Studies on the function and substrate permeation in ABC transporters by biomolecular simulations.

Dissertation presented to obtain the Doutoramento (Ph.D.) degree in Molecular Biosciences at the Instituto de Tecnologia Química e Biológica António Xavier - Universidade Nova de Lisboa

Bárbara Palma de Abreu Caldeira Pinto

Supervised by:

Dr. Cláudio M. Soares

Dr. Ana Sofia Oliveira

Oeiras, June 2021

“There is no choice but to proceed”

Joana Avillez

List of Contents

Acknowledgements

Even though a PhD is a “one-man show”, I have to express my gratitude to some people that have marked this journey.

To Cláudio Soares, my supervisor, not only for having accepted me as a PhD student, but also for all his support and guidance throughout this process, despite his busy schedule.

To the thesis committee for invaluable advice that contributed to shape the work here presented.

To Ana Sofia Oliveira, my co-supervisor, for a couple of useful opinions and advices.

To Carlos Cruz for his precious insights on the analysis of umbrella sampling simulations.

To Diana Lousa, not only for opinions and discussions related with work, but also for her friendship and great company in the lab.

To Dragana Barros, for being a great and caring friend, always with good energy and a young spirit.

To the members of the Protein Modelling, Molecular Simulation and Multiscale Modelling groups (both past and present members) that assisted me in the most various things, ranging from the hardware maintenance to coding advice. A special thanks to the ones that made my days a bit more joyful.

To my mother, who was my emotional cornerstone over these years, dealing with all my complaints, and for providing wisdom and support.

To João Carlos, my friend for almost 20 years, for his support, comprehension, positive energy and accurate opinions and viewpoints.

To ITQB, that is indeed, a great place for making a PhD, with excellent resources, facilities, and a truly scientific mindset, allowing to get the most out of scientific endeavours.

To FCT for financial support.

To all the people who gave me any kind word of advice and/or support during this period.

List of publications

Papers presented in this thesis

Abreu, B.; Lopes, E. F.; Oliveira, A. S. F.; Soares, C. M. The mutation F508del disturbs the dynamics of the nucleotide binding domains of CFTR before and after ATP hydrolysis, *Proteins Struct. Funct. Bioinforma.* **2019**, **88**, 113

Kaur, H.; **Abreu, B**.; Akhmetzyanov, D.; Lakatos-Karoly, A.; Soares, C. M.; Prisner, T.; Glaubitz, C. Unexplored nucleotide binding modes for the ABC exporter MsbA. *J. Am. Chem. Soc.* **2018**, **140**, 14112

Abreu, B., Cruz, C., Oliveira, A. S., Soares, C., ATP hydrolysis and nucleotide exit enhance maltose translocation in the MalFGK$_2$E importer, *Scientific Reports,* **2021, 11,** 10591

Abstract

The translocation of molecules across cell membranes often requires the participation of transmembrane proteins. The **A**TP **B**inding **C**assette (ABC) transporters are one of the major class of proteins dedicated to the translocation of molecules across membranes. ABC transporters make use of ATP hydrolysis and undergo a series of large-scale conformational changes in order to carry out their function. Despite being extensively studied, some molecular details regarding their function remain undisclosed. In this thesis, the function of three relevant ABC systems, CFTR, MsbA and the *Escherichia coli* $MalFGK_2E$ importer, was investigated using computational methods, such as molecular dynamics simulations. The effect of the CFTR mutation F508del, responsible for cystic fibrosis, was studied in the nucleotide binding domains of the CFTR receptor. In the MsbA protein, the molecular details of nucleotide binding during the adenylate kinase cycle were investigated in collaboration with experimental groups with expertise in solid state NMR and EPR. Finally, the details of substrate translocation in the $MalFGK_2E$ importer were also researched.

Regarding the role of the mutation F508del in the isolated nucleotide binding domains of CFTR (chapter 3), it was found that F508del increases the opening degree of the ATP binding pockets and decreases the amount of catalytically competent conformations, i.e, conformations able to carry ATP hydrolysis. The short-term conformational changes after 300 ps of hydrolysis were affected by F508del, showing lower magnitude in the mutant form. Nonetheless, after longer simulation time, the magnitude of the conformational changes has become similar in most regions, except on the F508del loop, the residues surrounding the catalytic site and the portion of NBD2 adjacent to ABP1. These findings show that F508del can influence NBD dynamics regardless of the presence of transmembrane domains.

Chapter 4 describes the study of nucleotide binding modes during the reverse adenylate kinase reaction in MsbA. MsbA is a model system for ABC exporters and it is able to perform a reverse adenylate kinase (AK) reaction, which involves phosphate transfer from 2 ADP molecules to generate ATP and AMP. It was discovered that the AK binding site is located in between the Q-loops of each NBD. The usage of molecular dynamics simulations showed that the nucleotide tends to use the adenine moiety to interact with the Q-loop motifs, while the phosphate is oriented towards the histidine residue from the H-loop. This binding mode seems to be essential for a proper orientation towards the catalytic site, leaving no space for binding of two nucleotides at the same time.

Substrate binding and transport in the maltose importer $MalFGK_2E$ (chapter 5), was studied in the pre-hydrolysis, post-hydrolysis and nucleotide-free states. The nucleotide free-state is

intended to mimic nucleotide exit after hydrolysis. It was concluded that ATP hydrolysis and nucleotide exit trigger conformational changes in the transmembrane helices that result in the decrease of energetic barriers to maltose translocation towards the cytoplasm, with a concomitant increase of the energy in the periplasmic side of the pore, contributing for the irreversibility of the process. ATP hydrolysis generates conformations similar to the nucleotide-free state, leading to increased similarity from the energetic viewpoint. Furthermore, the global minimum of all the free energy profiles, was observed when maltose was bound to its binding pocket on MalF, previously identified in the pre-hydrolysis X-ray structures. Key residues that aid in positioning and orientation of maltose were identified, as well as a novel binding pocket for maltose in MalG. Additionally, Y383 was observed to act as a first gatekeeper changing drastically its conformation upon hydrolysis, allowing increased diffusion towards the NBDs. This study sheds light on the ABC type I importer mechanism, explaining the contribution of ATP hydrolysis and nucleotide exit in the transport cycle.

All the three projects above described contribute for a deeper understanding of the mechanisms underlying the function of ABC transporters.

Resumo

As proteínas transmembranares estão frequentemente envolvidas no movimento de moléculas através de membranas celulares. Os transportadores **A**TP **B**inding **C**assette (ABC) constituem uma das principais classes de proteínas dedicadas ao transporte molecular através de membranas. Os transportadores ABC utilizam a hidrólise de ATP e sofrem uma série de alterações conformacionais de grande escala para poderem desempenhar a sua função.

Apesar destas proteínas serem exaustivamente estudadas, existem detalhes moleculares respeitantes ao seu funcionamento que permanecem largamente desconhecidos. Nesta tese, a função de três sistemas ABC: CFTR, MsbA e o importador $MalFGK_2E$ de *Escherichia coli*, foi investigada usando métodos computacionais como simulações de dinâmica molecular. O efeito da mutação F508del no receptor CFTR, responsável pela fibrose quística, foi estudada nos domínios de ligação a nucleótidos da mesma. Na proteína MsbA averiguou-se o modo de ligação dos nucleótidos durante o ciclo de adenilato cinase no transportador MsbA em colaboração com grupos experimentalistas na área de RMN em estado sólido e RPE. Por fim, os detalhes da translocação de substratos no importador de maltose $MalFGK_2E$ de *Escherichia coli* foram investigados.

Relativamente ao papel da mutação F508del nos domínios catalíticos isolados da CFTR (capítulo 3), descobriu-se que esta mutação estimula a abertura dos locais de ligação de ATP e conduz à diminuição do número de conformações cataliticamente competentes, i.e, conformações capazes de hidrolisar ATP. Na presença da F508del as alterações conformacionais provocadas pela hidrólise de ATP, após 300 ps de simulação, foram de menor magnitude. Contudo, após tempos maiores de simulação, a magnitude das alterações conformacionais tornou-se semelhante nos dois estados, excepto em algumas regiões como o *loop* onde a mutação está presente, os resíduos que rodeiam o centro catalítico e a porção do NBD2 adjacente ao ABP1. Estes resultados mostram que a mutação F508del pode influenciar a dinâmica dos domínios catalíticos, independentemente da presença de domínios transmembranares.

O capítulo 4 descreve o estudo dos modos de ligação dos nucleótidos durante o ciclo catalítico de adenilato cinase da MsbA. O transportador MsbA é um sistema modelo para os exportadores ABC e é capaz de realizar a reacção reversa de adenilato cinase, que consiste na transferência de um grupo fosfato entre duas moléculas de ADP para gerar ATP e AMP. Descobriu-se que os nucleótidos extra que participam nesta reacção ligam-se entre os *Q-loops* de cada domínio catalítico. As simulações de dinâmica molecular mostraram que a adenina tende a interagir com os *Q-loops*, enquanto que o fosfato fica orientado na direcção do resíduo de histidina do *H-loop*.

Desta forma, este modo de ligação garante uma orientação correcta em direção ao local catalítico, mas parece não deixar espaço para a ligação simultânea de dois nucleótidos neste local.

Por fim, a ligação e transporte de substrato no importador de maltose $MalFGK_2E$ (capítulo 5) foi estudada nos estados pré-hidrolíse, pós-hidrólise e na ausência de nucleótido. O estado sem nucleótidos nos domínios catalíticos tem o objectivo de mimetizar a saída dos mesmos após a hidrólise. Concluiu-se que a hidrólise de ATP e a saída dos productos da reacção levam a alterações conformacionais nos domínios transmembranares, resultando na diminuição de barreiras energéticas à translocação da maltose na direcção citoplasmática, com um simultâneo aumento da energia livre na direcção periplasmática, contribuindo para a irreversibilidade do processo. A hidrólise de ATP tende a gerar conformações semelhantes às amostradas pelo estado sem nucleótidos, conduzindo a semelhanças do ponto de vista energético. Adicionalmente, o mínimo de energia global corresponde à ligação da maltose no local de ligação observado nas estruturas cristalográficas. Foi possível identificar resíduos chave que participam na orientação e posicionamento correcto da maltose, assim como um local de ligação extra na MalG, não reportado previamente. Observou-se que o resíduo Y383 muda drasticamente de conformação após a hidrólise, permitindo o aumento da difusão vertical na direcção citoplasmática. Desta forma, este estudo contribui para o esclarecimento do mecanismo transporte de substrato típico de importadores ABC de tipo I, explicando a contribuição da hidrólise de ATP e a saída de nucleótido na actividade destes transportadores.

Os três projectos acima descritos contribuem para uma maior compreensão dos mecanismos subjacentes à função de transportadores ABC.

List of abbreviations and symbols

Aminoacids

Alanine	A	Ala	Leucine	L	Leu
Arginine	R	Arg	Lysine	K	Lys
Asparagine	N	Asn	Methionine	M	Met
Aspartate	D	Asp	Phenylalanine	F	Phe
Cysteine	C	Cys	Proline	P	Pro
Glutamine	Q	Gln	Serine	S	Ser
Glutamate	E	Glu	Threonine	T	Thr
Glycine	G	Gly	Tryptophan	W	Trp
Histidine	H	His	Tyrosine	Y	Tyr
Isoleucine	I	Ile	Valine	V	Val

Abbreviations

ABC	ATP binding cassette
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
AK	Adenylate kinase
C-α	Carbon α
CE	Continuum Electrostatic
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
Cryo-EM	Cryogenic electron microscopy
DNA	Deoxyribonucleic acid
EPR	Electronic paramagnetic resonance
FRET	Förster resonance energy transfer
GROMACS	Groningen Machine for Chemical Simulations
GROMOS	Groningen Molecular Simulation
ICL	Intracellular loop
LINCS	Linear constraint solver

MBP	Maltose Binding Protein
MD	Molecular Dynamics
Mg^{2+}	Magnesium ion
MM	Molecular mechanics
µs	Microsecond
µm	Micrometer
ns	Nanosecond
nm	Nanometer
NBD	Nucleotide Binding Domain
MAS-NMR	Magic angle spinning - Nuclear magnetic resonance
LPS	Lipopolyssacharide
PBC	Periodic Boundary Conditions
PDB	Protein Data Bank
PMF	Potential of Mean Force
PKA	Protein kinase A
Pi	Inorganic Phosphate
ps	Picosecond
QM-MM	Quantum Mechanics-Molecular Mechanics
RI	Regulatory insertion
RE	Regulatory extension
RESP	Restrained Electrostatic Potential Fitting
RNA	Ribonucleic acid
RMSD	Root Mean Square Deviation
SBP	Substrate Binding Protein
TM	Transmembrane
TMD	Transmembrane Domain
TMH	Transmembrane Helix

Latin and Greek Symbols

V	Potential energy
k_b	Force constant for bond strength
k_θ	Force constant for the θ angle
k_ξ	Force constant for the ξ dihedral
k_ϕ	Force constant for the ϕ dihedral
σ	Separation distance for a pair of atoms in which the energy is zero in the Lennard-Jones equation.
ϵ	Minimum potential energy between two atoms in the Lennard-Jones equation.
ϵ_0	Electric permittivity in vacuum
ϵ_r	Electric permittivity in the considered medium
r	Position vector
r_{ij}	Interatomic distance between atoms *i* and *j*
m_i	Mass of atom i
a_i	Acceleration of atom i
t	Time
v_i	Velocity of atom i
K	Kinetic energy
k_B	Boltzmann constant
T	Temperature
τ_T	Coupling constant of the thermostat
E	Total energy
P	Pressure
β	Inverse thermodynamic temperature
Q	Partition function
A	Helmholtz free-energy
U	Internal energy
w_ξ	Umbrella potential
ξ	Reaction coordinate ξ
ΔG	Gibbs free energy
R	Gas constant

W_{ij}	Electrostatic contribution for free energy
$\Phi(r)$	Electrostatic potential at point r

Chapter 1

Introduction

1.1-ABC transporters

Cells are delimited by membranes, which are structures formed by amphipathic lipids, such as phospholipids. These molecules are arranged in a bilayer, in which the hydrophilic part faces the aqueous medium (both intra and the extra cellular media) while the hydrophobic one is facing the hydrophobic portion of the opposite layer. Membranes form a hydrophobic barrier, mostly allowing the permeation of hydrophobic molecules. In this way, transmembrane proteins are essential to allow hydrophilic molecules and ions to cross this barrier efficiently. The nature of molecular transport across membranes can vary: in passive transport, substrates are transported down the concentration (or electrochemical) gradient without any spending of energy. An example of passive transport are ion channels. On the other hand, active transport makes use of energy input to translocate substrates against the concentration gradient [1]. If the energy source is derived from ATP, redox or light energy, it is considered primary transport, while secondary transporters use energy stored in electrochemical gradients. **A**TP **B**inding **C**assette (ABC) transporters are an example of active and primary transport.

ATP binding cassette proteins are one of the largest protein superfamilies. ABC transporters make use of ATP hydrolysis to power the import and export of allocrites across the membrane. The substrates transported range from aminoacids, sugars, lipids, metabolites, large cytotoxic compounds and even ions[2]. Hence, these proteins play a large variety of roles such as drug excretion in bacteria and cancer cells, lipid export for membrane building and assembling, nutrient intake and even maintenance of transmembrane gradients. Therefore, ABC proteins play an important role in phenomena such as: brain and organ homeostasis (ABC transporters are abundant on the blood-brain barrier and other barrier tissues), multidrug resistance and in diseases as cystic fibrosis, anaemia, cholesterol/bile defects, eye disorders and even neurodegenerative diseases[3].

1.2-Structure and function of ABC proteins

ABC proteins can be functionally divided in three categories: exporters, importers and proteins with functions non-related to transport. The latter ones play roles in translation, DNA repair and regulation of gene expression. Exporters carry allocrites from the intracellular medium to the external medium, while importers perform transport of solutes in the opposite direction, aided by soluble accessory proteins, the substrate binding proteins[4].

All members share the existence of two highly conserved nucleotide binding domains, with ATPase activity, possessing characteristic ATP binding sequences: such as the ABC motif, exclusive of ABC transporters, as well as the Walker A and Walker B motifs, typical of ATPases.

These nucleotide binding domains face each other and possess complementary motifs relevant for nucleotide binding, which are sandwiched between the two domains[5]. ABC transporters present in humans are divided by subclasses: ABCA, ABCB, ABCC, ABCD, ABCE, ABCF and ABCG[3].

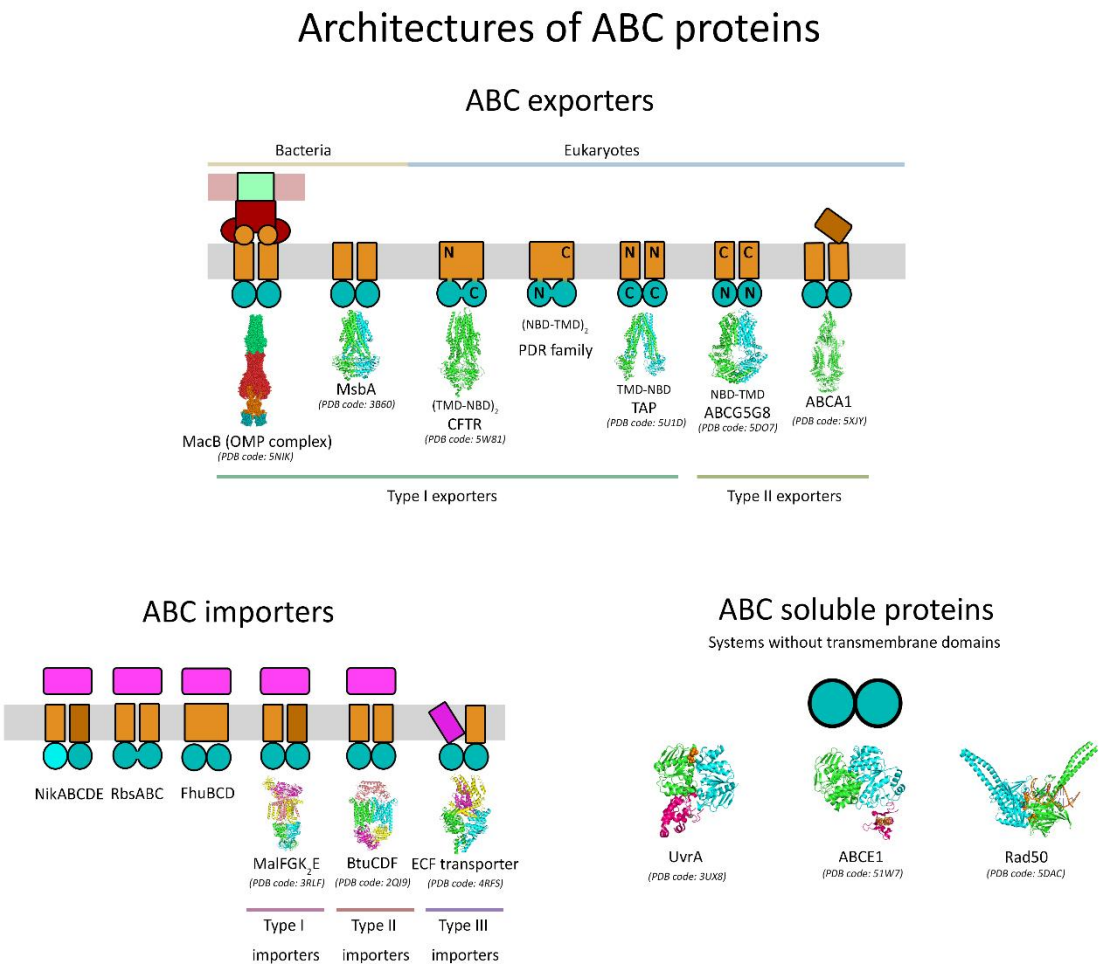


Figure 1.1 *- Variety in ABC proteins architectures. ABC importers and exporters have transmembrane domains that allow the transport of substrates. The transmembrane domains are represented in orange, while the nucleotide binding domains are represented in teal and substrate binding proteins in ABC importers are represented in purple. Within ABC exporters several architectures can be observed: $(TMD\text{-}NBD)_2$ – architecture in which the transmembrane domains and the nucleotide binding domains are fused in one polypeptide, but the N terminal is located in one of the TMDs. $(NBD\text{-}TMD)_2$ - architecture in which the nucleotide binding domains and the transmembrane domains are fused in one polypeptide, with the N terminal is located in one of the NBDs. TMD-NBD – architecture in which the transporter is constituted by two chains and the N-terminal is on the transmembrane domains. NBD-TMD – architecture in which the transporter is constituted by two chains and the N-terminal is on the nucleotide binding domain. The MacA tube channel in the OMP complex is represented in red, the periplasmic domain in ABCA1 is displayed in brown. ABC importers can have fused TMDs (FhuBCD) or NBDs (RbsABC), or they can be constituted by several different chains as seen in NikABCDE and $MalFGK_2E$ (indicated the cyan NBD and dark brown TMD, respectively).*

A detailed description of the location and function of these transporters can be found in Holland *et al*[3].

