].

Several studies have provided evidence in favour of the central role of human Organic Anion Transporter 1 (*SLC22A6*/OAT1) in this dual physiology/pharmacology context [11]. Originally known as the New Kidney Transporter (NKT), *h*OAT1 is a multi-specific transporter mostly expressed in kidneys [12], at the basolateral membrane of proximal tubular cells (PTC) where it participates in the substrate uptake phase of blood-urine PTC exchanges [13]. *h*OAT1 transports mostly anionic compounds, including xenobiotics such as antiviral acyclic nucleoside phosphonates (*e.g.*, tenofovir, adefovir) [14], endogenous compounds and metabolites (*e.g.*, mono- and di-carboxylates) including uremic toxins, especially protein-bound uremic toxins (*e.g.*, indoxyl sulfate, *p*-cresol sulfate) [6,7,15,16]. Therefore, *h*OAT1 dysfunctions are not only associated with the impairment of xenobiotic elimination, but also with pathophysiological conditions owing to increased systemic retention of uremic toxin such as in chronic kidney disease [6,7,16]. Furthermore, a large diversity of substrates may compete between them for *h*OAT1 transport [2]. Likewise, several xenobiotics act as *h*OAT1 inhibitors and affect *h*OAT1-mediated detoxification processes [16]. These competition events can also impair drug therapeutic efficacy or lead to adverse effects [15,17]. *h*OAT1 impairment has long been assumed to have a limited impact owing to the redundant expression of *h*OAT3, with which it has a significant substrate overlap. However, this importance of *h*OAT1/*h*OAT3 duality should not be overestimated owing to the recently described substrate selectivity regarding metabolites [18]. This explains the recommendation from the International Transporter Consortium about the evaluation of *h*OAT1 activity in drug discovery [19,20], followed by the Food and Drug Administration [21], the European Medicine Agency and the Japan Pharmaceutical and Medical Devices Agency, at least in term of inhibition studies [22,23].

Despite the great importance of *h*OAT1 in terms of xenobiotic renal clearance, knowledge about the transport mechanism remains fragmented. *h*OAT1 is an antiporter, translocating substrates from blood to the intracellular compartment in exchange for at least one α-ketoglutarate (αKG) [13]. Substrate translocation is expected to be driven by αKG concentration which is governed by both the Na^+/dicarboxylate transporter (*SLC13A3*/NaDC3) and intracellular metabolism [12]. It is worth mentioning that *h*NaDC3 activity in PTCs is strongly related to Na^+/K^+-ATPase, leading to a "tertiary" active transport involving the Na^+/K^+-ATPase – NaDC3 – OAT1 triad [12]. At the nanoscale, substrate translocation is expected to follow alternating access involving at least three conformational states, namely the outward-facing (OF), occluded and open inward-facing (IF) states [3]. Regarding the unknown folding of *h*OAT1, only two structural models of *h*OAT1 in lipid bilayer membranes have been reported so far. They were obtained by refining homology models with short 100ns+ molecular dynamics (MD) simulations [24,25]. Both models adopted the IF state, using bacterial *E. coli* Glycerol-3-phosphate Transporter (GlpT) resolved structure as the initial template [26]. A high-confidence IF model was recently released using the machine-learning structural prediction tool AlphaFold 2 (AF2) [27].

Even though key residues can be identified from computational as well as experimental studies the dynamic and atomic features of *h*OAT1 structure still remain unclear [18,24,28,29] (see Supplementary Table S1 for details). The absence of a robust *h*OAT1 OF model precludes the thorough atomistic rationalization of substrate binding events as well as the investigation of lipid-protein interactions, which have been shown to be of major importance for several MFS transporters and other membrane proteins by either experimental or computational techniques [30–33]. Furthermore, within the frameworks of pharmacogenetics (PGx), atomic-scaled and dynamic pictures of MFS transporters enable the investigation of the structural (and possible functional) modifications arising from Single Nucleotide Polymorphisms (SNPs) or rare pharmacogenetic mutations. In the present study, we propose a protein threading-based model of the missing OF state of *h*OAT1. Microsecond-long MD simulations of *h*OAT1 inserted in several lipid bilayer membranes were performed in order to: (i) refine the initial protein threading static model; (ii) explore the local conformational space of *h*OAT1; and (iii) assess lipid-protein interactions. Topology and structure of the proposed models were carefully analyzed and systematically confronted to AF2 model, as well as to experimental observations. We propose here mechanistic and structural insights into *h*OAT1 transport including the role of lipid-protein interactions.

III.2. Methods

III.2.1. Putative structure of *h*OAT1 in outward-facing state

The amino acid sequence of *h*OAT1 was obtained from UniProt database [34] with the accession number Q4U2R8, using isoform 1 as the canonical sequence. The initial three-dimensional model of wild-type *h*OAT1 was achieved using the automated protein structure prediction tool I-TASSER webserver [35]. Three relevant resolved MFS proteins were identified as templates, namely *h*GLUT3 (PDB ID: 5C65, 2.65Å resolution) [36], *r*GLUT5 (PDB ID: 4YBQ, 3.27Å resolution) [37] and XylE (PDB ID: 4GBY, 2.81Å resolution) [38], for which sequence identities and similarities are reported in Supplementary Table S2 as well as sequence alignments in Supplementary Fig.S1. It is worth mentioning that the initial I-TASSER *h*OAT1 model exhibited a salt bridge between Asp112 and Thr540. This would lead to an implausible direct polar interaction between the extracellular loop (ECL) 1 and the intracellular C-terminal

domain in the lipid bilayer. This artifact was thus corrected by means of steered MD simulations in which both Asp112 and Thr540 were pulled apart from each other. A steered MD simulation in pure 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer membrane (see section 2.2. regarding the embedding procedure used) was first performed to smoothly increase the distance between Asp112 and lipid bilayer membrane centers-of-mass (COM) while maintaining Thr540 by positional restraints. Then, the distance between Thr540 and the lipid bilayer membrane COMs was increased. Both simulations were carried out for 2 ns, applying a restraint force constant potential of 35 kcal/mol/Å^2 in the z-direction and a pulling velocity of 10 Å/ns. In order to improve our initial model, μs-scaled MD simulations were performed including the surrounding environment (*i.e.*, lipid bilayer membrane, water and ions) following an approach similar to that previously used for the human multidrug resistance-associated protein 4 (*ABCC4*/MRP4) [39]. Besides, the released AF2 *h*OAT1 IF model was also considered for MD simulations.

III.2.2. Model preparation for MD simulations

Protonation states of charged residues were assigned, using the PROPKA software [40], at pH = 7.4. Special attention was paid to histidines to which protonation states were assigned in accordance with their calculated pKa as well as potential H-bond networks with surrounding residues by visual inspection. The ε-protonated state was used for His47, His130, His217, His246, His249 and His546, the δ-protonated state for His48, His275 and His337, while the cationic double ε/δ-protonated state was assigned to His34. The C-terminal domain (549-563) was cut out of the model to avoid unexpected interactions owing to its high flexibility. The resulting *h*OAT1 model was then embedded in lipid bilayer membranes using the CHARMM-GUI membrane builder tool [41]. Four different POPC-based lipid bilayer membranes were considered, representing different molecular ratios of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) and Cholesterol (Chol): POPC, POPC:Chol (3:1), POPC:POPE (3:1), and POPC:POPE:Chol (2:1:1). The POPC:POPE:Chol (2:1:1) membrane was chosen to mimic the plasma membrane while the others were used to investigate the specific role of PC, PE and Chol lipids. Only POPC:POPE:Chol (2:1:1) membrane was considered for *h*OAT IF conformation as a comparative model. All systems were solvated in water and neutralized with 154 mM NaCl ions to match physiological conditions.

III.2.3. MD simulation setup

Amber FF14SB [42], Lipids17 [43] and TIP3P [44] forcefields were used to model protein, lipids and water, respectively. TIP3P-compatible parameters of Na^+ and Cl^- counterions were obtained from Joung and Cheatham [45,46].

MD simulations were carried out with the Amber18 package [47] using both CPU and GPU codes for minimization and equilibration, while MD production was performed exclusively on GPU code [48]. Periodic boundary conditions were applied. Non-bonded interactions were explicitly described within a cut-off distance of 10 Å using electrostatic and Lennard-Jones potentials. Long-distance electrostatic interactions were treated using the Particle Mesh Ewald (PME) method [49]. SHAKE algorithm was used to fix bonds involving hydrogen atoms allowing to set the integration time to 2 fs. Production temperature was set at 310K and maintained using a Langevin thermostat [50]. Constant pressure boundary conditions were initially maintained under semi-isotropic conditions using Berendsen barostat [51].

All systems were initially equilibrated by first minimizing all atomic positions. Then, water molecules were smoothly thermalized from 0 to 100K during 200 ps under (*N,V,T*) conditions. Additional system thermalization up to 310K was then carried out under semi-isotropic (*N,P,T*) conditions for which pressure control was ensured using Berendsen barostat. System boxes were then equilibrated during 5.5 ns. System details (number of atoms and box sizes) are reported in supporting information (Supplementary Table S3). Three independent replicas per lipid bilayer membrane were performed with up to 2 μs (for OF model) and 1 μs (for IF model) MD simulation each, leading to a total of *ca.* 27 μs. Trajectory snapshots were saved every 10 ps.

III.2.4. Analysis

Structural analyses. Given the high-confidence model provided by AF2 [27], the reliability of the present OF model folding was evaluated on the secondary structure and the global MFS folding obtained [3]**.** Structural analyses were performed using the PyTRAJ and CPPTRAJ AMBER modules [52], VMD [53] and in-house python scripts. Analyses were performed on equilibrated 1.5 μs long trajectories according to the evolution of time-dependent backbone root-mean squared deviations (Supplementary Fig. S2 and S3). z-Dependent pore radius was calculated using the Hole program [54]**.** 500 snapshots were considered for each trajectory. Interhelical and interdomain interactions were monitored focusing on contacts (< 4 Å) and H-bond interactions. The latter was considered using distance and angle cutoffs set at 3.0 Å and 135°. The minimum fraction threshold was set at 0.1 given the known uncertainties for side chain rotameric states in protein threading techniques. The dynamic cross correlation matrices were calculated over the OF and IF *h*OAT1 MD trajectories in POPC:POPE:Chol (2:1:1) separately, considering only the MFS core.

Principal Component Analysis (PCA). In order to confirm the OF state of the hereby proposed *h*OAT1 model, trajectories were projected on the MFS conformational space obtained from experimentally resolved MFS structures. This conformational space was obtained by performing PCA over a structural data set consisting of MFS proteins available in the Protein Data Bank [3,55]. The MFS dataset including all alternating access states, *i.e.*, OF, OF^{occ}, IF, IF^{occ} states, are listed in Supplementary Table S4. PCA was achieved by only considering Cα of the MFS twelve transmembrane helices (TMH, see Supplementary Table S5 for transporter MFS core definitions). Dimensionality reduction by PCA points to the main sources of structural variability in the MFS dataset, which thus allows distinguishing IF and OF states as recently proposed [3,55]. Besides, to monitor the OF subspace sampled during MD simulations, a second PCA was also carried out on the MFS backbone using *h*OAT1 OF trajectories. Every membrane was considered.

Clustering. Clustering was performed to identify the different subspaces sampled during MD trajectories. Clustering was achieved using the InfleCS approach which take advantage of Gaussian Mixture Models (GMM) to provide insights into the free energy surface [56]. Clustering was carried out by focusing on the first principal component obtained from PCA performed on all trajectories as well as the extracellular distances between TMH1 and TMH7 of *h*OAT1 OF model (residues 212 – 219 and 447 – 454, respectively). Clustering was achieved by using a grid size of 80x80, 5 iterations and from 2 to 16 gaussian functions for GMM. For further details about this method, see Ref. [56].

III.3. Results and Discussion

III.3.1. Structural patterns of *h*OAT1

III.3.1.1. Topological overview of the *h*OAT1 model

In agreement with previous studies [24,25] as well as AF2 prediction [27], the present MD-refined model of *h*OAT1 adopted the MFS fold. Despite the low sequence similarity within the MFS superfamily, they share a common architecture. MFS fold mostly consists of 12 transmembrane helices (TMHs) divided into two bundles of 6 TMHs each. The so-called N- and C-bundles comprise TMH1-6 and TMH7-12, respectively (see Fig.1a&b) [57]. As expected for an MFS-fold transporter [3,55,58,59], N- and C-domains exhibited pseudo-symmetry perpendicularly to the plane of the membrane (Fig.1a). The present model revealed at least 6 intracellular helices (ICHs, Fig.1a&b), as observed in the AF2 model [27] and other experimentally resolved mammalian MFS transporters (*e.g.,* GLUT1/*SLC2A1*, GLUT3/*SLC2A3*, GLUT5/*SLCA5*) [37,60,61]. These ICHs are in close contact with TMHs for which local details are discussed in section 3.2. Finally, *h*OAT1 topology suggested a long extracellular loop (ECL1) made of *ca.* 90 amino acids (40-126) between TMH1 and 2. The secondary structure of ECL1 appeared more disordered in the present model than in AF2, leading to lower confidence for ECL1 than for MFS core. However, this is not expected to strongly affect the MFS core structure discussed in the present study. Furthermore, the glycosylation of known sites in ECL1 is not required for the transporter function [14,28].

III.3.1.2. *h*OAT1 model adopts OF state conformation according to MFS conformational space

MFS alternating access is expected to follow the rocker-switch mechanism [3] in which N- and C-bundles rearrange between OF and IF states. This large-scale conformational change along the transport cycle was also shown to affect intra-bundle TMH arrangements [3]. Overall, the MFS tertiary structure is modified along the transport cycle by rocking N- and C- bundles to alternatingly expose substrates to both sides of the lipid bilayer membrane (see Ref. [3] for further details). Visual comparison between the MD-refined *h*OAT1 model and AF2 prediction suggested two distinct states. The present model adopted a “V”-shape conformation, while AF2 clearly exhibited a “Λ”-shape (Fig.52b&c) as proposed for the IF conformation of resolved MFS transporters [24,25]. This was confirmed by projecting MD trajectories and the AF2 structure onto the MFS conformational space obtained by PCA (see Fig.52c and Supplementary Fig.S4). Besides, the OF state was also confirmed by exhibiting significantly larger (respectively smaller) Met358-Ser139 (Gly446 - Val211) distances with respect to the previous IF model obtained by Tsigelny *et al*. [25]. These distances were suggested to picture opening at either the extra- or intra-cellular sides, respectively (see Supplementary Fig.S5).

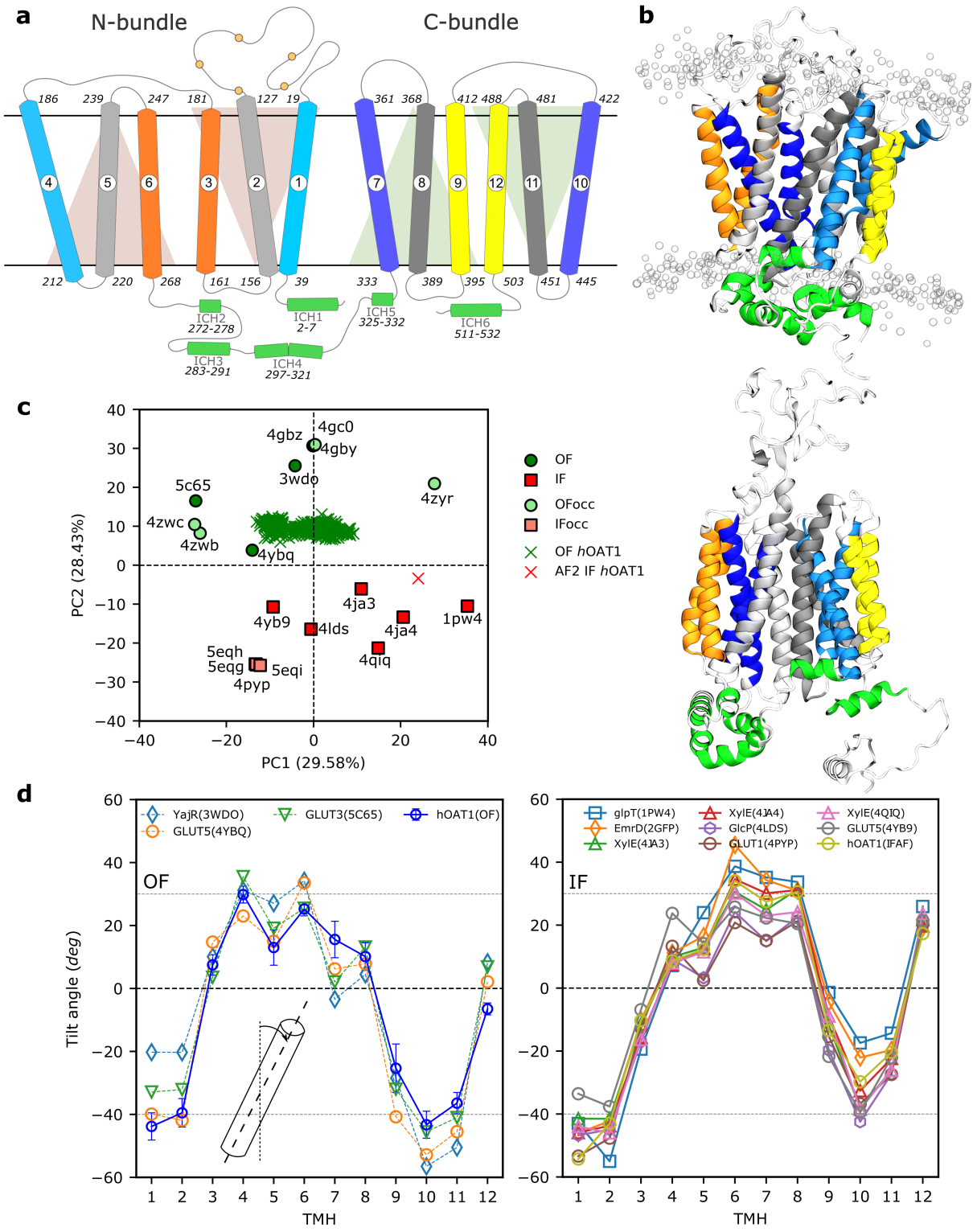


Figure 52. Overview of the *h*OAT1 transporter. (a) The topology scheme shows *h*OAT1 adopting the canonical MFS fold that consists of 12 transmembrane helices (TMH) divided into N- and C- bundles, connected by an intracellular loop rich in intracellular helices (ICHs). Each bundle is constructed of 3-TMH inverted segments. TMH1 and TMH2 are connected by a long extracellular loop possessing 5 glycosylation sites (Arg39, Arg56, Arg92, Arg97, Arg113). The so-called A-, B- and C-helices are depicted in blueish, grayish, and yellowish, respectively. (b) 3D model of *h*OAT1 obtained from MD simulation and AlphaFold2 prediction in OF (top) and IF (bottom) conformational states, respectively. (c) *h*OAT1 projected onto the conformational space obtained via PCA of resolved MFS transporters in OF, OF^{occ}, IF, IF^{occ} conformations. (d) Tilt angle profile of MFS transporters in OF (left) and IF (right)

conformations. The TMH tilt angle profile for the *h*OAT1 OF model was averaged over MD simulations considering all replicas.

Building upon the concept of typical structural features for MFS proteins, tilt angles between TMHs and the lipid bilayer axis normal were monitored along MD simulations (see Supplementary Fig.S6) and compared with experimentally resolved MFS transporters. Tilt angle profiles were averaged over MD trajectories, exhibiting good agreement with the profiles obtained with experimentally resolved OF state MFS transporters (see Fig.52d). The present OF *h*OAT1 model exhibited the well-known 3-TMH repeated segment fold observed in MFS proteins. Within each bundle, two 3-TMH segments are related by approximately 180° rotation around the lipid bilayer normal [3,57,58]. This leads to three sets of TMHs, namely A-, B- and C-helices (see Fig.52 and Supplementary Fig.S7) for which different functional roles were suggested [59]**.** The dynamic interplay of interactions between helices is an imperative part of alternation between OF and IF states, including the existence of intermediate occluded states [59]. The averaged contact maps of OF and IF states were also compared along MD simulations, focusing on the MFS core (see Supplementary Fig.S8). Overall, intra-bundle contacts were conserved between the two states. This is in agreement with the suggested rocker-switch mechanism [3] for MFS transport cycle. Furthermore, in line with the large-scale conformational changes occurring during the transport cycle, the strong interdependency between TMHs was suggested from MFS core dynamic cross correlation matrix; showing significant motion correlations between TMHs along simulations (see Supplementary Fig.S9).

III.3.1.3. Structural arrangement of TMHs in *h*OAT1 OF model

It is important to note that A- and B-helices were suggested to act by pairs across N- and C-bundles. Therefore, in the present section, particular attention was paid to inter-bundle interactions for A- and B-helices in contrast to C-helices.

The central cavity of *h*OAT1 consists of A-helices, namely TMH1 and 4 for the N-bundle and TMH7 and 10 for the C-bundle (see Fig.52 and Supplementary Fig.S7). These helices play a role in substrate binding and release events in the OF and IF states, respectively [59,61,62]. They interact by pairs across bundles *i.e.,* TMH1 with TMH7 and TMH4 with TMH10. Key non-covalent interactions between TMH1 and 7 occur on the extracellular site. z-Dependent cavity pore radii exhibited the expected OF pattern, *i.e.*, greater opening in the extracellular side than in the intracellular one (see Supplementary Fig.S10). However, large standard deviations suggest the existence of OF occluded structures during MD simulations. This was confirmed by monitoring (i) the extracellular distance between TMH1 and TMH7 (see Supplementary Fig.S11) and (ii) performing PCA considering all lipid bilayers (see Supplementary Fig. S12 and S13). The first two principal components (PC) were assigned to the opening and closing of the extracellular gate, for which TMH contributions are reported in Supplementary Fig.S14. PC1 and the extracellular distance between TMH1 and TMH7 were used to picture the OF subspace sampled during MD simulations by means of the InfleCS method [56]**.** It is important to note that the free energy barriers obtained from InfleCS should be carefully considered given the low sampling of intermediate regions. However, three main state populations were clearly identified, namely extracellular open, intermediate, and closed conformations (see Fig.53). This confirms the central role of the TMH1/TMH7 pair which is expected to be involved in extracellular gating event prior to the occlusion of the extracellular gate along transport cycle. MD simulations revealed the following interacting residues in *h*OAT1: Asn35, Thr36, Asn39, Phe40 for TMH1 and Tyr353, Tyr354, Leu356, Val357 for TMH7. MD simulations also revealed kinking of A-helices, in agreement with previous studies [57,59]. Investigations of TMH

helicities in OF *h*OAT1 model exhibited discontinuity in TMH1 and 10 (see Supplementary Table S6) leading to elbow-shape TMHs. A-helix discontinuities were used to picture the structural adaptability of MFS core along the OF-to-IF transition and *vice versa* [38,57,59]. It provides flexibility allowing side chains of gating residues to interact within paired A-helices. Structural analyses performed on OF *h*OAT1 model enabled the identification of dispersive, electrostatic, and H-bond interactions between the aforementioned residues involved in the so-called "gating" events. MD simulations and the AF2 model support the key role of Tyr354/Tyr353 placed on the "elbow" point of TMH7 for gating as it was shown for conserved tyrosine in sugar porters (*e.g.*, conserved Tyr292 and Tyr293 in *h*GLUT1, Tyr290 and Tyr291 in *h*GLUT3) [3]. MD simulations stressed out that interactions between A-helices are highly dynamic as pictured by H-bond network analysis (see Supplementary Tables S7 and S8), especially for extracellular occlusion event in OF *h*OAT1. Our simulations suggest that the interchange between the OF open and occluded states can dynamically occur even in the absence of substrate, owing to local flexibility and thermal fluctuation [55] as pictured by the aforementioned clusters (see Fig.53 and Supplementary Fig.S7).

B-helices (TMHs 2, 5, 8 and 11, see Fig.52a and Supplementary Fig.S7) are expected to play a role in maintaining the interface between the N- and C- bundles [59]. As shown for A-helices, B-helices might be considered as pairs, *i.e.*, TMH2/TMH11 and TMH5/TMH8. Therefore, particular attention was paid to non-covalent interactions between bundles. The present *h*OAT1 model is in agreement with these findings as pictured for instance by strong H-bond interactions between TMH5 and TMH8, which are maintained for more than 80% of the time during MD simulations (see Supplementary Tables S7 and S8). The most frequent residues involved in TMH2/TMH11 and TMH5/TMH8 H-bond networks are reported in Supplementary Table S8. Besides, in line with previous experimental observations, OF *h*OAT1 B-helices are likely to participate in substrate binding and translocation along transport cycle thanks to: (i) their "banana-shape" bending (see *e.g.,* TMH2 in Supplementary Fig. S7 and S13) [38]; and (ii) their residues exposed at the substrate cavity (*e.g.*, Arg466, Ser469, Arg131, Arg134). In OF state conformation, TMH5/TMH8 interactions are preserved all along the helices. Bending of B-helices displays a different profile for AF2 IF with respect to OF *h*OAT1 conformation. The helices differ in the curvature at the helical ends, and the most pronounced variation between states was found for TMH11 and TMH8 (Supplementary Fig.S7). This suggests that large-scale conformational changes along the *h*OAT1 transport cycle are asymmetric. The C-bundle is likely to be more flexible, in line with previous findings regarding other MFS proteins such as LeuT, *h*GLUT3, and *h*GLUT5[3,38,61,63]. This hypothesis was strengthened by both PCA, and TMH tilt angle profiles obtained with the OF *h*OAT1 model, which showed larger flexibility for the C- than for the N-bundle (see Supplementary Fig. S4 and S13). This is also the case for AF2 (see Supplementary Fig.S3).

Finally, the C-helices (TMH3, TMH6, TMH9 and TMH12, see Fig.52a and Supplementary Fig.S7) are located out of the central *h*OAT1 core. In contrast to A- and B-helices, C-helices stand by each other in each bundle, *i.e.,* TMH3/TMH6 and TMH9/TMH12 for the N- and C-bundles, respectively. They support the structure integrity of *h*OAT1 by interacting with the lipid bilayer. In the present OF model, inter-helical interactions between TMH3 and TMH6 are mostly located at the intracellular side. This is not the case for TMH9 and TMH12 which exhibit contacts over the whole helices. Interestingly, the opposite trend seems to occur with the AF2 IF *h*OAT1 model: TMH9/TMH12 exhibit less contact than TMH3/TMH6, likely due to a conformational change along the transport cycle.

