



Università degli Studi di Cagliari

PHD DEGREE

Physics

Cycle XXX

TITLE OF THE PHD THESIS

**Transport Properties in Specific
Porins from *A. baumannii* and Ion
Channels**

Scientific Disciplinary Sector(s)

FIS 07 / Fisica Applicata

PhD Student: Dehbia BENKERROU

Coordinator of the PhD Programme

Prof. Alessandro De Falco

Supervisor Prof. Matteo CECCARELLI

Final exam. Academic Year 2016 – 2017

Thesis defence: February-March 2018 Session

Transport Properties in Specific Porins from *A. baumannii* and Ion Channels

Submitted by: Dehbia BENKERROU

11 December 2017



Università degli Studi di Cagliari Facoltà di Scienze
Dipartimento di Fisica

Tutor: Prof. Matteo Ceccarelli

Coordinator: Prof Alessandro De Falco

Table of Contents

Abstract	4
Acknowledgements	5
Chapter 1: General Introduction	6
<i>Organization and main objective of this thesis</i>	<i>12</i>
<i>References</i>	<i>15</i>
Chapter 2: <i>In silico</i> characterization of small molecules permeation through porins of <i>A. baumannii</i>	17
<i>Introduction</i>	<i>18</i>
<i>The intrinsic resistance via porins</i>	<i>22</i>
General (non-specific) porins	22
Specific porins	23
Porins of <i>A. baumannii</i>	24
OccAB porins	25
Treatment options of <i>A. baumannii</i> infections	25
OM permeability challenge for antimicrobial drug development.....	26
<i>The aim and objectives of this study</i>	<i>27</i>
<i>Results and discussion</i>	<i>28</i>
Investigation of the substrates translocation and their selectivity by OccAB porins	28
Dynamic and shape of the pore dimension.....	28
The unusual pore flexibility of OccAB2 porin	29
Electrostatic features of OccAB porins.....	32
Investigation of the substrates diffusion and free energy calculations	35
The translocation of substrates through OccAB1.....	35
The translocation of substrates through OccAB2.....	37
The translocation of substrates through OccAB3.....	38
The translocation of substrates through OccAB4.....	41
Antibiotics translocation through OccAB1 porin.....	44
Internal electric field in OccAB1	45
Free energy surfaces of antibiotics penetration	45
Imipenem and Meropenem	46
Ampicillin.....	47
β -lactamase inhibitors.....	48
Avibactam and sulbactam	48
<i>Conclusion and perspectives.....</i>	<i>55</i>
<i>Appendix 1.....</i>	<i>57</i>
<i>References</i>	<i>63</i>
Chapter 3 On the modelling of human TPC2 channel: homology modelling and docking simulations.....	69
<i>Introduction.....</i>	<i>70</i>
Cellular and sub-cellular localization	70
Topology, function and regulation.....	70
The aim of this study	73

<i>Results and discussion</i>	74
I. Model building, structure refinement and validation	74
C-terminal domain(CTD) prediction	75
The Quality of the model.....	75
Pocket detection using the PockDrug server.....	78
II. Molecular Docking simulations	79
Validation of the docking protocol.....	79
Docking results of naringenin and verapamil.....	80
Results of the docking I: blind docking.....	80
Results of the docking II (selective filters & voltage-sensitive domains (VSD)).....	83
<i>Conclusion and perspectives</i>	88
<i>Appendix</i>	90
<i>References</i>	94
Chapter 4 Material &Methods	99
<i>Molecular modelling</i>	100
Molecular dynamics simulations	100
Force-fields contribution	102
AMBER.....	103
CHARMM	103
Ensemble thermodynamics	106
Periodic boundary conditions	106
Main Steps of classical MD simulations runs and details used in our systems	107
System preparation and Simulations details.....	107
OccAB1 systems:	107
PCA analysis.....	108
HsTPC2 channel model system:	108
Methods for free energy calculations	108
Metadynamics.....	109
Homology modeling and molecular docking.....	112
Steps in molecular modelling	113
Model validation[39]	114
Molecular docking.....	115
Protein-ligand interactions.....	115
Scoring functions approaches	115
Force field-based methods[45]	116
Involve the flexibility in the process.....	117
Details of the Protocol of AutoDock docking	117
<i>References</i>	120
Chapter 5	124
<i>General conclusion</i>	124

Abstract

This work is about the computational study of two cases of ion channel systems; OccAB porins from gram-negative bacterium and human TPC2 channel. In particular, describe the biophysical properties of transport processes in OccAB porins, the TPC2 channel-ligand binding interactions on the basis of molecular modelling approaches.

The first part of this thesis focuses on the study of OccAB porins that are ion channels located in the outer membrane of gram-negative pathogen *A. baumannii*. These channels work as filters to the small molecule and antibiotics translocation through the outer membrane. In order to study, at atomic scale, the parameters that modulate their selectivity to specific substrates, we applied different molecular dynamics simulation methods. In particular, we reported here the surface free energy of the translocation barriers of each channel of OccAB1-4, first, to natural substrates (arginine, glutamic acid and glycine) in order to probe the importance of the molecular charge and size in the discrimination of small molecules. We then evaluated the free energy translocation barriers of five molecules of antibiotics through OccAB1, the most open channel among OccAB porins. This latter investigation, revealed the crucial role of the internal transversal electric field located in the constriction region of OccAB1 in guiding the translocation of molecules of antibiotics through this channel. Here, we reported a quantification of the electric field, the free energy surfaces of the translocation barriers and the interactions involved in the molecules-porins complexes.

The second part of the thesis, treated the human TPC2 gated sodium ion channel playing an essential role in several cellular functions and signaling regulation. As a result of its involvement in several important diseases such as Parkinson's disease and Ebola virus, the TPC2 channel has emerged as an important therapeutic target in early drug discovery and development. As part of this thesis, we focused on the study of two molecules, naringenin and verapamil, known to act as blockers in similar gated-ion channels in order to check whether these selected molecules have the same binding sites in the human TPC2 channel, by means of molecular modelling and docking method. Our results give atomistic information about putative binding sites of these potential blockers. We reported the molecular basis and the major modes of binding that can be used as supporting information for mutagenesis experiments for the understanding of the complex mechanism of these channel with two proposed small molecules to act as blockers.

Overall, the use of molecular modelling methods for the study of these two membrane systems has contributed to elucidate complex biological processes such as transport through the outer membrane and to describe physical-chemical parameters involved in the formation of the functional complex systems.

Keywords: Membrane ion channels, OccAB porins, human TPC2 channel, classical molecular dynamics, metadynamics, homology modelling, molecular docking.

Acknowledgements

I would first like to thank those who gave me the opportunity to come to the University of Cagliari to implement this doctoral project including the University of Cagliari,

I would like to thank Pr. Matteo CECCARELLI the tutor of this thesis for having me in his group during almost four years, for allowing me to carry out this research work and to have followed my work progressively during this thesis.

I would also like to thank Pr. Paolo RUGGERONE, who supported me during the program of this thesis by giving me good advice, encouragements and for his precious help. Thank you for everything.

I would like to thank Dr. Giuliano MALLOCI for his advice and for helping to complete the writing of this thesis manuscript. I thank him for helping me to implement molecular docking simulations on human TPC1 channel.

I would also like to thank Pr. Anne-Claude CAMPROUX and Dr. Gautier MOROY with home I spent 3 months of Erasmus placedoc placement. Thank you for having me in their laboratory in Paris Diderot-Paris 7 and for your help to realize the part of this thesis on building a sequence based homology model of the human TPC2 channel as well as the molecular docking simulations. Thank you for everything.

Finally, I would like to thank my family who has always supported me. Thank you for your eternal love and support.

Chapter 1

General Introduction

Despite the huge costs spent each year by the pharmaceutical industry for drug discovery, the success of the development of new drugs through clinical tests is stagnating. It is estimated that less than 40% of new molecular entities (NMEs) synthesized by pharmaceutical companies are candidates for preclinical testing [1]. Of these, only about 10% pass clinical trials and are approved for marketing [2]. The development of a new drug implies an interdisciplinary process driven by a funnel of several increasingly complex and strict stages constituting severe barriers that must be escaped by NMEs. This process operates according to the double principle of intuition and iteration. From a hundred basic therapeutic ideas (intuitions), several million compounds can be considered potentially interesting. They are then screened by automatic test protocols that eliminate most of them. At the end of this process 1 or 2 molecules will emerge perhaps and will arrive on the market 10 or 15 years later [3].

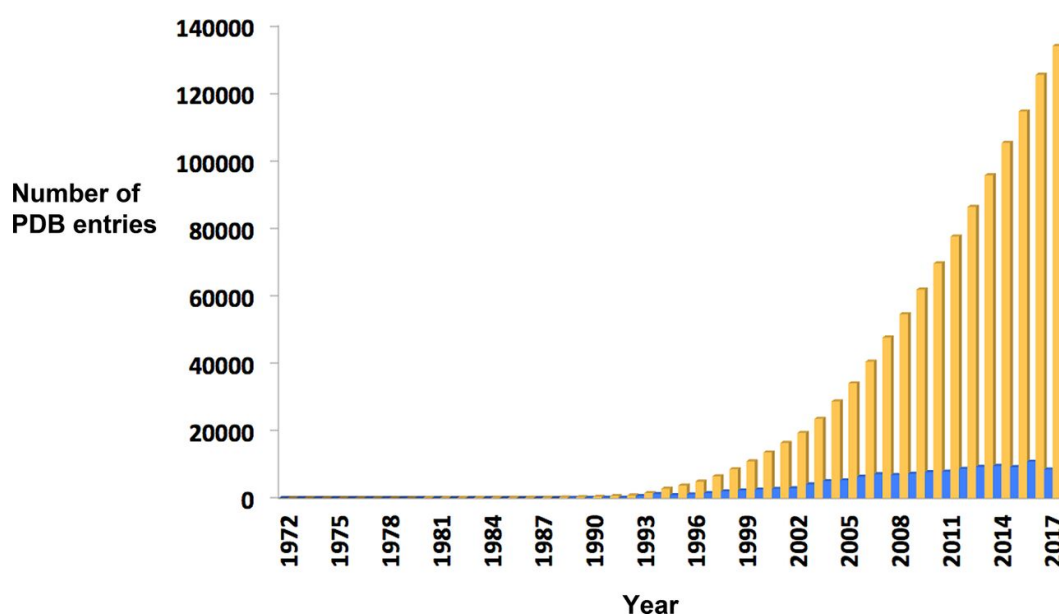


Figure 1: Yearly growth of structures in the Protein Data Bank

Number of structures available in the PDB per year through May, 2017. Beige bars indicate the number of X-ray, NMR, electron microscopy and modelled structures in the database. Blue bars indicate the total number of structures deposited per year. Source: RCSB Protein Data Bank.

www.rcsb.org/pdb/home/home.do

One of the major barriers to develop new effective drugs is the many long and complex requirements that must be respected during the drug development process, including administrative, legislative and ethical aspects, adding to that the chemical, biological and

physiological conditions that have to be respected. Indeed, for a molecular product to be considered as an effective drug, it must be easily administered to patients and must be stable as long as needed to reach its target. In addition, to avoid undesirable physiological effects, a drug must not modulate the properties of biomolecules other than its target. All of these requirements involve high costs for a new product to be allowed on the market, significantly limiting the number of compounds likely to become clinically useful drugs. As a result, the research-based pharmaceutical industry currently spends more than 149.8 billion dollars in R & D per year [3] [2] [4]. In 2011-2015, the number of new chemical or biological entities launched on the world market increased from 146 a decade earlier to 226 [3]. The most worrying example illustrating the lack of progress in the development of new therapeutic molecules, is the lack of new efficient antimicrobial drugs against antimicrobial resistance.

Moreover, since the 90s, the drug development process has been facing a technological revolution in several domains, including genomics, proteomics, combinatorial chemistry, computer-aided drug design, robotics and high-throughput screening [5] [6]. This has significantly transformed and improved research methods for NMEs. During which, the lack knowledge of the therapeutic targets is not improving that much and represent another important issue that drug industry has to deal with. Indeed, among all the estimated therapeutic targets, very few are known and the modes of action of the majority of them remain to be discovered. This challenge is even more true for the membrane proteins in the both eukaryotic and prokaryotic organisms. Being a major component of real membranes involved in a variety of important biological functions, the membrane proteins emerged as a major therapeutic target for the development of new drugs for the treatment of many emerging diseases. In the top 100 drugs, more than 60% target membrane proteins. Whereas, when approximately 3000 therapeutic targets are estimated, 2000 are membrane proteins but only 10% are known [7].

Yet, the biological activity of the membrane proteins depends mainly on their spatial conformation. Therefore, the development of an effective new drug requires an intimate knowledge of the target's 3D-structure that makes the latter active in its cellular environment. The knowledge of the 3D molecular structure of a protein target gives access to information about the active site and possible ligand binding in the target. In addition, structural information on the target also provide crucial details at the atomic scale on its biophysical properties, such as the electrostatic and geometric properties that are important for predicting the possible interaction partners involved in regulation and complexation [8]. Such information allow to

gain fundamental insights in our understanding of the relationship between the structure-function and the disease development [8].

In Jun 2017, the number of the solved crystal structure in PDB data base reached 131400 proteins, but only 2.4% of them are transmembrane proteins [9] (Figures 1 [10] and 2). The gap between the numbers of available structures and the number of potential membrane protein targets is mainly due to the particular difficult conditions in which new membrane proteins are obtained. Thus, experimental data at the atomic resolution for membrane proteins are very rare making the design of appropriate new drugs more prevented. Indeed, the solubilisation,

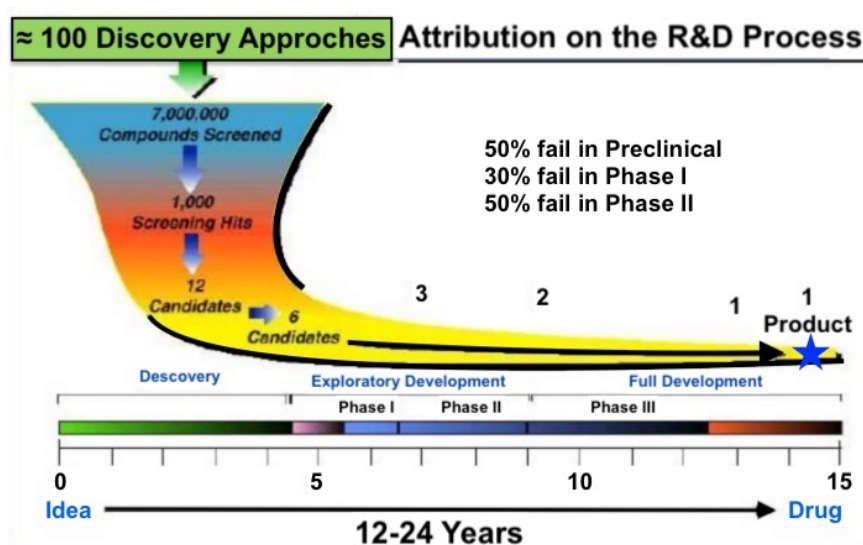


Figure 2: **Drug development process**

example of drug discovery and development testing funnel. Adapted from [1]

purification and extraction of membrane proteins while guaranteeing their functionality is particularly difficult due mainly to their particularity of being inserted into fluid lipid membranes within the cell envelopes. Hence, the widely used method, X-ray crystallography, to obtain the 3D structure for a large number of soluble protein has limitations in producing single crystals for membrane proteins which explains the relatively low number of crystals in the Protein Data Bank (PDB).

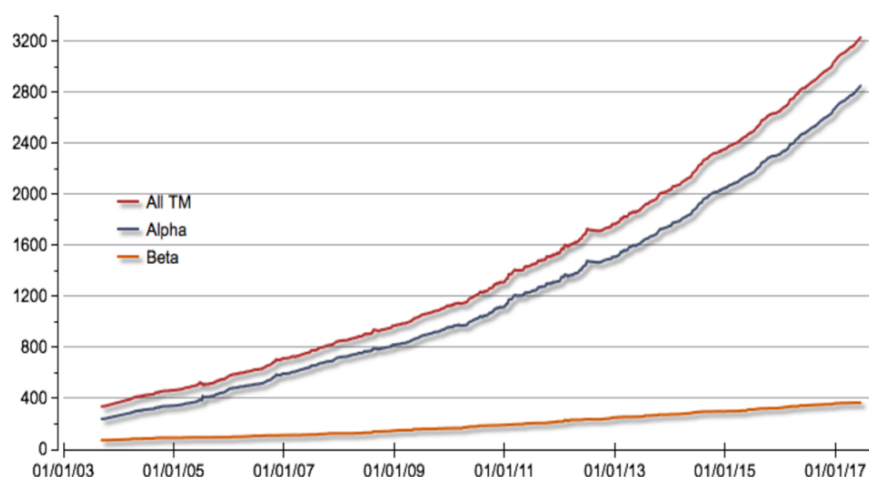


Figure 3: The number of transmembrane protein in the Protein data bank: (3227 in total of which alpha: 2848, beta: 366). Adapted from http://pdbrtm.enzim.hu/?_=/statistics/growth

Nevertheless, to speed up the process of drug design development, the pharmaceutical industry started to realize the necessity to integrate the parallel development of various technologies, including methods that are inquiringly fast and sophisticated whilst keeping the requirements of an effective drug development for lower costs. More particularly, the rise of computational methods such as molecular modelling with improved algorithms highly contributed to generate 3D models of important drug targets based on sequence similarity and structural conservation with the available experimental data coming mainly from X-ray data [11] [12]. This homology modelling approach helped to increase greatly the number of available structures of membrane proteins. In parallel to the 3D structure modelling, *in-silico* approaches have been also specifically developed to predict and rationalise binding motifs of many ligand molecules to these X-ray/3D modelled structure such as the so-called structure based drug design (SBDD) approach [13] [14]. However, available X-ray structures or the 3D model can only provide a snapshot of one particular conformational state of a protein without any information about its flexibility and overall dynamical behaviour [15].

On the other hand, dynamical properties could play an important role at various structural levels for the functionality of many biological machineries, such as ion channels during substrates transport process. In general, the most flexible regions are regions that are unfolding, rather disordered such as protein loops that may contain interaction sites for small molecules as is often the case in the outer vestibule of many channels. The flexibility of the residues lining

transport channels is often involved in the selectivity and diffusion mechanism, for example the residues located in the narrowest region of bacterial porins could have an effect on the passage of small molecules such as antibiotics. In the case of the specific gated-ions channels, in sodium channel for instance, the flexibility of the residues of the selective filter could have a direct effect on the discrimination of the different ions based on the size that would pass through the channel by changing the size of the filter.

Widely used experimental methods which study the transport process such as electrophysiology and liposome swelling assay cannot give detailed atomic information about the conformational changes that are involved in several biological processes such as transport of small molecules by membrane proteins and ligand binding into specific pockets inside the target. All of these conformational changes including atomic fluctuation, domain rearrangements, and the movement of the loops are often crucial for the function of protein channels and transport process [16] [17].

The computational method mostly used to study such biological phenomenon is molecular dynamics (MD) simulations. The principle of a MD simulation is based on the use of computer algorithms to simulate the motion of a large number of atoms and follow their evolution over time. MD simulations provide a complementary view that can help the understanding of experimental data and can, in principal, guide and suggest new dedicated experiments.

Nowadays, computer simulations raised the number of accurate algorithms for modelling the 3D structure and study their function. For biological systems involving rare events that are inaccessible in the time scale imposed by classical MD simulations such as transport process of ions and small molecules through protein channels inserted in their biological membrane, different computational approaches (accelerated sampling methods such as metadynamics [18], umbrella sampling [19]...etc.) are widely used for large scale-simulations (Figure 3) [20] [21]. The use of such methods is important to investigate the biomolecular flexibility associated with ligand recognition [22] [23] and also to discover and characterize explore binding sites on the target receptor that are not evident from X-ray structures [24] (Figure 3).

This work is focused on a class of membrane proteins called ion channels. Ion channels are a group of integral membrane proteins that allow the diffusion of hydrophilic molecules that cannot cross the lipid bilayer at sufficient rates to achieve the cell's needs. When molecules make use of these proteins to move from one compartment of the cell membrane to the other by spontaneous diffusion and other natural processes of passive transport, they don't require any energy costs for the cell and the transport mechanism is classified as passive.

Structurally, ion channels are membrane proteins that show two different common secondary structures found in many membrane proteins: α -helices and β -sheets (Figure 2). α -helical membrane proteins are observed for transmembrane proteins of various functions and origins, illustrating that this class of membrane proteins is widespread in all organisms while β -sheets are mainly observed for outer membrane proteins forming cylinder of beta-barrel conformation. To date beta-barrel proteins have exclusively been found in the outer membrane of the Gram-negative bacteria cell envelopes and in the outer membranes of mitochondria and chloroplasts. In particular, in this thesis, we focus on the translocation process through two specific ion channels; the Human TPC sodium ion channel (α -helical) and the most abundant bacterial ion channels in the outer membrane of the Gram-negative bacterium *A. baumannii*, namely the OccAB porins. (β -sheets). Despite recent experimental investigations giving information on the structures and selectivity to specific substrates or ions of these selected proteins, the exact mechanisms of function of both systems are still unknown. These two selected membrane systems are of interest due to their implication in serious diseases.

In the case of OccABs porins, *A. baumannii* is one of the most dangerous pathogen that causes resistance against all of the currently available antibiotics, representing thus an increasing global health problem. The low permeability of the outer membrane (OM) is a major factor contributing to resistance of this Gram-negative pathogen. OccAB porins that are inserted in the OM, play a major role in the membrane permeability by filtering the antibiotics and natural substrates that can pass in the cell pathogen. It is thus important to study at the molecular scale the mechanism of action of these porins in the translocation process in order to improve our understanding of the OM permeability of *A. baumannii* [25] [26].

TPC (Two-pore channels) is responsible of Ca^{2+} release from lysosomes and endolysosomes. A better understanding of its biophysical properties is required due to the implication of this protein in various diseases including Parkinson's disease [27], and Ebola infection [28]. Different studies suggest that this channel is activated by NAADP (nicotinic acid adenine dinucleotide phosphate), $\text{Pi}(3,5)\text{P}_2$ (phosphatidylinositol 3,5-bisphosphate) and cytosolic Ca^{2+} . However, the crystal structure of TPCs human is not yet available making the structure-based investigation at the atomic level more difficult. In order to fill this gap in this work we built, based on the homology modelling, a model of a 3D-structure of TPC2 Human. Then, in order to investigate *In-Silico* their mechanisms of action in their membrane environment we used an integrated approach combining molecular docking and molecular dynamic simulations methods

(MD).

Organization and main objective of this thesis

The present work contributes to the understanding of structure-function relationship of a class of membrane proteins, namely the ion channels, that attracted a great interest within the biophysicists community in the last years. In particular, we focus on two protein channels expressed in the membrane of two different living organisms, the porins from Gram-negative bacteria (*Acinetobacter baumannii*) and the TPCs channels from *Homo Sapiens*.

The whole Thesis is organized into Five chapters. After a first introductory chapter (Chapter 1) the two main central chapters (Chapter 2 and 3) present original research data on membrane proteins of the two different organisms (prokaryotic (bacteria), eukaryotic (human)) under investigation. The Thesis ends with a material and methods chapter (Chapter 4) and a general conclusions and perspectives chapter (Chapter 5).

Chapter 1: Introduces the state of art and the general context on which the topics of this thesis are important for the development of new therapeutic drugs.

Chapter 2: *OccAB porins system:* This chapter is focused on the investigation and analysis of 4 porins from *Acinetobacter baumannii*. Porins are the most important systems for the translocation of small molecules through the bacteria cell envelope. In particular, they are involved in the translocation of antibiotics, hence directly implicated in the drug discovery development. The porins under study have been very recently resolved at high resolution which gives the opportunity to study their atomic structure and investigate the translocation of small molecules at atomic level by using MD simulations. The main aim of the work was to assess the influence of the combined size-constriction and electrostatic interactions in the selectivity/transport of the OccAB porins toward small substrates and antibiotics.

Chapter 3: *TPCs channels:* This chapter concerns the human TPC2 channel. This protein belongs to a class of classified as alpha-helices trans-membrane proteins which are sodium selective and modulated by calcium. The work has started about one year ago, when the crystal of TPC1 from *A. thaliana* was first published at high resolution. The first interest was to build the 3D structure of the human TPC2 channel using sequence homology structural methods. In

addition, few experimental data are now available concerning the binding of small molecules that show inhibition of *HsTPC2* in the case of serious diseases such as Melanoma, Parkinson and Ebola virus.

Chapter 4: describes briefly the numerical methods and algorithms used in this thesis discussing advantages and limitations for the selected systems.

Chapter 5: contains a general conclusion and some future perspectives of this thesis.

References

- [1] E. H. Ohlstein, R. R. R. Jr, and J. D. Elliott, "D d n m," no. 2, 2000.
- [2] IFPMA, "the Pharmaceutical Industry and Global Health," 2017.
- [3] PhRMA, "2016 Biopharmaceutical Research Industry Profile," *Pharm. Res. Manuf. Am.*, p. 86, 2016.
- [4] A. Mullard, "2014 FDA drug approvals," *Nat. Rev. Drug Discov.*, vol. 14, no. 2, pp. 77–81, 2015.
- [5] G. D. Geromichalos, C. E. Alifieris, E. G. Geromichalou, and D. T. Trafalis, "Overview on the current status on virtual high-throughput screening and combinatorial chemistry approaches in multi-target anticancer drug discovery; Part II," *J. B.U.ON.*, vol. 21, no. 6, pp. 1337–1358, 2016.
- [6] E. Maréchal, "Chemogenomics: a discipline at the crossroad of high throughput technologies, biomarker research, combinatorial chemistry, genomics, cheminformatics, bioinformatics and artificial intelligence.," *Comb. Chem. High Throughput Screen.*, vol. 11, no. 8, pp. 583–586, 2008.
- [7] A. L. Hopkins and C. R. Groom, "Hopkins200Nat_Druggable Genome," vol. 1, no. September, pp. 7–10, 2002.
- [8] S. Napier and M. Bingham, *Transporters as Targets for Drugs*. .
- [9] F. C. Bernstein, "The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures," *Alcohol*, vol. 112, pp. 535–542, 1977.
- [10] P. AP, "Global Challenges in Cardiovascular Drug Discovery and Clinical Trials," *Mol. Biol.*, vol. 6, no. 3, pp. 1–4, 2017.
- [11] A. Fiser and A. Šali, "MODELLER: Generation and Refinement of Homology-Based Protein Structure Models," *Methods Enzymol.*, vol. 374, pp. 461–491, 2003.
- [12] N. Eswar, D. Eramian, B. Webb, M.-Y. Shen, and A. Sali, "Protein structure modeling with MODELLER," *Methods Mol. Biol.*, vol. 426, pp. 145–59, 2008.
- [13] M. D. Varney *et al.*, "Crystal-Structure-Based Design and Synthesis of Benz[cd]indole-Containing Inhibitors of Thymidylate Synthase," no. 3, 1992.
- [14] B. D. Dorsey *et al.*, "L-735,524: The Design of a Potent and Orally Bioavailable HIV Protease Inhibitor," *J. Med. Chem.*, vol. 37, no. 21, pp. 3443–3451, 1994.
- [15] S.-H. Chung and B. Corry, "Conduction properties of KcsA measured using brownian dynamics with flexible carbonyl groups in the selectivity filter.," *Biophys. J.*, vol. 93, no. 1, pp. 44–53, 2007.
- [16] A. Pohorille, M. A. Wilson, and G. Shannon, "Flexible Proteins at the Origin of Life," *Life*, vol. 7, no. 2, p. 23, 2017.
- [17] R. G. Tsushima, R. A. Li, and P. H. Backx, "P-loop flexibility in Na⁺ channel pores revealed by single- and double-cysteine replacements," *J. Gen. Physiol.*, vol. 110, no. 4, pp. 59–72, 1997.
- [18] A. Laio and F. L. Gervasio, "Metadynamics: A method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science," *Reports Prog. Phys.*, vol. 71, no. 12, 2008.
- [19] H. Hassen and P. H. Hunenberger, "Using the Local Elevation Method to Construct Optimized Umbrella Sampling Potentials: Calculation of the Relative Free Energies and Interconversion Barriers of Glucopyranose Ring Conformers in Water," *Wiley Intersci.*,

- 2009.
- [20] J. C. Phillips *et al.*, "Scalable molecular dynamics with NAMD," *J. Comput. Chem.*, vol. 26, no. 16, pp. 1781–1802, 2005.
 - [21] A. Aksimentiev and K. Schulten, "Imaging α -Hemolysin with Molecular Dynamics: Ionic Conductance, Osmotic Permeability, and the Electrostatic Potential Map," *Biophys. J.*, vol. 88, no. 6, pp. 3745–3761, 2005.
 - [22] P. C. Nair, A. K. Malde, and A. E. Mark, "Using theory to reconcile experiment: The structural and thermodynamic basis of ligand recognition by phenylethanolamine N -methyltransferase (PNMT)," *J. Chem. Theory Comput.*, vol. 7, no. 5, pp. 1458–1468, 2011.
 - [23] P. C. Nair, A. K. Malde, N. Drinkwater, and A. E. Mark, "Missing fragments: Detecting cooperative binding in fragment-based drug design," *ACS Med. Chem. Lett.*, vol. 3, no. 4, pp. 322–326, 2012.
 - [24] M. Stepniewski, *Computational studies on membrane proteins and membrane-drug interactions by*. 2016.
 - [25] M. Zahn, S. P. Bhamidimarri, A. Baslé, M. Winterhalter, and B. van den Berg, "Structural Insights into Outer Membrane Permeability of *Acinetobacter baumannii*," *Structure*, vol. 24, no. 2, pp. 221–231, 2016.
 - [26] J. Morán-Barrio, M. M. Cameranesi, V. Relling, A. S. Limansky, L. Brambilla, and A. M. Viale, "The *Acinetobacter* outer membrane contains multiple specific channels for carbapenem β -lactams as revealed by kinetic characterization analyses of imipenem permeation into *Acinetobacter baylyi* cells," *Antimicrob. Agents Chemother.*, vol. 61, no. 3, pp. 1–15, 2017.
 - [27] S. Feijóo-Bandín *et al.*, "Two-pore channels (TPCs): Novel voltage-gated ion channels with pleiotropic functions," *Channels*, vol. 11, no. 1, pp. 20–33, 2017.
 - [28] Y. Sakurai *et al.*, "Two-pore channels control Ebola virus host cell entry and are drug targets for disease treatment," *Science (80-.)*, vol. 347, no. 6225, 2015.

Chapter 2

***In silico* characterization of small molecules
permeation through porins of *A. baumannii***

Introduction

Most of the existent bacteria are harmless or beneficial to the host organisms. However, since the end of the 19th century, many pathogenic species causing many infectious diseases have been established, yet, no real treatment was available or at least known at that time, although the search for treatment has started to interest the microbiological and medical fields when in 1877, Pasteur and Joubert have noticed that the injection of anthrax (*Bacillus anthracis*) bacteria into animals prevented the development of bacterial diseases. This bacterial cell lysis was later attributed, in 1928 by Alexander Fleming, to penicillin substance secreted by the fungi hypha, it was the discovery of the first antibiotic. Antibiotics are molecules naturally produced by microorganisms such as bacteria (especially actinomycetes) and fungi to kill other microorganisms (bacteria) competing with their biotic environment[1]. Nowadays, several families of antibiotics composed of natural, semi-synthetic or synthetic molecules are available, and classified on the basis of their mode of action or with respect to their target. The discovery of antibiotics has revolutionized the modern medicine and infectious disease treatments. Widely used since the Second World War, antibiotics have significantly reduced the mortality associated with infectious diseases such as tuberculosis or plague during the 20th century [2]. However, their effectiveness has been accompanied by their over use, misuse and repeated use in public health [3]. These practices have constrained bacteria to develop defence mechanisms against antibiotics through a selection pressure, that lead to favour and promote the appearance of resistant bacteria strains. Nowadays, the resistance increases become the most dangerous public health, while the development of new drugs stagnates [3] [4][5]. According to the World Health Organization, antibiotic resistance is one of the three most important public health threats of the 21st century (figure 4)[6]. The resistance has become an increasingly massive phenomenon of concern in particular in gram-negative bacteria in contrast to gram-positive bacteria that are less resistance. The origin of this classification name comes from the Gram stain method invented by Hans Christian Gram, in 1884, to detect and identify bacteria[7]. The Gram stain procedure makes gram-positive bacteria appear dark purple and gram-negative bacteria appear pink to red under the microscope, based on the differences in the chemical and physical properties of the cell wall.

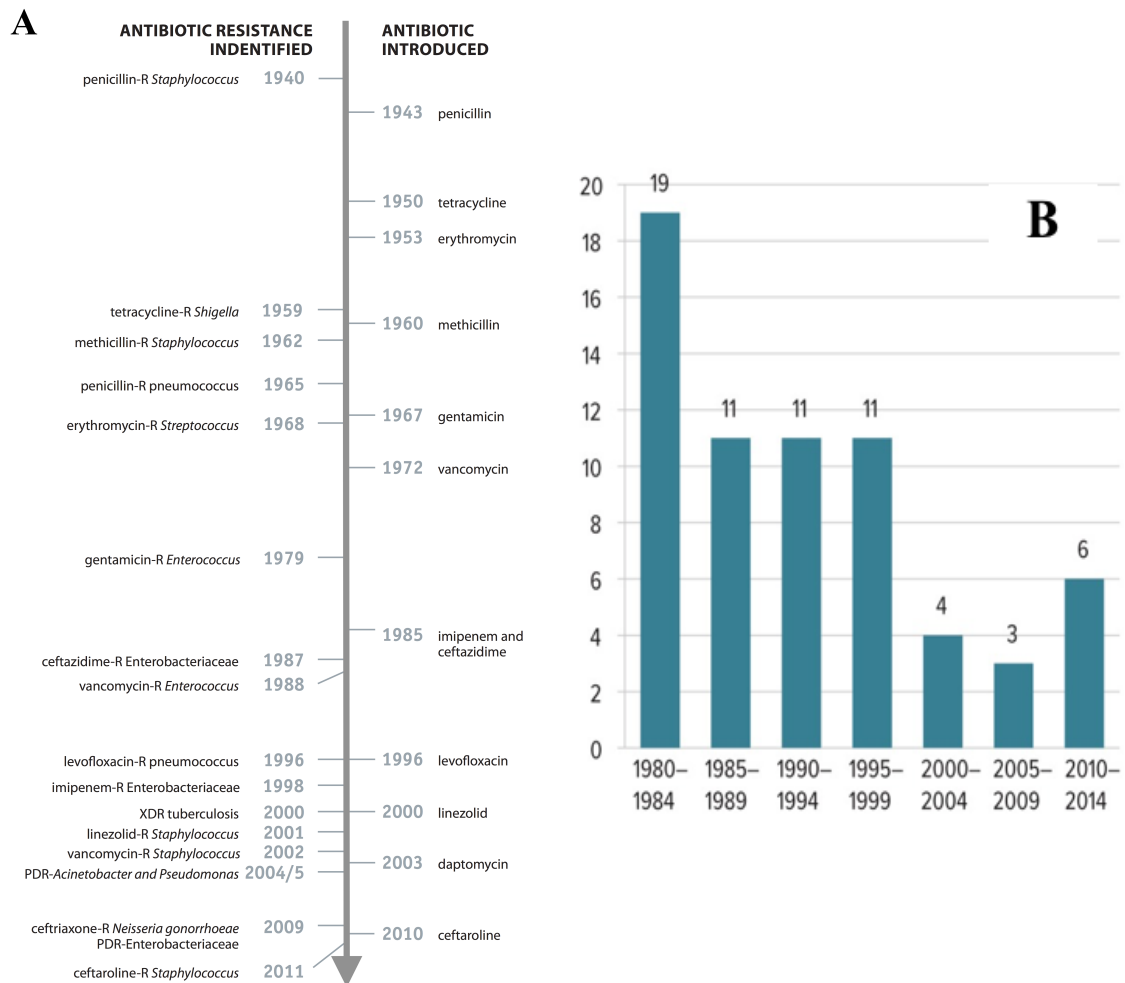


Figure 4: (A) Developing Antibiotic Resistance: A Timeline of Key Events

PDR = pan-drug-resistant; R = resistant; XDR = extensively drug-resistant. Dates are based upon early reports of resistance in the literature. In the case of PDR *Acinetobacter* and *Pseudomonas*, the date is based upon reports of health care transmission or outbreaks. Note: penicillin was in limited use prior to widespread population usage, (B) number of antibacterial new drug application approvals versus year intervals: the number of new antibiotics developed and approved has decreased steadily over the past three decades (although four new drugs were approved in 2014), leaving fewer options to treat resistant bacteria. Adopted from [4]

In Gram-positive bacteria, the wall is thick and dense; due to the presence of layers of peptidoglycan (50-90% of cell wall), which stains the cells as purple. These layers of peptidoglycan are adjacent to the unique plasma membrane layer resulting in the rigidity of the cell wall. In contrast, gram-negative bacteria possess a thin and sparse wall; due to the presence of a thinner layer of peptidoglycan (10% of the cell wall), thus they lose easily the crystal violet-

iodine complex during decolourization with the alcohol rinse, but retain the counter stain Safranin, thus appearing reddish or pink (figure 5).

This group of gram-negative bacteria represent the most dangerous pathogens developing resistance increasingly to a large number of the available antibiotics. Some bacterial strains among gram-negative bacteria group have become resistant to multiple antibiotics, they are

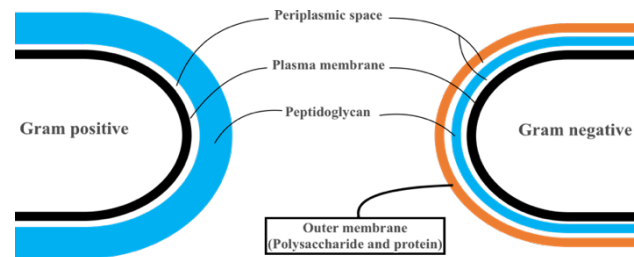


Figure 5: Schematic representation of the main differences between the cell envelope of gram-negative and gram-

called MDR for Multi-drug-resistant and some of the strains are resistant to all known antibiotics, they are referred as “superbugs” [8]. The phenomenon of antimicrobial resistance emergency, is particularly pronounced in Gram-negative bacteria. They possess an additional membrane, the so-called outer membrane (OM), which superposes to the thin layer of peptidoglycan above the plasma membrane. The space in between the former and the OM is called periplasmic space in which are included different proteins (nutrient binding protein, lipoprotein, enzymes...etc.) (figure 5, 6).

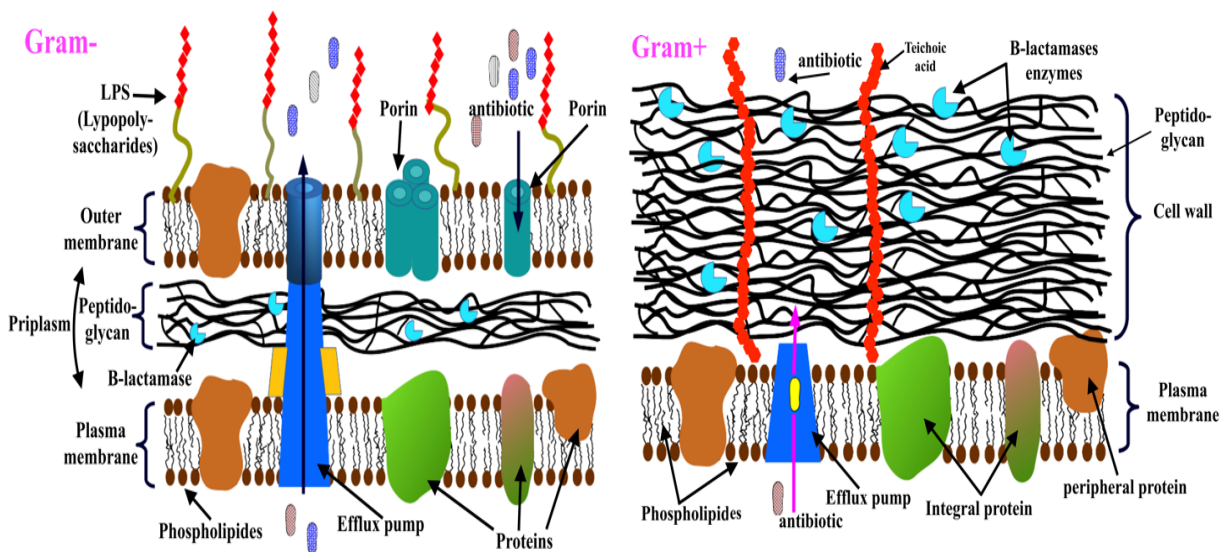


Figure 6: Schematic representation of Gram-positive and Gram-negative cell envelope

The OM is an asymmetrical bilayer of lipopolysaccharides negatively charged (LPS) and phospholipids. The presence of layers of both LPS and peptidoglycan confers a high rigidity to the cell wall slowing the diffusion of hydrophobic molecules. The outer membrane itself act as a protective barrier against the outside environment as well as a selective barrier for hydrophilic molecules. Moreover, several integral proteins are embedded in the OM in particular, water-filled pores so-called “porin” are found in OM of gram-negative bacteria[9][10]. Porins are designed by nature to act as translocation pathways in and out of the cell. They allow the passive passage of polar molecules, such as nutrients and several classes of antibiotics that would not go through the LPS barrier [9]. The resistance of bacteria to antibiotics is demonstrated through several mechanisms (figure 7)[11];

1. The most biochemical effective mechanism of resistance is the deactivation by hydrolysis of the antibiotic by β -lactamase enzymes produced by the bacteria [12]. β -Lactamases are widespread among many bacterial species (both Gram positive and Gram negative) [13].
2. mutation may also have changed the target profile of the antibiotic (molecular target). The drug can no longer attach to it, and becomes ineffective.
3. Efflux pump that are responsible of increasing the exclusion of antibiotics outside the bacteria
4. An additional resistance mechanism is also mediated, specifically in gram-negative pathogens, by reducing the entry of antibiotics into the periplasm.