Regarding ABC proteins with transporter function, not only they contain the two nucleotide binding domains, but also have two transmembrane domains. Despite this common architecture, they display a series of variations.

Exporters can have a series of different architectures (figure 1.1): they can be composed by two separate protein chains, that correspond to a nucleotide binding domain and a transmembrane domain (TMD-NBD or NBD-TMD). The TAP transporter is an example of an exporter with two chains. Alternatively, the NBDs and the TMDs can be fused in a single polypeptide chain. An example of such a protein is CFTR, with the architecture $(TMD\text{-}NBD)_2$, in which the N-terminal is located at the TMDs. The N and C terminal can either be at any of the TMDs or the NBDs[3,6]. The architecture $(NBD\text{-}TMD)_2$ in which the NBDs are fused with TMDs, and the NBDs contain the N-terminal, is exclusive of eukaryotic systems and it is characteristic of pleiotropic drug resistance (PDR) transporters, which are found only in plants and fungi. Additionally, some exporters can also have extra domains that allow the interaction with other proteins, such as the case of MacB in bacteria that has periplasmic domains that allow the connection with the MacA tube channel, connecting the inner membrane to the outer membrane in gram-negative bacteria[7]. An example in eukaryotes is ABCA1, that also displays an extra periplasmic domain for interaction with lipoprotein[8]. ABCA and ABCG are type II exporters due to the TMD structure, in which the folding of each TMD chain is different and there is no intersection between the helices of the two halves of the transporter, in opposition to type I exporters. Additionally, the coupling helices between TMDs and NBDs lay in the membrane boundary rather than being projected to the cytoplasm, as in type I importers [8].

Importers are exclusive of bacteria and not only have two transmembrane domains and two nucleotide binding domains, but they also count with a soluble substrate-binding protein (SBP). SBP binds the substrate in solution and carries it to the protein complex. Most importers are constituted by several independent polypeptides[9]. Nonetheless, there are complexes composed by fused TMDs or fused NBDs, such as the cases of FhuBCD or RbsABC, respectively. Structurally and functionally, importers can be divided in three categories: type I, II and III importers [10,11]. Type I importers have a lower number of transmembrane helices (≈ 12 helices), displaying a common helical core, and they also display a weak constitutive ATPase activity, being stimulated by the binding of liganded SBP. Nonetheless, the affinity of liganded SBP to the protein complex is variable. Examples of type I importers are the maltose importer $MalFGK_2E$ and the methionine importer MetNI. In contrast, type II importers generally display a larger number of transmembrane helices (≈ 20 helices), and also contain a common helical core. However, in

contrast with type I importers, the liganded SBP shows very high affinity to the protein complex [10]. In addition, type II importers are able to translocate large-size molecules such as vitamin B_{12}. Examples of type II importers are the vitamin B_{12} importer BtuCD and the heme importer HmuUV. Some importer proteins possess covalently bound SBPs, increasing transport efficiency[11]. Type III importers include the energy coupling factor (ECF) proteins. ECF proteins have two nucleotide binding domains, but only one transmembrane domain (EcfT or T-component) bound to the NBD dimer. The S component or EcfS subunit is an integral binding protein that diffuses across the membrane delivering the substrate and, when bound to the NBD-EcfT complex, acts as another TMD. Some energizing modules can bind to several S components, while others are specific for a certain S component[10].

Regarding ABC proteins with no transporter functions, besides the nucleotide binding domains, these proteins can have extra domains that interact with other proteins or even DNA. Members of this class are involved in translation regulation, ribosome biogenesis and DNA repair by nucleotide excision[3,6]. These include the eukaryotic ABCE and ABCF families. Examples of class 2 proteins include ABCE1 protein, the UvrA protein and Rad50. ABCE1 is responsible for signalling translation termination and ribosome recycling[3,5]. ABCE1 also contains a N-terminal domain with two non-equivalent $[4Fe\text{-}4S]^{2+}$ clusters[12]. The iron-sulfur centers are essential for activity but they are not redox active[13]. The UvrA protein is involved in DNA repair and antibiotic resistance in *Escherichia coli*. In UvrA there is a zinc finger domain intercalated in between the ABC and walker A motifs. Rad50 is an eukaryotic protein involved in DNA double-strain repair. It forms a complex with the nuclease Mre11 and it possesses two coiled-coil domains that are 15 to 50 nm long, each with a terminal Zn hook dimerization motif, that play roles in DNA-end processing and activation of other proteins[14].

Concerning the phylogenetic classification of ABC transporters, Saurin *et al*[6] classified all ABC transporters based on the nucleotide sequences of the NBDs and found three groups: the first one contemplates transporter systems with fused nucleotide binding domains (NBDs) to transmembrane domains (TMDs), the second contemplates ABC proteins with no transmembrane domains and that are constituted by one chain, while the third contemplates ABC proteins with transmembrane domains with several independent polypeptide chains (including importers) [6,9]. Recent analysis by Xiong *et al*[15] on eukaryotic ABC transporters show that eukaryotic transporters can be divided into three groups: one that includes ABCA and ABCG families, other that only contemplates ABCC, ABCB and ABCD families and a last one with ABCE and ABCF families[15].

Concerning the evolutionary path of ABC transporters, it is thought that the last universal common ancestor already possessed the NBD dimer, due to the similarities in nucleotide and

protein sequences between prokaryotic and eukaryotic half transporters and in between prokaryotic NBDs and ABCF proteins[3,6,9,16].

The large diversity of ABC proteins can be explained by genetic events of fusion and duplication of genes. It is possible that gene fusion happened between the NBD units fused to the transmembrane domain, while the half transporters with two copies of the same chain are an example of gene duplication. Moreover, most eukaryotic transporters are found in organelle membranes, which means that some systems might be acquired by symbiotic bacteria that were the ancestors of organelles or by horizontal gene transfer[3,6,9,16].

1.2.1-The nucleotide binding domains

The nucleotide binding domains (NBDs) are responsible for ATP binding and hydrolysis. The NBDs are essential for transporter function and occur in a dimeric form (figure 1.2A). These domains are highly conserved, showing high sequence identity – from 30 to 50%[17]. Each NBD monomer can be divided in two regions: a catalytic core domain that contains all the motifs important for ATP binding and hydrolysis, and an α-helical subdomain, exclusive of the ABC subfamily (figure 1.2B). The catalytic subdomain (also known as the RecA-like subdomain) is similar to other ATPases and possess important motifs for ATP binding and hydrolysis: the A-loop, Walker A motif (also known as P-loop), the Walker B motif, the Q-loop and the H switch. The A-loop contains a highly conserved aromatic residue that is involved in π-π stacking with adenine. The Walker A motif (GxxGxGKS/T) participates in the binding to the β and γ phosphate groups of ATP and forms a loop engulfing the latter. The lysine together with the backbone amide groups of the remaining residues form a network of interactions with the β and γ phosphate groups [18]. The Walker B motif (ϕ ϕ ϕ ϕDE), in which ϕ is an hydrophobic residue, contains a catalytic glutamate and a conserved aspartate residue that coordinates the magnesium ion. Despite several mechanisms having been proposed for ATP hydrolysis in ABC proteins, it is generally accepted that the glutamate residue is responsible for the polarization of the attacking water molecule. The D-loop (SALD) not only contributes for the NBD dimer stability, but it also coordinates with the attacking water molecule ensuring a proper orientation for attack. This happens via an hydrogen bond between the alanine backbone and the catalytic water [19,20]. The Q-loop contains a conserved glutamine and is located in the interface between the catalytic domain and the α-helical one. The conserved glutamine coordinates the Mg^{2+} ion and the attacking water [21,22].The Q-loop can interact with the transmembrane domains, contributing for signal transmission[23]. The H switch is a highly conserved histidine residue that can interact with the γ-phosphate of the ATP, the aspartate from the D-loop and the catalytic glutamate (figure 1.2C) [18,24]. In this way, it assists with the correct positioning and orientation of

the water and the remaining catalytic residues [18]. Both Q-loop and H switch are located between the interface of the two domains. Exporters still possess an extra conserved region called the X-loop (TEVGERG), which is preceded by the ABC motif and is in proximity with the transmembrane domains[25]. The ABC motif (or signature motif) plays a key role in ATP binding and ATP-induced dimerization, because it allows ATP to bind the Walker A motif in the opposite dimer interface [26]. This motif is contained in the α-helices of the helical subdomains, which are in close contact with the transmembrane domains and are essential for the transmission of the hydrolysis signal [27]. This α-helical subdomain is highly conserved among ABC proteins and suffers significant conformational changes when ATP binds or is hydrolysed [22]. The ECF proteins also contain an exclusive sequence on the NBDs: the Q-helices. Q-helices are six-residue long and contain the motif xPD/ExQφ (where φ is a hydrophobic residue). These are located in the vicinity of the signature motif. The Q-helices contain a conserved Gln residue and are able to interact with the residues adjacent to the ABC sequence on the opposite NBD. Q-helices are hypothesized to be important to couple the energy of ATP hydrolysis with the conformational changes of the T subunits (the permanent transmembrane component), as well as the regulation of ATP hydrolysis in each cycle, but are not directly involved in ATP binding and hydrolysis[28]. The α-helical subdomain contains the ABC motif, also known as the signature motif, (LSGGQ) which is only found in ABC transporters.

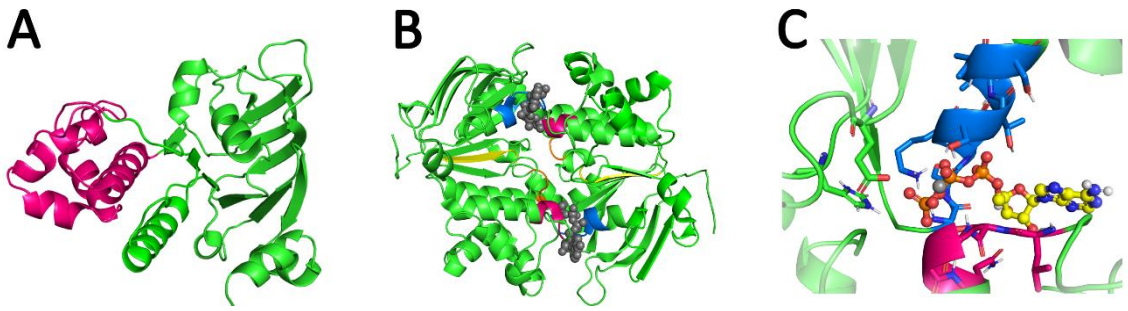


***Figure 1.2**-Structure of the nucleotide binding domains. A- Structure of a NBD monomer. The α-helical subdomain is portrayed in pink, while the catalytic subdomain is portrayed in green. B- Structure of a NBD dimer. The ABC motifs are represented in pink, the Walker A motifs are portrayed in blue, the Walker B in yellow and the nucleotides in grey. C- Close view of the active site, in which the ATP is portrayed in yellow and in spheres and sticks, the ABC motif in pink and the Walker A motif in blue. The catalytic glutamate is represented in green sticks as well as the H switch histidine*.

Both NBDs align in a head-to-tail fashion, in which an ATP molecule is sandwiched between the Walker A motif of one NBD and the ABC motif of the other. The NBD dimerization plays a key role in the mechanistic cycle of ABC transporters [4].

These domains can occur either as a homodimer or a heterodimer. Examples of transporters with heterodimeric NBDs are CFTR, TM287/288, SUR1 and the TAP transporter. In fact, most human eukaryotic transporters have heterodimeric NBDs. In some heterodimeric transporters, the Walker B glutamate, the H-switch histidine and the ABC motif show differences from their consensus sequence and these sites are referred as being degenerate[29]. The ATP binding pockets possessing altered sequences are not catalytically competent, being able to bind ATP, but not hydrolyse it. This will lead to an asymmetric NBD dimer, in which one of the pockets is closed with ATP bound, while the ATP bound in the other pocket is hydrolysed leading to pocket opening. CFTR and TM287/288 are examples of transporters with degenerate pockets, in which one ATP is bound during prolonged periods of time. In CFTR, ATP binding in this pocket has allosteric effects and regulates the transporter kinetics[30]. Structural evidences point that the communication between the two pockets is done via the D-loops [31–33]. Asymmetric NBD states can also happen in the context of asynchronous ATP hydrolysis, as it will be further discussed in section in the context of catalytic cycle mechanisms. P-glycoprotein and MJ0796 are examples of such transporters. However, asymmetric NBD dimers can arise in apparently symmetric NBD dimers such as the HlyB NBD dimer, in which the two ATP molecules display different binding modes in each pocket despite its symmetry[24]. On the maltose ($MalFGK_2E$) and histidine transporters ($HisQMP_2$), the non-equivalence of the NBDs was also reported due to different behaviour in the reaction of cysteine residues with either crosslinking or thiol-specific reagents [34,35]. Nevertheless, the fact that these transporters show different TMDs can also exert some influence in the asymmetric NBD behaviour, as well as the binding of SBPs, which are asymmetric by nature.

Finally, the nucleotide binding domains can also exhibit regulatory domains, such as the maltose transporter $MalFGK_2E$, in which the MalK subunits, that behave as NBDs have regulatory domains for interaction with unphosphorylated $EIIA^{Glc}$, which is a subunit of the phosphotransferase (PTS)-type glucose transporter, leading to the inhibition of maltose transport[36,37].

1.2.2- ATP hydrolysis in ABC transporters

1.2.2.1- Mechanism of ATP hydrolysis in ABC transporters

The mechanism of ATP hydrolysis is already characterized on several ABC transporters, such as HlyB, $MalK_2$ and BtuCD. Despite the high degree of similarity of the NBDs and the knowledge of which residues participate in hydrolysis, the precise mechanism for ATP hydrolysis in ABC transporters is still not consensual and it is not certain that all transporters share a common mechanism.

Two general mechanisms prevail for ATP hydrolysis in ABC transporters: the associative and dissociative mechanism. In the associative mechanism, a water molecule attacks the γ-phosphate of ATP leading to its release, while in the dissociative one, ATP captures a proton from the water, followed by the dissociation of the γ-phosphate (figure 1.3). Numerous studies point that mutants of the H-switch histidine and the Walker B glutamate abolish hydrolysis [24,38–40].

X-ray structures with non-hydrolysable ATP analogues in $MalFGK_2E$ suggested a general base mechanism in which the catalytic glutamate polarizes the catalytic water, which attacks the γ-phosphate, leading to the cleavage of the bond between the β and γ phosphates. This mechanism is associative and claims the existence of a trigonal bipyramidal intermediate. The H-switch histidine acts as a "linchpin" holding the participating groups in the correct conformations[41]. However, the mutation of glutamate to glutamine in P-glycoprotein, HlyB and GlcV does not abolish hydrolysis completely, leading to alternative models for explaining ATP hydrolysis[21,24,42,43] .

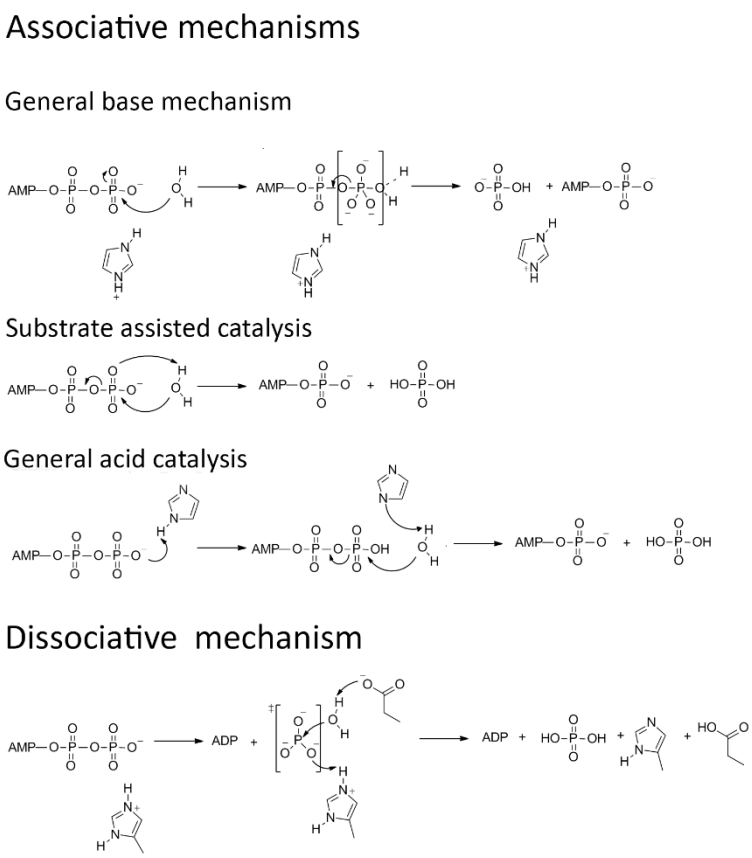


***Figure 1.3**- Scheme of some of the proposed mechanisms for hydrolysis in ABC transporters. Associative mechanisms include the general base mechanism proposed by Oldham* et al[41]*, the substrate assisted catalysis proposed by Zaitseva* et al[24]*, and the general acid catalysis proposed by Zhou and coworkers[45], based on QM-MM calculations. The dissociative mechanism portrayed is the one suggested by Huang* et al[47]*.*

A substrate assisted mechanism was suggested for HlyB in which a water proton is directly transferred to one of the γ-phosphate oxygens, and the OH- group attacks the γ-phosphate and there is release of the latter[24]. The H-switch histidine once again acts as a "linchpin". This mechanism is supported by HlyB x-ray structures, kinetic data in D_2O and ATPase activity measures in different pH values[24]. In addition, kinetic studies at several pH values on the human TAP complex also support this model in which H-switch histidine is directly involved in ATP bond breaking, while the glutamate may serve to properly orient the catalytic histidine[44].

Kinetic studies on bacteriophage T4 Rad50 (an ABC DNA repair protein) support a dissociative mechanism, in which an asymmetric transition state is formed, but without the aid of a catalytic base because kinetic data in water and D_2O revealed no differences between mutants of the catalytic glutamate and histidine.

Several theoretical studies using QM-MM methods, suggest similar mechanisms to the ones presented above. Zhou *et al* have studied the hydrolysis mechanism in HlyB NBDs[45] and using potential energy surfaces concluded that hydrolysis happens via a generalized acid catalysis (GAC) mechanism. In this mechanism, the H-loop histidine mediates proton transfers, by firstly donating a proton to ATP and afterwards, in a concerted step, it captures another from the catalytic water, and the resulting OH^- attacks the γ-phosphate, leading to its dissociation from ATP. This mechanism has raised some criticism[46] because it implicates the existence of a doubly negative histidine, implying a high energy barrier (22 kcal/mol) and it does not involve the catalytic glutamate present in Walker B motif, whose mutation is known to impair hydrolysis[38,39,41] Other two theoretical studies approached hydrolysis in MalK. Huang *et al*[47] have concluded that hydrolysis happens via a dissociative mechanism in which a trigonal bipyramidal transition state when the bond Pγ-Os is cleaved, followed by a concerted step in which the catalytic glutamate abstracts a proton from water and the resulting OH^- attacks the intermediate (figure 1.3). This mechanism is in agreement with kinetic data with D_2O[48]. Hsu *et al*[49] have come up with a similar mechanism in which a trigonal bipyramidal transition state is also formed, with a concerted action of the water molecule and the catalytic glutamate. Interestingly, the authors reported the formation of a neutral lysine generated by the transfer of a proton to the phosphate intermediate, when the catalytic water attacks the intermediate. Finally, Prieß *et al*[46] have simulated hydrolysis on the BtuCD importer, finding that hydrolysis happens via a dissociative mechanism in which a transition state intermediate is firstly formed, and in a second step, the catalytic glutamate also activates a water molecule yielding $H_2PO_4^-$. Remarkably, the authors estimated a ΔG of +1.8 kcal/mol for the overall reaction, suggesting that ATP hydrolysis is not the power stroke of BtuCD, but rather nucleotide binding.