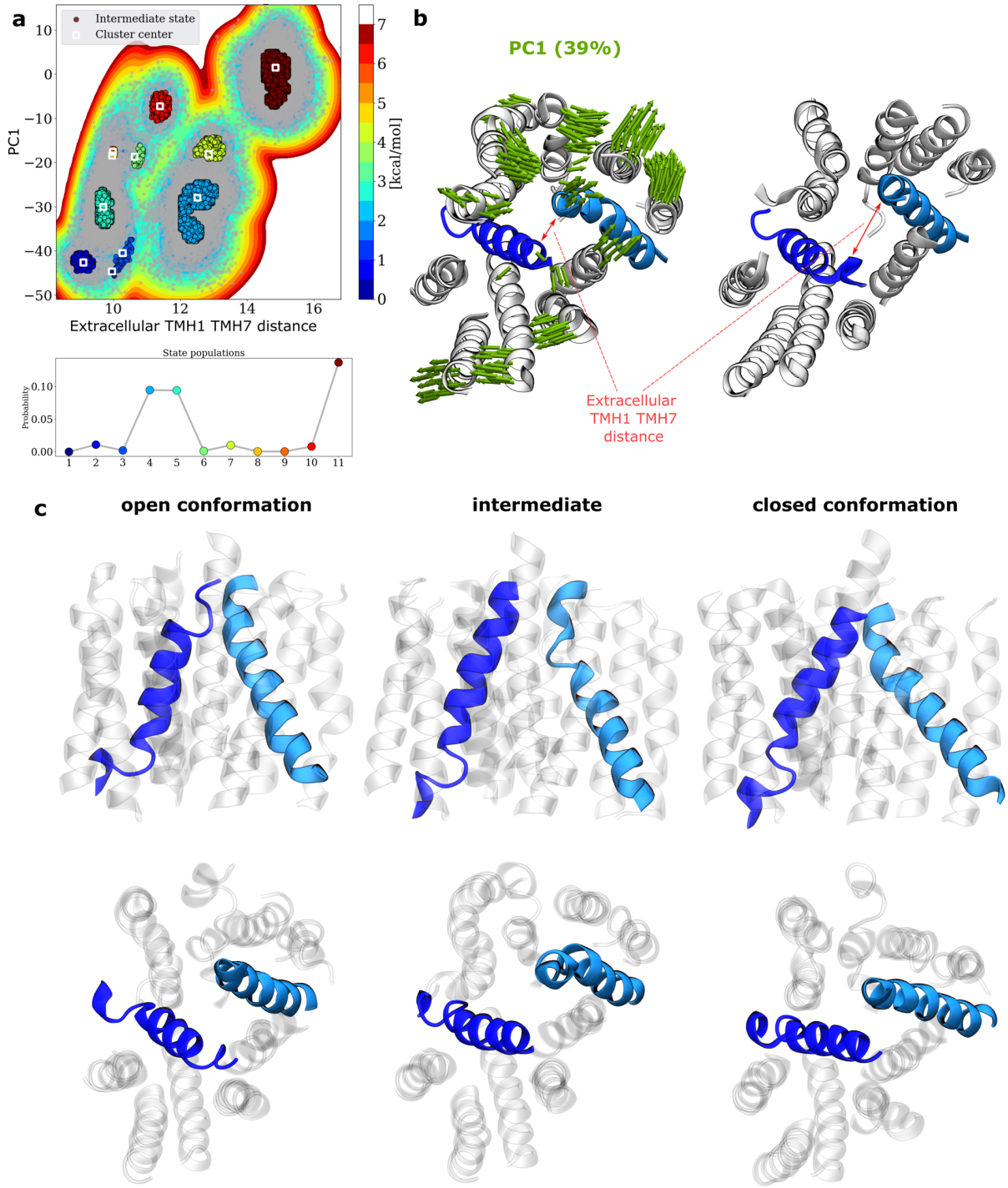


Figure 53. Conformational sampling of extracellular gating events of *h*OAT1 OF model. (a) Insights into the free energy surface (top) sampled during MD simulations according to PC1 and extracellular distance between TMH1 and TMH7 as well as cluster probabilities (bottom). Porcupine plot (left) obtained from PCA performed considering *h*OAT1 model embedded in all lipid bilayer membranes and corresponding evolution of extracellular distance between TMH1 and TMH7. (c) Representative snapshots of the three main clusters in which TMH1 and TMH7 are highlighted to feature occlusion states (side and top are respectively shown on top and bottom panels.

III.3.2. The importance of MFS conserved sequences on the "charge-relay" system of *h*OAT1

III.3.2.1. MFS conserved motifs as central components of the charge-relay system

*h*OAT1 shares with other MFS transporters conserved sequences across species, which were shown to act as "molecular switches" during the transport cycle by *e.g.*, triggering large-scale conformational changes [31,64,65]. The so-called MFS signature motifs are located at the intracellular interface, *i.e.*, in intracellular loops (ICLs) and TMHs as observed in other MFS transporters [57,59,64,66,67]. These motifs are duplicated in the N- and C-bundles. MFS signature motifs identified in *h*OAT1 may slightly differ in terms of sequence between the two bundles (see Table 3 and Fig.54), as well as with other MFS proteins [24,68]. The so-called A-motifs [3,64,65] are located in the ICLs between TMH2 and TMH3 in the N-bundle, and between TMH8 and TMH9 in the C-bundle. The N-bundle A-motif matches with the canonical sequence, *i.e.,* G[X3]D[R/K]XGR[R/K]. The C-bundle A-motif sequence is shorter, whilst the final LGRR pattern is conserved (Table 3 and Fig.54). The E[X_6]R sequence (also known as ELYPT[66,68] for the N-bundle) is observed in the ICLs connecting TMH4 and TMH5 in the N-bundle and TMH10 and TMH11 in the C-bundle. The PETL motif is located in the C-terminal domain, after TMH12. Finally, the conserved [P/X]ESXRW[L/X] sequence[66,68] is also observed in *h*OAT1 after TMH6, in the intracellular domain connecting the N- and C-bundles. In spite of significantly different primary sequences, OF and IF *h*OAT1 structural models support that PETL and [P/X]ESXRW[L/X] motifs are expected to behave similarly in the C- and N-bundles, respectively [3]. Thereby, it must be stressed that, for sake of readability, both motifs will be referred to as PETL.

MFS signature motifs are rich in charged and polar amino acids (mostly arginine, aspartate, and glutamate) leading to strong H-bond and salt-bridge networks. MD simulations showed that the so-called "charge-relay system" [64,65] is highly dynamic since salt-bridges and H-bonds can be exchanged along the simulation. The IF AF2 and MD-refined OF models were then used to identify shared patterns and conformation-dependent rearrangements.

The "charge-relay system" can be divided into two building blocks in the N- and C-bundles, each made of A-motif, E[X6]R, PETL motifs (Fig.54a). These motifs share a similar structural arrangement regardless of the conformational state. H-bond analyses highlighted the central role of the last two arginine residues of A-motifs in maintaining the supramolecular arrangement with the other two motifs (see Fig.54b&c). N-bundle Arg161 and Arg162 interact with Glu212 and Glu270 from the E[X_6]$R_{N\text{-bundle}}$ and $PETL_{N\text{-bundle}}$ motifs, respectively. Likewise, in C-bundle, Arg394 and Arg395 interact with Glu447 and Glu506, respectively in the E[X_6]$R_{C\text{-bundle}}$ and $PETL_{C\text{-bundle}}$ motifs. It is worth mentioning that our findings are supported by a directed site-mutagenesis experiment where the mutation of Glu506 led to complete inactivation of *h*OAT1 transport [29]. Glu212 and Glu447 in the N- and C-bundle E[X_6]R motifs also interact with PETL motifs. Using MD simulations on the OF *h*OAT1 model, H-bond fractions over time were also calculated to measure the strength of the local H-bond network in each triad (Fig.54c).

Table 3. The description of MFS signature intracellular motifs found in *h*OAT1 divided into N- and C- bundles showing the pseudosymmetry of the transporter.

	Motif	Topological Location	Sequence
N-bundle	**A-motif** G[X3]D[R/K]XGR[R/K]	ICL2-3	G153, Y154, L155, A156, D157, R158, L159, G160, R161, R162
	ELYPT E[X6]R	ICL4-5	E212, TRP213, M214, P215, I216, H217, T218, R219
	PETL [P/X]ESXRW[L/X]	ICL6-ICH2	I269, E270, S271, A272, R273, W274, H275
C-bundle	**A-motif** [D/N][R/H]LGRR	ICL8-9	N390, S391, L392, G393, R394, R395
	ELYPT E[X6]R	ICL10-11	E447, L448, Y449, P450, T451, M452, I453, R454
	PETL PET[K/L]	ICL12-C-terminal	P505, E506, T507, L508

Salt-bridges between A- and E[X_6]R motifs exhibit highly conserved interactions for Arg162-Glu212 and Arg394-Glu447; time fractions respectively being above 1.0 along MD simulations (see Fig.54c and Supplementary Table S9). Interestingly, contact analysis of static AF2 IF conformation suggested similar H-bond pattern (see Fig.54b). This was confirmed by monitoring H-bond during MD simulations performed on the IF *h*OAT1 model (see Supplementary Table S9). Similar H-bond network as OF conformation was observed, supporting the existence of the conformation-independent motif arrangement within each bundle. Interestingly, interactions between motifs across bundles differ significantly in IF and OF models. The AF2 IF *h*OAT1 model does not exhibit non-covalent interactions, nor even contacts, between motifs of N- and C-bundles (see Fig.54a&b). In contrast, *h*OAT1 OF model exhibits strong H-bond and salt-bridge networks (see Supplementary Table S9). The supramolecular arrangement of OF *h*OAT1 relies on the interactions between the two E[X_6]R motifs, as pictured by the strong salt bridge between Arg219 and Glu447 (H-bond fraction of 1.334). In agreement with previous observations on MFS proteins [3,37,55,59,65], MD simulations show that cross-bundle interactions also involved the intracellular side of TMHs with MFS signature motifs, but to a lesser extent. For example, H-bonds were observed between N-bundle A-motif and TMH11 (Asp157-Gln455, fraction = 0.224) or between C-bundle A-motif and TMH5 (Thr224 or Ala220 with Asn390, with fractions of 0.776 and 0.134, respectively). H-bond interactions were also monitored on IF *h*OAT1 trajectories; revealing the absence of inter-bundle H-bond (see Supplementary Table S9) leading to a greater distance

between E[X_6]R motifs (*ca.* 12 and 17Å, respectively for OF and IF conformations, see Supplementary Fig.S15).

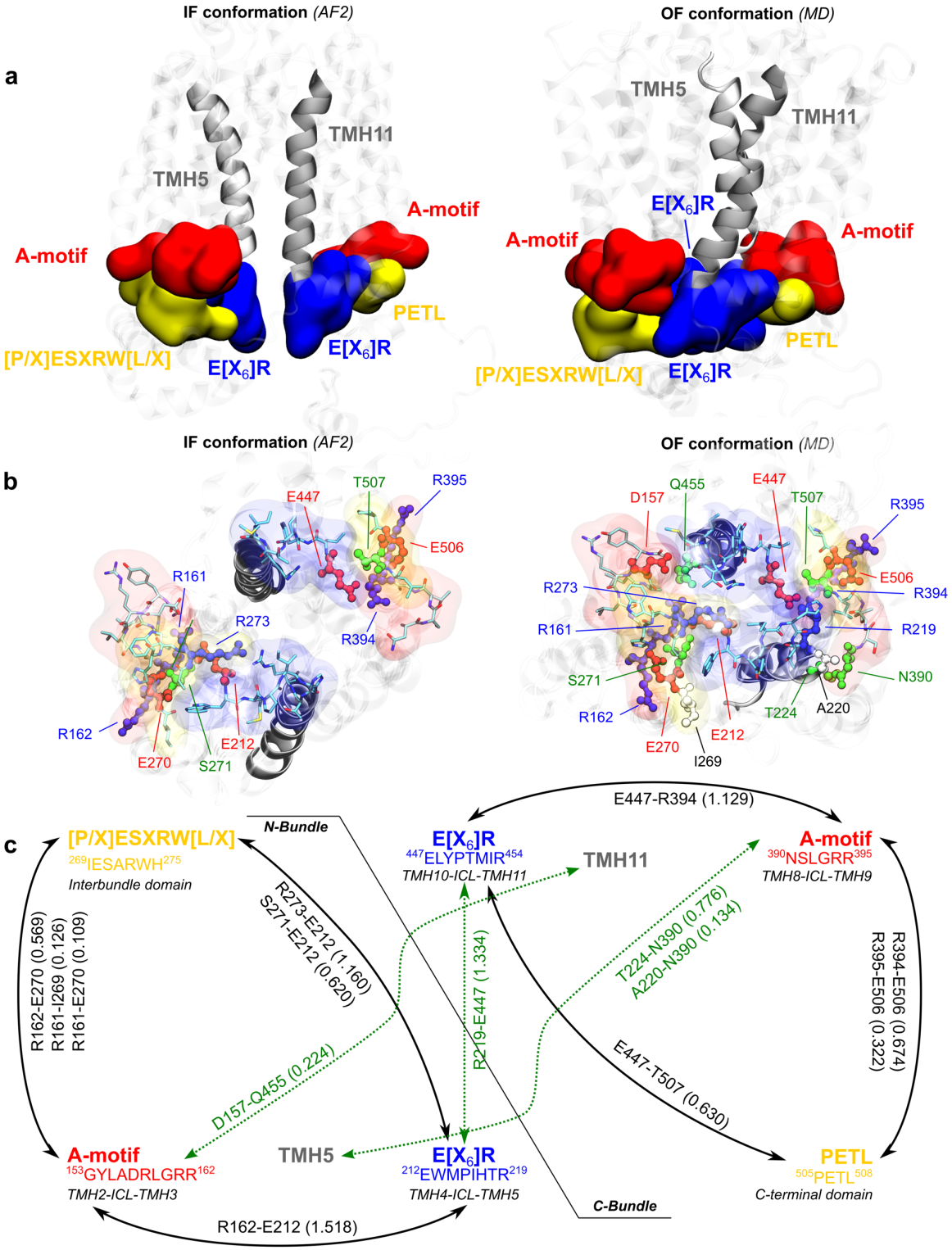


Figure 54. Intracellular motifs conserved among MFS. (a) Charge-relay system of *h*OAT1 as a triad made of A-motif, [P/X]ESXRW[L/X] / PETL and E[X_6]R symmetrically in the N- and C-bundles visualized in IF (left) and OF (right) conformations. (b) Intracellular view of the charge-relay system with highlighted residues involved in interactions. (c) The map of each motif interactions emphasizing the symmetry in

bundles. The communication within motifs is demonstrated by the strength of hydrogen bonds. Green dotted lines represent the missing interactions in the IF model, crucial for conformational changes. It must be stressed that values above 1.0 highlight salt-bridges in which more than one H-bond is possible (*e.g.*, between arginine and glutamate/aspartate residues).

This suggests that A-motifs might be involved in locking intracellular gate, which in turn maintain the OF conformation. This is in agreement with previous observations on resolved MFS proteins adopting OF confirmations, such as YajR and GLUT1 [3,37,55,59,65]. For example, high-throughput single directed mutagenesis performed on glycine and aspartic acid in the first and the fifth position of the YajR MFS transporter A-motif showed conformational transition from OF-to-IF, while other single point mutations only destabilized the protein [31,55,62,65]. Likewise, the E[X_6]R motif may play an important role in local arrangement of TMHs across bundles in OF conformation. Interestingly, structural analyses revealed only interactions of the N-bundle E[X_6]R motif with TMH11 through the H-bond interaction between Glu212 and Gln455 (fraction=0.464). Comparatively it was shown for the YajR transporter, where the interactions between TMH2 and TMH11 would be essential for the OF conformation [65]. However, no interaction was observed between C-bundle E[X_6]R and TMH5 in spite of the pseudosymmetry of the MFS transporter. This may be due to the resolution of the present OF model. Besides, it may also suggest an asymmetrical behavior in *h*OAT1 between the N- and C-bundles, which requires further investigations.

Despite the lower confidence of our model regarding the resolution of intracellular loops and helices in the OF state, the MD simulations as well as the comparison with AF2 structure provided hints regarding the cytoplasmic arrangement. As observed for resolved GLUTs [38,55,61,62], both models suggest that intracellular helices (ICHs) are in contact with the MFS signature motifs. In the AF2 IF model, ICHs are separated between the N- and C-bundles, while the OF model suggest contacts between ICHs as well as with intracellular loops. These interactions are expected to play a key role along the transport cycle. In case of sugar porters (*e.g.,* GLUT1 or GLUT3), ICHs and TMHs were proposed to lock the transporter in the OF conformation, precluding the exposure of the intracellular gate to the environment [3,55,61].

Altogether, present MD simulations findings line up with the putative role of tightly arranged intracellular interactions engaging the ICHs that are likely involved in substrate access to the intracellular gate. It is consistent with previous hypotheses that the intracellular interactions of *h*OAT1 are also prompt to stabilize the OF conformation. Therefore, the eventual breakage of these interactions may be directly involved in the conformational change along the transport cycle [12,59,61,62,64,65].

III.3.3. The impact of membrane lipid components

III.3.3.1. On the interplay between lipid composition and the *h*OAT1 conformational space

MFS transporter structures and functions were both computationally and experimentally shown to strongly depend on membrane composition [3,30–32,59]. This is particularly true for membranes made of PC and mixtures of PE phospholipids, which showed different behaviors in term of non-covalent interactions with membrane proteins [30–32,69,70]. In the present study, MD simulations were used to provide insights in protein-membrane interactions. The OF *h*OAT1 model was embedded in various lipid bilayers, *i.e.*, POPC:POPE:Chol (2:1:1), POPC:Chol (3:1), POPC:POPE (3:1), and POPC, the first one presenting the closest amounts of PE lipids and cholesterol to actual cell membranes[71]. Although the membranes used in

the present study do not comprehensively account for the whole complexity of cell membranes in terms of composition and asymmetry, they are expected to faithfully catch the main features of membrane-protein interactions for the most abundant lipids, *i.e.*, PC, PE and cholesterol.

In order to assess the overall lipid composition-structure relationship, trajectories obtained from MD simulations in different lipid bilayer membranes were all projected onto the MFS conformational space obtained using PCA on the resolved structure. Regardless of the membrane composition, all systems conserved the expected OF conformation along simulations as shown by PCA projection as well as TMH tilt angle profiles (see Supplementary Fig. S16 and S17). However, PCA projections revealed that protein dynamics and conformational space are slightly different in pure POPC and binary lipid bilayer (*i.e.,* POPC:POPE and POPC:Chol), as compared to the POPC:POPE:Chol (2:1:1) lipid bilayer membrane. Trajectories were also projected on the OF subspace obtained from the PCA calculated using all lipid bilayer membrane (see Supplementary Fig.S12). PCA projection supports a differential behavior according to lipid bilayer composition (see Supplementary Fig.S18). In absence of POPE and/or cholesterol, replicas sampled different subspaces with no overlap between them. This suggest that PE lipids and cholesterol play a central role in *h*OAT1 dynamics. MD simulations also suggested that lipid composition is also expected to slightly modulate gating events as pictured by the extracellular distances between TMH1-TMH7 (see Supplementary Fig.S11). Although no clear conclusion in terms of function can be drawn from these results, differential protein dynamics according to lipid bilayer composition are in agreement with previous observations [30,31] stressing the importance of protein-lipid non-covalent interactions.

Both intracellular and extracellular openings were monitored by respectively measuring Met358-Ser139 and Val211-Gly446 distances in *apo h*OAT1 simulations (Supplementary Fig.S5). Given that *h*OAT1 adopts the OF conformation, intracellular distances were expected to exhibit low variability owing to the structural intracellular arrangement maintained by the charge-relay system (see Supplementary Fig.S11). Interestingly, intracellular distances exhibit slightly higher variability in a PE-free than in a PE-based lipid bilayer membrane. This may picture a looser packing of intracellular loops which may in turn modulate *h*OAT1 function.

III.3.3.2. Non-covalent interactions between lipid components and the *h*OAT1 transporter

The ability of lipid bilayer membranes to form H-bond networks is expected to contribute to protein stability and dynamics. To probe protein-lipid interactions, H-bond networks were monitored along trajectories. It is important to note that protein-lipid H-bond interactions are highly dynamic [30,31] leading to lipid-lipid exchange along the trajectories. The strongest network was observed in the POPC:POPE:Chol (2:1:1) membrane (Fig.55a). This is explained by the higher H-bond donor feature of PE polar heads, with their ammonium N-atom, than of PC polar heads. Interestingly, in the absence of PE lipids, PC contribution to H-bond networks increases. It is worth mentioning that many less H-bonds were observed in cholesterol-protein interactions, owing to the single OH group of cholesterol. However, the presence of cholesterol in lipid bilayer membranes tends to favor PC- and PE-protein H-bond interactions. Cholesterol is known to modulate lipid dynamics by *e.g.,* increasing lipid order and dynamics in fluid lipid bilayer membranes [59]. Furthermore, the presence of cholesterol in artificial membranes is associated with local low and high lipid packing through lipid-lipid interactions [3,59]. Therefore, the presence of cholesterol is expected to decrease PC and PE lipid dynamics, which in turn increases presential lifetime of surrounding lipids. Given the high dynamic feature of lipid-

protein interactions, distributions of surrounding lipids were also calculated focusing on cholesterol and PE polar heads.

The calculated protein-lipid density maps suggested several hotspots for B- and C-helices where *h*OAT1 residues are preferentially in contact with either PE polar heads or cholesterol for more than 80% of the simulations (see Fig.55b and Supplementary Fig.S19). It suggests that specific lipid binding sites exist, in agreement with observations made for other MFS transporters [30–32]. Present analyses exhibited that cholesterol might have strong binding in the TMH1 region of the inner leaflet, as well as in the TMH8 and TMH10 regions of the upper leaflet. Other hot spots were observed between TMH9 and TM12. Regarding PE lipids, a binding site was observed involving residues located on the extracellular site of TMH2 and TMH11, in agreement with observations made for XylE and LacY transporters [31]. Direct interactions between PE and TMH2 and TMH11 were suggested to modulate the conformational state dynamics [30–32,69,70]. For instance, in GLUT transporters, PE lipids were shown to compete over the salt-bridges between N- and C-bundles [31]. In the present OF *h*OAT1 model, PE lipids disrupt the salt bridge between Glu480 and Arg131/Arg138 which ultimately may stabilize the OF state (Fig.55). Regarding protein-lipid interactions on the intracellular side, conformational changes along the transport cycle were reported to be facilitated by lipids through lipid-A-motif interactions [31].

MD simulations proposed that PE lipids preferentially interact with Tyr154, Asp157 and Arg158 which are involved in the charge-relay system. This event was not observed with PC lipids. Even though the present results should be considered carefully, they stress out the central role of PE lipids in *h*OAT1 dynamics and function in agreement with observations made for other MFS transporters. For instance, PE lipids were shown to act as a chaperon facilitating the folding of LacY transporter [72]. Function-wise, lipids were shown to disrupt key salt-bridges which in turn may favor conformational changes along the transport cycle [3]. For example, LacY transporter activity was increased in the presence of PE lipids [73,74]. The same was shown for the xylose (XylE) and Glycerol-3-phosphate (GlpT) transporters, the conformational states of which were also stabilized by PE lipids [31]. Several transporters possess a cholesterol binding site with a distinct role [59]. In GLUT transporters, the presence of cholesterol has been found to stabilize the protein and potentially promote oligomerization [75,76]. Besides, the presence of PE lipids is known to increase membrane fluidity and thus contribute to lipid packing defects [59]. PE components would facilitate the transport cycle by direct interactions with key residues of the transporter [31].

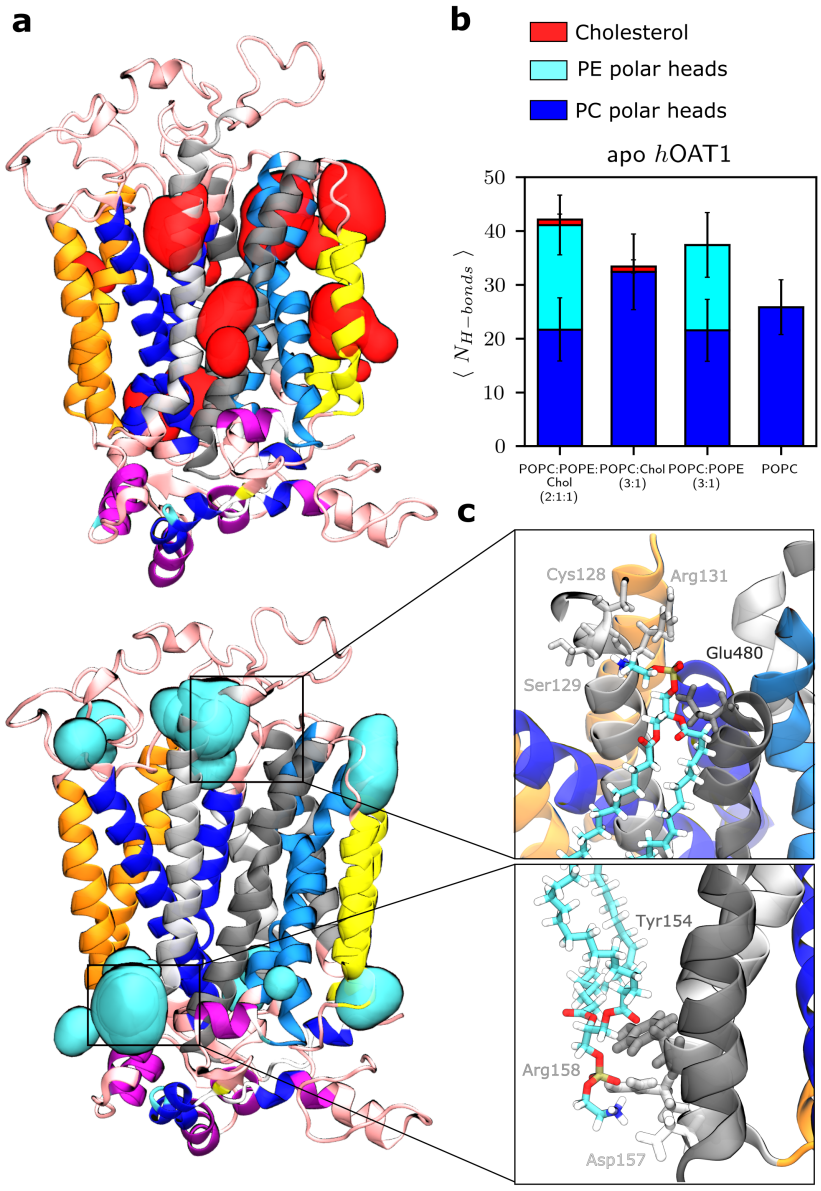


Figure 55. Impact of the membrane lipid components. (a) Hot spots for lipid-protein interactions appearing over 80% of simulations. (b) Number of hydrogen bonds between lipid polar heads and *h*OAT1 for each membrane. (c) Close-up frame points for specific interactions: PE polar heads disrupt salt-bridges between gating residues placed on the extracellular ends of TMH2 and TMH11 (top); PE polar heads interacting with Tyr154, Asp157 and Arg158 by the A-motif (bottom).

III.4. Conclusion

We propose a novel, full molecular model of the human *SLC22A6*/OAT1 transporter in the outward-facing conformation. The present model was thoroughly compared with the recently proposed IF *h*OAT1 model obtained using AF2 and validated by using the conformational space from experimentally resolved MFS transporters. Particular attention was paid to the transmembrane domain for which TMH arrangements are consistent with the OF conformation and literature reports. The role of *h*OAT1 intracellular charge-relay system was investigated, highlighting key residues involved in salt bridges. Comparison with the AF2 IF *h*OAT1 model suggests the existence of two local intracellular arrangements in which conserved motifs may lock *h*OAT1 in the OF conformation. The conformational change is likely facilitated by specific

interactions of PE lipid components with gating and motif residues confirming the dependency of MFS proteins on the composition of lipid bilayer. The present model can be used for further investigation of drug(-drug) interactions (inhibitory studies) by providing atomic pictures and binding affinities for given drugs.