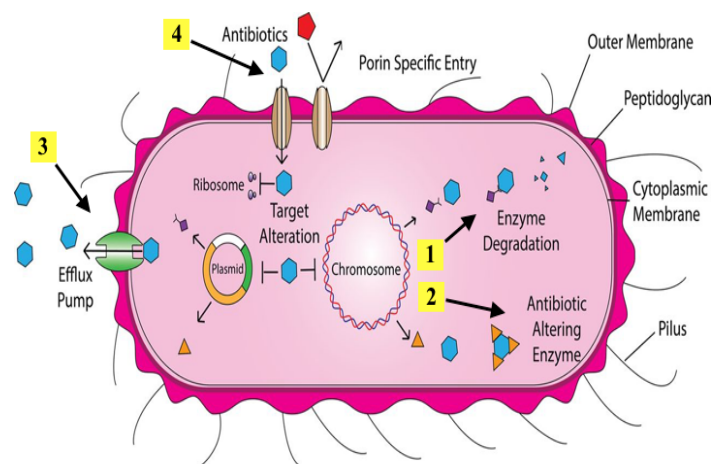


Figure 7: Illustration of major antimicrobial resistance (AMR) mechanisms used by Gram-negative bacteria 1) and 2) enzymatic degradation, target alteration, 3) porin specificity and 4) efflux pumps. Adapted from [8]

These two latter mechanisms consist of reducing the concentration of antibiotics inside the bacterial cell through the net decrease of the outer membrane permeability. In gram-negative bacteria, this occurs by either increasing the exclusion of antibiotics via efflux pumps and/or reducing their influx through porins namely “resistance via porins” [14][15][16]. This latter alteration of membrane permeability acts either modifying the porins size by point mutations and/or reducing the number of porins in the OM [17][10].

The intrinsic resistance via porins

Porins represent the main entry gate for substrates such as nutrient and other molecules necessary for bacteria. Their main function consists in the regulation of the OM permeability through selectivity features of the pore [9][18]. Indeed, the permeability of hydrophilic molecules through the OM is limited by the size so that only small hydrophilic molecules, which size is <600 Dalton [18], are selected to diffuse through, while larger substances are excluded. Several classes of antibiotic targeting specifically gram-negative bacteria include aminoglycosides and carbapenems.

As for natural substrates, polar antibiotics must enter through porins to reach their target, thus any modification of these porins by mutations that can alter the size or the selectivity for certain compounds might increase the capacity of gram-negative bacteria to develop resistance. In addition, bacteria can resist down-regulating the number of porins expressed in the cell, therefore reducing the internal concentration of antibiotics and thus their action on the target. These intrinsic mechanisms of resistance, in particular towards β -lactam, are more developed in bacteria such as *Pseudomonas aeruginosa*, and *Acinetobacter baumannii*, whose outer membrane is rich of many and specific diverse porins, monomeric and small in size, that can be expressed alternatively depending on the condition. In contrast, other gram-negative bacteria like *Escherichia coli*, *Enterobacter species*, and *Klebsiella pneumoniae*, are more susceptible toward β -lactam molecules due to the presence of nonspecific diffusion porins such as OmpF and OmpC, which are trimeric and present a larger size [9][19].

General (non-specific) porins

According to their function, in general, porins are divided into two major groups; (i) the general diffusion porins that are non-specific, and (ii) the specific porins that allow the diffusion of selected substrates.

General porins are open water-filled ions channels expressed in the OM of gram-negative bacteria to allow the diffusion of small hydrophilic molecules at high rate. They represent the most studied and characterized porins in contrast to the specific ones. The X-ray

crystallographic structures revealed that these porins consist of homotrimers. Each monomer is composed by 16-stranded anti-parallel beta-barrels that all together form a central pore [20][21]. The β -strands are connected by short periplasmic turns and longer extracellular loops, which are exposed at the cell surface but the loop (L3). This loop folds inward at half the height of the channel and together with the opposite barrel wall forms the so-called constriction zone conferring an hourglass-shape[22][23]. This constriction zone is characterized by the presence of residues with opposite charges, acidic residues on the loop L3 facing a cluster of basic residues on the opposite side of the pore. This segregation of charges creates a local electrostatic field suggested to act as polar filters for molecules during their diffusion through the pore[24]. However, the mechanism of permeation through the porins of the OM is not completely clear and we do not have any rational rule yet to explain why a molecule can permeate [23].

Specific porins

Pathogens like *P. aeruginosa* (PA) and *A. baumannii* (AB) do not possess general trimeric porins but produce a larger fraction of substrate-specific porins, strongly induced to control the selectivity to different substrates. Though, a direct correlation between their expression and their specific role is still under debate [9] [14] [23]. Specific porins are described as β -barrels channels with a 3D folding similar to those of general porins, though with a smaller pore. In some cases, the X-ray structures do not show a permanent open pore. Thus, it is believed that the substrates diffusion is facilitated by the presence of specific binding sites [25] that might also open the pore. Some of these specific porins have an established role in the uptake of amino acids and organic acids [26]. This is the case of OprD from *P. aeruginosa* (PA), specifically expressed to transport basic amino acids. Interestingly, the OprD porin is also known to represent the main path for antibiotics such as imipenem and meropenem [27], because of their structural similarity to arginine, the specific substrate transported by OprD. The specificity of these porins for related classes of molecules seems to be related to internal affinity sites that facilitate their uptake under nutrient-limited conditions [27][26].

Additional specific porins from PA similar to OprD (OprD homologues) were characterized. In total, about 19 crystal structures from PA are available and grouped as the Occ family (outer membrane carboxylate channel) [26]. The Occ family is divided into two subfamilies (OccK and OccD sharing 40-50 % similarity) according to their substrate specificities. For instance, OccD members are known to have a specific role in the uptake of positively charged molecules and OccK are anion selective [26].

The availability of the 3D structure of *Pseudomonas aeruginosa* porins has largely contributed in our understanding of their structure-function relationship and physico-chemical parameters that could explain their substrate specificity. In contrast, the intrinsic resistance via OM porin permeability is poorly characterized in *Acinetobacter baumannii*, yet it is one of the most dangerous pathogens of the gram-negative family posing a real therapeutically impasse treatment due to its ability to rapidly develop resistances to all available antibiotics. In addition, the lack of information concerning the porins of the OM and their structural properties is one of the major limitations of our knowledge of *A. baumannii* intrinsic resistance.

Porins of *A. baumannii*

The gram-negative bacterium *Acinetobacter baumannii* is the most important member among *Acinetobacter* species associated with hospital-acquired infections worldwide [28][29]. It is a successful pathogen responsible for opportunistic infections of the skin, bloodstream, urinary tract, and other soft tissues [30] [31] [32]. The share of nosocomial infections related to AB resistant to imipenem has increased from 2/3% in 2008 to 11% in 2011[3]. A significant increase in the incidence of multidrug-resistant (MDR) strains has raised the profile of this emerging opportunistic pathogen to develop strategy to effectively escape the effects of all available antibacterial drugs [33] [34] [32]. More importantly, this pathogen is equipped with the presence of all existing mechanisms used to develop resistance to antibiotics. Therefore, antibiotic resistance of this pathogen with the other gram-negative species is considered to be one of the biggest public health concern with economic and social implications around the world. As mentioned earlier, one of the major factor in increasing the intrinsic resistance in AB is the reduced permeability of its porins [35]. To date, few porin from AB are reported and little is known about their role as main path for translocation of polar molecules, with respect to porins from Enterobacteriaceae and *P. aeruginosa*. Indeed, for a very long time, only CarO (carbapenem- associated outer membrane protein) was supposed to provide a path for polar molecules in *A. baumannii* [36][37]. However, the CarO structure recently resolved with X-ray crystallography unveiled a very small pore formed by 8 beta strands, which raised some doubt on its role of carbapenem pathway for crossing the OM [27] [38] [39][35] [36]. A more recent study by Zahn et al indicated the existence of a series of 5 OM channels encoded on the AB genome, and these channels were shown to have small pores which may contribute to the low permeability of the outer membrane [40]. Of these porins, four were crystalized and termed OccAB1-4 [40], the fifth being very similar to OccAB4.

OccAB porins

The X-ray structures of OccAB porins show many similarities with *P. aeruginosa* specific porins: they share the monomeric 18-stranded β -barrels shape and external loops folding. In addition, a phylogenetic analysis of *A. baumannii* OccAB1-4 porins amino acid sequences shows higher similarity between Occ family members (33% with OccK, 25% with OccD). Therefore, the specific activity of OccAB porins was found to be similar to that of monomeric porins from PA [40]. Despite their generally similar core topologies, significant differences between the 3D structures of OccAB porins do exist [26]. The two regions with the highest variability are the extracellular mouth, due mainly to the lengths and the conformations of several loops, and the constriction region, created by two additional inserted long loops (loop L3 and loop L7), which folds back at halfway into the barrel. The variability in the structure and in the folding of both loops determines variability in the dimensions and shapes of the constrictions regions so that the size limits of the solute translocation may change from porin to porin. The common characteristic between all of these porins is the presence, along the channel wall, of a “basic ladder” of arginine and lysine which constitute a favourable energetic path for antibiotics possessing a carboxylic group [40]. The liposome swelling assay together with electrophysiology technique showed that OccAB1 is able to efficiently translocate all investigated antibiotics with rates comparable with those of OmpF, while OccAB2-4 are rather anion selective [40]. However, these methods have a finite time scale resolution and do not allow a direct access to the information at atomic scale about the interactions between the small molecules transported and the OccAB porins. Thus, the molecular basis of the mechanism of the translocation of small molecules and their selection by these porins is yet not defined.

In this work, we planned to use In-silico study to reveal the interactions of the small molecules with these channels and the effect of their dynamics on the translocation process. Molecular dynamic simulations are the best method that can meet such time resolution requirement. The long-term goal of this study is to establish the structural and physicochemical parameters that could increase the permeation of small molecule drugs through OccAB porins. This would help the development of new efficient antimicrobial drug development to combat the resistance of AB to current antibiotics.

Treatment options of *A. baumannii* infections

Indeed, *A. baumannii* infections today have few treatment options left; the last option used was based on the carbapenem association with β -lactamase inhibitor such as sulbactam with colistin. In fact, the Carbapenems have been used as the first treatment option against *Acinetobacter*[41] followed later by various antimicrobials including polymyxins, tetracyclines and

glycylcyclines, aminoglycosides, fluoroquinolones. Since the monotherapy option is no longer effective, the strategy based on the combination therapy has emerged. In the case of *A. baumannii* infections, the molecule of sulbactam, a β -lactamase inhibitor, is combined with another β -lactam antibiotic such as Ampicillin/sulbactam and Amoxicillin/sulbactam [42][43]. It has been shown that such combination has a better efficiency because sulbactam is the main active molecule with antimicrobial properties against *Acinetobacter* [44]. β -Lactamase inhibitors, including sulbactam, bind to β -lactamase thereby increasing the susceptibility of *A. baumannii* to the co-administered β -lactam antibiotic. Indeed, most β -lactamase inhibitors do not exert antimicrobial activity if given alone[43]. Sulbactam, however, has been demonstrated to have antimicrobial properties, including against *A. baumannii*, which are thought to be mediated by binding to penicillin binding-proteins[45]. Moreover, it has been reported that sulbactam given in combination with colistin may offer a significant benefit over colistin monotherapy [46][47]. However, resistance have been already emerging against the both treatments. In addition, the high toxicity level generated by colistin indicated that this treatment is not an appropriate option for treating this pathogen [28]. Thereby, novel approaches and new molecules to fighting infections are urgently needed.

OM permeability challenge for antimicrobial drug development

In the field of the drug design development of new molecules having an antimicrobial effect, the OM porins is the promising targets to fight antimicrobial resistance. In fact, the interest for the design of molecules that targets membrane permeability to escape the OM barrier, both for influx and efflux, has particularly increased.

This is mainly due to the presence of porin exclusively in the OM of gram-negative bacteria, therefore the new developed molecules that must pass through these porin to reach their target into the cell will be specific to the OM of the pathogen. This notion of specificity is an important advantage when one wants to develop new effective drugs and at the same time prevent the undesirable secondary effects on the treated subject. In the case of gram-negative bacteria, the presence of a specific OM has two main advantages; First, the antimicrobial molecules will have an effect on the OM pathogens more than that of the plasma membrane of the hosting cells (human or any eukaryotic cell host). Secondly, it would increase the diffusion rate resulting in increasing the concentration of the molecules which will consequently increase the probability of the molecule to reach its target in the cell, thereby, increase the susceptibility of the micro-organism to the treatment. In order to make this new drug development possible, more in-depth

studies of permeability through porins are necessary to elucidate their role in multidrug resistance and further predict new molecule having better permeation.

The aim and objectives of this study

This study aims to contribute in the understanding of the permeation of molecules through the OM in *A. baumannii* specifically via OccAB porins, the major influx system for antibiotics diffusion. Here we take the advantage of having high-resolution structures of the OccAB porins as an opportunity to study at atomic level the basic molecular parameters that govern their specificity towards molecules. To do so we modelled these porins embedded in a membrane-mimetic environment:

- **In the first step, we used standard molecular dynamics simulations**

To investigate comparatively the structural and dynamical properties of the four OccAB porins, by identifying the structural and biophysical parameters related to their atomic structures. In particular, to have information on the size, the chemical and structural properties of the small molecules that would be selected and translocated by these porins, we focused on the calculation of the pore dimension with its fluctuations and the electrostatic characterization of the residues lining the pore. Further, the comparative analysis would assess potential similarities of each OccAB with porins from related organism such as PA, providing information on potential common mechanism of translocation and selectivity of substrates.

- **In the second step, we used metadynamics simulations¹⁶**

1. To investigate the translocation of selected substrates (arginine, glutamic acid and glycine) focusing on molecular scaffolds whose complexity is the simplest possible. In the selection of substrates, we considered simple descriptors most frequently used in drug modelling research field, the charge, size, and shape. Thus, we reconstructed multi-dimensional free energy surfaces to estimate the barrier imposed by each porin of OccAB for translocation of substrates. From the reconstructed free energy substrates, we identified affinity sites of these substrates in the OccAB porins that might accommodate more complex molecules.
2. To investigate the translocation of few antibiotics selected among those used in the treatment of *A. baumannii* infections. The aim was to evaluate the potentiality of these porins to transport antibiotics. The use of atomistic methods allowed dissecting interaction patterns involved in substrates permeability, focusing on both the porins and

the substrates, correlating simple molecular descriptors to transport. Further, it allowed to identify putative antibiotic binding sites that should represent a common path through which the antibiotics can pass [48].

Results and discussion

Investigation of the substrates translocation and their selectivity by OccAB porins

Dynamic and shape of the pore dimension

We characterized structurally the pore by computing the internal dimension of each OccAB channel. In Fig. 8 we reported the profile radius of the pore along the diffusion axis z obtained from the time-averaged surface area accessible to the solvent [48]. OccAB1 is the largest pore with an average minimum radius of 2.4 Å (blue curve in Fig. 8A), located at $z=5$ Å from the center of the porin. Interestingly, OccAB2 shows two constriction regions (red curve in Fig. 8).

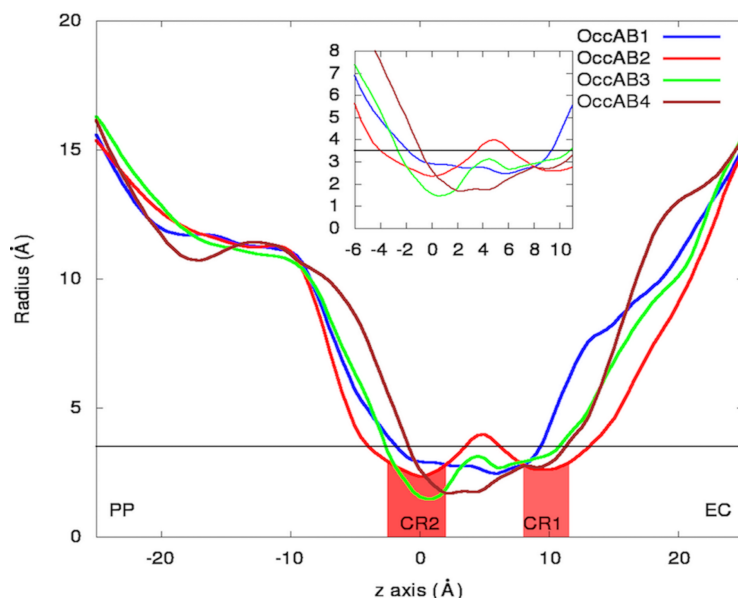


Figure 8: (A) Pore radius profiles along the z -axis of OccAB porins. The radius is obtained from the surface accessible area averaged over the simulations time. “CR” indicates the Constriction region. PP: Periplasm, EC: Extra-cellular

The first constriction zone (CR1) has a radius of 2.6 Å and is located at $z=10$ Å at the entry of the channel (extracellular vestibule). The second (CR2) has dimension of 2.3 Å and is located exactly at the center of the porin, $z=0$ Å. In between the two CRs the value of the radius increases attending a maximum value of ~ 4 Å. OccAB3 with a high similarity with OccAB2 (40%), is the smallest pore with a minimum radius of 1.5 Å. OccAB4 has a minimum radius of approximately 1.7 Å. The size of the minimum radius and its position along the channel axis

as a function of simulation time are shown in fig. S1 A, B, respectively. OccAB2 shows a large variation of the minimum radius, from complete closure to a maximum of 3.2 Å (open state), on average 1.8 ± 0.5 Å. As shown in the table S1, the minimal dimension occurs along z in the range -5 Å to 14 Å, a region covering both CR1 and CR2.

Interestingly, the crystal structure of OccAB2 does not show any open pore, when observing from the extracellular side, because of the curvature of its CR. To quantify this curvature, the CHEXVIS [49] server was used to evaluate the straightness of each channel, whose results are reported in the table 2 and fig. S2. OccAB2 is the most curved channel with the lowest straightness value followed by OccAB1, OccAB3, and finally OccAB4 being the straightest channel (0.89) (fig. S2 E). The length of the external loops largely affects the curvature of the pore. In OccAB2 and OccAB3, the loop L4 is the longest, which function, with the loop L9, as a “lid” hiding the entry of the channel to avoid its direct accessibility to the solvent. This folding of the loop as a “lid” is controlled by the salt bridge Arg173-Glu296 and Arg171-Asp294 between loop L4 and L7 in OccAB2 and OccAB3 respectively. In OccAB1, the loop L4 is shorter whereas the loop L6 is longer making the pore less curved than OccAB2. In OccAB4 the extracellular exposed loops L4 and L9 are shorter than that of OccAB1-3, explaining the straightness of this pore and thus its visibility when viewed from the extracellular side.

With the help of the same server, we evaluated the hydrophobic character of internal pores, reported in fig. S2. OccAB2 and OccAB4 have the more pronounced hydrophobic region in their CRs, consistent with the number of hydrophobic residues lining the channels there (yellow residues in the Fig. 10A). Of note, the hydrophobic region in OccAB2 is located at CR2 ($z=0$ Å), whereas in OccAB4 it is at the entry of the pore from the extracellular side, suggesting a probable hydrophobic barrier for permeation of ions and polar substrates.

The unusual pore flexibility of OccAB2 porin

Due to the curved shape of the pore of OccAB2 and its dynamical properties, we performed additional analysis to better characterize the two CRs and their potential role in the permeation process. We focused our analysis on the fluctuations of extracellular and channel-constricting loops. The negatively charged residues on the loop L7 (Glu296, Glu285) and residues of the basic ladder (Arg21, Arg125, Arg157), contribute to the formation of the CR1 and its electrostatic characterization. Whereas, a salt bridge between the residues Asp291 on the loop L7 and residues Arg368 and Arg389 of the “basic ladder” characterizes the CR2. The calculation of the root means square fluctuations (*RMSF*) of the alpha carbon (fig. 9A) in all OccAB porins, show that the cytosolic N and C termini as well as the turns have greater

flexibility compared with the external loops regions, in agreement with b-factor values. For OccAB2 the most considerable fluctuations among the external loops are for the loop L1, loop L2 and loop L9, where the residues implicated in the flexibility are Glu71-His72, Glu25-Tyr29, Leu377-Thr379, respectively.

From fig. 9B, the channel-constricting loop L7 on OccAB2 shows considerable fluctuation among the 4 porins. This is due mainly to the flexibility of the region around the three-key residue Asp291, Thr292 and Trp293 that contribute to the dynamic of the CR and may affect the translocation events. Indeed, previous single-channel electrophysiology experiments revealed that the OccAB2 channel exists in a non-stable conductance state which makes it difficult to measure a single conductance value, suggesting a correlation with the dynamical character of the loop L7, namely a high b-factors[40] (Fig. 9B). Further analysis of the structural alignment of the loop L7 for all porins reveals major substitutions (figure 9A). Interestingly, among the key residues (Asp291, Thr292 and Trp293) of the loop L7 in OccAB2, the Thr292, due to its polar character and small size, provides high flexibility around the Asp291 and Trp293; this might explain the high b-factor values. In OccAB1, these residues are replaced by a small polar amino acid (Ser307), an aromatic polar residue (Tyr308) and a non-polar residue (Met309). Together, these residues maintain the stability of the loop L7 through formation of electrostatic interactions with residues of the barrel as well as with water.

In OccAB3, the substitution of the residues Thr292 and Trp293 with Phe290 and Met291 provide a high steric effect reducing its flexibility. In OccAB4, the stability of the loop L7 is maintained due to the substitution of the Trp293 with a small polar residue (Ala294), surrounded by two asparagines (Asn291, Asn293) that are polar and can make hydrogen bonds in the CR. Moreover, due to the probable involvement of the loop L7 in the translocation process of small molecules through OccAB2, we investigated its dynamic properties. To capture its movements along the MD simulations, we applied a PCA analysis to the 1 μ s MD trajectory at 300K, focusing on the three eigenvectors with lowest frequency. As shown in the Fig. 11, the radius profile exhibits an alternate opening/closing of CR1 and CR2, through the movement of the loop L7. The residues more involved in the modulation of pore size are Glu285, Arg157, Arg125, Arg21 in the CR1 and Asp291, Arg368, Thr292 and Trp293 in the CR2. As shown in Fig. 11, the anti-correlated movements mapped by the eigenvectors are limited for CR2 and quite important for CR1. When CR1 reaches maxima values, CR2 decreases with less intensity. In contrast, when CR1 reaches minima values, CR2 opens slightly without reaching the maximum opening of CR1. The most important structural motions of the

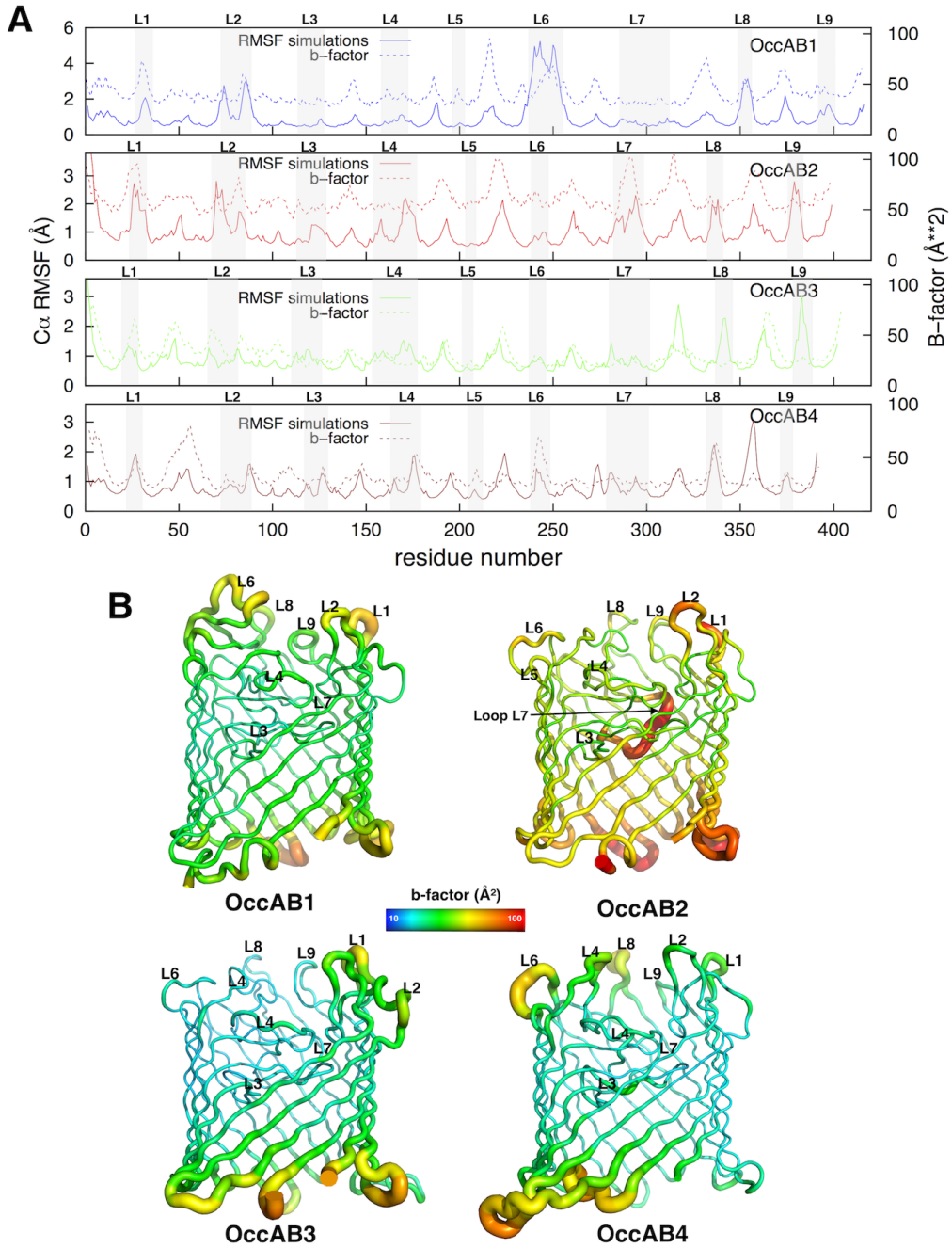


Figure 9: The RMSF ($\text{\AA}$) of the backbone atom (C-alpha) averaged per residue during simulation of molecular dynamics are plotted indicating the structure stability of OccAB1-4 porin. The grey regions highlight the extracellular loops, labelled from 1 to 9. B) The b-factor ($\text{\AA}^2$) of backbone of each crystal of OccAB represented as the color and thickness of the tube (from blue = 10 to red = 100).

three first eigenvectors are represented in fig. 11a. Due to the probable implication of these residues in the translocation events, the opening/closing are related to movements of the loop L7 (Fig. 11b). From these findings, OccAB2 behaves similarly to the OccK1 porin, a specific channel belonging to OprD family of PA, exhibiting different sub-conductance states[50] [51]. An in silico study identified two residues as responsible for the flexibility of loop L7, Phe291

and Arg284 in OccK1, and able to modulate the ionic current[52], these residues correspond to Glu285 and Trp293 in OccAB2 respectively (fig. S5C).

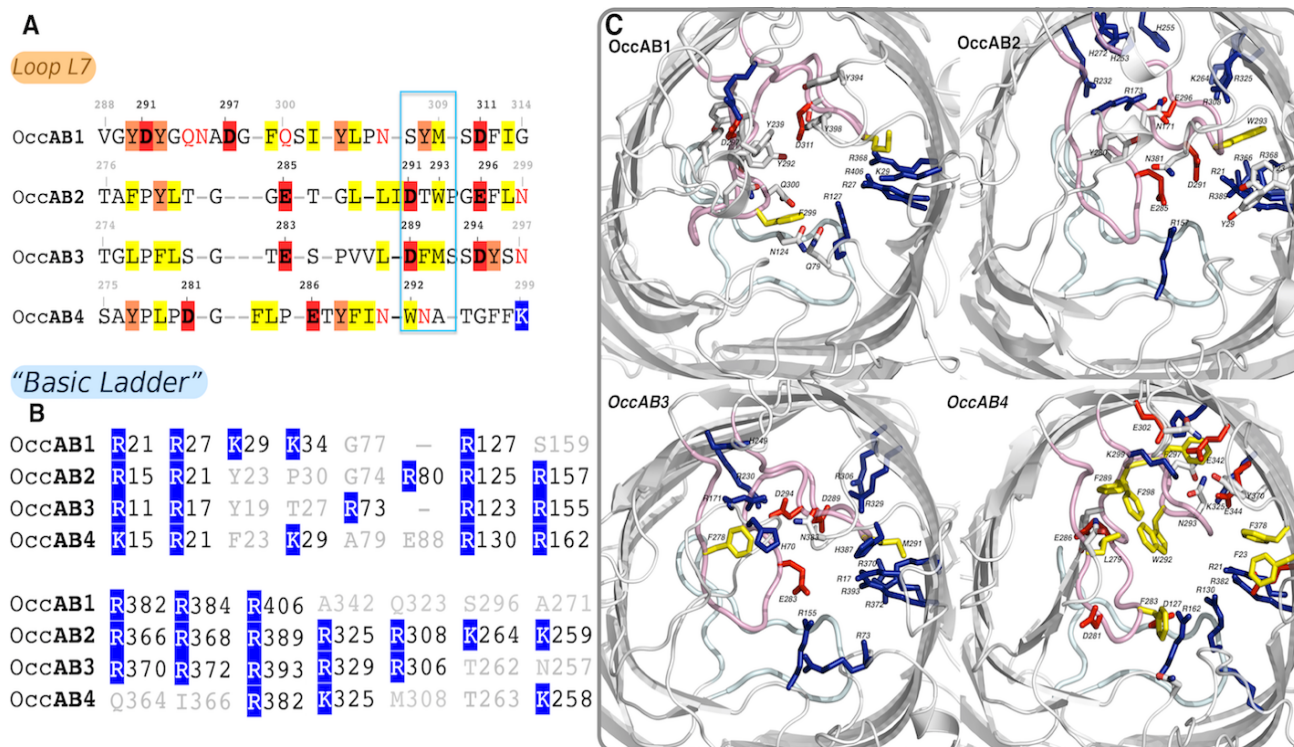


Figure 10: (A) The alignment of the loop L7sequence of OccAB porins; (B) listed residues of the “basic ladder” in OccAB1-4 (C)Top view of the crystal structure of OccAB1-4 porins, the loop L7 represented as tube in pink colour and basic residues in blue sticks, acidic residues in red sticks and hydrophobic residues in Yellow.

Electrostatic features of OccAB porins

OccAB porins contain many charged amino-acid residues that contribute to create a permanent dipole moment, called macro-dipole[53] [54], that measures the global distribution of electronic charges of the whole protein. Previous studies reported its role in the integration and orientation of OM porins in the transmembrane environment[55]. Moreover, another study has reported the implication of the macro-dipole during the ligand-protein binding recognition at large distance, by guiding the substrate to reach its binding site[56]. In the case of porins, the micro-dipole might guide the substrate to orient toward the proper entry point of the pore as soon as it reaches its extracellular surface. This would increase the diffusion rate and the concentration of the substrate at the mouth. Thus, for each porin and using the X-ray structure, we reported the magnitude and direction of the macro-dipole using the method described in the work of Ripoll D.R et, al[57], see Fig. S3 A, B. OccAB porins have major differences in the magnitude

as well as in the direction of the macro-dipole. OccAB1 has a net charge of -7 and a dipole of 335 D and direction almost transversal, similarly to OmpF (which has 271 D for monomer and 540 D for the trimetric complex). OccAB4 has a net charge of -10 and its dipole is four times higher (1389 D) than that of OccAB1, showing a marked longitudinal direction. OccAB2 and OccAB3 have a net charge of -7 and -2 and a dipole moment of 640 D and 903 D, respectively (Table 2). A comparison with porins from *P. aeruginosa* shows that OccK8 and OccK1 have dipoles of 605 D and 854 D, respectively. The common point between all OccAB porins regarding their macro-dipole is their direction towards the periplasmic side which confronts the above hypothesis on its role (Fig. 3S A, B).

In order to complete the electrostatic description, the local charge density profile along z-axis was calculated and averaged along the simulations time for each channel, as shown in fig. 12, using the method described in the vmd corresponding tool[58]. The profiles can be divided into 3 regions along z-axis; two external wide regions (extracellular for $z > 5 \text{ \AA}$ and periplasm for $z < -5 \text{ \AA}$) and the constriction region for $(-5 \text{ \AA} < z < 5 \text{ \AA})$. In the extracellular side, the figure 12 shows three negatives charge-density peaks in both OccAB2 and OccAB3, and only one peak in OccAB1 and OccAB4. For OccAB4 the negative peak is widely extended along the z axis (charge density $\sim -3 \text{ e/\AA}^3/10^4$).

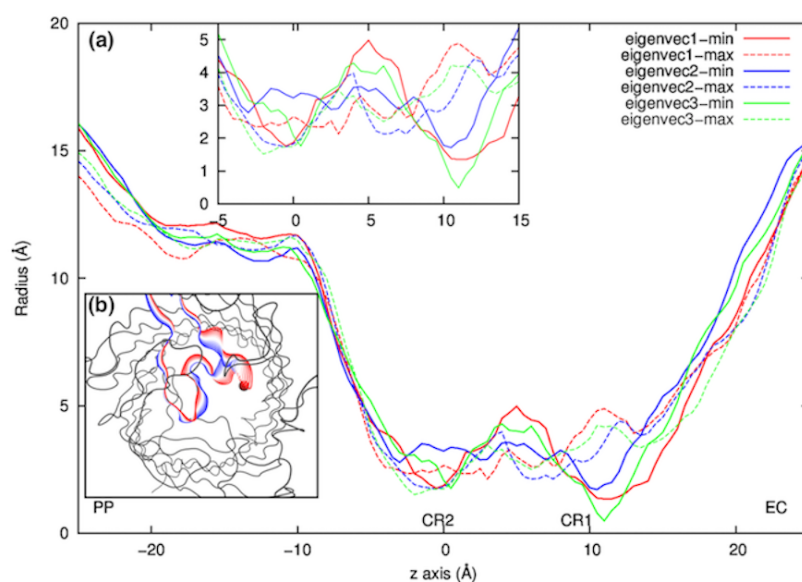


Figure 11: (a) Pore radius profile along the z-axis of OccAB2 porin of the MD motions along the 3 first eigenvector obtained from PCA analysis. (b) representation of 25 projections (from blue to red color) Of OccAB2 porins backbone of the loop L7, of the MD motions along the first eigenvector obtained from PCA analysis for the fluctuation (RMSF).

The strong negative charge density in the extracellular side in all the OccAB porins indicates similarity with the negative environment of the LPS region that have a role of stabilizing the OM proteins integrated in the membrane[20]. It might as well correspond to affinity sites for positively-charged groups or substrates while translocating through the channels if we consider the effect of the electrostatics interactions at long distances. In the periplasmic side, the charge density profiles show rather positive charges for OccAB3 and OccAB4 and an oscillation of small positive and negative peaks for OccAB1 and OccAB2, indicating neutrality. In the constriction region of OccAB1, OccAB3 and OccAB4, the profile of the charge density shows two separated positives peaks with a negative peak at $z=0$ Å. Contrarily, the OccAB2 charge density profile shows rather small positive peaks, suggesting a pore prone to select for negative substrates (see below). The macro-dipole moments agree with the distribution of charges. In all OccABs but OccAB1 the negative residues are mainly pronounced in the extracellular top of the porins while the end periplasmic side is mainly positively charged. Moving from OccAB2 to OccAB3 and OccAB4, we see a progressive increase of charge separation, with more negative residues on the extracellular side, and more positive on the periplasmic space.

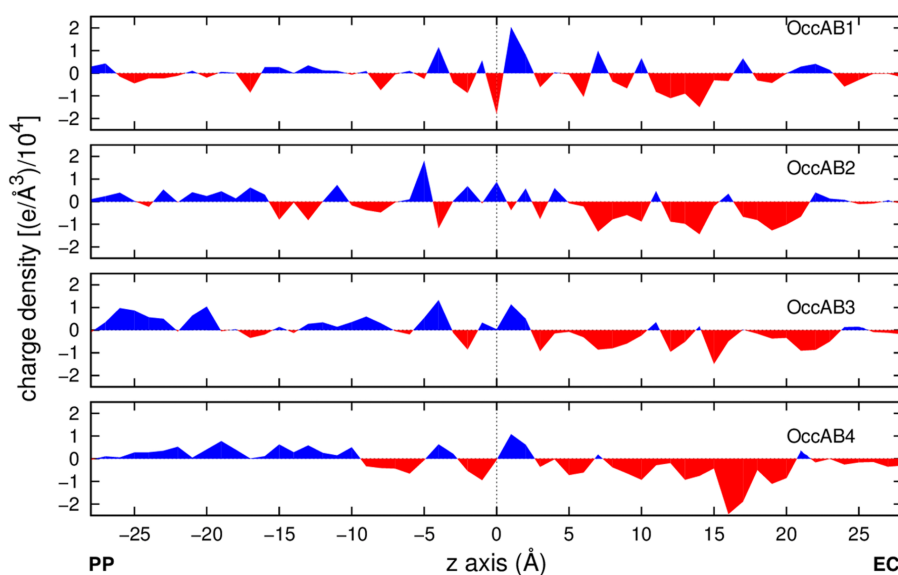


Figure 12: One dimensional charge-density profiles along z-axis of OccAB porins, charge density values are averaged over the simulations time. PP: Peri-plasm, EC: Extra-cellular

In order to characterize the charge distribution of the surface area exposed to the solvent in the pore, we calculated the electrostatics map [59] [60] [61] of all OccABs porins, using two structures: the crystal structure and the representative structure of the most populated cluster obtained from the MD simulations. This allows us to see the difference in the charge distribution and the effect of the

dynamics of the residues lining the pore on the charge distribution. As shown in the fig. 13, the differences in the electrostatic map between the crystals and the representative structure of the MD simulations (fig. 13A, D) are minimal but for OccAB2. Here we see a large change of the distribution of the negative and positive charges between the two structures. In the crystal, a negative charge distribution is present in the CR (Fig. 13B. 1) while the representative structure shows rather positive charges lining the CR (Fig. 13B. 2). This is due to the dynamics of the pore and to the fluctuation of the residues lining the pore.

Looking more in details to the two elements characterizing the constriction region, the loop L7 and the basic ladder, the conservation of charges highlights interesting electrostatics features. The basic ladder of OccAB2 is the largest with 12 positively charged residues, followed by OccAB3 having 10 residues and finally, OccAB1 and OccAB4 possess only 8 positively charged residues (Fig. 10B). The difference in the positive charges agree with the charge density, showing the CR of OccAB2 rather positively charged during the MD simulations. On the other hand, the sequence alignment of the loop L7 shows conservation of three negative charges in all OccAB porins but OccAB4 that exhibits a substitution of one negative charge with an asparagine. In addition, the exact position of negative charges differs between the porins: they are more spaced in OccAB1 and more confined and exposed to the basic ladder for possible electrostatic interactions in OccAB2 and OccAB3 (fig. 10 A, B, and C), that might explain the closure of these two porins and its effect on the process of substrates translocation.

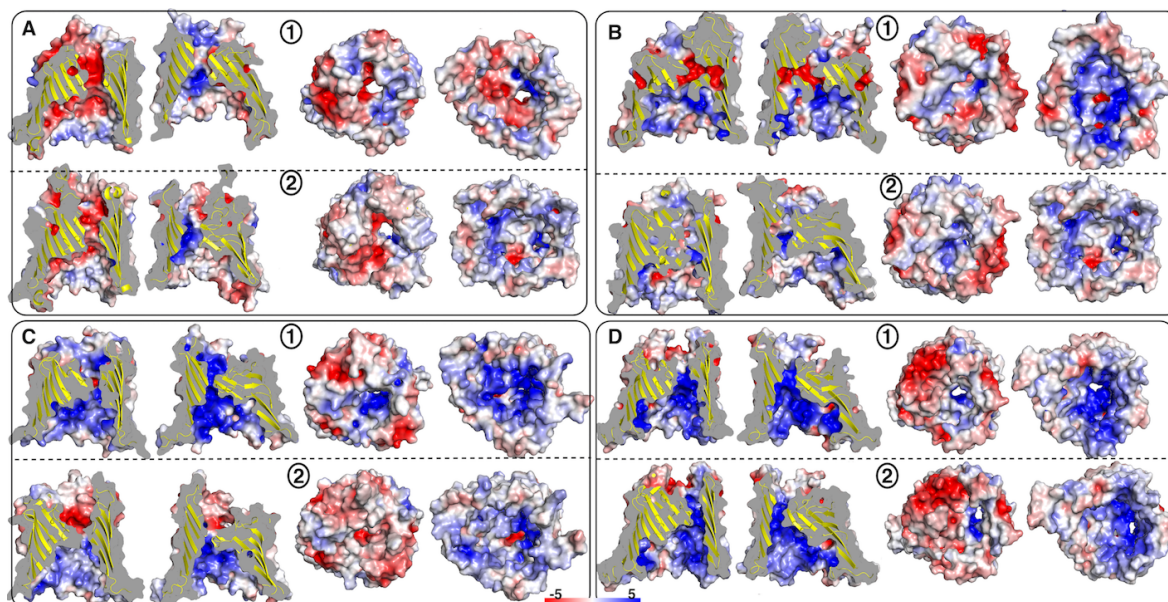


Figure 13: Electrostatic potentials from negative ($-5kT/e$ in red) to positive ($+5kT/e$ in blue) mapped, using the programs PDB2PQR^{31,33} and APBS³², on the surfaces of OccAB1 to 4 porins (A, B, C and D respectively), On the crystal structures (top labelled (1)). On the representative structure of the most populated cluster (bottom labelled (2)).