Overall, most studies seem to point to a dissociative mechanism, in which the catalytic glutamate acts as a base, and the H-switch histidine contributes for keeping the correct orientations of the intervening groups. The above data, namely the x-ray structures and QM-MM calculations also suggest that only one water is involved in catalysis, in contrast with other NTPases.

1.2.2.2-Stoichiometry of ATP hydrolysis

Considering the NBD dimer structure and the existence of two ATP binding sites, one can question whether how many ATP are hydrolysed by transport cycle. In the case where one of the ATP binding pockets possesses degenerate sequences, it is possible that hydrolysis of one ATP suffices to complete a transport cycle. Such an example is the TM287/288 transporter.

When both nucleotide binding sites are catalytically competent, it becomes harder to measure the stoichiometry of ATP hydrolysis. This happens because the transporters display high ATPase activity even in the absence of substrate and it is difficult to isolate ATP and substrate on opposite sides of the membrane. In this way, there are large discrepancies on the values obtained, ranging from 2 to 17 ATP hydrolysed per transport cycle. For importer OpuA, a ratio of 2 ATP hydrolysed per substrate was obtained [50], while values between 0.5, 2 and 17 were found for the maltose importer[51–53]. Muir *et al*[51] observed in the maltose importer that stoichiometry ratios were substrate dependent, ranging from 0.5 ATP per lactose transported to 1-1.2 per maltose/maltodextrin. Other importer systems such as BtuCD showed a stoichiometry of 50 ATP hydrolysed per substrate [54], but this value was contested raising the hypothesis of futile hydrolysis[55]. In exporters, the peptide transporter Mdl1p can hydrolyse two ATP molecules, but at ATP limiting conditions the NBD dimer dissociates before the hydrolysis of the second ATP[56]. Other transporters such as P-glycoprotein also hydrolyse two ATP molecules per transport cycle, although not concomitantly [57].

1.2.3- Beyond ATP hydrolysis – Adenylate kinase activity

Some ABC transporters were reported to display adenylate kinase (AK) activity. In the AK reaction ATP and AMP react and originate two ADP molecules (ATP + AMP → 2ADP). Examples are MsbA and CFTR, where adenylate kinase activity was observed[58,59]. In both transporters the reverse adenylate kinase reaction was enough to power the normal function of the transporter. In the case of CFTR, channel gating was observed in the presence of mM concentrations of AMP and intermediate ATP concentrations [58]. On MsbA, transport of lipid A and of the dye Hoechst-33342 was observed in the presence of ADP[59]. AK activity was also observed in the DNA-repair ABC protein Rad50 [60] and SMC protein- structural maintenance of chromosome protein[61].

The AK reaction is characterized by the direct transfer of a phosphate from one nucleotide to another, in contrast with the ATPase reaction in which a phosphate is released to the cytoplasm. The molecular details of the adenylate kinase reaction have been characterized using mutagenesis assays, x-ray structures, solid-state NMR and molecular dynamics. Studies with the isolated NBD2 of CFTR showed that existed two nucleotide binding sites, one of them being the canonical ATP-binding site and the other being the AMP-binding site [58,62]. Additionally, mutation of N1303 halted AK gating activity, but not ATPase gating activity[62]. Mutagenesis studies of CFTR have found that Walker A is essential for nucleotide binding for the AK reaction and that the presence of AMP is dependent of ATP[63].

The crystal structure of SMC from *Pyrococcus furiosus* complexed with the adenylate kinase inhibitor Ap5A (that emulates two nucleotides) showed that the Q-loop is involved with

nucleotide binding at the AMP-binding site, and the Walker A motif is important to mediate binding of the phosphate groups[61]. On MsbA, solid-state NMR and mutagenesis studies confirmed the potential binding site of AK-dedicated nucleotides near the Q-loop[59].

Regarding the physiological role of the adenylate kinase reaction, it was suggested that the reverse AK reaction is a way to circumvent limiting ATP or high ADP concentrations[60]. In this way, although ATP hydrolysis is the primary reaction, it is possible to exist a coupled ATPase-AK equilibrium which is shifted when necessary. In the case of MsbA it may allow bacteria to endure environmental stress, since the inactivity or absence of MsbA leads to lipid A accumulation and toxicity resulting cell death[64]. In Rad50, a mutation that reduced AK activity, but not ATPase activity in yeast, led to impaired meiosis and telomere maintenance, a phenotype similar to a null Rad50 [60]. In the CFTR channel, the role of the reverse AK reaction is still not clear[62].

1.2.4-The transmembrane domains

In contrast with the nucleotide binding domains, the transmembrane domains (TMDs) display little sequence conservation. This may reflect the structural diversity of the substrates transported, because the transmembrane domains determine substrate specificity. Nonetheless, transporters within a class share a similar domain topology[17]. All ABC transporters have two transmembrane domains but show different protein topologies. Both domains can be identical or not. This may be the result of genetic events such as gene duplication or gene transfer[15]. ABC exporters most often show two sets of six α-helices in two domains, making a total of twelve helices[18]. In type II exporters (ABCA and ABCG families), both TMDs display different folding and the two domains do not intersect. Each domain of type I importers has between 5 to 8 helices, making a total of 10 to 14 helices per transporter. In type II importers each TMD comprises 10 helices each, making a total of 20 transmembrane helices [2] (figure 1.4). On ECF transporters, one of the TMDs, the T-component (or EcfT subunit), is associated with the NBDs, while the other, the S-component (or EcfS subunit) diffuses freely across the membrane binding to substrates. In the available structures, the EcfT has 5 transmembrane helices, while EcfS has a minimum of 6 α-helices[18]. Additionally, three helices are projected into the cytoplasm and two of them are longer, crossing in a "X" shape. The longer ones contain the conserved sequence XRX, which interacts directly with the α-coupling helices participating in signal transmission from hydrolysis in the NBDs to the remaining protein [10]. Nonetheless, recent cryo-EM structures of bacterial importers YbtPQ and Mtb, responsible for the intake of siderophores and B_{12} vitamin, respectively, show structures typical of exporter proteins[65,66]. This suggests that the structural classification of ABC proteins might be more flexible than initially thought.

The TMDs also possess α-helices that interface with the nucleotide binding domains, known as coupling helices. The coupling helices stand parallel to the membrane and interact with a groove in between the subdomains of each NBD. Despite their widespread existence in all ABC transporters, their sequence is not conserved [67].

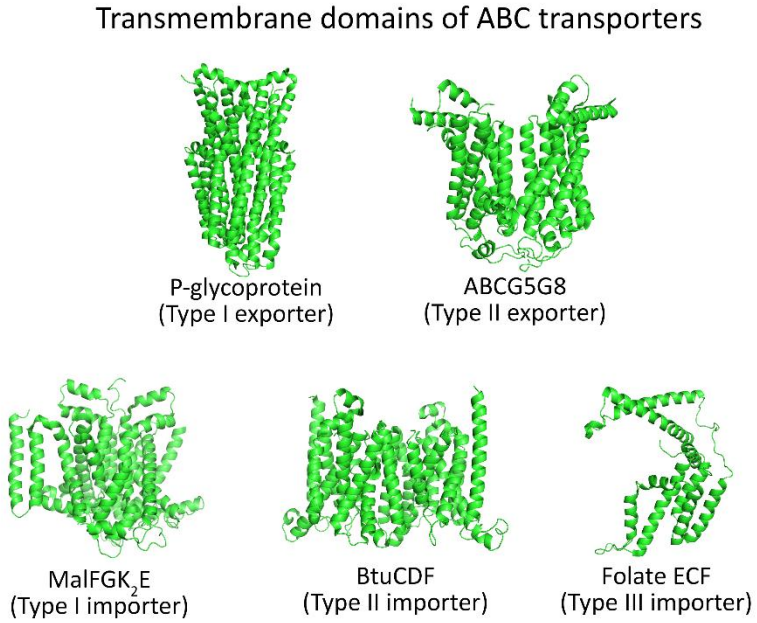


***Figure 1.4**- Structures of transmembrane domains in ABC transporters. A representative member of each class is portrayed. P-glycoprotein, PBD code: 3B60. ABCG5G8, PBD code: 5DO7. MalFGK$_2$E, PBD code: 3RLF. BtuCDF, PDB code: 2QI9. Folate ECF, PDB code: 4RFS.*

Regarding substrate specificity, while some transporters are highly specific towards a single substrate, such as BhuT (heme transporter) and BtuCD (vitamin B_{12} transporter), other transporters exhibit the ability to bind to more than one substrate – polyspecificity. This can be observed in transporters such as P-glycoprotein, the aminoacid transporter ArtI-Art(QN)$_2$, and MsbA, to state a few examples. Many exporters show a high degree of polyspecificity. Polyspecificity becomes especially relevant in the context of multidrug resistance, by allowing binding of different xenobiotics leading to their efflux. Exporters like MDR1, P-glycoprotein, MsbA and ABCG2 have a transmembrane cavity with independent hydrophilic and hydrophobic regions, that allow binding of amphipathic, hydrophilic and hydrophobic molecules. In the case of P-glycoprotein, it is even suggested that the rearrangement of the TM helices may lead to changes in the surface topology of the transmembrane cavity leading to polyspecificity [68]. The transmembrane subunits of importers can also bind to a range of substrates: the Art(QN)$_2$ importer was crystallized with either arginine and histidine and it is hypothesized that it can transport any aminoacid with a positive side chain, because the transmembrane channel is lined

with negative charged residues. It is also known that the maltose importer $MalFGK_2E$ can also bind to several malto-oligosaccarides of different sizes linked through α-1, 4 glycosidic bonds, ranging from maltose with two glycosyl units, to maltoheptatose with seven glycosyl units. However, this system is highly selective for derivatives of maltose, since closely related sugars, such as glucose, lactose and isomaltose are not recognized[69]. Furthermore, type II importers cannot differentiate between ligand-bound and unliganded the substrate-binding protein, while type I importers are able to distinguish liganded and unliganded substrate-binding protein, displaying different binding affinities[70].

1.2.5-Substrate binding proteins

Substrate binding proteins (SBPs) are soluble proteins that assist ABC importers by collecting and delivering molecules in the cytoplasm to the membrane complexes. Remarkably, SBPs are not exclusive of ABC transporters, as their participation was also found in other membrane proteins, such as the tripartite ATP-independent periplasmic (TRAP) transporters[71], peptide receptors, G-protein coupled receptors end even ligand-gated ion channels[72,73]. Additionally, substrate binding domains can also be found in other proteins not related with transport, such as prokaryotic DNA-binding proteins responsible for gene regulation. SBPs can be soluble in the cytoplasm but they can also be attached to the membrane via a lipid anchor or they can be fused to the transmembrane domains. The latter is more frequent in Gram-positive bacteria, while Gram-negative bacteria tend to show soluble SBPs[74]. Generally, one or two SBPs are fused to the membrane complex, but complexes with two and three SBPs fused in tandem to each TMD, resulting four to six total SBPs have also been observed. Fused SBPs occur in both homodimeric and heterodimeric TMDs and allow increased transport efficiency, due to the increase of substrate binding sites near the transporter[11]. An alternative uptake mechanism was observed in the MetNI importer, in which the soluble SBP is able to perform substrate uptake when bound to the complex, via a channel with direct access from the SBP to the transmembrane domains, with higher efficiency than the regular mechanism[75].

Type I SBPs are present in large excess relatively to the transmembrane complex, while type II SBPs appear to exist in a stoichiometric ratio[11]. SBPs contain a binding pocket between the N and C termini lobes of the protein. Both lobes are connected by a linker α-helix or a hinge domain. The N and C terminal are composed by a central β-sheet core with five β-sheets, flanked by α-helices. Crystal structures showed that SBPs alternate between the open unliganded conformation, in which the two lobes are separated, making the binding site exposed to the solvent, the closed liganded conformation, in which the substrate lies in the middle of the two lobes, the closed unliganded conformation and the open liganded one. The last two

conformations exist mainly when SBP is bound to the transporter. The closed-liganded conformation is the one that binds to the transmembrane complex for delivering the ligand. Two models attempt to explain the mechanism of ligand binding in SBPs: the induced fit model and the conformational selection model[76]. In the induced fit model, the protein is initially in the open unliganded conformation, and ligand binding triggers the closure of the protein, leading to the formation of the closed liganded form. On the other side, in the conformational selection model, the protein is able to sample closed or partially closed states in the absence of ligand[76]. Ligand binding stabilizes these states and drives the conformational equilibrium to an increase of the closed liganded form[76]. Studies of the maltose binding protein MalE and of the substrate binding domains of GlnPQ showed that SBPs transit intrinsically from open to closed conformations and that ligand binding stimulates closure, combining the two aforementioned mechanisms[76–79].On GlnPQ it was shown that the lifetime of the closed liganded form is the same as the closed unliganded form, when high affinity ligands are bound, being possible to say that the ligand binding does not stabilize the closed form[79].

Regarding the structural characteristics of SBPs and substrate binding, it is known that type I SBPs are much more flexible than type II SBPs, exhibiting large amplitude rigid-body motions up to 60 Å, while on type II SBPs movements are much more limited (figure 1.5), due to a rigid α-helix that serves as a hinge. In the presence of substrate, the lobes close in a Venus flytrap fashion[80,81]. A classification of the SBPs was done based on a pairwise alignment of sequences. This alignment allowed to obtain six clusters: A, B, C, D, E and F[82]. Cluster A SBPs are exclusive of type III ABC importers and are involved on the import of metal ions. Cluster B contains SBPs related with type I importers and that bind peptides, carbohydrates, and aminoacids. Cluster C contains SBPs that belong to type II importer systems but have an extra domain that often allows to bind larger substrates[82]. These SBPs also bind to various peptides, aminoacids, nickel ions and cellobiose. Cluster D also comprehends SBPs exclusive of type II importers with very short hinge regions, ranging from 4 to 5 aminoacid residues. Cluster E contains SBPs that bind TRAP transporters only[82]. These SBPs have an additional β-sheet as well as a long helix connecting both domains. Some known substrates include lactate and sialic acid. Cluster F SBPs contains type II SBPs only, which have a hinge segment that consists in two segments, with 8-10 residues each that allow extra flexibility. Known substrates include aminoacids, compatible solutes and ions, such as nitrate and carbonate [82].

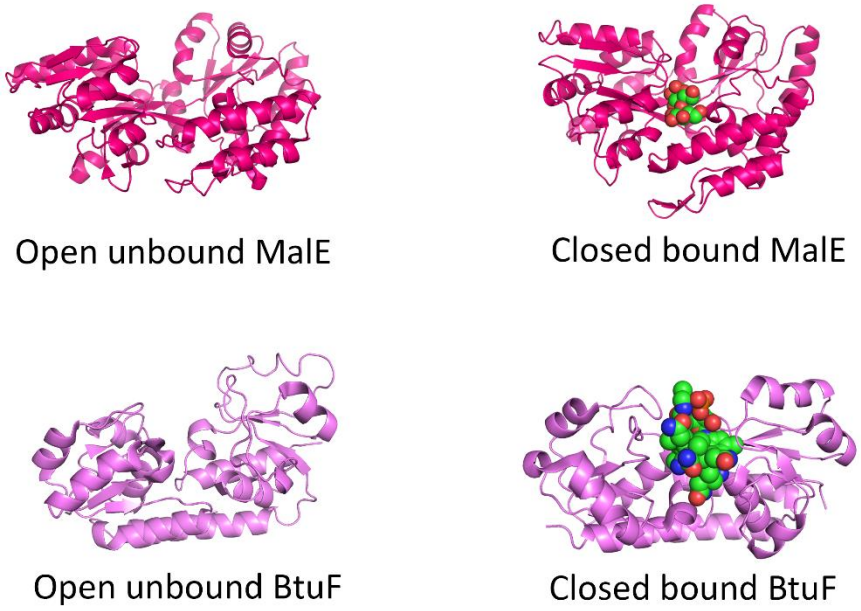


***Figure 1.5**-Structures of type I and II substrate binding proteins. MalE is a type I SBP and it is portrayed in pink, in the upper portion of this figure. BtuF is a type II SBP and it is portrayed in purple. The substrates are depicted in spheres, coloured according to the atom type. Maltose is bound to MalE, whereas B_{12} vitamin is bound to BtuF.*

1.2.6-The S-components of ECF transporters

The S components of ECF transporters are transmembrane mobile components that scavenge substrates for ECF transporters (type III importers). Despite binding these substrates, S-components do not translocate them across the membrane[83]. There are over 21 different families of S-component proteins[84], which are able to bind to a wide range of molecules, mainly vitamins and nutrients. The known structures share a common core of six hydrophobic helices, with connecting loops that trespass the membrane leaflets and are projected to solution, as shown in figure 1.6. The transmembrane helix 1 contains a conserved AxxxA motif and TM helices 1 to 3 are structurally more conserved, while TM helices 4 to 6 are structurally different and are a large part of the binding site. The AxxxA motif participates in complex formation and substrate transport [11,85]. Remarkably, the S protein NikM2 from a nickel ECF transporter has an additional N-terminal α-helix that inserts into the binding site, aiding to the nickel coordination[86]. The substrate binding site is located in a cleft formed by residues of the helical C-terminal alongside with some residues of the connecting loops[87]. S components bind substrates with high affinity, with K_d in the low nanomolar range[11]. EPR and MD simulation data show that there are not large conformational changes triggered upon substrate binding, but rather loop movements that act as a lid[83]. MD simulations along with experimental data, show

that the lipid composition, along with membrane deformations ,play a key role in binding of the S-component to the remaining complex[88].

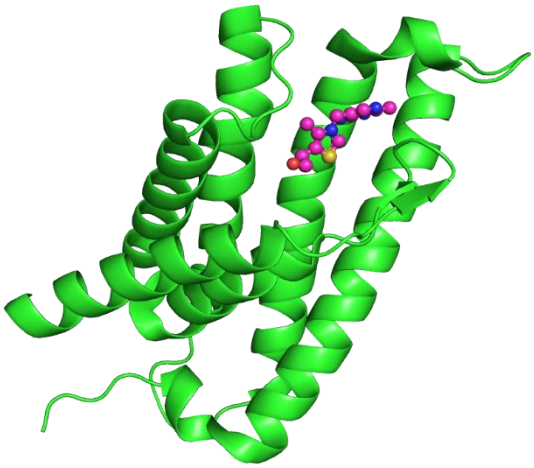


***Figure 1.6**-Structure of the thiamin specific S-component ThiT with thiamin bound. Thiamin is represented in spheres.*

1.3- Mechanisms for function of ABC transporters

The first general mechanism proposed for substrate transport across membranes was the "alternating access" model[89]. In this model, it is assumed that the protein contains a cavity in which the binding of a molecule is possible and that the transporter alternates between an inward facing conformation, in which the cavity is exposed to intracellular side of the membrane, and an outward facing conformation in which the cavity is exposed to the extracellular medium. It is also expected that the affinity of the transported allocrite is different in both conformations. This model was firstly suggested for the Na^+/K^+ pump in 1966[89]. In ABC transporters, the transmembrane domains interconvert between the inward and outward facing conformations, allowing substrate binding and transport. On the nucleotide binding domains there is ATP binding, hydrolysis and subsequent ADP and phosphate release. Most early models for the transport cycle were based on data about the nucleotide-binding domains. Still, it was possible to find a plethora of different models, based on biochemical, spectroscopic data, x-ray crystallography and molecular dynamics simulations. More recently, cryo-EM has also joined these tools to provide a clearer picture of the underlying mechanisms. An extensive description of the existing models for isolated NBD function will be done in the subsequent section, followed by the current models for the coupling of the substrate transport with hydrolysis on full-length ABC transporters.