Finally, the present model should help to better understand *h*OAT1 function at the molecular level, pending experimental resolution by means of *e.g.,* cryo-EM techniques. This model should help rationalize known polymorphism or rare mutation by *e.g.*, simply replacing amino acid of importance and achieving MD relaxation. It can also be used to investigate local binding models of small molecules to support substrate/inhibitor competitive experiments.

III.5. References

[1] Q. Faucher, H. Alarcan, P. Marquet, C. Barin-Le Guellec, Effects of Ischemia-Reperfusion on Tubular Cell Membrane Transporters and Consequences in Kidney Transplantation, JCM. 9 (2020) 2610. https://doi.org/10.3390/jcm9082610.

[2] S.K. Nigam, What do drug transporters really do?, Nat Rev Drug Discov. 14 (2015) 29–44. https://doi.org/10.1038/nrd4461.

[3] D. Drew, R.A. North, K. Nagarathinam, M. Tanabe, Structures and General Transport Mechanisms by the Major Facilitator Superfamily (MFS), Chem. Rev. 121 (2021) 5289–5335. https://doi.org/10.1021/acs.chemrev.0c00983.

[4] A. Mateus, L.J. Gordon, G.J. Wayne, H. Almqvist, H. Axelsson, B. Seashore-Ludlow, A. Treyer, P. Matsson, T. Lundbäck, A. West, M.M. Hann, P. Artursson, Prediction of intracellular exposure bridges the gap between target- and cell-based drug discovery, Proc Natl Acad Sci USA. 114 (2017) E6231–E6239. https://doi.org/10.1073/pnas.1701848114.

[5] H. Derendorf, B. Meibohm, Modeling of Pharmacokinetic/Pharmacodynamic (PK/PD) Relationships: Concepts and Perspectives, Pharmaceutical Research. 16 (1999) 176–185. https://doi.org/10.1023/A:1011907920641.

[6] M.A. Risso, S. Sallustio, V. Sueiro, V. Bertoni, H. Gonzalez-Torres, C.G. Musso, The Importance of Tubular Function in Chronic Kidney Disease, IJNRD. Volume 12 (2019) 257–262. https://doi.org/10.2147/IJNRD.S216673.

[7] C.-Y. Sun, M.-S. Wu, C.-C. Lee, S.-H. Chen, K.-C. Lo, Y.-H. Chen, A novel SNP in the 5' regulatory region of organic anion transporter 1 is associated with chronic kidney disease, Scientific Reports. 8 (2018) 8085. https://doi.org/10.1038/s41598-018-26460-y.

[8] J.D. Momper, S.K. Nigam, Developmental regulation of kidney and liver solute carrier and ATP-binding cassette drug transporters and drug metabolizing enzymes: the role of remote organ communication, Expert Opinion on Drug Metabolism & Toxicology. 14 (2018) 561–570. https://doi.org/10.1080/17425255.2018.1473376.

[9] S.K. Nigam, V. Bhatnagar, The systems biology of uric acid transporters: the role of remote sensing and signaling, Current Opinion in Nephrology and Hypertension. 27 (2018) 305–313. https://doi.org/10.1097/MNH.0000000000000427.

[10] D. Martinez, K. Muhrez, J.-B. Woillard, A. Berthelot, E. Gyan, S. Choquet, C.R. Andrès, P. Marquet, C. Barin-Le Guellec, Endogenous Metabolites-Mediated Communication Between OAT1/OAT3 and OATP1B1 May Explain the Association Between *SLCO1B1* SNPs and Methotrexate Toxicity, Clin. Pharmacol. Ther. 104 (2018) 687–698. https://doi.org/10.1002/cpt.1008.

[11] G. Burckhardt, B.C. Burckhardt, In Vitro and In Vivo Evidence of the Importance of Organic Anion Transporters (OATs) in Drug Therapy, in: M.F. Fromm, R.B. Kim (Eds.), Drug Transporters, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011: pp. 29–104. https://doi.org/10.1007/978-3-642-14541-4_2.

[12] S.K. Nigam, K.T. Bush, G. Martovetsky, S.-Y. Ahn, H.C. Liu, E. Richard, V. Bhatnagar, W. Wu, The Organic Anion Transporter (OAT) Family: A Systems Biology Perspective, Physiological Reviews. 95 (2015) 83–123. https://doi.org/10.1152/physrev.00025.2013.

[13] W. Wu, K.T. Bush, S.K. Nigam, Key Role for the Organic Anion Transporters, OAT1 and OAT3, in the in vivo Handling of Uremic Toxins and Solutes, Sci Rep. 7 (2017) 4939. https://doi.org/10.1038/s41598-017-04949-2.

[14] T. Cihlar, D.C. Lin, J.B. Pritchard, M.D. Fuller, D.B. Mendel, D.H. Sweet, The Antiviral Nucleotide Analogs Cidofovir and Adefovir Are Novel Substrates for Human and Rat Renal Organic Anion Transporter 1, Mol Pharmacol. 56 (1999) 570–580. https://doi.org/10.1124/mol.56.3.570.

[15] X. Huo, K. Liu, Renal organic anion transporters in drug–drug interactions and diseases, European Journal of Pharmaceutical Sciences. 112 (2018) 8–19. https://doi.org/10.1016/j.ejps.2017.11.001.

[16] S.M. Mihaila, J. Faria, M.F.J. Stefens, D. Stamatialis, M.C. Verhaar, K.G.F. Gerritsen, R. Masereeuw, Drugs Commonly Applied to Kidney Patients May Compromise Renal Tubular Uremic Toxins Excretion, Toxins. 12 (2020) 391. https://doi.org/10.3390/toxins12060391.

[17] C.-H. Hsueh, K. Yoshida, P. Zhao, T.W. Meyer, L. Zhang, S.-M. Huang, K.M. Giacomini, Identification and Quantitative Assessment of Uremic Solutes as Inhibitors of Renal Organic Anion Transporters, OAT1 and OAT3, Mol. Pharmaceutics. 13 (2016) 3130–3140. https://doi.org/10.1021/acs.molpharmaceut.6b00332.

[18] A.K. Nigam, J.G. Li, K. Lall, D. Shi, K.T. Bush, V. Bhatnagar, R. Abagyan, S.K. Nigam, Unique metabolite preferences of the drug transporters OAT1 and OAT3 analyzed by machine learning, J. Biol. Chem. 295 (2020) 1829–1842. https://doi.org/10.1074/jbc.RA119.010729.

[19] K.M. Hillgren, D. Keppler, A.A. Zur, K.M. Giacomini, B. Stieger, C.E. Cass, L. Zhang, Emerging Transporters of Clinical Importance: An Update From the International Transporter Consortium, Clin Pharmacol Ther. 94 (2013) 52–63. https://doi.org/10.1038/clpt.2013.74.

[20] M.J. Zamek-Gliszczynski, M.E. Taub, P.P. Chothe, X. Chu, K.M. Giacomini, R.B. Kim, A.S. Ray, S.L. Stocker, J.D. Unadkat, M.B. Wittwer, C. Xia, S.-W. Yee, L. Zhang, Y. Zhang, International Transporter Consortium, Transporters in Drug Development: 2018 ITC Recommendations for Transporters of Emerging Clinical Importance, Clin. Pharmacol. Ther. 104 (2018) 890–899. https://doi.org/10.1002/cpt.1112.

[21] S.-M. Huang, R. Temple, D.C. Throckmorton, L.J. Lesko, Drug Interaction Studies: Study Design, Data Analysis, and Implications for Dosing and Labeling, Clin Pharmacol Ther. 81 (2007) 298–304. https://doi.org/10.1038/sj.clpt.6100054.

[22] European Medicines Agency's Guideline on the Investigation of Drug Interactions, (2012). https://www.ema.europa.eu/en/investigation-drug-interactions.

[23] Japan Pharmaceutical and Medical Devices Agency, (n.d.). https://www.pmda.go.jp/files/000206158.pdf.

[24] J.L. Perry, N. Dembla-Rajpal, L.A. Hall, J.B. Pritchard, A Three-dimensional Model of Human Organic Anion Transporter 1: AROMATIC AMINO ACIDS REQUIRED FOR SUBSTRATE TRANSPORT, J. Biol. Chem. 281 (2006) 38071–38079. https://doi.org/10.1074/jbc.M608834200.

[25] 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. https://doi.org/10.1007/s12013-011-9191-7.

[26] Y. Huang, Structure and Mechanism of the Glycerol-3-Phosphate Transporter from Escherichia coli, Science. 301 (2003) 616–620. https://doi.org/10.1126/science.1087619.

[27] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature. 596 (2021) 583–589. https://doi.org/10.1038/s41586-021-03819-2.

[28] K. Tanaka, W. Xu, F. Zhou, G. You, Role of glycosylation in the organic anion transporter OAT1, J Biol Chem. 279 (2004) 14961–14966. https://doi.org/10.1074/jbc.M400197200.

[29] W. Xu, K. Tanaka, A. Sun, G. You, Functional Role of the C Terminus of Human Organic Anion Transporter hOAT1, J. Biol. Chem. 281 (2006) 31178–31183. https://doi.org/10.1074/jbc.M605664200.

[30] C. Martens, R.A. Stein, M. Masureel, A. Roth, S. Mishra, R. Dawaliby, A. Konijnenberg, F. Sobott, C. Govaerts, H.S. Mchaourab, Lipids modulate the conformational dynamics of a secondary multidrug transporter, Nat Struct Mol Biol. 23 (2016) 744–751. https://doi.org/10.1038/nsmb.3262.

[31] C. Martens, M. Shekhar, A.J. Borysik, A.M. Lau, E. Reading, E. Tajkhorshid, P.J. Booth, A. Politis, Direct protein-lipid interactions shape the conformational landscape of secondary transporters, Nat Commun. 9 (2018) 4151. https://doi.org/10.1038/s41467-018-06704-1.

[32] C. Martens, M. Shekhar, A.M. Lau, E. Tajkhorshid, A. Politis, Integrating hydrogen–deuterium exchange mass spectrometry with molecular dynamics simulations to probe lipid-modulated conformational changes in membrane proteins, Nat Protoc. 14 (2019) 3183–3204. https://doi.org/10.1038/s41596-019-0219-6.

[33] A.M. Westerlund, O. Fleetwood, S. Pérez-Conesa, L. Delemotte, Network analysis reveals how lipids and other cofactors influence membrane protein allostery, J. Chem. Phys. 153 (2020) 141103. https://doi.org/10.1063/5.0020974.

[34] The UniProt Consortium, UniProt: a hub for protein information, Nucleic Acids Research. 43 (2015) D204–D212. https://doi.org/10.1093/nar/gku989.

[35] Y. Zhang, I-TASSER server for protein 3D structure prediction, BMC Bioinformatics. 9 (2008) 40. https://doi.org/10.1186/1471-2105-9-40.

[36] Pike, A.C.W., Quigley, A., Chu, A., Tessitore, A., Xia, X., Mukhopadhyay, S., Wang, D., Kupinska, K., Strain-Damerell, C., Chalk, R., Burgess-Brown, N.A., Edwards, A.M., Arrowsmith, C.H., Bountra, C., Carpenter, E.P., hGLUT3, (n.d.). https://www.rcsb.org/structure/5C65.

[37] N. Nomura, G. Verdon, H.J. Kang, T. Shimamura, Y. Nomura, Y. Sonoda, S.A. Hussien, A.A. Qureshi, M. Coincon, Y. Sato, H. Abe, Y. Nakada-Nakura, T. Hino, T. Arakawa, O. Kusano-Arai, H. Iwanari, T. Murata, T. Kobayashi, T. Hamakubo, M. Kasahara, S. Iwata, D. Drew, Structure and mechanism of the mammalian fructose transporter GLUT5, Nature. 526 (2015) 397–401. https://doi.org/10.1038/nature14909.

[38] 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. https://doi.org/10.1038/nature11524.

[39] 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. https://doi.org/10.1016/j.phrs.2018.02.029.

[40] C.R. Søndergaard, M.H.M. Olsson, M. Rostkowski, J.H. Jensen, Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p K_a Values, J. Chem. Theory Comput. 7 (2011) 2284–2295. https://doi.org/10.1021/ct200133y.

[41] 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. https://doi.org/10.1002/jcc.23702.

[42] 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. https://doi.org/10.1021/acs.jctc.5b00255.

[43] 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. https://doi.org/10.1021/ct4010307.

[44] 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. https://doi.org/10.1063/1.445869.

[45] I.S. Joung, T.E. Cheatham, Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations, J. Phys. Chem. B. 112 (2008) 9020–9041. https://doi.org/10.1021/jp8001614.

[46] I.S. Joung, T.E. Cheatham, Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters, J. Phys. Chem. B. 113 (2009) 13279–13290. https://doi.org/10.1021/jp902584c.

[47] D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman, AMBER 2018, (2018).

[48] A.W. Götz, M.J. Williamson, D. Xu, D. Poole, S. Le Grand, R.C. Walker, Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born, J. Chem. Theory Comput. 8 (2012) 1542–1555. https://doi.org/10.1021/ct200909j.

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

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

[51] 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. https://doi.org/10.1063/1.448118.

[52] 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. https://doi.org/10.1021/ct400341p.

[53] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, Journal of Molecular Graphics. 14 (1996) 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

[54] O.S. Smart, J.G. Neduvelil, X. Wang, B.A. Wallace, M.S.P. Sansom, HOLE: A program for the analysis of the pore dimensions of ion channel structural models, Journal of Molecular Graphics. 14 (1996) 354–360. https://doi.org/10.1016/S0263-7855(97)00009-X.

[55] T. Galochkina, M. Ng Fuk Chong, L. Challali, S. Abbar, C. Etchebest, New insights into GluT1 mechanics during glucose transfer, Sci Rep. 9 (2019) 998. https://doi.org/10.1038/s41598-018-37367-z.

[56] A.M. Westerlund, L. Delemotte, InfleCS: Clustering Free Energy Landscapes with Gaussian Mixtures, J. Chem. Theory Comput. 15 (2019) 6752–6759. https://doi.org/10.1021/acs.jctc.9b00454.

[57] N. Yan, Structural Biology of the Major Facilitator Superfamily Transporters, Annu. Rev. Biophys. 44 (2015) 257–283. https://doi.org/10.1146/annurev-biophys-060414-033901.

[58] S. Radestock, L.R. Forrest, The Alternating-Access Mechanism of MFS Transporters Arises from Inverted-Topology Repeats, Journal of Molecular Biology. 407 (2011) 698–715. https://doi.org/10.1016/j.jmb.2011.02.008.

[59] 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. https://doi.org/10.1038/nrm.2015.25.

[60] T.F. Custódio, P.A. Paulsen, K.M. Frain, B.P. Pedersen, Structural comparison of GLUT1 to GLUT3 reveal transport regulation mechanism in sugar porter family, Life Sci. Alliance. 4 (2021) e202000858. https://doi.org/10.26508/lsa.202000858.

[61] 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. https://doi.org/10.1038/nature14655.

[62] 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. https://doi.org/10.1038/nature13306.

[63] K. Kazmier, D.P. Claxton, H.S. Mchaourab, Alternating access mechanisms of LeuT-fold transporters: trailblazing towards the promised energy landscapes, Current Opinion in Structural Biology. 45 (2017) 100–108. https://doi.org/10.1016/j.sbi.2016.12.006.

[64] X.C. Zhang, Y. Zhao, J. Heng, D. Jiang, Energy coupling mechanisms of MFS transporters: Energy Coupling Mechanisms of MFS Transporters, Protein Science. 24 (2015) 1560–1579. https://doi.org/10.1002/pro.2759.

[65] D. Jiang, Y. Zhao, X. Wang, J. Fan, J. Heng, X. Liu, W. Feng, X. Kang, B. Huang, J. Liu, X.C. Zhang, Structure of the YajR transporter suggests a transport mechanism based on the conserved motif A, Proc Natl Acad Sci USA. 110 (2013) 14664–14669. https://doi.org/10.1073/pnas.1308127110.

[66] 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. https://doi.org/10.1038/nsmb.2569.

[67] Å. Västermark, B. Lunt, M. Saier, Major Facilitator Superfamily Porters, LacY, FucP and XylE of ***Escherichia coli*** Appear to Have Evolved Positionally Dissimilar Catalytic Residues without Rearrangement of 3-TMS Repeat Units, J Mol Microbiol Biotechnol. 24 (2014) 82–90. https://doi.org/10.1159/000358429.

[68] C. Zhu, K.B. Nigam, R.C. Date, K.T. Bush, S.A. Springer, M.H. Saier, W. Wu, S.K. Nigam, Evolutionary Analysis and Classification of OATs, OCTs, OCTNs, and Other SLC22 Transporters: Structure-Function Implications and Analysis of Sequence Motifs, PLoS ONE. 10 (2015) e0140569. https://doi.org/10.1371/journal.pone.0140569.

[69] M.P. Muller, T. Jiang, C. Sun, M. Lihan, S. Pant, P. Mahinthichaichan, A. Trifan, E. Tajkhorshid, Characterization of Lipid–Protein Interactions and Lipid-Mediated Modulation of Membrane Protein Function through Molecular Simulation, Chem. Rev. 119 (2019) 6086–6161. https://doi.org/10.1021/acs.chemrev.8b00608.

[70] V. Corradi, E. Mendez-Villuendas, H.I. Ingólfsson, R.-X. Gu, I. Siuda, M.N. Melo, A. Moussatova, L.J. DeGagné, B.I. Sejdiu, G. Singh, T.A. Wassenaar, K. Delgado Magnero, S.J. Marrink, D.P. Tieleman, Lipid–Protein Interactions Are Unique Fingerprints for

Membrane Proteins, ACS Cent. Sci. 4 (2018) 709–717. https://doi.org/10.1021/acscentsci.8b00143.

[71] R.A. Zager, B.M. Sacks, K.M. Burkhart, A.C. Williams, Plasma membrane phospholipid integrity and orientation during hypoxic and toxic proximal tubular attack, Kidney International. 56 (1999) 104–117. https://doi.org/10.1046/j.1523-1755.1999.00533.x.

[72] M. Bogdanov, W. Dowhan, Phospholipid-assisted protein folding: phosphatidylethanolamine is required at a late step of the conformational maturation of the polytopic membrane protein lactose permease, EMBO J. 17 (1998) 5255–5264. https://doi.org/10.1093/emboj/17.18.5255.

[73] H.R. Kaback, L. Guan, It takes two to tango: The dance of the permease, Journal of General Physiology. 151 (2019) 878–886. https://doi.org/10.1085/jgp.201912377.

[74] 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. https://doi.org/10.1016/j.str.2012.08.021.

[75] R.J. Zottola, E.K. Cloherty, P.E. Coderre, A. Hansen, D.N. Hebert, A. Carruthers, Glucose Transporter Function Is Controlled by Transporter Oligomeric Structure. A Single, Intramolecular Disulfide Promotes GLUT1 Tetramerization, Biochemistry. 34 (1995) 9734–9747. https://doi.org/10.1021/bi00030a011.

[76] D.N. Hebert, A. Carruthers, Cholate-solubilized erythrocyte glucose transporters exist as a mixture of homodimers and homotetramers, Biochemistry. 30 (1991) 4654–4658. https://doi.org/10.1021/bi00233a003.

Chapter IV. The modulation of human Organic Anion Transporter's 1 functionality impacted by lipids, substrates and single nucleotide polymorphism

Angelika Janaszkiewicz,[1] Ágota Tóth,[1] Quentin Faucher,[1] Hélène Arnion,[1] Nicolas Védrenne,[1]

Chantal Barin-Le Guellec,[1,2] Pierre Marquet,[1,3] Florent Di Meo,[1,*]

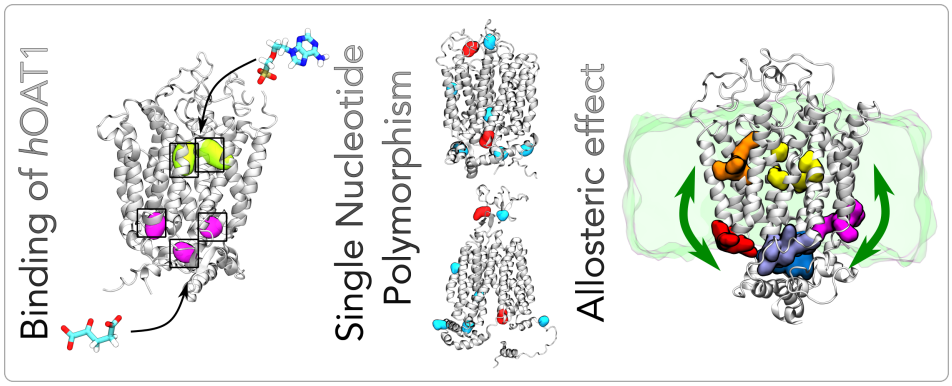
Published in bioRxiv, doi: https://doi.org/10.1101/2022.07.14.500056
Supplementary Information are available here

Abstract

The Organic Anion Transporter 1 is a membrane transporter known for its central role in drug elimination by the kidney. *h*OAT1 is an antiporter, *i.e.*, it translocates substrate in exchange for α-ketoglutarate. The pharmacological attention is drawn by *h*OAT1-mediated drug-drug interactions which may lead to adverse effects. Nonetheless, the understanding of *h*OAT1's structure and function remains limited due to the absence of resolved structure of *h*OAT1. However, owing to conserved structural and functional patterns shared within Major Facilitator Superfamily (MFS) that OAT1 belongs to, an opportunity appeared for structure modelling by protein structure prediction tools and molecular dynamics simulations.

Taking advantage of formerly validated *h*OAT1 models, the present work investigated substrate and co-substrate binding to *h*OAT1, paying attention to allostery between key *h*OAT1 domains. Dynamic pictures were provided by microsecond-scaled molecular dynamic simulations of *h*OAT1 OF bounded to adefovir and/or α-ketoglutarate embedded into POPC-based membranes. Our computational approach has revealed key residues of two binding pockets for adefovir. α-Ketoglutarate might bind to the intracellular charge-relay system highly conserved motifs within MFS. Allostery between these binding pockets was also investigated highlighting the active role of the surrounding lipid bilayer. Lastly, using the structure of *h*OAT1 in OF as well as inward-facing conformation obtained from AlphaFold 2 (AF2) structure prediction tool, we proposed a structural rationalization of transport impairments experimentally observed for 2 significant single nucleotide polymorphisms, namely Arg50His and Arg454Gln.

[1] *Inserm U1248 Pharmacology & Transplantation, Univ. Limoges, 87000 Limoges, France*
[2] *CHU de Tours, 2 Boulevard Tonnellé, 37044 Tours, France*
[3] *Department of Pharmacology and Toxicology, CHU Limoges, F-87000 Limoges, France*
** Corresponding Author: Florent Di Meo – florent.di-meo@inserm.fr, INSERM U1248, Pharmacology & Transplantation, CBRS, Faculté de Médecine et Pharmacie, Univ. Limoges, 2 rue du Prof. Descottes, 87000 Limoges, France*



Graphical abstract. The present work (from left): (i) reveals binding modes of adefovir (top) and a-ketoglutarate (bottom) to hOAT1; (ii) maps Single Nucleotide Polymorphisms on outward-facing (top) and inward-facing (bottom) conformation of *h*OAT1; (iii) asses the allosteric effect of lipidic environment and presence of substrates.

Highlights

Adefovir has at least two binding pockets on *h*OAT1 in the outward-facing conformation.

The highly conserved B-motif within MFS is strongly involved in substrate binding.

α-Ketoglutarate binds to the intracellular domain of *h*OAT1 and destabilizes its OF conformation.

The lipid membrane bilayer plays an active role in the allosteric communication between intracellular and extracellular domains of *h*OAT1.

Keywords: Membrane Transporters; Structural Pharmacology; Molecular Dynamics; Protein-lipid interactions; Major Facilitator Superfamily

IV.1. Introduction

Membrane transporters such as ATP-binding cassette (ABC) proteins and solute carriers (SLCs) are responsible for substrate translocation across cell membranes in a myriad of pharmacological and physiological events. At the cellular level, membrane transporters are involved in the selective influx and efflux of a broad range of compounds, including xenobiotics and endogenous substances. From the physiological perspective, the transepithelial transport of hormones [1], toxins [2], metabolites [3], and signalling molecules [4] at the interface between biofluids (e.g., blood/urine, blood/bile) plays an active role in body homeostasis as proposed by the "Remote Sensing and Signalling Theory" (RSST) which describe the transporter-mediated multi-organ regulatory system [4]. Membrane transporters are also involved in pharmacological events such as the drug absorption, distribution, metabolism, and elimination (ADME). Membrane transporters directly govern local pharmacokinetics (PK), *i.e.*, drug concentration at the target sites, whether linked with drug therapeutic or adverse effects [5,6]. Understanding the role of membrane transporters in systemic drug PK is thus essential since defects in their expression and/or function might alter drug exposure, disposition, and response [4,6]. The International Transporter Consortium has listed transporters of "emerging clinical importance" for which the functional evaluations are recommended for new drug development [6,7]. The human organic anion transporter 1 (*SLC22A6*/*h*OAT1) belongs to this list given its central role in drug elimination as well as in the RSST [6–8]. *h*OAT1 is localized at the basolateral membrane of kidney proximal tubular cells (PTCs), standing as a central mediator of renal elimination [9,10]. *h*OAT1 is involved in the translocation of a broad range of endogenous substrates such as metabolites and signalling molecules. *h*OAT1 is also responsible for the elimination process of xenobiotics including antiviral (*e.g.*, adefovir), anticancer and antituberculosis drugs [1].

*h*OAT1 is an antiporter, *i.e.*, the uptake of substrates from the blood into PTCs is coupled to the excretion of α-ketoglutarate (αKG) as a co-substrate into blood circulation [11]. This requires a high intracellular concentration of αKG in PTCs. That is maintained thanks to the so-called tertiary active transport model in which *h*OAT1 activity is coupled with Na^+/K^+-ATPase pump and NaDC3/*SLCA13A3* transporter [1,12]. Atomic-scaled mechanism of *h*OAT1-mediated transport remains unclear owing to the lack of experimentally resolved structure. Structural patterns of *h*OAT1 might provide insights into its roles regarding drug efficacy, toxicity, and elimination [13,14]. We recently proposed a structural and dynamic model of *h*OAT1 embedded in lipid bilayers [15]. *h*OAT1 is expected to adopt the Major Facilitator Superfamily (MFS) fold for which substrate translocation follows the alternating access model. It requires large-scale conformational changes between at least two major conformational states, namely outward-facing (OF) or inward-facing (IF) conformations responsible for substrate binding and release events, respectively [16]. *h*OAT1 structure encompasses 12 transmembrane helices (TMH) organized into N- (TMH1-6) and C-bundles (TMH7-12). *h*OAT1 transmembrane domain exhibits pseudo symmetrical repeats [15] of functional helices, namely A-, B- and C-helices. Briefly, the inner cavity is made of A-helices (TMH1, 4, 7 and 10), which are expected to bind substrates and structurally adjust along the transport cycle. B-helices (TMH2, 5, 8, 11) stand at the interface of the bundles, contributing to the large conformational changes along transport cycle. Finally, the out-of-the-core C-helices (TMH3, 6, 9 and 12) were suggested to maintain the structural integrity of the transporter [17] through interactions with surrounding lipids. *h*OAT1 also exhibits a large extracellular loop (ECL) between TMH1 and TMH2 in which four glycosylation sites were described. ECL was suggested to be involved in trafficking; however, its glycosylation is non-

essential for the transport function [18,19]. On the other side of the membrane, the intracellular region is involved in regulations of transport function and expression [19,20]. The intracellular region was suggested to consist in 6 intracellular helices which are tightly connected to TMH intracellular regions. It is featured by the presence of charged amino acids in conserved motifs within MFS proteins. The so-called A-, E[X_6]R and PETL motifs are repeated in both N- and C-bundles. The interactions between these motifs differ between IF and OF conformations [15]. Knowledge about binding modes of *h*OAT1 substrates remain fragmented while they are of particular importance for the understanding and prediction of transporter-mediated drug-drug interactions (DDIs) [21–23]. Likewise, investigating the structural impact of single nucleotide polymorphisms (SNPs) in *h*OAT1 may help to understand the mild transport impairments of *h*OAT1 observed experimentally, even though their clinical impacts may be rather limited owing to substrate overlap with other PTC influx transporters (e.g., *h*OAT3).