Investigation of the substrates diffusion and free energy calculations

In order to study at atomic level the specificity of these four porins for the uptake of molecules based on their size and charge, we selected three molecules: (i) arginine as a basic amino acid (arginine, with an ϵ -group and a δ -guanidino group, respectively, which are protonated at physiological pH), (ii) the negatively charged glutamic acid, and (iii) glycine as the smallest molecule having one aliphatic carbon. The arginine and glutamic acid were selected to investigate the impact of charge, while glycine is selected to assess the impact of the size.

The translocation of substrates through OccAB1

The free energy surface (FES) of Arginine's translocation through OccAB1 shows clearly two minima (minima 2 and 3 in the FES of the fig. 14A) in the constriction region with z in the range [0-5] Å (Fig. 14A). In these minima, Arginine enters pointing down either the amino group (NH_2) interacting with a negative pocket (Tyr292 and Asp311) or the carboxylic group

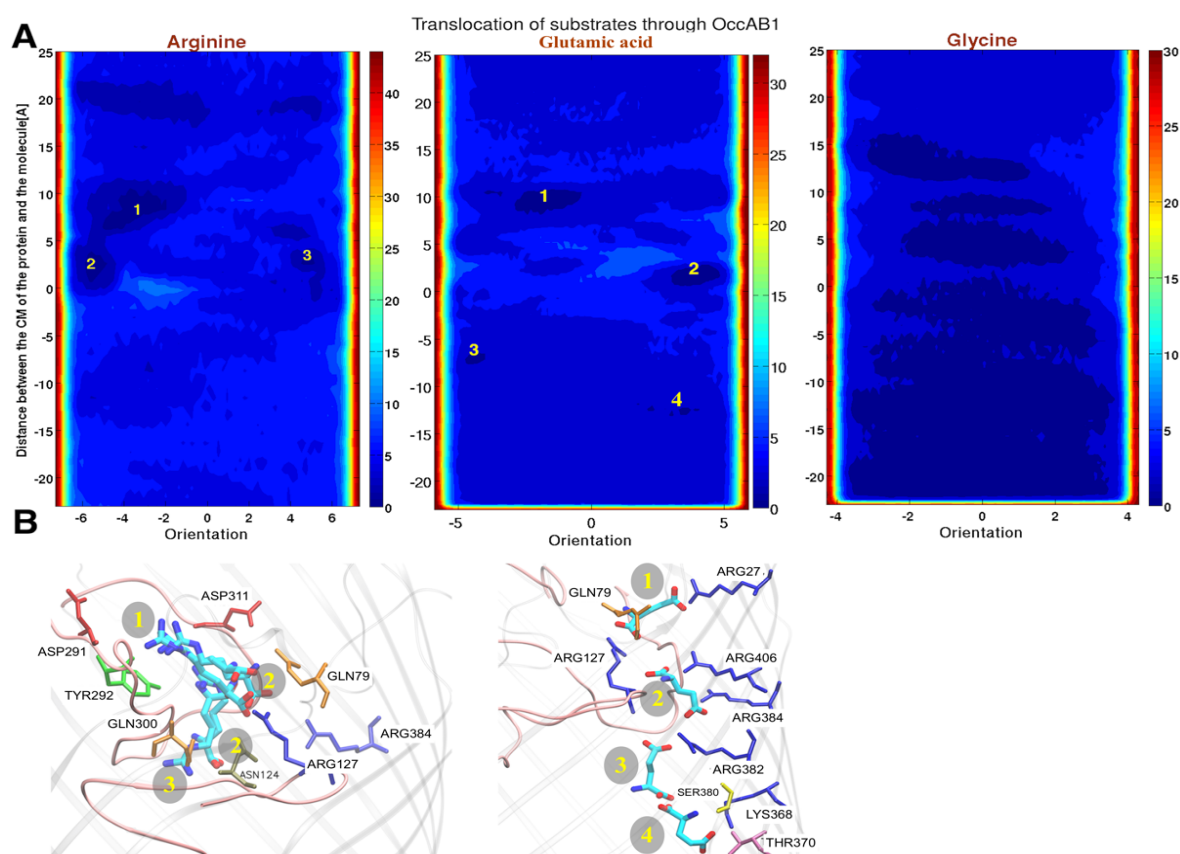


Figure 14: (A) FESs of Arginine, Glutamic. Acid and glycine showing the translocation path through OccAB1 porin. The minima are enumerated in yellow. (B) Top front view of OccAB1 in complex with each molecule in the main minima displayed in the figure FES, the molecules are shown in sticks representation with nitrogen coloured in blue and oxygen coloured in red at their most stable interactions. The pink tube highlights the internal loops L3 and L7. The residues involved in the minima energy are displayed with residue number.

(COO) favouring the interaction with the basic ladder of positively charged residues (R21, R27, K29, K34, R127, R382, R384, R406), in particular the residues R127/R406/R384 (Fig. 13B). glutamic acid is displayed in Figure 14A. Two main minima are recognizable corresponding to 2 affinity sites with different orientation; the first minimum, located at $z=10$ Å (just above the constriction zone), corresponds to the glutamic acid entering with COO group of the backbone pointing down to interact with the basic ladder. In the constriction zone, glutamic acid changes its orientation and diffuse maintaining interactions of the side chain with the basic ladder (Arg127/Arg382) (Fig. 14B); this corresponds to minimum 2 in the free energy surface. The energy barrier is ~ 4 kcal/mol, like for arginine. The glycine is simulated to represent small polar molecule. According to its small size that fits largely with the radius of the channel, it adopts all possible orientations showing a flat free energy surface, reported in Fig. 14A.

The free energy profiles of all the three substrates translocating through OccAB1 channel indicate the presence of a very weak energy barrier and similar when comparing the substrate with opposite charges (with $\Delta G \sim 4$ kcal/mol). This suggests that OccAB1 clearly does not discriminate charged substrates. The reason for that is probably due to the symmetric distribution of the opposite charges in the CR so that this porin interacts with both positive and negative substrate. The free energy surfaces for translocation through OccAB1 indicate a similar free energy barrier for arginine and glutamic acid as low as ~ 4 kcal/mol compared to ~ 2 kcal/mol for the small glycine. The lack of size selectivity is due to the large size of OccAB1 relatively to the size of the molecules under study. These results on OccAB1 are in agreement with the liposome swelling assay and electrophysiological experiment that indicate OccAB1 as not selective and large channel[40]. However, while it was supposed to be similar to *OprD* from *P. aeruginosa*, which selects strongly for basic amino acids, we see that our analysis point to a lack of strong selectivity.

The translocation of substrates through OccAB2

Figure 15 shows the FESs of the substrates translocating through OccAB2. The FES of arginine in complex with OccAB2 featured 2 minima; the first large minimum at the first constriction region corresponds to a strong interaction of the carboxylic (COO) group with Arg389 and Arg157 of the basic ladder and the NH₂ group pointing down towards the periplasmic side to interact with the residues of the negative pocket, especially Glu285, Asn77 and Thr286. The positive values of the variable “Orientation” in the FES indicate this orientation of arginine. This position of the minimum along the z-axis ($z = [5:10]$ Å), qualitatively correlate to the charge-density distribution profile of OccAB2 along z-axis showing a negative charge site at

the same position, which could explain the strong binding of arginine in the FES. The second minimum corresponds to a binding of the amino group (NH) of arginine with the Glu296 residue on the loop L7, while the carboxylic group is pointing down towards the periplasmic side interacting with Arg389 and Arg368 of the basic ladder, as supported by the negative value of the orientation in the associated FES. The transition from the first minimum to the second requires a change in the orientation of the molecule with a cost of $\Delta G \sim 2$ kcal/mol.

Finally, the third minimum is located below the CR2 at $z = -10$ Å and corresponds to the interaction between COO group of arginine and Arg308, Arg325, Arg368 residues of the basic ladder (R15/R21/R80/R125/R157/R173/R389/R368/R366/R308). The transition from the second to the third minimum requires $\Delta G \sim 8$ kcal/mol. The FES for glutamic acid is characterized by an energetic barrier of ~ 4 kcal/mol, moving inside with the side chain COO group ahead. The key residues involved in its minimum energy are R21, R157, Y23, T282, S168 and Y23. Despite its small size, glycine exhibits a single minimum energy in the constriction region located at z in the range $[0:5]$ Å which corresponds to the molecule entering

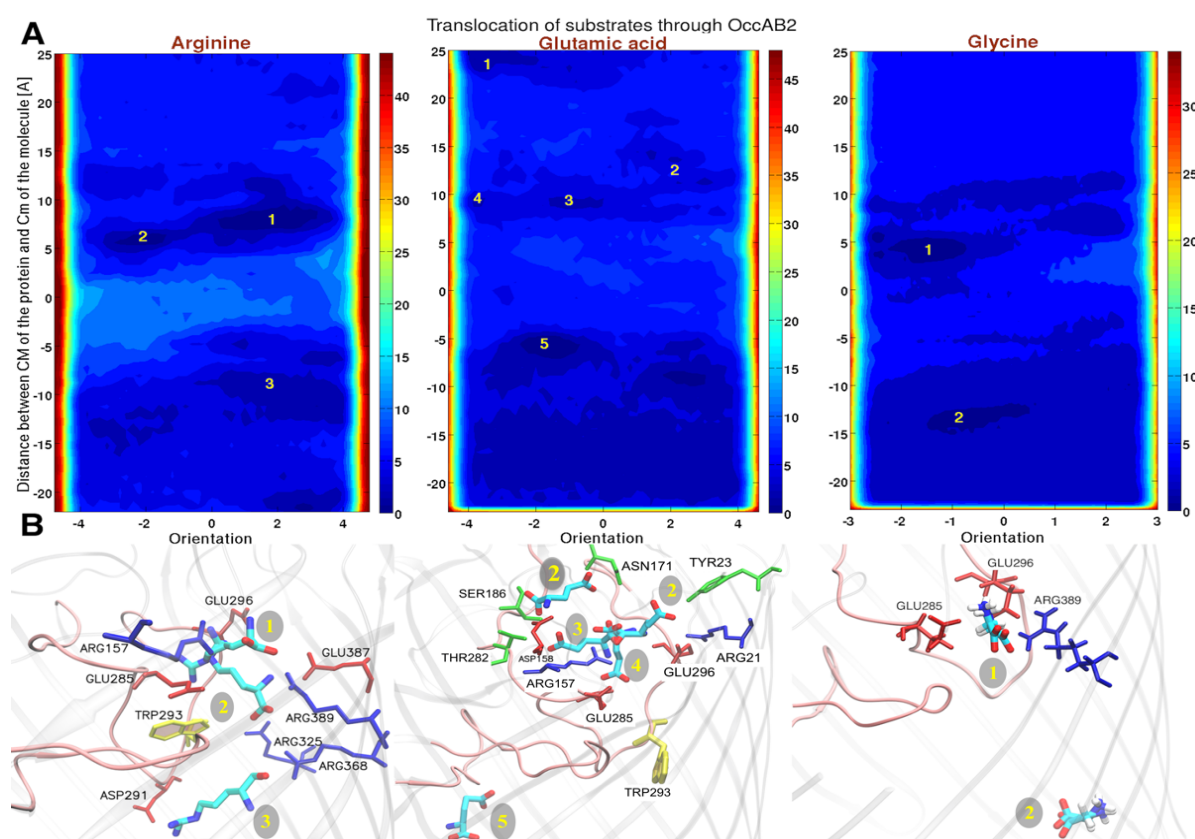


Figure 15: (A) Free energy surfaces (FESs) showing the translocation path of each substrate in OccAB2 porin (each color is 2 kcal/mol). The minima energies are enumerated in yellow, (B) The front view of OccAB2 in complex with each molecule in the main minima displayed in or near the CR of the figures FESs. The substrates are shown as sticks with nitrogen colored in blue and oxygen colored in red. The pink tube highlights the position of the internal loops L3 and L7. The residues involved in the minima energy are displayed with residue number.

with the COO group and interacting with Arg389 of the basic ladder and Glu296, Glu285 of the negative pocket, overcoming the CR1 and CR2 with $\Delta G \sim 4$ kcal/mol. The case of glycine diffusing with preferable orientation (with its COO group) in OccAB2 indicates the preference of this channel for negatively charged molecule. All substrates share a common COO group orientation path to translocation through OccAB2 (fig. 15A, B). Most interestingly, the presence of two barriers, at $z=10$ Å and $z=0$ Å indicated in the radius profile, has an effect in the FES which confirms the existence of two constriction regions and their importance in the translocation process of small molecules in OccAB2. The both electrostatic feature of CR1 and the flexibility of the loop L7 via the residue Thr291, Trp292 and Asp291 affect the substrates diffusion through OccAB2.

The dynamic of the loop L7 in OccAB2 involves the fluctuation of the residues Trp292 and Asp 291 affecting drastically the mechanism of the translocation of a small molecule. The RMSF and the radius profile calculated from the simulation of OccAB2 show a structural reorganization of the constriction region through the formations and disruption of hydrogen bonds between the residues lining the CRs. An alternation in opening-closing states of the pore caused by the fluctuations of the residue 293 (Trp293) as well as the formations alternatively of three electrostatics interactions (Arg157-Glu296, Arg125-Glu285, Arg368-Asp291) at the entry of the channel, are shown by both the cluster and PCA analyses (Fig. 11). The flipping of the Trp293 affects drastically the motion of the loop L7 that causes the variation in the radius profile thereby the opening or closing of the pore as suggested by the high b-factors of the loop L7 in the crystal structure. This finding indicates that not only the size and the electrostatics features are important for the translocation events but also the structural dynamic of the pore play a key role. The Loop L7 flexibility could be a key element in the opening of the channel and could explain the electrophysiology experiments described in previous study[40]. These results indicate that the substrates translocation through the porins could be affected by increasing the flexibility of the key residues forming the CR such as the residues Asp291, Thr292 and Trp293 of loop L7 in OccAB2.

The translocation of substrates through OccAB3

Similarly, to OccAB2, the OccAB3 porin conserves negatively charged residues Asp294, Glu283, Asp289 on the loop7 and Asp166 on the loop L4 forming a negative pocket opposite to the 'basic ladder' composed by (R11, R17, R73, R123, R155, R370, R372, R393, R329, R306) see figure 16B. These residues are supposed to play a role in the substrates translocation.

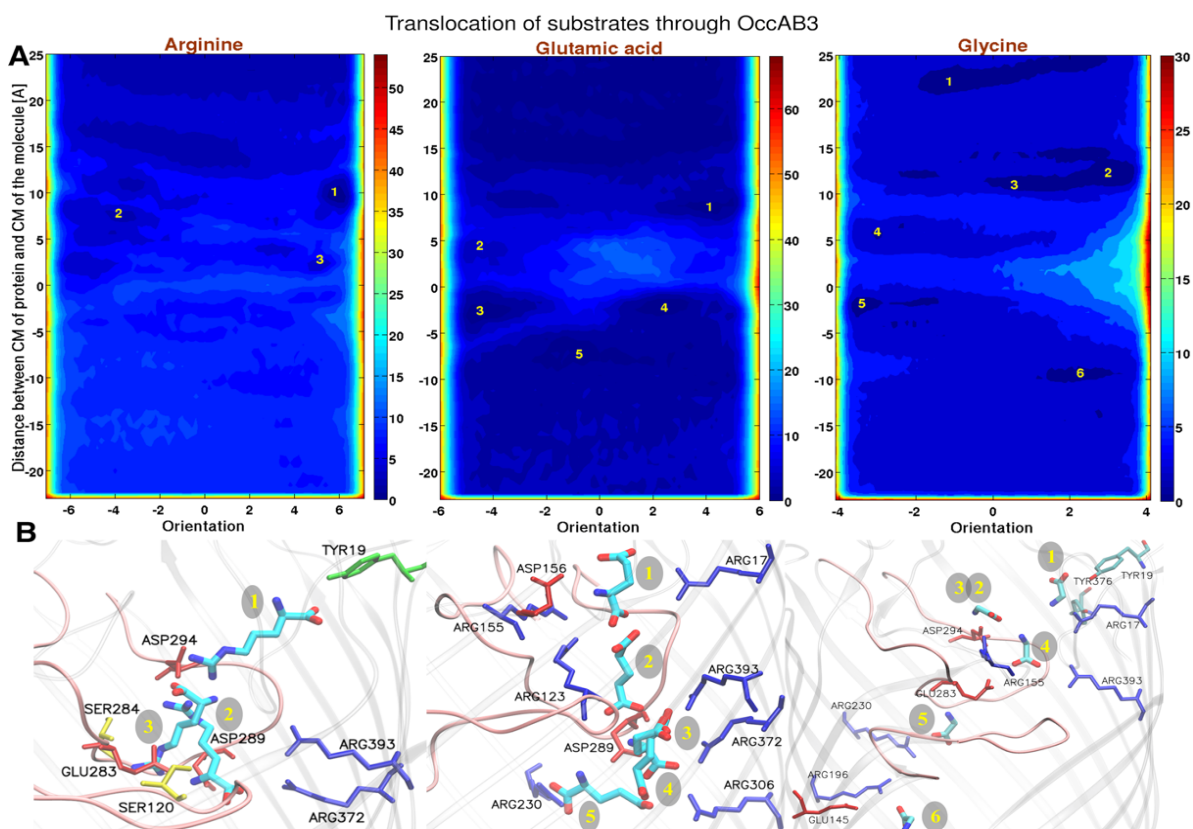


Figure 16: (A) Free energy surfaces (FESs) showing the translocation path of each substrate in OccAB3 porin (each color is 2 kcal/mol). The minima energies are enumerated in yellow, (B) The front view of OccAB3 in complex with each molecule in the main minima displayed in or near the CR of the figures FESs. The substrates are shown as sticks with nitrogen colored in blue and oxygen colored in red. The pink tube highlights the position of the internal loops L3 and L7. The residues involved in the minima energy are displayed with residue number.

In addition, as mentioned above, the second main similarity between OccAB2 and OccAB3 is the presence of the large external loop L4 acting as a ‘lid’ that contributes in decreasing the size of the CR1 at the mouth of the channel. The ‘lid’ in the OccAB3 is stabilized by an interaction of its residues with that of the loop L7 and the loop L9. Especially the formation of the hydrogen bonds between Arg171 (on the loop L4) and Asp294 (on the loop L7) and the interaction between Asn169 (on the loop L4) and Thr389 (on the loop L9). On the other hand, the presence of the residues Phe278 and Phe386 on the loop L7 and loop L9 respectively, form an aromatic stacking group at the mouth. This might negatively affect the translocation of polar substrates by decreasing the size and the polarity of the pore (Fig. 16 A, B, and C).

The FES for the translocation of the three molecules, arginine, glutamic acid, and glycine, through OccAB3 are summarized in fig. 16.

The FES of arginine shows that all energy minima are in the extracellular side of the protein at $z > 0$ Å of the FES which reveals favourable interactions of arginine in this area of the protein. This is in agreement with the negative surface electrostatics map of OccAB3 (Fig. 13) as well as with the charge-density profile of OccAB3 showing a preference for negative charges in the extracellular side (fig. 12). In the constriction region with z in the range $[0:5]$ Å, arginine enters with the NH group pointing towards the periplasmic side. In this minimum, arginine forms electrostatic interactions with the negatively charged residue Asp294 and Tyr19 making a stable conformation. To translocate, the molecule must change its orientation with the COO group ahead. Yet, the energetic barrier is high around ΔG 6~ kcal/mol.

The FESs of the both glutamic acid and glycine show a clear preferred orientation with the negatively charged group (COO) ahead (Fig. 16A, B), with a barrier of ΔG 6~ kcal/mol and ΔG 4~ kcal/mol respectively, in good agreement with their size. These results are also in consistence with the positively charged electrostatic potential surface of the CR, as shown in the Fig. 13C. Despite the high similarity between OccAB2 and OccAB3, the translocation mechanisms of the selected substrates are slightly different due to the differences parameters that are not necessary revealed by a sequence alignment, such as the electrostatic potential, the dimension of the CR and the dynamic of its residues.

The translocation of substrates through OccAB4

One of the structural features that could favour the diffusion of a small molecules in OccAB4 is the shortness of its external loops. Indeed, this favours the exposure of the CR to the external environment, which is located at $z = [2: 5]$ Å. Also, the basic ladder in this case is shorter compared to the other member of OccAB pore with only six positive residues grouped in the extracellular side of the CR at $z > 0$ (fig S5D). As mentioned earlier, the loop L7 of OccAB1-3 possess three negatively charged residues forming a negative pocket. In OccAB4, one of the negative charges is missing. Moreover, the arrangement of the two conserved negatively charged residues on the loop L7 is different from those in OccAB1-3. In OccAB4, the negative charges are located next to the basic ladder rather than in front of it (fig S5 A, B, D). These structural features of the loop L7 impact on the formation of an evident negative pocket that should affect the substrates translocation through this channel.

Regarding the translocation of arginine through OccAB4, the FES revealed two strong minima (Figure 17). The first minimum corresponds to the amine (NH₂) group of the arginine interacting with a negative residue (Asp281) of the loop L7 which tends to be more exposed to the zone of the extracellular solvent at $z \sim 15$ Å (fig. S5 A). The second minimum is located at

$z = [5: 10]$ Å. The stability of arginine in this minimum is due to the interactions of its carboxylic group (COO) with the basic ladder (K15, R21, K29, R130, R162, R382, K325, K258) and the interaction of the amine group (NH₂) with the residues Glu302, Glu342 and Glu344 (fig. S5 A). These residues, located on the beta barrel behind the loop L7, are conserved in all *baumannii* porins except for OccAB3, which shows a deletion of one negative charge. In OccAB1-3, these barrel residues are hidden from the external side because of the length of their external loops (loop L6 and L9 in OccAB1 and loops L4 and L9 in OccAB2-3). Whereas, in OccAB4 they are more exposed to the extracellular side (figure S5 A, B, D) due to the shortness of its external loops. Therefore, they act as an alternative negative pocket (“barrel negative pocket”) to the missing negative charges on the loop L7 in OccAB4.

Indeed, despite the very low flexibility of the loop L7 in the RMSF analysis, the Trp292 revealed different conformation to that of the crystal structure, through slight movement during the MD simulations, as supported by the Phi and Psi angles values shown in the fig. S4. This behaviour of the Trp292 during the translocation affects the formation of the CR2 with z in the

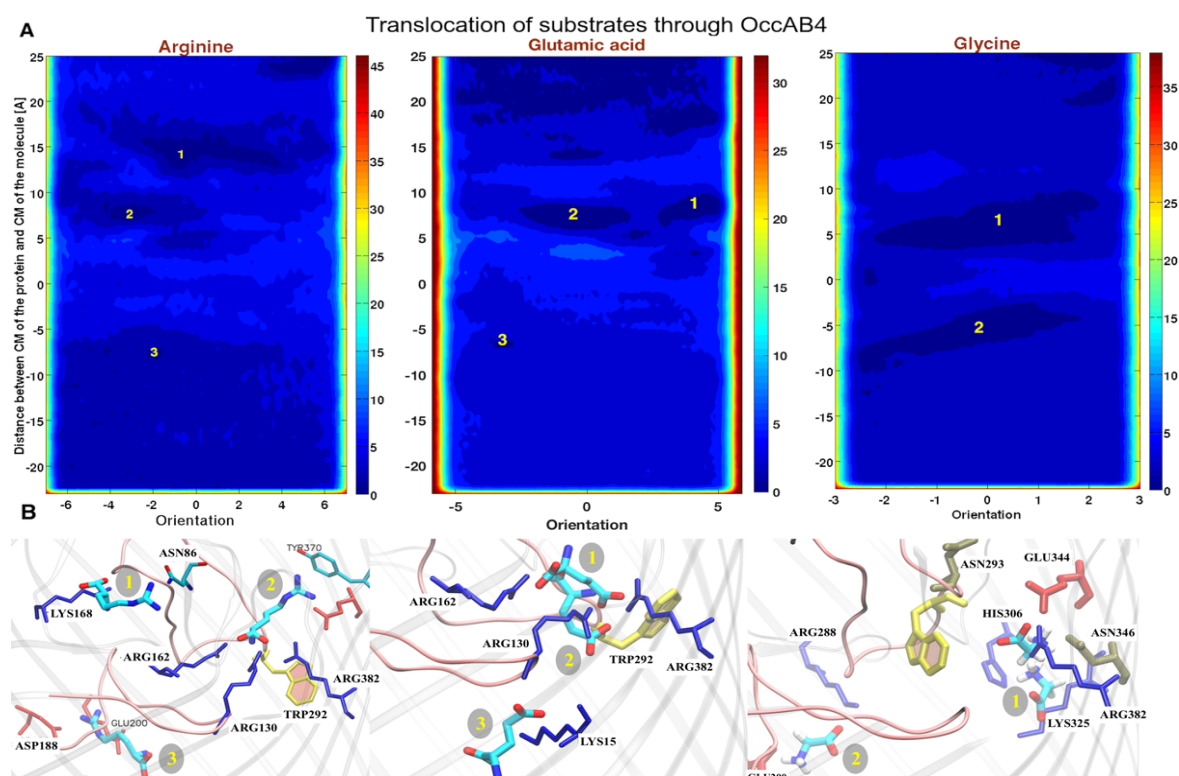
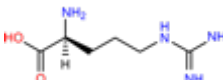
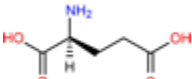
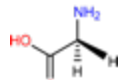


Figure 17: (A) Free energy surfaces (FESs) showing the translocation path of each substrate in OccAB4 porin (each color is 2 kcal/mol). The minima energies are enumerated in yellow, (B) The front view of OccAB4 in complex with each molecule in the main minima displayed in or near the CR of the figures FESs. The substrates are shown as sticks with nitrogen colored in blue and oxygen colored in red. The pink tube highlights the position of the internal loops L3 and L7. The residues involved in the minima energy are displayed with residue number.

range [-4: 4] Å in the FESs. In this region, there is no minimum energy observed in the FES of arginine translocating through OccAB4 similarly to the FES of arginine in complex with OccAB2. The lack of affinity sites in this region of OccAB4 is mainly due to the small size of its CR whereas, in OccAB2, it is related to the shape and the hydrophobicity of the CR. Thus, the free energies needed for arginine translocating through OccAB4 is lower (~6 kcal/mol) compare to that in OccAB2 (~8 kcal/mol) (table 1). The accessibility of the barrel negative pocket to the entry of OccAB4 porin might explain the two strong minima in the FES of arginine. In addition, the entry pore of OccAB4 is surrounded by a high number of hydrophobic residues (Leu279, Phe283, Leu284, Phe289, Ile290, Trp292, Phe297, Phe298) (highlighted yellow residues in the fig. 10C). This improves the hydrophobic interactions to the detriment of polar ones. In fact, mainly the short basic ladder of OccAB4 establish the very few electrostatic interactions inside the pore giving a strong positively charged electrostatic potential as shown in the fig. 13D.

The FES of glutamic acid shows two large minima at z in the range [5:10] Å, the energy needed to escape the barrier of the CR1 is $\Delta G \sim 4$ kcal/mol. The low free energy of glutamic acid translocation is mainly attributed the straightness of the OccAB4 pore and the positive electrostatic potential of its CR (fig. 13D). The glycine displayed two large minima energy in the FES (figure 17) separated with an energy barrier of $\Delta G \sim 2$ kcal/mol, the first minimum is in the CR at $z=5$ Å and the second minimum is in the periplasmic side at $z= [-4: -10]$. In the both minima, the glycine does not have preference for specific orientation when entering due to its small size. Thus, the hydrophobic region in OccAB4 acts as a barrier (CR1) for the polar substrate translocation through OccAB4 as shown in the FESs at $z= [10:15]$ Å. However, the CR1 in OccAB4 are not visible in the radius profile averaged along the classical MD time, which could be explained by the lack of the extended loop L4 that contributed to the formation of the CR1 in OccAB2.

Table 2: Energy barriers (in Kcal/mol) for translocation of substrates through OccAB porins. MPA: minimum projection area. HD: Hydrogen donor, HA: hydrogen acceptor, RB: rotatable bonds, N_{ca} : number of carbon aliphatic.

Molecule	Charge	N_{ca}	MPA $\text{\AA}^2$	RB	HD	HA	OccAB1	OccAB2	OccAB3	OccAB4
							Free energy (Kcal/mol)			
arginine 	+1	6	30.9	5	4	4	6	8	6	6
glutamic acid 	-1	5	24.3	4	3	5	4	4	6	4
glycine 	0	2	18.2	1	2	3	2	4	4	2

Antibiotics translocation through OccAB1 porin

The previous molecular dynamics simulations study of substrates translocation through OccAB porins confirmed the experimental data that revealed OccAB1 having the highest permeation rate for polar molecules, without discrimination between positively or negatively charged substrates. In addition, OccAB1 is considered up to now the *A. baumannii*'s porin with the largest radius, (2.4 Å), comparable that of the trimeric general porin of *E. coli*, namely OmpF and OmpC.

In line with the investigation of substrates diffusion through OccAB porins, we investigated the passage of selected antibiotics through OccAB1, the supposed OprD homologue, using metadynamics simulations. We considered the zwitterionic ampicillin from the penicillin class,

the zwitterionic meropenem and imipenem from the carbapenem class, and avibactam and sulbactam as β -lactamase inhibitors. The β -lactam ring (Fig.18) is a common structural feature of the class of β -lactam antibiotics, including penicillin derivative, cephalosporins, carbapenems, monocyclic β -lactam...etc. Most of these antibiotics act by affecting the bacterial cell wall of gram-negative bacteria. From the structural point of view, we considered antibiotics with small differences: (1) the double/single bond on the five-membered ring that modulate

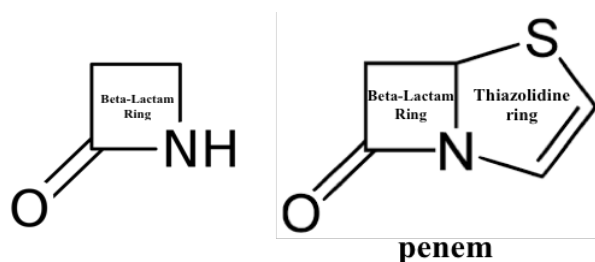


Figure 18: β -lactam structure representation and penem complex ring as a common structure in the β -lactam class

ring strain, (2) the sulfur or carbon atom in the five-membered ring, (3) the presence/absence of acylamino sidechain, (4) a different trans- stereochemistry of the hydrogens.

Ampicillin can react with amines to form inactive amides, which has implication on co-administration for instance with sulbactam against *A. baumannii*. Unlike avibactam, sulbactam is an example of β -lactamase inhibitors that contain the β -lactam ring in its structure. β -lactamases are serine protease enzymes expressed by bacteria to act against β -lactam drugs through a hydrolysis of the β -lactam ring. Our *In-silico* study will allow to understand their interactions with the OccAB1 porin and the mechanisms of translocation. For this purpose, for each antibiotic a metadynamics simulation was carried out up to 2 μ s and the free energy surfaces associated are reconstructed in the space of the two collective variables (the distance along the diffusion axis and the orientation of antibiotics's dipole moment).

Internal electric field in OccAB1

A thorough electrostatic study of OccAB1 has shown the conservation, similarly to the general porins, of a transversal segregation of charges in its constriction region. A negative pocket is formed by a set of negatively charged amino-acids mainly located on the loop L7 and include Asp291, Asp293, Asp311, Tyr303 and Tyr308. On the opposite side at the CR, a series of positively charged amino-acids composed by several arginines and lysines (R21, R27, K29, K34, R127, R382, R384, R406) are located on the β -barrel, forming a strong positively charged

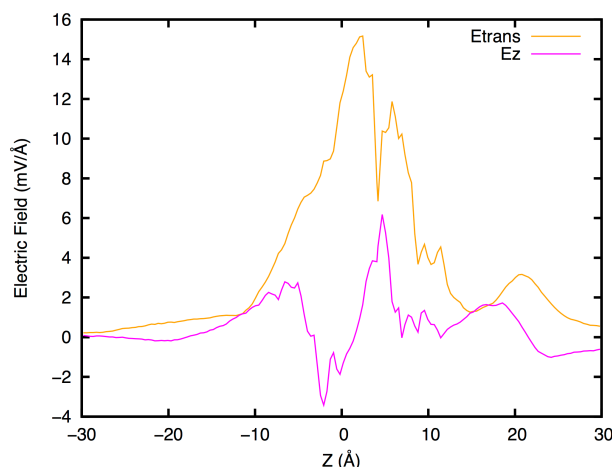


Figure 19: The electrostatic field along the z-axis channel of OccAB1 during classical MD simulations measured by using the method developed by (S Acosta-Gutierrez et al.,2015). This figure shows a strong transversal field (xy component) in the constriction region suggesting the alignment of the molecules to this field during their translocation in the narrowest part of the pore.

region facing the negative pocket. This arrangement of the negative and positive charges is conserved in general porins and described to create a transverse electric field in CR, such as in OmpF from *E. coli* [62]. Indeed, recently, an in-silico method was proposed in our lab to quantify this internal electric field by using the polarization of water molecules as probe[63]. The same method was applied here to OccAB1 (fig.19) to quantify its intrinsic electric field. This transversal electric field is naturally present in porins with wide radius to filter molecules that pass through the channels: those molecules possessing a dipole moment can align it to the electric field in such a way they pass faster the CR barrier and with low energy. On the other hand, non-polar molecules are excluded even in channels with a wide CR [24].

Free energy surfaces of antibiotics penetration

The free energy surface of each antibiotic was reconstructed based on the metadynamics of collective variables: (1) the position of its center of mass with respect to the axis of diffusion Z and (2) the orientation of its dipole moment along the same axis Z . Overall, all the FESs of the selected antibiotics molecules showed an asymmetry between the minima energy in the extracellular and intracellular side, the deepest minima energies are rather located in the intracellular side of OccAB1 channel (periplasmic side) except for the ampicillin that displayed a symmetry between the two sides of the porin (fig.20). This has been already observed in the previous study of substrates translocation through OccABs porins.

Imipenem and Meropenem

The FES and molecular pathway of imipenem and meropenem translocation show two free energy barriers one in the constriction region at z from -5 Å to $+3 \text{ Å}$ and the second barrier at

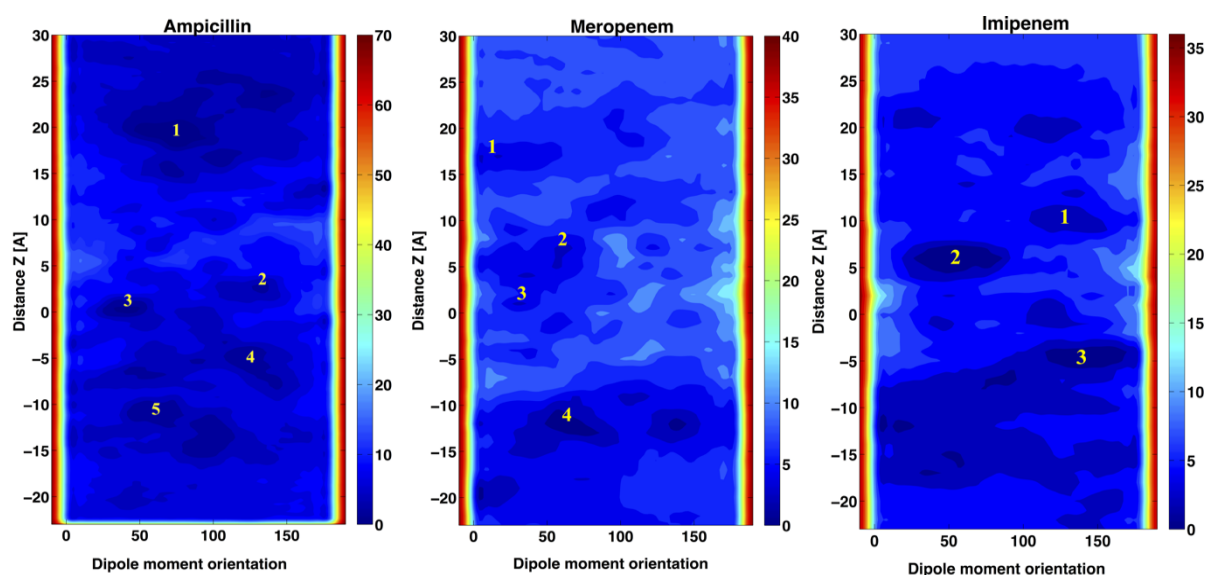


Figure 20: Free energy (FESs) reconstructed from metadynamics simulations for the translocation of ampicillin, meropenem and imipenem in OccAB1. The FESs are plotted with respect to the selected collective variables: the orientation of the dipole moment of antibiotics (x-axis) and the distance along the pore (y-axis). Each color in the FES corresponds to a free energy difference of 2 kcal/mol.

z in from 10 Å to 16 Å. These two barriers were identified also in the FES of arginine and glutamic acid (fig. 20), this similarity might be related to the similar dipole moment of these molecules. The both imipenem and meropenem enter the constriction region ($z=5$ Å) with their negatively charged group with the dipole moment orientation values <90 on the x-axis of the FES. However, whereas the minima energy of the translocation path of meropenem (from 1 to 4 on the FES) is rather straight with a dipole moment oriented towards the extracellular side, Imipenem changes its reorientation in the CR at $z=0$ Å in order to exit with its positively charged group following the path from 1 to 3 on the FES (Fig.20, Fig.21). The interactions porin-antibiotics involved during the translocation mechanism are reported in the figure 24, Figure 25 and table III, where the positively charged residue of the “basic ladder” interacts with the negatively charged group of antibiotics while entering the porin, whereas the negatively charged residues of the negative pocket on L7 interacts with the positive group of the imipenem and meropenem. In the both imipenem and meropenem, the FESs displayed a common minimum energy (minimum 2) at $z=5$ Å involving a set of consensus residues in the translocation process and are mainly: Arg127, Arg406 (of the “basic ladder”), Asn124, interacting with the negatively charged groups of the antibiotics, and Phe299 to make hydrophobic interaction with the rings of the b-lactam groups (see figure 6 and 7 for minimum labelled 2). Moreover, by comparison with the previous free energy calculations of natural substrates (arginine, glutamic acid, glycine), the FES of imipenem and meropenem share the same minima energies with the FES of arginine (Fig.14) in the CR at z from 5 Å to 10 Å. This

indicates that these 3 molecules imipenem, meropenem and arginine have the same binding site in their translocation path, they all involve interactions with residues R127, R406 and R384 of the basic ladder. In this finding, Arginine was served as probe to reveal a common path for molecule structurally related such as meropenem and imipenem.

Ampicillin

The FES of ampicillin translocation through OccAB1 is reported in the Fig.20 and revealed a barrier at $z = 5 \text{ \AA}$, which corresponds to the CR in the radius profile over time simulations (fig.1). The FES shows also 5 energy minima through a path 1 to 5 during the translocation of ampicillin. In the minima 2, Ampicillin penetrates with the phenyl ring group down interacting with an arginine cluster at the constriction zone, (Fig. 20).

As the pore narrows, in the minima 2, ampicillin changed its orientation (Fig. 21) to point down the negatively charged group interaction with the positive charged residues of the “basic ladder” in particular, the Arg406, Arg127, Arg384 and Arg27 but also polar residues such as Asn124 (Fi. 23). In the minima 4 the ampicillin changes again its dipole orientation to exit with phenyl group pointing towards the intracellular side. The mechanism of translocation of ampicillin in AB1 required several changes of orientation to find stable poses allowing it to diffuse easily with a lower energy cost. This is due to the fact that its dipole is aligned along the long axis and cannot adapt well to the small CR, as

already noticed with Norfloxacin through OmpF [64]. Interestingly, this orientation corresponds to the direction of the electric field of the component z . In fact, in the minimum 2 the ampicillin has reoriented itself to align its dipole to the electric field. In the same way, ampicillin changes its orientation to reach minimum 3 in order to be aligned with the electric field in the region z from $2 \text{ \AA}$ to $-10 \text{ \AA}$. Interestingly, the same mechanism is observed for the imipenem which also needs to change its orientation when it penetrates the CR and successively align with the electric field along z -axis (path 1-2-3 similar to path 2-3-4 for ampicillin). It appears that in the case of ampicillin and imipenem, the internal electric field with its z component guide their translocation. However, this was not observed in the case of meropenem probably due to the different dipole that has a strong component transversal to the long axis: with minimal rotation of the long axis the dipole is aligned to the transversal field. Furthermore, different minima in the FESs related to the translocation path of each antibiotic were analysed in detail to identify and characterize antibiotic-OccAB1 interactions and are labelled with consecutive numbers in Fig. 20 and Fig.21.