1.3.1-Models for NBD function that assume complete separation of the nucleotide binding domains

Tweezers model

The tweezers-like model[90] was suggested based on the x-ray structures of the maltose $MalFGK_2E$ importer, and most data supporting this model is related with this importer. The first structures of the MalK NBD dimer showed that it could adopt several conformations: a fully open one with both domains fully separated and with no nucleotides, a closed conformation with nucleotides bound and a semi-open open conformation with no nucleotides bound. With this information a model was suggested in which the MalK dimer acted like tweezers controlling the opening and closing of the transmembrane domains depending of the presence of nucleotides[90] (figure 1.7).

Processive clamp model

The processive clamp model was first suggested by Janas *et al*[56], based on biochemical studies of the NBD dimer of MDR1p. In this model, ATP binding the NBD dimer induces dimerization, similarly to the ATP switch model, but hydrolysis of both ATP molecules occurs in an alternated way. This leads to the genesis of intermediate states with both ADP and ATP bound. According to this model, hydrolysis of one of the ATP can be followed by phosphate release or hydrolysis of the other ATP molecule (figure 1.7). After release of both phosphates from the active site, the dimer dissociates, and ADP is released. It was hypothesized that the transmembrane domains remain in the outward-facing conformation[91].

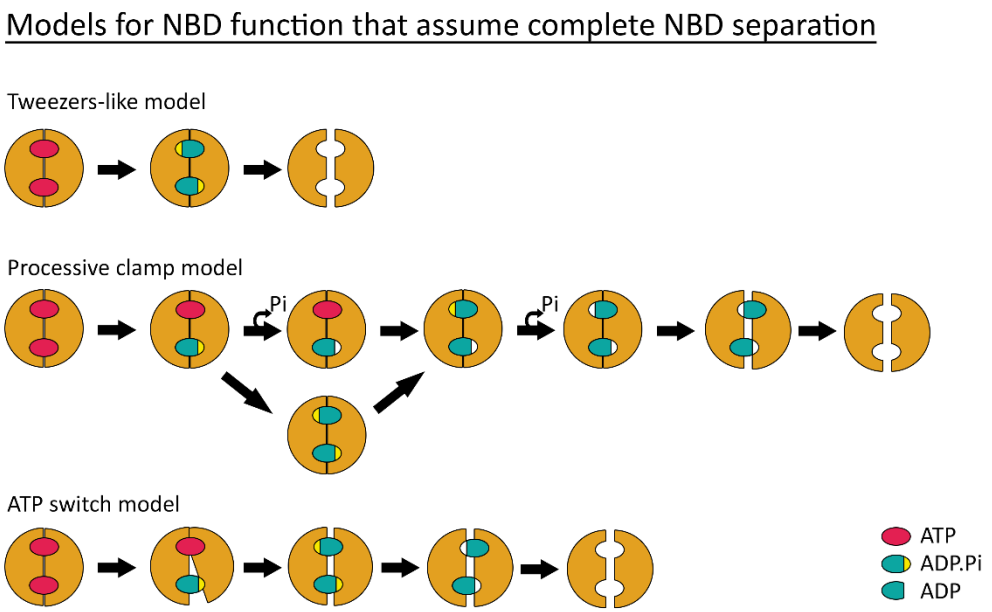


***Figure 1.7**- Scheme of the models for NBD function that assume complete NBD separation. The nucleotide binding domains are represented in orange. ATP is portrayed in pink, ADP in blue and phosphate (Pi) in yellow. The represented models are cyclic, when the last state is achieved, the cycle restarts over again.*

ATP switch model

The ATP switch model was proposed by Higgins *et al*[92] and it states that the NBD dimer switches between two conformations that trigger the transport "on" and "off"[92]. The conformation in which ATP is bound to both sites has the NBDs in close contact, also called a closed conformation. In contrast, when ATP hydrolysis occurs and there is the release of ADP and Pi, the NBD dimer adopts an open conformation, where the NBDs are separated from each other. In the closed NBD conformation, the TMDs are closed and the binding site is inaccessible, while in the open NBD conformation the TMDs are separated exposing the substrate binding site (figure 1.7).

1.3.2-Models for NBD function that require contact between the NBDs throughout the transport cycle

Alternating sites model

The alternating sites model was firstly suggested by Senior *et al* for P-glycoprotein, based on biochemical data[93]. In this model, there is always one ATP bound to the NBD dimer, implying that the active site is closed. This happens in the following way: one ATP is bound in the NBD2 active site and another ATP binds to the NBD1 site, followed by hydrolysis of the NBD2 ATP and

phosphate release. This causes the transporter to become outward-facing and release the substrate. The release of the ADP from NBD2, causes the transporter to return to the inward-facing conformation, and the whole cycle repeats again, this time with NBD1 ATP being hydrolysed first (figure 1.8). According to this model, the active site changes conformation depending on the nucleotide bound.

Constant contact model

The constant contact model was suggested by Jones *et al* based on molecular dynamics simulations of the MJ0796 NBD dimer[94]. This model not only contemplates the states with different nucleotides bound as well as an empty state, but also sub-states where the active site can either be occluded or open, allowing nucleotide exchange (figure 1.8). In the case of the empty site, it can alternate between a state of high and low affinity. Overall, the full cycle starts with an ATP bound to the one of the sites in a closed sub-state, while the other site is a low-affinity open conformation. After ATP hydrolysis, the empty site converts to a high affinity sub-state and it is ready to bind ATP. Afterwards, while the second site has ATP bound, but is in an

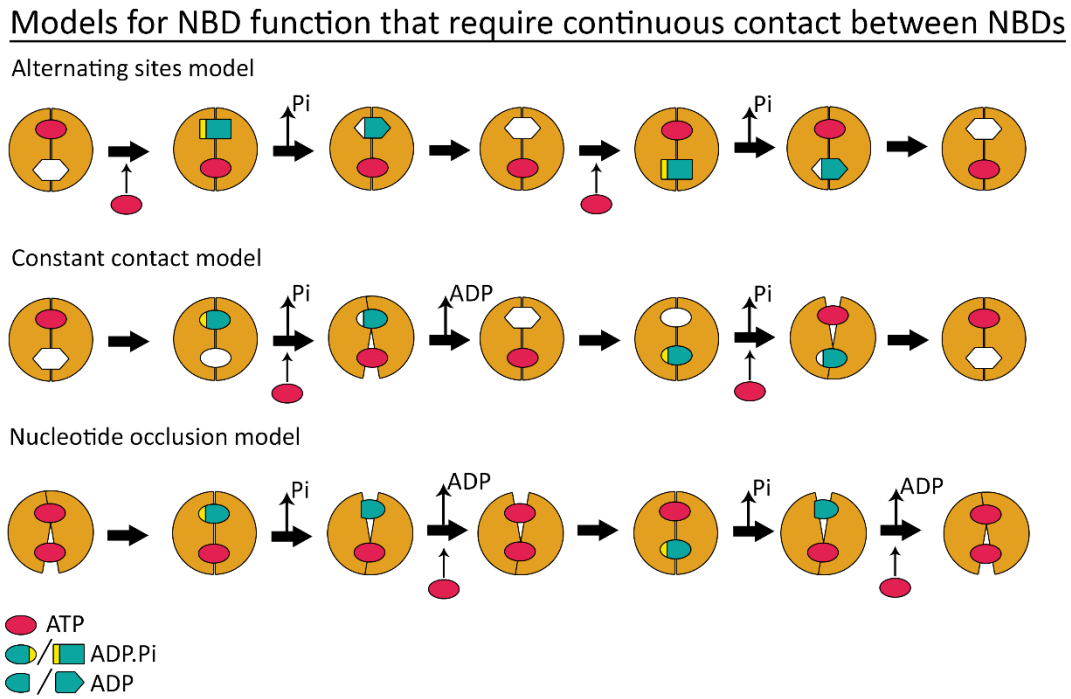


***Figure 1.8**- Scheme of the models that assume continuous contact between NBDs. In the alternating sites model, the hexagonal and rectangular shapes of the nucleotides mean different conformations of the active site, as suggested by the authors. In the context of the constant contact model, the hexagonal shape means a low affinity binding site, while the round one means a high affinity site. The represented models are cyclic, when the last state is achieved, the cycle restarts over again.*

open sub-state, there is phosphate release in the other one (figure 1.8). This is followed by closure of the ATP-bound site and transition of the empty one to a low-affinity state, and the cycle repeats itself all over again.

Nucleotide occlusion

This model was originally proposed by Sauna *et al* to explain P-glycoprotein function[95]. It contemplates the existence of an asymmetric state with two ATP molecules bound (figure 1.8). The initial state has two ATP molecules bound, but only one of the pockets is occluded, being the latter the site where hydrolysis is going to take place. After hydrolysis, there is closure of the ATP-bound site and release of the hydrolysis products ADP and Pi. This is followed by the binding of another ATP at the empty site, and subsequent hydrolysis of the other one, leading to the repetition of the cycle (figure 1.8).

1.3.3-Mechanisms of coupling of hydrolysis with substrate transport in ABC exporters

Molecular structures and biochemical data of the full transporters allowed to support and complement the aforementioned models, as well as deduce mechanisms to the coupling between hydrolysis with substrate transport.

The first full-length X-ray structures of ABC exporters were the Sav1866 from *Staphylococcus aureus*[27] and MsbA from *Escherichia coli*, *Vibrio cholerae*, and *Salmonella typhimurium*[96]. In the spirit of the alternating access model, the Sav1866 structure was labelled as being "outward-facing" as well as the *S. typhimurium* MsbA structure. These structures show nucleotides bound to their NBDs. In these structures, the transmembrane domains are crossed in a "V"-shape, forming a cavity that faces the extracellular medium, with the nucleotide binding domains in a closed conformation. On the other hand, in the *E.coli* and *V. cholerae* inward-facing MsbA structures, the transmembrane domains show an inverted "V" shape, disclosing a large transmembrane cavity with the NBDs far apart. Other exporter structures have followed up, such as the eukaryotic P-glycoprotein. The inward-facing structures of MsbA and P-glycoprotein comprised a large cavity, in which NBDs are separated by approximately 30 Å[97], raising even some criticism on whether such structure is physiologically possible and not an artefact due to the purification process. These concerns take in consideration the high intracellular ATP concentrations, that may increase the probability of ATP binding to the NBDs, leading to shorter NBD separations. Other puzzling structures include ABCB10, a mitochondrial transporter, which was also crystallized with AMP-PNP (an ATP analogue), but did not show a closed NBD dimer[98] ,and the heterodimeric transporter TM287/288 in which the degenerate binding pocket shows

a significantly larger intermotif distance than the conserved pocket[29], even in the absence of nucleotide in both pockets.

These structures are assumed to represent intermediates of the catalytic cycle. The similarity between exporter structures may lead to the temptation of generalizing a common NBD-TMD coupling mechanism for all exporters, based on an alternating access model. The switch model can also be extrapolated to a full transporter and is in agreement to with most of these structures[11], mainly the MsbA and Sav1866 ones. In this case, translocation begins with substrate binding to the TMDs, followed by ATP binding to the NBDs inducing a series of conformational changes that culminate in the closure of the NBDs and the egress of the substrate. This conversion is followed by ATP hydrolysis, which resets the transporter to the initial state by destabilizing the NBD dimer and inducing the reverse transition. Amid this step, there is ADP and Pi release. Nonetheless, the TM287/288 seems to agree with a model that implicates contact between the NBDs, throughout the cycle[29].

However, the models presented above, along with more biochemical and structural evidences, greatly reduce the probability of existing a unifying mechanism for all ABC transporters (figure 1.9). Nonetheless, EPR data on MsbA confirmed the hypothesis of an alternating access mechanism in this transporter, in which the ATP-bound dimer is associated to the outward-facing state and that ATP binding is the power stroke that triggers the interconversion between states [99–101]. These data also confirmed the existence of large amplitude motions in the order of 10-20 Å, confirming the x-ray data about the existence of wide-open apo structures. Extensive MD simulations of apo MsbA[102] also revealed that the most likely transition pathway starts from the outward facing conformation. NBD dissociation facilitates closure of the periplasmic portion of TMDs, which consequently leads to twisting of the NBDs. These events precede the closure of periplasmic end of the TMDs, culminating in the fully open inward-facing state[102].

EPR studies on P-glycoprotein, made by Verhalen *et al*[57] , allowed to confirm and incorporate the elements of nucleotide occlusion and the alternating sites models[57] (figure 1.9). In this case the conformational cycle proceeds in the following way: ATP binding causes the approximation of the NBDs in the inward-facing conformation. Afterwards, substrate binding leads to full NBD dimerization. Nonetheless, an asymmetric NBD dimer conformation was observed, confirming the observations made by Sauna *et al*[95] on P-glycoprotein NBDs that led to nucleotide occlusion model. One of the pockets is fully occluded, while the other is not. ATP hydrolysis happens first on the fully occluded pocket, leading to the occlusion of the second one. Hydrolysis of the first ATP leads to the closure of the intracellular side of the TMDs, while hydrolysis of the second one is the driving force to the opening of the extracellular side of the TMDs, leading to substrate release.

Mishra *et al*[103] have also used EPR assays with the goal of assessing mechanistic difference between homodimeric and heterodimeric transporters, namely MsbA and BmrCD (figure 1.9), respectively [103]. They concluded that there are striking mechanistic differences depending on the structure of the transporter.

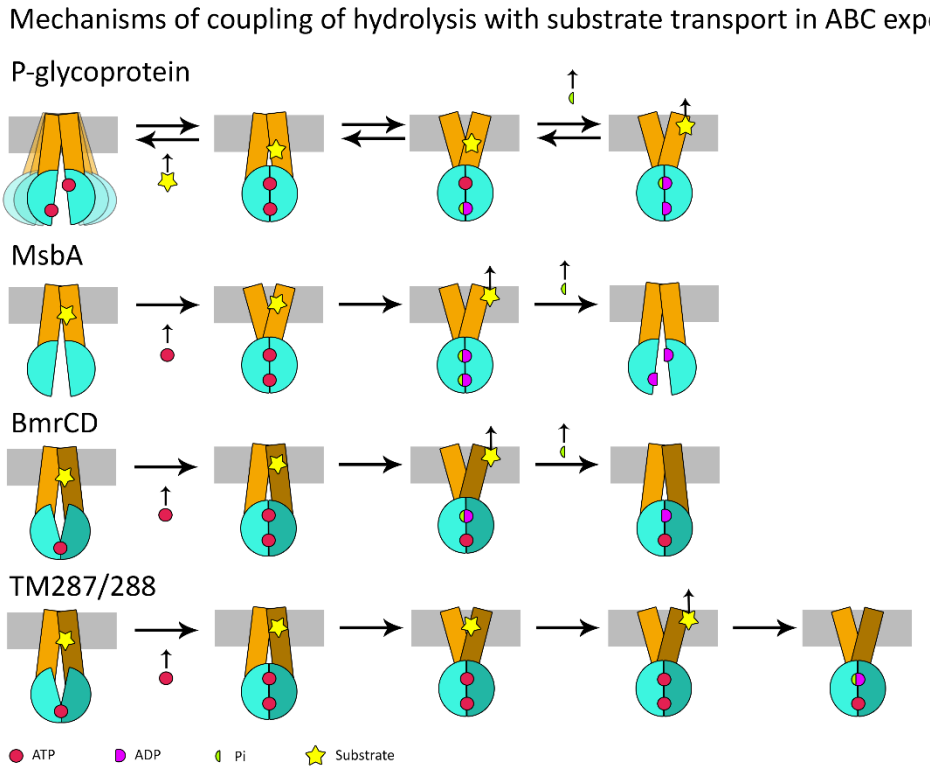


Figure 1.9 *– Scheme depicting the mechanisms of coupling of hydrolysis with substrate transport in some ABC transporters. The represented proteins are P-glycoprotein and MsbA, BmrCD and TM287/288. MsbA is an homodimer. BmrCD and TM287/288 are heterodimers. The NBDs are portrayed in blue, while the TMDs are in orange. In the heterodimer proteins, the different subunits are coloured in different shades. The substrate is portrayed as a yellow star, ATP in red, ADP in purple, phosphate (Pi) in green. In the P-glycoprotein mechanism, the transitions are represented as equilibrium. In the first state, the protein oscillates between wide open and open states. The represented models are cyclic, when the last state is achieved, the cycle restarts over again.*

In heterodimeric transporters, there is an inward-facing pre-hydrolysis state in which ATP is bound to both NBDs, while the substrate can be already bound to the transmembrane domains. ATP hydrolysis in the consensus site triggers the transition to the outward-facing state and consequent substrate release. After the release of hydrolysis products, the transporter returns to its initial inward-facing state and a whole new cycle begins. Transport in homodimeric transporters also begins with an inward-facing state of the transporter, with the substrate already bound to the transporter, but without any nucleotides bound, creating a fully open NBD

dimer[103]. The conformational cycle for MsbA resembles the switch model previously mentioned (figure 1.7). TM287/288 is a heterodimeric transporter with only one competent catalytic site (figure 1.9). ATP binding triggers a large conformational change that leads to the NBD closed state and the outward-facing state. ATP binding to the consensus site leads to substrate exit. After ATP hydrolysis and phosphate and ADP release, the importer returns to its initial state. This study showed than transporter structure influences the transport cycle, and that the diversity of NBD mechanisms (described above) is intrinsically related with this structural diversity. This study showed that heterodimeric transporters show a less efficient coupling between the NBDs and TMDs - a "conformational leak", that enables NBD movement without transmitting it to the TMDs[103]. A posterior EPR study on TM287/288 confirmed the mechanism above described for this transporter[104]. Cryo-EM and biochemical data on the heterodimeric transporter TmrAB revealed a similar mechanism[105]. ATP binding leads to NBD dimerization and to the conversion from the inward-facing conformation to the outward-facing one, allowing substrate release[105,106]. After substrate release, ATP is hydrolysed, in the single catalytic pocket. In the particular case of TmrAB, the transporter samples a conformational subspace of the outward-facing conformation, switching between outward-facing open and occluded conformations, with different TMDs opening degrees. In addition, phosphate release results in an asymmetric NBD conformation, which leads to the opening of an intracellular gate in the TMDs ultimately resulting in the inward-facing conformation[105]. In this way, ATP hydrolysis and phosphate release are key to the directionality of the transport reaction. Interestingly, a wide inward-open conformation was observed similarly to P-glycoprotein and Sav1866, even in the presence of one ATP molecule. Additionally, the inward-facing conformation can alternate between a wider form and a narrower one, varying the size of the transmembrane cavity, eventually accommodating substrates of different sizes.

1.3.4-The unique mechanism of CFTR

The CFTR channel is an ABC transporter that plays a key role in regulating the fluidity of secretions in epithelial tissues in organs such as lungs and intestines. Its malfunction in humans leads to cystic fibrosis. CFTR transports mainly Cl^- ions, but can transport HCO_3^- as well. CFTR is a single-chain protein with 1480 residues. In addition to the TMDs and NBDs, CFTR also possesses a regulatory domain. If this domain is not phosphorylated, it will be located between the two TMDs, blocking ion transport. Upon phosphorylation, this domain suffers a conformational change that leads to its diffusion away from the TMDs [107].