The present study aimed to model the binding modes of adefovir, an acyclic nucleoside phosphonate (ANP) antiviral, to *h*OAT1. Adefovir was chosen as a *h*OAT1 substrate prototype given the extensive experimental data available in the literature [1,24,25]. To better understand the interplay with co-substrate translocation, attention was paid to the binding of co-substrate αKG on the intracellular region as well. The allosteric communication between substrate and co-substrate through their binding sites was also studied, considering the contribution of surrounding lipids (i.e., phosphatidylcholine and phosphatidylethanolamine lipids as well as cholesterol) since it was previously described as being key for other MFS proteins [26–28].

IV.2. Methods

IV.2.1. Binding of (co-)substrates to *h*OAT1 model

From our previous μs-scaled MD simulations performed on the apo *h*OAT1 OF model, representative snapshots were extracted to investigate the binding modes of αKG and adefovir as co-substrate and substrate, respectively. Initial poses for (co-)substrates were obtained by means of molecular docking calculations using AutoDock Vina 1.1.2 [29]. αKG and adefovir were modelled as dianionic compounds given the physiological pH. αKG and adefovir initial structures were obtained from PubChem database [30] and optimized by quantum mechanical methods at the (CPCM)-M06-2X/6-31+G(d,p) level of theory using the Gaussian16 Rev. A package [31]. Molecular docking search volumes were defined with 66560 and 64768 $Å^3$ for the *h*OAT1 extracellular and intracellular sides respectively for adefovir and αKG given the antiporter transport (see Table S1 and Figure S1 for the definitions of box size and centre). For each substrate, 20 molecular docking calculations were carried out providing 20 molecular poses each, leading to 400 poses per substrate for each membrane. For each substrate, three initial binding poses were selected accounting for the calculated affinity scores as well as for the experimentally-identified key binding residues (see Figure S2) [32–34]. To confirm their relevance, molecular docking poses were then used as initial positions for μs-scaled MD simulations.

Three different systems were considered for further MD simulations, namely αKG-, adefovir- and αKG-adefovir-bound *h*OAT1. For each system, three binding modes were considered for further MD simulations (see Figure S2). In total, nine *h*OAT1 systems were used for MD simulations (see Figure S2). It is worth mentioning that two αKG molecules were systematically considered for αKG-based systems in order to optimize sampling. All systems were embedded into three types of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC)-based

membranes namely POPC, POPC and cholesterol (POPC:Chol (3:1)), POPC with cholesterol and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPC:POPE:Chol (2:1:1)). In total, 27 different systems were considered in the present study. Systems were all solvated with water using a 0.154M NaCl concentration to mimic extracellular physiological conditions. System compositions and sizes are reported in supporting information (Table S2). It is worth mentioning that one position (namely, position 3 in Figure S2) was excluded since adefovir left the binding cavity in all simulations, except in POPC:POPE:Chol (2:1:1) membrane. However, in these simulations, adefovir mostly interacted with lipid components at the lipid-protein-water interface rather than with *h*OAT1 residues.

IV.2.2. MD simulation setup

For the definitions of the protein, lipids and water, the following forcefields were applied Amber FF14SB [35], Lipids17 [36] and TIP3P [37], respectively. The parameters for counterions (Na^+, Cl^-) were obtained from Joung and Cheatham [38,39]. The definition of the co-substrate and substrate required a parametrization using GAFF2 and DNA.OL15 forcefield's parameters, considering the latter being a purine derivative. MD simulations were performed using a similar setup as previously established [15]. The simulations were carried out with using the CPU and GPU codes of the Amber18 package [40], [41]. The system was treated with periodic boundary conditions. Non-covalent interactions were considered within the cut-off of 10 Å for electrostatic and Lennard-Jones potentials. Long-range electrostatic interactions were treated using the particle mesh Ewald (PME) method [42]. The bonds involving hydrogen atoms were fixed by applying the SHAKE algorithm. Integration time step was set at 2 fs. To maintain the physiological conditions, the temperature was set at 310K and carried on using a Langevin thermostat [43]. Whereas the Monte Carlo barostat was applied to maintain the constant pressure boundary condition under semi-isotropic conditions [44].

The equilibration of the system was pursued by minimizing all atomic positions and followed by a smooth two-step thermalization. The system was heated up from 0 to 100 K during 200 ps under (N, V, T) conditions, while the second step of the thermalization up to 310K was carried out under semi-isotropic (N, P, T) conditions. Subsequently, equilibration simulations were performed for 5.5 ns. Finally, the MD simulations of substrate-bound *h*OAT1 resulted in a duration of 1.5 μs each, leading to a total of *ca.* 40.5 μs. Trajectory snapshots were saved every 10 ps.

IV.2.3. Analysis

Given the previous validation of the model [15], structural analyses were based on our recent findings, as well as on the general knowledge of MFS patterns [16]**.** Structural analyses were performed using the PyTRAJ and CPPTRAJ AMBER modules [45] as well as the VMD software [46]. Based on the time-dependent backbone root-mean squared deviations (Figure S3), analyses were performed on the equilibrated section of the present trajectories, i.e., over the last 800 ns (Figure S3). H-bond analyses were performed using distance and angle cutoffs set at 3.0 Å and 135°. The minimum fraction threshold was set at 0.1 given the known uncertainties for side chain rotameric states in protein threading techniques. MD trajectories obtained from our previous study was also used for sake of comparison between apo and bound *h*OAT1 systems.

To unravel the allosteric communications between key regions of *h*OAT1, we carried out network analyses using Allopath tool [47]. Since lipids or ligands are known to modulate the

activity of certain proteins, here we calculated communication between distant domains using the Allopath tool, including substrates and lipids as cofactors to measure their impact. Atoms are grouped as nodes, and connections between them are described by edges, which is the term used to describe residue network interaction. However, to include cofactors such as substrates and lipids, atoms are grouped into domains, represented as interactors, where in the case of the lipids, a lipid component is divided into a head group and two hydrophobic tails. The allopath approach takes into account the dynamics of the lipids in the membrane, so regardless of the lipid exchange, it calculates its contribution. The so-called network adjacency matrix is obtained from the contact map and positional mutual information (correlation matrix) to model spatial proximity and correlation of residue movements. After defining the distant domains as a source and a sink, the information flow (also referred to as the current flow) is calculated to measure the communication between the source and the sink. The network is aiming to find the most efficient pathway between the two domains, therefore calculating betweenness, the involvement of nodes and interactors in the pathway, and closeness. Both are used to extract the allosterically important residues, but also to explore the allosteric roles of lipids and substrates. Current flow closeness centrality, also known as efficiency, describes communication between distant domains. Communication efficiencies were calculated unidirectionally from “source” to “sink” residues (see Ref. [47] for more details). In the present work, allosteric communications between binding pocket residues (B-like motif and the inner binding cavity, see Section 3.3) and the charge-relay system (A- and E[X_6]R motifs, see Table S3 and Section 3.3) were calculated. The contributions of interactors were assessed by calculating communication efficiencies on systems made of: (i) *h*OAT1 *apo*, (ii) *h*OAT1 substrate-bound, (iii) *h*OAT1 *apo* with lipids, and (iv) *h*OAT1 substrate-bound with lipids.

IV.3. Results and Discussion

IV.3.1. Interactions of the *in silico h*OAT1 model with (co-)substrates

IV.3.1.1. Substrate interactions with *h*OAT1 cavities

*h*OAT1 was shown to play a key role in the uptake of anionic endogenous and exogenous substrates, such as antiviral acyclic nucleoside phosphonates (ANP) (*e.g.*, tenofovir, adefovir) [34,48], urate, *p*-aminohippurate (PAH), β-lactam antibiotics and sulfate conjugates [49]. The presence of numerous cationic residues (i.e., lysine, arginine, and histidine, see Figure 56A) at the *h*OAT1 extracellular interface in OF and IF *h*OAT1 models are consistent with the expected high affinity for anionic substrates, by favouring electrostatic interactions. Cationic residues were found in the ECL between TMH1 and TMH2 as well as in the water-exposed cavity in the OF *h*OAT1 model. Molecular docking calculations suggested three possible binding positions for adefovir as well as for αKG in the intracellular regions (Figure S2). MD simulations confirmed the importance of A-helices in the water-exposed cavity, mainly TMH1 and TMH4. The latter was previously reported as key for substrate binding and translocation owing to the presence of the conserved motif RXX[Q/S]G [50]. The so-called B-like motif was observed in several MFS antiporters, even though it cannot be considered as an antiporter fingerprint [50]. MD simulations reveal that adefovir may strongly bind to the B-like motif Arg192 residue thanks to a salt-bridge between phosphonate and guanidinium moieties as pictured by calculated large H-bond fraction at 1.950 over the MD simulation (see Figure 56B and Table S4).

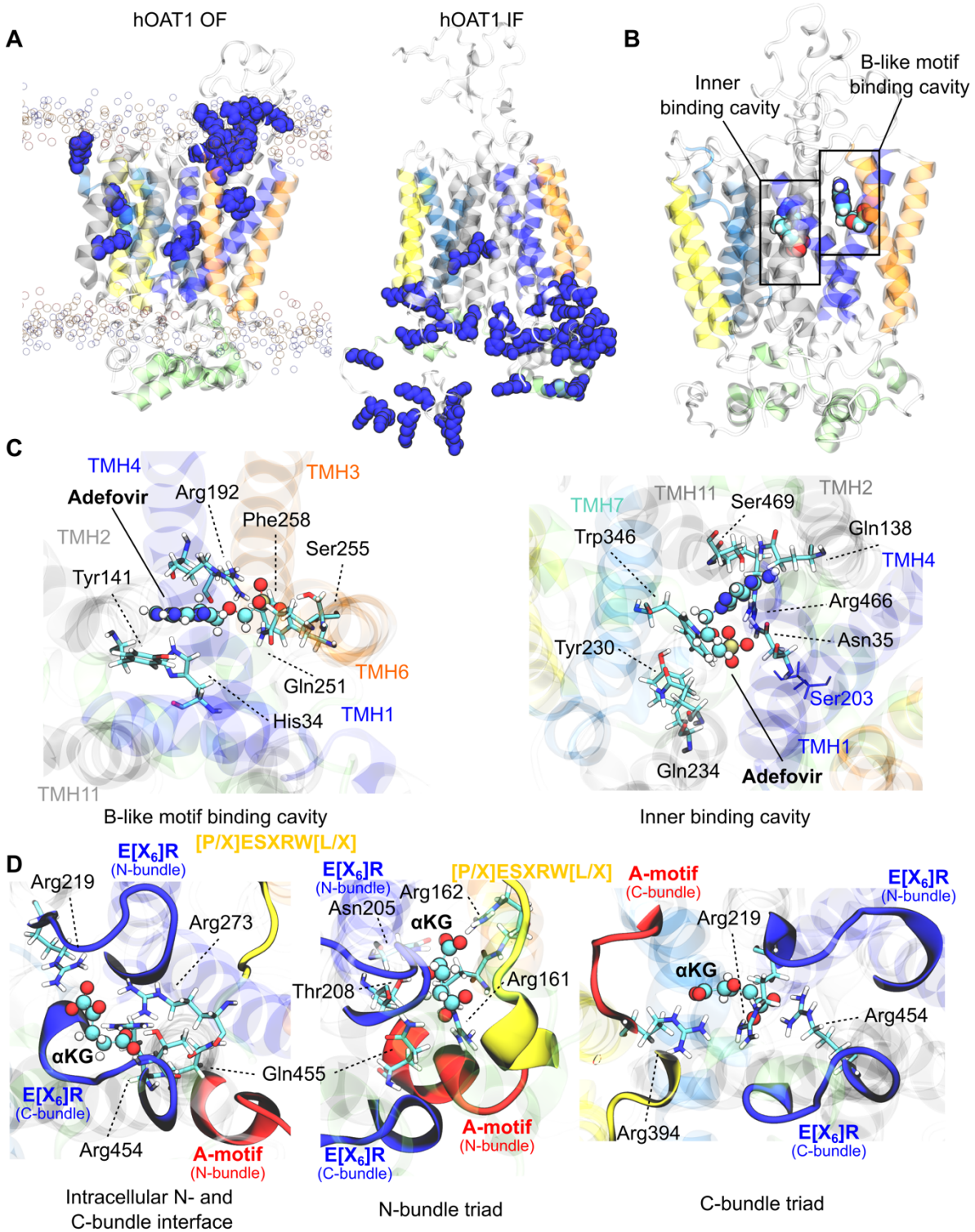


Figure 56. Interactions with (co-)substrate. A) Representation of cationic residues present in extracellular and intracellular sites of the OF (left) and IF (right) models, respectively. B) The water-exposed cavity exhibits 2 binding pockets for adefovir: the B-like motif and the inner binding cavity involving TMH1 and TMH4 (dark blue). C) Residues involved in adefovir binding in the B-like motif (left) and inner (right) binding cavity. D) Three mains αKG binding spots and the most frequently interacting residues. A-, B- and C-helices are coloured blue-, grey- and yellow-ish, respectively.

H-bond analyses also highlighted the key role of several polar and aromatic residues, namely Ser255, Gln251, Ser195 and Tyr141. π-Stacking interactions with Tyr141 were also observed between the adefovir adenosyl moiety and tyrosine phenol moiety contributing to the stabilization of the substrate. This is in line with the *h*OAT1 substrate spectrum which includes various anionic aromatic compounds (*e.g.*, ANPs, PAH, urate). It is worth mentioning that the *h*OAT1 OF model suggested two possible accesses to the B-like motif cavity. On one hand, substrates may enter directly from the water phase through the ECL cationic residue network. On the other hand, the OF *h*OAT1 model suggests that there may exist an access channel between TMH3 and TMH6 connecting the high-density polar head region of the outer leaflet membrane to the B-like motif cavity. This is in line with biophysical studies showing that amphiphilic substrates might partition below polar head region of the lipid bilayer membrane [51].

Molecular docking calculations also suggested the existence of an inner binding cavity at the interface between N- and C-bundles, involving mostly TMH1 and TMH7. The substrate is again stabilized thanks to a strong H-bond network (Figure 56D). Cationic Arg466 is likely to play a central role for anionic substrate translocation (Table S4 and Figure 56C). This is in good agreement with experimental observations in which site-directed mutagenesis of Arg466 was associated to substrate-dependent loss of transport capacity (Table S5) [52]. Likewise, several residues of the site-directed mutagenesis which were associated with the loss of transport function are present in either the first- or the second shell around adefovir (*e.g.*, Asn39 [18], Tyr230 [32], Tyr353 [53], Trp346 [53], see Table S5). This confirms the relevance of the region obtained from MD-refined molecular docking calculations. Particular attention was also paid to Ser203 (Figure 56C) which was described as central for the binding of several ANPs. In our simulations, Ser203 belongs to the second-shell of contact residues with adefovir. Therefore, no direct interactions between adefovir and Ser203 were observed in our calculations. However, it is worth mentioning that binding to Ser203 was suggested from static structural investigations performed using an IF *h*OAT1 homology model [34]. Therefore, in line with experimental investigations, we can hypothesize that Ser203 is likely involved in the substrate translocation by carrying substrate from the outer to the inner leaflet as part of the OF-to-IF large-scale conformational transition. Furthermore, this is strengthened by the recent AF2 IF model [54] in which Ser203 is deeply located in the MFS core, precluding direct substrate binding in the OF conformation.

MD simulations suggest that the inner binding cavity is slightly less accessible than the B-like motif pocket, likely due to the absence of an access channel. Interestingly, the two cavities are contiguous, around TMH1. Therefore, present results as well as experimental observations regarding mutations of residues located in the inner binding cavity [32,33,52,53] suggest that substrate binding might occurs first in the B-like motif cavity. Substrate binding to B-like motif pocket is thus expected to rapidly lead to local conformational changes which may open an access channel to the inner binding cavity. Binding to inner cavity is in turn expected to be then pivotal for substrate translocation events along the transport cycle. Finally, rocker-switch large-scale TMH conformational changes occurring during the translocation event might then decrease substrate binding affinity, favouring substrate release in the intracellular medium.

IV.3.1.2. Investigations on α-ketoglutarate binding at the *h*OAT1 intracellular interface

*h*OAT1 transport cycle requires large conformational changes for the OF-to- IF transition. *h*OAT1 being an antiporter, substrate uptake is coupled with the co-transport of αKG in the

opposite direction [11]. However, the sequence of events (*i.e.*, αKG efflux and substrate influx) remains unclear as well as triggers. However, our present model can be used to assess binding modes of αKG in the intracellular region of *h*OAT1. As stated for adefovir binding at the extracellular interface, intracellular domains are also rich in cationic residues favouring electrostatic interactions with anionic αKG molecules. These cationic residues are exposed to the intracellular medium independently on the conformational state (Figure 56A and 1B). Even though αKG translocation is expected to reset the OF conformation, our simulations suggest that αKG can also bind intracellular domains, regardless of the *h*OAT1 conformational state.

MD-refined molecular docking calculations also exhibited three preferential binding modes for αKG: (i) on the N- and (ii) C-bundle motif triads; or (iii) at the interface between N- and C-bundle E[X_6]R motifs (Figure 56D). Owing to its two anionic carboxylate moieties, αKG favours electrostatic interactions with the charge-relay system cationic residues, as pictured by described H-bond network from MD simulations (Figure 56D, Tables S4 and S6). For instance, strong salt-bridges were observed along the simulations between αKG and A-motif arginine residues: Arg162 and Arg161, or Arg394 for binding poses in the N- or C-bundle triad, respectively (H-bond fractions=2.019, 1.698 and 1.376 for Arg162, Arg161 and Arg394, respectively). Likewise, the intracellular charge-relay system may play a central role in αKG binding at the interface between N- and C-bundle E[X_6]R motifs (*e.g.*, Arg219, Arg454). In turn, MD simulations with αKG bound system to the intracellular domain also reveal that the charge-relay system might be disrupted by the presence of αKG (Table S6). Assuming that αKG can bind *h*OAT1 in the OF conformation, present results suggest that αKG may play favour substrate translocation by destabilizing intracellular N- and C-bundle interactions, which is necessary for OF-to-IF transition along with substrate translocation. These findings pave the way for further investigations to establish whether the IF-to-OF transition resetting *h*OAT1 conformation is driven by αKG efflux or if αKG and the other substrate are simultaneously transported; even though the latter is expected to be less likely [16].

IV.3.2. Allosteric communication between the substrate binding pockets and intracellular domains

The comparison of the intracellular structural arrangement between OF and IF apo *h*OAT1 models revealed that the charge-relay system must be disrupted along transport cycle. This might ultimately unlock the IC gating required for substrate release [15]. Conformational changes are expected to be triggered upon substrate and/or co-substrate binding, suggesting the existence of a distant communication between substrate binding pockets and MFS conserved intracellular motifs. Allosteric communications were thus monitored providing structural insights about the efficiency regarding the different domains located across the membrane (see Section 2.3 and Ref. [47] for technical details). Particular attention was paid to the plausible communication between binding domains (B-like motif and inner substrate binding site) with the charge-relay system (A- and E[X_6]R motifs). It is worth mentioning that (i) N- and C-bundle IC motifs were considered separately and (ii) allostery was monitored in both directions, *i.e.*, from EC to IC regions and *vice versa* leading to 16 different pathways (Figures 57A). The role of substrate, co-substrate and surrounding lipid bilayer were also taken into account. However, results provided here are only qualitatively discussed since the resolution of the present model, especially regarding side chains, precludes quantitative conclusions (Figure 57 and S4-S11).

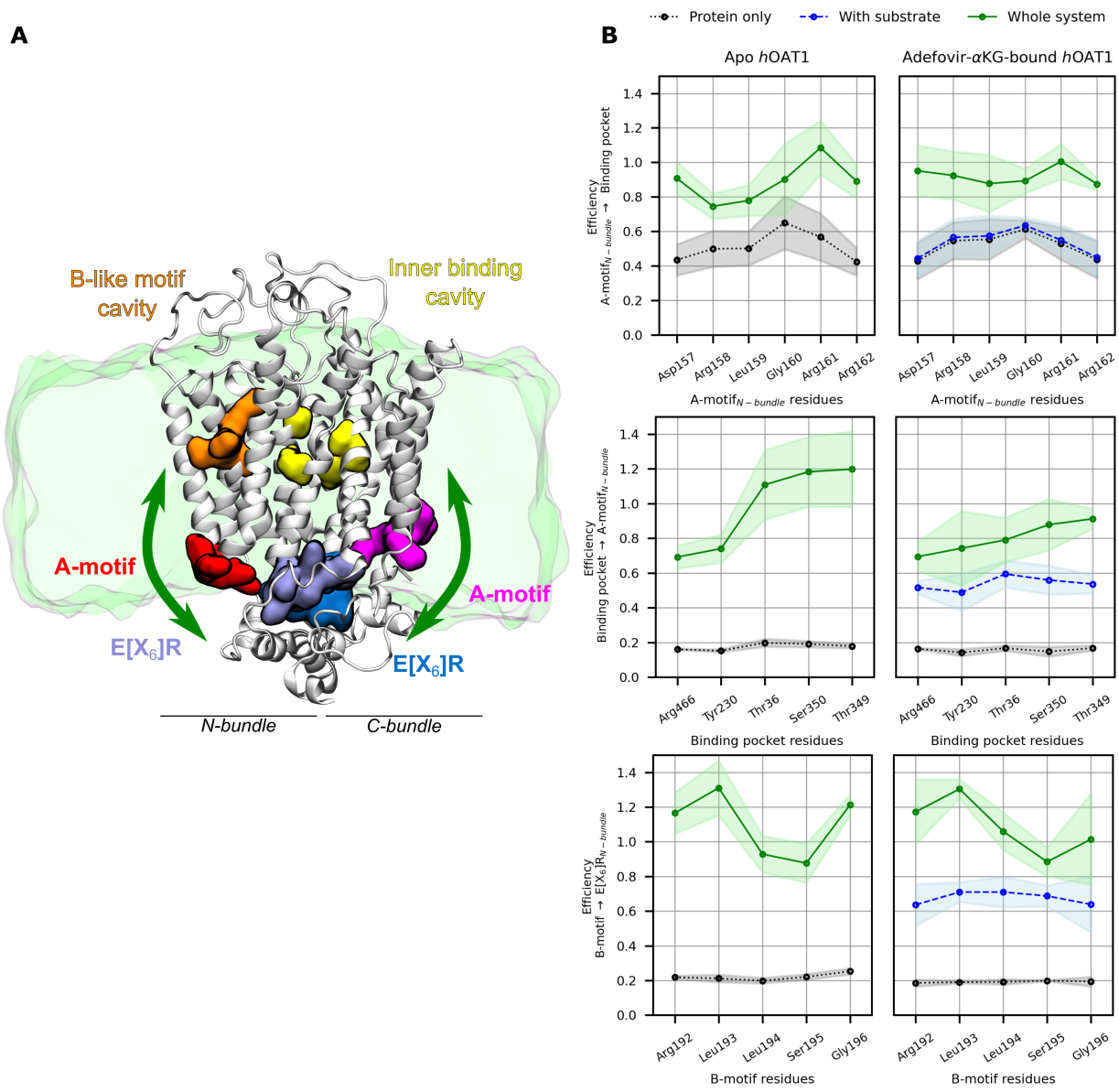


Figure 57. The allosteric effect of the membrane lipid bilayer in the presence and absence of substrates on the *h*OAT1 communication efficiency from biding pockets toward charge-relay system and reverse. (A) Visualisation of binding pockets (B-like motif cavity on left and inner binding cavity on right) and motifs of charge-relay system divided into N and C domain. The wide green arrows represent strong communication in presence of lipids. (B) The schemes plot the communication efficiency calculated for pure protein (black dashed line), protein in presence of substrates (blue dashed line) and protein accounting lipids in apo (left column) or substrate-bound state (right column) in green line.

As an example, Figure 57B shows efficiencies of selected allosteric pathways. MD simulations revealed the existence of a reciprocal distant communication between intracellular motifs and substrate binding pocket. While considering only the role of *h*OAT1 protein, allosteric pathway analyses revealed an efficient communication from A-motifs to substrate binding pockets, while other pathways remain inherently low. Furthermore, these allosteric pathway efficiencies exhibit very similar profiles for both B-like motif and inner binding pockets. Interestingly, efficiencies calculated for allosteric pathways from A-motifs to substrate binding pockets were not affected by the presence of substrates. In contrast, the presence of substrate and/or co-substrate strongly potentiated the efficiency of the allosteric communications from binding

pockets to intracellular motifs. In other words, substrate binding events are thus expected to modulate the dynamics of intracellular motifs. Such observations are in line with the proposed mechanism of *h*OAT1 transport for which binding events might trigger the conformational changes in the IC region required for substrate translocation. The contributions of *h*OAT1 residues to allosteric networks were also calculated underlining the pivotal roles of TMH2, TMH3 and TMH4 which are involved either in binding pockets or in A-motifs.

Besides, it is important to note that MFS protein functions have been experimentally shown to strongly depend on the membrane composition [26,27,55]. Likewise, PE lipids were shown to non-covalently bind with the residues standing at the interface between N- and C bundles in apo *h*OAT1 structure [15]. Therefore, out of the structural interplay between surrounding lipid bilayer and *h*OAT1 protein, the active role of lipid bilayer membrane in the communication between extracellular and intracellular regions was also investigated. Efficiencies drastically increased while considering the membrane. This confirms that the membrane plays an active part in allosteric signalling as recently shown for other membrane proteins [47], including transporters [56]. This also underlines the central role of protein-lipid interactions and the need for representative and realistic membranes, including PE and cholesterol, for computational and experimental investigations on hOAT1. At the body level, in situ alteration or modulation of cell membranes may affect the activity of hOAT1 either by modifying the global structure and dynamics of hOAT1, or its intrinsic function [57,58]. Such observations may be extended to other MFS transporters given the high protein-lipid dependency.

IV.3.3. Structural mapping of *h*OAT1 single nucleotide polymorphisms

Beyond pharmacologically relevant mechanistic insights into *h*OAT1-mediated drug transport, the present structural model substrate can also be used to understand the modulation of substrate transport across cell membranes by naturally-occurring polymorphisms in genes coding transporters [59–62]. Several SNPs in *h*OAT1 were identified and classified according to ethnical origins and species [34]. Since most SNPs are located in untranslated intronic regions, the focus will be given here to non-silent mutations (Table 4 and Figure 58), which correspond to a limited number of SNPs reported in the literature. Moreover, most of them exhibit no or limited alteration of transport function. The present IF and OF *h*OAT1 models can help understand the structural changes induced by genetic polymorphisms. Particular attention will be paid to those which are not associated with either protein expression or altered membrane trafficking. Indeed, protein structure prediction tools should be carefully considered when variants exhibit significant folding differences [63].