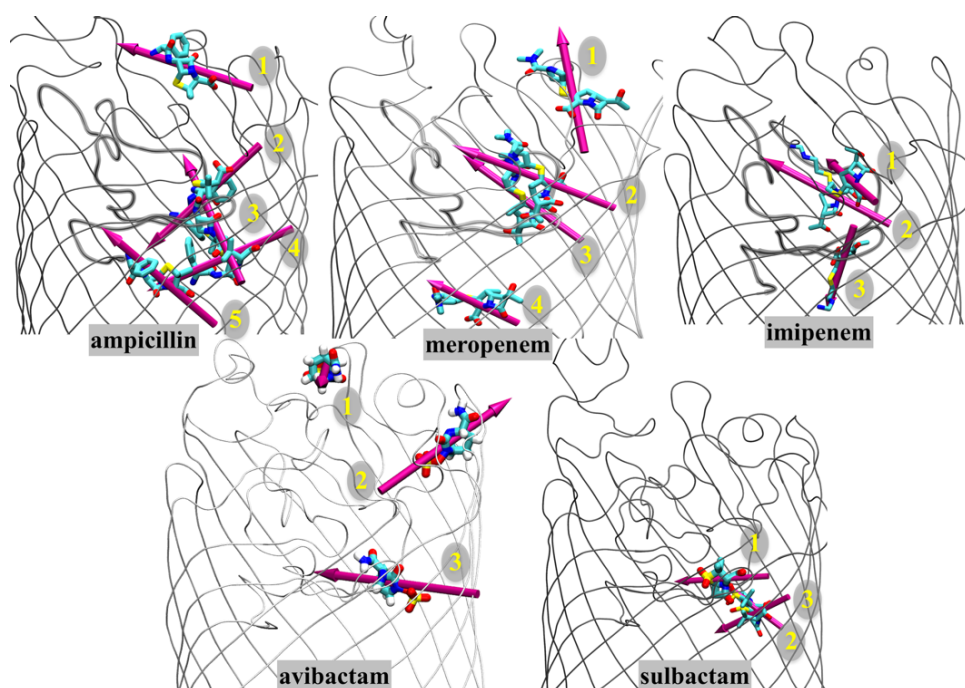


Figure 21: Antibiotics translocation through OccAB1. Top front view of OccAB1 (protein backbone in grey) in complex with each antibiotic in the main minima displayed in the figure FES, the antibiotics are shown in sticks representation with nitrogen coloured in blue and oxygen coloured in red. The magenta arrow indicates the orientation of the dipole moment of antibiotic in each minimum.

β -lactamase inhibitors

Avibactam and sulbactam

The FES of avibactam translocation through OccAB1 (Fig. 22) displayed in total 4 energy minima; two strong minima in the extracellular side (minima 1 and 2), one minimum in the intracellular side (minimum 4) and one main minimum (minimum 3) in the CR at $z=0$ Å. Avibactam adapts reorienting its dipole to the transversal electric field in the CR of OccAB1, as shown in Fig.19. In the figure 27 is reported the diagram of the avibactam-OccAB1 interactions in the main minima extracted from the FES (Fig.22). As we can see, in the minimum 3 located in the CR, avibactam interacts mainly with the positive residues of the basic ladder by its negatively charged group, especially with the residues Arg406, Arg382, Arg127 but also Gln366 and Thr125. On the other hand, the NH₂ group of avibactam is oriented towards the negative pocket on the loop L7.

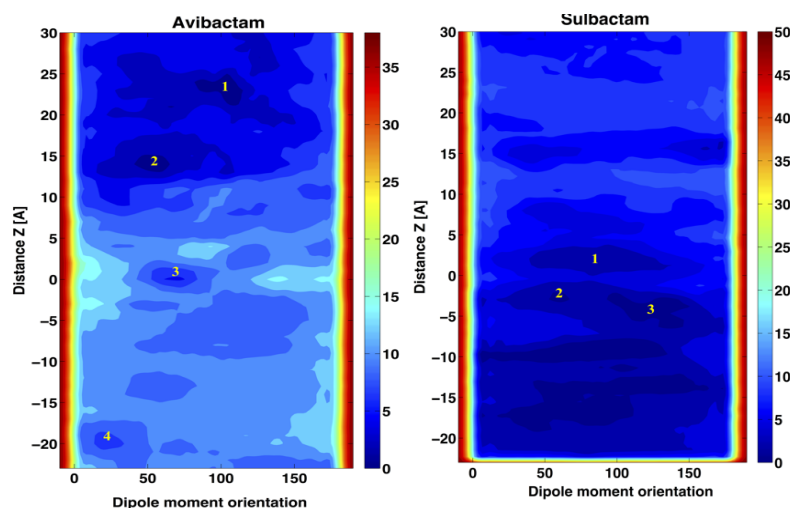


Figure 22: Free energy (FESs) reconstructed from metadynamics simulations for the translocation of ampicillin, meropenem and imipenem in OccAB1. The FESs are plotted with respect to the selected collective variables: the orientation of the dipole moment of antibiotics (x-axis) and the distance along the pore (y-axis). Each color in the FES corresponds to a free energy difference of 2

The FES of sulbactam has revealed several large energy minima in the intracellular side of the protein, in agreement with the positive charge distribution of the protein along z-axis (Fig.5 section I) that favors the interaction with the negative charge of sulbactam. As we can see in the FES, the highest energetic barrier located at $z \approx 10$ Å corresponds to region of the protein rather negatively charged, thus unfavourable to the sulbactam. In the CR between $z \approx 0$ and 5 Å, sulbactam adopts a transversal position so that it aligns the dipole moment (7.9 Debye) to the transversal internal electric field of the protein, as reported in the Fig.21 (magenta arrow) and shown in the minima 1 on the figure FES (Fig.22). In this minimum, sulbactam points its negatively charged group towards the basic ladder to interact mainly with Arg384, Arg406 and Arg127, while its sulfate group interacts with Asn306 on the loop L7 and Thr366 of the loop L3 (Fig.22 and 26). In the minima below the CR at $z=0$ Å to -5 Å, avibactam changes its orientation of its dipole moment towards the extracellular side, maintaining its sulfate group towards loops L7 and L3 and its negatively charged group moves down the pore axis to be released from the CR following the basic ladder (Fig.21, 22). Then it changes again its orientation to have more stable conformation (minimum 3) by maintaining interactions with positively charged of the basic ladder mainly with residue Arg384, Arg382 (Fig.27). Despite, the structural and charge similarity between avibactam and sulbactam, they show different free energies related to their translocation path through OccAB1 porin, this might be due to the convergence issue of the FES of avibactam for which the metadynamics simulations would

probably need larger time scale to explore better the selected collective variables. Indeed, the simulations of sulbactam were extended up to 2us whereas those of avibactam reached only 1us simulations. Therefore, further extended simulations are needed to confirm the above FES.

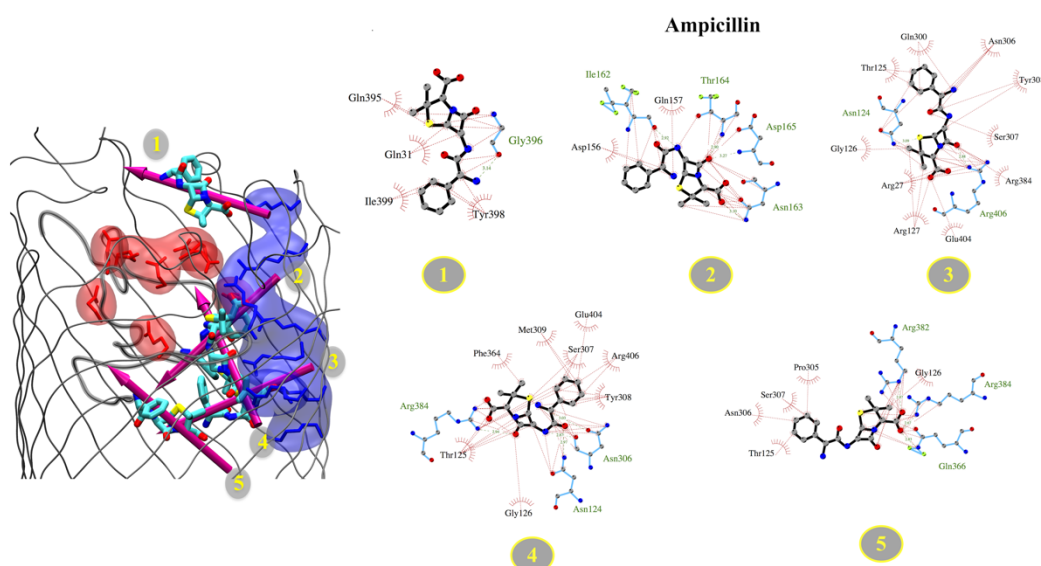


Figure 23: (A) Top front view of OccAB1 (protein backbone in grey) in complex with each ampicillin in the main minima displayed in the figure FES, the antibiotics are shown in sticks. The magenta arrow indicates the orientation of the dipole moment of antibiotic in each minimum. (B) interaction diagram from LigPlot, shows the interaction of ampicillin (grey color) with OccAB1 residues (cyan color).

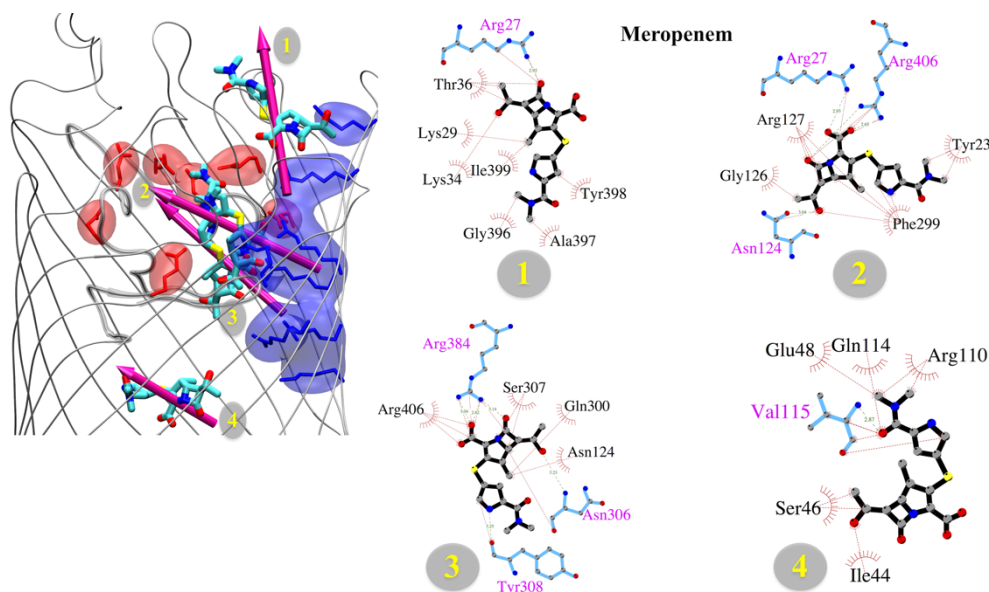


Figure 24: (A) Top front view of OccAB1 (protein backbone in grey) in complex with each meropenem in the main minima displayed in the figure FES, the antibiotics are shown in sticks. The magenta arrow indicates the orientation of the dipole moment of antibiotic in each minimum. (B) interaction diagram from LigPlot, shows the interaction of meropenem (grey color) with OccAB1 residues (cyan color).

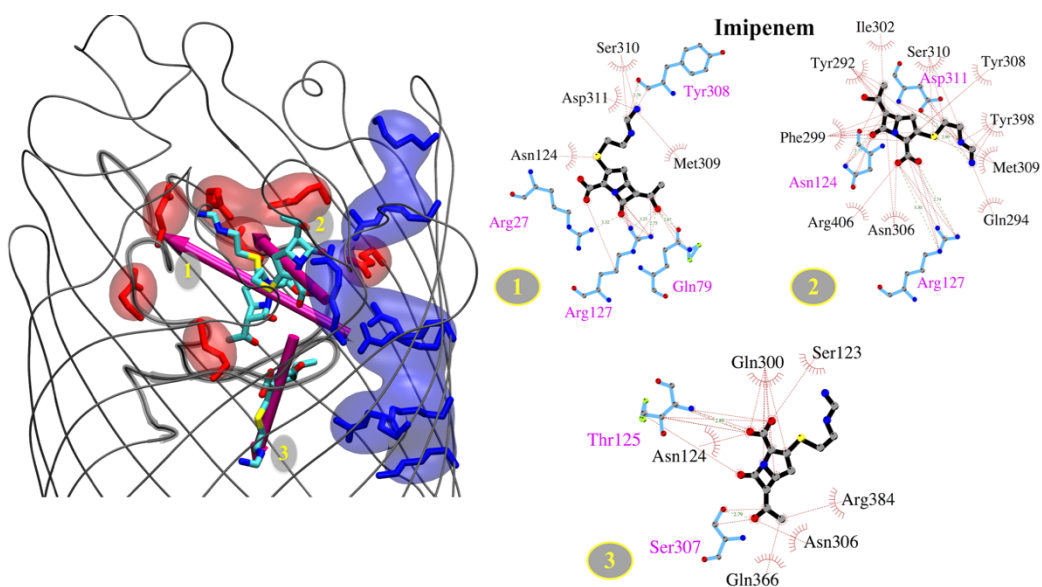


Figure 25: (A) Top front view of OccAB1 (protein backbone in grey) in complex with each imipenem in the main minima displayed in the figure FES, the antibiotics are shown in sticks. The magenta arrow indicates the orientation of the dipole moment of antibiotic in each minimum. (B) interaction diagram from LigPlot, shows the interaction of imipenem (grey color) with OccAB1 residues (cyan color).

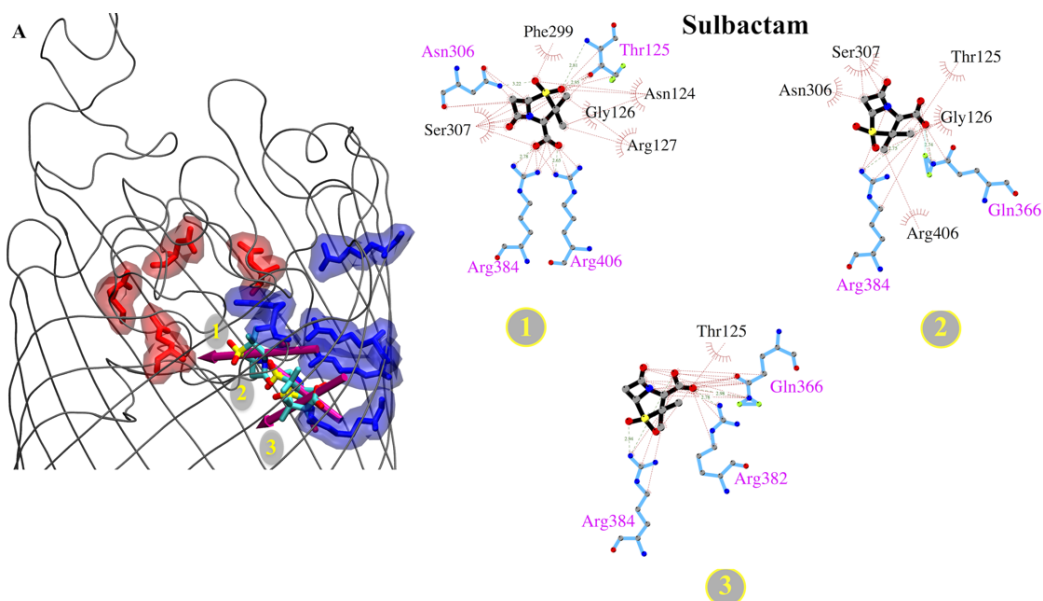


Figure 26: (A) Top front view of OccAB1 (protein backbone in grey) in complex with each sulbactam in the main minima displayed in the figure FES, the antibiotics are shown in sticks. The magenta arrow indicates the orientation of the dipole moment of antibiotic in each minimum. (B) interaction diagram from LigPlot, shows the interaction of sulbactam (grey color) with OccAB1 residues (cyan color).

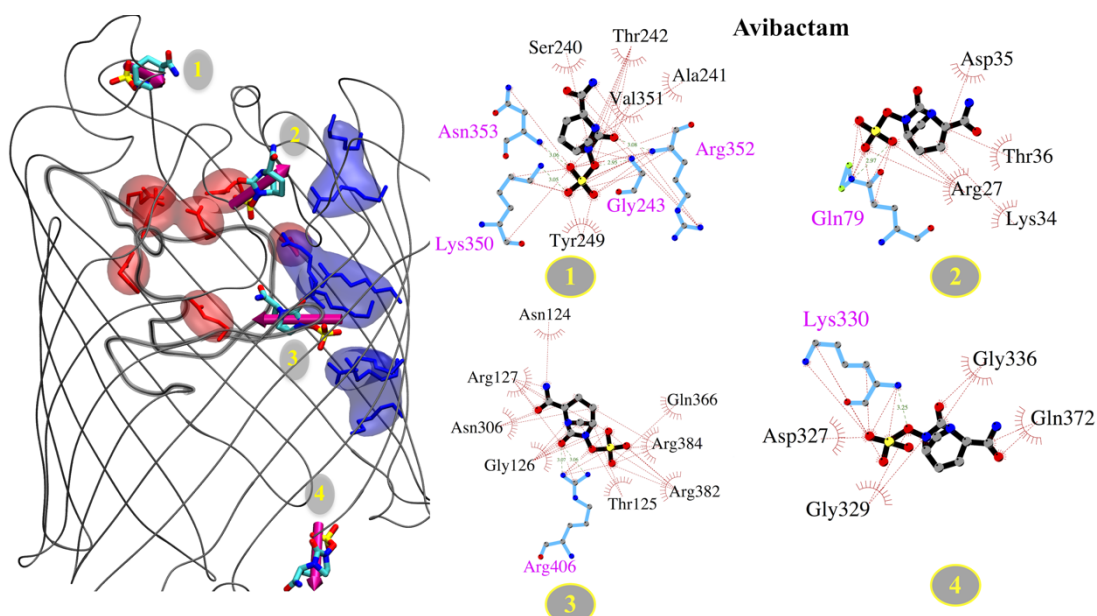


Figure 27: (A) Top front view of OccAB1 (protein backbone in grey) in complex with each avibactam in the main minima displayed in the figure FES, the antibiotics are shown in sticks. The magenta arrow indicates the orientation of the dipole moment of antibiotic in each minimum. (B) interaction diagram from LigPlot, shows the interaction of avibactam (grey color) with OccAB1 residues (cyan color).

Table II: Structure representation of the selected antibiotics of this study and their main physico-chemical properties.

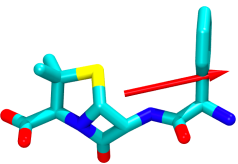
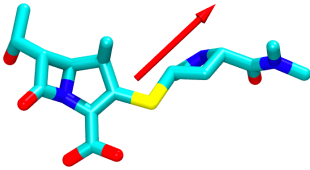
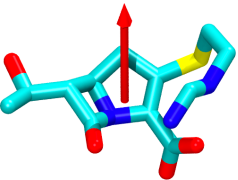
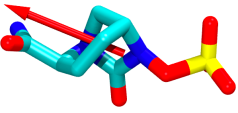
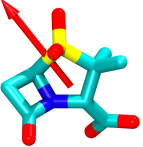
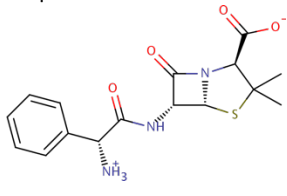
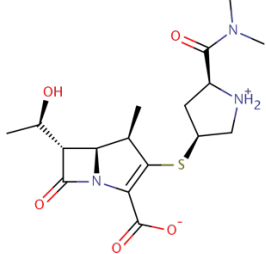
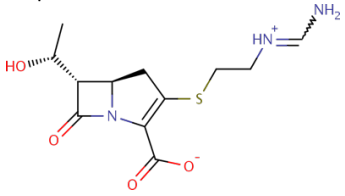
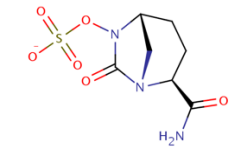
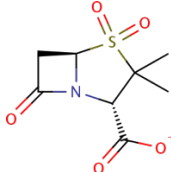
	Minimal projection area (Å)	Dipole in vacuum (Debye)	H-bonds donors/acceptors	Rotatable bonds	Molecular surface (Å ²)	Physiological charge
Penicillins						
Ampicillin 	50.47	35.11	2 & 4	4	303.84	0
Carbapenems						
Meropenem 	50.89	30.95	2&5	5	344.19	0
Imipenem 	44.68	13.75	3&5	6	258.84	0
β-lactamase inhibitors						
Avibactam 	35.25	13.28	1&5	3	207.31	-1
Sulbactam 	39.61	7.95	0&5	1	178.93	-1

Table III: the difference free energy (in kcal/mol) needed to cross the barrier in OccAB& related to the FESs of each antibiotic and the residues involved in the OccAB1-antibiotics interactions.

	Free energy Difference Kcal/mol	Residue implicated in the translocation process	Residues implicated in H-bonds
Ampicillin 	8	1: GLY396, TYR398, ILE399, GLU31, GLN395, 2: ASN163, ASP165, THR164, GLU57, ILE162, ASP156 3: ARG406, ARG384, SER307, TYR308, ASN306, GLN300, THR125, ASN124, GLY126, ARG27, ARG127, GLU404 4: ASN124, ASN306, TYR308, ARG406, SER307, GLU404, MET309, PHE364, ARG384, THR125, GLY126 5: GLN366, ARG384, GLY126, ARG382, PRO305, SER307, ASN306, THR125	GLY396, ASN163, ASP165, THR164, ILE162, ASP156, ARG406, ASN124, ASN306, ARG384, GLN366, ARG382
Meropenem 	6	1: ALA397, TYR398, ARG27, THR36, LYS29, ILE399, LYS34, GLY396 2: PHE299, TYR239, ARG406, ARG27, ARG127, GLY126, ASN124 3: TYR308, SER307, GLN300, ASN124, ASN306, TYR308, ARG406 4: ILE44, SER46, VAL115, GLU48, GLN114, ARG110	ARG27, ARG406, ASN124, ARG384, TYR308, ASN306, VAL115
Imipenem 	4	1: ARG127, GLN79, TYR308, ARG27, MET309, SER310, ASP311, ASN124 2: ARG127, ASP311, ASN124, GLN294, MET309, TYR398, TYR308, SER310, ILE302, TYR292, PHE299, ARG406, ASN306 3: THR125, SER307, GLN366, ASN306, ARG384, SER123, GLN300, ASN124	ARG127, GLN79, TYR308, ARG27, ARG127, ASP311, ASN124, THR125, SER307
Avibactam 	10	1: ASN353, LYS350, GLY243, ARG352, TYR249, VAL351, ALA241, THR242, SER240 2: GLN79, ASP35, THR36, ARG27, LYS34 3: ARG406, GLY126, ASN306, ARG127, GLN366, ARG384, ARG382, THR125 4: LYS330, ASP327, GLY329, GLY336, GLN372	ASN353, LYS350, GLY243, ARG352 , GLN79, ARG406, LYS330
Sulbactam 	4	1: Arg384, Arg406, Arg127, Asn306, Thr125, Ser307, Asn124 2: Arg384, Arg406, Gln366, Asn306, Ser307, Thr125 3: Arg384, Arg382, Gln388, Thr125	Asn306, Arg384, Arg382, Arg406, Thr125, Gln366

Conclusion and perspectives

In this study, we carried out metadynamics simulations of four OM porins from *A. baumannii*, OccAB[1-4], in presence of different simple substrates in order to evaluate the energetics of substrates during translocation. The three small substrates used in this study were selected such that the selectivity for size and charge of these porins could be investigated. The reconstructed free energy surfaces displayed similar binding sites in the extracellular loops of OccAB porins for the selected positively charged substrate (arginine) indicating the stability of the latter in that region. This suggests a correlation with the macro-dipole moment that would allow the attraction of positive substrates, and is supported by the negative charge distribution that is particularly pronounced in the extracellular area of the porins.

Our results showed that the OccAB1 porin can translocate both positive and negative substrates without any clear preference and at low energy cost. Our results are in agreement with the experimental data showing OccAB1 having the best translocation rates among OccAB porins[40]. The large size of the pore radius and the symmetry of the charge distribution along its pore axis contribute mainly to the lack of selectivity of OccAB1. The similarity of the macro-dipole of OccAB1 with that of the general porin OmpF [62] being in both proteins rather transversal to the z-axis with a low magnitude, suggests for the future to focus on the internal electric field which has been suggested to play a key role in general channels from *Enterobacteria* [65] [62]. Our analysis highlights OccAB1 as a good candidate for further investigation to explore the effect of the internal electric field on the translocation of polar antibiotics.

In OccAB2-4, the translocation of the selected substrates is limited by the size of their pore such that only small molecule can pass through easily with low energy costs as shown with glycine. Specifically, in the CR of OccAB3 and OccAB4, the glutamic acid enters with the side chain (negative), whereas arginine prefers entering with the negative backbone group. This indicates that these porins are clearly selective to negative substrates as suggested by liposome swelling assays. This is due mainly to the dominated positive charge distribution in their CR despite the rearrangements of the charge distribution occurring above the CR in the extracellular of OccAB3 (in the mouth of the channel) as mentioned earlier in the figure 5 and 6. In addition to the electrostatic effect, the translocation of substrates in OccAB4 is also affected by the hydrophobicity of the pore, acting as an additional barrier for polar substrate uptake. Therefore, OccAB4 is expected to have also better permeation for hydrophobic molecules.

In the case of OccAB2, not only the charge and the size can affect the passage of molecules but the flexibility of the extracellular loops is an additional key element that contributes to the uptake of substrates. Indeed, the flexibility of the internal loop L7 in OccAB2 makes this channel different compared to the other OccABs. The OccAB2 channel is unstable and shows two constriction regions, which can open and close alternately. In addition, differences in the charge distribution in the surface electrostatics map have been observed between the representative structure of the MD trajectory obtained by clustering and that of the crystal structure. (figure 6 B). This indicates that the flexibility of the external loops changes the charge distributions at the surface so that positive or negative affinity sites could be hidden or exposed to consequently affecting the selectivity of this pore to substrates.

This first step contributes to our understanding of the structure-function relationship of these specific channels of *A. baumannii*. We provided a detailed in-silico investigation of structural, dynamical, and physicochemical properties that are involved in the discrimination, uptake, and translocation mechanisms of specific substrates of each OccAB porins by using molecular dynamics simulations.

The second step of this study focused on the OccAB1, and the investigation of antimicrobial molecules (antibiotics molecules) through this channel. Our results showed that the FES of imipenem and meropenem have similar minima to that of substrates translocation through OccAB1. Thus, common residues are involved during the translocation process, and the analysis of substrate served as a probe for the imipenem and meropenem investigations. This similarity is due to the structure similarity of these molecules (especially between arginine, meropenem and imipenem). A common trend observed for all molecules is that the internal transversal electric field in OccAB1 has an effect on the translocation mechanism of the small polar molecules through the pore. Even the small molecule such as sulbactam/avibacam has to align its dipole moment to this transversal electric field in order to overcome the barrier in the CR. Also, the orientation of the dipole moment with respect to the long axis of the molecules matters: ampicillin shows several reorientations inside the CR because the longitudinal dipole cannot orient to the electric field, similarly to norfloxacin through OmpF. Also, imipenem has the same behaviour, though its high flexibility helps it to adapt. On the other hand, the transversality of dipole in meropenem allows its passage without any reorientation.

Eventually, we have reported the main interaction in the binding sites involved in the mechanism of the translocation of each molecule through OccAB1. This second step investigation on the antibiotics translocation through OccAB1, gives support information for

further experimental data such as mutagenesis of these reported residues in antibiotics binding sites, in order to evaluate the importance of single point mutations on this channel and their effect on the uptake of these antibiotics.

Appendix 1

Table S 1: (D) Statistic values over time trajectories of the distribution of the minimum radius and its position along z-axis (Sd: Standard deviation).

Minimum radius Å				
	mean	Sd	min	max
OccAB1	2.03	0.39	0.28	3.21
OccAB2	1.76	0.51	0	3.15
OccAB3	1.08	0.47	0	2.64
OccAB4	1.25	0.43	0	3.19
Constriction region position along z axis Å				
OccAB1	4.73	2.67	-2	10
OccAB2	5.58	5.4	-5	14.5
OccAB3	1.47	2.02	-2	11.5
OccAB4	3.09	1.66	-0.5	11

Table S 2: Physico-Chemical properties of OccAB porins for the macro dipole moment calculations, to be compared with OccK8, OmpF and OccK1. OccK8, 2OMPf, OccK1 (604.85D, 271.40 (trimer=540.38), 853.7 D).

	AB1	AB2	AB3	AB4
Number of atom	6288	6251	6388	6239
Total charge	-7.0	-7.0	-2.0	-10.0
Geometrical centre	0.28 -0.88 -0.31	0.15 -0.32 -0.54	-0.29 0.10 -0.27	-0.17 0.08 -0.44
Centre of masses	0.27 -0.86 -0.35	0.127 -0.303 -0.555	-0.28 0.09 -0.28	-0.17 0.06 -0.45
Centre of charges	0.15 -1.03 0.17	0.09 -0.24 -0.13	-0.50 0.14 0.09	-0.22 0.13 0.24
Macro-Dipole moment magnitude	334.94 D	640.05 D	902.75 D	1389.06 D

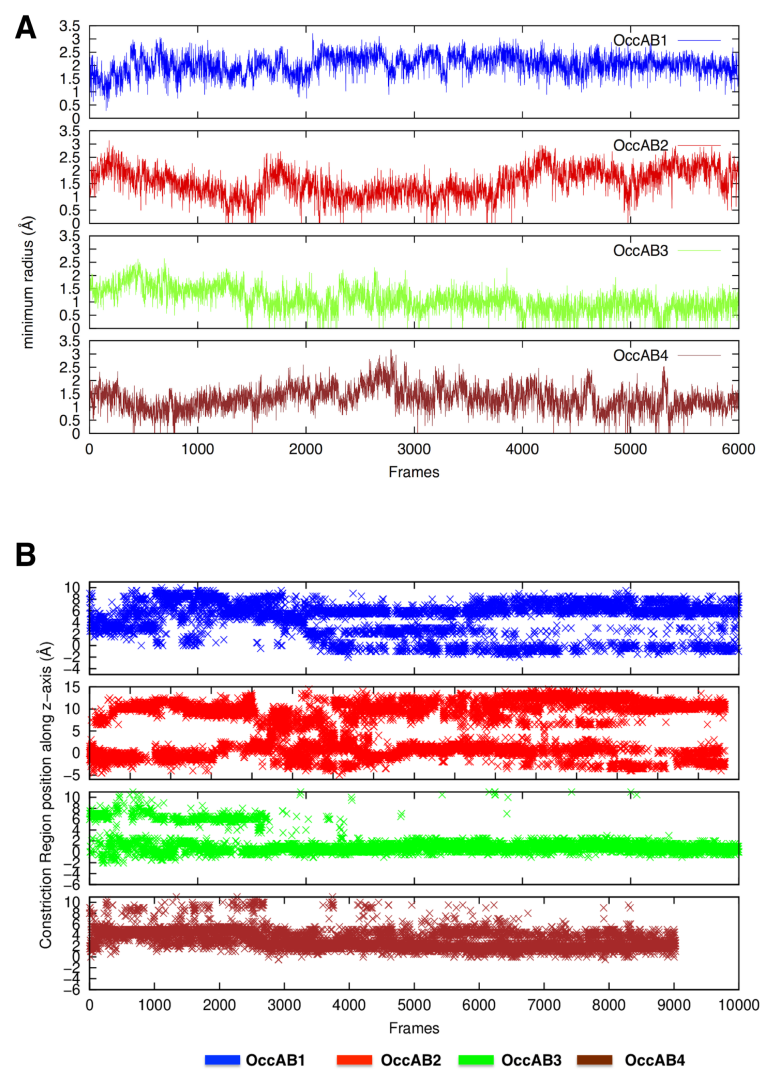
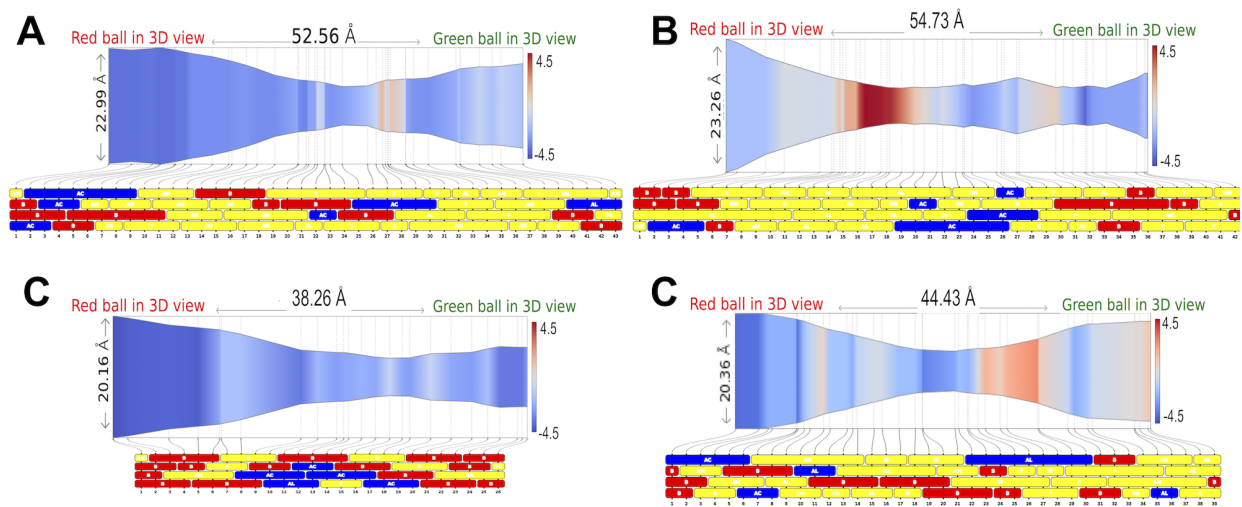


Figure S 1: Radius profiles of OccAB. (A) Minimum radius over time simulation. (B) distribution of the minimum radius area (constriction region: CR) along the pore axis (z-axis).



Current profile: Hphob. / Kyte & Doolittle

Boxes colored based on: Charge

Neutral Positive Negative

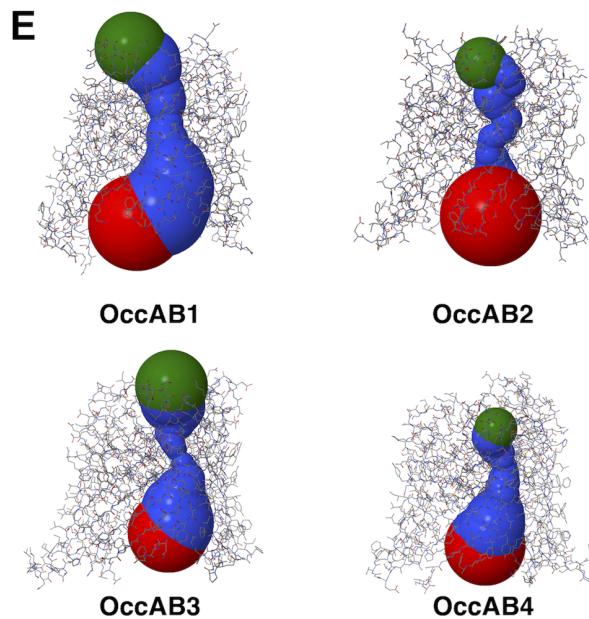


Figure S 2: Structural and biophysical properties of OccAB channel from *A.baumannii*. TOP (A, B, C, and D) The 2D profiles of the radius of OccAB crystals in the figure are colored by hydrophobicity of residues along with the channels. The boxes are colored according to the Amino Acid charge property and labeled by Amino acid. (E) The transmembrane channels identified by ChExVis and extracted in OccAB (5DL5, 5DL6, 5DL7 and 5DL8 crystals) pore shown as sphere (blue), the green ball corresponds to the entrance of the channel (extracellular) and the red ball corresponds to wide pore end (periplasm side).

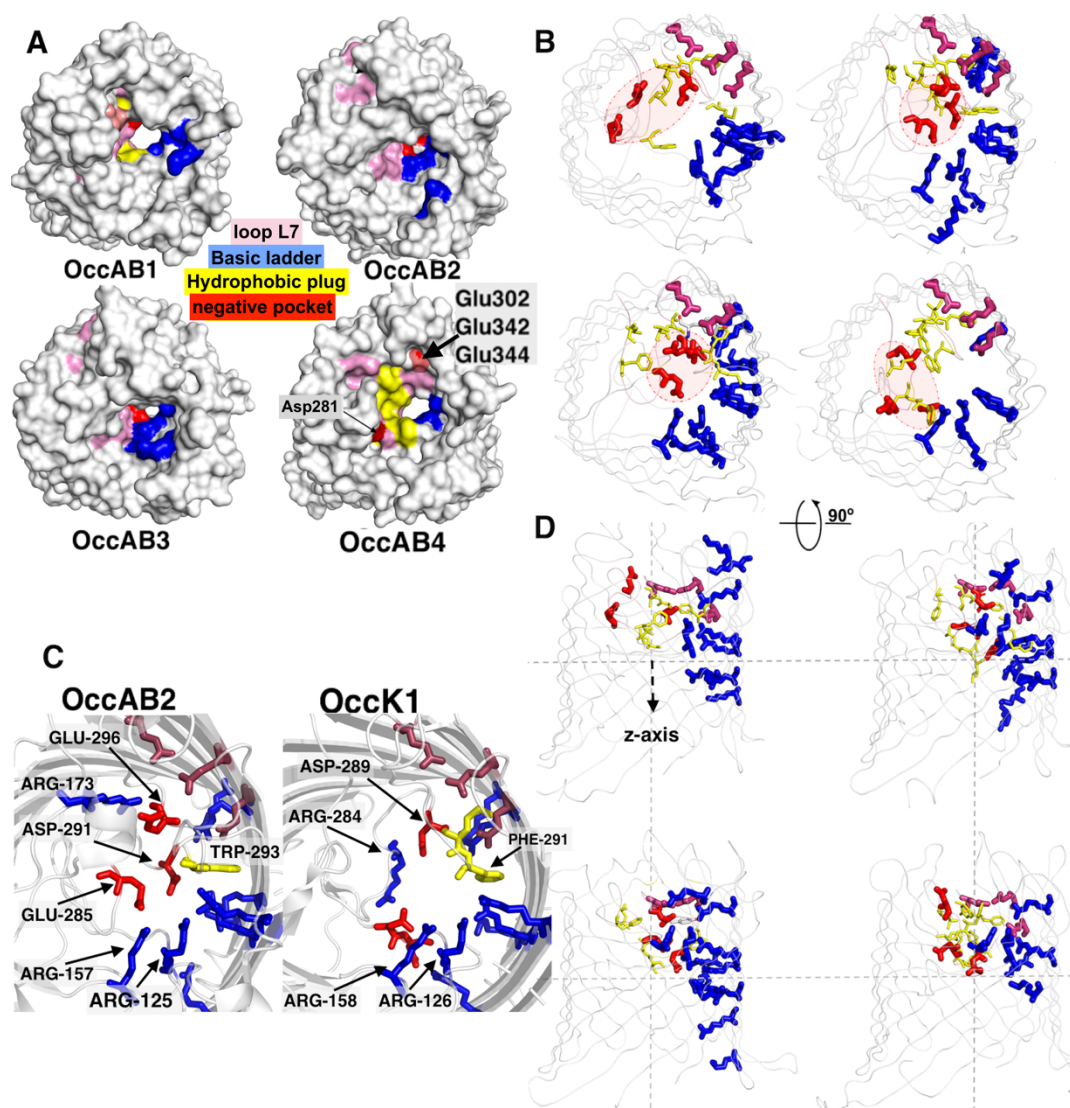


Figure S 3: The basic ladder and negative pocket of OccAB porins. (A) surface representation of the Top view of OccAB crystals, (B) (D) the transversal view of the porins shown as ribbon, the residues involved in the “basic ladder” are shown as blue stick, the residues of the negative pocket are shown as red sticks and the barrel negatively, the hydrophobic residues are represented in yellow. The pink colour in the panel A corresponds to the loop L7. (D) Structural comparison between residues lining the pore involved in the conductance state in OccAB2 and OccK1

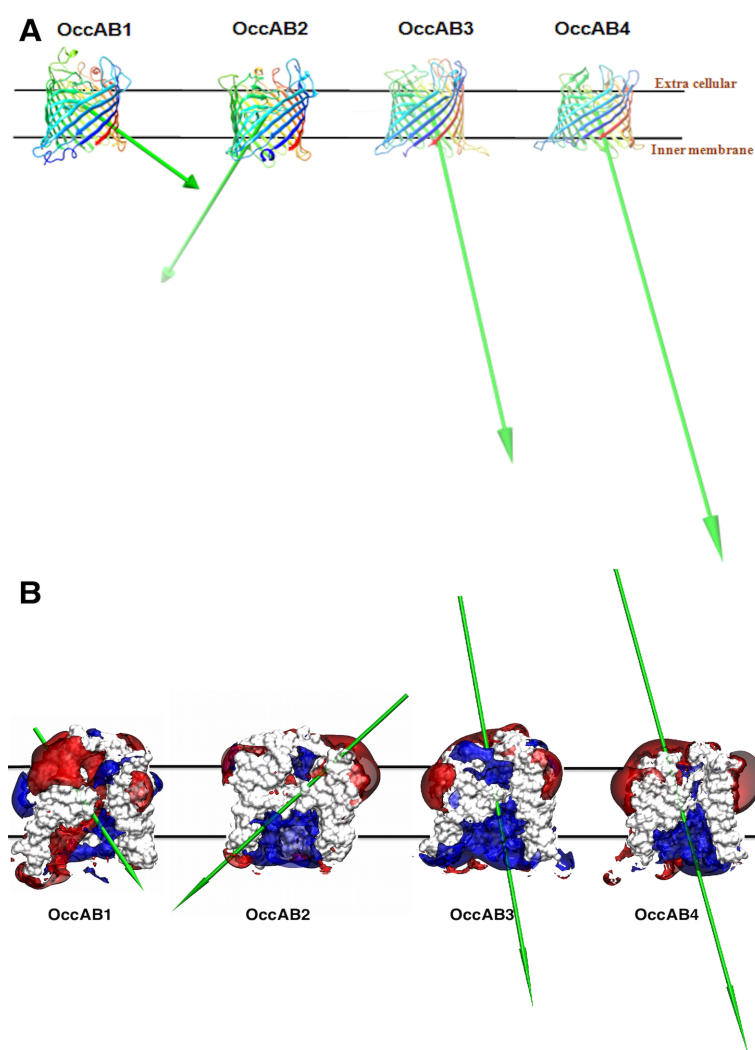


Figure S4: A) Calculated electric dipole moment vector of the OccAB crystals porins which is represented by the green arrow. The center of the arrow corresponds to the Center of Mass of the porin, the tip indicates the positively charged end and the base of the arrow corresponds the negatively charged end, the length of the arrow corresponds to the magnitude of the marco-dipole moment. B) Half transversal surface representation of OccAB porins with the averaged dipole moment vector over time simulations. As red surface representation is the negative electrostatic potential (in -2 KT/e isocontour), as blue surface is the positive electrostatic potential (in $+2 \text{ KT/e}$ isocontour) averaged over times simulations.