The NBDs are structurally different, since one of them possesses residue substitutions in the conserved sequences, such as in the Walker A, and ABC motifs. This will cause one of the ATP

binding pockets to be non-hydrolytic, just like a heterodimeric transporter. Despite being an ion transporter CFTR still relies on ATP hydrolysis to carry its function, in contrast with other ion channels and transporters that do not rely on ATP hydrolysis. At this moment there are two models to explain CFTR function, the strict coupling model and the energetic coupling mechanism (figure 1.10)[108]. The strict coupling model defends that hydrolysis and gating are tightly coupled [109,110]. In this model, CFTR alternates between three states: the closed state, the open O_1 state and open state O_2. The open O_1 and O_2 states are distinguishable in electrochemical measurements, with O_1 showing a lower current than O_2. The most frequent transitions are found from the closed state to the O_1, followed to the O_2 and back to the closed state. The reverse pathway is highly unlikely indicating the need for energy input in the transition from O_1 to O_2, in the form of ATP hydrolysis. In the O_1 state, both ATP are bound to both pockets. In the closed state, the NBD dimer is open. This implicitly leads to the conclusion that one ATP is hydrolysed in each cycle. In contrast, the energetic coupling mechanism[111,112] defends that the transport cycle is controlled by allosteric coupling between the NBDs and the TMDs. This model is based on assays with CFTR mutants that showed the existence of a second post-hydrolysis state in which the channel is open, but the hydrolytic site is already empty (O_2) [112]. Therefore, there is an allosteric coupling between NBDs and TMDs in which conformational changes in the NBDs facilitate conformational changes on TMDs. Interestingly, mutations, such as R352C, that alter local electrostatic potential in the TMDs lead to repeated cycles that alternate between the O_1 and O_2 states[113]. On the other hand, Yu *et al*[114] showed that the pathological R117H mutation leads to a closed state with dimerized NBDs (C_2AD state in figure 1.10)[114]. These observations support this latter model.

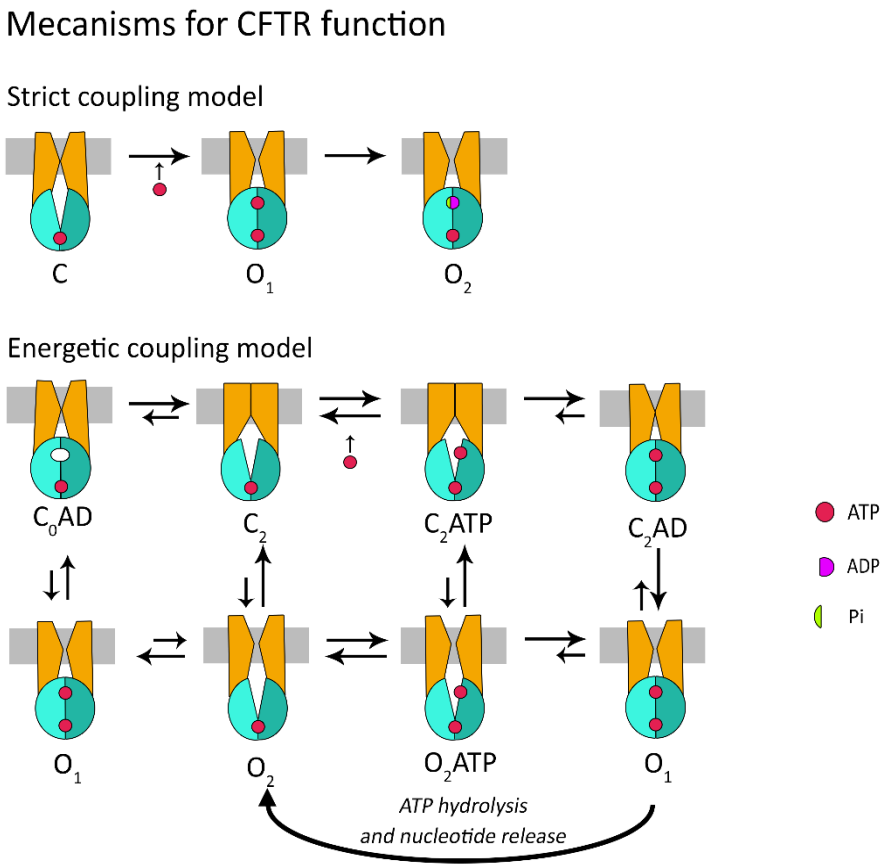


Figure 1.10- *Models for CFTR function. The strict coupling model is cyclic, when the last state is achieved, the cycle restarts over again, while the energetic coupling model assumes allosteric modulation in which a conformational change in one domain make changes in other domain possible, therefore all these transition are in a dynamic equilibrium, in which the length of the arrows represents the probability of the transition. ATP is represented in red, ADP in purple and Pi in green. C stands for closed and O stands for open. The TMDs in orange and the NBDs in blue. Picture adapted from Hwang* et al[108].

1.3.5- ABC transporters whose substrate transport is coupled to electrochemical gradients

Transport studies with artificial substrates on LmrA TMDs concluded that the transport activity of this protein is influenced by pH and electrochemical gradients and showed that apparent symport of substrate and protons was happening[115]. More recent data on LmrA that, in fact, there is a coupling of chemiosmotic mechanisms with the transport mechanism of LmrA. LmrA is able to transport one ethidium molecule, along with one proton and a Cl^- ion, in counter-transport with two Na^+ ions[116]. This way of transport happens while ATP is bound to the NBDs, while ATP hydrolysis interrupts this process.

Remarkably, in the absence of nucleotide or mutations that annihilate ATP hydrolysis, substrate transport still proceeds, while a decrease in transport activity is observed in the

absence of an electrochemical gradient of chloride. This coupling between electrochemical gradients, ATP hydrolysis and substrate transport was also observed in MsbA. Singh *et al*[117] found that ATP-dependent drug transport in MsbA is associated with proton antiport, in the presence of a proton or a reversed electrochemical gradient[117]. Similarly to LmrA, MsbA also shows proton coupled transport in the absence of NBDs. The presence of mutations that invalidate hydrolysis, as well as the addition of non-hydrolyzable ATP analogues, severely interfered with molecular transport. In contrast, the presence of ATP in the absence of a proton gradient did not lead to substrate transport. It is believed that ATP hydrolysis enhances the directionality of the transport reaction in comparison with the plain transport based on chemical gradients only. In fact, similar mechanisms of coupling ATP hydrolysis with concentration gradients have been reported on other ATPases, as well as the existence of proteins with shared transport paths for ions and lipids[118].

The existence of ABC transporters whose transport is dependent on electrochemical gradients adds an extra layer of complexity to their mechanism of transport and add up to the alternating access mechanism proposed by Jardetzky[89]. These findings raise many questions on the role of the electrochemical gradients on the transport thermodynamics and whether and how they affect conformational changes. It is also pertinent to ask how many ABC transporters use membrane potentials.

1.3.6-Substrate recognition and binding in exporters

ABC transporters can carry a large diversity of substrates, such as vitamins, cofactors, lipids, peptides, ions, drugs, among others. Substrate recognition and transport in ABC exporters depend not only of the substrate to be transported, but also of the transporter itself. There is a plethora of biochemical and biophysical data regarding substrate interaction with ABC transporters. Nonetheless, x-ray structures and more recently, cryo-EM structures have provided invaluable insights on the mechanisms of substrate transport, allowing the obtention of structures with substrates bound.

Most substrate binding and transport studies concern P-glycoprotein, due to its clinical importance, since P-glycoprotein is involved in multidrug resistance in cancer. P-glycoprotein is able to bind to numerous substrates, such as xenobiotics and toxic metabolites[119]. There are several hypothesis for ligand binding at the TMDs, while some data suggest the existence of multiple binding sites and that multiple substrates compete for the same binding sites[120], while other studies propose that the TMDs collectively form a single non-specific binding pocket with weak binding of the molecules[121] .

The flippase model was proposed by Higgins *et al*[122] for the transport of substrates by P-glycoprotein. In this model, drugs which are already partitioned into the membrane and located in the inner leaflet, are flipped to the outer leaflet of the lipid membrane, and eventually the other way around. In this way a concentration gradient is generated between the two leaflets[122]. This model was based on partition and NMR studies. Additionally, fluorescence assays showed that P-glycoprotein is able to flip native membrane lipids such as phosphatidylcholine, phosphatidylethanolamine and sphingomyelin[123]. The hydrophobic vacuum cleaner model[119] suggests that drugs partition into the membrane, spontaneously diffuse to the cytoplasmic leaflet and bind to the substrate binding pocket. Afterwards, molecules are effluxed to the cytoplasm. This model is supported by FRET studies and on molecule partitioning studies[119]. These models make the important assumption that the substrate is already partitioned into the lipid membrane before binding to the protein. In fact, it is known that the local lipid composition and cholesterol can play a major role in substrate partition and transport[119]. In addition, lipid content also affects ATPase activity of P-glycoprotein. One of the strategies to overcome multidrug resistance is to develop drugs or optimize drug delivery with the goal of increasing passive diffusion, in order to maximize the intracellular concentration of the molecule and create an opposite gradient to the one created by P-glycoprotein[124]. Despite the existence of these two models, it is experimentally impossible to distinguish them, because the same equilibrium final state is reached in both cases. Nonetheless, it is believed that both can coexist, depending of the substrate and on the thermodynamics of the partition and binding process[119].

ABCG2 is another ABC transporter involved in pharmacokinetics and drug delivery in humans[125]. It also plays a role in uric acid transport, and its malfunction is related with several diseases such as gout, hypertension and kidney disease[126,127]. Cryo-EM structures with a sterol lipid, show substrate binding in a hydrophobic cavity in the absence of ATP[125]. On the other hand, a structure with ATP shows a closed cavity with no space to accommodate the ligand. The authors hypothesized that ATP binding is sufficient for the substrate extrusion, while hydrolysis leads to transporter reset. The conformation with substrate bound is transient, in contrast with the stable occluded conformations found in other transporters, such as P-glycoprotein. Computational studies found two symmetrical binding sites in the transmembrane domains[128]. These sites are similar, but with different lipophilic characters. ABCG2 also causes membrane deformations that can affect substrate entry and diffusion towards the protein[128].

Substrate transport in other ABC transporters has also been studied. It was possible to obtain a structure of the MsbA transporter in an inward-facing conformation with a lipopolyssacharide (LPS) molecule bound to the TMDs, while the NBDs were empty[129]. LPS is located at the bottom of the transmembrane cavity and the binding site for LPS is composed by a hydrophobic pocket

at the bottom of the TMDs, in which the acyl chains bind, and a hydrophilic cavity towards the cytoplasm, in which the glycosidic part binds. Based on these data and on the previous MsbA structures, the authors have proposed a model for LPS translocation in which LPS enters MsbA in the ADP or nucleotide-free state, followed by ATP binding and hydrolysis, that lead to conformational changes culminating in LPS flipping and release. After nucleotide exit, the transporter resets to the inward-facing conformation. In a similar fashion to MsbA, the PglK transporter is thought to transport lipid-linked oligosaccharides using a flipping mechanism[130].

The $LptB_2FG$ transporter is another ABC transporter that also carries LPS and its structure was also solved by cryo-EM and x-ray crystallography[131–133]. $LptB_2FG$ is part of a larger protein complex that exports LPS from the inner membrane to the outer membrane in gram-negative bacteria. This implicates the existence of an extra protein LptC present in the transmembrane domains. Cryo-EM structures reveal that in the presence of ANP-PNP (an ATP analog) the transmembrane cavity is closed, while in the absence of the analog, the transmembrane cavity is open with substrate bound[132]. It is speculated that ATP hydrolysis leads to TMD opening leading to substrate expulsion towards the outer membrane, resetting the conformational cycle. It is thought that substrate leaves the transporter in one step only. In this way, ATP hydrolysis assumes a more regulatory function[132].

Other ABC transporters such as the McjD transporter, the PCAT1 transporters and the antigen processing transporter (TAP), that are responsible for peptide transport, were also characterized. The TAP protein transports peptides that are going to be bound to the major histocompatibility complex I (MHC-I). The MHC-I complex is key to the recognition of virus infected cells and subsequent destruction. TAP is able to carry peptides up to 40 residues long. A cryo-EM structure of TAP with a competitive inhibitor peptide, from herpes virus was obtained[134]. The peptide binds in a parallel fashion to the transmembrane domains, occupying approximately half of the transmembrane cavity. This peptide competes for the peptide binding site and it is believed it acts by blocking the conformational cycle of TAP, leaving it in an inward-facing conformation with the two NBDs separated and it blocks the transmembrane pathway [134,135].

McjD is dedicated to the export of microcin J25, an antibacterial peptide with 21 residues. MD simulations showed that the substrate occupies the entire transmembrane chamber. Despite the large size of the peptide, the transporter undergoes relatively small conformational changes to excrete microcin J25[136,137]. PCAT1 exports quorum-sensing or antimicrobial peptides named bacteriocins, which are up to 90-residues long. Interestingly, PCAT1 possesses peptidase domains, that lie on the membrane inner leaflet, in a parallel orientation to the membrane. These domains cleave a secretion signal of the peptides[138].

The ion transport in the ABC transporters was also examined. Regarding ion transport in CFTR, molecular dynamics simulations showed that ion entry may occur sideways through lateral entrances in the TMDs[139]. There are two lateral clefts, but one of them is thought to be closed in the ATP-bound state. The internal pore in CFTR is composed by a large internal vestibule, a narrow region that serves as a gate and selectivity filter, and a shallow external vestibule. In this way the internal pore has the shape of an asymmetric hourglass. The internal vestibule and the region adjacent to the supposed entrance is lined with positively charged residues, which create a positive potential. The narrow region is flanked by polar residues such as serine, threonine, but also phenylalanine. The external vestibule is also composed by positively charged residues, such as arginine and lysine. This portion of the pore is much smaller than the internal vestibule[108,140]. The mechanism of transport was already explained in section 1.2.4. Recently, a cryo-EM structure of CFTR with a potentiator molecule was obtained[141]. Interestingly, the potentiator binds in a cleft formed by three transmembrane helices, near the membrane-protein interface and outside the transmembrane pore.

1.3.7-Mechanisms for importers function

1.3.7.1-Type I importers

In both type I and II importers the first step of molecular translocation is the substrate recognition and binding to the substrate-binding protein. As seen on section 1.1.5, type I SBPs are quite flexible and bind the substrate in a Venus flytrap fashion[80,81]. Afterwards, the SBP-substrate complex binds to the membrane complex[142]. At the moment of binding, the maltose importer is on the inward-facing conformation with ATP bound, but the NBDs are in the open conformation[143] (figure 1.11). Upon binding of the SBP, there is a reorientation of the transmembrane helices, allowing the transition to the outward-facing conformation[143]. SBP binds to the membrane complex with low affinity, leading to a relatively transient interaction [144–146]. During that transition, the two lobes of the SBP open to release the substrate to the transmembrane cavity (figure 1.11). After SBP docking to the membrane complex, NBD closure

happens[144,147]. A closed NBD dimer makes ATP hydrolysis possible. ATP hydrolysis leads to NBD dimer opening, SBP release and, consequently, substrate diffusion[144,148].

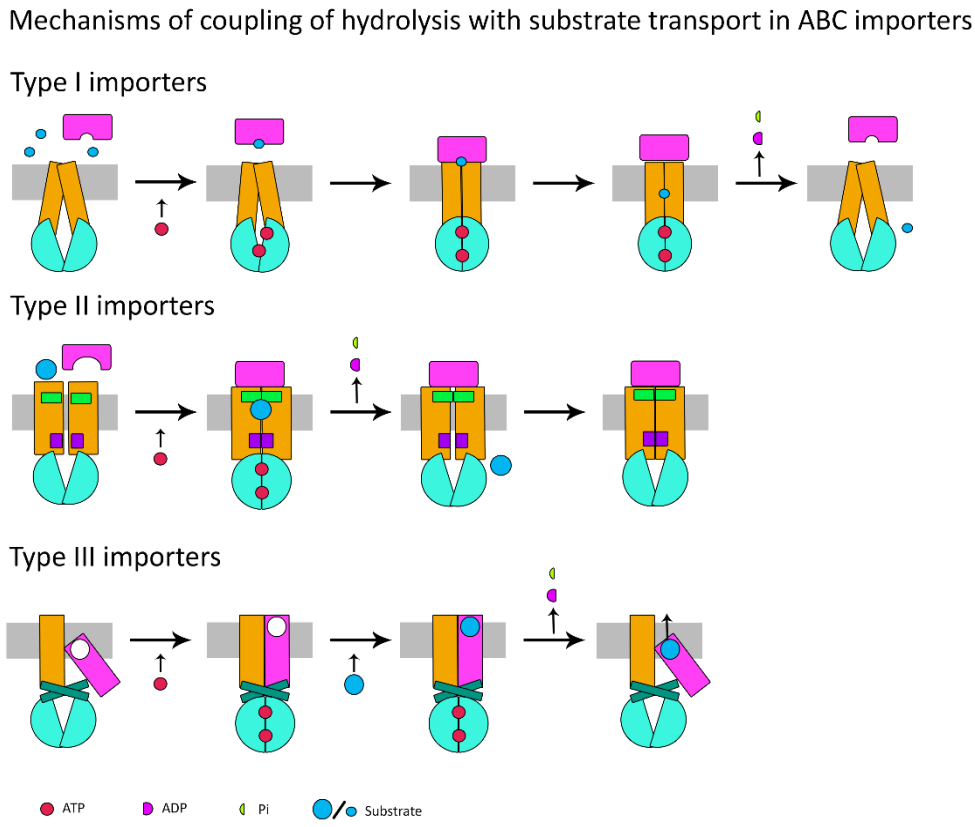


Figure 1.11- *Models predicting the function of ABC importers. The transmembrane domains are represented in orange, the SBPs and the S-component in magenta, ATP in red, ADP in purple, Pi in green, the substrate in blue, the NBDs in cyan. Protein-specific characteristics are also represented, the periplasmic gate in type II importers is represented in green, while the cytoplasmic gate is coloured in dark purple. Additionally, the transmembrane helices on type III importers are coloured in teal. Add models are cyclical, when the last state is achieved, the cycle restarts over again.*

1.3.7.2- Type II importers

The mechanism of molecular translocation for type II importers is drastically different from type I proteins, starting with substrate binding to SBP. Type II SBPs are much more rigid, and no significant conformational changes happen upon substrate binding, meaning that the liganded and unliganded form of type II SBPs are quite similar[149,150]. Upon SBP binding to a type II membrane complex, little conformational changes occur in the transmembrane domains, regardless of the substrate being bound or not[151] (figure 1.11). Additionally, in striking contrast with type I importers, these proteins exhibit a very high basal ATP hydrolysis activity in the absence of both substrate and SBP [54,152]. Also, ATP binding alone can lead to full closure of the

NBDs[70]. Nonetheless, SBP binding to a type II membrane complex has high affinity and forms highly stable complexes[70,152,153]. Upon SBP binding, there is diffusion of the substrate from the SBP (figure 1.11). It is hypothesized that diffusion of substrate happens due to the concerted movement of residues that serve as gates in the periplasmic and cytoplasmic sides[55]. The sequential closing of the periplasmic gate, followed by the opening of the cytoplasmic gate creates a peristaltic movement, allowing substrate translocation[55]. In this way there is no allosteric coupling between the NBDs and SBP, because the conformational changes in SBP do not influence the NBD conformation. After substrate transport, the full complex becomes highly stable, but in an asymmetric conformation [151,153]. In this case, ATP binding and hydrolysis provide the energy necessary for system reset. It is hypothesized that several rounds of ATP binding and hydrolysis are required to do so[54,70]. In fact, a QM/MM study on BtuCD showed that ATP hydrolysis has a free energy change close to zero, proposing that ATP hydrolysis does not deliver a power stroke for conformational changes, but can affect the complex reset[46].

1.3.7.3- Type III / ECF importers

In sharp contrast with type I and II importers, type III or ECF importers, have a mobile transmembrane protein (S component), rather than a soluble binding protein. The S component binds the substrate and delivers it to the ECF module, which is a transmembrane complex composed by a fixed transmembrane protein (T component) and two nucleotide binding domains. In comparison with type I and II importers there is much less information concerning ECF proteins. X-ray structures of full complexes show the S component in close interaction with the coupling helices and in an almost parallel orientation relatively to the membrane [154–156]. The S component interacts with the transmembrane helices of the T component, but not with the NBDs.