Table 4. List of *h*OAT1 SNPs and rare variants. The reported impairment of *h*OAT1 function is shown in bold.

Nucleotide change	Variant	Position on OAT1 gene sequence[a]	Amino acid change	*h*OAT1 location	Effects	References
T>C	rs1415632329	20 (Exon 1)	p.Leu7Pro	N-terminal region	No data	[64–66]
G>A	**rs11568626**	**149 (Exon 1)**	**p.Arg50His**	**ECL1**	**No change in transport of PAH, OTA, MTX; Decreased transport affinity for AFV, CDF and TNF**	[59,62,64,65,67]
C>T	rs11568627	311 (Exon 1)	p.Pro104Leu	ECL1	No change in transport of PAH, OTA, MTX	[62,65]

T>C	rs11568623	677 (Exon 4)	p.Ile226Thr	TMH5	No change in transport of PAH, OTA, MTX	[62,65]
C>T	rs11568624	767 (Exon 4)	p.Ala256Val	TMH6	No change in transport of PAH, OTA, MTX	[62,65]
C>T	rs45607933	877 (Exon 5)	p.Arg293Trp	Intracellular region	No change in transport of PAH, OTA, MTX	[62,65]
G>A	**rs11568634**	**1361 (Exon 8)**	**p.Arg454Gln**	**C-bundle E[X_6]R**	**Loss of transport activity of PAH, OTA and MTX; not associated with PKTD induced by AFV treatment for chronic hepatitis B**	**[24,62,65]**
A>T	nd	1575 (Exon 10)	p.Lys525Ile	C-terminal region	No change in transport of PAH, AFV, CDF, TNF	[59,65]

[a] Gene sequence position are related to the ATG start site.
Abbreviations: AFV, adefovir disproxyl fumarate; CDF, cidofovir; MTX, methotrexate; OTA ochratoxin A; PAH, p+-aminohippurate; PKTD, proximal kidney tubular dysfunction; TNF, tenofovir, not determined (structures of the drugs are shown in Figure S12).

Eight SNPs have been considered and mapped on the present IF and OF *h*OAT1 models (Figure 58). They can be classified according to their location on the model, as follows: (i) in the N-terminal domain (p.Leu7Pro); (ii) in the long ECL between TMH1 and TMH2 (p.Arg50His and p.Pro104Leu); (iii) in TMH5 and TMH6 (p.Ile226Thr and p.Ala256Val); and (iv) in the intracellular domain (p.Arg293Trp, p.Arg454Gln and p.Leu525Ile). SNPs leading to p.Pro104Leu, p.Ile226Thr, p.Ala256Val, p.Arg293Trp and p.Lys525Ile substitutions were not associated with alteration of substrate intake [24,59,62,64,67]. Only two SNPs were reported to impair *h*OAT1 function, namely rs11568626 and rs11568634 leading to p.Arg50His and p.Arg454Gln protein mutations, respectively. The rs11568626 SNP was found to be specific to the African population. The resulting p.Arg50His mutation was shown to be associated with decreased transport affinities (K_m) of phosphate analogues such as adefovir, cidofovir and tenofovir in Xenopus oocytes-expressed p.Arg50His variant as compare to wild-type *h*OAT1 (rs**15914676)**. However, Xenopus oocytes-expressing p.Arg50His variant exhibited normal uptake of PAH, ochratoxin A and methotrexate [62].

Interestingly, Arg50 is located in a conserved motif on the aforementioned ECL1 pattern, rich in cationic residues and likely involved in anionic substrate access to the B-like motif binding pocket. Therefore, we can hypothesize that the p.Arg50His substitution may lead to a lower electrostatic potential between the anionic substrate and the binding access channel given the poor ECL resolution of the present OF model, this must be confirmed by further experiments or through the experimental structure resolution of *h*OAT1. The rs11568634 SNP leads to the substitution of Arg454 by a glutamine residue. Interestingly, this SNP was associated to decreased uptake of PAH, ochratoxin A and methotrexate, suggesting a loss of function [62]. This is in perfect agreement with present structural observations which underlined the central role of Arg454 in the intracellular charge-relay system: Arg454 is likely involved in salt-bridges between the C-bundle A-motif and the PETL motif. The p.Arg454Gln substitution might weaken the local supramolecular arrangement, which in turn is likely to disrupt the charge-relay system essential for the MFS transport cycle.

It is worth mentioning that no *h*OAT1 SNP was associated with pathological conditions at the clinical level. Owing to the central role of OATs in RSST [4], the substrate overlap between *e.g., h*OAT1 and *h*OAT3 [11] might lead to compensation activity between transporters.

However, systemic compensation between transporters might not be sufficient to overcome *h*OAT1 impairment in chronic conditions such as chronic kidney diseases. Recently, particular attention has been paid to the loss of tubular function in CKD [68], where *h*OAT1 impairment is expected to play a key role [69]. Further clinical investigations are required to examine *h*OAT1 SNPs as risk factors for the still unclear patient variability in *e.g.*, long-term ANP nephrotoxicity [25,59,62,70].

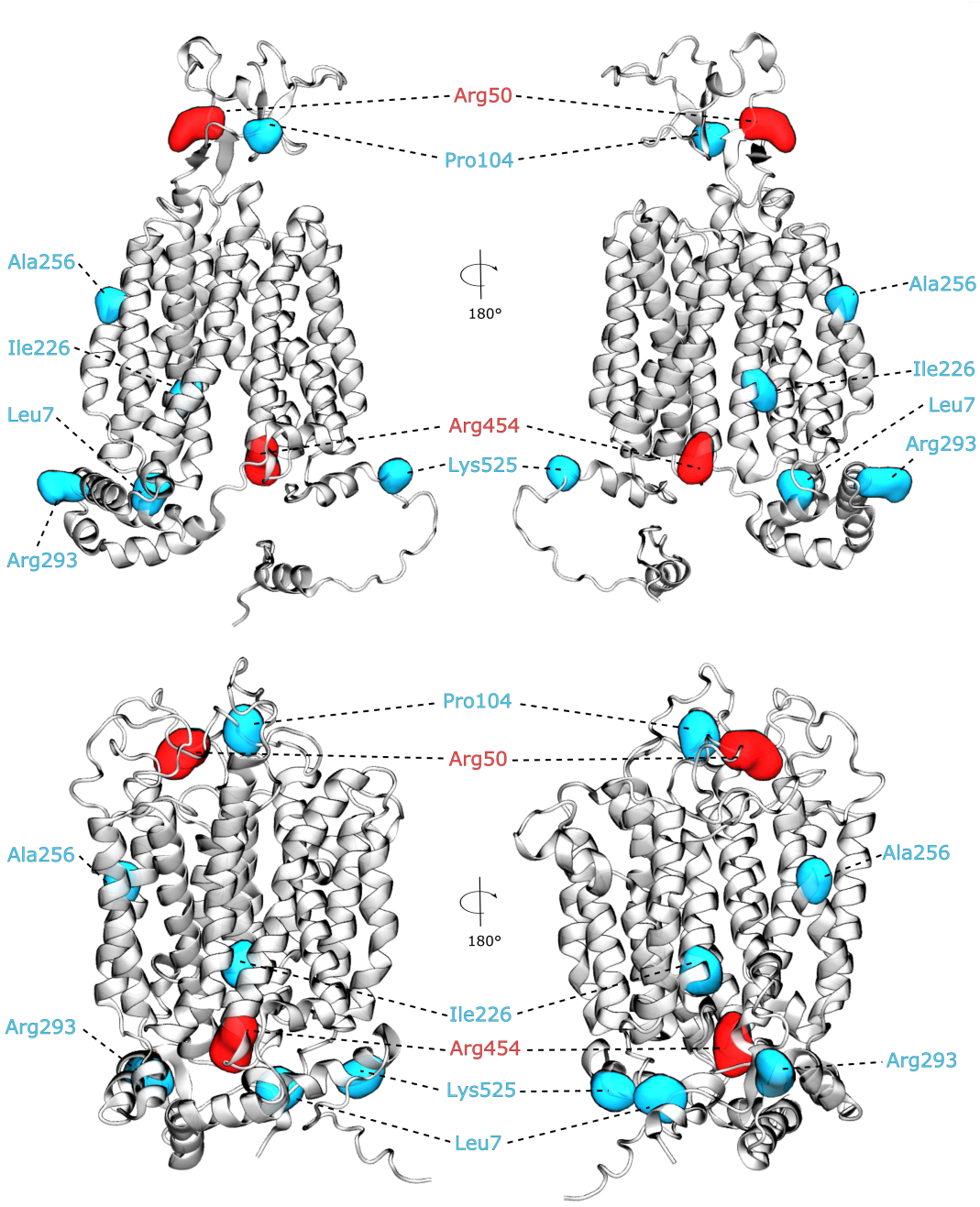


Figure 58. Mapped Single Nucleotide Polymorphisms on IF (top) and OF (bottom) conformations of *h*OAT1. Substitutions of Arg50 and Arg454 were shown to impact drug transport while no impact was reported for substitutions of Pro104, Ala256, Arg293, Ile226 and Leu7.

IV.4. Conclusion

The present study investigated substrate binding events of both substrate and co-substrate to the human *SLC22A6*/OAT1 at the atomic scale. Molecular docking calculations and μs-scaled MD simulations revealed two plausible binding spots for adefovir, consistent with experimental observations, in the B-like motif and inner binding pockets. A-helices (essentially TMH1, TMH4) residues were shown to play an essential role for substrate binding event. In the intracellular region, MD simulations also suggested the binding of αKG involved residues of the charge-relay system, located either within the motif triads (within A-motif, E[X_6]R and PETL) or at the interface between them. The presence of αKG may interfere with the salt-bridge network of intracellular conserved motifs. This is expected to favour the opening of IC gating which might be key in driving large-scale conformational changes required for *h*OAT1 alternated access. Besides, the allosteric pathways between substrate cavities and intracellular motifs were revealed on *h*OAT1 dynamic pictures obtained using MD simulations and highlight the central role of *e.g.*, TMH2, 3 and 4. Likewise, our simulations showed that the surrounding lipids play an active role in *h*OAT1 transport function. Non-synonymous SNPs were then mapped onto the *h*OAT1 structural model. Transport impairments experimentally observed for p.Arg50His and p.Arg454Gln were respectively attributed to (i) the decrease of substrate binding affinity or (ii) disrupted intracellular domain interactions owing to the weakening of electrostatic interactions. Each new SNP or site-direct mutagenesis mutant can now be implemented in the present model to elucidate its role at an atomistic resolution.

Altogether, the present study paves the way for the structural understanding of *h*OAT1 function in a pharmacological context. The present model can be used to understand and predict transporter-xenobiotic interactions as well as DDIs involving this membrane transporter. The allosteric pathways described may be involved in non-competitive DDIs, by which a *h*OAT1 inhibitor may interact with an allosteric site involved in the communication between key regions.

Supplementary Information

In Supplementary information are reported: (i) molecular docking and MD technical details; (ii) calculated H-bond and contact fractions; (iii) list of site-directed mutagenesis reviewed from the literature; (iv) Allosteric communication efficiencies for all systems. Data are available upon reasonable request.

IV.5. References

[1] A.L. VanWert, M.R. Gionfriddo, D.H. Sweet, Organic anion transporters: discovery, pharmacology, regulation and roles in pathophysiology, Biopharm. Drug Dispos. (2009) n/a-n/a. https://doi.org/10.1002/bdd.693.

[2] D.H. Sweet, Organic anion transporter (Slc22a) family members as mediators of toxicity, Toxicol. Appl. Pharmacol. 204 (2005) 198–215. https://doi.org/10.1016/j.taap.2004.10.016.

[3] Q. Faucher, H. Alarcan, P. Marquet, C. Barin-Le Guellec, Effects of Ischemia-Reperfusion on Tubular Cell Membrane Transporters and Consequences in Kidney Transplantation, J. Clin. Med. 9 (2020) 2610. https://doi.org/10.3390/jcm9082610.

[4] S.K. Nigam, What do drug transporters really do?, Nat. Rev. Drug Discov. 14 (2015) 29–44. https://doi.org/10.1038/nrd4461.

[5] R. Ho, R. Kim, Transporters and drug therapy: Implications for drug disposition and disease, Clin. Pharmacol. Ther. 78 (2005) 260–277. https://doi.org/10.1016/j.clpt.2005.05.011.

[6] The International Transporter Consortium, Membrane transporters in drug development, Nat. Rev. Drug Discov. 9 (2010) 215–236. https://doi.org/10.1038/nrd3028.

[7] M.J. Zamek-Gliszczynski, M.E. Taub, P.P. Chothe, X. Chu, K.M. Giacomini, R.B. Kim, A.S. Ray, S.L. Stocker, J.D. Unadkat, M.B. Wittwer, C. Xia, S.-W. Yee, L. Zhang, Y. Zhang, International Transporter Consortium, Transporters in Drug Development: 2018 ITC Recommendations for Transporters of Emerging Clinical Importance, Clin. Pharmacol. Ther. 104 (2018) 890–899. https://doi.org/10.1002/cpt.1112.

[8] K.M. Hillgren, D. Keppler, A.A. Zur, K.M. Giacomini, B. Stieger, C.E. Cass, L. Zhang, Emerging Transporters of Clinical Importance: An Update From the International Transporter Consortium, Clin. Pharmacol. Ther. 94 (2013) 52–63. https://doi.org/10.1038/clpt.2013.74.

[9] G. Burckhardt, B.C. Burckhardt, In Vitro and In Vivo Evidence of the Importance of Organic Anion Transporters (OATs) in Drug Therapy, in: M.F. Fromm, R.B. Kim (Eds.), Drug Transp., Springer Berlin Heidelberg, Berlin, Heidelberg, 2011: pp. 29–104. https://doi.org/10.1007/978-3-642-14541-4_2.

[10] H.C. Liu, N. Jamshidi, Y. Chen, S.A. Eraly, S.Y. Cho, V. Bhatnagar, W. Wu, K.T. Bush, R. Abagyan, B.O. Palsson, S.K. Nigam, An Organic Anion Transporter 1 (OAT1)-centered Metabolic Network, J. Biol. Chem. 291 (2016) 19474–19486. https://doi.org/10.1074/jbc.M116.745216.

[11] S.K. Nigam, K.T. Bush, G. Martovetsky, S.-Y. Ahn, H.C. Liu, E. Richard, V. Bhatnagar, W. Wu, The Organic Anion Transporter (OAT) Family: A Systems Biology Perspective, Physiol. Rev. 95 (2015) 83–123. https://doi.org/10.1152/physrev.00025.2013.

[12] M. Roth, A. Obaidat, B. Hagenbuch, OATPs, OATs and OCTs: the organic anion and cation transporters of the SLCO and SLC22A gene superfamilies: OATPs, OATs and OCTs, Br. J. Pharmacol. 165 (2012) 1260–1287. https://doi.org/10.1111/j.1476-5381.2011.01724.x.

[13] L. Wang, D.H. Sweet, Renal Organic Anion Transporters (SLC22 Family): Expression, Regulation, Roles in Toxicity, and Impact on Injury and Disease, AAPS J. 15 (2013) 53–69. https://doi.org/10.1208/s12248-012-9413-y.

[14] S.K. Nigam, W. Wu, K.T. Bush, M.P. Hoenig, R.C. Blantz, V. Bhatnagar, Handling of Drugs, Metabolites, and Uremic Toxins by Kidney Proximal Tubule Drug Transporters, Clin. J. Am. Soc. Nephrol. 10 (2015) 2039–2049. https://doi.org/10.2215/CJN.02440314.

[15] A. Janaszkiewicz, Á. Tóth, Q. Faucher, M. Martin, B. Chantemargue, C. Barin-Le Guellec, P. Marquet, F. Di Meo, Insights into the structure and function of the human organic anion

transporter 1 in lipid bilayer membranes, Sci. Rep. 12 (2022) 7057. https://doi.org/10.1038/s41598-022-10755-2.

[16] D. Drew, R.A. North, K. Nagarathinam, M. Tanabe, Structures and General Transport Mechanisms by the Major Facilitator Superfamily (MFS), Chem. Rev. 121 (2021) 5289–5335. https://doi.org/10.1021/acs.chemrev.0c00983.

[17] 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. https://doi.org/10.1038/nrm.2015.25.

[18] K. Tanaka, W. Xu, F. Zhou, G. You, Role of glycosylation in the organic anion transporter OAT1, J. Biol. Chem. 279 (2004) 14961–14966. https://doi.org/10.1074/jbc.M400197200.

[19] J. Zhang, H. Wang, Y. Fan, Z. Yu, G. You, Regulation of organic anion transporters: Role in physiology, pathophysiology, and drug elimination, Pharmacol. Ther. 217 (2021) 107647. https://doi.org/10.1016/j.pharmthera.2020.107647.

[20] C. Zhu, K.B. Nigam, R.C. Date, K.T. Bush, S.A. Springer, M.H. Saier, W. Wu, S.K. Nigam, Evolutionary Analysis and Classification of OATs, OCTs, OCTNs, and Other SLC22 Transporters: Structure-Function Implications and Analysis of Sequence Motifs, PLOS ONE. 10 (2015) e0140569. https://doi.org/10.1371/journal.pone.0140569.

[21] A. Gessner, J. König, M.F. Fromm, Clinical Aspects of Transporter-Mediated Drug–Drug Interactions, Clin. Pharmacol. Ther. 105 (2019) 1386–1394. https://doi.org/10.1002/cpt.1360.

[22] Y. Liang, S. Li, L. Chen, The physiological role of drug transporters, Protein Cell. 6 (2015) 334–350. https://doi.org/10.1007/s13238-015-0148-2.

[23] X. Huo, K. Liu, Renal organic anion transporters in drug–drug interactions and diseases, Eur. J. Pharm. Sci. 112 (2018) 8–19. https://doi.org/10.1016/j.ejps.2017.11.001.

[24] M. Shimizu, N. Furusyo, H. Ikezaki, E. Ogawa, T. Hayashi, T. Ihara, Y. Harada, K. Toyoda, M. Murata, J. Hayashi, Predictors of kidney tubular dysfunction induced by adefovir treatment for chronic hepatitis B, World J. Gastroenterol. 21 (2015) 2116–2123. https://doi.org/10.3748/wjg.v21.i7.2116.

[25] C.C. Wong, N.P. Botting, C. Orfila, N. Al-Maharik, G. Williamson, Flavonoid conjugates interact with organic anion transporters (OATs) and attenuate cytotoxicity of adefovir mediated by organic anion transporter 1 (OAT1/SLC22A6), Biochem. Pharmacol. 81 (2011) 942–949. https://doi.org/10.1016/j.bcp.2011.01.004.

[26] C. Martens, R.A. Stein, M. Masureel, A. Roth, S. Mishra, R. Dawaliby, A. Konijnenberg, F. Sobott, C. Govaerts, H.S. Mchaourab, Lipids modulate the conformational dynamics of a secondary multidrug transporter, Nat. Struct. Mol. Biol. 23 (2016) 744–751. https://doi.org/10.1038/nsmb.3262.

[27] C. Martens, M. Shekhar, A.J. Borysik, A.M. Lau, E. Reading, E. Tajkhorshid, P.J. Booth, A. Politis, Direct protein-lipid interactions shape the conformational landscape of secondary transporters, Nat. Commun. 9 (2018) 4151. https://doi.org/10.1038/s41467-018-06704-1.

[28] M.P. Muller, T. Jiang, C. Sun, M. Lihan, S. Pant, P. Mahinthichaichan, A. Trifan, E. Tajkhorshid, Characterization of Lipid–Protein Interactions and Lipid-Mediated Modulation of Membrane Protein Function through Molecular Simulation, Chem. Rev. 119 (2019) 6086–6161. https://doi.org/10.1021/acs.chemrev.8b00608.

[29] O. Trott, A.J. Olson, AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. (2009) NA-NA. https://doi.org/10.1002/jcc.21334.

[30] PUBCHEM, (n.d.). https://pubchem.ncbi.nlm.nih.gov.

[31] 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.V. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J.V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, Williams, 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. Throssell, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M.J. Bearpark, J.J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, T.A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A.P. 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 16 Rev. A.03, Wallingford, CT, 2016.

[32] J.L. Perry, N. Dembla-Rajpal, L.A. Hall, J.B. Pritchard, A Three-dimensional Model of Human Organic Anion Transporter 1: AROMATIC AMINO ACIDS REQUIRED FOR SUBSTRATE TRANSPORT, J. Biol. Chem. 281 (2006) 38071–38079. https://doi.org/10.1074/jbc.M608834200.

[33] A.N. Rizwan, W. Krick, G. Burckhardt, The chloride dependence of the human organic anion transporter 1 (hOAT1) is blunted by mutation of a single amino acid, J. Biol. Chem. 282 (2007) 13402–13409. https://doi.org/10.1074/jbc.M609849200.

[34] L. Zou, A. Stecula, A. Gupta, B. Prasad, H.-C. Chien, S.W. Yee, L. Wang, J.D. Unadkat, S.H. Stahl, K.S. Fenner, K.M. Giacomini, Molecular Mechanisms for Species Differences in Organic Anion Transporter 1, OAT1: Implications for Renal Drug Toxicity, Mol. Pharmacol. 94 (2018) 689–699. https://doi.org/10.1124/mol.117.111153.

[35] 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. https://doi.org/10.1021/acs.jctc.5b00255.

[36] 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. https://doi.org/10.1021/ct4010307.

[37] 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. https://doi.org/10.1063/1.445869.

[38] I.S. Joung, T.E. Cheatham, Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations, J. Phys. Chem. B. 112 (2008) 9020–9041. https://doi.org/10.1021/jp8001614.

[39] I.S. Joung, T.E. Cheatham, Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters, J. Phys. Chem. B. 113 (2009) 13279–13290. https://doi.org/10.1021/jp902584c.

[40] D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman, AMBER 2018, University of California, San Francisco, 2018.

[41] A.W. Götz, M.J. Williamson, D. Xu, D. Poole, S. Le Grand, R.C. Walker, Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born, J. Chem. Theory Comput. 8 (2012) 1542–1555. https://doi.org/10.1021/ct200909j.

[42] 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. https://doi.org/10.1063/1.464397.

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

[44] J. Åqvist, P. Wennerström, M. Nervall, S. Bjelic, B.O. Brandsdal, Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm, Chem. Phys. Lett. 384 (2004) 288–294. https://doi.org/10.1016/j.cplett.2003.12.039.

[45] 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. https://doi.org/10.1021/ct400341p.

[46] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

[47] A.M. Westerlund, O. Fleetwood, S. Pérez-Conesa, L. Delemotte, Network analysis reveals how lipids and other cofactors influence membrane protein allostery, J. Chem. Phys. 153 (2020) 141103. https://doi.org/10.1063/5.0020974.

[48] T. Cihlar, D.C. Lin, J.B. Pritchard, M.D. Fuller, D.B. Mendel, D.H. Sweet, The Antiviral Nucleotide Analogs Cidofovir and Adefovir Are Novel Substrates for Human and Rat Renal Organic Anion Transporter 1, Mol. Pharmacol. 56 (1999) 570–580. https://doi.org/10.1124/mol.56.3.570.

[49] T. Sekine, H. Miyazaki, H. Endou, Molecular physiology of renal organic anion transporters, Am. J. Physiol.-Ren. Physiol. 290 (2006) F251–F261. https://doi.org/10.1152/ajprenal.00439.2004.

[50] X.C. Zhang, Y. Zhao, J. Heng, D. Jiang, Energy coupling mechanisms of MFS transporters: Energy Coupling Mechanisms of MFS Transporters, Protein Sci. 24 (2015) 1560–1579. https://doi.org/10.1002/pro.2759.

[51] F. Di Meo, G. Fabre, K. Berka, T. Ossman, B. Chantemargue, M. Paloncýová, P. Marquet, M. Otyepka, P. Trouillas, In silico pharmacology: Drug membrane partitioning and crossing, Pharmacol. Res. 111 (2016) 471–486. https://doi.org/10.1016/j.phrs.2016.06.030.

[52] S. Li, Q. Zhang, G. You, Three ubiquitination sites of organic anion transporter-1 synergistically mediate protein kinase C-dependent endocytosis of the transporter, Mol. Pharmacol. 84 (2013) 139–146. https://doi.org/10.1124/mol.113.086769.

[53] M. Hong, F. Zhou, K. Lee, G. You, The putative transmembrane segment 7 of human organic anion transporter hOAT1 dictates transporter substrate binding and stability, J. Pharmacol. Exp. Ther. 320 (2007) 1209–1215. https://doi.org/10.1124/jpet.106.117663.

[54] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature. 596 (2021) 583–589. https://doi.org/10.1038/s41586-021-03819-2.

[55] V. Corradi, B.I. Sejdiu, H. Mesa-Galloso, H. Abdizadeh, S.Yu. Noskov, S.J. Marrink, D.P. Tieleman, Emerging Diversity in Lipid–Protein Interactions, Chem. Rev. 119 (2019) 5775–5848. https://doi.org/10.1021/acs.chemrev.8b00451.

[56] K. Kapoor, S. Pant, E. Tajkhorshid, Active Participation of Membrane Lipids in Inhibition of Multidrug Transporter P-Glycoprotein, BioRxiv. (2020) 2020.11.15.383794. https://doi.org/10.1101/2020.11.15.383794.

[57] D. Casares, P.V. Escribá, C.A. Rosselló, Membrane Lipid Composition: Effect on Membrane and Organelle Structure, Function and Compartmentalization and Therapeutic Avenues, Int. J. Mol. Sci. 20 (2019) 2167. https://doi.org/10.3390/ijms20092167.

[58] P.V. Escribá, X. Busquets, J. Inokuchi, G. Balogh, Z. Török, I. Horváth, J.L. Harwood, L. Vígh, Membrane lipid therapy: Modulation of the cell membrane composition and structure as a molecular base for drug discovery and new disease treatment, Prog. Lipid Res. 59 (2015) 38–53. https://doi.org/10.1016/j.plipres.2015.04.003.

[59] K. Bleasby, L.A. Hall, J.L. Perry, H.W. Mohrenweiser, J.B. Pritchard, Functional Consequences of Single Nucleotide Polymorphisms in the Human Organic Anion Transporter hOAT1 (*SLC22A6*), J. Pharmacol. Exp. Ther. 314 (2005) 923–931. https://doi.org/10.1124/jpet.105.084301.

[60] Z. Li, Current Updates in the Genetic Polymorphisms of Human Organic Anion Transporters (OATs), J. Pharmacogenomics Pharmacoproteomics. 03 (2012). https://doi.org/10.4172/2153-0645.1000e127.

[61] S.W. Yee, D.J. Brackman, E.A. Ennis, Y. Sugiyama, L.K. Kamdem, R. Blanchard, A. Galetin, L. Zhang, K.M. Giacomini, Influence of Transporter Polymorphisms on Drug Disposition and Response: A Perspective From the International Transporter Consortium, Clin. Pharmacol. Ther. 104 (2018) 803–817. https://doi.org/10.1002/cpt.1098.

[62] T. Fujita, C. Brown, E.J. Carlson, T. Taylor, M. de la Cruz, S.J. Johns, D. Stryke, M. Kawamoto, K. Fujita, R. Castro, C.-W. Chen, E.T. Lin, C.M. Brett, E.G. Burchard, T.E. Ferrin, C.C. Huang, M.K. Leabman, K.M. Giacomini, Functional analysis of polymorphisms in the organic anion transporter, SLC22A6 (OAT1):, Pharmacogenet. Genomics. 15 (2005) 201–209. https://doi.org/10.1097/01213011-200504000-00003.