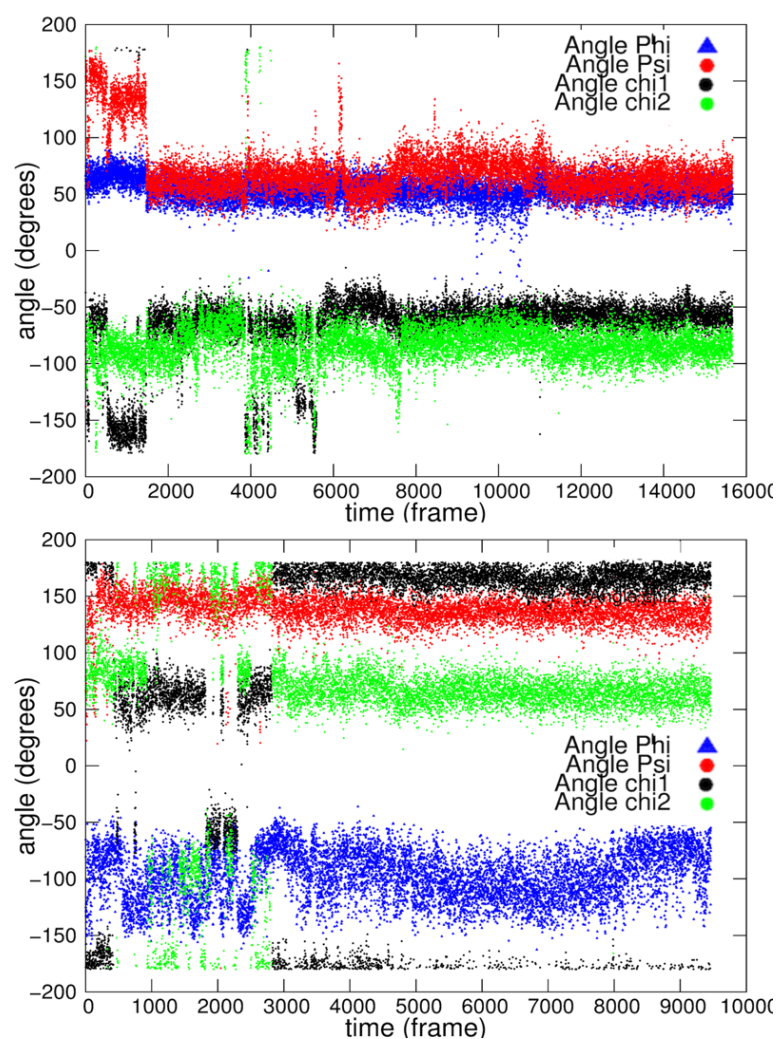


Figure S 5: The flexibility of the pore in OccAB2 and OccAB4. The angles Phi, Psi, Chi1 and Chi2 of Trp293 in OccAB2 (Top) and Trp292 in OccAB4 (bottom) plotted against time showing the flipping of the tryptophan lining the pore during the MD simulations.

Reference:

- [1] M. E. Hibbing, C. Fuqua, M. R. Parsek, and S. B. Peterson, "Jungle," *Natl. Rev. Microbiol.*, vol. 8, no. 1, pp. 15–25, 2010.
- [2] R. Quinn, "Rethinking antibiotic research and development: World War II and the penicillin collaborative," *Am. J. Public Health*, vol. 103, no. 3, pp. 426–434, 2013.
- [3] J. Avorn, J. Barrett, P. Davey, S. McEwen, T. O'Brien, and S. Levy, "Antibiotic resistance: synthesis of recommendations by expert policy groups," *Alliance Prudent Use Antibiot. WHO/CDS/...*, vol. 24, no. 2, pp. 240–7, 2000.
- [4] J. R. Lonks, J. Garau, and A. A. Medeiros, "Implications of antimicrobial resistance in the empirical treatment of community-acquired respiratory tract infections: The case of macrolides," *J. Antimicrob. Chemother.*, vol. 50, no. SUPPL. 2, pp. 87–91, 2002.
- [5] A. Docze *et al.*, "Penicillin resistance in viridans streptococcal bacteremia is related with high mortality.," *Scand. J. Infect. Dis.*, vol. 35, no. 11–12, pp. 916–917, 2003.
- [6] CDC, "Antibiotic resistance threats in the United States, 2013," *Centers for Dis. Control Prev.*, p. 114, 2013.
- [7] R. Coico, "Gram Staining," *Curr. Protoc. Microbiol.*, p. 18770544, 2005.
- [8] M. E. Falagas, I. A. Bliziotis, S. K. Kasiakou, G. Samonis, P. Athanassopoulou, and A. Michalopoulos, "Outcome of infections due to pandrug-resistant (PDR) Gram-negative bacteria," *BMC Infect. Dis.*, vol. 5, no. 1, p. 24, 2005.
- [9] H. Nikaido, "Molecular basis of bacterial outer membrane permeability revisited.," *Microbiol. Mol. Biol. Rev.*, vol. 67, no. 4, pp. 593–656, 2003.
- [10] J.-M. Pagès, C. E. James, and M. Winterhalter, "The porin and the permeating antibiotic: a selective diffusion barrier in Gram-negative bacteria," *Nat. Rev. Microbiol.*, vol. 6, no. 12, pp. 893–903, 2008.
- [11] S. Mukerji, M. O'Dea, M. Barton, R. Kirkwood, T. Lee, and S. Abraham, "Development and transmission of antimicrobial resistance among Gram-negative bacteria in animals and their public health impact," *Essays Biochem.*, vol. 61, no. 1, pp. 23–35, 2017.
- [12] A. a Medeiros, "Recent increases in resistance: mechanisms and organisms. Evolution and Dissemination of / 3-Lactamases Accelerated by Generations of (3-Lactam Antibiotics," *Clin. Infect. Dis.*, vol. 24, no. Suppl 1, pp. 19–45, 1997.
- [13] D. M. Livermore, "beta-Lactamases in laboratory and clinical resistance.," *Clin. Microbiol. Rev.*, vol. 8, no. 4, pp. 557–84, 1995.
- [14] H. Nikaido and J. M. Pagès, "Broad-specificity efflux pumps and their role in multidrug resistance of Gram-negative bacteria," *FEMS Microbiol. Rev.*, vol. 36, no. 2, pp. 340–363, 2012.
- [15] S. Galdiero *et al.*, "Microbe-Host Interactions: Structure and Role of Gram-Negative Bacterial Porins," *Curr. Protein Pept. Sci.*, vol. 13, no. 8, pp. 843–854, 2012.
- [16] H. A. Terzi, C. Kulah, A. R. Atasoy, and I. H. Ciftci, "Investigation of oprd porin protein levels in carbapenem-resistant *Pseudomonas aeruginosa* isolates," *Jundishapur J. Microbiol.*, vol. 8, no. 12, pp. 10–14, 2015.
- [17] M. Winterhalter and M. Ceccarelli, "Physical methods to quantify small antibiotic molecules uptake into Gram-negative bacteria," *Eur. J. Pharm. Biopharm.*, vol. 95, pp. 63–67, 2015.

- [18] A. H. Delcour, "Outer Membrane Permeability and Antibiotic Resistance," *Biochim Biophys Acta.*, vol. 1794, no. 5, pp. 808–816, 2009.
- [19] E. M. Nestorovich, E. Sugawara, H. Nikaido, and S. M. Bezrukov, "Pseudomonas aeruginosa porin OprF. Properties of the channel," *J. Biol. Chem.*, vol. 281, no. 24, pp. 16230–16237, 2006.
- [20] A. Arockiasamy, P. D. Kumar, N. Sundara Baalaji, M. R. Rukmini, and S. Krishnaswamy, "Folding and structural stability of OmpC from Salmonella typhi: Role of LPS and environment," *Curr. Sci.*, vol. 87, no. 2, pp. 197–202, 2004.
- [21] W. Welte, U. Nestel, T. Wacker, and K. Diederichs, "Structure and function of the porin channel," *Kidney Int.*, vol. 48, no. 4, pp. 930–940, 1995.
- [22] M. Ceccarelli, C. Danelon, A. Laio, and M. Parrinello, "Microscopic Mechanism of Antibiotics Translocation through a Porin," *Biophys. J.*, vol. 87, no. 1, pp. 58–64, 2004.
- [23] M. A. Scorciapino *et al.*, "Rationalizing the permeation of polar antibiotics into Gram-negative bacteria," *J. Phys. Condens. Matter*, vol. 29, p. aa543b, 2017.
- [24] H. Bajaj *et al.*, "Molecular Basis of filtering carbapenems by porins from β -lactam-resistant clinical strains of Escherichia coli," *J. Biol. Chem.*, vol. 291, no. 6, pp. 2837–2847, 2016.
- [25] S. Samanta, M. A. Scorciapino, and M. Ceccarelli, "Molecular basis of substrate translocation through the outer membrane channel OprD of Pseudomonas aeruginosa," *Phys. Chem. Chem. Phys.*, 2015.
- [26] E. Eren *et al.*, "Toward Understanding the Outer Membrane Uptake of Small Molecules by Pseudomonas aeruginosa," vol. 288, no. 17, pp. 12042–12053, 2013.
- [27] M. Zahn, T. D'Agostino, E. Eren, A. Baslé, M. Ceccarelli, and B. van den Berg, "Small-Molecule Transport by CarO, an Abundant Eight-Stranded β -Barrel Outer Membrane Protein from Acinetobacter baumannii," *J. Mol. Biol.*, vol. 427, no. 14, pp. 2329–2339, Jul. 2015.
- [28] H. Giamarellou, A. Antoniadou, and K. Kanellakopoulou, "Acinetobacter baumannii: a universal threat to public health?," *Int. J. Antimicrob. Agents*, vol. 32, no. 2, pp. 106–119, 2008.
- [29] C.-R. Lee *et al.*, "Biology of Acinetobacter baumannii: Pathogenesis, Antibiotic Resistance Mechanisms, and Prospective Treatment Options," *Front. Cell. Infect. Microbiol.*, vol. 7, no. March, 2017.
- [30] A. Y. Peleg, H. Seifert, and D. L. Paterson, "Acinetobacter baumannii : Emergence of a Successful Pathogen," vol. 21, no. 3, pp. 538–582, 2008.
- [31] L. Dijkshoorn, A. Nemec, and H. Seifert, "An increasing threat in hospitals: Multidrug-resistant Acinetobacter baumannii," *Nat. Rev. Microbiol.*, vol. 5, no. 12, pp. 939–951, 2007.
- [32] N. C. Gordon and D. W. Wareham, "Multidrug-resistant Acinetobacter baumannii: mechanisms of virulence and resistance," *Int. J. Antimicrob. Agents*, vol. 35, no. 3, pp. 219–226, 2010.
- [33] J. A. Hindler and R. M. Humphries, "Colistin MIC Variability by Method for Contemporary Clinical Isolates of Multidrug-Resistant Gram-Negative Bacilli," vol. 51, no. 6, pp. 1678–1684, 2013.
- [34] H. W. Boucher *et al.*, "Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America," *Clin. Infect. Dis.*, vol. 48, no. 1, pp. 1–12, 2009.
- [35] Y. Smani and J. Pachón, "Loss of the OprD Homologue Protein in Acinetobacter

- baumannii : Impact on Carbapenem Susceptibility," vol. 57, no. 1, p. 2013, 2013.
- [36] M. Catel-ferreira *et al.*, "Deciphering the Function of the Outer Membrane Protein OprD Homologue of *Acinetobacter baumannii*," vol. 56, no. 7, pp. 3826–3832, 2012.
- [37] A. Siroy *et al.*, "Channel Formation by CarO , the Carbapenem Resistance-Associated Outer Membrane Protein of *Acinetobacter baumannii*," vol. 49, no. 12, pp. 4876–4883, 2005.
- [38] A. Mussi, Maria, A. S. Limansky, and A. M. Viale, "Acquisition of Resistance to Carbapenems in Multidrug-Resistant Clinical Strains of," *Society*, vol. 49, no. 4, pp. 1432–1440, 2005.
- [39] M. I. Hood, A. C. Jacobs, K. Sayood, P. M. Dunman, and E. P. Skaar, "*Acinetobacter baumannii* increases tolerance to antibiotics in response to monovalent cations," *Antimicrob. Agents Chemother.*, vol. 54, no. 3, pp. 1029–1041, 2010.
- [40] M. Zahn, S. P. Bhamidimarri, A. Baslé, M. Winterhalter, and B. van den Berg, "Structural Insights into Outer Membrane Permeability of *Acinetobacter baumannii*," *Structure*, vol. 24, no. 2, pp. 221–231, 2016.
- [41] M. E. Jiménez-Mejías, J. Pachón, B. Becerril, J. Palomino-Nicás, A. Rodríguez-Cobacho, and M. Revuelta, "Treatment of multidrug-resistant *Acinetobacter baumannii* meningitis with ampicillin/sulbactam," *Clin. Infect. Dis.*, vol. 24, no. 5, pp. 932–935, 1997.
- [42] A. S. Levin, C. E. Levy, A. E. I. Manrique, E. A. S. Medeiros, and S. F. Costa, "Severe nosocomial infections with imipenem-resistant *Acinetobacter baumannii* treated with ampicillin/sulbactam," *Int. J. Antimicrob. Agents*, vol. 21, no. 1, pp. 58–62, 2003.
- [43] H. M. Lode, "Rational antibiotic therapy and the position of ampicillin/sulbactam," *Int. J. Antimicrob. Agents*, vol. 32, no. 1, pp. 10–28, 2008.
- [44] M. . Kitzis, F. . Goldstein, R. Labia, and J. . Acar, "Synergic activity between sulbactam and clavulanic acid on *Acinetobacter calcoaceticus* .," *Ann. Microbiol. (Paris)*, no. 2, p. pp.163-168 Abstract, 1983.
- [45] P. I. Rafailidis, E. N. Ioannidou, and M. E. Falagas, "Ampicillin / Sulbactam Current Status in Severe Bacterial Infections," vol. 67, no. 13, pp. 1829–1849, 2007.
- [46] H. Chu, L. Zhao, M. Wang, Y. Liu, T. Gui, and J. Zhang, "Original article Sulbactam-based therapy for *Acinetobacter baumannii* infection : a systematic review and meta-analysis," vol. 7, no. 12, pp. 389–394, 2013.
- [47] M. Kempf and J. M. Rolain, "Emergence of resistance to carbapenems in *Acinetobacter baumannii* in Europe: Clinical impact and therapeutic options," *Int. J. Antimicrob. Agents*, vol. 39, no. 2, pp. 105–114, 2012.
- [48] A. Kumar, E. Hajjar, P. Ruggerone, and M. Ceccarelli, "Structural and dynamical properties of the porins OmpF and OmpC: insights from molecular simulations.," *J. Phys. Condens. Matter*, vol. 22, no. 45, p. 454125, 2010.
- [49] T. Bin Masood, S. Sandhya, N. Chandra, and V. Natarajan, "Open Access CHEXVIS : a tool for molecular channel extraction and visualization," pp. 1–19, 2015.
- [50] B. R. Cheneke, B. Van Den Berg, and L. Movileanu, "Analysis of gating transitions among the three major open states of the OpdK channel," *Biochemistry*, vol. 50, no. 22, pp. 4987–4997, 2011.
- [51] B. R. Cheneke, B. van den Berg, and L. Movileanu, "Quasithermodynamic Contributions to the Fluctuations of a Protein Nanopore," *ACS Chem. Biol.*, vol. 10, no. 3, pp. 784–794, Mar. 2015.

- [52] K. R. Pothula and U. Kleinekathöfer, "Theoretical analysis of ion conductance and gating transitions in the OpdK (OccK1) channel," *Analyst*, vol. 140, no. 14, pp. 4855–4864, 2015.
- [53] D. Porschke, "Macroipoles Unusual electric properties of biological macromolecules," *Biophys. Chem.*, vol. 66, no. 2–3, pp. 241–257, Jun. 1997.
- [54] J. Antosiewicz and D. Porschke, "The nature of protein dipole moments: experimental and calculated permanent dipole of .alpha.-chymotrypsin," *Biochemistry*, vol. 28, no. 26, pp. 10072–10078, Dec. 1989.
- [55] G. Vasanthi and S. Krishnaswamy, "Dipole moment analysis of membrane proteins suggests role in orientation in the membrane.," *Indian J. Biochem. Biophys.*, vol. 39, no. 2, pp. 93–100, Apr. 2002.
- [56] D. R. Ripoll, C. H. Faerman, P. H. Axelsen, I. Silman, and J. L. Sussman, "An electrostatic mechanism for substrate guidance down the aromatic gorge of acetylcholinesterase.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, no. 11, pp. 5128–32, 1993.
- [57] C. E. Felder, J. Prilusky, I. Silman, and J. L. Sussman, "A server and database for dipole moments of proteins," *Nucleic Acids Res.*, vol. 35, no. SUPPL.2, pp. 512–521, 2007.
- [58] T. Giorgino, "Computing 1-D atomic densities in macromolecular simulations: The density profile tool for VMD," *Comput. Phys. Commun.*, vol. 185, no. 1, pp. 317–322, 2014.
- [59] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. A. Baker, "PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations," *Nucleic Acids Res.*, vol. 32, no. WEB SERVER ISS., pp. 665–667, 2004.
- [60] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, "Electrostatics of nanosystems: Application to microtubules and the ribosome," *Proc. Natl. Acad. Sci.*, vol. 98, no. 18, pp. 10037–10041, Aug. 2001.
- [61] T. J. Dolinsky *et al.*, "PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations," *Nucleic Acids Res.*, vol. 35, no. SUPPL.2, pp. 522–525, 2007.
- [62] S. Acosta-gutierrez, M. A. Scorciapino, I. Bodrenko, and M. Ceccarelli, "Filtering with Electric Field : The Case of E . coli Porins," 2015.
- [63] S. A. Gutiérrez, I. Bodrenko, M. A. Scorciapino, and M. Ceccarelli, "Macroscopic electric field inside water-filled biological nanopores," *Phys. Chem. Chem. Phys.*, vol. 18, no. 13, pp. 8855–8864, 2016.
- [64] H. Bajaj *et al.*, "Bacterial Outer Membrane Porins as Electrostatic Nanosieves: Exploring Transport Rules of Small Polar Molecules," *ACS Nano*, vol. 11, no. 6, pp. 5465–5473, 2017.
- [65] D. P. Tieleman and H. J. C. Berendsen, "A molecular dynamics study of the pores formed by Escherichia coli OmpF porin in a fully hydrated palmitoyloleoylphosphatidylcholine bilayer," *Biophys. J.*, vol. 74, no. 6, pp. 2786–2801, 1998.

Chapter 3

*On the modelling of human TPC2 channel:
homology modelling and docking simulations*

Introduction

Ion channels are transmembrane proteins that enable the passage of ions across the membrane. One of the most relevant properties of those channels is their ability of switching between open-and-closed states. Several gating mechanisms are processed in the cell membrane to control the channel gating and are classified regarding the external stimulus. Voltage-gated ion channels open and close in response to changes in transmembrane voltage or electrochemical (pH-gated) field. Ligand-gated ion channels open and close in response to a binding of two or more ligand molecules that induce a conformational change in the protein. As a result, a central, water-filled pore opens and conducts ions at high rates, up to 10^7 ions per second [1]. Another mechanism concerns the mechanosensitive channels that are activated by a modification of the pressure made by the lipid membrane on their surface or by other mechanical factors. Here, the membrane-protein interaction plays a crucial role. In this work, we focus on the Two Pore Channels (TPCs) family, a member of voltage-gated channels. We consider, in this thesis, the human TPC2 member channel.

Two-Pore Channels (TPCs) [1] are members of the eukaryotic intracellular voltage-gated cation channel superfamily that contains over 140 structurally related pore-forming proteins [2], including family of voltage-gated Ca^{2+} (CaVs), Na^{+} (Navs), K^{+} (Kvs), and TRP channel family [3]. The membership of TPCs channels to this subfamily of voltage-gated channels is based on the sequence similarity of the pore [4]. In plant, there is only one two-pore channel gene expressed, TPC1 [5] whereas, animals hold three isoforms (TPC1, TPC2, TPC3). Although TPC3 is present in most mammals but mice, rats, and primates [6] (Morgan et al., 2011) [4].

Cellular and sub-cellular localization

Two Pore channels (TPCs) are present in different tissues; in the human, TPC2 is localized in kidney and liver and TPC1 is found in endosomes [6] [5]. Whereas plant TPC1 is present in leaf and stem [8] [9] [10] [11] [12] [6]. Moreover, the sub-cellular localization of plant AtTPC1 and animal TPC2 are found in the membrane of the acidic compartment, which is the vacuole in plants or the lysosome in animals, respectively [13]. This sub-localization has a relevant meaning with respect to their functions.

Topology, function and regulation

Different studies have predicted the topology of TPCs channels containing two, instead of four with respect to other family members, six-transmembrane (6TM) domains[2]. In particular,

phylogenetic analysis revealed that TPCs are most closely related to TRPML and TRPP channels [14] [3] [15] and may represent evolutionary intermediates between TRP channels and voltage-gated Ca^{2+} and Na^+ channels. Recent crystal structure of TPC1 from *Arabidopsis Thaliana* (*AtTPC1*) has confirmed the 3D structural folding of these channels and helped on giving proposition on their structural function based mechanisms. The functional topology of the pore is formed by a homodimer, and each monomer is composed of two Shaker-like subdomains with six transmembrane domains (TM) each and a pore region between transmembrane 5 and 6, and 11 and 12, respectively as shown in the figure 28 [3]. The two Shaker-like subdomains are connected by a cytosolic linker including two EF-hand motifs. Also, both the N- and the C-terminus are exposed to the cytosol [10]. EF-hands are helix-loop-helix structural domains that provide a Ca^{2+} -binding loop between the so-called E-helix and F-helix and change their conformation upon Ca^{2+} -binding [16][3] [15]. *AtTPC1* channel contain multiple positively charged amino acid residues in their voltage sensor domains (S4-S5 TM domain)[17], and negatively-charged amino acid residues in their pore domains, similarly to Nav and Cav channels [18]. These positively charged voltage-sensing motifs in the S4 transmembrane domain are universally present [19], although *HsTPC2* does not appear to be voltage-sensitive, it is not regulated by voltage [20] [21] [22].

TPCs channels play an important role in vesicular fusion and trafficking in endo-lysosomes[23], by promoting the local Ca^{2+} release to regulate the cellular excitability [24]. However, with respect to ion selectivity the two pore channels (TPCs) are controversy[25]. Recently they have been claimed to be sodium selective rather than calcium selective, and that the main cation they release from the endosomes/lysosomes is sodium and not calcium [26] [27] [3]. Indeed, studies have identified residues in the pore loop of human TPC channels that select for Na^+ confirming that these channels are sodium channel with low calcium permeability [26] [27]. Furthermore, they confirmed that *HsTPC2* is a highly Na^+ -selective channel activated by PI(3,5) P2 [26][27]. In the work of [26], they have used the recent structure of *AtTPC1*, to solve mutant structure that contain in its pore selectivity the residues of the *HsTPC2*, by doing so, they converted the non- selective *AtTPC1* to a highly Na^+ -selective, *HsTPC2*-like channel and identified key residues in the TPC filters that differentiate the selectivity between *AtTPC1* and *HsTPC2* [26] [27]. From a structural point of view the most studied TPCs member is the plant *AtTPC1* thanks to the recent availability of the high-resolution structure. It is suggested that the EF-hand motifs of *AtTPC1* are responsible for activation of the channel by binding cytosolic calcium, however, the mechanism in which conformational changes in the EF-hands induces the opening of the

channel gate is still unknown. Whereas, animal TPCs do not have EF-hands in their linker domain and were shown to be activated by the nicotinic acid adenine dinucleotide phosphate molecule (NAADP¹) and/or by the phospholipid (phosphatidylinositol 3,5 bisphosphate (PI (3,5) P2²)) [28] [27] [21] [29]. The activation of TPCs activated by NAADP or by PI(3,5)P2 mediates several function such as the cell differentiation, metabolism, autophagy and in the regulation of pigmentation[30]. They are also implicated in the control of various diseases including control of Ebola virus host cell entry and Parkinson's disease [31] [32] [33].

From the pharmaceutical point of view and in the context of diseases directly involving the function of the animals TPCs channels, only few molecules are proposed to act as inhibitors, mainly due to the lack of knowledge about the structure-function relationship of these channels: their crystal structures are still not available which makes the investigations at atomic scale more difficult. Since the TPCs channels family has been discovered, they have become a major therapeutic target and have attracted the attention of a large number of biophysicists and have

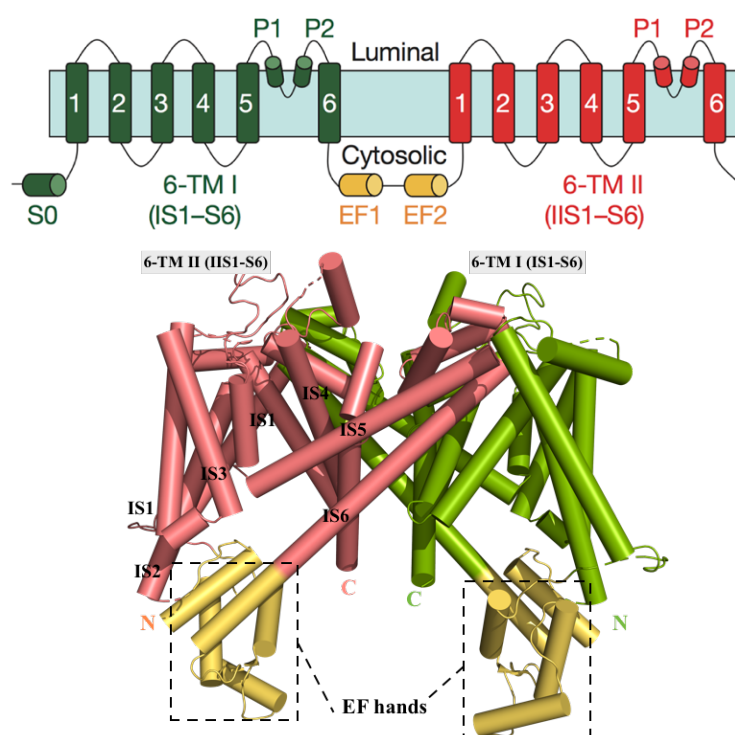


Figure 28: Topology diagram of AtTPC1, shows two 1-6-TM domains (1-6.TM I and 1-6.TM II) and an intervening cytosolic EF-hand domain per AtTPC1 subunit, two of which assemble into a functional channel equivalent to a tetrameric voltage-gated channel. Adapted from (Jiangtao Guo et al., 2015)

¹ NAADP is molecules that acts as a potent Ca²⁺ mobilizing that binds to and gates Ca²⁺ channels in intracellular organelles [12] [63] [64].

² PI (3,5) P2 is a membrane anchored endo-lysosome specific phosphoinositide of low abundance that plays a role in membrane trafficking and regulation of size, shape and acidity of endo-lysosomes [12] [63] [64].

been devoted to a very large number of experimental studies and scientific journals. This fervor is mainly due to their involvement in various diseases[34] [35] [36] [37] [11] [38].

Moreover, based on the similarity of their pore to that of TPCs channels, different families of Ca^{2+} channels inhibitors (blockers) have been reported to be implicated in the inhibition of TPCs channels. Among these molecules, verapamil belongs to Phenylalkylamines (PAAs) class of Ca^{2+} channels inhibitors suggested to have a similar binding site in TPC2 channel [38] [27]. Indeed, it has been shown that PAAs directly block the pore lumen [38] of these channels. Indeed, it was shown that PAAs bind glutamates in P-loop and by this way constrict a pore making Ca^{2+} passing impossible[38]. Another interesting molecule is Ned19 that was recently co-crystallized with AtTPC1 channel from *A. thaliana* and due to the functional similarity between AtTPC1 and HsTPC2 channels, NED19 is supposed to act with a similar mechanism as in AtTPC1. Interestingly, NED19 shows promise in treating cancers [39] and filo-viral infections [34] [33]. Finally, naringenin belongs to a group of flavonoids also known as blocker of potassium channels. It was recently studied experimentally and showed ability to inhibit the TPC2 channel [40]. Although these intense experimental and theoretical investigations, basic relationships between their structure and their function are not understood sufficiently, not even at atomic scale.

The aim of this study

The 3D structure of the human TPCs channels do not exist yet, and it might take time to find the good experimental condition to solve the structure by crystallization experiments. In this case, the use of computational method is appropriate and the only way to build a 3D model. Indeed, since the availability of the X-ray structure of plant TPC1 channel at high resolution (3.3 Å) [3] [41] [27], the construction of a homology model 3D-structure of the unknown Human TPC2 channel is possible, as it has been already largely done in the past on different ions channels[42][43].

As part of this thesis, we carried out docking simulation to study the interaction mode of naringenin and verapamil with TPC2 and investigate their binding site. The results are shown and discussed in the following section. Here, we used the X-ray structure of the mutant AtTPC1 from *A. thaliana* as the 3D-structure template to build the 3D models of human TPC2 channels based on the homology modeling approach (see material and methods chapter for more details). These molecular modeling methods predict and construct the 3D conformations of proteins based on the amino acid sequence homology with the most closed available X-ray structures to

the protein of interest. This 3D model will then serve in the second step as an initial configuration atomic structure for molecular dynamics simulations to study the dynamics and extract the most representative structures to carry out docking simulations, allowing to include conformational changes and flexibility aspect in the docking process. All the details concerning the methods used in this study are described in chapter 4.

The main objective of this study is (i) to build a 3D model structure of the human TPC2 channel, (ii) perform molecular dynamics simulations to analyze the dynamics and the conformational changes that could affect the conformation of the selective pore, (iii) then use the representative structure of the MD as initial structure to dock the selected small molecules by applying molecular docking simulations.

We aim to check whether well-known conventional calcium channel inhibitors (naringenin, verapamil) and the co-crystallized Ned19 with the plant AtTPC1 can have similar potential binding site to HsTPC2 channel.

Results and discussion

I. Model building, structure refinement and validation

The homology model of *HsTPC2* from *Homo sapiens* was generated based on the recent high-resolution (3.3 Å) chimer X-ray structure of *AtTPC1* mutant (pdb: **5TUA**) from *A. thaliana*. The two proteins share about 25 % of sequence identity (see Table IV below). Sequence alignment was generated using the *CLUSTALW* web-server. The Highest sequence identity is observed in the pores filters that are highly conserved in TPC channels (see the figure SIII).

Table IV: Sequence identity (%) performed by CLUSTALW

	<i>AtTPC1</i>	<i>HsTPC1</i>	<i>HsTPC2</i>
<i>AtTPC1</i>	100		
<i>HsTPC1</i>	26.38	100	
<i>HsTPC2</i>	24.93	27.18	100

C-terminal domain(CTD) prediction

As detailed below, we used the program Phyre2 to create a reliable homology model of the CTD of TPC2 human. The global alignment does not cover the C terminal region, which is longer in *HsTPC2* than in *AtTPC1*. In order to improve the modeling of the whole protein, an external server, Phyre2, was used to generate a homology model of the C-terminal sequence



Figure 29: C-terminal alignment used by Phyre2 and added to the alignment used to build the complete structure by MODELLER.

only. Interestingly, the best template detected was a bacterial sodium channel, with 3.8 Å resolution (PDB: 4LTP). Further, the sequence alignment with the best template selected by this server was included in the homology modelling building procedure with *MODELLER*. The structure detected by Phyre2 shares 26% of sequence identity with C-terminal of HsTPC2 (Figure 29). Experimentally, the C-terminal channel seems to play a key role for the functionality of the HsTPC2 channel. A recent paper from *Christopher J. Penny and all 2016* [44] studied the functionality of the pore-only domain. They reproduced the activity of the channel by using only the central pore region including the selectivity filters and the gated pore of the two-pore domains in which they have removed the SV domains as well as the EF-hand domain (fig.28). The figure 30 shows the last model obtained by *MODELLER* by using the 5TUA and 4LTP proteins as templates.

The Quality of the model

The stereo-chemical quality of the model shown in Figure 31 has been checked using the PROCHECK v.3.5 server [45]. The stereo-chemical quality plots and residue-by-residue listing main-chain parameters are shown in the figure 30 and Table V.

Table V: Parameters of the TPC2 homology model validation as obtained with the PROCHECK server

Protein model	Template From plant <i>A. thaliana</i>		Sequence identity (%)	Procheck analysis of the Ramachandran plot details for chain (A, B), the numbers indicate the number of amino acid residues, with the corresponding %, reported in the figure 29				
	protein	Pdb code			Most favored regions (red regions)	Additional allowed regions (yellow regions)	Generously allowed regions (light yellow region)	Disallowed regions (white region)
hsTPC2	AtTPC1	5TUA	25.0	A	588 (90.9%)	42 (8.0%)	9 (0.5%)	8 (0.8%)
				B	586 (90.6%)	52 (8.08%)	3 (0.46%)	6 (0.92%)

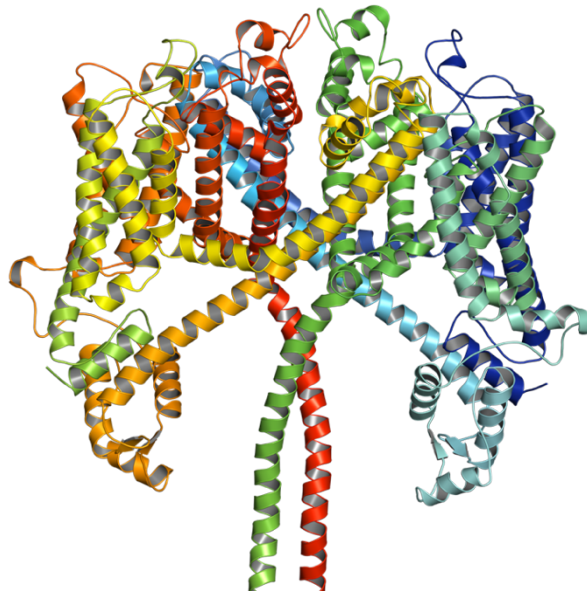


Figure 30: 3D-structure of the homology model of the Human TPC2 channel.

As we can see in the table 4 and figure 29, the Ramachandran plot of the TPC2 homology model represent 90% of amino acid residues falling in the allowed region indicating a good local quality of the 3D-structure of the model. Moreover, it has been reported that when several models are calculated for the same target, the best model can be selected by picking the model

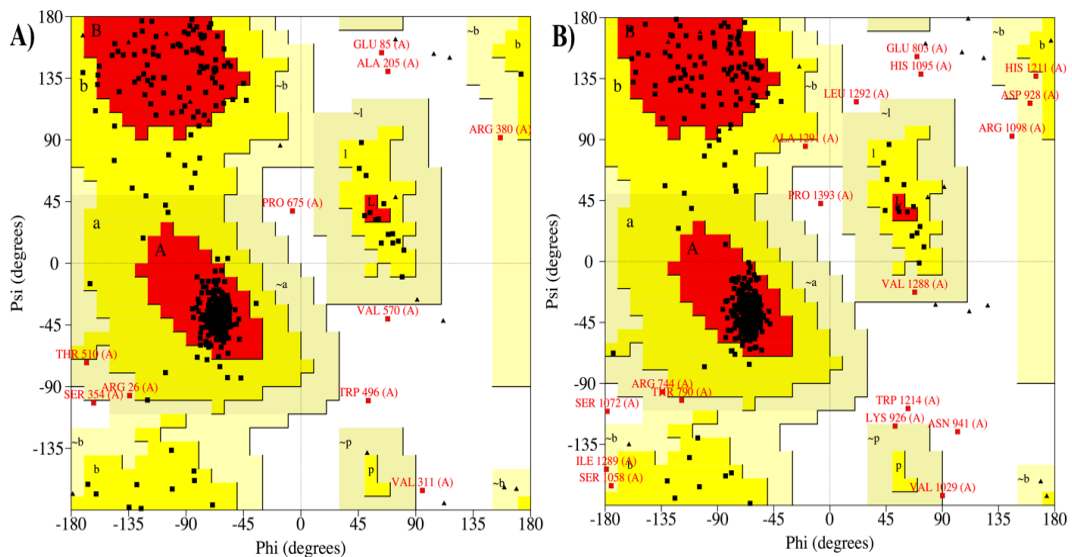


Figure 31: Phi-psi and chi1-chi2 distributions Main Ramachandran plot, White region are sterically disallowed for all amino acids except glycine. The red regions correspond to allowed regions namely the α -helical and β -sheet conformations, the yellow areas correspond to outer limit (G. N. Ramachandran., 1963), The red dots on the plot correspond to the *outlier's* residues

with the lowest value of the *MODELLER* objective function of the DOPE (Figure 32)(Discrete Optimized Protein Energy) [46] [47]. The DOPE potential in *MODELLER* is used to evaluate the fold of the selected model. However, this score is not an absolute measure, in the sense that it can be used only to rank models calculated from the same alignment [46]. In our case, over 1000 models were generated and we selected only the top scored one for our subsequent investigation.

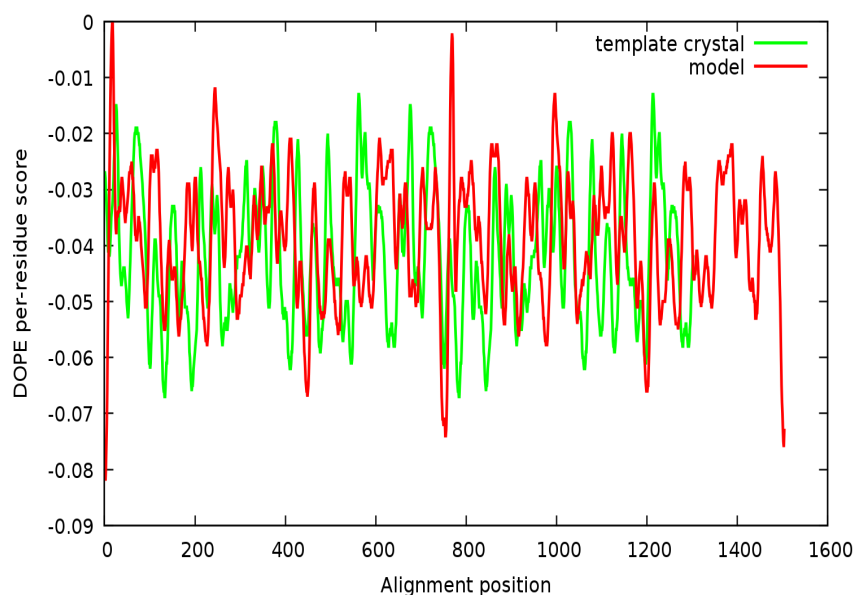


Figure 32: Plot of the DOPE per-residue score obtained from *MODELLER* for the TPC2 model compared to the crystal structure of TPC1 from plant

Pocket detection using the PockDrug server

The druggability of a protein corresponds to its ability to bind a drug-like molecule, orally available, with a good affinity [48][49][50]. Drug-like molecules are defined by Lipinski as follows “Drug-like are defined as those compounds that have sufficiently acceptable ADME properties and sufficiently acceptable toxicity properties to survive through the completion of human Phase I clinical trials” [51]. The PockDrug website uses a PockDrug model that is an optimized druggability model developed by *Borrel, et al.*[52]. It is obtained from three statistical steps: (i) optimization of linear discriminant analysis models combining a selection of pocket descriptors from a pool of 52 geometrical and physicochemical descriptors by cross-validation using a training part of NRDL set estimated by fpocket, (ii) the selection of the even most stable and efficient models using NRDL independent test set estimated using four different pocket estimation methods, (iii) the construction of one unique consensus druggability

model from the best models. The PockDrug server provides, as output, one druggability probability corresponding to the average of the probabilities of the seven best models and its associated standard deviation. In the table 5 below the pocket predicted for the TPC2 transmembrane regions of the model are listed. The ranking of the pocket is based on their drugability probability and are represented in the Figure 33 [53] [52].

Table 3: Druggable pockets (P#) in TPC2 channel homology model detected by PockDrug program.