The S component is able to detach from the complex and diffuse in the membrane to scavenge the substrate[157]. When the S component binds to the complex, it suffers a reorientation of almost 90º, from being in a perpendicular orientation relatively to the membrane to an almost parallel one. The loop that acts as a lid moves away, opening the binding site. Additionally, in the EcfT from *Lactococcus lactis* transporter, several S components compete for the same ECF module [157]. It is hypothesized that ATP binding leads to NBD closure and to reorientation of the S component that makes it competent for substrate binding. This is followed by ATP hydrolysis and release of the substrate to the cytoplasm[85].

Finally, according to the studies and models above presented, it is highly likely that different ABC transporters operate by different transport mechanisms and reinforce the idea that there is not a unifying mechanism for the action of these proteins, but rather different mechanisms for different groups of transporters that are dependent of the structural characteristics on the protein.

1.4-Computational studies of ABC transporters

Molecular modelling and simulation along with other biophysical techniques, such as x-ray crystallography, EPR and NMR allowed to obtain important structural details on the structure and function of proteins. More recently, cryo-EM has also joined this array of methodologies. ABC transporters are membrane proteins and the need of a membrane environment used to be a nuisance, complicating the isolation of these proteins as well as the obtention of reliable data. In this way, early structural studies of ABC transporters contemplated only the NBDs or even the individual monomers of the NBDs. The first MD simulation study of an ABC protein was a 390 ps simulation of a single NBD of histidine transporter, HisP in 2002, with the goal of studying monomer dynamics [158]. Since then, numerous studies on ABC NBDs were performed, from which a few will be highlighted henceforth. In 2004, the previously referred study was extended to an approximate total of almost 80 ns of MD simulation in the presence and absence of ATP and Mg[159]. This study concluded that ATP and MgATP affected the conformational changes within HisP, mainly in the α-helical subdomains and at the interface between the latter and the RecA subdomain. Later on, the function of the MJ0796 NBD dimer was investigated using MD simulations [160,161]. In one study, different nucleotide states were simulated: ATP bound, ADP/ATP bound and the apo state, concluding that in the presence of ATP in both pockets affects the dimer in a symmetrical manner, while the presence of ADP leads to asymmetrical degrees of interaction in both sites[160]. Other study showed that the energy released by ATP hydrolysis is transferred to the helical subdomains[161]. MD simulations also allowed to investigate the influence of nucleotide in pocket opening in different systems, such as MalK, MJ0796 and BtuCD [160,162,163]. All these studies converged to the conclusion that the presence of ATP in a catalytic site leads to closure, while ADP and phosphate results in a semi-open pocket. Jones *et al*[160] have observed the rotation of the helical sub-domain relatively to the catalytical one, allowing pocket opening. Other studies using enhanced sampling MD simulations assessed the mechanism of ATP binding to the NBDs, concluding that it happens via an induced-fit mechanism [164]. Hayashi *et al* [165] have performed an energetic analysis of dimer opening and concluded that water molecules around the NBD dimer play a key role on the thermodynamics of dimer formation/dissociation by increasing or decreasing its conformational entropy, respectively[165].

The catalytic mechanism of ATP hydrolysis was also subject of several studies. Jones *et al*[33] have defined geometric criteria between the catalytic water and residues that increase the probability of hydrolysis[33], while several other studies have used QM/MM methods to study the hydrolytic mechanism, in which most of them defend a dissociative mechanism and general acid catalysis[46,47,49].

Regarding the study of full-length ABC transporters, increases in the computational power have allowed longer simulations to be performed, in the scale of μs. Nonetheless, the timescale of conformational transitions and molecular transport in ABC transporters often reaches ms. To circumvent these limitations, several methodologies have been employed such as enhanced sampling methods[102,137] and the usage of coarse grain simulations[166]. Most MD simulations contemplate the ABC model systems and make use of the available x-ray/cryo-EM structures. However, when necessary and if possible, homology modelling was also used.

The first molecular dynamics simulation of a full-length ABC transporter was a 1 ns long simulation of MsbA in a lipid mimetic – octane, which raised questions about the stability and correctness of this structure[167]. This MsbA simulation was followed by other studies that aimed to study the interactions between the membrane and the transporter [168]. Weng *et al*[169] have used targeted molecular dynamics to study the molecular details of the interconversion from the outward-facing to the inward-facing state revealing the sequence of events that result in the final state. It starts with the opening of the ATP-binding sites, leading to the disruption of the NBD dimer, whose movements will be transmitted to transmembrane domains via the x-loops resulting in the rearrangement of the TMDs, leading to the exposure of the substrate binding sites [169]. Moradi *et al*[102] also answered this question using non-equilibrium simulations and retrieved the relevant transition pathways using measurements of the non-equilibrium work performed between states, obtaining a more complete picture of the transition. The authors concluded that upon NBD dissociation, there is closure of the periplasmic end of the TMDs, followed by the twisting of the NBDs, ending in the opening of the cytoplasmic portion of the TMDs, naming it the "doorknob mechanism"[102]. Recently, MD simulations have been used to study spontaneous lipid binding into the central cavity of MsbA[170]. It was seen that several phospholipids are able to bind concomitantly in various orientations[170].

ABC importers have also been subjected to extensive MD simulations. Oliveira *et al*[171] have studied the conformational changes triggered by ATP hydrolysis in the maltose $MalFGK_2E$ importer, concluding that hydrolysis induces conformational changes mainly in the NBDs, especially the helical subdomains[171]. Wen *et al*[172] have investigated the conformational coupling between the NBDs and TMDs identifying critical structural motifs of the NBDs whose motion is correlated with the structural changes in the TMDs[172]. More recently, Weng *et al*[173] used

metadynamics simulations to explore the effect of MalE on different conformational states and found that MalE energetically favours closure of the of the NBDs and promotes the vertical motion between MalF and MalG helical cores[173]. The BtuCD B_{12} vitamin importer has also been subject to several simulation studies, ranging from targeted MD to study the conformational changes from the inward-facing to outward-facing state[174], to equilibrium MD simulations to assess the ATP hydrolysis effect on the complex[175]. It was observed that ATP hydrolysis leads to the opening of the ATP binding pockets and the opening of one the cytoplasmic gates as well as to the establishment of an interaction network between the NBDs and the TMDs[175]. Prieß *et al*[176] have analysed the force networks in the wild-type and hydrolytic mutants of BtuCD and concluded about the pathways of force transmission from the NBDs to the TMDs and found that the two NBDs show asymmetric behaviour[176].

The appearance of structures of asymmetric transporters such as TM287/288 also motivated MD studies: Furuta *et al*[177] have studied the nucleotide and substrate effect on NBD closing. The authors concluded that two ATP molecules are required to allow complete dimer closure and that the presence of substrate stimulates NBD closure. On the other hand, the absence of nucleotide induced a wide-open conformation, with exposed α-subdomains[177]. Gödekke *et al*[178] performed microsecond simulations of the same at different temperatures and observed a transition from the inward-facing state to the outward-facing state[178]. Despite all the studies mentioned above, a significant body of research is focused on ABC systems with a direct impact on human health, mainly P-glycoprotein, CFTR and, more recently, ABCG2.

P-glycoprotein (P-gp) still is the subject of extensive research. For instance, investigations regarding ligand binding range from docking of substrates with the goal of identifying binding hotspots[124,179], passing by the usage of free-energy methods to study binding pathways for some substrates, to structure-activity relationship (SAR) models are also quite used when identifying inhibitors and substrates and distinguish them from non-substrates[180]. Additionally, homology modelling was also used to model human P-gp structures using mouse P-gp as template[120,181]. Nonetheless, O'Mara *et al*[182] showed that one of the mouse P-gp structures is unstable in a membrane embedded MD simulation, becoming stable only in the presence of detergent[182]. Barreto-Ojeda *et al*[166] have used a coarse-grain model of P-gp embedded in bilayers of different compositions with the goal of understanding lipid entry and binding to the central cavity. The authors found lipid entry is favoured by positive residues at the entry of portals and within the cavity and that only lipids from the lower leaflet can access the protein in the inward-facing state[166].

CFTR has also been extensively studied with molecular simulation methods. The first homology models of the CFTR NBD dimer were made based on the structure of the MJ0796 NBD dimer[183]

and others[184]. Other models were done of the NBD dimer using different approaches, such as protein-protein docking[185]. The appearance of the Sav1866 structure allowed the creation of models of the full-length protein. Such examples are the models of Mornon *et al*[186] and Serohijos *et al*[187]. Yet, these studies showed significative differences at the interface between NBDs and TMDs, which are the result of the usage of different modelling templates and alignments, arising from the low sequence identity of the TMDs. Molecular dynamics simulations were also used to study the structure of the CFTR cavity and entry portals, often supported by experimental data[188,189]. Corradi *et al*[190] already made use of the available cryo-EM structures to study the transmembrane cavity and the shape of TM8, as well as Strickland *et al*[191] who used cryo-EM maps to refine an homology model and proceed to simulation with ATP bound. The structural knowledge of CFTR allows to predict the effect of cystic fibrosis related mutations. Regarding the study of CFTR mutations, one of the most studied mutations is F508del. Early full-length CFTR homology models already suggested that F508 should be involved in interactions in the NBD-TMD interface[187,188]. On the other hand, MD simulations of NBD1, where the mutation is located, did not reveal any significant structural changes[192], although it was found that F508del NBD1 has higher aggregation potential, by populating states associated with partially misfolded conformations[193] and replica-exchange MD simulations showed that F508del increases local and global fluctuations of NBD1[194]. Full-length CFTR models also allowed to conclude on the structural effect of other mutations[195]. It was found that CFTR mutations tend to cluster around critical regions, such as the interface between NBDs and TMDs, the intracellular loops, the canonical ATP-binding site along with other residues lining the pore[188,195].

Models of the NBD dimer and of the full-length CFTR have also been used to study the binding of modulators. Moran *et al*[196] have used an NBD dimer model to perform blind docking and found three hotspots, two of them corroborated by experimental results[196]. Other *in silico* screening studies contemplating the full-length F508del CFTR were also done and found that many molecules bind at the interface between the two NBDs, the interface between NBD1 and ICL4 (where F508del is located) and the interface between the NBDs and the TMDs[197,198]. Nonetheless, it was shown that other molecules are able to bind to the TMDs [199,200]. Additionally, docking studies allowed to understand the mechanism of action of already known drugs such as lumacaftor that increases the amount of properly folded CFTR, binding mainly on the interface NBD1-ICL4[200,201]. On the other hand, ivacaftor, a potentiator already used was suggested to bind at the interface between NBDs[202]. Docking studies are also being used to develop small-molecule inhibitors for the treatment of secretory diarrhoea or other CFTR-associated diseases, such as polycystic kidney disease [199,203].

A comprehensive review of the screening studies of ligands on CFTR can be found in Callebaut *et al*[195].

Finally, another system that has received some attention recently is ABCG2, which is involved in multidrug resistance. Before the obtention of the cryo-EM structure, several homology models were done based on the ABCG5/G8 structure and docking of various compounds was done. In that way it was possible to identify binding hotspots[128,204] and substrate translocation pathways. With the appearance of more cryo-EM structures further studies are expected to appear.

1.5-Studied systems on this thesis

1.5.1-CFTR

The CFTR (cystic fibrosis transmembrane conductance regulator) channel is an ABC transporter that plays a key role in maintaining the correct equilibrium of water and electrolytes in epithelial tissues[205]. The CFTR channel is mainly responsible for chloride transport, but it is also involved in the regulation of other channels, being involved in the inhibition of sodium channels in epithelia and the production of acidic environments in macrophages[205–207].

Structurally, CFTR is 1480 residues-long and contains the core elements of ABC transporters, such as two NBDs and two transmembrane domains, but it also has an extra regulatory domain between the NBDs and the TMDs, with multiple PKA-phosphorylation sites. For a long time, the full-length structure of CFTR has remained elusive, and only x-ray structures of isolated NBDs were available. However, advances in cryo-EM allowed to obtain structures with both NBDs and TMDs (figure 1.12 A-C). At the moment of writing, there are over seven structures/density maps of CFTR with transmembrane domains. Even though, none of these structures displays a full-length protein since the R domain is intrinsically disordered and it is hard to obtain proper electron density. The CFTR NBDs show a similar structure to other ABC transporters but form a heterodimer. Nonetheless, there are two exclusive segments to CFTR, the regulatory insertion (RI) and the regulatory extension (RE)[208]. The RI is a loop in NBD1, while the RE is part of the C-terminal of NBD1 and connects the NBD1 to the N-terminal of the R domain[209,210]. Both RI and RE contain additional phosphorylation sites. One of the ATP binding pockets in the NBDs is catalytically inactive, containing residue substitutions in the canonical motifs, detailed in table 1.1.

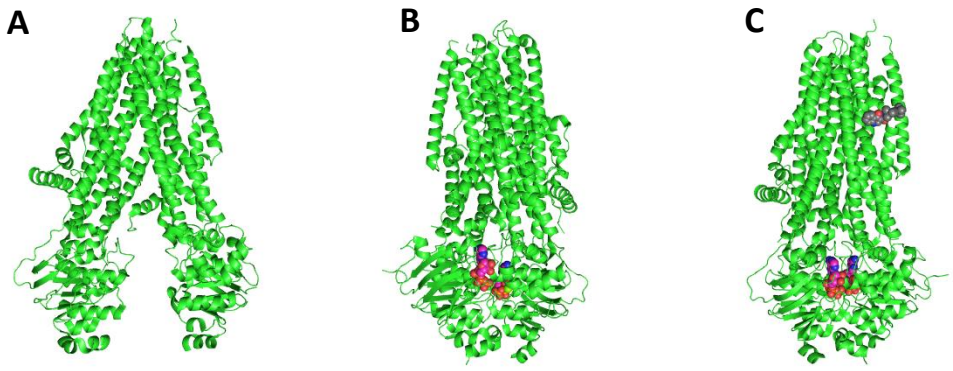


***Figure 1.12**-Examples of structures of representative conformational states of CFTR. A-Human CFTR with the regulatory domain dephosphorylated in a nucleotide-free state. In this state the channel pore is closed. PBD code: 5UAK[211]. B- Human CFTR with the phosphorylated R domain and ATP bound. PBD code: 6MSM[107]. C-Human CFTR with the phosphorylated R domain and ivacaftor and ATP bound. ATP is represented in coloured spheres, while ivacaftor is represented in grey spheres. PDB code:6O2P[141]. The regulatory domain is disordered and does not produce a stable electron density.*

***Table 1.1-** Table comparing the typical ABC motifs in NBD1 and NBD2. The canonical motifs are Walker A in NBD2, the ABC sequence in NBD1, Walker B sequence in NBD2 and the H-loop on NBD2. The residues highlighted in red correspond to the catalytic residues and the corresponding replacements in the degenerate sequences.*

	NBD1	**NBD2**
Walker A	GSTGAKTS	GRTGSGKST
ABC sequence	LSGGQ	LSHGH
Walker B	RAVYKDADLYLLDS	RSVLSKAKILLLDE
D-loop	D	D
H-loop	S	H
A-loop	W	Y
Q-loop	Q	Q

The most important differences between NBD1 and NBD2 (table 1.1) are highlighted in red. The H-loop histidine in NBD2, was replaced by a serine residue in NBD1, while the catalytic glutamate in the Walker B motif of NBD2 was replaced by a serine residue. As a consequence, when the two NBDs are together, the ATP binding pocket that includes the non-canonical sequence will not be catalytically active. However, the ATP binding modes are also similar to other ABC transporters, making similar interactions. Nonetheless, weaker interactions are observed in the non-catalytic site and the CFTR structures show that this site has a larger degree of opening, leading to a higher degree of exposure of this site to the solvent[110]. ATP can bind in this pocket for tens of seconds[212,213] . Mutations in this site decrease the transition rate from a

pre-hydrolytic state to the post-hydrolytic one, and high affinity analogues bound delay channel closure[214–216], suggesting that ATP binding to this pocket might have allosteric effects in gating energetics and channel kinetics.

CFTR has twelve transmembrane helices and four intracellular loops (ICL) that stem from the TMDs and make the interface between TMDs and NBDs. Each transmembrane domain includes six transmembrane helices that origin two ICLs each. The transmembrane domains form the ion entrance, pathway and exit area. Transmembrane helix 8 contains a short unstructured segment in the second half of the helix, breaking the symmetry of the protein and shaping the ionic pathway. The ion entrance is located in a lateral cavity formed by TM4 and TM6, but a second and smaller entry was reported formed by TM10 and TM12[195]. The larger lateral entry is lined with positively charged residues, such as arginines and lysines, believed to attract Cl^- anions. The inner cavity is asymmetrical and contains a large internal vestibule and a narrower region that serves as gate and selectivity filter. The large internal vestibule also contains positive residues that create a favourable electropositive potential to direct anions towards the exit[108]. Some authors also regard these residues as anion binding sites[217]. It is believed that the gate region of CFTR is formed by helices TM12, TM1 and TM6, and it is a narrow region lined with bulky and charged residues that serve as size and charge filter, near the external end of the pore[108]. It was shown that this region is controlled by the presence of ATP[218]. Other authors proposed the existence of a separate size and charge filters, in which the charge filter is located more towards the cytoplasmic side, while the size one is located towards the external side[219]. The external vestibule is a much smaller cavity lined with positive residues, with the role of concentrating chloride ions to the pore[220]. It is defined by the external segments of the transmembrane helices. Mutations of the positive residues in this region lead to decreased open state duration as well as reduced conductance[220,221]. In the absence of mutations, channel opening results in ions crossing the pore at a rate larger than 1 million/second[108].

The R domain is essential for CFTR activation via its phosphorylation by PKA kinase, but the mechanism of activation is still unknown. This domain is intercalated between the NBD1 and the TM7 and it is thought to have arisen from a noncoding sequence in the genome[222]. This segment possesses a high net charge as well as low hydrophobicity, contributing for making it an intrinsically disordered structure. It also contains two clusters of negative charges in which one of them is highly conserved and interacts with the transmembrane domains[223], as it was confirmed by the human CFTR cryo-EM structure. Mutations in this region were found in patients with cystic fibrosis[224]. The R domain phosphorylation triggers a conformational change that allows NBD dimerization. In fact, unphosphorylated zebrafish and human cryo-EM structures show electron density of the R domain in between the NBDs, blocking NBD

dimerization (figure 1.12 B and C)[107,141]. Additional density was observed near the interface between NBD1 and ICL4, which is remarkable, since some disease-associated mutations are located in this region[108,211].

The existing models for the conformational cycle of CFTR were already described in section 1.2.4.

The malfunction of CFTR leads to cystic fibrosis, which is characterized by the formation of viscous secretions in epithelial tissues, leading to obstruction, inflammation, and infections[225], resulting in poor life quality and decreased life expectancy. There are over 1700 known mutations that lead to cystic fibrosis. Mutations can affect different stages: protein synthesis (Class I mutations), maturation (Class II), gating (Class III), conduction (Class IV) and insufficient expression (Class V). Class II mutations are the most frequent ones and the most common mutation in the Caucasian population is the deletion of phenylalanine 508 (F508del)[207]. F508del affects protein folding in the endoplasmic reticulum leading to its degradation. Nonetheless, a small portion is able to reach the membrane, showing an impaired function and decreased half-life residency [226–229]. F508 is located in NBD1 within a loop that interacts with ICL4. X-ray structures of the isolated F508del NBD1 showed little difference in comparison with the wild-type structure, with some variations on the loop where it is inserted. Nonetheless, MD simulations show that the F508del NBD1 in solution displays larger fluctuations and has tendency to form misfolded intermediates prone to aggregation[193,194]. In the context of the full-length protein, it is not clear what the effect of F508del is, but some studies show that F508del impairs domain communication between NBD1 and the transmembrane domains[187,230]. Indeed, the F508 residue is inserted in a hydrophobic pocket composed by the ICL4 residues.