[63] G.R. Buel, K.J. Walters, Can AlphaFold2 predict the impact of missense mutations on structure?, Nat. Struct. Mol. Biol. 29 (2022) 1–2. https://doi.org/10.1038/s41594-021-00714-2.

[64] R.A. McPherson, M.R. Pincus, eds., Henry's clinical diagnosis and management by laboratory methods, 23rd edition, Elsevier, St. Louis, Missouri, 2017.

[65] A. Emami Riedmaier, A.T. Nies, E. Schaeffeler, M. Schwab, Organic Anion Transporters and Their Implications in Pharmacotherapy, Pharmacol. Rev. 64 (2012) 421–449. https://doi.org/10.1124/pr.111.004614.

[66] G. Xu, V. Bhatnagar, G. Wen, B.A. Hamilton, S.A. Eraly, S.K. Nigam, Analyses of coding region polymorphisms in apical and basolateral human organic anion transporter (OAT) genes [OAT1 (NKT), OAT2, OAT3, OAT4, URAT (RST)], Kidney Int. 68 (2005) 1491–1499. https://doi.org/10.1111/j.1523-1755.2005.00612.x.

[67] I.M. da Rocha, A.S. Gasparotto, R.K. Lazzaretti, R.K. Notti, E. Sprinz, V.S. Mattevi, Polymorphisms associated with renal adverse effects of antiretroviral therapy in a Southern Brazilian HIV cohort, Pharmacogenet. Genomics. 25 (2015) 541–547. https://doi.org/10.1097/FPC.0000000000000169.

[68] J. Lowenstein, J.J. Grantham, The rebirth of interest in renal tubular function, Am. J. Physiol.-Ren. Physiol. 310 (2016) F1351–F1355. https://doi.org/10.1152/ajprenal.00055.2016.

[69] K.T. Bush, P. Singh, S.K. Nigam, Gut-derived uremic toxin handling in vivo requires OAT-mediated tubular secretion in chronic kidney disease, JCI Insight. 5 (2020) 133817. https://doi.org/10.1172/jci.insight.133817.

[70] T.T.G. Nieskens, J.G.P. Peters, M.J. Schreurs, N. Smits, R. Woestenenk, K. Jansen, T.K. van der Made, M. Röring, C. Hilgendorf, M.J. Wilmer, R. Masereeuw, A Human Renal Proximal Tubule Cell Line with Stable Organic Anion Transporter 1 and 3 Expression Predictive for Antiviral-Induced Toxicity, AAPS J. 18 (2016) 465–475. https://doi.org/10.1208/s12248-016-9871-8.

Chapter V. Membrane dependency on the structural dynamics of human major facilitator superfamily transporters

Angelika Janaszkiewicz[1], Ágota Tóth[1], Veronica Crespi[1], Florent Di Meo[1]

Note to reviewers:

The present manuscript is in preparation to be published. The Electronic Supplementary Information will be provided in a separate file.

[1] *Inserm U1248 Pharmacology & Transplantation, Univ. Limoges, 87000 Limoges, France*

V.1. Introduction

Membrane transporters play a central role in a myriad of biological events, including physiological and pharmacological processes. They regulate the translocation of small to medium-sized molecules across the cell membrane, either into or out of the cell [1]. Membrane transporters are typically divided into two main superfamilies, namely ATP-Binding Cassette (ABC) transporters and SoLute Carriers (SLC). In humans, more than 420 SLC transporters, divided into approximately 65 subfamilies, have been annotated so far [2], representing one of the largest membrane protein families [3]. They are involved in the translocation of a broad range of small molecules, including endogenous compounds (e.g., sugars, amino-acids, nucleotides, uremic toxins) as well as xenobiotics. ABC proteins typically depend on ATP hydrolysis, while SLC proteins rely on concentration gradients to drive downhill or uphill substrate transport [4]. The former is referred to as "facilitated diffusion" transport. The latter often requires the cotransport of another substrate either in the same direction (symporter) or in the opposite direction (antiporter) [4]. Over the past decades, research on the structures, dynamics, and functions of SLCs has been a very active field of research, taking advantage of joint biochemical, biophysical, and structural approaches supported by computational techniques. Among the 16 SLC families, the Major Facilitator Superfamily is one of the most significant superfamilies. The deep understanding of MFS function is of particular importance in pharmacology. Indeed, many MFS transporters are involved in xenobiotic membrane-crossing events, which are key to determining the local concentration in a given compartment or close to its target site, whether it is associated with therapeutic or adverse effects. In this context, MFS transporters located in the kidneys and liver are of particular importance since these organs are involved in the metabolism and elimination of most xenobiotics (e.g., members of the SLC22 and SLC21 families) [5–7]. This has been stressed out by the International Transporter Consortium, which has defined some of them as of "emerging clinical importance", e.g., OATP1B1, 1B3 (SLCO1B1, 1B3); OCT2 (SLC22A2); MATE1, 2-K, 2 (SLC47A1-2); OAT1, 3 (SLC22A6, 8) [8,9]. Even though in vitro functional studies were carried out to better understand their physiological and pharmacological roles, to date, none has been structurally resolved, precluding a robust atomistic and dynamic picture of their functions. Only a few human MFS transporter structures have been resolved so far, most of them belonging to the Sugar Porter (SP) family.

The typical fold of MFS consists in twelve transmembrane helices (TMH) divided into two bundles, namely N- (TMH1-6) and C-bundles (TMH7-12) [10]. Both N- and C-helices are connected by intracellular domains consisting of at least three intracellular helices (ICH). Interestingly, ICH and intracellular regions of TMHs were described to interact thanks to a strong H-bond and salt-bridge networks referred to as the "charge-relay system" [11,12]. MFS-mediated substrate translocation was widely described in the literature according to the "alternating-access", in which MFS fold undergoes large-scale conformational changes alternatively exposing the inner binding cavity to either extracellular or intracellular compartments. Recently, the free energy surface of a prototypical MFS transporter according to the "rocker-switch" alternating access mechanism" was reconstructed for the prototypical Glucose Transporter 5 (SLC2A5/GLUT5) by means of biased molecular dynamics simulations [13]. Such a robust approach provides robust clues about the main driving forces behind large-scale conformation changes, e.g., confirming the central roles of the TMH1-TMH7 and TMH4-TMH8 pairs for outward-facing (OF) and inward-facing (IF) conformations [14–16].

Over the past year, particular attention has been paid to the interplay between lipid and MFS proteins, showing that they are shown to be particularly sensitive to lipid bilayer environments.

For example, biophysical investigations based on Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS) highlighted the central role of PE lipids in the dynamics of the MFS transport cycle [17]. Likewise, as shown in other membrane proteins, surrounding lipid were also computationally suggested to actively play an allosteric role in signal transduction from substrate binding to charge-relay system. Computational studies often use models in which transporters are embedded in symmetric lipid bilayers, including at least phosphatidylcholine, phosphatidylethanolamine lipids, and cholesterol. However, it is well known that in situ lipid cell membranes are more complex asymmetric systems if considering only lipid composition.

Taking advantage of recent advances made for MFS transporters (namely GLUTs), we here propose to computationally investigate the impact of lipid bilayer composition on MFS transporters, focusing on *human* GLUT1, GLUT3 as prototypes, as well as human Organic Anion Transporter 1 (*h*OAT1). By means of long μs-scaled MD simulations, the impact of protein-lipid interactions was investigated, focusing on (i) the overall dynamics of chosen MFS transporters, (ii) substrate and inhibitor binding, and (ii) charge-relay system network. It is worth mentioning that *h*GLUT1 was used as a reference model since it is the only transporter considered here with almost all conformational states along the transport cycle, namely apo OF, substrate-bound OF, inhibitor-bound OF, occluded IF, and IF open.

V.2. Materials and methods

V.2.1. Overview of MFS transporter of interest

The initial structures of *human* glucose transporter 1 and 3 (*h*GLUT1 and *h*GLUT3) and *human* Organic Anion Transporter 1 (*h*OAT1) were obtained from PDB database and AlphaFold2 (AF2) [18] predicted structures database. *h*GLUT1 was used in initial conformations: IF_{open} (6THA [19]), $IF_{occluded}$ (5EQG [20]), and OF_{open} (AF2). The structures of *h*GLUT3 were used in OF_{open} (AF2, 4ZWC [21]) and $OF_{occluded}$ (4ZWB [21]) initial conformations. Whereas (*h*OAT1) was utilised in the IF conformation available in AF2 database. The PDB structures of *h*GLUT3 (4zwc and 4zwb) were mutated (T43N) back to the wild type. Special attention was paid to the protonation states of histidine, which were assigned at pH=7 using the propKa server [22] resulting in ε-protonated histidine for glucose transporters and *h*OAT1 δ-protonated (His47) and ε-protonated (His48, His34, His130, His217, His246, His249, His275, His337, His546) histidine. The AF2 structures of *h*GLUT1 and *h*GLUT3 were N and C truncated resulting in 8-456 and 2-471 sequence for *h*GLUT1 and *h*GLUT3, respectively, to match the sequence obtained from PDB database.

The substrates, D-α-maltose (alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose) and D-α-glucose were built and described using GLYCAM_06j-1force field [23]. By aligning 4YB9 [16] and 4ZWC [16], where D-α-glucose (glucose) and D-α-maltose (maltose), respectively are present in the PDB database, the molecules were inserted to OF structures of *h*GLUT1 (AF2) and *h*GLUT3 (4ZWC, 4ZWB [16]). The simulated systems are reported in Table S1. Proteins were aligned to OPM (5EQG) structures and embedded in the lipid bilayer membranes: POPC:POPE:Chol (2:1:1) and asymmetric membrane (composition reported in Table 2) using the webserver CHARMM-GUI membrane builder tool [24].

V.2.2. System preparation

Based on experimental and theoretical studies [25–28], the membrane composition and leaflet distribution of the asymmetric membrane was established. The membrane composition considers cholesterol (Chol), phosphatidic acid (PA), phosphatidylcholine (PC),

phosphatidylethanolamine (PE), phosphatidylserine (PS) and acyl groups such as palmityoyl (PA, 16:0), oleoyl (OL, 18:1), arachidonyl (AR, 20:4), docosahexaenoyl (DHA, 22:6). The number of lipids in the membrane was approximately 500, however it may slightly differ within systems to match the lateral leaflet tension. The composition of the asymmetric membrane is presented in Table S1. To mimic physiological conditions, all systems were dissolved in water and neutralised with 154 mM NaCl.

V.2.3. MD technical setup

The simulations were carried out using Amber20 [29] package on CPU and GPU code [30]. The system was described by FF14SB [31], Lipid17 [32], TIP3P [33] and GLYCAM_06j-1 [23] to model protein, lipids, water and substrates, respectively. The ions Na^+ and Cl^- were described by TIP3P-compatible parameters obtained from Joung and Cheatham [34,35].

The simulations were performed with applied periodic boundary conditions. The electrostatic and Lennard Jones potential described the non-covalent interactions within 10Å cut-off distance. The long-range interactions are treated using the Particle Mesh Ewald (PME) method [36]. The bonds of hydrogen atoms were restrained using the SHAKE algorithm in order to increase the integration time to 2 fs. The simulations were performed in constant temperature (310K) and under constant pressure maintained by Langevin thermostat [37] and Berendsen barostat [38], respectively, under semiisotropic conditions. All systems were initially equilibrated by minimising the positions of all atoms. The molecules of water were then thermalized smoothly from 0 to 100K in 200 ps under (N,V,T) conditions. The system was then heated to an additional 310K under semi-isotropic (N,P,T) conditions in which the pressure was governed using Berendsen barostat. The system boxes were then equilibrated for 5,5 ns. Systems apo were performed in 5 replicas, while the systems with a substrate were performed in 3 replicas, each one $2\mu s$-long, providing 42 μs for GLUT1 (IF_{open} $10\mu s$; $IF_{occluded}$ $10\mu s$; OF_{open} $10\mu s$; $OF_{glucose}$ $6\mu s$; $OF_{maltose}$ $6\mu s$), 54 μs for GLUT3 (OF_{open} $20\mu s$; $OF_{ocludded}$ $10\mu s$; $OF_{glucose}$ $12\mu s$; $OF_{maltose}$ $12\mu s$) and $10\mu s$ for *h*OAT1 per membrane, giving 212 μs in total (Table S1) .

V.2.4. Analysis

PyTRAJ and CPPTRAJ AMBER modules [39], as well as VMD [40] and custom Python scripts, were used to conduct the analyses. The evolution of time-dependent backbone root-mean squared deviations was used to monitor the system equilibration resulting in of 1.5 μs long equilibrated trajectories, on which the analyses were performed (Fig S1). To ensure populated state, the Principal Component Analysis (PCA) was performed on the equilibrated trajectories of systems listed in Table S1 using so-called MFS core (transmembrane domain) using backbone masks listed in Table S3. The PCA was projected onto the conformational map of resolved structures of MFS proteins (section IV.2.4) representing OF/IF conformational state (Fig. S2). The equibilibrated MD simulations of IF_{open} (6THA); $IF_{occluded}$ (5EQG); OF_{open} (AF2); $OF_{glucose}$ (AF2); $OF_{maltose}$ (AF2) of *h*GLUT1 were used to discover the variability of conformations along the transport cycle by PCA employing the mask of backbone residues of *h*GLUT1. The PCA of *h*GLUT3 was projected on *h*GLUT1 conformational map (Fig. S3, S4). InfleCS clustering was performed to identify the different subspaces sampled during MD trajectories.

A new approach, inflection core state (InfleCS) clustering, has been used in our study to investigate subpopulations of the system. The InfleCS approach clusters the free energy minima, the so-called metastable core states, and estimates the relative energy barriers of

transitions between them. The advantage that distinguishes InfleCS is the use of the gaussian mixture free energy estimator, which provides a density landscape for the ensemble. The predicted density landscape (the shape of the estimated Gaussian mixture density) is used to identify clusters as the density peaks, which refer to the metastable core states. The number of density peaks refers to the number of clusters. Since the Hessian density reflects the curvature of the landscape, it is calculated at each point assigning each point as a core state or a transition state. To accurately define core states, their boundaries are identified by second-order derivatives of Gaussian mixture density. The clusters are naturally shaped by building graphs. Finally, the methods allow for free energy estimation that helps to understand the foundations of mechanisms by visualising a possible pathway from one metastable state to another [41]. InfleCS analysis was performed using the eigenvalues obtained from PCA of *h*GLU1 and *h*GLUT3 (Fig. S5). Clustering was achieved by using a grid size of 80×80, 5 iterations and from 2 to 16 gaussian functions for GMM. For further details about this method, see Ref [41].

V.3. Results and Discussion

V.3.1. Slight Membrane-dependency of *h*GLUT1 dynamics along transport cycle states.

Five *h*GLUT1 conformations and states were initially considered per membrane in total, namely two IF structures (from PDB IDs 5EQG [20] and 6THA [19]) and three OF structures from AF2 (namely apo, substrate-bound and inhibitor-bound). Along MD simulations, monitoring time-dependent root-mean square deviations (RMSD) suggest stable structure regardless of lipid bilayer membrane composition (Fig. S1). MFS core exhibited significantly lower structural variability than the whole protein in agreement with the expected flexibility of extra- and intracellular domains and loops. This was confirmed by monitoring per-domain root mean square fluctuations (RMSD) which stressed out the larger flexibility of TMH1-to-TMH2 extracellular loop as well as ICHs (Fig. S6). RMSF Differences between POPC:POPE:Chol (2:1:1) and asymmetric membranes were also calculated (Fig. S7). Globally, membrane composition seems not to strongly affect the overall flexibility of the protein whatever the conformation state. In spite of relatively long MD simulations (2μs) and the use of 3 replica for each substrate-bound system, no substrate translocation event was unfortunately observed, regardless of the membrane composition.

When comparing all system MFS cores with the same reference (namely initial IF *h*GLUT1 occluded structure), MD simulations highlighted the expected structural variabilities between all IF and OF states. However, RMSD analyses struggled to decipher structural differences between IF or OF sub-conformations, *i.e.*, occluded or open, substrate or inhibitor bound states (Fig. S1). Likewise, such analyses were unable to show differences between system embedded in POPC:POPE:Chol (2:1:1) and asymmetric membranes.

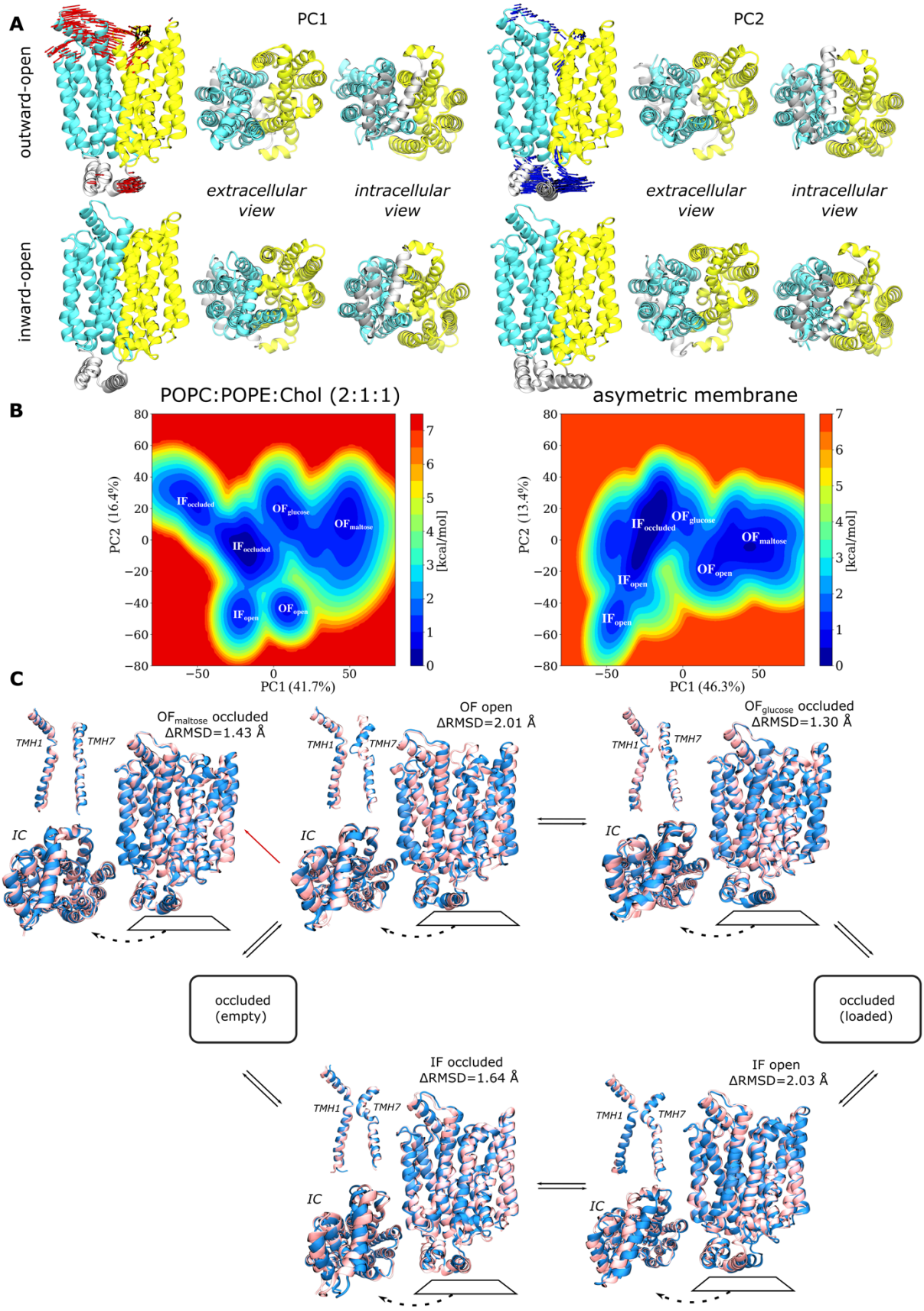


Figure 59. *h*GLUT1's structural variability. A) Principal elements are displayed. B) The PCA analysis of symmetric and asymmetric membrane conformational landscapes revealed representative structures

(C) of the transport cycle. Representative structures of symmetric (blue) and asymmetric (pink) membrane simulation have been aligned with their corresponding structures.

Therefore, inter-conformation principal component analyses (PCA) were performed by considering the whole protein backbone for each membrane separately, each conformation and state being equally represented in the whole dataset. The first two PCs explained ca. 60% of the overall structural variability (58.1% and 59.7, respectively, see Fig. 59, S3, S4). Regardless of membrane composition, PC1 and PC2 of each system match with the main structural fingerprints of IF and OF conformation, i.e., EC and IC opening/closing motions. Calculated PC1 for symmetric and asymmetric are very similar, overlap being 0.90 (See Fig. S8). They respectively explained 41.7 and 46.3% of the structural variability. PC1 might be mostly associated to the EC closing/opening of *h*GLUT1. In this event, per-residue contributions to PC1 highlighted the central role of TMH1 and the continuous extracellular helix between TMH1 and TMH2, sometimes called TMH1a and b. They are involved in 14.9 to 17% of PC1, for symmetric and asymmetric membranes, respectively). Likewise, ICHs and TMH8 are also strongly involved in PC1 modes (more than from 13.7 to 15.1% of PC1). In line with what has been proposed for MFS transport cycle, PCA revealed that the *h*GLUT1 rocker-switch motion is asymmetric. N-bundle is more involved in the structural variability than C-bundle (e.g., respectively 42.6% vs 33% in asymmetric membrane, see Figure S9). It is important to note that ICH motion is also coupled to N-bundle motion (e.g., 13.7% in asymmetric membrane). The second principal component PC2 might be associated to the intracellular closing *h*GLUT1 transport in POPC:POPE:Chol (2:1:1) and asymmetric membranes. Interestingly, residue contributions exhibit larger differences between membranes than for PC1. For instance, ICHs are significantly more involved in asymmetric membrane than in POPC:POPE:Chol (2:1:1) (35.8% vs 19.1%, respectively). This might denote a more important *h*GLUT1 structural variability in asymmetric membrane than in symmetric POPC:POPE:Chol.

To better rationalize structural transitions between all considered states, PC-based clustering was carried out using the InfleCS framework [41] (See Fig. 59B and S5). It is important to note that in the present work, the actual free energy surface along transport cycle cannot be captured since the sampling was not sufficient and no substrate translocation events were observed. However, InfleCS clustering can be used to provide atomic pictures of shared structural patterns between the adjacent conformational states according to their structural variability (See Fig. 59C). Regardless of the membrane composition, projections of all conformations and state are relatively consistent with expected transitions along MFS transport cycle. Each state and conformations are well-defined by local minima and cluster centres. These representative snapshots were compared according to membrane composition (RMSD ranging from 1.30 to 2.03 Å). The largest differences were observed for IF and OF open conformations. For OF_{open} conformation, the main difference was observed for the EC gating pictured by closer TMH1-TMH7 in POPC:POPE:Chol (2:1:1) membrane. Contacts between TMH1 and TMH7 were shown to be essential for sugar binding and transport in several many MFS including GLUTs (e.g., GLUT5 [13], and in GLUT1 [42]) as well as distant relative *h*OAT1 [11]. On the other hand, IF_{open} structures mostly differ in the ICH arrangements suggesting subtle membrane impact onto the intracellular charge-relay system, for which salt bridge network were shown to be associated with large scale conformational transitions between IF and OF conformations.

Interestingly, MD simulations carried out in asymmetric membrane tend to populate the conformation space of an *h*GLUT1 IF occluded state (see Fig. 59B and S5). Interestingly, IF_{open} and, to a lesser extent, OF_{open} *h*GLUT1 MD simulations mostly populate this region of the

conformational space. The former tends to close IC while the latter showed one closed EC gate subpopulation. Given that PC1 and PC2 mostly described IC and EC closing motions, it is likely that the present occluded space gathers both $IF_{occluded}$ and $OF_{occluded}$ conformations, for which, if they differ the present approach cannot distinguish one from the other. This anyway may suggest that, in absence of substrate, occluded *h*GLUT1 states might be the most predominant. It is important to note that occluded GLUT states were spontaneously resolved by means of cryo-EM structures [20,21,43–45].

In POPC:POPE:Chol (2:1:1) membrane, there exist a subpopulation from simulations starting with OF_{open} conformation which exhibits a small overlap with other OF conformations (i.e., substrate and inhibitor bound OF conformations) as well as with occluded region. Interestingly, this OF_{open} subpopulation exhibit similar local arrangement of the IC domains pictured by PC2 with IF open. On the other hand, EC opening is different as pictured by PC1. In asymmetric membrane simulations, these populations are more distant. This may suggest that symmetric membrane may uncorrelated IC closing with EC opening as compared to asymmetric membrane.

Projections of MD simulations onto the PCA conformational space in asymmetric membrane also reveal that transition from OF_{open} to occluded conformation is likely trigger by substrate-binding event. Indeed, MD simulations also exhibited a subpopulation of stable wide OF_{open} conformation. This subpopulation is distant from the aforementioned occluded subpopulation starting from the same initial OF_{open} conformation and for which only few snapshots were projected between them. Lipid membrane composition might alter substrate and inhibitor binding event in GLUTs

In asymmetric membrane, the conformational space separating OF_{open} and occluded states is mostly populated by OF glucose-bound conformations (see Fig. 59 and S5). This suggests that glucose binding event might favour EC closing event as picture by closer contact between TMH1 and TMH7 (see Figure 59C). MD simulations were thoroughly analysed in order to decipher glucose binding modes. Glucose binding to *h*GLUT1 is driven by strong H-bond network between polar residues and the numerous glucose OH moieties. Such network is highly dynamics as pictured by H-bond fractions along MD trajectories. In the binding pocket, key residues were identified, in agreement with what has been previously described in GLUTs [21,21,43]. For instance, the key residue Asn288 (*h*GLUT1) was observed in our simulations while it has been ascribed as a pivotal residue for substrate binding and substrate translocation.

Similar H-bond networks between *h*GLUT1 residues and glucose were observed in POPC:POPE:Chol (2:1:1) and asymmetric membranes. In term of H-bond fractions, the following sequence was obtained namely Gln161 ≈ Asn288 > Gln283 ≈ Gln282 > Glu380. MD simulations interestingly suggests that H-bond network is weaker in asymmetric membrane than in POPC:POPE:Chol (2:1:1) membrane (total H-bond fractions being 4.9 and 3.6 per glucose molecule, respectively). This may be due to biophysical properties of surrounding lipid bilayer which may apply lower restraints to *h*GLUT1 in asymmetric membrane than in POPC:POPE:Chol (2:1:1). Our MD simulations are in agreement with the known role of substrate binding in MFS transport cycles. Indeed, glucose binding systematically led to increased contacts between TMH1 and TMH7 which is associated to EC gating event (Fig. 2B&C). This is slightly more pronounced in POPC:POPE:Chol (2:1:1) membrane in which more contacts were observed than in asymmetric membrane likely due to different membrane properties. It is worth mentioning that EC gating was also associated with increased TMH2-

TMH11 contact. The same trend as observed for TMH1-TMH7 pair was observed for TMH2-TMH11 one. Glucose-bound *h*GLUT1 simulations performed in POPC:POPE:Chol (2:1:1) membrane exhibit more contacts between EC regions of TMH2 and TMH11 as compared to asymmetric membrane.