Pockets	Vol. Hull*	Hydroph. Kyte*	Polar Res.*	Aromatic Res.*	Otyr atom	Nb. Res.*	Drugg Prob*	Standard Deviation
P 0	32277.5	1.65	0.31	0.21	0.0	160.0	1.0	0.0
P 20	1209.1	2.32	0.22	0.11	0.0	18.0	1.0	0.0
P 21	2108.59	1.66	0.36	0.14	0.01	28.0	1.0	0.0
P 29	743.01	1.64	0.43	0.21	0.0	14.0	1.0	0.0
P 58	738.88	1.91	0.25	0.06	0.0	16.0	1.0	0.0
P 63	1257.73	2.76	0.13	0.19	0.0	16.0	1.0	0.0
P 24	1391.32	1.43	0.38	0.13	0.0	16.0	0.99	0.0
P 4	5750.39	1.21	0.29	0.15	0.0	41.0	0.99	0.0
P 40	1464.92	1.04	0.36	0.14	0.02	14.0	0.99	0.0
P 42	560.27	1.49	0.29	0.21	0.0	14.0	0.99	0.0

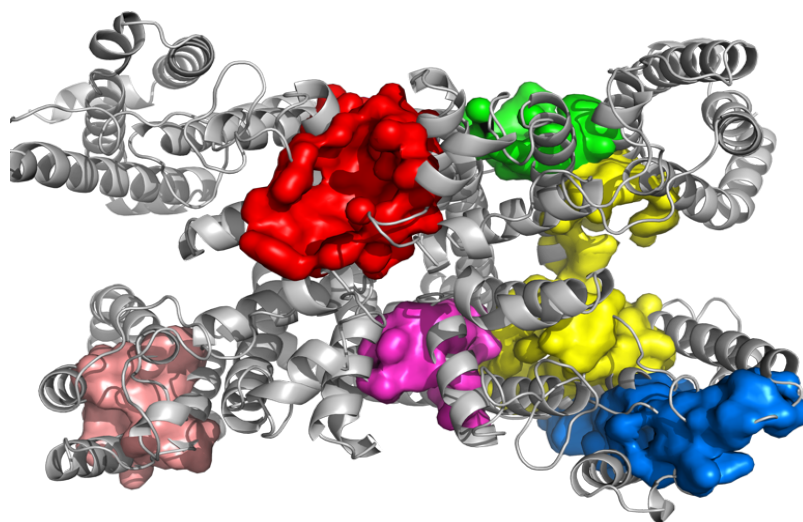


Figure 33: Pockets detected with the PockDrug webserver. The pocket list is ranked according to the druggability probability. The different pockets are represented as colored surfaces.

II. Molecular Docking simulations

Molecular docking is used to predict the relative configuration and the associated energetics of a small molecule (ligand) and a big macromolecular target (receptor). Here, as possible ligands targeting TPC2 we considered two molecules: naringenin and verapamil, whose relevance has been described before. In particular, verapamil (PAA) was the first calcium channel blocker to be used in clinic.

Validation of the docking protocol

Our docking protocol can be considered valid and reliable if the method used is able to reproduce the binding mode of the small molecule co-crystallized with the AtTPC1 structure. To do so, we removed the small molecule (trans-Ned19) from its active site in the structure of AtTPC1 channel and re-docked it back to the same structure. The conformation and the orientation of the docked trans-Ned19 were compared with the conformation of the original X-ray structure.

The docking results showed that the RMSD is within 1.91 Å of the co-crystallized conformation, with an estimated free energy of binding of -6.58 kcal/mol and the estimate inhibition-constant, $K_i = 4.90 \mu\text{M}$. To note that, the reported K_i values are based on theoretical prediction by the Autodock algorithm.

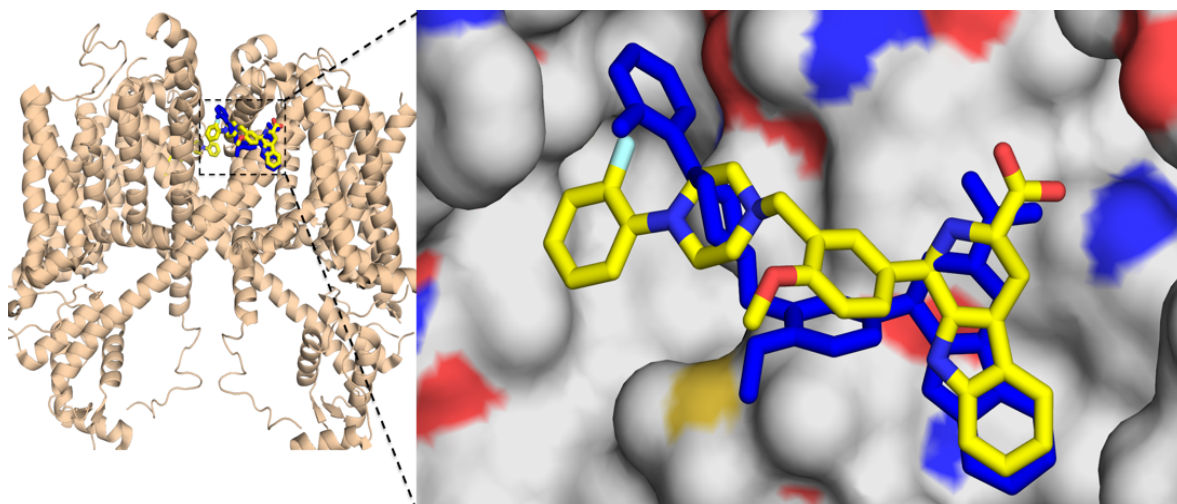


Figure 34: superposition of the predicted docking pose of trans-Ned19 (blue sticks) in the crystal structure of TPC1 and the corresponding X-ray configuration (yellow sticks). The crystal structure of the protein is displayed in cartoon representation

The Figure 34 displays the superposition of the two poses in their binding site of the crystal structure of TPC1. The ability to re-dock trans-Ned19 in the know X-ray configuration confirms that the docking method adopted can be used to dock systematically the selected molecules to TPC1.

Docking results of naringenin and verapamil

In the following steps, we used 2 different procedures of docking, namely docking I and II.

- I) In the docking I protocol, we used a blind docking with a box size (124 Å, 124 Å, 126 Å) that includes the entire protein, the results are shown in Table 4. Then we refined the results of the previous docking by centering the box size on the top scored previous clusters and reduced the size of the box (50 Å, 50 Å, 82 Å) for verapamil and (45 Å, 45 Å, 62 Å) for naringenin to include only the binding site and run 1000 docked run.
- II) In the docking II protocol, we performed two docking in which we suppose a priori that the two molecules bind specifically to two specific regions; i) one by considering only the corresponding voltage-gated (VSD) region and of the channel, ii) the other by considering only the central pore of the channel including the selective filter as well as the gating pore. For these specific regions, the quality of the model is more confident in that region and stable compared to the external and internal regions that share low similarity with the plant structure and are supposed to be more flexible and unstable. For each docking i) and ii) we used box size of (78 Å, 36 Å, 126 Å) and (78 Å, 78 Å, 126 Å) respectively

Results of the docking I: blind docking

Overall, all the docking simulations of the naringenin and verapamil revealed several binding sites in different location in the channel, as shown in Figure 35. However, according to the predicted binding-energy and the predicted K_i (inhibition-constant), naringenin is more active than verapamil. As revealed in the first step of blind docking, the result showed that the best binding site for naringenin is located in the corresponding EF-hand region of the protein in the plant AtTPC1. The binding-energy of the top scored cluster is -6.6 kcal/mol. In the case of verapamil, the binding site is located at the interface between the TM channel and the membrane. In this pose, the molecule shows lower binding energy than that of naringenin, and is 5.1 kcal/mol (see Table VI). Then we refined the previous top scored cluster of each docked molecule by performing a docking (1000 docking run) in which the docking size box including only the binding site of the previous top scored cluster centers. For the two ligands, the energies

of binding are improved from the last docking, and the best docked molecule is naringenin with the best docking energy of -8.2 Kcal/mol (Table VII). The best pose for each molecule is shown in the figure 36.

Table VI: Docking I analysis of the blind docking, the pose of the best cluster is reported, obtained from docking of each inhibitor against the HsTPC2 model

Ligand	binding-energy (kcal/mol)	Enclosed contacts with distance cutoff of 3.2 Å
Naringenin	-6.6	loop EF-hand: V786, V1117, L1118, Q1119, HIS1163, PRO1164, PRO1165, ARG1166, PRO1167, E1168, ARG1246, ILE805, GLN810, ARG817, MET797, GLN800, ASP807
Verapamil	-5.1	VS loop: V919, S920, V922, D923, V926, L930, E934, P935, L936, R937, I938, R939, L942

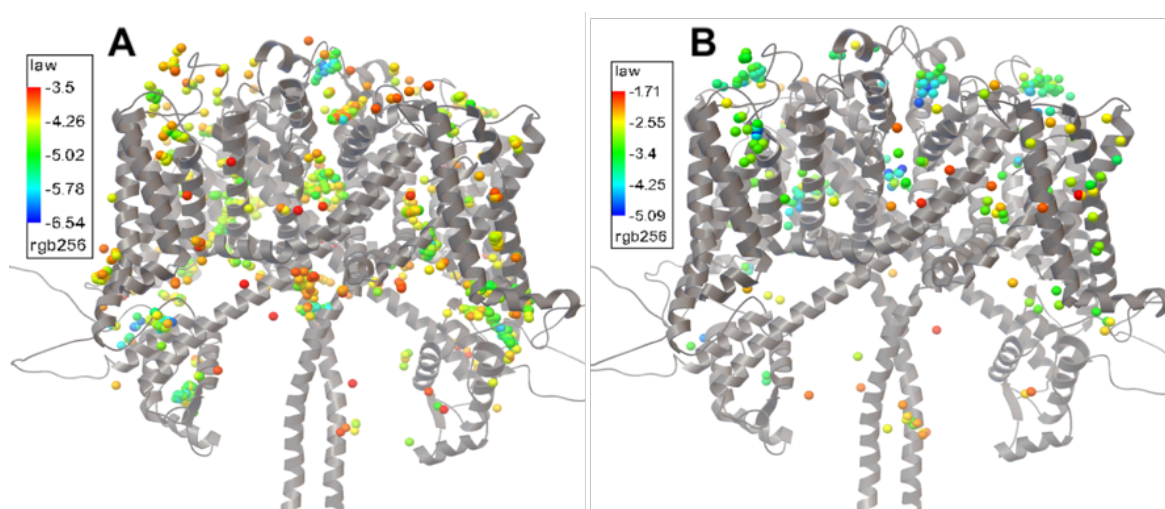


Figure 35: Docked molecules of naringenin (A) and verapamil (B) in the entire TPC2 channel homology model. The obtained scored clusters in the docking I (blind docking) are represented as sphere and colored according to their binding energy in kcal/mol.

Table VII: Docking analysis of the top scored cluster obtained from docking of each molecule against HsTPC2 model channel

ligand	kcal/mol	Predicted Ki (μM)	binding residues	Hydrogen bonds
naringenin	-8.2	0.947	G765, G766, V786, P788, A792, L796, L1114, V1117, L1118, P1167, E1168, Y1169	(EF-hand-loop) E1168 G766
Verapamil	-6.6	15.06	V19, V922, D923, V926, S927, L930, E934, P935, L936, R937, I938, R939	R937

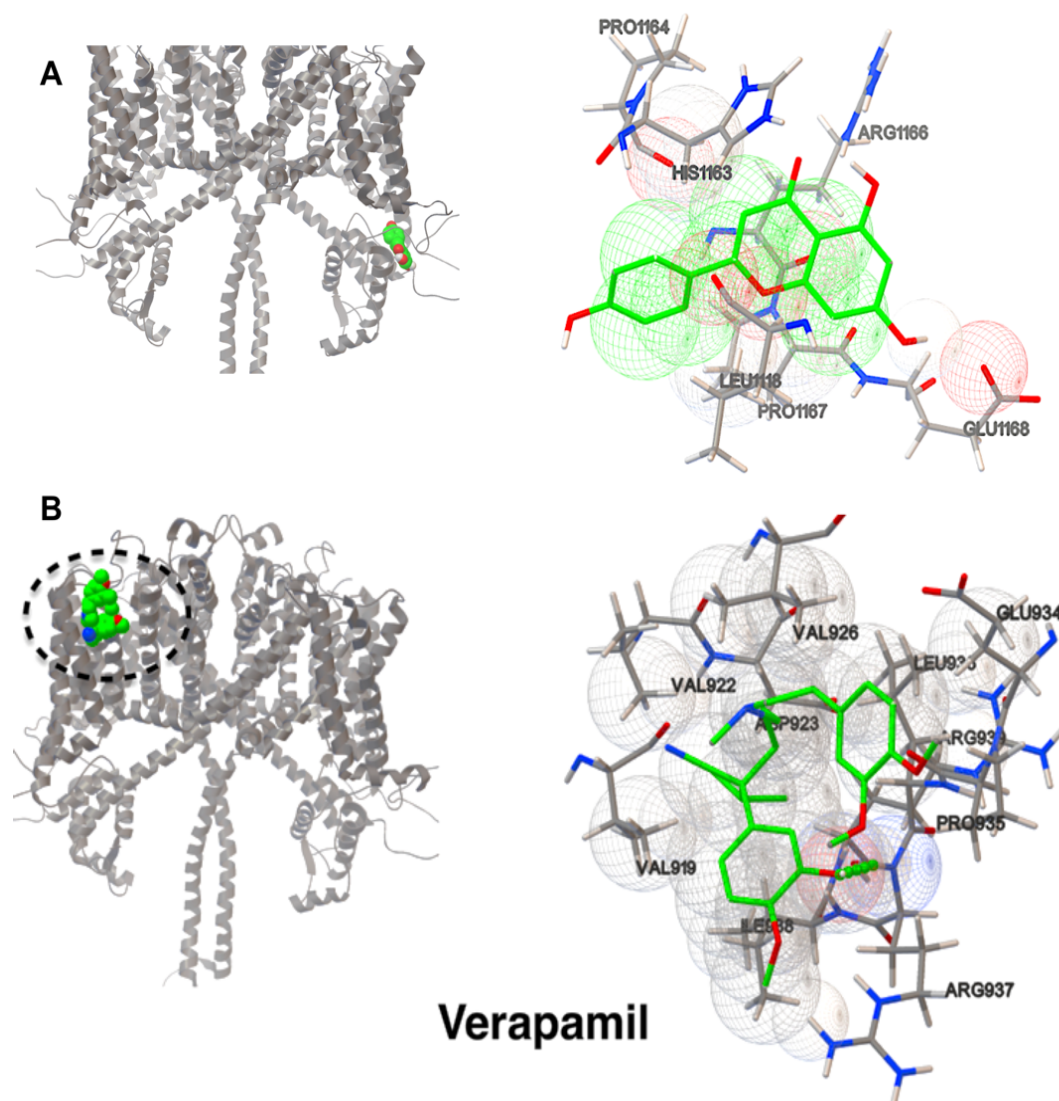


Figure 36: Molecular docking of (A) Naringenin, and (B) Verapamil into the TPC2 model. The left panel of the figure shows the binding pockets. The right panel corresponds to 3D representation of the molecules' interactions in the receptors, the residues involved in the active site are labelled and represented as sticks, the hydrogen bonds are represented as green spheres. The ligands are represented as green sticks.

At this stage, the docking simulations showed rather low energy and a large number of potential binding sites, especially in unordered regions such as the external loops, which in general show high flexibility that might affect the predicted pose. In this case the results are somehow truncated or at least cannot be reproduced. The other problem identified at this stage is due to the large sampled space due to the big size of the protein. The main risk in this procedure is to include inadequate conformations of the high flexible regions such as the external loops, where even minor conformational changes of the receptor are expected to have a major influence on the binding process. On the other hand, studies have been conducted by applying docking on similar molecules on the TPC2 by considering only the central pore cavity, this obviously

supposed as a binding site for molecules blockers due to the implication of the central cavity in the selectivity of the ion diffusion [44]. Thus, in order to improve the revealed poses in the binding sites, we limited the docking simulation by using a smaller box including only the binding sites. In case one knows in advance the binding pocket smaller the box, more the results are confident and accurate.

Results of the docking II (selective filters & voltage-sensitive domains (VSD))

In this step, we restricted the docking search space on small regions of interest in the protein. Several studies have shown, by point mutations [41], the implication of the pore filter in slowing the permeation of ions or inactivation of the channel [54]. Several cases of blockers including verapamil have been shown to bind in the selective filter or in the pore-central hydrophobic pocket and thus block the channel activity [55] [56] [57].

Table VIII: Results of the top scored clusters obtained from docking II for each molecule docked against *HsTPC2* model in both the Pore center and the VSD regions.

	Naringenin		Verapamil	
	Pore center	VSD regions	Pore center	VSD regions
Initial structure	-7.68	-7.46	-5.44	-5.93
Energy kcal/mol	-8.25	-8.23	-7.29	-7.89
Ki predicted (μ M)	0.899 (experimental: $180 \pm 20 \mu$ M)	0.932	4.56	1.64
Hbonds	ASN1406, SER1434	GLY766, GLU1168	ASN305, ASN1439	No hydrogen bond formed
Closed contacts with a distance cutoff of 3.2 Å	VAL651, THR1023, VAL1438, VAL1435, VAL652, TRP1430, TRP1407, ASN653, THR1022, GLN1408, VAL1404, ALA1024, SER1434, ASN1405, MET1056, ASN1406	PRO788, LEU1114, VAL1117, ALA792, LEU1118, GLU1168, GLY766, TYR1169, PRO1167, VAL786	VAL1435, THR271, MET304, ASN305, VAL1438, SER300, LEU1021, TRP1430, LEU1442, VAL1403, LEU1059, ASN1439, LEU301, SER1434, MET1059	LEU1302, PHE1290, PHE295, LEU229, PHE1256, ILE298, PHE291, ARG1306, PHE1305, SER1291, VAL1328, PHE230, LEU303, HIS226, VAL294, ILE1303

The distribution of all the clusters obtained from this docking for each molecule of naringenin and verapamil are illustrated in Figure 37. According to the predicted binding energy, naringenin seems to prefer the EF-hand domain, in particular the loops connecting the VSD channel and the EF-hand region (blue color of the panel) (see Figure 37). The results of the docking are reported in Table VIII.

The docking II results revealed that naringenin has the lowest binding-energy to the pore cavity of TPC2 channel, indicating its affinity towards this protein. Naringenin displayed -8.25

kcal/mol free binding energy and only 0.899 μ M inhibition-constant predicted by AutoDock4.2. In the docking analysis, the interaction involved in the binding site are reported in Table 8 and located in the selective filter II formed by the pore P1-II and P2-II (magenta colored loops in Fig.38 B and D) forming the subunit of the central pore (Figure 38). As shown in Figures 38 (A, B) and fig. 39A, naringenin binds to the filter by making hydrogen bond interactions with residues ASN1406, VAL651 and SER1434. This vertical position (parallel to the axis of the ions diffusion) of naringenin might block the entry of the channel thus inactivating the ion current.

The docked structure of naringenin in the VSD region including the EF-hand appeared most stable in the loop between the VSD domain and the EF-hand (Fig.40 A) with a binding energy of -8.2 kcal/mol and $K_i \approx 0.93 \mu$ M. It should be noted that the binding energy in this region of the TPC2 channel is similar to that of the central pore (selective filter II). This indicates that naringenin has at least two possible binding sites. The closed interactions (with distance cutoff of 3.2 Å) of the complex protein-naringenin and the corresponding docking poses are shown in the Figure 41.A and Table 8.

The molecule of verapamil showed the highest affinity with -7.9 kcal/mol binding free energy and 1.64 μ M inhibition-constant predicted in AutoDock. The binding site is located in between the transmembrane helices IS4 and IS5 (fig.40 B) that have been proposed to play a role in the mechanism of opening the channel in the plant AtTPC1. In this conformation, no hydrogen bonds are formed, the main interaction being hydrophobic contacts and based on geometrical complementarity (Fig.41.B). In addition, this binding site of verapamil was predicted by the program PockDrug as a large druggable pocket (yellow pocket in Fig.33).

At this position, verapamil could serve as a mechanical blocker of a possible movement of the two helices IS4 and IS5, thus blocking the opening of the channel.

The second binding site of verapamil in TPC2 channel is located in the central cavity of the channel as shown in Fig.38.C and D, with a binding energy of -7.3 kcal/mol and inhibition constant of 4.56 μ M (Table 8). In The second pose, verapamil formed 2 hydrogen bonds with the central pore filters, residues ASN305 and ASN1439 as shown in Figure 38.D and Fig.39B. Here, two binding sites are possible for the binding of verapamil into TPC2 channel with similar binding energies but slightly different inhibition-constants. That of verapamil docked in the

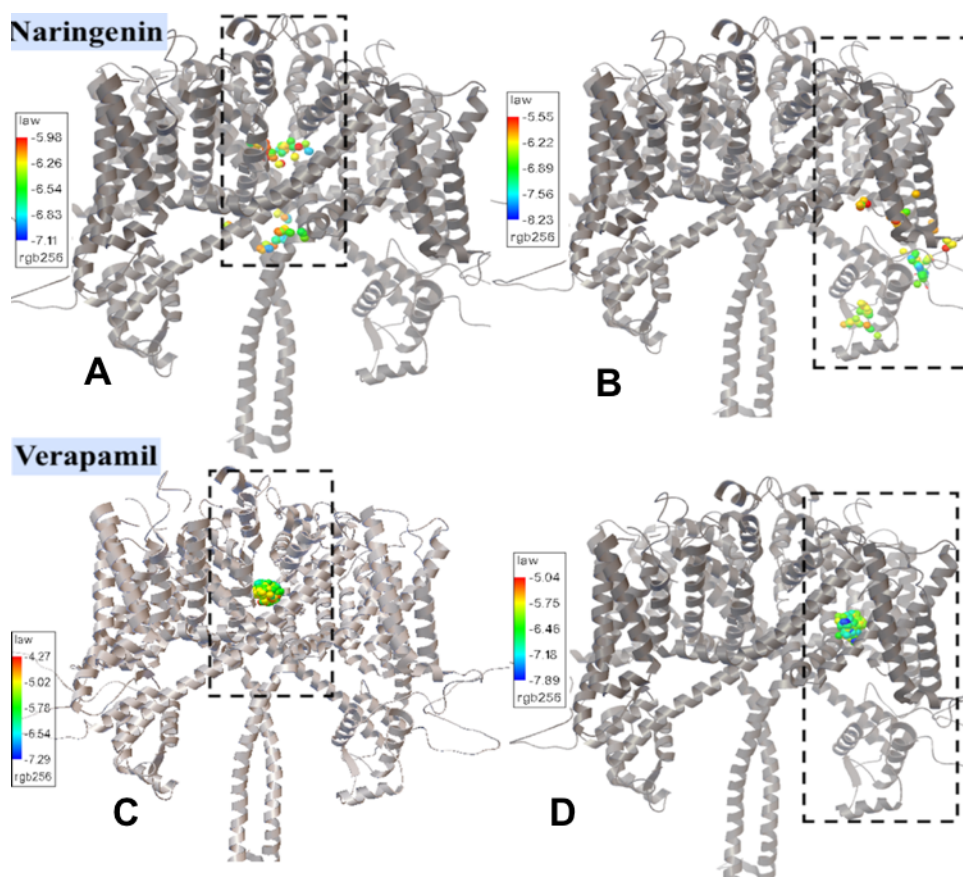


Figure 37: Docked molecules of naringenin (top) and verapamil (bottom) in the selective and cavity filters (A, C) and in the VSD (B, D) of TPC2 channel homology model. The obtained scored clusters in the docking II are represented as sphere and colored according to their energy of binding in kcal/mol.

VSD is lower, less than $2\mu\text{M}$, indicating a strong potentiality to inhibit the channel activity. To note that the binding energy model implemented in AutoDock exhibits a residual standard error of 2 kcal mol^{-1} .

On the other hand, different studies have shown that verapamil, being structurally similar to tetrandrine molecules blocker of gated-ion channels that bind in the central pore of the channel, it likely might bind TPCs in a similar location [58], which fit with our docked structure in the central cavity (fig.38C,D, 39B).

Overall, in the pore binding site we can observe differences between the binding energy of naringenin (-8.3 kcal/mol) and verapamil (7.3 kcal/mol), in particular, a significant difference is observed for the K_i which is five times lower for the docked structure of naringenin ($0.89\mu\text{M}$) than that of verapamil ($4.65\mu\text{M}$). Moreover, several similar molecular docking studies on the gated ions channels blockers, such as potassium channel hERG, have reported the existence of three principle conformations of the molecules binding the channels cavities with respect to the diffusion axis: “curled”, parallel and perpendicular [59] [60] [61] [62]. From our data, the

docked structure of naringenin can be considered rather in the parallel position with respect to the filter II of TPC2 channel (Fig.38 A, B).

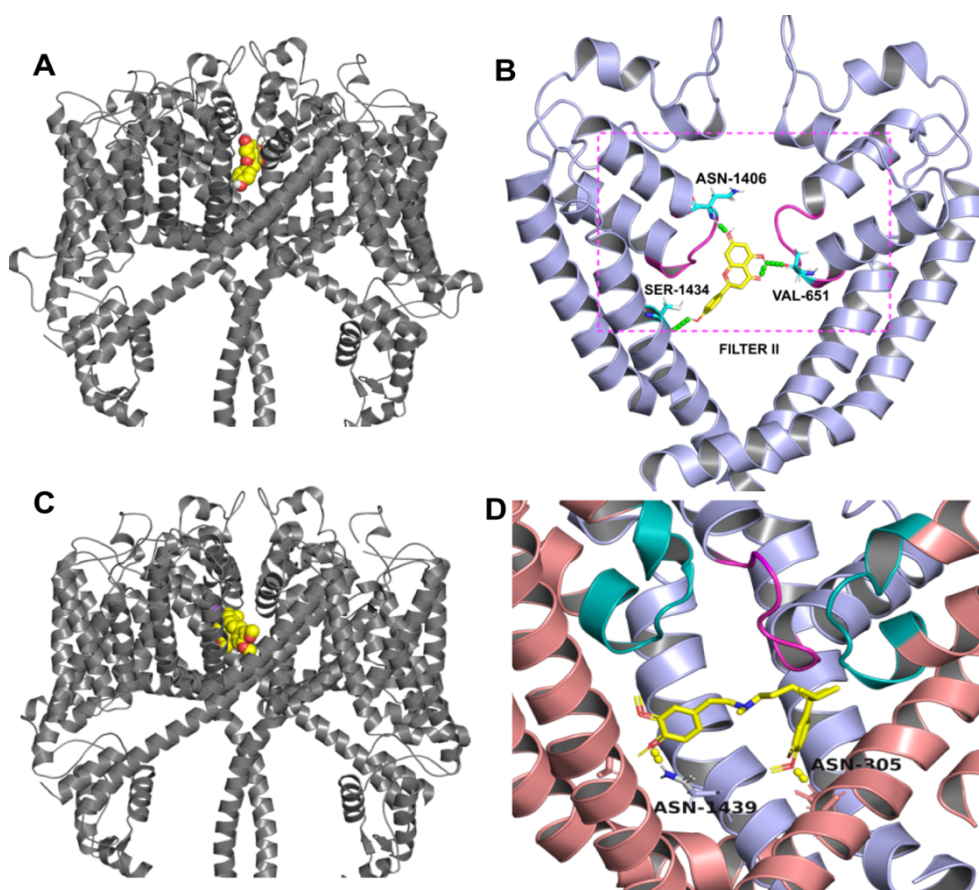


Figure 38: Docking of (A, B) Naringenin and (C, D) verapamil within the inner cavity and the selectivity filter of a TPC2 channel homology model. Side view of the top scored cluster. The TPC2 channel is shown as a cartoon representation. Naringenin and verapamil are represented as yellow sticks. The cyan sticks represent the residues implicated in hydrogen bonds with naringenin. The magenta and cyan colored region just above the two molecules represent the loops of the selectivity filter II and filter I of TPC2 channel respectively.

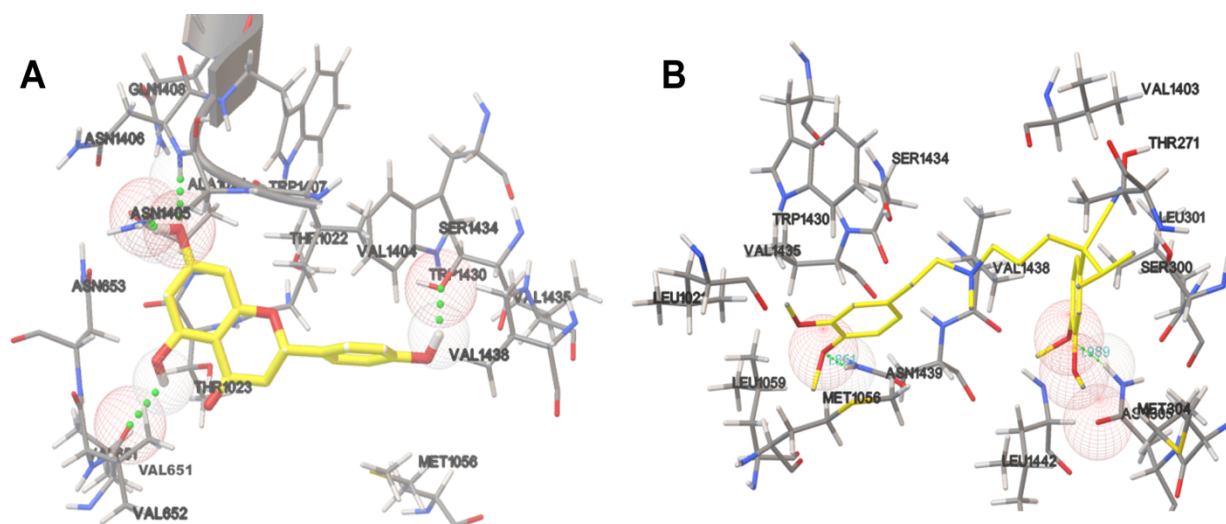


Figure 39: Docked structure of (A) naringenin and verapamil (B) (yellow sticks) in the pore cavity of the model of TPC2 channel. The closed contacts are shown as gray sticks and those involved in hydrogen bonds are highlighted by a sphere, the hydrogen bonds are represented as dashed green line. The cartoon representation juts above the molecule of naringenin corresponds to the loop of the selective filter II of the channel.

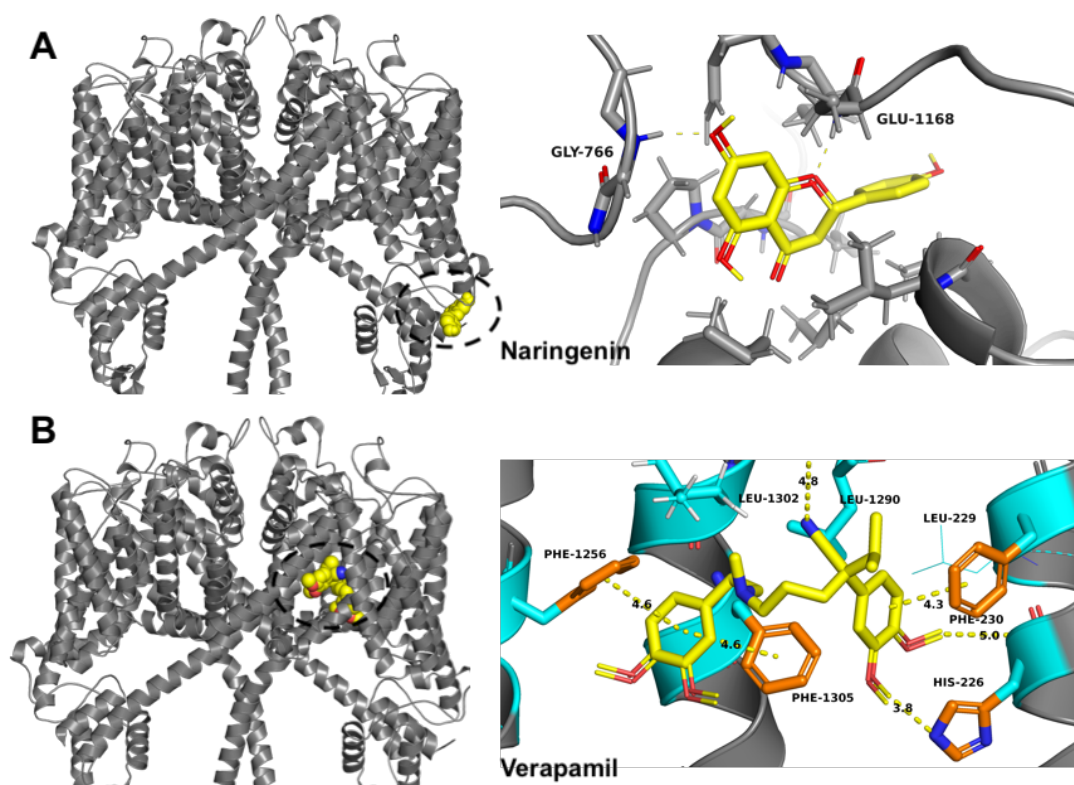


Figure 40: Docking of (A) Naringenin and (B) verapamil within the VSD regions of TPC2 channel homology model. Side view of the top scored cluster. The TPC2 channel is shown as a cartoon representation. Naringenin and verapamil are represented as yellow sticks. The cyan sticks represent the residues implicated in hydrogen bonds with naringenin. The magenta and cyan colored region just above the two molecules represent the loops of the selective filter II and filter I of TPC2 channel respectively.

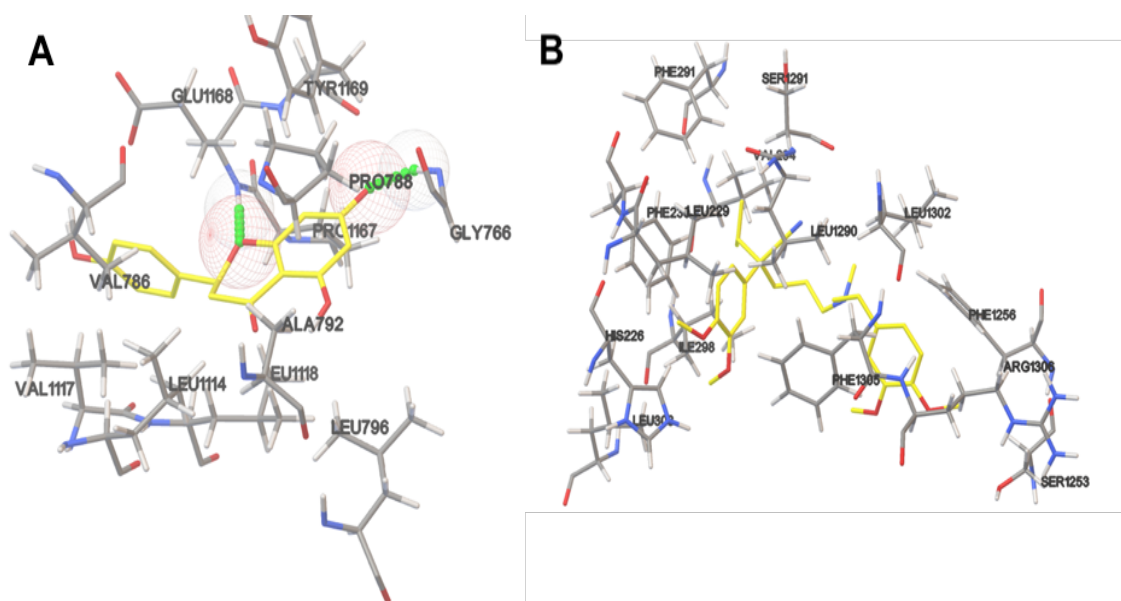


Figure 41: Docked structure of Naringenin (A) and verapamil (B) in the VSD regions of the model of TPC2, hydrogen bonds are represented by green dashed red line, the small molecules are represented as yellow sticks and gray colored stick are closed residues in the binding site. In each representation model the oxygen is colored in red and the atom of nitrogen are colored in blue. The sphere representations highlight the residues implicated in hydrogen bonds.

Conclusion and perspectives

In this study, we succeeded in constructing a reliable 3D model of the TPC2 protein based on the crystal structure of the plant's TPC1 protein. The Ramachandran plot of the TPC2 homology model represented 90% of amino acid residues falling in the allowed region indicating a good local quality of our human TPC2 channel model and showed good DOPE profile (Fig.32) with respect to that of the template. These results indicate the ability of the homology modeling methods to build a model of 3D-structure based on the sequences similarity even if the percentage of the similarity is not very high. In addition, we showed here the possibility of using homology modelling technique to model separately different regions and combine the modeled regions to construct a complete model. Moreover, the PockDrug server was able to detect pockets that were found by some small molecules selected for docking (the VSD binding site of verapamil shown in fig.40 B).

Furthermore, we carried out molecular docking in which we used two different small molecules; naringenin and verapamil. Our results show that the both molecules can bind to the TPC2 channel. In the pore central cavity, both molecules show binding with similar affinity but with different positions; naringenin has a parallel pose with respect to the axis of the selectivity filter (axis of diffusion) whereas verapamil showed rather a “curled”-perpendicular pose with

respect to the selectivity filter. In addition, the docked structure of naringenin is stabilized by interactions of the residues in the filter II whereas verapamil involves residues in the filter I. In the VSD regions of the protein, the binding sites predicted for the two molecules are different but with similar binding energies; naringenin prefers to bind in the EF-hands loop relying the VSD and EF-hand domain, yet the inhibition-constant are similar (Table VIII).

Given the very low predicted K_i in all the binding sites, naringenin has a higher affinity and is more potent to likely inhibit the channel than verapamil. However, further studies are needed to confirm such results. Further docking runs on the clusters extracted from the MD trajectories are needed to check whether the flexibility of the residues will not rearrange the detected binding site and thus modify the affinity of these molecules. By using conformations from MD simulation as starting structure for docking runs we will potentially reveal binding sites that are not accessible in one conformation of the TPC2 homology model. This might fit better the selected small molecules to their active sites previously detected. In addition, experimental studies, such as mutagenesis of the involved residues in the binding sites, might confirm the mechanism of the interaction revealed in this study.

Appendix

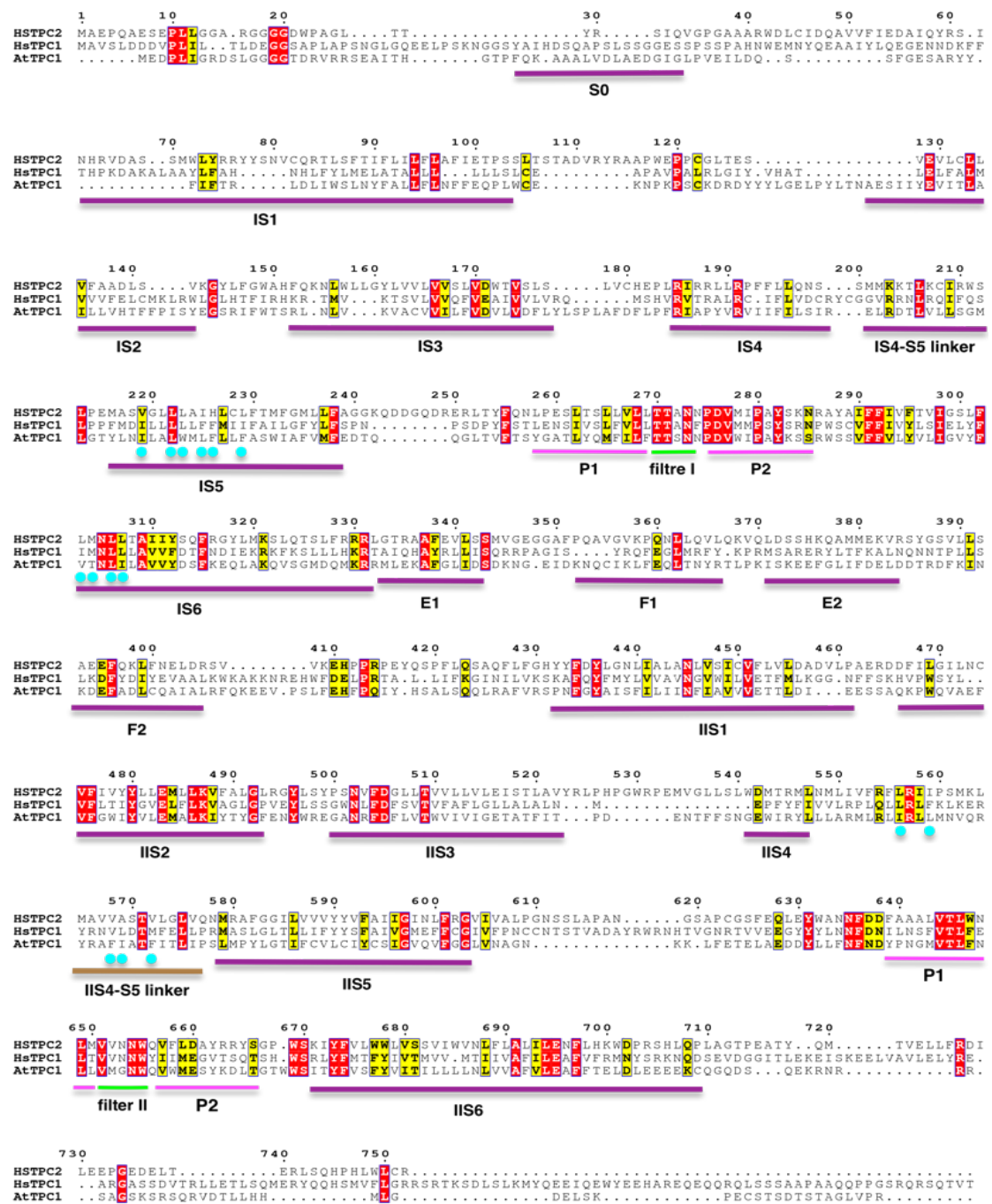


Figure S III: multiple alignment, using clustalW, of the Human TPC1 and TPC2 and plant AtTPC1 sequences.