In order to overcome F508del effects, combinations of several compounds are used. These combinations contain molecules with different effects, such as potentiators and correctors. Correctors increase the amount of mature and functional CFTR on the membrane, while potentiators have the purpose of restoring the channel conducting activity. The most common combinations are elexacaftor+tezacaftor+ivacaftor, lumacaftor+ivacaftor and tezacaftor +ivacaftor. Elexacaftor and lumacaftor are correctors and ivacaftor and tezacaftor are potentiators[195]. A recent cryo-EM structure of CFTR with ivacaftor and other potentiator drug, GLPG1837, showed that both drugs bind in the transmembrane domains, in a pocket formed by TM helices 4, 5 and 8 in the transmembrane domain 2 (figure 1.12-C). This binding site includes the loop region of TM8, but only about 40% of the molecular surface of the inhibitors contacts directly with the protein, while the remaining 60% are exposed to the membrane. It is hypothesized that potentiators help stabilizing the rotation of TM8 upon ATP binding[141]. The mechanism of correctors is still unknown, although some authors claim that these enhance

interactions between TM1 with NBD1, as well as restore TM1 folding in F508del CFTR , while it is being synthesized[231]. Other computational studies claim that potentiators bind to pockets at the interface of ICL4 and NBD1, near the place of F508del mutation[232]. However, these data hardly explain how correctors increase CFTR expression.

Compounds that target CFTR have interest, not only because of cystic fibrosis, but also to combat hyperactivity of CFTR due to bacterial toxins that result in secretory diarrhoea, mainly in developing countries[233].

1.5.2-MsbA

MsbA is an ABC transporter involved in lipopolysaccharide (LPS) export in gram-negative bacteria, such as *Escherichia coli*. LPS molecules constitute the outer leaflet of the outer membrane, while the inner leaflet, as well as the inner membrane, is constituted by glycerophopholipids[234]. LPS contributes to the protection from environmental agents, such as antibacterial molecules and evasion from the host immune systems. LPS is synthetized in the inner leaflet of the inner membrane. MsbA plays a key role in flipping LPS from the inner leaflet to the outer one, allowing its transport to the outer membrane by the Lpt complex [235]. This latter complex also contains the ABC exporter complex $LptB_2FG$[236]. MsbA is one of the most studied ABC transporters, resulting in a vast collection of biochemical and structural data including x-ray structures and, more recently cryo-EM data (figure 1.13). Early X-ray structures of MsbA were obtained in the inward and in the outward-facing states[96]. One of the inward-facing structures without nucleotides showed a large separation between NBDs, while the other one displays a shorter distance between NBDs, yet they are twisted against each other[96].

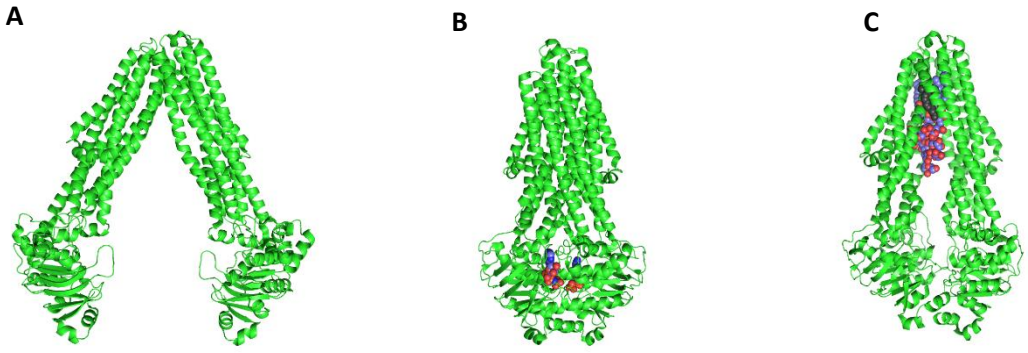


Figure 1.13- *Examples of representative structures of MsbA. A- A wide-open, inward-facing, MsbA structure, PDB code: 6O30*[237]*. B-Closed, outward-facing MsbA with ATP analogues bound. The ligands are represented in purple spheres. PDB code: 3B60*[96]*. C-Semi-open MsbA with LPS and an inhibitor bound. The inhibitor is represented in gray spheres, while LPS is shown in purple spheres. PDB code: 6BPP*[238]*.*

These structures provided important mechanistic details on exporter function, since they were some of the early full-length ABC transporter structures. More recently, cryo-EM allowed to obtain an inward-facing structure with LPS inside the transmembrane cavity[238]. LPS is bound to the last half of the transmembrane domains, with the lipid chains bound to a hydrophobic portion of the cavity towards the extracellular medium, while the glucosamine moiety binds to an hydrophilic region located in the opposite direction. Another structure was resolved with ADP-vanadate bound, mimicking a transition state after ATP hydrolysis[238]. This structure shows two helices that move towards the centre of the cavity, annihilating this cavity, in stark contrast with the outward-facing structure previously obtained with ATP, which showed a clear separation between the transmembrane domains. The cryo-EM map for ADP bound MsbA showed that it already adopts an inward-facing conformation[238]. Hence, the authors confirmed a transport mechanism for MsbA, in which the LPS binds to the protein in the absence of nucleotide, whereas ATP binding leads to the transition to the outward-facing state with consequent release of the substrate. ATP hydrolysis will then reset MsbA to the inward facing state[238].

MsbA shows adenylate kinase activity, as previously explained in section 1.1.3. It is known that adenylate kinase activity support substrate translocation[59], yet the mechanistic details are still unknown.

1.5.3-MalFGK$_2$E from Escherichia coli

The *Escherichia coli* maltose importer MalFGK$_2$E is one the most studied ABC transporters systems. It is mainly used as a model system for type I importers and is involved in the intake of malto-oligossacharides. In *E.coli* this protein is located in the inner membrane, transporting molecules from the periplasm to the intracellular medium. It acts in conjunction with the outer membrane protein LamB, which is a β-barrel protein with 18 antiparallel β-helices, containing strategically located aromatic residues to facilitate translocation via hydrophobic interactions – the greasy slide[239]. MalFGK$_2$E is a protein complex with two separate and different transmembrane domains: MalF and MalG. However, the NBDs are constituted by two copies of the same chain: MalK, which also contains a regulatory domain. The substrate binding protein (SBP) is MalE. In *E.coli*, MalE freely diffuses in the periplasm, yet in *Bdellovibrio bacteriovorus* MalE is fused to the MalF protein[240].

MalE has a typical structure of a type I SBP with two globular lobules connected by a hinge domain, constituted by three flexible loops. MalE binds to the substrate using a “Venus-flytrap” mechanism[241]. The substrate binding pocket is lined with polar and aromatic residues, in which the latter are predominantly located in domain I. Smaller sugars with one ring only can bind only

in hinge region, while longer substrates require binding to both lobules. Four subsites were found in the substrate binding pocket in which the first binds to the first ring in maltose, the next two to the second ring and the last one to longer substrates. Remarkably, MalE is specific for malto-oligossacharides due to the particular orientations and conformations of the hydrogen bonding residues and of the aromatic residues that stack with the glycosidic rings, and the ability of maltose derivates to fulfil these interactions[241].

MalF and MalG are the transmembrane units, which are constituted by eight and six helices, respectively. MalF possesses four periplasmic loops and three cytoplasmic loops. The P2 loop is the largest with a length of 3 nm with an Ig-like folding[242]. The P2 loop interacts with MalE via the C-terminal lobule, when it binds to the membrane complex. The coupling helix is located between TM6 and TM7, and interacts with a groove in MalK, transmitting the conformational changes from the NBDs to the remaining protein[11]. MalG contains three periplasmic loops, in which two of them are inserted in MalE and two cytoplasmic loops. The MalG coupling helix is in between TM4 and TM5. The TM helices 3 to 8 of MalF are related with helices 1 to 6 of MalG, forming a core region with pseudotwofold symmetry, which form a pocket able to hold the substrate. On the periplasmic end, this pocket is terminated by four helical kinks. The interior of the cavity is lined with polar and hydrophobic residues[41,243].

MalK contains the canonical motifs typical of ABC transporters and the common folding of ABC transporters, but it also contains a regulatory domain on the C-terminal. The $EIIA^{Glc}$ protein in its unphosphorylated state binds to the regulatory domains, stabilizing the inward-facing conformation of the transporter and preventing any structural rearrangements[244]. As a consequence of this inhibition, substrate transport in $MalFGK_2E$ is halted. This regulation is part of the carbon catabolite repression strategy in bacteria, that allows *E.coli* to favour carbon sources that support rapid growth[244].

The crystal structures of $MalFGK_2E$ show MalE docked to the interface of MalF and MalG. In the inward facing structures MalE is in an open unliganded conformation, while in the outward facing ones encounters itself in the closed liganded conformation. In addition to the interaction interface with MalF and MalG, MalE also binds to the P2 loop of MalF. MalG also contributes for binding with interactions from two of the three periplasmic loops. The P3 loop (also known as the scoop loop) of MalG insert into the binding site of MalE during the transport cycle to force the transfer of maltose to TMDs and the P1 loop is known to interact with the N-terminal lobe of MalE.

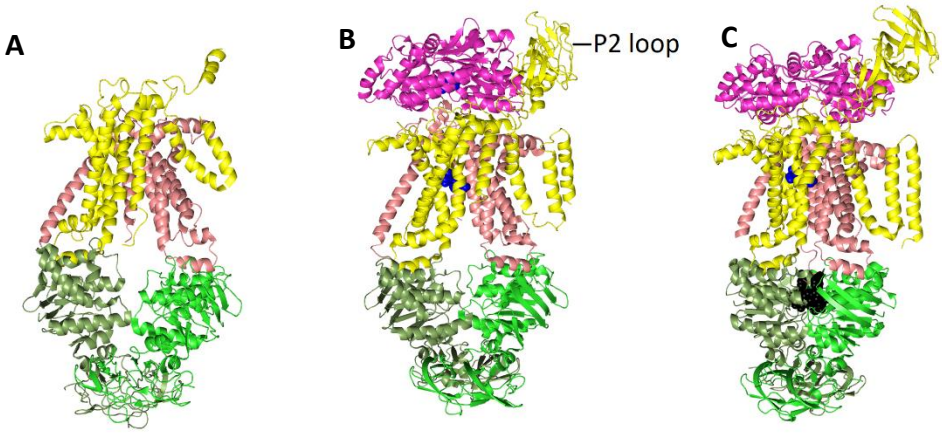


Figure 1.14- *Representative structures of the MalFGK$_2$E importer from* E.coli. MalE is coloured in magenta, MalF in yellow, MalG in salmon and MalK in green. *A- MalFGK$_2$ importer without MalE nor nucleotides bound. PBD code:3FH6*[245]*. B-MalFGK$_2$E importer in a pre-translocation state, with maltose bound on the transmembrane domains and MalE. PBD code: 3PV0*[143]*. C-MalFGK$_2$E with ATP analogues and maltose bound. Maltose and ATP are displayed in blue and black spheres, respectively.PBD code: 3RLF*[41]*.*

The crystal structures of MalFGK$_2$E can give a structural insight on the catalytic cycle of the maltose importer, which has already been described in section 1.2.6.1. Prior to MalE binding and without ATP bound, the importer is in the inward-facing state with the separated NBDs and TMDs (figure 1.14A), while MalE docking and even before ATP binding, the separation between the NBDs and TMDs decreases (figure 1.14B). Upon ATP binding, there is NBD and TMD closure, leading to the outward-facing in which the TMDs form a cavity to accommodate the substrate (figure 1.14C).

Crystallographic structures containing molecules that mimic catalytic intermediates of the ATP hydrolysis show that the hydrolysis process does not trigger any substantial conformational changes on the protein[41]. The structures of post-hydrolytic intermediates are not known, yet cryo-EM data[148] revealed the existence of a semi-open asymmetric MalK conformation, when bound to ADP and other analogues, but in the absence of MalE. Nonetheless, the sequence of events after hydrolysis is still not clear[148].

1.6-Scope and contributions

As seen so far, ABC transporters are ubiquitous proteins that are involved in phenomena of interest. Therefore, they are extensively studied using biochemical and biophysical methodologies.

The main aim of this thesis and contributions is to provide molecular details about the function of ABC transporters using computational methods. Three different systems (CFTR, MsbA and the *E.coli* MalFGK$_2$E maltose importer) were used to investigate key aspects regarding the mode of action of this class of proteins, allowing for a more complete understanding of their function.

The final goal was to simulate the actual substrate translocation process and extract knowledge from these simulations, which was done using the *E.coli* $MalFGK_2E$ maltose importer.

In this chapter (chapter one), a comprehensive introduction to ABC transporters was presented in order to allow the understanding of the fundamental concepts underlying the function of these proteins.

In chapter two, the theoretical background for the computational methodologies employed will be reviewed.

The study of the effect of F508del mutation in the CFTR NBD dimer is the subject of study of the third chapter. Extensive molecular dynamics (MD) simulations of the wild-type and mutant were performed in order to evaluate the impact of this mutation on the isolated dimer, in the ATP state and upon ATP hydrolysis. This work showed that F508del does have an impact on the properties of the ATP-bound dimer, such as the increase in the opening of the ATP binding sites before hydrolysis, the decrease in the number of catalytically competent conformations and the alteration in the magnitude of the conformational changes upon hydrolysis. This work was entirely developed by me and published in *Proteins*[32].

Chapter four will approach the molecular details related with the adenylate kinase (AK) reaction in MsbA. This work was a collaboration with the group of Prof. Clemens Glaubitz at Goethe University in Germany, in which a combination of solid-state NMR, EPR and MD simulations were used to answer the proposed questions. It was concluded that the AK site is located halfway between both ATPase sites in close proximity to the Q-loop and H-switch, and despite that the binding of two nucleotides is possible, it is highly likely that only one nucleotide at a time might be able to productively react with the other on the ATPase site, as seen by the final nucleotide conformations obtained in MD simulations. These data assessed and confirmed the location of the AK site in MsbA and that the NBDs are able to accommodate more than two nucleotides. The MD simulations in this work were my contribution. This work resulted in a joint publication in the *Journal of the American Chemical Society*[64].

The fifth chapter reports the effect of ATP hydrolysis and nucleotide exit on substrate binding and translocation on the maltose $MalFGK_2E$ importer. It was found that ATP hydrolysis and nucleotide exit lead to conformational changes throughout the protein, that cause the decrease of energetic barriers to maltose translocation towards the NBDs, while simultaneously increasing the energetic barriers on the opposite direction towards MalE. Therefore, ATP hydrolysis and nucleotide exit contribute for the irreversibility of the process. A novel binding pocket for maltose in MalG was identified and Y383 was identified as a gatekeeper residue, being its conformation highly nucleotide dependent. This work was developed entirely by me and was submitted to *Scientific Reports.*

The sixth and last chapter will summarize the main findings in each work and discuss them in the wider ABC transporters context.

Chapter 2

Theory and methods

2.1- Molecular modelling and computer simulations of molecules

Molecular modelling embodies all methodologies to simulate and study the molecular behaviour using models. These models are simplified representations (or abstractions) of the systems of interest [246,247]. Early models of molecules involved physical objects, such as glass or gelatine spheres[248]. In 1953, for the first time, a computer was used to simulate a molecular model. It was a Monte Carlo simulation of a two-dimensional rigid-sphere system, carried out at the Los Alamos National Laboratories in the United States, using the most powerful computer at the time, the MANIAC [249]. In 1975, Levitt and Warshel simulated the folding of the bovine pancreatic trypsin inhibitor (BPTI), and in 1977, McCammon and Karplus successfully simulated BPTI dynamics [250,251]. Since then, an increase in computational power and model complexity has allowed the study of complex systems such as membranes, protein complexes and even entire viruses.

Molecular modelling can provide detailed information at a molecular or atomic level, such as the conformational space sampled by the molecules and, in contrast with most experimental techniques, do not involve averaging over time and/or space. In addition, it is possible to obtain thermodynamic and statistical properties, and relate them with the conformations sampled by the molecules and, eventually, with their function. Molecular simulations can also be used to predict experimental outcomes and guide experiments, such as in the case of drug design[252].

2.2-Molecular mechanics (MM) models

Although atoms have a quantum nature, it is not always necessary to model the quantum behaviour of molecules, and it often becomes impractical to do so in biomolecules due to the large number of atoms and degrees of freedom. Thus, depending on the questions to be answered, several approximations can be done by decreasing the level of detail and number of degrees of freedom of the system. Nonetheless, all models need to be validated by properties that were already measured experimentally, in order to ensure reliable predictions.

In this work, molecular mechanics (MM) models were mostly used. Continuum models were also used to perform continuum electrostatics calculations.

In molecular mechanics models, atoms or sets of atoms are treated as single particles with fixed charges, van der Waals radii and masses. These models are used in the framework of classical mechanics, in which forces, potential and kinetic energies are calculated with this assumption. This simple description relies on major underlying assumptions:

- The Born-Oppenheimer approximation
- That molecules are described by empirical force-fields

The Born-Oppenheimer approximation states that the motions of nuclei and electrons can be decoupled[253]. This happens because the nuclei have a larger mass and their inertia will be larger, while electrons are much smaller and move faster. If nuclear positions change, electrons are supposed to instantly adjust their positions. As a consequence, it is possible to consider the energy of a configuration as a function of the nuclear coordinates only, and that electrons are in their ground state[246].

A force field model simply consists on a set of equations and parameters chosen to evaluate the potential energy of molecular conformations. Force fields are empirical because they attempt to reproduce quantum mechanical behaviour in a series of separate equations. This approach will always be an approximation and will have a limiting accuracy[254]. The potential energy is split in several terms: terms related to "bonded" interactions, such as bonds, angles and dihedrals[246] and "nonbonded" terms that account for the van der Waals and electrostatic interactions (equations 2.1,2.2 and 2.3). In this way, force fields are said to be additive, because they can be written as a sum of different terms. Generally, force fields are developed and tested on a small set of molecules and are applied to other molecules. It is possible to have different combinations of parameters that aim to reproduce the same potential energy landscape. However, generally, force fields optimized and developed in different ways are applicable to different systems[255 254].

The force field used in this work was GROMOS 54A7, which is a force field developed for biomolecular simulations[256]. GROMOS 54A7 is an united atom force field, which means that all aliphatic hydrogen atoms are collapsed towards the bound carbon.

$$V_{total} = V_{bonded} + V_{non-bonded} \quad 2.1$$

$$V_{bonded} = V_{bonds} + V_{angles} + V_{proper\ dihedrals} + V_{improper\ dihedrals} \quad 2.2$$

$$V_{non-bonded} = V_{van\ der\ Waals} + V_{electrostatic} \quad 2.3$$

In GROMOS 54A7 the potential energy is given as the sum of bonded and non-bonded terms as described in equations 2.1 to 2.3.

2.2.1-Bonded terms

The bonded terms can be further decomposed in several contributions: bonds, angles, and proper and improper dihedrals (equation 2.2).

$$V_{bonds} = \sum_{bonds} \frac{1}{4} k_b (b^2 - b_0^2)^2 \qquad 2.4$$

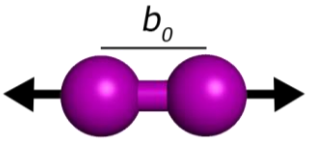


***Figure 2.1**- Bond vibrational behaviour. The arrows represent the atomic stretching motion around b_0 , which is the reference length.*

The equation 2.4 represents the potential used by the GROMOS54A7 force field to model bond stretching (figure 2.1). It is a fourth power potential function that describes the behaviour of a spring. In this case k_b is the force constant for bond stretching, b is the bond length and b_0 is the reference length. Harmonic potentials are also commonly used to model bond behaviour. Generally, constraint algorithms are used in simulations. These algorithms maintain bond lengths constant regardless of the forces acting on them. This approximation is made because the high frequency modes (such as bond vibrations) are not of interest in comparison to other potential conformational changes, which are much slower. Furthermore, a constrained bond shows a more similar behaviour to a quantum oscillator than a classical (harmonic) oscillator[257]. This also allows to increase the integration timestep and speeding up the simulations. The constraint algorithms used in this work are SETTLE[258] for the water molecules and LINCS [259] for the remaining molecules.