MD simulations were also carried out considering a known *h*GLUT1 inhibitor, namely maltose. Interestingly, in symmetric and asymmetric membrane, maltose-bound *h*GLUT1 state systematically sample a conformational space distant from occluded GLUT conformations (Fig. 1B and S5). PC1 tends to significantly increase suggesting a so-called wide-open OF conformation in which TMH1-TMH7 occlusion is unlikely. This was also confirmed by assessing TMH1-TMH7 contact map in which very few contacts were observed in contrast to glucose binding (Fig. 60). Maltose binding mode is similar to glucose, same residues but Asn288 were observed to be involved in H-bond network between *h*GLUT1 and maltose. However, sum of H-bond fractions are significantly lower in maltose binding than in glucose (ca. 2.2 in POPC:POPE:Chol (2:1:1) and asymmetric membrane, see Table 5. Considering (i) the lower sugar binding and (ii) the absence of TMH1-TMH7, maltose binding is expected not to trigger conformational changes required for OF-to-IF transition. This was confirmed by comparing per-residue dynamic cross correlation matrices (DCCM) between apo, glucose-bound and maltose-bound state for *h*GLUT1 (see Figure S10). In presence of maltose, overall cross-correlation between different domains of *h*GLUT1 exhibit similar pattern with apo state, while glucose-bound system increased cross-correlation along the protein. This is particularly true for the communication between intracellular domain and N-bundle. This is in line with former studies which have shown [13,14] that the IC gate-opening event is coupled with substrate binding.

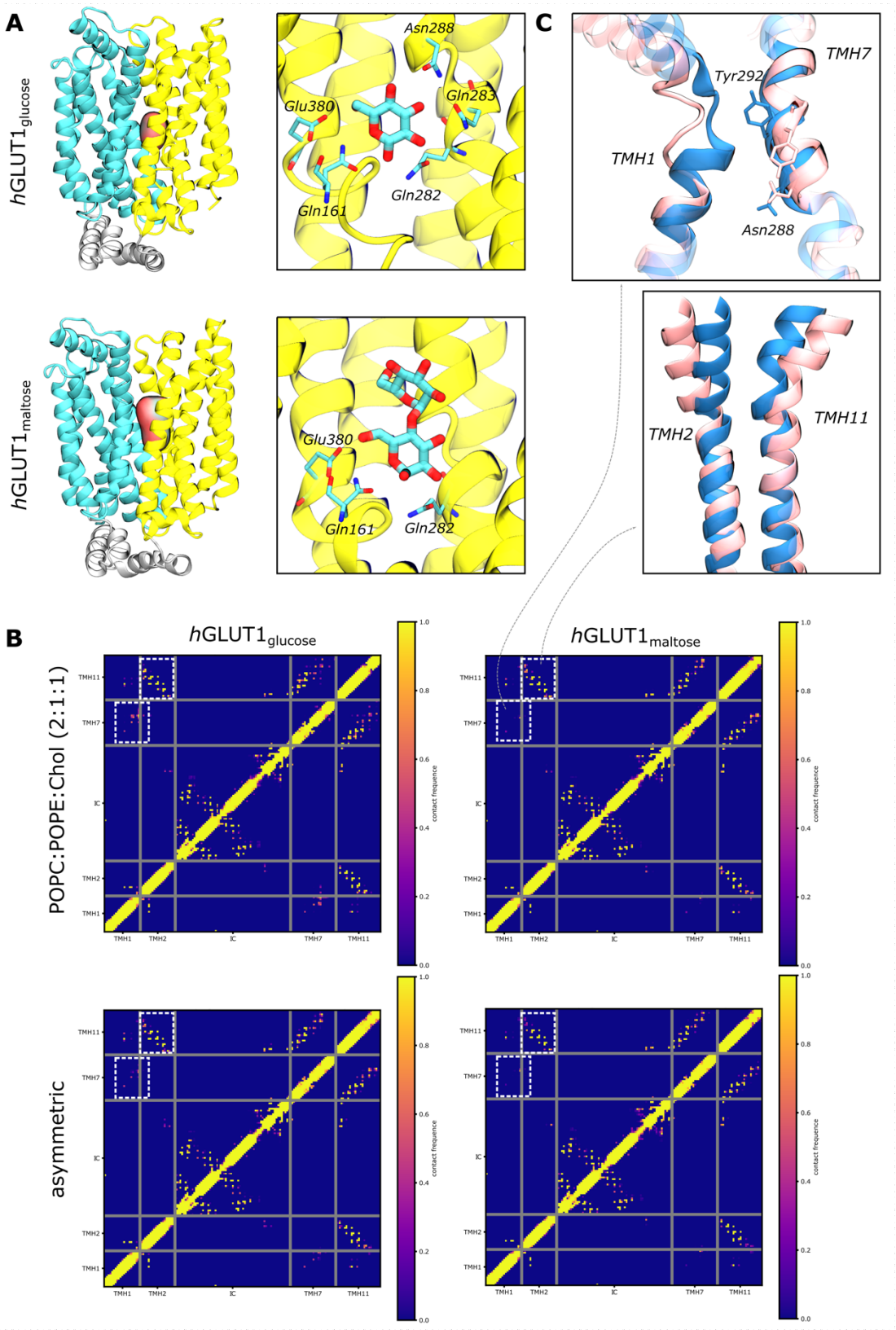


Figure 60. The comparison of *h*GLUT1 structures in the presence of maltose and glucose (substrate) in the OF conformation (inhibitor). (A) The presence of maltose sterically prevents the occlusion due to the second sugar moiety, stabilising the OF conformation in the open state, whereas the presence of glucose tends to lead to the occluded state (closed conformation). The interaction of glucose and maltose differ mainly in Asn288, which has been suggested to be responsible for occlusion. (B) Stripped contact maps of *h*GLUT1 with glucose and maltose in symmetric and asymmetric membranes reveal

significant differences in contact, primarily in (C) TMH7 and TMH1 (top) and TMH11 and TMH2 (bottom).

It is worth mentioning that DCCM maps significantly differ between POPC:POPE:Chol (2:1:1) and asymmetric membrane. In asymmetric membrane, cross-correlation between the different *h*GLUT1 domains (e.g., N-bundle, C-bundle and ICHs) is clearly well defined in presence of glucose. In contrast, in presence of maltose, the absence of cross-correlation between ICH and TMHs is more pronounced than in POPC:POPE:Chol (2:1:1). This strongly suggest a central role of membrane composition in the overall dynamics of *h*GLUT1 including distant communications between subdomains.

Table 5. Hydrogen bond fractions of protein-sugar interactions comparing the membranes in (a) *h*GLUT1 (b) *h*GLUT3.

(a)

Residues	Glucose		Maltose	
	POPC:POPE:Chol (2:1:1)	Asymmetric	POPC:POPE:Chol (2:1:1)	Asymmetric
Gln161	1.43	1.19	0.97	1.00
Gln282	0.85	0.73	0.41	0.42
Gln283	0.71	0.57	0.57	0.34
Asn288	1.41	0.91	-	-
Glu380	0.52	0.20	0.25	0.46

(b)

Residues	Glucose		Maltose	
	POPC:POPE:Chol (2:1:1)	Asymmetric	POPC:POPE:Chol (2:1:1)	Asymmetric
Gln159	1.44	1.55	1.26	1.32
Gln280	0.76	0.60	0.67	0.80
Gln281	0.36	0.57	0.49	0.58
Asn286	0.60	0.57	1.01	0.97
Glu378	0.52	0.69	0.50	0.40
Trp386	0.52	0.25	0.24	0.17

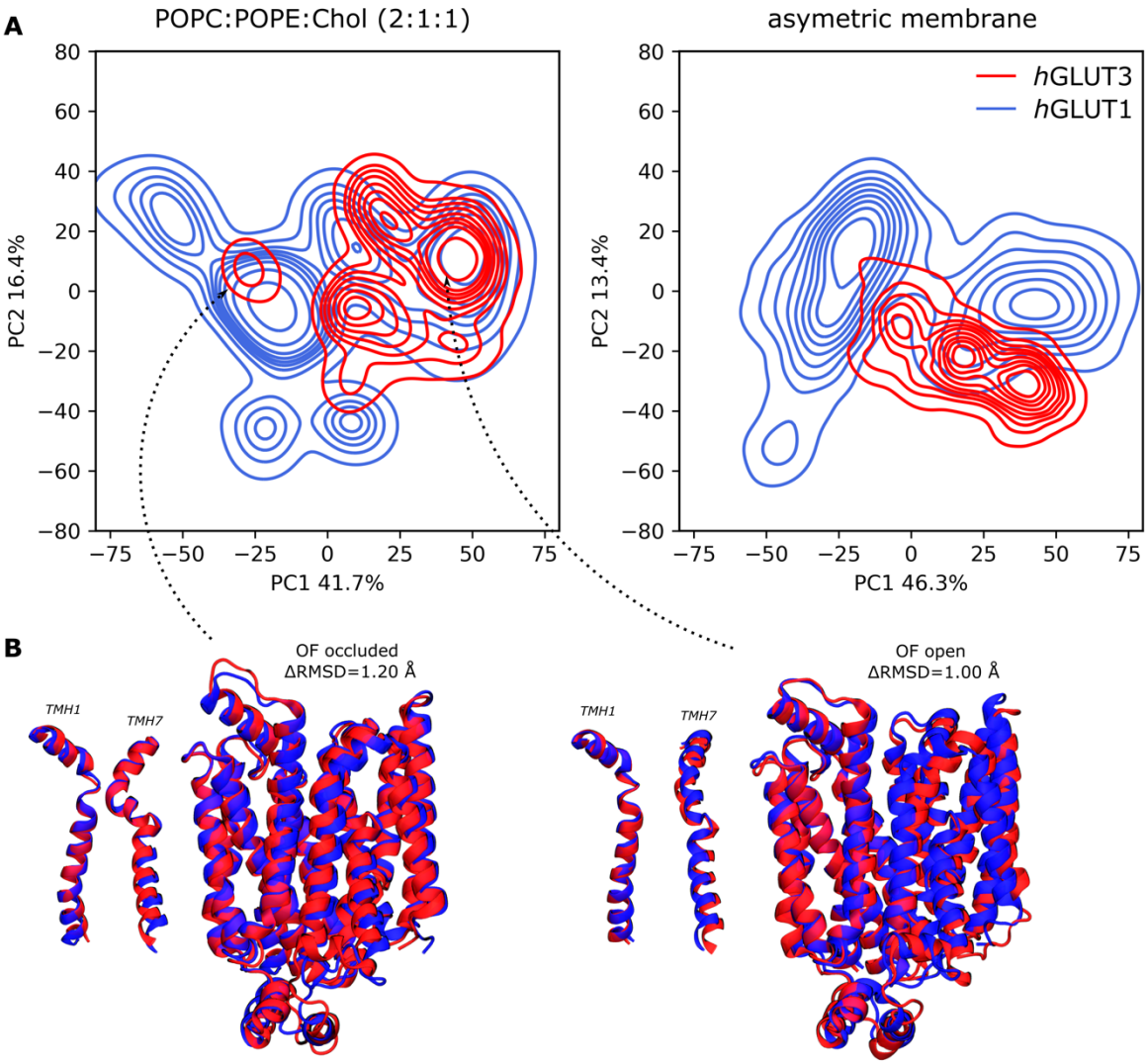


Figure 61. (A) The PCA of *h*GLUT3 projected onto the *h*GLUT1 in symmetric (left) and asymmetric (right) membrane. (B) The OF (apo) occluded conformation (left) is the least populated while the most populated is the OF (apo) open conformation. The respective conformations of *h*GLUT1 (blue) and *h*GLUT3 (red) are aligned.

MD simulations were also carried out considering *h*GLUT3 adopting OF conformation and in apo, glucose- and maltose-bound states. As observed for *h*GLUT1, projection of maltose-bound *h*GLUT3 onto the *h*GLUT1 PC-based conformational space is associated with open OF conformation contrary to glucose-bound and apo state (Fig. 61 and S4). Likewise, glucose-bound MD simulations tends to populate occluding conformations, therefore suggesting the EC gating event upon substrate binding. Glucose binding modes are relatively similar with *h*GLUT1, i.e., mostly driven by polar residues in the binding cavity, namely, Gln159, Gln280, Gln281, Asn286, Glu378 as well as Trp386. However, in *h*GLUT3, glucose binding mode seems not to be affected by membrane composition, overall fractions being almost identical in POPC:POPE:Chol (2:1:1) and asymmetric membranes (i.e., 4.2, see Table 4B). Likewise, membrane difference in terms of DCCM are less obvious (see Fig. S10). Very similar patterns were observed for POPC:POPE:Chol (2:1:1) and asymmetric membrane. We can thus hypothesize that *h*GLUT3 is less sensitive to membrane composition than *h*GLUT1. This may be associated with tissue distribution of these transporters. *h*GLUT1 is specific to the blood brain barrier, being responsible for glucose uptake at the basolateral membrane of epithelial

cells [46]. On the other hand, *h*GLUT3 is mostly distributed in brain, likely in different regions of neurons. We may thus assume that the lower sensitivity of *h*GLUT3 to membrane composition might be associated with maintaining glucose transporter activity regardless of the membrane composition.

V.3.2. On the importance of the charge-relay system in GLUTs

GLUTs have conserved structural patterns and motifs in the IC domains, maintaining a tight network [10,47]. The so-called A-motif, E[X6]R and PETL, are present in both N- and C-bundle, distributed in a pseudo-symmetric manner (Fig. S11). Charge-relay system is expected to strongly differ between MFS transporter using co-substrate with others. Therefore, for sake of comparison, MD simulations carried out on the *h*OAT1 transporter adopting IF conformations were here also considered. *h*OAT1 transport cycle is antiport, which requires the efflux of a-ketoglutarate to reset OF conformation. Recent investigations suggested that the binding of α-ketoglutarate in the IC domain of OF *h*OAT1 may break interbundle charge relay system to favour *h*OAT1 substrate release [11].

In each bundle, there exists a salt-bridge network between the three motifs (Table 5) maintaining them into a triad. For *h*GLUT1, *h*GLUT3 and *h*OAT1, inter-motif distances were monitored along MD simulations for each lipid bilayer membrane and each stat. Membrane composition seems to only slightly modulate the local arrangement of each bundle, given that only small shifts are observed in A-motif-E[X6]R, A-motif-PETL and E[X6]R- PETL distance distributions, (see Fig. S12). This is in agreement with aforementioned RMSF calculations which showed that per-residue flexibility in this region was not strongly affected by membrane composition. Particular attention was paid to the interaction between N- and C-E[X6]R motifs which was shown to fingerprint MFS conformations along the transport cycle of *h*OAT1[11] as well as other transporters. Our MD simulations are in agreement with this hypothesis since, in *h*GLUT1, N- and C-E[X6]R distances tend to be smaller in OF than IF conformation. Interestingly, IF conformations shown larger variabilities in line with expected flexible gating event. However, it is important to note that these distances are relatively small suggesting that IC gate may spontaneously close after substrate release. Interestingly, *h*GLUT1 and *h*GLUT3 exhibited a different membrane dependence regarding the distance between N- and C-E[X6]R motifs. While membrane composition does not show significant difference in *h*GLUT3 MD simulations, *h*GLUT1 shows more variability in POPC:POPE:Chol (2:1:1) membrane than in asymmetric membrane. Interestingly, MD simulations performed using IF conformations suggest that the spontaneous closing of IC gate is more likely than in POPC:POPE:Chol (2:1:1). However, this must be considered carefully since no simulation was performed considering IF state for *h*GLUT3. It is anyway worth mentioning that *h*GLUT3 and *h*GLUT3 exhibited different behaviours regarding interbundle distances for OF conformations. While *h*GLUT3 IC gate remains strictly close, regardless of the membrane composition, *h*GLUT1 exhibits more IC opening event in POPC:POPE:Chol (2:1:1) while it tends to remain close in asymmetric membrane. IC gating seems to be driven by double salt bridge interaction between glutamate and arginine residues of N- and C-bundle E[X6]R motifs (Fig. 62). Our MD simulations suggest that this double interaction is asymmetric. The interactions between the N-bundle E[X6]R glutamate residue with the C-bundle E[X6]R arginine residue (e.g., Glu144 and Arg398 in hGLUT1) is weaker than the one between C-bundle E[X6]R glutamate residue with the C-bundle E[X6]R arginine residue (Fig. 62). Furthermore, the former does not show significant difference between the different IF and OF conformations, in contrast to the latter.

We can here hypothesize that substrate release may require to only break one interaction rather than both, decreasing in turn the energy barrier.

To a lesser extent, *h*OAT1 seems to also be sensitive to membrane composition as *h*GLUT1. Indeed, IC spontaneous closing seems to be favoured in asymmetric membrane as compared to symmetric POPC:POPE:Chol (2:1:1). This is in line with aforementioned hypothesis regarding transporter cell distribution, *h*OAT1 being located exclusively at the basolateral membrane of kidney proximal tubular cells. Furthermore, in contrast to *h*GLUT1, almost no salt bridge were observed between N- and C-bundle E[X6]R motifs in POPC:POPE:Chol (2:1:1) or asymmetric membranes (calculated total H-bond fraction being respectively, 0.03 and 0.03). Therefore, it might mean that the proper IC closing to reset *h*OAT1 to OF conformation may require the transport of *h*OAT1 co-substrate.

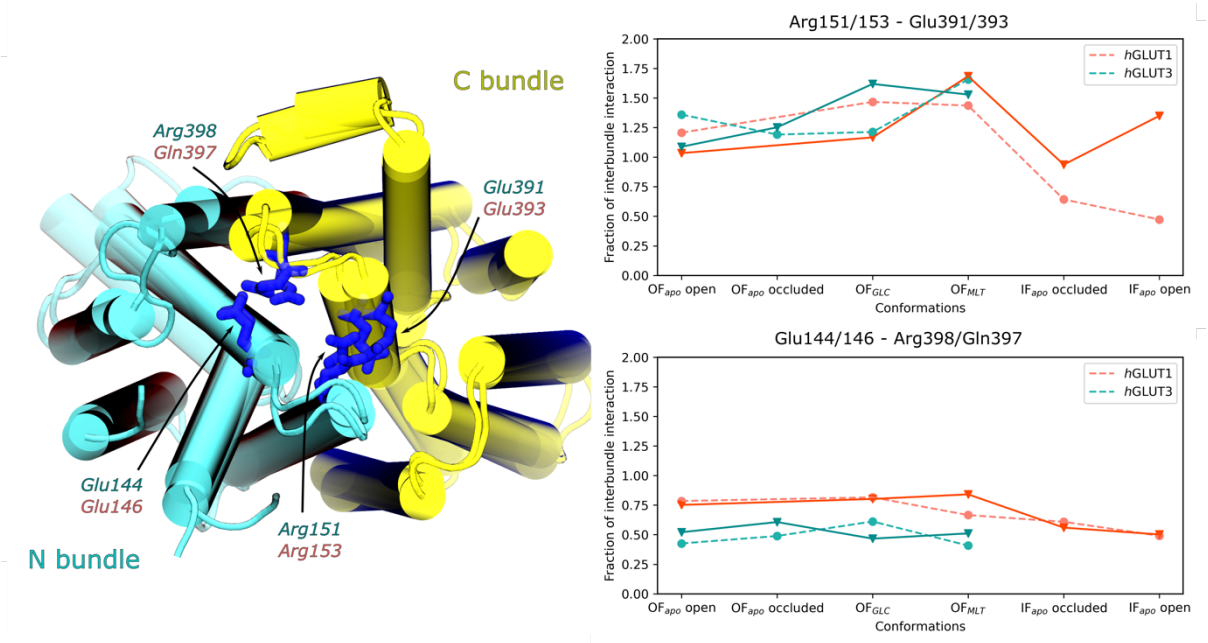


Figure 62. The interbundle interactions that is getting formed and broken along the transport cycle. On the right are presented the fraction from each simulation in symmetric (dashed line) and asymmetric (regular line) for *h*GLUT1 (orange) and *h*GLUT3 (cyan).

Table 6. The intracellular motifs that form the charge-relay system in *h*GLUT1, *h*GLUT3 and *h*OAT1.

	N-bundle			C-bundle		
	A-motif G[X3]D[R/K]XGR[R/K]	E[X6]RG	„PETL" [P/X]ESXRW[L/X]	A-motif [E/D]RAGRR	E[X6]R	„PETL" PET[X]G
*h*GLUT1	GLFVNRFGRR	EVSPTALRG	PESPRFL	ERAGRR	ELFSQGPR	PETRG
	G84-R93	E146-G154	P208-L214	E329-R335	E393-R400	P453-G457
*h*GLUT3	GLFVNRFGRR	EISPTALRG	PESPRFL	ERAGRR	ELFSQGPR	PETKG
	G82-R91	E144-G152	P206-L212	E327-R332	E391-R398	P451-G455
*h*OAT1	GYLADRLGRR	ETMPIHTRA	IESARWH	NSLGRR	ELYPTMIR	PETLG
	G153-R162	E212-A220	I269-H275	N390-R395	E447-R454	P505-G509

V.3.3. On the interplay between MFS transporters and bilayer membranes

Membrane properties were also investigated in order to better understand the slight structural differences observed in POPC:POPE:Chol (2:1:1) and asymmetric membranes. Overall membrane thicknesses were approximated by measuring distances between upper and lower leaflet P-atoms (Fig. 63 and S13). As expected, no difference were observed between POPC:POPC:Chol (2:1:1) and asymmetric membrane thicknesses (ca. 42 Å). This may be easily explained because (i) PC and PE lipids as well as Chol are the main polar hear components of asymmetric membranes (ca. 0.33:0.22:0.38) as well as palmityl and oleyl lipid tails are present for 80% of phospholipids used in the asymmetric membrane model. However, our MD simulations that lipids tails affect the membrane fluidity. Indeed, lipid tail order were also calculated and shown in ESI (Fig. 63 and S14). For sake of comparison, focus was paid only to palmityl and oleyl tails since they are the only lipid tails present in POPC:POPE:Chol (2:1:1) membrane. Interestingly, Lipid order profile are systematically lower in POPC:POPE:Chol (2:1:1) membrane for all MFS transporters (i.e., *h*GLUT1, *h*GLUT3 and *h*OAT1) regardless of their conformations. This suggests that asymmetric membrane is slightly less fluid than POPC:POPE:Chol (2:1:1) which may then modulate MFS transporter function.

Finally, direct interaction between surrounding lipid and MFS transporters were monitored. Protein-lipid interactions were shown to be highly dynamics, since one lipid can exchange with each other. Therefore, lipid distribution occupancies were calculated in order to assess the probability that a given lipid is located around the protein (Fig. 63). *h*GLUT1 and *h*GLUT3 clearly exhibit similar patterns in term of lipid distribution. In symmetric POPC:POPE:Chol (2:1:1), MD simulations tend to overestimate lipid-protein interactions with PE lipids in upper leaflet. Indeed, several PE hotspots were observed for almost all conformations of *h*GLUT1 and *h*GLUT3 while they are not observed in asymmetric membrane. This is consistent with in situ membrane composition since PE lipids tends to be preferentially distributed in inner leaflet than in the upper leaflet [25,26,48]. In asymmetric models, PE-ratio is lower than in POPC:POPE:Chol (2:1:1) (25% versus 16%). Furthermore, we can hypothesize that PE lipid tails may also favour or not lipid-protein interactions, since POPC lipids represent only 2.0% of lipid in asymmetric membrane models, PAPE being the most abundant PE lipids in the upper leaflet. In the inner leaflet, PE distribution are relatively similar between POPC:POPE:Chol (2:1:1) and asymmetric membrane. In this case, POPE:POPE:Chol (2:1:1) may be considered as more representative of in situ cell membrane. Indeed, in asymmetric membrane PE lipids represent 28% of lipid components (POPE being 82.5% of them). PS and PA lipids are exclusively distributed in the inner leaflet, representing respectively ca. 5.5% and 1.0% of the total lipid bilayer membrane composition (or respectively 11 and 2% of inner leaflet). In spite of their significantly lower ratios, MD simulations revealed binding hotspots for PS and PA lipids suggesting that they are closely distributed around MFS transporters considered in the present study. It might be easily explained by strong electrostatic contributions between MFS transporters and these lipids. Indeed, *h*GLUT1, *h*GLUT3 and *h*OAT1 exhibit an asymmetric distribution of cationic residues (i.e., arginine, protonated histidine and lysine) along membrane normal. Domains at the interface between membrane and IC compartment are significantly richer in cationic residues than domain at the EC interface. In turn, this favours electrostatic interactions with PA and PS lipids which are anionic lipids in contrast to neutral zwitterionic PC and PE phospholipids.

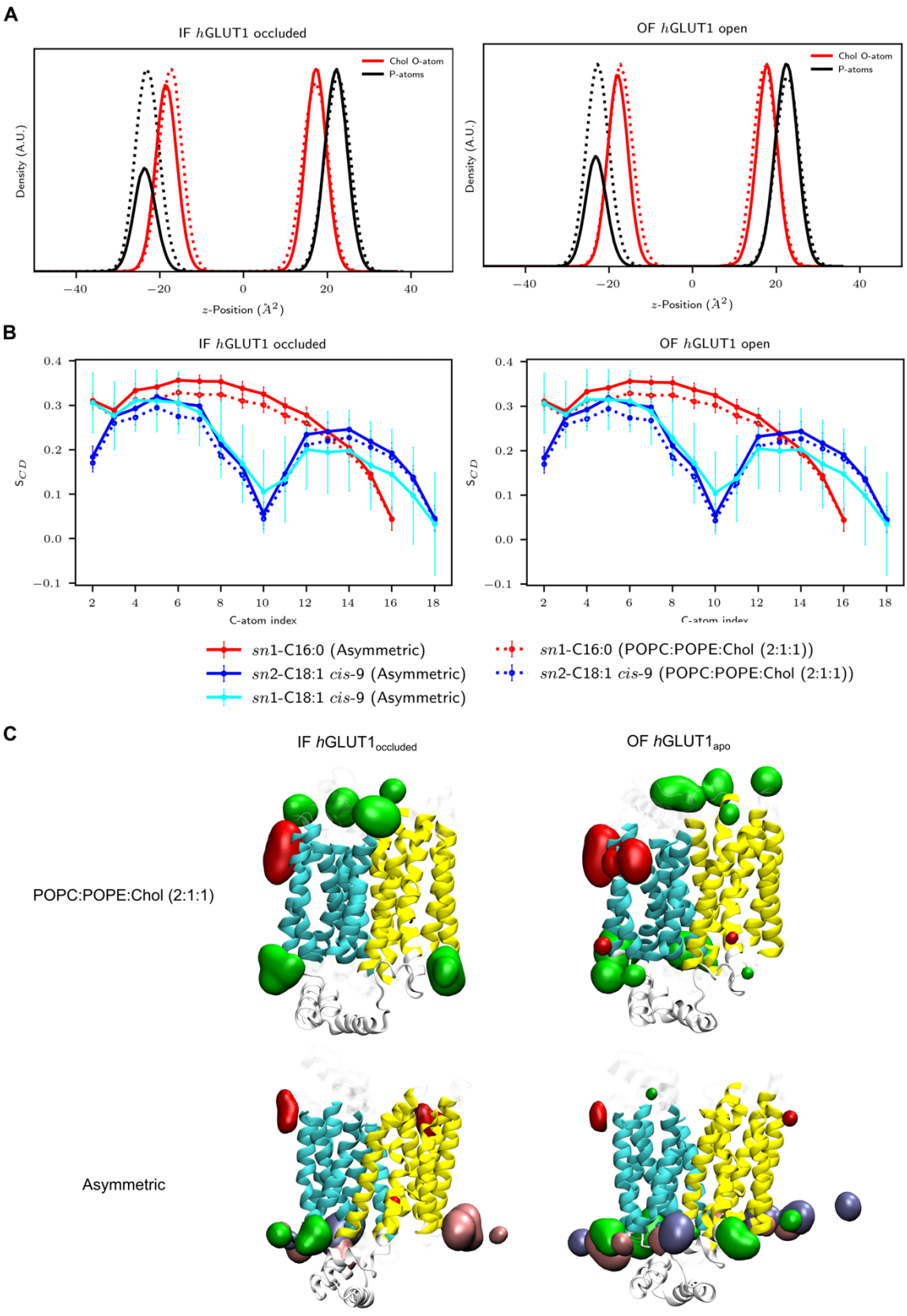


Figure 63. Membrane-protein interactions of *h*GLUT1 in POPC:POPE:Chol (2:1:1) (dashed lines) and asymmetric (solid lines) membrane. (A) The thickness, and (B) lipid order of membranes. (C) The lipid

distribution of *h*GLUT1 shows cholesterol (red), phosphatidylethanolamine (green), phosphatidic acid (iceblue), phosphatidylserine (pink) hot spots.