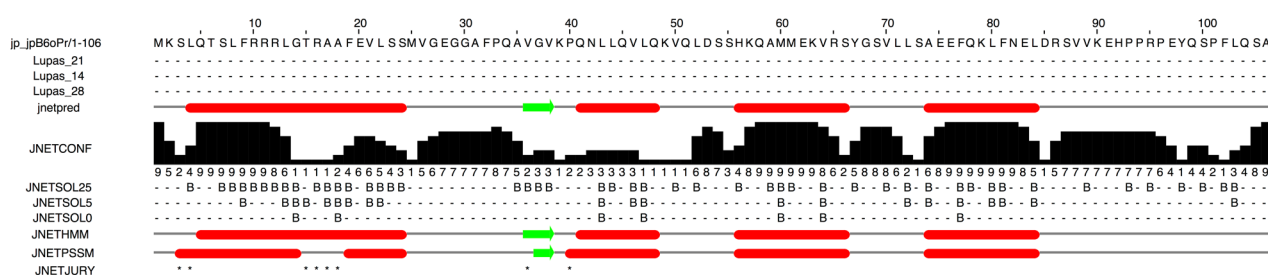
EF-hand domain prediction

The main difference between plant and animal TPCs is the presence of two EF hands in the linker region connecting the two domains in plants, which in other proteins bind Ca^{2+} . Indeed, it is established that the SV (slow voltage) currents in plants, which are mediated by TPCs, are Ca^{2+} -regulated. The lack of EF hands in animal TPCs is consistent with the insensitivity of NAADP-evoked Ca^{2+} release to cytosolic Ca^{2+} [11].

It is important to stress that Jpred (ref) does Skip searching 3D structures in the PDB before prediction. Four helices preceded by a beta structure were predicted with Jpred and are similar to those generated by MODELLER based on the sequence identity of EF-hand from AtTPC1 (fig.SII). It seems that despite the low sequence identity between the Calmodulin and other members of the EF-hand family, MODELLER was able to generate in HsTPC2 the four helices folded in these structures (fig.SIII). However, the negatively charged residues, conserved between the EF-hand family as well in AtTPC1 channel, located in the calcium binding sites are not conserved in the Human TPC sodium channels as shown in the figure SIII.

Percent Identity Matrix created by ClustalW

The multiple sequence alignment between the sequences of EF-hand and Calmodulin proteins collected from the *PDB* and the EF-hand domain from *HsTPC1&2* from Human and *AtTPC1* from plant AT are reported in the figure SIII.



EF-hand domain II structure prediction

Figure S IV: Secondary structure prediction of the HsTPC2 EF-hand sequence by Jpred sever. The helices are shown in red tube and green represent beta sheet.

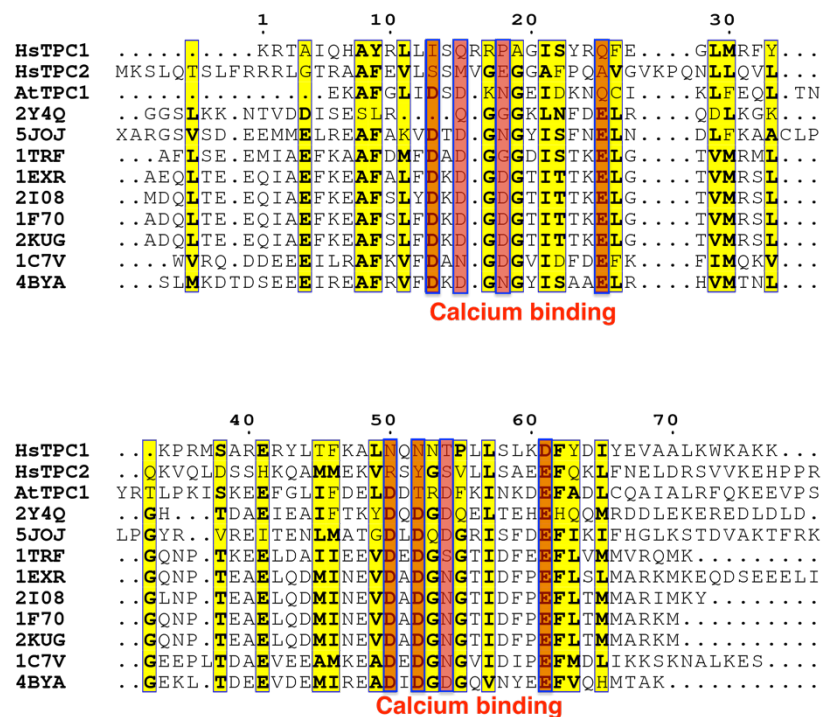


Figure S V: Multiple sequence alignment of the EF-hand available structures (Calmodulin) and TPC2 Ef-hand from TPC channels

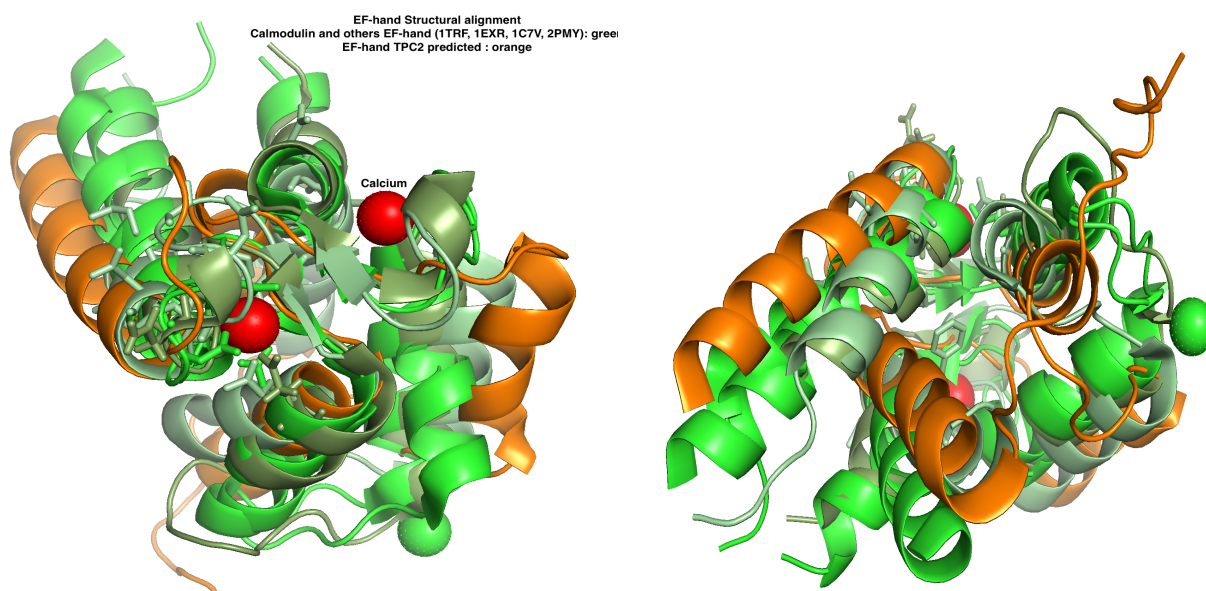


Figure S VI: Multiple structural alignment of EF-hand family (green) and the HsTPC2 modeled (orange). The Calcium is shown as a red sphere. The most helices predicted by Jpred web server are as well predicted by internal MODELLER modeling program.

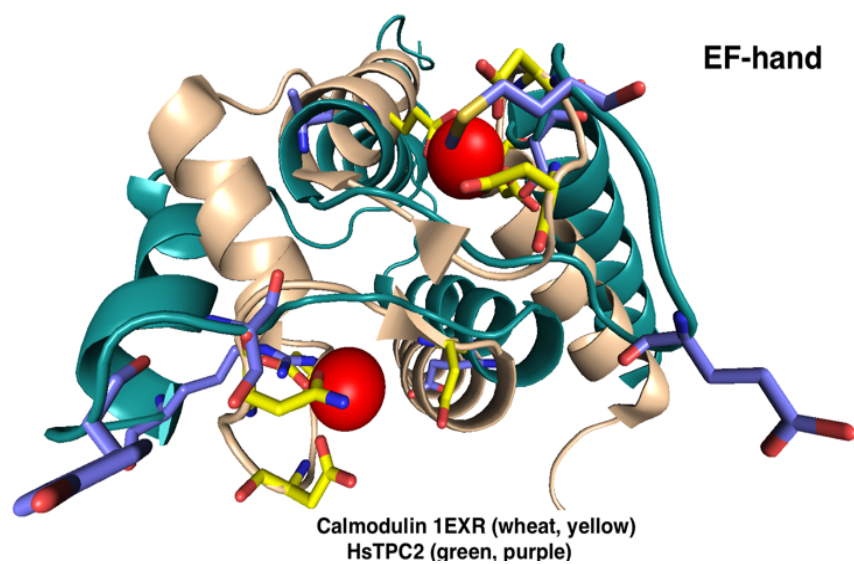


Figure S VII: Structural alignment of Calmodulin EF-hand (PDB; 1EXR) and EF-hand predicted by MODELLER for HsTPC2 model. The Calcium is shown as a sphere, and the Calmodulin as cartoon in wheat color and the calcium binding residues as yellow sticks

References

- [1] M. Rivera, "Two-pore channels (TPCs): Novel voltage-gated ion channels with pleiotropic functions," vol. 11, no. 1, pp. 20–33, 2017.
- [2] F. H. Yu and W. A. Catterall, "The VGL-Chanome: A Protein Superfamily Specialized for Electrical Signaling and Ionic Homeostasis," *Sci. Signal.*, vol. 2004, no. 253, p. re15-re15, 2004.
- [3] J. Guo *et al.*, "Structure of the voltage-gated two-pore channel TPC1 from *Arabidopsis thaliana*," *Nature*, 2015.
- [4] T. Rahman, X. Cai, G. C. Brailoiu, M. E. Abood, E. Brailoiu, and S. Patel, "Two-pore channels provide insight into the evolution of voltage-gated Ca²⁺ and Na⁺ channels," *Sci. Signal.*, vol. 7, no. 352, p. ra109-ra109, 2014.
- [5] R. Hedrich and I. Marten, "TPC1 - SV channels gain shape," *Mol. Plant*, vol. 4, no. 3, pp. 428–441, 2011.
- [6] M. X. Zhu, J. Ma, J. Parrington, P. J. Calcraft, A. Galione, and A. M. Evans, "Calcium signaling via two-pore channels: local or global, that is the question," *AJP Cell Physiol.*, vol. 298, no. 3, pp. C430–C441, 2010.
- [7] A. J. Morgan, F. M. Platt, E. Lloyd-Evans, and A. Galione, "Molecular mechanisms of endolysosomal Ca²⁺ signalling in health and disease," *Biochem. J.*, vol. 439, no. 3, pp. 349–378, 2011.
- [8] R. Hedrich *et al.*, "General Mechanisms for Solute Transport Across the Tonoplast of Plant Vacuoles: a Patch-Clamp Survey of Ion Channels and Proton Pumps," *Bot. Acta*, vol. 101, no. 1, pp. 7–13, 1988.
- [9] I. Ghai *et al.*, "General Method to Determine the Flux of Charged Molecules through Nanopores Applied to β -Lactamase Inhibitors and OmpF," *J. Phys. Chem. Lett.*, vol. 8, no. 6, pp. 1295–1301, 2017.
- [10] T. Furuichi, "A Putative Two Pore Channel AtTPC1 Mediates Ca²⁺ Flux in Arabidopsis Leaf Cells," *Plant Cell Physiol.*, vol. 42, no. 9, pp. 900–905, 2001.
- [11] E. Peiter *et al.*, "The vacuolar Ca²⁺-activated channel TPC1 regulates germination and stomatal movement," *Nature*, vol. 434, no. 7031, pp. 404–408, 2005.
- [12] P. J. Calcraft *et al.*, "NIH Public Access," *Nature*, vol. 459, no. 7246, pp. 596–600, 2009.
- [13] M. X. Zhu, J. Ma, J. Parrington, A. Galione, and A. Mark Evans, "TPCs: Endolysosomal channels for Ca²⁺ mobilization from acidic organelles triggered by NAADP," *FEBS Lett.*, vol. 584, no. 10, pp. 1966–1974, 2010.
- [14] F. H. Yu, V. Yarov-Yarovoy, G. A. Gutman, and W. A. Catterall, "Overview of molecular relationships in the voltage-gated ion channel superfamily.," *Pharmacol. Rev.*, vol. 57, no. 4, pp. 387–95, 2005.
- [15] K. Ishibashi, M. Suzuki, and M. Imai, "Molecular cloning of a novel form (two-repeat) protein related to voltage-gated sodium and calcium channels.," *Biochem. Biophys. Res. Commun.*, vol. 270, no. 2, pp. 370–376, 2000.
- [16] A. Marchler-Bauer *et al.*, "CDD: Specific functional annotation with the Conserved Domain Database," *Nucleic Acids Res.*, vol. 37, no. SUPPL. 1, pp. 205–210, 2009.
- [17] S. Patel, C. J. Penny, and T. Rahman, "Two-pore Channels Enter the Atomic Era:

- Structure of Plant TPC Revealed," *Trends Biochem. Sci.*, vol. 41, no. 6, pp. 475–477, 2016.
- [18] E. R. Liman, P. Hess, F. Weaver, and G. Koren, "Voltage-sensing residues in the S4 region of a mammalian K⁺ channel," *Nature*, vol. 353, no. 6346, pp. 752–6, 1991.
 - [19] E. Brailoiu *et al.*, "An NAADP-gated two-pore channel targeted to the plasma membrane uncouples triggering from amplifying Ca²⁺ signals," *J. Biol. Chem.*, vol. 285, no. 49, pp. 38511–38516, 2010.
 - [20] S. J. Pitt *et al.*, "TPC2 is a novel NAADP-sensitive Ca²⁺ release channel, operating as a dual sensor of luminal pH and Ca²⁺," *J. Biol. Chem.*, vol. 285, no. 45, pp. 35039–35046, 2010.
 - [21] E. Brailoiu *et al.*, "Essential requirement for two-pore channel 1 in NAADP-mediated calcium signaling," *J. Cell Biol.*, vol. 186, no. 2, pp. 201–209, 2009.
 - [22] M. Schieder, K. Rötzer, A. Brüggemann, M. Biel, and C. A. Wahl-Schott, "Characterization of two-pore channel 2 (TPCN2)-mediated Ca²⁺ currents in isolated lysosomes," *J. Biol. Chem.*, vol. 285, no. 28, pp. 21219–21222, 2010.
 - [23] C. Grimm, C. Chen, C. Wahl-schott, and M. Biel, "Two-Pore Channels : Catalyzers of Endolysosomal Transport and Function," vol. 8, no. February, pp. 6–11, 2017.
 - [24] J. Parrington, P. Lear, and A. Hachem, "Calcium signals regulated by NAADP and two-pore channels - their role in development , differentiation and cancer," vol. 355, no. July, pp. 341–355, 2015.
 - [25] A. J. Morgan and A. Galione, "Two-pore channels (TPCs): Current controversies," *BioEssays*, vol. 36, no. 2, pp. 173–183, 2014.
 - [26] J. Guo, W. Zeng, and Y. Jiang, "Tuning the ion selectivity of two-pore channels," 2016.
 - [27] Z. Liu, H. Chen, J. Zhou, and F. Laezza, "Tuning the Ion Selectivity of Two-Pore Channels," *Biophysj*, vol. 112, no. 3, p. 242a, 2017.
 - [28] X. Wang *et al.*, "TPC Proteins Are Phosphoinositide- Activated Sodium-Selective Ion Channels in Endosomes and Lysosomes," *Cell*, vol. 151, no. 2, pp. 372–383, 2012.
 - [29] M. Ruas *et al.*, "Expression of Ca²⁺-permeable two-pore channels rescues NAADP signalling in TPC-deficient cells," vol. 34, no. 13, pp. 1743–1758, 2015.
 - [30] P. Sulem *et al.*, "Two newly identified genetic determinants of pigmentation in Europeans," *Nat. Genet.*, vol. 40, no. 7, pp. 835–837, 2008.
 - [31] S. Feijóo-Bandín *et al.*, "Two-pore channels (TPCs): Novel voltage-gated ion channels with pleiotropic functions," *Channels*, vol. 11, no. 1, pp. 20–33, 2017.
 - [32] L. N. Hockey *et al.*, "Dysregulation of lysosomal morphology by pathogenic LRRK2 is corrected by two-pore channel 2 inhibition," no. November, 2014.
 - [33] Y. Sakurai *et al.*, "Two-pore channels control Ebola virus host cell entry and are drug targets for disease treatment," *Science (80-.)*, vol. 347, no. 6225, 2015.
 - [34] Y. Sakurai *et al.*, "drug targets for disease treatment," vol. 347, no. 6225, pp. 995–998, 2016.
 - [35] C. Grimm *et al.*, "High susceptibility to fatty liver disease in two-pore channel 2-deficient mice," *Nat. Commun.*, vol. 5, 2014.
 - [36] C. Grimm, E. Butz, C. C. Chen, C. Wahl-Schott, and M. Biel, "From mucopolidosis type IV to Ebola: TRPML and two-pore channels at the crossroads of endo-lysosomal trafficking and disease," *Cell Calcium*, vol. 67, pp. 148–155, 2017.
 - [37] M. H. Rashid *et al.*, "A potent and selective peptide blocker of the Kv1.3 channel: Prediction from free-energy simulations and experimental confirmation," *PLoS One*,

- vol. 8, no. 11, 2013.
- [38] A. A. Kochegarov, "Pharmacological modulators of voltage-gated calcium channels and their therapeutical application," *Cell Calcium*, vol. 33, no. 3. pp. 145–162, 2003.
 - [39] O. N. P. Nguyen *et al.*, "Two-pore channel function is crucial for the migration of invasive cancer cells," *Cancer Res.*, vol. 77, no. 6, pp. 1427–1438, 2017.
 - [40] I. Pafumi *et al.*, "Naringenin Impairs Two-Pore Channel 2 Activity And Inhibits VEGF-Induced Angiogenesis," no. May, pp. 1–11, 2017.
 - [41] A. F. Kintzer and R. M. Stroud, "On the structure and mechanism of two-pore channels," *FEBS J.*, 2017.
 - [42] S. Tayefeh, "Computational study of the Kcv potassium channel," pp. 1–87, 2007.
 - [43] C. E. Capener, H. J. Kim, Y. Arinaminpathy, and M. S. P. Sansom, "Ion channels: structural bioinformatics and modelling," *Hum. Mol. Genet.*, vol. 11, no. 20, pp. 2425–2433, 2002.
 - [44] C. J. Penny, T. Rahman, A. Sula, A. J. Miles, and B. A. Wallace, "Isolated pores dissected from human two-pore channel 2 are functional," *Nat. Publ. Gr.*, no. November, pp. 1–11, 2016.
 - [45] R. Laskowski, J. A. Rullmann, M. MacArthur, R. Kaptein, and J. Thornton, "AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR," *J. Biomol. NMR*, vol. 8, no. 4, pp. 477–486, 1996.
 - [46] N. Eswar *et al.*, *Comparative protein structure modeling using Modeller*, vol. Chapter 5. 2006.
 - [47] A. Fiser and A. Šali, "MODELLER: Generation and Refinement of Homology-Based Protein Structure Models," *Methods Enzymol.*, vol. 374, pp. 461–491, 2003.
 - [48] P. Schmidtke and X. Barril, "Understanding and predicting druggability. A high-throughput method for detection of drug binding sites," *J. Med. Chem.*, vol. 53, no. 15, pp. 5858–5867, 2010.
 - [49] E. Perola, L. Herman, and J. Weiss, "Development of a rule-based method for the assessment of protein druggability," *J. Chem. Inf. Model.*, vol. 52, no. 4, pp. 1027–1038, 2012.
 - [50] F. N. B. Edfeldt, R. H. A. Folmer, and A. L. Breeze, "Fragment screening to predict druggability (ligandability) and lead discovery success," *Drug Discov. Today*, vol. 16, no. 7–8, pp. 284–287, 2011.
 - [51] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Adv. Drug Deliv. Rev.*, vol. 64, no. SUPPL., pp. 4–17, 1997.
 - [52] A. Borrel, L. Regad, H. Xhaard, M. Petitjean, and A. C. Camproux, "PockDrug: A model for predicting pocket druggability that overcomes pocket estimation uncertainties," *J. Chem. Inf. Model.*, vol. 55, no. 4, pp. 882–895, 2015.
 - [53] H. A. Hussein, A. Borrel, C. Geneix, M. Petitjean, L. Regad, and A. C. Camproux, "PockDrug-Server: A new web server for predicting pocket druggability on holo and apo proteins," *Nucleic Acids Res.*, vol. 43, no. W1, pp. W436–W442, 2015.
 - [54] D. J. Posson, R. Rusinova, O. S. Andersen, and C. M. Nimigean, "Calcium ions open a selectivity filter gate during activation of the MthK potassium channel," *Nat. Commun.*, vol. 6, no. May, 2015.
 - [55] E. Kutluay, B. Roux, and L. Heginbotham, "Rapid Intracellular TEA Block of the KcsA Potassium Channel," *Biophys. J.*, vol. 88, no. 2, pp. 1018–1029, 2005.

- [56] Z. Chen, B. H. Ong, N. G. Kambouris, E. Marbán, G. F. Tomaselli, and J. R. Balser, "Lidocaine induces a slow inactivated state in rat skeletal muscle sodium channels.," *J. Physiol.*, vol. 524 Pt 1, pp. 37–49, 2000.
- [57] M. D. Cahalan and W. Almers, "Interactions between quaternary lidocaine, the sodium channel gates, and tetrodotoxin," *Biophys. J.*, vol. 27, no. 1, pp. 39–55, 1979.
- [58] A. F. Kintzer and R. M. Stroud, "Structure, inhibition, and regulatory sites of TPC1 from *Arabidopsis thaliana*," vol. 531, no. 5, pp. 258–262, 2016.
- [59] R. Farid, T. Day, R. A. Friesner, and R. A. Pearlstein, "New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies," *Bioorganic Med. Chem.*, vol. 14, no. 9, pp. 3160–3173, 2006.
- [60] P. J. Stansfeld, P. Gedeck, M. Gosling, B. Cox, J. S. Mitcheson, and M. J. Sutcliffe, "Drug Block of the hERG Potassium Channel: Insight From Modeling," *Proteins*, vol. 68, no. 2, pp. 568–580, 2007.
- [61] M. Cavalli, A. Cavalli, and M. Recanatini, "Modeling the hERG Potassium Channel in a Phospholipid Bilayer: Molecular Dynamics and Drug docking Studies," *Wiley Intersci.*, vol. 29, no. 5, pp. 795–808, 2007.
- [62] U. Zachariae, F. Giordanetto, and A. G. Leach, "Side chain flexibilities in the human ether-a-go-go related gene potassium channel (hERG) together with matched-pair binding studies suggest a new binding mode for channel blockers," *J. Med. Chem.*, vol. 52, no. 14, pp. 4266–4276, 2009.
- [63] G. Di Paolo and P. De Camilli, "Phosphoinositides in cell regulation and membrane dynamics," *Nature*, vol. 443, no. 7112, pp. 651–657, 2006.
- [64] S. K. Dove, K. Dong, T. Kobayashi, F. K. Williams, and R. H. Michell, "Phosphatidylinositol 3,5-bisphosphate and Fab1p/PIKfyve under PIP₂ in endo-lysosome function," *Biochem. J.*, vol. 419, no. 1, pp. 1–13, 2009.

Chapter 4

Material & Methods

Molecular modelling

The studies of this thesis are all based on molecular modelling approaches. They concern the *microscopic* description of systems (governed by molecular and atomic interactions) to understand their *macroscopic* behavior, through the theory of statistical mechanics. This approach allows to explore the macroscopic properties of a system through microscopic simulations, for example, to calculate changes in the binding free energy of a particular drug candidate, or to examine the energetics and mechanisms of conformational changes.

Molecular modelling methods include[1]: quantum methods, molecular mechanics and molecular dynamics; i) The quantum mechanics methods are based on the calculation of molecular orbitals and for which the complexity increases rapidly with the number of atoms (and electrons); ii) molecular mechanics is a technique that calculates the energy of atoms without considering explicitly the electrons, it is thus based on semi-classical (or semi-empirical) approximations. The resulting considerable simplification of the calculations makes it possible to work on large molecules, such as biological macromolecules (proteins, nucleic acids), or on systems comprising a large number of molecules. iii) The molecular dynamics technique aims to calculate the motions of molecules, most often from the energies of molecular mechanics, by applying the laws of classical mechanics.

Molecular dynamics simulations

Molecular dynamics (MD) [2] [1] is a technique for the investigation of molecular systems at atomic detail. It consists in generating a time-dependent trajectory for a finite set of particles, by means of the numerical integration of classical equations of motion. Newton's equation of motion are solved under a model potential $V(\mathbf{r})$ [2] [3]. Thus, the equation of motion for the i molecule is written:

$$\mathbf{F}_i(t) = m_i \mathbf{a}_i(t) = m_i \ddot{\mathbf{r}}_i(t) = -\nabla V_i(\mathbf{r}) = -\frac{\partial V_i(\mathbf{r})}{\partial \mathbf{r}_i} \quad (1)$$

Where:

i : is the particle index,

$\mathbf{r}$: the position,

m : the mass,

$\mathbf{F}$: the force,

V: the potential energy
a: is the acceleration.

Here a molecule is considered as an assembly of atoms whose spatial positions vary over time. Each atom is considered as a point mass whose movement is determined by all the forces acting on it. This results in an atomic motion corresponding to oscillations around a minimum of energy or the transition from one of these minima to another. Molecular dynamics thus has the ability to extract the molecule from a local minimum. In such methods two main initial parameters are crucial to start calculations, because these equations of motion are deterministic:

- i) the determination of an initial configuration of the system of interest, the initial configuration corresponds to a set of spatial coordinates (experimental or model)
- ii) determine the initial velocities

The initial positions can be obtained from experimental structures, such as the x-ray crystal structure of the protein or the solution structure determined by NMR spectroscopy. For the coordinates of proteins for which the experimental data are unavailable, we can estimate coordinates by using computational tools based on available experimental data, the so-called homology modelling of protein. This approach has been used in this thesis to build a 3D-model for the human TPC channel and the details about the method will be discussed later in this chapter.

The initial velocities can be directly obtained from the probabilistic Maxwell-Boltzmann distribution at a given temperature, which gives the probability $\rho(v)$ that an atom i has a velocity v_x in the x direction at a temperature T (eq. 2)[4] [5]:

$$\rho(v) = 4\pi \left(\frac{m}{2\pi K_B T} \right)^{3/2} v^2 e^{-mv^2/2K_B T} \quad (2)$$

From the initial coordinates and velocities, the Molecular Dynamics (MD) procedure consists in generating a long series of configurations responding to well-defined external constraints such as pressure, temperature, total energy, density, volume, etc. The configurations are obtained by solving numerically the classical equations (1) of the motion of the N molecules of the system over a time interval. The most widely used method of integrating Newton's equations of motion is the discretization proposed by the Verlet [6] algorithm. With this algorithm, the

positions at $t+\delta t$ are calculated from the knowledge of the positions at $t-\delta t$ and the positions and accelerations at time t , calculated from the following equations (3)[6]:

$$\mathbf{r}_i(t + \delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \delta t) + \delta t^2 \mathbf{a}_i(t) \quad (3)$$

Where δt is the so-called timestep of the simulation. The integration step must be chosen small compared to the period of the highest frequency vibrational movement (fastest movement), in this case that of the C-H bond, which is 10 femtoseconds. We choose a step approximately 10 times less than this period, the molecular dynamics simulations are therefore generally performed with an integration step of 1 fs. This step can be increased to 2 fs or larger when using the SHAKE algorithm during simulations whose function is to freeze the X-H bonds, where X is any heavy atom.[7] [8]

Force-fields contribution

All the physical parameters characterizing the system are described in what is called a force field or potential functions (Potential V in the eq.1). The total potential energy is decomposed into a sum of additive terms each representing bonding interactions and non-bonding interactions terms. Bonding interactions depend only on the internal coordinates of the molecules, i.e., bonds, valence angles, and torsions. The non-binding potential is expressed in two terms: Van der Walls term and a term for electrostatic interactions. Electrostatic interactions have the longest range in this force field model[9]. The calculation of interaction terms that are long-range is computationally expensive. A common way to handle long-range potentials is the Ewald sum [10] [11]. To save computational time, a grid-based technique called "Particle Mesh Ewald Method" [10] [11] was used in this work to calculate the electrostatic interaction during the MD simulations.

The terms of the force field are decomposed as follows:

$$\begin{array}{cc} \boxed{\text{Bonded Energy}} & \boxed{\text{Non-Bonded Energy}} & (4) \\ \hline E = E_{\text{stretching}} + E_{\text{bending}} + E_{\text{torsion}} + E_{\text{vdW}} + E_{\text{elec}} + E_{\text{hydrogen}} \end{array}$$

The force fields parameters are set up from experimental results and quantum mechanical calculations of small model compounds, in such a way that all the different contributions allow

to reproduce a series of experimental results. A force field must therefore always be considered as an indivisible entity and it is excluded to combine the terms and/or the parameters of two force fields in the hope of obtaining a better set. They must be chosen to obtain agreement with experimental data. Several force fields were developed and well characterized rather for biological systems. In this work, we used two popular force fields: AMBER (AMBER14 for protein and lipid, GAFF for substrates) and CHARMM36 force fields.[12] [13] [14] [15] [16]. The interactions potential terms included in Eq. 4 are illustrated in figure 42.

AMBER

The standard AMBER (Assisted Model Building with Energy Refinement) is a family of force fields [17], initially written by Kollman [14]. The field is parametrized to reproduce proteins and nucleic acids interactions[17] starting from ab initio data from small fragments. It has been used for polymers and other small molecules. As part of this work, we used the force field ff99 for the systems of OccAB porins. GAFF (General AMBER force field) was used for simulating small molecules (substrates and antibiotics)[14]

The expression of the AMBER force field is of the following form (equation 5):

$$\begin{aligned}
 \mathcal{V}(\mathbf{r}) = & \sum_{\text{bond stretch}} k_r (r - r_0)^2 + \sum_{\text{bond torsion}} k_\theta (\theta - \theta_0)^2 \\
 & + \sum_{\text{dihedral}} \sum_n \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\
 & + \frac{1}{k^{1-4}_{\text{vdW}}} \sum_{\substack{i < j \\ \{i,j\} \in 1-4}} \varepsilon_{ij} \left[\left(\frac{R_{ij}^*}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^*}{r_{ij}} \right)^6 \right] + \frac{1}{k^{1-4}_{\text{Coulomb}}} \sum_{\substack{i < j \\ \{i,j\} \in 1-4}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \\
 & + \sum_{\substack{i < j \\ \{i,j\} > 1-4}} \varepsilon_{ij} \left[\left(\frac{R_{ij}^*}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^*}{r_{ij}} \right)^6 \right] + \sum_{\substack{i < j \\ \{i,j\} > 1-4}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \\
 & \quad \quad \quad \text{Hydrogen-bond term}
 \end{aligned} \tag{5}$$

The terms 1 to 3 in equation (5) handle the bonded interactions of bonds, angles, and dihedrals. The terms 4 and 5 account for the non-bonded interactions including van der Waals and electrostatic interactions. The final term, 6, is an optional hydrogen-bond term that augments the electrostatic description of the hydrogen bond. This term adds only about 0.5 kcal mol⁻¹ to the hydrogen-bond energy in AMBER [12].

CHARMM

CHARMM (Chemistry at Harvard Macromolecular Mechanics), developed initially by Karplus et al[1][15] for the calculation of biomolecules, is similar to AMBER. It was parameterized on the basis of energies and geometries of small organic models. CHARMM uses a flexible and comprehensive empirical energy function that is a summation of many individual energy terms. The energy function is based on separable internal coordinate terms and pairwise non-bonding interaction terms [18](Brooks et al. 1983). The total energy is expressed by the following equation (6):

$$U_{CHARMM} = U_{bonded} + U_{non-bonded} \quad (6)$$

where U_{bonded} consists of the following terms,

$$U_{bonded} = U_{bond} + U_{angle} + U_{UB} + U_{dihedral} + U_{improper} + U_{CMAP} \quad (7)$$

$$\begin{aligned} U_{bond} &= \sum_{bonds} K_b(b - b^0)^2, \\ U_{angle} &= \sum_{angles} K_\theta(\theta - \theta^0)^2, \\ U_{UB} &= \sum_{Urey-Bradley} K_{UB}(b^{1-3} - b^{1-3,0})^2, \\ U_{dihedral} &= \sum_{dihedrals} K_\varphi((1 + \cos(n\varphi - \delta))), \\ U_{improper} &= \sum_{improvers} K_\omega(\omega - \omega^0)^2, \text{ and} \\ U_{CMAP} &= \sum_{residues} u_{CMAP}(\Phi, \Psi) \end{aligned} \quad (8)$$

$U_{non-bonded}$ consists of two terms,

$$U_{non-bonded} = U_{LJ} + U_{elec}$$

$$U_{LJ} = \sum_{nonb.pairs} \epsilon_{ij} \left[\left(\frac{r_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^{min}}{r_{ij}} \right)^6 \right] \quad (9)$$

$$U_{elec} = \sum_{nonb.pairs} \frac{q_i q_j}{\epsilon r_{ij}}$$

The CMAP term is used in CHARMM force field to treat conformational properties of protein backbones. The so-call CMAP term is a cross-term for the (backbone dihedral angle) values, realized by grid based energy correction maps. It is used in the current CHARMM force field to improve the conformational properties of protein backbones [16].

Overall, the same empirical potential (roughly) is being used in both force fields (AMBER and CHARMM), however the parameters were generated in a different manner (i.e. different charge

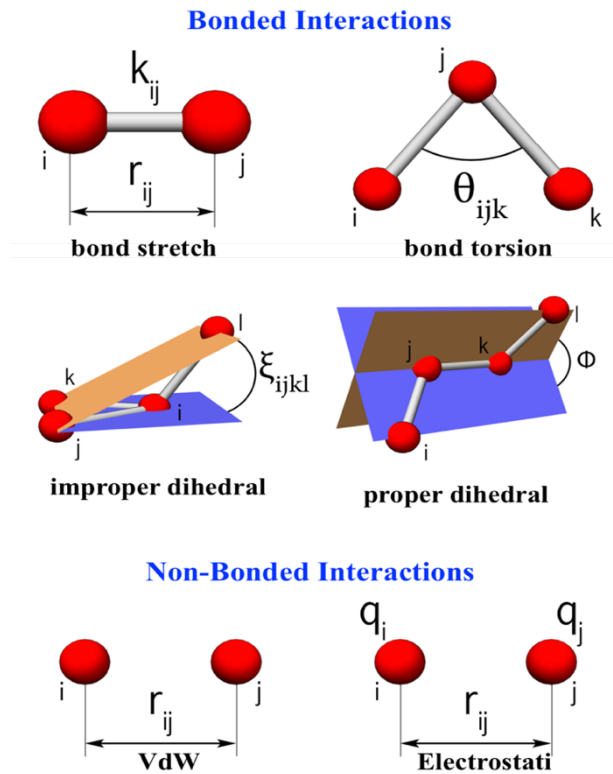


Figure 42: Types of interactions expressed in the potential energy in both CHARMM and AMBER force-fields. (A) The interactions between bonded atoms: bond stretch, bond torsion, proper dihedral and improper dihedral. (B) non-bonded interaction; vdw interaction, electrostatics interactions. Picture adapted from (<http://cbio.bmt.tue.nl/pumma/index.php>)

model, different vdw, different torsions, different geometries, etc...). As part of this study, we used this force field for the study of the Human TPC channels.

Ensemble thermodynamics

The integration of Newton's equations of motion allows us to explore the free energy landscape of a system. In order to mimic a more realistic behavior and to obtain results comparable with the experimental data available, it is necessary to consider a control of the thermodynamic variables of this system such as volume (V), temperature (T) and/or pressure (P). Several different methods that use different approaches to control T and P during the MD simulations have been proposed (ref). The application of these algorithms allows the simulated system to sample different and more realistic thermodynamic ensembles, such as the canonical ensemble (NVT) and the isobaric-isothermal ensemble (NPT). In the case of proteins inserted in lipid bilayers[19], the most commonly used ensemble for modeling their behavior by molecular dynamics is (NVT). Indeed, it allows the conservation of the experimental conditions characteristic of the simulated phase, namely the volume and temperature of the system. [7] [8] [19]

Periodic boundary conditions

The molecular dynamics simulations of a system in solution are carried out in a solvent box with dimension R_{box} (figure 43) and different shapes (parallelepiped, octahedral, etc.). In order to avoid or reduce the edge effects in the calculation of interactions between non-bonded atoms and solute-solvent calculations, the simulation box is duplicated in all directions so that the fluid is seen as an infinite periodic reproduction of cells, each of which contains N molecules. The simulation is applied to a single region called simulation cell. During a simulation, when a molecule (black sphere in figure 43) moves in the original cell, its periodic image in each of the neighboring cells moves in exactly the same way in the same direction and with the same speed. In other words, everything happens as if the same simulation takes place in all the boxes (cells). Thus, the number of molecules in the original cell as well as throughout the system is preserved. To do so, a distance limit is applied namely R_{cut} which is the cutoff radius that is normally applied when calculating the force between two atoms, its value must be chosen so that an atom can interact with only one image of any given atom. This means that R_{cut} cannot be greater than half the R_{box} . On the other hand, the interactions with particles in the central cell as well as all periodic images must be taken into account. A variety of methods have been developed to handle long-range forces for periodic cell, the widely used one is the Ewald sum technique [9] [10] and its numerical implementation for large systems, *Particle Mesh Ewald* [10] [11].

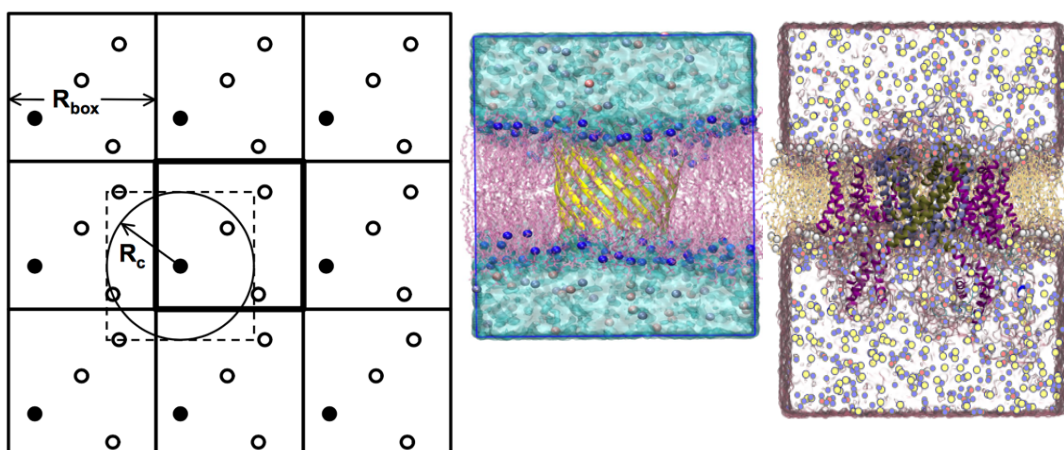


Figure 43: Periodic condition representation and the common geometrical box used in this work.

Main Steps of classical MD simulations runs and details used in our systems

1. Initial input
2. Energy minimization
3. Heating to 300K
4. Equilibration (NPT)
5. Production MD simulation (NVT)

System preparation and Simulations details.

Each protein was inserted in a pre-equilibrated membrane aligned with the pore along the z-axis, with its center at $x=y=z=0$ Å, removing those lipid molecules that overlapped with the protein coordinates. This system was minimized with 500 steps of conjugate gradient to relax locally atomic coordinates after hydrogen addition. Then, a short NVT MD run of 1ns at 10 K was performed followed by several NPT equilibrations, decreasing restraints on the C-alpha protein and head groups of POPC (*Palmitoyl Oleoyl Phosphatidyl Choline*) lipid model provided by *CHARMMGUI* program[20]. During these steps, the temperature was increased from 100K to 300K using the Langevin thermostat; the pressure was controlled by the Berendsen barostat at 1 atm. With this relaxation, we equilibrated the lipid molecules around the external surface of the protein. Afterward, the system was neutralized and solvated with the addition of counterions and TIP3P[21] water molecules.

OccAB1 systems:

All our systems were set up starting from crystal structures of OccAB. The missing atoms on the loop L9 of OccAB2 porin were built using *MODELLER* program [22]. The systems were hydrated with ~16000 TIP3P water molecules and ionized with KCl ions molecules at a concentration of 0.2 M, adding K⁺ counter-ion for charge neutrality. The final system comprises 86500 atoms in a cell of about 98×98×88 Å³. Starting again from 100K we used the same steps to relax both temperature and density, followed by 20 ns in NPT ensemble at 300K and 1 atm. Finally, the restraints were removed and each system was subject to a 400 ns NVT production run. All simulations were carried out with ACEMD software version 3212[23] and plumed 1.3[24] and using the Lipid14 force field[13] for the membrane and ff14SB force field[13] [12] for the protein and GAFF (General Amber Force Field)[14] for the substrates preparation. The final equilibrated box size was of 97.6×96.7×85.1 Å³.

PCA analysis. In the classical MD simulations, collective motions (slow motion) and local fluctuations (fast motion) occur simultaneously making the identification of the relevant movement difficult. In this case, we use principal components analysis method (PCA) [25] that can represent the principal motion by a set of eigenvectors that allows the capture of the global movements. This is done by building a so-called covariance matrix of the protein atomic fluctuations. Diagonalization of this matrix gives a set of eigenvectors and relative eigenvalues, which describe collective modes of fluctuations of the protein. The eigenvectors are sorted with descending eigenvalues, the largest eigenvalues are the principal components as they represent the largest-amplitude collective motions. To see the motion corresponding to each of eigenvectors, we filtered the original trajectory and projected out a selected eigenvector, then we extracted structures projecting on the extreme (maximum and minimum value) of the selected eigenvectors. Finally, for each extracted structure we calculated the pore radius. These analyses were performed using GROMACS software [26].

HsTPC2 channel model system:

This system was prepared as the other OccAB systems. After relaxation of the system in the box of water, we added 300mM NaCl and 200mM CaCl salt concentration, adding counter-ion for imposing charge neutrality. Then the system was subjected to a ~1000 ns NVT production run. We used CHARMM36 force field.

Methods for free energy calculations

In the following, a short overview of the three techniques applied in our calculations to bias the system towards the transitions of interest are presented with the purpose of pointing out the theoretical assumptions and the computational strategies that can possibly produce qualitative

and quantitative effects on the obtained results. Details are also provided about the values of the parameters used in our simulations for the small molecules translocation through OccAB channels.

Metadynamics

Many physical, chemical and biological processes occur at timescales that are larger than those available through molecular dynamics simulation. Events such as the translocation of a substrate through an entire protein channel happen at long time scale, they cannot be described with standard MD simulations. Calculation of free energy differences by means of atomistic simulations is the central subject of this thesis. Indeed, the most part of this work focuses on the process of translocation of small molecules through the membrane of gram-negative bacterium (*A. baumannii*) via protein channels, the so-called porins (OccAB porins). This type of biological process is classified as a rare event for simulations. In the same category fall the ions diffusion or the folding of proteins. These processes occur so rarely so that is impossible to quantify them by conventional trajectory calculations (CMD simulations).

In general, a rare event can occur when a high free-energy barrier[27], much larger than $k_B T$, separates two metastable states (Figure 44), this would require large computational time, not accessible with common computer architectures. One way to calculate the difference in free energy between two states is simply to direct the system from one state to another by means of enhanced sampling methods. Several methods have been developed such as umbrella sampling and adaptive reaction coordinate forces[28]. In our case, we focus on the metadynamics method [27] used to explore the diffusion of small molecules through bacterial outer membrane porins.

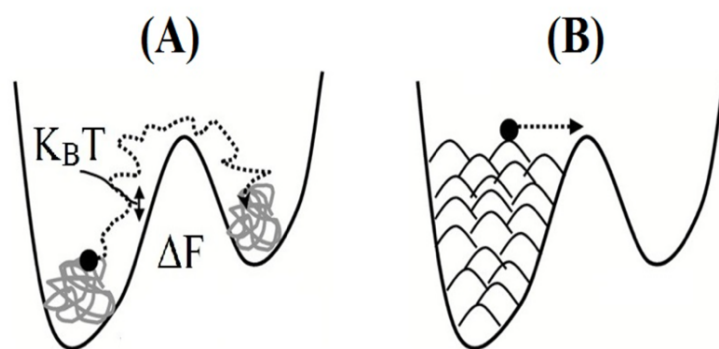


Figure 44: Schematic scheme of the biased potential *used* in Metadynamics method. A system may require free energy jumps much greater than the average fluctuations in energy at a given temperature (A); In Metadynamics an history-dependant potential is added discouraging the system from visiting already explored conformations (B). Reprinted from²⁹

A metadynamics [27] [29] simulations consists of a standard MD run in which history dependent biases are added as external forces on a set of a priori defined collective variables. It allows exploring the conformational space of collective variables (CV) faster and characterizes rare events by reducing the energy barriers. Looking at energies, the bias terms, during the simulation, fill the minima of the surface energy landscape by a series of consecutives Gaussians deposited in the CVs space during the trajectory. Once the energy landscape is completely flattened by the bias terms, one can finally reconstruct the Free Energy Surface (FES) as a function of the chosen collective coordinates. However, the risk of adding the bias is to overfill the region of the FES and consequently increases the error on the calculation of free energy differences proportionally to the added potential energy. The total bias that has been added in the simulation at a given time is:

$$V(\vec{s}, t) = \sum_{k\tau < t} W(k\tau) \exp\left(-\sum_{i=1}^d \frac{(s_i - s_i(q(k\tau)))^2}{2\sigma_i^2}\right) \quad (10)$$

Where;

$\vec{s}(q)$: corresponds to the selected collective variables (CVs)

τ : is the Gaussian deposition stride,

σ_i : the width of the Gaussian for the i th CV

$W(k\tau)$: the height of the Gaussian.

The starting point of the method is an appropriate choice of the set of collective variables (CV) able to completely characterize the process at hand, to separate the initial from the final state, and also to take into account all the slow modes, identified by slower time scales. Typical examples of collective coordinates are distances, angles, coordination numbers or, in general, any function that can be expressed by means of the original physical atomic coordinates x , y , z accessible with MD simulations. As collective variables for our studies, we first considered the axis of diffusion, namely the z -axis, as the direction perpendicular to the plane of the membrane. We selected three collective variables (CVs) and we implemented them in the MD code using the PLUMED plugin[30]:

1. CV1 as the projection of the distance between the center of masses of protein and substrate along the axis of diffusion, so-labelled z -axis.

2. CV2 as the orientation of the substrate, with respect to z-axis indicated by the distance between the positive (NH₂⁺) and negative (CO₂⁻) groups of the substrate. A positive value corresponds to the substrate pointing toward the periplasmic space with its positive group.

3. Dipole moment of the antibiotics

As described in PLUMED[24], the Gaussians of constant height in standard metadynamics simulations, are added for the entire course of a simulation. As a result, the system is eventually pushed to explore high free-energy regions and the estimate of the free energy calculated from the bias potential oscillates around the real value which increases the errors. In *Well-Tempered Metadynamics* (WTMeta)[31], the height of the Gaussian is decreased with simulation time according to:

where W_0 is an initial Gaussian height, ΔT an input parameter with the dimension of a

$$W(k\tau) = W_0 \exp\left(-\frac{V(s(\vec{q}(k\tau)), k\tau)}{k_B \Delta T}\right), \quad (11)$$

temperature, and k_B the Boltzmann constant. With this rescaling of the Gaussian height, the bias potential smoothly converges in the long-time limit, but it does not fully compensate the underlying free energy:

$$V(s, \vec{t} \rightarrow \infty) = -\frac{\Delta T}{T + \Delta T} F(s) + C. \quad (12)$$

where T is the temperature of the system. In the long-time limit, the CVs thus sample an ensemble at a temperature $T + \Delta T$ which is higher than the system temperature T . The parameter ΔT can be chosen to regulate the extent of free-energy exploration: $\Delta T = 0$ corresponds to standard molecular dynamics, $\Delta T \rightarrow \infty$ to standard metadynamics. In well-tempered metadynamics literature and in PLUMED[24], you will often encounter the term "biasfactor" which is the ratio between the temperature of the CVs ($T + \Delta T$) and the system temperature (T)[31]:

$$\gamma = \frac{T + \Delta T}{T}. \quad (13)$$

Thus, in WTMmeta [32] [31], the overfilling and fluctuation problems of standard Metadynamics are avoided and the exploration of the FES can be tuned varying the $\gamma/\Delta T$ parameter. In our studies, for the small molecule translocation through OccAB porins, we used Well-tempered metadynamics simulations[32] as an enhanced sampling technique in order to accelerate the sampling of the substrates translocation through OccAB channel. The parameters set for the Well-tempered metadynamic simulations in our systems are:

- 1) The initial width of the deposited Gaussian potential; 1.2 Kcal/mol
- 2) The energy deposition rate at for the Gaussian potential;
- 3) The width of Gaussian at mid eight, 0.25 Angstrom for CV1 and 0.05 Angstrom for CV2
- 4) The ΔT value or so-called bias factor that determines the rescaling factor for the Gaussian height along the simulation; 3000 K.

Moreover, we used Multiple Walker metadynamics to parallelise our calculations: multiple simulations of the same system are run in parallel using metadynamics on the same set of selected collective variables. The deposited bias is shared among the replicas in such a way that the history dependent potential depends on the history. In total 4 walkers were used biasing the same CVs.

Homology modeling and molecular docking

For several systems in particular membrane proteins, the elucidation of a 3D structure by the previous experimental techniques may often be delayed by difficulties in obtaining sufficient amount of material (cloning, expression and purification of milligram quantities of the protein) and difficulties associated with crystallization. In this case, methods dealing with the prediction of protein structure are often used[33]. One of the most important theoretical techniques used for the determination of the initial conformation of the protein, involves the conformational modelling of proteins by homology of primary sequences[34]. This method is based on the

observation that two proteins belonging to the same family (a sharing similar amino acid sequences), will have similar three-dimensional structures[35]. In fact, it has been shown that the degree of conservation of protein three-dimensional structure within a family is much higher than conservation of the sequence[22]. This method is used in this work to model the initial conformation of human TPC channels for which the experimental structure is not yet available, the details about the method will be described in the next paragraph. The program package MODELLER [36] [35] offers all these steps. It was used in this work for the creation of all initial homology models of HsTPC2.

Steps in molecular modelling

The concept and the main steps in homology modeling are illustrated in Figure. 45

1. Identification of structures that will form the template for the target structure (model).
2. Alignment – the most important step. Alignment of low homology sequences can be improved using secondary structure prediction (align-model-realign-remodel).
3. Transfer of coordinates from the template(s) to the target of structurally conserved regions.
4. Modelling variable regions or missing residues; loops: by using MODELLER and homology modelling
5. Modelling side chains (practically a virtual step)
6. Minimisation:
7. Checking the correctness of the model.
 - Stereochemical properties: **PROCHECK**[37] [38] [37]
 - Bond angles
 - Bond Length
8. Molecular Dynamics: To study regional flexibility.

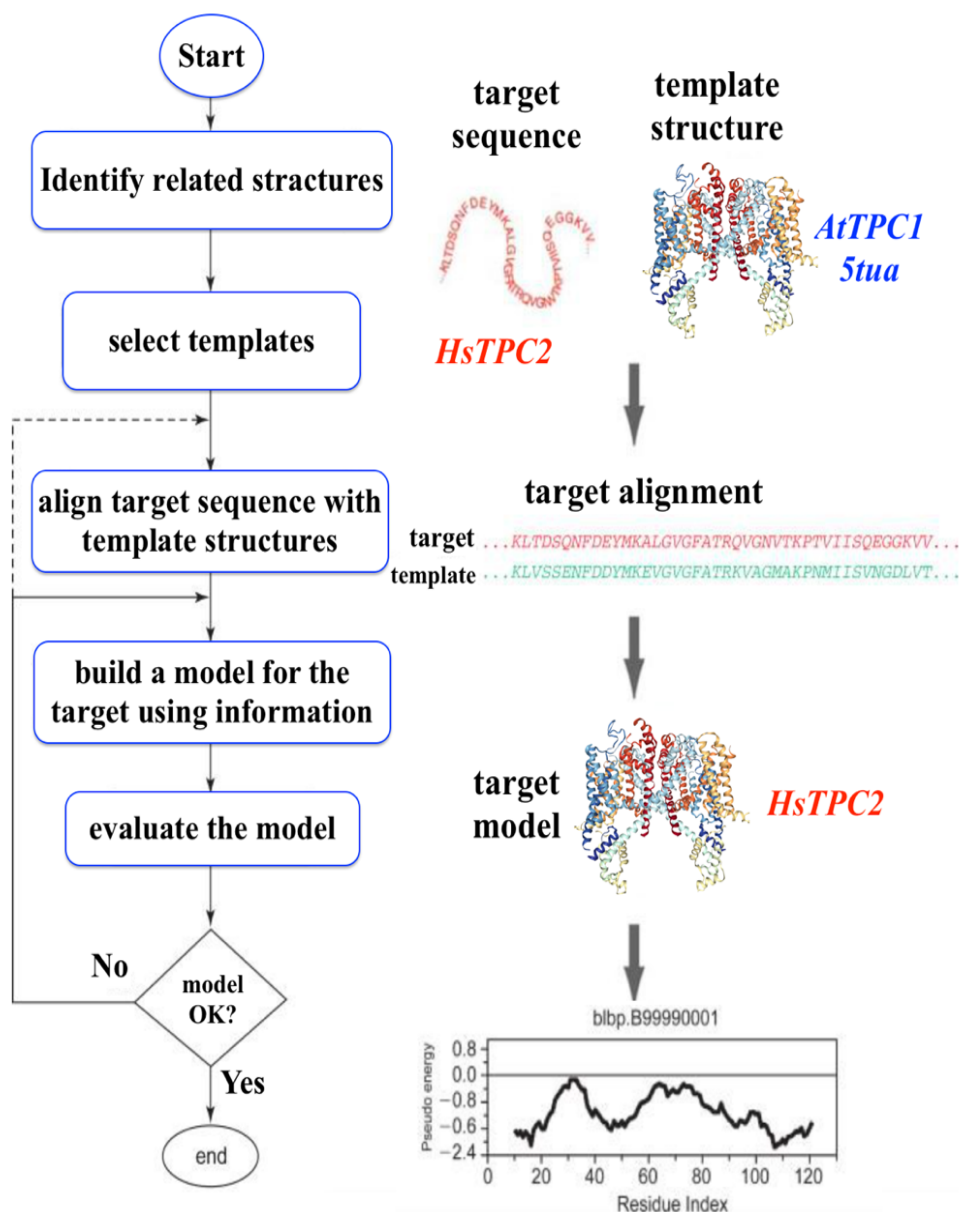


Figure 45: flow illustrates the major steps of obtaining structure from sequence.

Model validation[39]

The homology model of *Homo sapiens* TPC2 channel was generated from the x-ray crystal structure of *A. thaliana* TPC1 (pdb: 5DQQ) (Figure. 46). The latter was used as a template, since both proteins share sequence homology, particularly in the selective channel. The

resulting model was evaluated using the online 3D refinement server and checkprot server[37], through the calculations of the Q-mean[40], score, Ramachandran calculation and Z-score [39] that all together reflect the acceptability of the models (Table 1). Hence, they were used for further docking studies.

The stereo-chemical quality of the final models was evaluated using three different servers ERRAT[41], which are used to determine the statistical significance of a protein 3D model considering spatial position of amino-acids and overall stability of the structure. Usually more than 90% accuracy is expected for the validity of a model. Furthermore, template-target superposition based on structural alignment has been used in Superpose to understand the coverage between template and target-protein, and the reliability of the model.

Molecular docking

Molecular docking is a computer simulation technique that is used to model the interaction between two molecules, with the aim to find the best position for a ligand (small molecules) in the binding site of a receptor (protein), upon optimization of the interactions with the receptor. This approach allows also the calculation of the binding free energy and compares it to the experimental data. A ligand is in general defined as a small molecule structure that binds to a binding site and the receptor as a structure usually a protein that contains the active binding site. The binding site is the active zone of the receptor that physically interacts with the ligand for docking compound formation.

Protein-ligand interactions

The interactions between a receptor (protein) and a ligand (small molecule) are in general non-covalent including: the vdw interactions for the estimations of the geometric complementarity and steric surface between the ligand and the receptor, and electrostatics interactions, including hydrogen bonds. All of these components are included in the calculation of Gibbs free energy

$$\Delta G = -RT \ln K_i = \Delta H - T\Delta S$$

according to the following equation:

Were:

R: constant gases (8.314 J/K/mol)

T: Temperature (K)

K_i: Inhibition constant

Scoring functions approaches

The evaluation and ranking of predicted ligand conformations is a crucial point[42]. Even when binding conformations are correctly predicted, the calculations ultimately do not succeed if they do not differentiate correct poses from incorrect ones. Thereby, the design of reliable scoring functions and schemes is a fundamental step in the development of a successful molecular docking method. Several free-energy simulation techniques have been developed for quantitative modelling of protein–ligand interactions and the prediction of binding affinity [43] [44].

However, these expensive calculations remain impractical for complex systems and are not always accurate. Scoring functions implemented in docking programs make various assumptions and simplifications in the evaluation of modelled complexes and do not fully account for a number of physical phenomena that determine molecular recognition, for example, and entropic effects (ref). Essentially, three types or classes of scoring functions are currently applied: FORCE-FIELD-based, empirical and knowledge-based scoring functions [45]. As part of this work we used autodock4 [46] to perform our docking calculations in which the force-field based scoring function is implemented, discussed in the following section.

Force field-based methods[45]

In this scoring function method, the ranking of the docking conformation and their affinity are estimated by summing the strength of intermolecular vdW and electrostatic interactions between all atoms of the two molecules in complex by using the noncovalent energy terms of a force field[17].

However, for this purpose, the hydrogen bonding is taken into account with an additional term or being included implicitly in the electrostatic energy term. For example, early versions of AutoDock4 employed energy functions based on the AMBER force field[13] [14], which were computed with either Poisson–Boltzmann (PB) or Generalized Born (GB) continuum solvation models[18]. In AutoDock4, the force field was parameterized using a large number of protein-inhibitor complexes for which both structure and inhibition constants, or K_i , are known. The force field evaluates binding in two steps. The ligand and protein start in an unbound conformation. In the first step, the intramolecular energetics are estimated for the transition from these unbound states to the conformation of the ligand and protein in the bound state. The

second step then evaluates the intermolecular energetics of combining the ligand and protein in their bound conformation [47] [48].

The general functional form below is used in autodock4 ([46]):

$$\Delta G_{\text{binding}} = \Delta G_{\text{vdW}} + \Delta G_{\text{elec}} + \Delta G_{\text{hbond}} + \Delta G_{\text{desolv}} + \Delta g_{\text{tors}}$$

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

Involve the flexibility in the process

The old concept of the key and the lock, used to describe respectively the ligand and the receptor, is nowadays obsolete. A more adapted image of such a complex is rather that of a hand in a glove. Within this view each of the ligand and receptor can modify its shape, by adopting physical complementarity in order to limit the clash interactions and increase quality of the contact between the two entities. One of the main difficulties for the available docking methods is how to treat flexible zones near rigid regions detected by the docking in the majority of proteins. The flexible zones generally correspond to loops locking the ligand. On the other hands, rigid zones mainly correspond to catalytic sites. The main weakness of the docking process is the absence of the inherent movements of proteins. This does not therefore allow optimal co-adaptation of ligand and receptor. Molecular dynamics is able to handle the flexibility of the protein. Therefore, the coupling of the two powerful techniques of docking and molecular dynamics should in principle increases the predictive results.[47] [49]

Details of the Protocol of AutoDock docking

The small molecules used in this study were all collected from PubChem webserver as three dimensional SDF files. AutoDock 4.0 program[46] was used to perform the computational molecular docking. The AutoDockTools (ADT) [46] package was employed to prepare the protein (the TPC model as receptor) and the small molecules. The molecular docking is a computational method used to predict binding sites that are region on the protein that form complex with a small flexible molecule. The program uses a stochastic search method to generate conformations of the ligand at the surface of the protein. Since we cannot a priori define where the inhibitors are able to bind at the surface of the TPC2, the size of the grid is

defined using a grid size of 126Å°, 126Å°, 126 Å° with a 1 Å° spacing including the whole protein to perform “blind docking”. For ligand conformational searching, we used the Lamarckian genetic search algorithm (LGA)[47]. During the docking computations, the maximum of torsions angles was allowed to be rotatable. The docking results were visualized using AutoDock tools and were clustered using 2Å RMS. The most populated/lowest energy poses were examined.

All the initial structures used in this thesis (OccAB1-4 porins with the pdb code respectively: 5DL5, 5DL5, 5DL7, 5DL8 and for *AtTPC1* channel: 5DQQ, 5TUA) have been downloaded from the Protein Data Bank (PDB) [50], and used as initial structures in the simulations and for HsTPC2 homology modelling. All the antibiotics molecules used in chapter 2 were collected from the data base developed in our laboratory [51]. The substrates molecules were collected from Pubchem[52] and prepared using antechamber tool and Gaussian [53]. The small molecules used in the chapter 3 to study the TPC channels were as well collected from PubChem [52]and prepared for docking protocol using autodock4 tools[46]. For parameterization of the natural substrates (arginine, glutamic acid, glycine, ornithine), we followed the standard approach using Generalized AMBER Force Field (GAFF) in addition to RESP charges calculated from a Gaussian 09 structure optimization at the HF-6/31-g*(d) level of theory [53] [54] [14]. The reference data were obtained from the *ab-initio* quantum calculations following the procedure mentioned below:

- The optimized geometry of the molecules was obtained using Gaussian 03 program and at HF/6-31G* level of theory.
- The outputs of the QM simulations were processed using the ‘Antechamber’ package to obtain the initial bonding and nonbonding parameters. The Lennard-Jones potentials, force constants for the bonds and the angles were obtained from the existing GAFF database.
- Partial charges were calculated using CHELPG procedure[55]. The calculation was done on a grid of points around the molecule and the data was used to identify a set of atomic point charges able to fit the electrostatic potential (RESP charges) [53] [54].
- The obtained force field files were checked for missing parameters; if any, they were added manually.

The POPC membrane for the TPC channels was built with CHARMM-GUI [20] using CHARMM36 force field[16]. The sequences files of *HsTPC* channels, included among the

materials, is the file containing the sequence of *hsTPC*, the so called “fasta” format collected from *UniProt* webserver [56]. Two crystal structures of atTPC1 are selected for the second part of the study presented in this thesis: 1) Structure, inhibition and regulation of two-pore channel TPC1 from *Arabidopsis thaliana*- PDB ID: 5DQQ at 2.87 Å resolution. 2) structure of a Na⁺-selective mutant of two-pore channel from *Arabidopsis thaliana* AtTPC1- PDB ID: 5TUA at 3.3 Å resolution.

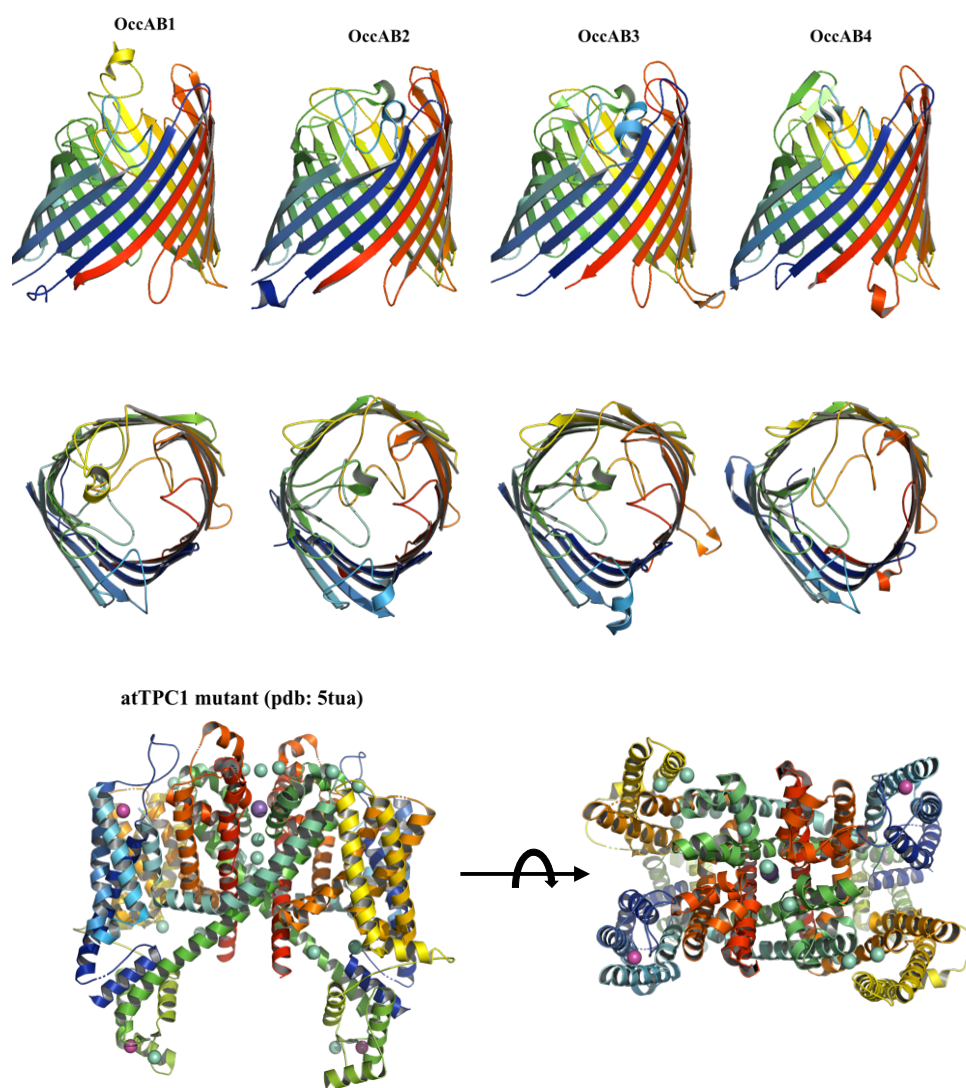


Figure 46: Cartoon representation of pdb structures used in the studies of this thesis. TOP: OccAB [1-4] porins from *A. baumannii* a gram-negative pathogen. Bottom: structures of TPC1 channel from *A. thaliana*, mutant (5TUA) used as template to of build the model HsTPC2 channel.

References

- [1] A. D. Mackerell *et al.*, "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins [†]," *J. Phys. Chem. B*, vol. 102, no. 18, pp. 3586–3616, 1998.
- [2] D. Frenkel, B. Smit, and M. A. Ratner, "Understanding Molecular Simulation: From Algorithms to Applications," *Phys. Today*, vol. 50, no. 7, pp. 66–66, 1997.
- [3] J. Crabbe, "Molecular modelling: Principles and applications: A.R. Leach. Adison Wesley Longman (1996)," *Comput. Chem.*, vol. 21, no. 3, p. 185, 1997.
- [4] J. C. Maxwell, "Illustrations of the dynamical theory of gases. Part II. On the process of diffusion of two or more kinds of moving particles among one another.," *Philos. Mag. J. Sci.*, vol. 20, pp. 21–37, 1860.
- [5] J. C. Maxwell, "Illustrations of the dynamical theory of gases. Part I. On the motions and collisions of perfectly elastic spheres," *Philos. Mag. J. Sci.*, vol. 19, pp. 19–32, 1860.
- [6] L. Verlet, "Computer 'experiments' on classical fluids. II. Equilibrium correlation functions," *Phys. Rev.*, vol. 165, no. 1, pp. 201–214, 1968.
- [7] K. Tu, D. J. Tobias, and M. L. Klein, "Constant pressure and temperature molecular dynamics simulation of a fully hydrated liquid crystal phase dipalmitoylphosphatidylcholine bilayer," *Biophys. J.*, vol. 69, no. 6, pp. 2558–2562, 1995.
- [8] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, "Constant pressure molecular dynamics simulation: The Langevin piston method," *J. Chem. Phys.*, vol. 103, no. 11, pp. 4613–4621, 1995.
- [9] P. P. Ewald, "Die Berechnung optischer und elektrostatischer Gitterpotentiale," *Ann. Phys.*, vol. 369, no. 3, pp. 253–287, 1921.
- [10] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems," *J. Chem. Phys.*, vol. 98, no. 12, pp. 10089–10092, 1993.
- [11] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," *J. Chem. Phys.*, vol. 103, no. 19, pp. 8577–8593, 1995.
- [12] D. A. Case *et al.*, "Amber 2014," no. March 2015, 2014.
- [13] C. J. Dickson *et al.*, "Lipid14: The Amber Lipid Force Field," 2014.
- [14] J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *J. Comput. Chem.*, vol. 25, no. 9, pp. 1157–1174, 2004.
- [15] K. Vanommeslaeghe *et al.*, "CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields," *J. Comput. Chem.*, vol. 28, no. 1, pp. 73–86, 2009.
- [16] J. Huang and A. D. Mackerell, "CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data," *J. Comput. Chem.*, vol. 34, no. 25, pp. 2135–2145, 2013.
- [17] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, "An all atom force field for simulations of proteins and nucleic acids," *J. Comput. Chem.*, vol. 7, no. 2, pp. 230–252, 1986.
- [18] S. Genheden and U. Ryde, "The MM/PBSA and MM/GBSA methods to estimate ligand-

- binding affinities," *Expert Opin. Drug Discov.*, vol. 10, no. 5, pp. 449–461, 2015.
- [19] C. Kandt, W. L. Ash, and D. Peter Tieleman, "Setting up and running molecular dynamics simulations of membrane proteins," *Methods*, vol. 41, no. 4, pp. 475–488, 2007.
 - [20] S. Jo, T. Kim, V. G. Iyer, and W. Im, "CHARMM-GUI: A web-based graphical user interface for CHARMM," *J. Comput. Chem.*, vol. 29, no. 11, pp. 1859–1865, Aug. 2008.
 - [21] M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *J. Chem. Phys.*, vol. 112, no. 20, pp. 8910–8922, May 2000.
 - [22] A. Sali and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints," *J. Mol. Biol.*, vol. 234, no. 3, pp. 779–815, Dec. 1993.
 - [23] M. J. Harvey, G. Giupponi, and G. De Fabritiis, "ACEMD: Accelerating biomolecular dynamics in the microsecond time scale," *J. Chem. Theory Comput.*, vol. 5, no. 6, pp. 1632–1639, 2009.
 - [24] M. Bonomi *et al.*, "PLUMED: a portable plugin for free-energy calculations with molecular dynamics," no. Md, pp. 1–15.
 - [25] C. C. David and D. J. Jacobs, *Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins*, vol. 1084. 2014.
 - [26] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: Fast, flexible, and free," *J. Comput. Chem.*, vol. 26, no. 16, pp. 1701–1718, 2005.
 - [27] A. Laio and F. L. Gervasio, "Metadynamics: A method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science," *Reports Prog. Phys.*, vol. 71, no. 12, 2008.
 - [28] C. B. Barnett and K. J. Naidoo, "Free Energies from Adaptive Reaction Coordinate Forces (FEARCF): An application to ring puckering," *Mol. Phys.*, vol. 107, no. 8–12, pp. 1243–1250, 2009.
 - [29] A. Laio and M. Parrinello, "Escaping free-energy minima," *Proc. Natl. Acad. Sci.*, vol. 99, no. 20, pp. 12562–12566, 2002.
 - [30] M. Ceccarelli, C. Danelon, A. Laio, and M. Parrinello, "Microscopic Mechanism of Antibiotics Translocation through a Porin," *Biophys. J.*, vol. 87, no. 1, pp. 58–64, 2004.
 - [31] A. Barducci, G. Bussi, and M. Parrinello, "Well-tempered metadynamics: A smoothly converging and tunable free-energy method," *Phys. Rev. Lett.*, vol. 100, no. 2, pp. 1–4, 2008.
 - [32] M. Bonomi and M. Parrinello, "Enhanced sampling in the well-tempered ensemble," *Phys. Rev. Lett.*, vol. 104, no. 19, pp. 1–4, 2010.
 - [33] B. Roux, T. W. Allen, S. Bernèche, and W. Im, "Theoretical and computational models of biological ion channels," *Curr Opin Struc Biol*, vol. 12, no. 2, pp. 182–189, 2002.
 - [34] S. Goldsmith-fischman, S. Goldsmith-fischman, B. Honig, and B. Honig, "Structural genomics: Computational methods for structure analysis," *Society*, pp. 1813–1821, 2003.
 - [35] M. A. Marti-Renom, A. C. Stuart, R. Sanchez, F. Melo, and A. Sali, "Comparative Protein Structure Modeling of Genes and Genomes," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 29, pp. 291–325, 2000.
 - [36] A. Fiser and A. Šali, "MODELLER: Generation and Refinement of Homology-Based Protein Structure Models," *Methods Enzymol.*, vol. 374, pp. 461–491, 2003.

- [37] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton, "PROCHECK: a program to check the stereochemical quality of protein structures," *J. Appl. Crystallogr.*, vol. 26, no. 2, pp. 283–291, 1993.
- [38] R. Laskowski, J. A. Rullmann, M. MacArthur, R. Kaptein, and J. Thornton, "AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR," *J. Biomol. NMR*, vol. 8, no. 4, pp. 477–486, 1996.
- [39] P. Benkert, M. Biasini, and T. Schwede, "Toward the estimation of the absolute quality of individual protein structure models," *Bioinformatics*, vol. 27, no. 3, pp. 343–350, 2011.
- [40] P. Benkert, S. C. E. Tosatto, and D. Schomburg, "QMEAN: A comprehensive scoring function for model quality assessment," *Proteins Struct. Funct. Bioinforma.*, vol. 71, no. 1, pp. 261–277, 2008.
- [41] M. W. MacArthur, R. A. Laskowski, and J. M. Thornton, "Knowledge-based validation of protein structure coordinates derived by X-ray crystallography and NMR spectroscopy," *Curr. Opin. Struct. Biol.*, vol. 4, no. 5, pp. 731–737, 1994.
- [42] H. J. Böhm, "The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure," *J. Comput. Aided. Mol. Des.*, vol. 8, no. 3, pp. 243–256, 1994.
- [43] P. Kollman, "Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena," *Chem. Rev.*, vol. 93, no. 7, pp. 2395–2417, 1993.
- [44] T. Simonson, "Free Energy Simulations Come of Age: Protein- Ligand Recognition," *Acc. Chem. Res.*, vol. 35, no. 6, pp. 430–437, 2002.
- [45] J. Liu and R. Wang, "Classification of current scoring functions," *J. Chem. Inf. Model.*, vol. 55, no. 3, pp. 475–482, 2015.
- [46] G. Morris and R. Huey, "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility," *J. ...*, vol. 30, no. 16, pp. 2785–2791, 2009.
- [47] D. S. Goodsell, G. M. Morris, and A. J. Olson, "Automated docking of flexible ligands: applications of AutoDock," *J Mol Recognit*, vol. 9, no. 1, pp. 1–5, 1996.
- [48] G. M. Morris *et al.*, "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function," *J. Comput. Chem.*, vol. 19, no. 14, pp. 1639–1662, 1998.
- [49] R. A. Friesner *et al.*, "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy," *J. Med. Chem.*, vol. 47, no. 7, pp. 1739–1749, 2004.
- [50] F. C. Bernstein, "The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures," *Alcohol*, vol. 112, pp. 535–542, 1977.
- [51] G. Mallocci, A. V. Vargiu, G. Serra, A. Bosin, P. Ruggerone, and M. Ceccarelli, "A database of force-field parameters, dynamics, and properties of antimicrobial compounds," *Molecules*, vol. 20, no. 8, pp. 13997–14021, 2015.
- [52] E. E. Bolton, Y. Wang, P. A. Thiessen, and S. H. Bryant, *Chapter 12 PubChem: Integrated Platform of Small Molecules and Biological Activities*, vol. 4, no. 8. Elsevier B.V., 2008.
- [53] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, "ARTICLES A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving," pp. 10269–10280, 1993.
- [54] A. C. A. M. I, A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, "Fast, Efficient

- Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method," vol. 21, no. 2, pp. 132–146, 1999.
- [55] C. M. Breneman and K. B. Wiberg, "Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis," *J. Comput. Chem.*, vol. 11, no. 3, pp. 361–373, 1990.
- [56] A. Bateman *et al.*, "UniProt: The universal protein knowledgebase," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D158–D169, 2017.

Chapter 5

General conclusion

General conclusion

The ion channel membrane proteins are fundamental in several biological functions of the surface cellular cells including cellular signaling and transport regulations. Their dysfunction or deregulation conduct to dramatic cellular damage such as the onset of serious and sometimes fatal illness, infectious disease or heart disease. Nowadays, these membrane proteins are one of the most challenging drug targets therapeutics. For the drug discovery industry, the most important step in the drug discovery early stage is the characterization of the protein-ligand binding that is crucial in developing drugs that can inhibit or activate the function of almost all ions channels.

The work of this thesis concerns two cases of ions channel implicated in diversified mechanisms and several diseases. The general goal is to describe, at atomic scale by using computational methods, the physicochemical parameters that govern the formation of optimal interactions in the channel-ligand complexation.

The first case concerns the most expressed ions channels in gram-negative bacteria, namely “porins”. Specifically, it concerns the OccAB porins of *A. baumannii* implicated in the transport of small molecule and potentially antibiotics too, all details are presented in the chapter 2. The second case concerns a gated ion channel expressed in acidic organelles of animal cells, the HsTPC2 channel specific to the translocation of sodium ions, the details concerned this study are presented in the chapter3. The main aim is to improve our understanding of the relationship between the 3D-structure of these channels with their functions. To achieve these objectives, we used several computational modelling methods such as homology modeling, classical MD simulations, enhanced sampling methods and molecular docking simulations. All of these methods have been largely used in the past to investigate similar biological membrane systems.

The data of the chapter 2 showed major differences among OccAB porins in their 3D structure organization, and in the behavior of their surface exposed loops, and reported a detailed comparison of the residues forming their CR that plays a major role in the translocation process. We also reported the effect of the dynamics of channels on the translocation process. Here, we provided an explanation of the unstable states of ions selectivity measurements revealed during the electrophysiology experiments on OccAB2. Indeed, the results of our MD simulations combined with PCA analysis confirmed that OccAB2 is the most dynamical pore through the dynamic of the residues lining the pore located on the constricting loop L7.

By the use of metadynamics simulations we were able to investigate at atomic scale the translocation of small molecules across OccAB[1-4]. For each selected molecule, we reconstructed the free energy surface controlling translocation. We thus provided the structural, dynamical, and physicochemical properties that are involved in the discrimination, uptake, and translocation mechanisms of specific substrates of each OccAB porins. More specifically, we demonstrated that the both electrostatics and size properties are important in the discrimination of the molecules that can pass through these channels. The OccAB1 porin can translocate both positive and negative substrates without any clear preference and at low energy cost, in agreement with the experimental data. In OccAB2-4, the translocation of the selected substrates is limited by the size of their pore such that only small molecule can pass through easily with low energy costs as shown with glycine. The free energy calculation also confirmed that OccAB[2-4] are selective to negative substrates as suggested by liposome swelling assays and electrophysiological experiments. Secondly, the results of antibiotics translocation through OccAB1 channel revealed the importance of the internal electric field present in the CR of OccAB1 in guiding the translocation process of antibiotics. The results of the FESs revealed a common trend observed for all the molecules investigated here. The internal electric field in OccAB1 has an effect on the translocation mechanism: the molecules have to align their dipole moment to the transversal electric field in order to overcome the barrier in the CR. The metadynamics simulations allowed reporting the main interaction in the binding sites involved in the mechanism of the translocation of each molecule through OccAB1.

As second part of this thesis, we reported a 3D-structure homology model of human TPC2 channel that allowed the investigation at atomic scale of binding by using molecular docking simulations: naringenin and verapamil known as blockers of gated-ions channels. Our results showed that the both molecules can bind to the TPC2 channel, in the central pore cavity as well as in the voltage-gated domains. More particularly, we reported that given the very low predicted K_i in all the binding sites, naringenin has a higher affinity and is more potent to likely inhibit the channel than verapamil. However, experimental studies are crucial in such computational studies to confirm and support our data. Methods such as mutagenesis experiment would help in checking the involved residues in the binding sites, and might confirm the mechanism of the interaction revealed in this study.

In summary, our results provide information at the atomic scale that helps to progress our knowledge about the molecular and biophysical properties of the human channel TPC2 and

OccAB porin. This information may serve as a support for conducting experimental investigations such as the mutagenesis of suggested key residues involved in the binding sites and their importance in the mode of binding proposed in this work. The work of this thesis aimed to contribute to gain understanding of the structure-function relationship of these ion channels.

Overall, we have shown that the use of different computational techniques can lead to describe complex biophysical systems and their performance in providing answers to biophysical questions that standard experimental techniques are currently not able to describe fully at atomic scale. The metadynamics was an appropriate method to describe the rare event of small molecules translocation through OccAB channels, by providing the free energy associated to this biological phenomenon. It also permits to identify the residue pairs forming hydrogen bonds in the molecules-channel complex. We also demonstrate that even if the experimental data do not exist, the methods of molecular homology modeling combined with the docking simulations by using autodock4 algorithm can rapidly determine the binding mode between the ligand and the channel and accurately calculate the its free energy. However, the both techniques used in the current thesis have their proper limitations. The metadynamic technique requires a large sampling of conformational space to be able to calculate the free energy of the whole process such as the process of transporting small molecules, therefore it is limited by a slow convergence.

In the case of docking simulation technique used to investigate potential inhibitors of the human TPC2 channel, although faster than metadynamics simulation in the free energy of binding calculation, it has serious limitations. The approximations used for scoring do not include all the parameter that would describe properly the protein-ligand complexes such as the effects of the solvation. The other issue of the docking calculations is the lack of flexibility of the receptor while performing the docking simulations. Finally, the huge size of biological membrane protein systems limits the accuracy of these computational methods in predicting results that reflect those in their natural environment, including the effects of the membrane in the protein-ligand binding mechanism. Therefore, all of these issues present today challenges that must be overcome for application in the early step of drug discovery, particularly in diseases involving membrane protein systems whose high-resolution structures are not always available.