Given the presence of charge lipids in close proximity with the charge relay system, we can expect that PA and PS lipids play an active role in salt-bridge network interactions along transport cycle. Furthermore, MD simulations also reveal an asymmetric behaviour in terms of direct H-bond between protein and lipids. Even though no clear difference was observed between asymmetric and POPC:POPE:Chol (2:1:1) membranes while counting the number of protein-lipid H-bonds (Fig. S15), such analyses clearly showed that lipids are likely to participate more actively in dynamics of IC regions of MFS transporters than EC.

V.4. Concluding remarks

In the present manuscript, the interplay between surrounding lipid and MFS transporter was investigated by means of MD simulations. Our results suggests that lipid composition is likely not to strongly affect the overall local minimum structures of MFS transporters as long as essential lipids are included in computational models (i.e., PE lipids, cholesterol). However, from the observed subtle differences observed between symmetric and asymmetric lipid bilayer models, we can here hypotheses that lipid composition may modulate conformational transitions along the transport cycle. Our results suggest that transport cycle kinetics may be affected by lipid composition, emphasizing the role of anionic phospholipids distributed in the inner leaflet. Interestingly, MD simulations here proposed that membrane dependence are also transporter dependent. Indeed, *h*GLUT1 and *h*OAT1 appeared more sensitive to lipid bilayer composition than *h*GLUT3. This might be correlated to specific tissue distribution of *h*GLUT1 and *h*OAT1 which were shown to be located exclusively in polarized cells. Similar assumption was made for ABC transporters which are known to be more distributed and less membrane specific than MFS transporters. However, this must be carefully considered and validated by means of joint experimental and computational approaches.

V.5. References

[1] S.K. Nigam, What do drug transporters really do?, Nat Rev Drug Discov. 14 (2015) 29–44. https://doi.org/10.1038/nrd4461.

[2] R.-A.A. Garibsingh, A. Schlessinger, Advances and Challenges in Rational Drug Design for SLCs, Trends in Pharmacological Sciences. 40 (2019) 790–800. https://doi.org/10.1016/j.tips.2019.08.006.

[3] C. Colas, P.M.-U. Ung, A. Schlessinger, SLC transporters: structure, function, and drug discovery, Med. Chem. Commun. 7 (2016) 1069–1081. https://doi.org/10.1039/C6MD00005C.

[4] T. Jiang, P.-C. Wen, N. Trebesch, Z. Zhao, S. Pant, K. Kapoor, M. Shekhar, E. Tajkhorshid, Computational Dissection of Membrane Transport at a Microscopic Level, Trends in Biochemical Sciences. 45 (2020) 202–216. https://doi.org/10.1016/j.tibs.2019.09.001.

[5] N. Yan, Structural Biology of the Major Facilitator Superfamily Transporters, Annu. Rev. Biophys. 44 (2015) 257–283. https://doi.org/10.1146/annurev-biophys-060414-033901.

[6] B.C. Burckhardt, G. Burckhardt, Transport of organic anions across the basolateral membrane of proximal tubule cells, in: Reviews of Physiology, Biochemistry and Pharmacology, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003: pp. 95–158. https://doi.org/10.1007/s10254-002-0003-8.

[7] 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. https://doi.org/10.1038/nrm.2015.25.

[8] M.J. Zamek-Gliszczynski, C.A. Lee, A. Poirier, J. Bentz, X. Chu, H. Ellens, T. Ishikawa, M. Jamei, J.C. Kalvass, S. Nagar, K.S. Pang, K. Korzekwa, P.W. Swaan, M.E. Taub, P. Zhao, A. Galetin, ITC Recommendations for Transporter Kinetic Parameter Estimation and Translational Modeling of Transport-Mediated PK and DDIs in Humans, Clin Pharmacol Ther. 94 (2013) 64–79. https://doi.org/10.1038/clpt.2013.45.

[9] M.J. Zamek-Gliszczynski, M.E. Taub, P.P. Chothe, X. Chu, K.M. Giacomini, R.B. Kim, A.S. Ray, S.L. Stocker, J.D. Unadkat, M.B. Wittwer, C. Xia, S.-W. Yee, L. Zhang, Y. Zhang, International Transporter Consortium, Transporters in Drug Development: 2018 ITC Recommendations for Transporters of Emerging Clinical Importance, Clin. Pharmacol. Ther. 104 (2018) 890–899. https://doi.org/10.1002/cpt.1112.

[10] D. Drew, R.A. North, K. Nagarathinam, M. Tanabe, Structures and General Transport Mechanisms by the Major Facilitator Superfamily (MFS), Chem. Rev. 121 (2021) 5289–5335. https://doi.org/10.1021/acs.chemrev.0c00983.

[11] A. Janaszkiewicz, Á. Tóth, Q. Faucher, M. Martin, B. Chantemargue, C. Barin-Le Guellec, P. Marquet, F. Di Meo, Insights into the structure and function of the human organic anion transporter 1 in lipid bilayer membranes, Pharmacology and Toxicology, 2022. https://doi.org/10.1101/2022.01.10.475390.

[12] D. Jiang, Y. Zhao, X. Wang, J. Fan, J. Heng, X. Liu, W. Feng, X. Kang, B. Huang, J. Liu, X.C. Zhang, Structure of the YajR transporter suggests a transport mechanism based on the conserved motif A, Proc Natl Acad Sci USA. 110 (2013) 14664–14669. https://doi.org/10.1073/pnas.1308127110.

[13] S.E. McComas, D. Mitrovic, C. Alleva, M. Bonaccorsi, D. Drew, L. Delemotte, Determinants of sugar-induced influx in the mammalian fructose transporter GLUT5, Biophysics, 2022. https://doi.org/10.1101/2022.06.17.495601.

[14] A.A. Qureshi, A. Suades, R. Matsuoka, J. Brock, S.E. McComas, E. Nji, L. Orellana, M. Claesson, L. Delemotte, D. Drew, The molecular basis for sugar import in malaria parasites, Nature. 578 (2020) 321–325. https://doi.org/10.1038/s41586-020-1963-z.

[15] N. Yan, A Glimpse of Membrane Transport through Structures—Advances in the Structural Biology of the GLUT Glucose Transporters, Journal of Molecular Biology. 429 (2017) 2710–2725. https://doi.org/10.1016/j.jmb.2017.07.009.

[16] N. Nomura, G. Verdon, H.J. Kang, T. Shimamura, Y. Nomura, Y. Sonoda, S.A. Hussien, A.A. Qureshi, M. Coincon, Y. Sato, H. Abe, Y. Nakada-Nakura, T. Hino, T. Arakawa, O. Kusano-Arai, H. Iwanari, T. Murata, T. Kobayashi, T. Hamakubo, M. Kasahara, S. Iwata, D. Drew, Structure and mechanism of the mammalian fructose transporter GLUT5, Nature. 526 (2015) 397–401. https://doi.org/10.1038/nature14909.

[17] C. Martens, M. Shekhar, A.M. Lau, E. Tajkhorshid, A. Politis, Integrating hydrogen–deuterium exchange mass spectrometry with molecular dynamics simulations to probe lipid-modulated conformational changes in membrane proteins, Nat Protoc. 14 (2019) 3183–3204. https://doi.org/10.1038/s41596-019-0219-6.

[18] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature. 596 (2021) 583–589. https://doi.org/10.1038/s41586-021-03819-2.

[19] T.F. Custódio, P.A. Paulsen, K.M. Frain, B.P. Pedersen, Structural comparison of GLUT1 to GLUT3 reveal transport regulation mechanism in sugar porter family, Life Sci. Alliance. 4 (2021) e202000858. https://doi.org/10.26508/lsa.202000858.

[20] K. Kapoor, J.S. Finer-Moore, B.P. Pedersen, L. Caboni, A. Waight, R.C. Hillig, P. Bringmann, I. Heisler, T. Müller, H. Siebeneicher, R.M. Stroud, Mechanism of inhibition of human glucose transporter GLUT1 is conserved between cytochalasin B and phenylalanine amides, Proc Natl Acad Sci USA. 113 (2016) 4711–4716. https://doi.org/10.1073/pnas.1603735113.

[21] 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. https://doi.org/10.1038/nature14655.

[22] C.R. Søndergaard, M.H.M. Olsson, M. Rostkowski, J.H. Jensen, Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p K_a Values, J. Chem. Theory Comput. 7 (2011) 2284–2295. https://doi.org/10.1021/ct200133y.

[23] 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: GLYCAM06, J. Comput. Chem. 29 (2008) 622–655. https://doi.org/10.1002/jcc.20820.

[24] 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. https://doi.org/10.1002/jcc.23702.

[25] A. Zachowski, Phospholipids in animal eukaryotic membranes: transverse asymmetry and movement, Biochemical Journal. 294 (1993) 1–14. https://doi.org/10.1042/bj2940001.

[26] 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. https://doi.org/10.1038/nrm2330.

[27] H.I. Ingólfsson, M.N. Melo, F.J. van Eerden, C. Arnarez, C.A. Lopez, T.A. Wassenaar, X. Periole, A.H. de Vries, D.P. Tieleman, S.J. Marrink, Lipid Organization of the Plasma Membrane, J. Am. Chem. Soc. 136 (2014) 14554–14559. https://doi.org/10.1021/ja507832e.

[28] R.A. Zager, B.M. Sacks, K.M. Burkhart, A.C. Williams, Plasma membrane phospholipid integrity and orientation during hypoxic and toxic proximal tubular attack, Kidney International. 56 (1999) 104–117. https://doi.org/10.1046/j.1523-1755.1999.00533.x.

[29] D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman, AMBER 2018, (2018).

[30] A.W. Götz, M.J. Williamson, D. Xu, D. Poole, S. Le Grand, R.C. Walker, Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born, J. Chem. Theory Comput. 8 (2012) 1542–1555. https://doi.org/10.1021/ct200909j.

[31] 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. https://doi.org/10.1021/acs.jctc.5b00255.

[32] 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. https://doi.org/10.1021/ct4010307.

[33] 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. https://doi.org/10.1063/1.445869.

[34] I.S. Joung, T.E. Cheatham, Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations, J. Phys. Chem. B. 112 (2008) 9020–9041. https://doi.org/10.1021/jp8001614.

[35] I.S. Joung, T.E. Cheatham, Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters, J. Phys. Chem. B. 113 (2009) 13279–13290. https://doi.org/10.1021/jp902584c.

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

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

[38] 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. https://doi.org/10.1063/1.448118.

[39] 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. https://doi.org/10.1021/ct400341p.

[40] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, Journal of Molecular Graphics. 14 (1996) 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

[41] A.M. Westerlund, L. Delemotte, InfleCS: Clustering Free Energy Landscapes with Gaussian Mixtures, J. Chem. Theory Comput. 15 (2019) 6752–6759. https://doi.org/10.1021/acs.jctc.9b00454.

[42] T. Galochkina, M. Ng Fuk Chong, L. Challali, S. Abbar, C. Etchebest, New insights into GluT1 mechanics during glucose transfer, Sci Rep. 9 (2019) 998. https://doi.org/10.1038/s41598-018-37367-z.

[43] N. Wang, S. Zhang, Y. Yuan, H. Xu, E. Defossa, H. Matter, M. Besenius, V. Derdau, M. Dreyer, N. Halland, K.H. He, S. Petry, M. Podeschwa, N. Tennagels, X. Jiang, N. Yan, Molecular basis for inhibiting human glucose transporters by exofacial inhibitors, Nat Commun. 13 (2022) 2632. https://doi.org/10.1038/s41467-022-30326-3.

[44] Y. Yuan, F. Kong, H. Xu, A. Zhu, N. Yan, C. Yan, Cryo-EM structure of human glucose transporter GLUT4, Nat Commun. 13 (2022) 2671. https://doi.org/10.1038/s41467-022-30235-5.

[45] 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. https://doi.org/10.1038/nature11524.

[46] R.A. Medina, G.I. Owen, Glucose transporters: expression, regulation and cancer, Biol. Res. 35 (2002). https://doi.org/10.4067/S0716-97602002000100004.

[47] K. Nagarathinam, Y. Nakada-Nakura, C. Parthier, T. Terada, N. Juge, F. Jaenecke, K. Liu, Y. Hotta, T. Miyaji, H. Omote, S. Iwata, N. Nomura, M.T. Stubbs, M. Tanabe, Outward open conformation of a Major Facilitator Superfamily multidrug/H+ antiporter provides insights into switching mechanism, Nat Commun. 9 (2018) 4005. https://doi.org/10.1038/s41467-018-06306-x.

[48] S.J. Marrink, V. Corradi, P.C.T. Souza, H.I. Ingólfsson, D.P. Tieleman, M.S.P. Sansom, Computational Modeling of Realistic Cell Membranes, Chem. Rev. 119 (2019) 6184–6226. https://doi.org/10.1021/acs.chemrev.8b00460.

Conclusion

The present work focuses on human proteins of the Major Facilitator Superfamily (MFS) that have been highlighted for their clinical and pharmacological importance. The description of structural patterns provided here provides insights into the function and its modulation considering certain conditions such as the presence of substrates, inhibitors, or the composition of the lipid bilayer membrane.

The present doctoral dissertation has highlighted the benefits of computational methods to support experimental investigations by providing atomistic pictures of difficult-to-catch events. Molecular dynamics simulations provided using *h*OAT1 as a prototype revealed (i) the structural overview of the membrane transporters, revealing important functional aspects; and (ii) plausible binding modes of the drug transporters. Furthermore, by including well-known MFS transporters such as *h*GLUT1 and *hGLUT3,* an atomic-level understanding of the inhibition mode of these MFS transporters and the impact of the lipid bilayer composition was proposed.

The structural investigation of *h*OAT1 has proved that the protein threading modelling methods and μs-MD refinement give a relatively accurate description of the transmembrane domains of the transporter, while the results of loop modelling require careful evaluation. However, due to recent machine-learning-based protein modelling techniques, the quality of the models has vastly improved, despite the fact that the resolution of loops remains relatively low due to high flexibility. Typically for the MFS family, the transmembrane domain of *h*OAT1, *h*GLUT1 and *h*GLUT3 comprises the so-called MFS core, a highly conserved secondary structure that undergoes rearrangement throughout the transport cycle. The rocker-switch mechanism in MFS involves nearly symmetrical movement of two pseudo-symmetrically related bundles around the centrally located substrate-binding cavity. However, the transitions between conformations with the cavity exposed toward the extracellular (OF) or intracellular (IF) compartments occur via metastable states that are mediated via "gating" helices. The gating event appeared to be fine-tuned by the presence of a substrate, where gating helices undergo a local rearrangement and, by interacting with one another, form an occlusion, preventing access to the binding cavity from either side of the membrane. Extracellular gating is dependent on TMH7 and TMH1, which initiate occlusion of the substrate-binding cavity, while intracellular occlusion is mediated by TMH4 and TMH10. Extracellular occlusion can occur even in the absence of the substrate, yet it does not lead to major conformational changes. Conformational change is closely associated with the formation and disassembly of the intracellular salt-bridge network, also known as the charge-relay system.

The charge-relay system is composed of E[X6]R, PETL, and A-motifs that are distributed pseudo-symmetrically in the N- and C-bundles and their distribution and interactions are highly conserved within the MFS proteins. However, crucial for the conformational changes are the interactions between E[X6]R motifs located in the N- and C-bundles, which are strongly preserved in the OF conformation but weaken throughout the transport cycle and are disrupted in the IF conformation. In the majority of, if not all, MFS transporters, it is hypothesized that the intracellular salt-bridge interactions determine the energy barriers that are overcome by substrate binding followed by the occlusion. MD simulations of *h*GLUT1 and *h*GLUT3 with maltose revealed, however, that the presence of an inhibitor prevents the EC occlusion by suppressing interactions with a highly conserved asparagine on the TMH7, which appears to be responsible for the event. Finally, the present investigation has established the significance

of the lipid bilayer composition and its impact on conformational states, in agreement with evidence supporting the regulating role of the lipid bilayer components [1–3].

Using recent computational techniques, the work has also demonstrated the allosteric influence of lipid bilayer components on protein function. In order to stabilise the OF conformation and possibly facilitate the transition between conformational states, phosphatidylethanolamine has been shown to interact directly with the protein, as well as anionic lipids in the IC domains. Our results are in line with experimental investigations which have clearly demonstrated the importance of such interactions in MFS functions Allosteric investigations including lipid bilayers have shown surrounding lipids actively participate in the communication between the binding pocket and the critical intracellular interactions.

In light of the substantially larger number of SLC family members as compared to ABC proteins (420 vs. 80), a wider range of structural, functional, and mechanistic variabilities may arise. However, among the subfamilies defined according to the fold and mechanism, the patterns appear to be relatively conserved in MFS superfamily. The structural patterns of the three MFS transporters under study - *h*OAT1, *h*GLUT1, and *h*GLUT3 - were relatively consistent and showed striking similarities. The provided insights into the structures and mechanisms will hopefully benefit the further investigation of the MFS family.

In view of the increased interest in rational drug design and individualised treatment, the theoretical tools used here for modelling drug-protein interactions have gained prominence. The benefits of the prediction and understanding provided by the computational methods allow for more conscious decisions in the above-mentioned processes. One can be confident that appropriate use of such methods can be relevant to deciphering atomic-scale events which are of importance in pharmacological processes. The present work aims at being extrapolated to other MFS transporters of pharmacological relevance, if structures are either experimentally resolved or if ML-based structural predictions are robust enough.

I am grateful for the opportunity to contribute knowledge to the scientific community that will hopefully benefit the pharmacological community by expanding our understanding of the mechanistic aspects of MFS proteins. During the studies, I contributed to the research on ABC drug transporters, specifically ABCB4, ABCC1, which are both important from a pharmacological standpoint. I am honoured to have been given the opportunity to work on such an intriguing topic with pharmacological impacts. I am grateful for the chance to learn and contribute to the scientific communities of pharmacology but also of computational modelling.

References

[1] C. Martens, R.A. Stein, M. Masureel, A. Roth, S. Mishra, R. Dawaliby, A. Konijnenberg, F. Sobott, C. Govaerts, H.S. Mchaourab, Lipids modulate the conformational dynamics of a secondary multidrug transporter, Nat Struct Mol Biol. 23 (2016) 744–751. https://doi.org/10.1038/nsmb.3262.

[2] C. Martens, M. Shekhar, A.J. Borysik, A.M. Lau, E. Reading, E. Tajkhorshid, P.J. Booth, A. Politis, Direct protein-lipid interactions shape the conformational landscape of secondary transporters, Nat Commun. 9 (2018) 4151. https://doi.org/10.1038/s41467-018-06704-1.

[3] C. Martens, M. Shekhar, A.M. Lau, E. Tajkhorshid, A. Politis, Integrating hydrogen–deuterium exchange mass spectrometry with molecular dynamics simulations to probe lipid-modulated conformational changes in membrane proteins, Nat Protoc. 14 (2019) 3183–3204. https://doi.org/10.1038/s41596-019-0219-6.

Appendices

List of publications:

Accepted publications

Insights into the structure and function of the human organic anion transporter 1 in lipid bilayer membranes
Angelika Janaszkiewicz, Ágota Tóth, Quentin Faucher, Marving Martin, Benjamin Chantemargue, Chantal Barin-Le Guellec, Pierre Marquet, Florent Di Meo
Scientific Reports, DOI: 10.1038/s41598-022-10755-2

Effect of CFTR correctors on the traffic and the function of intracellularly retained ABCB4 variants
Amel Ben Saad, Virginie Vauthier, Ágota Tóth, Angelika Janaszkiewicz, Anne-Marie Durand-Schneider, Alix Bruneau, Jean-Louis Delaunay, Martine Lapalus, Elodie Mareux, Isabelle Garcin, Emmanuel Gonzales, Chantal Housset, Tounsia Aït-Slimane, Emmanuel Jacquemin, Florent Di Meo, Thomas Falguières
Liver International, DOI: 10.1101/2022.01.10.475390

Ruthenium Olefin Metathesis Catalysts Bearing a Macrocyclic N-Heterocyclic Carbene Ligand: Improved Stability and Activity
Wioletta Kośnik, Dawid Lichosyt, Marcin Śnieżek, Angelika Janaszkiewicz, Krzysztof Woźniak, Maura Malińska, Bartosz Trzaskowski, Anna Kajetanowicz, Karol Grela
Angewandte Chemie International Edition, DOI: 10.1002/anie.202201472

The influence of the cationic carbenes on the initiation kinetics of ruthenium-based metathesis catalysts; a DFT study
Magdalena Jawiczuk, Angelika Janaszkiewicz, Bartosz Trzaskowski
Beilstein Journal of Organic Chemistry, DOI: 10.3762/bjoc.14.266

Preprint (Under revision)

Substrate binding and lipid-mediated allostery in the human organic anion transporter 1 at the atomic-scale
Angelika Janaszkiewicz, Ágota Tóth, Quentin Faucher, Hélène Arnion, Nicolas Védrenne, Chantal Barin-Le Guellec, Pierre Marquet, Florent Di Meo
Status: Under revision (Major revisions received)
bioRxiv, DOI: 10.1101/2022.07.14.500056

On the interplay between lipids and asymmetric dynamics of an NBS degenerate ABC transporter
Ágota Tóth, Angelika Janaszkiewicz, Veronica Crespi and Florent Di Meo,
Status: Under revision (Major revisions received)
bioRxiv, DOI: 10.1101/2022.05.16.492073v1

In preparation:

Membrane dependency on the structural dynamics of human major facilitator superfamily transporters
Angelika Janaszkiewicz, Ágota Tóth, Veronica Crespi and Florent Di Meo

Structural and dynamical differences shown by the pre- and post-hydrolysis states of ABCC1/MRP1
Ágota Tóth, Angelika Janaszkiewicz, Veronica Crespi and Florent Di Meo
Invited Special issue in Basic and Clinical Pharmacoloy & Toxicology

Taming the massive intestinal regeneration of mycophenolic acid to prevent mycophenolate-induced enteropathy
Clarisse Brossier, Manon Jardou, Angelika Janaszkiewicz, Hélène Arnion, Emilie Pinault, François-Ludovic Sauvage, Nicolas Picard, Florent Di Meo, Pierre Marquet, Roland Lawson

In silico Models of Pharmacologically Relevant Membrane Transporters: Focus on the Major Facilitator Superfamily Fold

La traversée des membranes par des médicaments influencent grandement la pharmacodynamique et la pharmacocinétique de ces derniers. Ces événements impliquent souvent des transporteurs membranaires qui sont classifiés en deux superfamilles, à savoir les Solute Carriers (SLC) et les transporteurs dits « ABC » (ATP-Binding Cassette). Malgré l'importance des transporteurs en pharmacologie clinique et les progrès récents dans la compréhension des interactions PK/PD au niveau local, les connaissances sur leurs structures et fonctionset les interactions médicamenteuses ou encore l'implication de la pharmacogénomique (PGx) restent assez limitées, surtout au niveau moléculaire.

Il n'existe actuellement aucune méthode expérimentale capable de donner une vue complète dynamique et structurale des transporteurs. La description atomique des transporteurs et de leur dynamique devrait permettre une meilleure connaissance de la structure des protéines, des changements de conformation et des mécanismes sous-jacent. Ceci pourra améliorer la compréhension de la liaison des substrats, des modes d'inhibition et de la cinétique de transport. Les avancées réalisées avec les simulations de dynamique moléculaire au cours des dernières décennies a été démontrée par leur capacité à compléter les données expérimentales sur les transporteurs et à fournir des images à l'échelle nanométrique. Dans ce contexte, le présent travail se concentre sur le décryptage des modèles structuraux des transporteurs de la Superfamille des Facilitateurs Majeurs (MFS) au moyen de simulations de dynamique moléculaire. L'accent a été mis sur un transporteur d'importance clinique et pharmacologique, le transporteur OAT1 (Organic Anion Transporter 1), qui fait partie de la liste des transporteurs membranaires humains considérés par le Consortium international des transporteurs (ITC) comme ayant une "importance clinique émergente". À des fins de comparaison, l'interaction entre les lipides et OAT1 ainsi que des prototypes de MFS (transporteurs de glucose 1 et 3 -SLC2A1/GLUT1 et SLCA3/GLUT3) a également été étudiée.

Mots-clés: Pharmacology, Personalized Medicine, Molecular Dynamics, Drug Transporters

In silico Models of Pharmacologically Relevant Membrane Transporters: Focus on the Major Facilitator Superfamily Fold

Drug pharmacodynamics (PD) and pharmacokinetics (PK) are both strongly impacted by drug membrane crossing events. Such an event often involves membrane drug transporters that are divided into two main superfamilies, namely Solute Carrier (SLC) or ATP-Binding Cassette (ABC) transporter superfamilies. Despite the importance of transporters in clinical pharmacology and recent advances in local PK/PD relationship, knowledge about their functions, but also drug-drug interactions or pharmacogenomics (PGx) involving membrane transporters is still rather limited, especially at the molecular level.

There are currently no experimental methods that can give an overall dynamic picture of transporter functions at the atomic level. The atomic description of transporters and their dynamics is anticipated to result in a greater knowledge of protein structure, conformational changes, and the underlying mechanism, which in turn will enhance our understanding of substrate binding, inhibition modes, and kinetics. The effectiveness of molecular dynamics (MD) simulations in recent decades has been demonstrated in their capacity to reveal structural properties, support experimental data on transporters, and provide pictures at the nanoscale. Therefore, the present work focuses on deciphering the structural patterns of Major Facilitator Superfamily (MFS) transporters by means of MD simulations. The focus was on a clinically relevant transporter, Organic Anion Transporter 1 (SLCA22A6/OAT1), which is among a group of human membrane transporters emphasized by the International Transporter Consortium (ITC) as being of "emerging clinical importance". For sake of comparison, the interplay between lipids and OAT1 as well as MFS prototypes, (i.e., Glucose Transporter 1 and 3) was also investigated.

Keywords: Pharmacology, Personalized Medicine, Molecular Dynamics, Drug Transporters