Angle bending is also modelled using the angle cosine (equation 2.5 and figure 2.2). Once more, k_θ is the force constant, $cos\ \theta$ and $cos\ \theta_0$ are the cosine of the angle and the reference angle, respectively.

$$V_{angles} = \sum_{i=1}^{N} \frac{1}{2} k_\theta (cos\theta - cos\theta_0)^2 \qquad 2.5$$

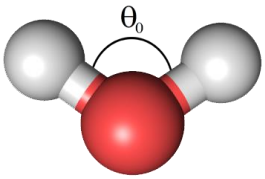


Figure 2.2*-Angle vibrational behaviour. The angle Θ_0 is composed by the three atoms displayed.*

Dihedrals are angles between planes. In GROMOS 54A7 improper dihedrals have the main function of keeping the planarity of chemical groups (figure 2.3 – first picture) and the tetrahedral geometries (figure 2.3- last two pictures). Improper dihedrals are also modelled with a harmonic potential of angle cosine (equation 2.6) in which k_ξ is the force constant, ξ and ξ_0 are the angle and reference angle, respectively.

$$V_{improper\ dihedrals} = \sum_{i=1}^{N} \frac{1}{2} k_\xi (cos\xi - cos\xi_0)^2 \qquad 2.6$$

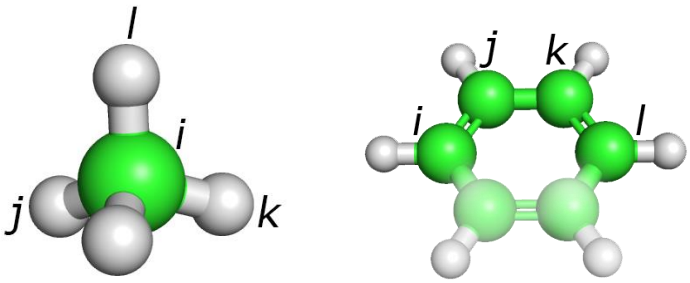


Figure 2.3*- Representation of the application of improper dihedral to maintain the tetrahedral geometry of molecules (first molecule at left) and the planarity of rings (second molecule at right) and other planar chemical groups. The improper dihedral angle ξ is defined as the angle between planes defined by atoms [I,j,k] and [j,k,l].*

Proper dihedrals model the rotation of chemical bonds. In this case, a periodic potential, given by a sinusoidal function is used (equation 2.7 and figure 2.4). k_θ is a force constant, σ is the phase shift, m the multiplicity and ϕ the angle.

$$V_{proper\ dihedrals} = \sum_{i=1}^{N} k_\phi \left[\left(1 + cos(\sigma) cos(m\phi)\right)\right] \qquad 2.7$$

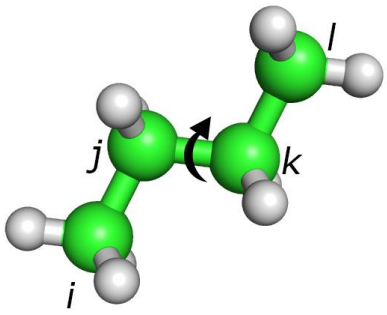


Figure 2.4- *Torsional behaviour of bonds, modelled by proper dihedrals The angle ϕ in equation 2.7 is defined by planes [l,j,k] and [j,k,l].The bond whose rotation is modelled in this case is the bond between atoms j and k.*

2.2.2-Non-bonded terms

The non-bonded terms account for the van der Waals and electrostatic interactions.

The van der Waals interactions describe deviations from the ideal gas behaviour. These interactions are described by attractive and repulsive contributions. The attractive (or dispersion) contribution arises from the distortion of electronic clouds, creating instantaneous dipoles and induced dipoles in neighbouring atoms or molecules, and it is considered to be a long-range contribution, due to the larger distances involved in comparison with the repulsive contribution. The repulsive contribution happens at very short distances (<3 Å) and it is due to the Pauli exclusion principle, arising from the repulsion between the atomic electronic densities. The van der Waals interactions were modelled using a Lennard-Jones potential (equation 2.8 and figure 2.5).

$$V_{LJ} = 4\,\epsilon_{i,j} \sum_{pairs\ i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{6} \right] \qquad 2.8$$

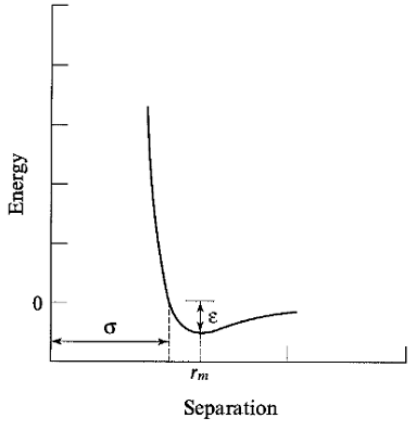


Figure 2.5- *Lennard-Jones potential between two atoms. The parameter σ is the separation distance for which the energy is zero, while ϵ is the minimum potential energy.*

The σ parameter (also called collision diameter) is the separation distance between two atoms for which the energy is zero. ϵ is the minimum potential energy for a pair of atoms. In equation 2.8 the repulsive term decays with r^{-12}, while the attractive one decays with r^{-6}. Hence, at very short distances the repulsive term leads to an increase of the potential energy. As distance increases, a minimum of energy is reached. At even longer distances, the interactions tend to zero (figure 2.5). Force-fields contain lists with σ and ϵ for each pair of atoms of the same type. In order to obtain parameters for interactions among different atom types, combination rules are used [257]. Additionally, in order to speed-up calculations, cut-offs are used, i.e, interactions are considered negligible beyond a certain distance.

The electrostatic properties of molecules are often represented by point charges usually centred on each nucleus: the partial atomic charges. The sum of the partial atomic charges is equal to the total charge of the molecule. The electrostatic interactions between atomic charges are modelled using a Coulombic potential (equation 2.9 and figure 2.6).

$$V_C = \sum_{pairs\ i,j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{i,j}} \qquad 2.9$$

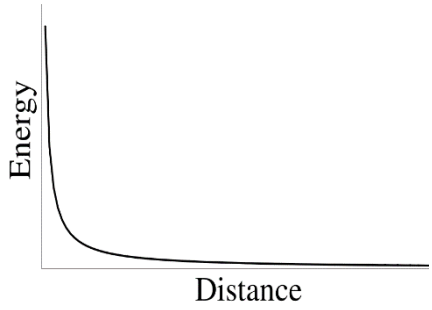


Figure 2.6 – *Representation of the Coulombic potential for particles with the same sign charge.*

In equation 2.9 q_i and q_j are the point charges of atoms i and j, respectively, r_{ij} is the distance between the two charges, ϵ_r is the permittivity in the considered medium and ϵ_0 is the electric permittivity in vacuum. The Coulomb potential decays at 1/r distance, which means that atoms will interact at long distances, typically distances greater than half the length of the simulation box. Increasing the system size is not a feasible solution because it will increase the simulation time by N^2 [260]. A plain cut-off for truncating these interactions will lead to artefacts in calculations. Therefore, other strategies have been developed to circumvent this problem. In this work the Particle Mesh Ewald (PME) method[261] was used to treat long-range interactions.

In the PME method the Coulomb potential is split in two parts: a short-range (within a pre-determined cut-off distance) and a long range one (beyond the cutoff distance). The short-range contribution is evaluated in real space (equation 2.9), while the long-range one is evaluated using a particle mesh approach [261]. The long-range contribution is evaluated by mapping the charges of the system on a grid and solving the Poisson equation using fast Fourier transforms. The electrostatic interactions will also be calculated through all the periodic images of the system. Furthermore, it is required that the system is neutral for the equations to converge.

2.3-Molecular dynamics simulation

After having a molecular description of a system of interest, it is possible to simulate atomic motion. Molecules are not static entities, they rather adopt a series of conformations and are always in constant motion due to thermal fluctuations and collisions with solvent molecules [262]. These conformations can be characteristic of a state of a protein and indicative of its function. Some molecules even need to suffer conformational changes in order to fulfil a specific role. The whole range of conformations that can be adopted by a molecule is called the phase space. A molecular system with *N* atoms can explore a 6*N* dimensional phase space, that arises from the combination of 3*N* positions and 3*N* momenta. For biological systems this results in a high dimensional space that cannot be fully explored. It is only possible to sample a statistically representative amount of conformations of the system (following a Boltzmann distribution, if sampling equilibrium): an *ensemble* [263]. The ergodic hypothesis states that a time average of some measurement is equal to its *ensemble* average. Therefore, measures done in the timespan of a simulation can be averaged and assumed that are equivalent to an *ensemble* average of the phase space [263].

Molecular dynamics simulations are one of the existent ways to obtain an *ensemble.* Alternative methodologies include Monte-Carlo sampling, for example. In a molecular dynamics simulation the motion of particles is calculated by solving Newton's equations of motion. The second law of Newton states that the force of a particle i is proportional to the product of its mass (m_i) and its acceleration (a_i) (equation 2.10).

$$F_i = m_i a_i \quad i = 1 \ldots N \tag{2.10}$$

Furthermore, it also possible to relate the potential energy of each particle with the force, by calculating the negative of derivative of the potential energy with respect to positions (equation 2.11).

$$F_i = -\frac{dV}{dr_i} \quad 2.11$$

In a molecular dynamics simulation, the forces are calculated using the potential energy of the system, given by the forcefield terms and the position of the particles (equation 2.11). Therefore, computed force also includes contributions from the surrounding bonded and non-bonded atoms.

Afterwards, the new positions will be calculated by integrating equation 2.12. This calculation is done in several timesteps using finite difference methods. These algorithms are called the MD integrators.

$$F_i = m_i \frac{d^2 r_i}{dt^2} \quad 2.12$$

The MD integrator used in this work was the leap-frog algorithm [264]. In the leap-frog algorithm the following equations are solved:

$$v_i\left(t + \frac{\Delta t}{2}\right) = v_i\left(t - \frac{\Delta t}{2}\right) + \Delta t \frac{F_i(t)}{m_i} \quad 2.13$$

$$r_i(t + \Delta t) = r_i(t) + v_i\left(t + \frac{\Delta t}{2}\right)\Delta t \quad 2.14$$

The velocities at $t + \frac{\Delta t}{2}$ are calculated from the forces at time t (equation 2.13). These velocities are then used to obtain positions at time $t + \Delta t$, which will be again used to calculate new forces at the new time $t + \Delta t$ (equation 2.14). In this way, molecular dynamics simulations are considered to be deterministic. The velocities at time t can be obtained by equation 2.15:

$$v_i(t) = \frac{1}{2}\left[v_i\left(t + \frac{1}{2}\Delta t\right) + v_i\left(t - \frac{1}{2}\Delta t\right)\right] \quad 2.15$$

In this method the calculation of velocities is lagged from the calculation of positions. Its main advantages are the increased numerical precision, efficiency and explicit estimation of velocity in equation 2.13[246]. Ideally, an integrator should preserve the phase-space volume and conserve energy [252].

In an atomistic molecular dynamics simulation, the timestep (Δt) used is about 1 to 2 fs. The timestep is also an important choice, as it should be an order of magnitude less than the smallest fluctuation of the system and it should be the largest possible for speeding up calculations. As

previously mentioned in section 2.2.1, the usage of constraints allows to increase the timestep. Other integration algorithms exist, such as the Verlet algorithm [265].

In the end of a molecular dynamics simulation, a trajectory is obtained with the time evolution of a system.

There are also non-sampling methods, such as the minimization algorithms like steepest descent or the conjugate gradient algorithm. These methods are used to achieve minimum energy structures, by calculating the derivative of the potential energy with respect to the system coordinates. This is an iterative process that stops when a certain number of steps is reached or when the energy (or force) falls below a certain threshold[257].

2.3.1-Periodic boundary conditions

The molecular models of biological systems can have variable sizes, depending on the nature of the molecule and its surroundings. Nonetheless, in order to mimic a biological bulk environment, periodic boundary conditions (PBC) can be applied [266]. This algorithm replicates the system in all three dimensions. As a result, multiple copies of system are emulated around the original cell. The particles of the original cell now interact with their periodic images as well as between themselves. While this allows the accurate calculation of bulk properties, it generates artifacts related to the increased artificial periodicity of the system and eliminates the conservation of the angular momentum. When using PBC, only the nearest image of each particle is considered for short-range interactions (such as Lennard-Jones interactions and the short-range part of the electrostatic interactions)[257], also known as the minimum image convention.

2.3.2-Temperature and pressure control

In a vacuum simulation, the total energy of a system is conserved. This assumes an isolated system and the sampling of a microcanonical *ensemble* (NVE) in which energy, volume and the number of particles remain unchanged[252]. Nonetheless, in biological conditions, there is exchange of energy and matter with the surroundings, as well as conditions of constant temperature and pressure. In most molecular simulation studies, it suffices to keep temperature and pressure constant to emulate biological conditions. Thermostats and barostats allow to maintain constant temperature and pressure, respectively. This allows to simulate an ensemble in the NPT conditions, giving the correct average temperature and the right amount of fluctuations. Thermostats act by changing the kinetic energy of the particles, which is given by:

$$K = \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2 \qquad 2.16$$

The kinetic energy of the particles is related to temperature via the equipartition theorem: each degree of freedom of a molecule of a system in thermal equilibrium has the energy of $\frac{kT}{2}$. Considering the three translational degrees of freedom it is possible to say the energy of a system can be given by equation 2.17. Equation 2.18 directly relates temperature with the velocity of the particles[247].

$$K = \frac{3}{2} N k_B T \qquad 2.17$$

$$\frac{3}{2} N k_B T = \left\langle \sum_{\mathrm{i=1}}^{\mathrm{N}} \frac{1}{2}\, m_i\, v_i^2 \right\rangle \qquad 2.18$$

The thermostats used in this work were the Berendsen thermostat and the modified velocity rescale thermostat[267,268]. In the Berendsen thermostat there is a weak coupling of the system to a heat bath in which the velocities are modified to match the target kinetic energy[267].

$$\frac{dT_t}{dt} = \frac{T_0 - T_t}{\tau} \qquad 2.19$$

In equation 2.19 T_0 is the target (heat bath) temperature, while T_t is the instantaneous temperature and τ is the coupling constant. The velocities are rescaled by a λ factor given by (equation 2.20):

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T}\left(\frac{T_0}{T} - 1\right)\right]^{1/2} \qquad 2.20$$

The Berendsen thermostat does not generate a thermodynamically correct ensemble because it does not generate the correct kinetic energy distribution and it can lead to simulation artifacts[269]. It was used in the equilibration phase to produce a relaxed system conformation, because it eliminates fluctuations faster.

In the same spirit of the Berendsen thermostat, the modified velocity rescale thermostat also performs the rescaling of the velocities, but it tries to achieve kinetic energy values that were previously chosen in a stochastic way from the kinetic energy distribution imposed by the canonical ensemble (NVT)[268]. In this case, the rescaling factor is different, and it is given by equation 2.21, where K_0 is the target kinetic energy drawn from the canonical equilibrium

function for the kinetic energy (equation 2.22). K_0 is the total kinetic energy of the system and N_{DOF} the number of degrees of freedom.

$$\lambda = \sqrt{\frac{K_0}{K}} \qquad 2.21$$

$$P(K_0) \propto K_0^{(N_{DOF}/2-1)} e^{-\beta K_0} \qquad 2.22$$

Regarding pressure control, barostats are the algorithms responsible for keeping a constant pressure. In a simulation, pressure is calculated via the virial theorem (equation 2.23)[246]:

$$W = -3\mathrm{PV} + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{dV(r_{ij})}{dr_{ij}} r_{ij} = -3\mathrm{Nk_B T} \qquad 2.23$$

$$P = \frac{1}{\mathrm{V}} \left[\mathrm{Nk_B T} - \frac{1}{3} \sum_{\mathrm{i=1}}^{\mathrm{N}} \sum_{\mathrm{j=i+1}}^{\mathrm{N}} \mathrm{r_{ij} f_{ij}} \right] \qquad 2.24$$

The virial W is the expectation value of the sum of the products between the coordinates of the particles and the forces acting on them (equation 2.23). The forces are given by the derivative of the potential energy to the distance between all pairs of particles on a system, plus the ideal gas part ($-3PV$). The virial theorem states that virial has to be equal to $-3Nk_BT$. The pressure (equation 2.24) can be deduced from equation 2.23.

In this work, the Berendsen and the Parrinello-Rahman barostats were used to control pressure[267,270]. The Berendsen barostat is analogue to the Berendsen thermostat previously mentioned. The system is also coupled to a pressure bath in which the volume is scaled by a scaling factor. Once more, it cannot be guaranteed whether Berendsen thermostats is sampling from any ensemble at constant pressure. Nevertheless, it is useful for the equilibration of a system. The Parrinello-Rahman barostat allows the simulation cell not only to change its size, but also to change its shape anisotropically[270].

$$\frac{db^2}{dt^2} = V\boldsymbol{A^{-1}} b'^{-1} \left(P - P_{ref}\right) \qquad 2.25$$

$$\boldsymbol{A^{-1}} = \frac{4\pi^2 \beta}{3\tau_p^2 L} \qquad 2.26$$

In equation 2.25, b is the matrix containing the box vectors, V is the box volume, P and P_{ref} the instantaneous and reference pressures, respectively and $\boldsymbol{A^{-1}}$ is the inverse mass parameter matrix that determines the strength of coupling and the deformation of the box and b is a matrix containing the box vectors. $\boldsymbol{A^{-1}}$ can be calculated using the isothermal compressibility β, the pressure time constant τ_p and L is the largest box matrix element (equation 2.26). In general, the Parrinello-Rahman barostat takes longer to reach equilibrium in comparison with the Berendsen barostat, yet it allows the obtention of thermodynamically correct ensembles.

2.4-Statistical mechanics

2.4.1-The ergodic hypothesis

Statistical mechanics is the theory that links the individual molecular properties and macroscopic properties measured in bulk environment. As previously explained, molecular dynamics simulations allow to obtain an *ensemble* of conformations, which is only a large set of conformations of our system under selected conditions of pressure and temperature. The ergodic hypothesis states that the time average of a variable in a system of interest is equal to its *ensemble* average over a large number of conformations representative of the system[271]. Therefore, this means that the arithmetic average of a property B over time is equal to its ensemble average (equation 2.27).

$$B_{time} = \frac{\sum_{i=1}^{N} B_i \dots B_n}{N} = \langle B_{ensemble} \rangle \qquad 2.27$$

The ensemble average is dependent of the visited phase space. Ideally, a simulation would visit the entire phase space accessible and an ergodic system should contemplate the whole phase space[271]. The *ensemble* average can also be written as the probability of occurrence of each microstate of interest, in which $\mathrm{P(s)}$ is the probability of occurrence of each microstate s and $\mathrm{B(s)}$ is the respective value of B (equation 2.28).

$$\langle B \rangle = P(s)B(s) \qquad 2.28$$

$\mathrm{P(s)}$ can either be interpreted as the fraction of systems of the state s at any time or the fraction of time spent in state s by a system.

2.4.2-Ensembles in simulation

Using statistical mechanics, it is possible to describe ensembles using probabilities and populations.

The Boltzmann distribution gives the probability of finding a system in a state as a function of the energy and temperature (equation 2.29), in which β is the inverse thermodynamic temperature, E_i is the energy of the corresponding state and k_B the Boltzmann constant (equation 2.30)[262]:

$$P_i = \frac{e^{-\beta E_i}}{\sum_i e^{-\beta E_i}} \quad 2.29$$

$$\beta = \frac{1}{k_B T} \quad 2.30$$

The denominator of equation 2.29 is the sum of all states and it is called the partition function. The partition function can also be deduced from the coordinates and momenta of particles (equation 2.31), allowing to relate positions and velocities obtained in simulations to the probabilities of any state. E_K is the kinetic energy, E_P the potential energy, p the momenta of the particles, r the positions and N the number of particles.

$$Q = \frac{1}{N!\,h^{3N}} \int dp\, e^{(-\beta E_K)} \int dr\, e^{(-\beta E_P)} \quad 2.31$$

Also, the equation 2.29 is valid for a canonical *ensemble* (NVT). This equation reveals that a state with low energy is more likely than a state with high energy. In conditions of constant temperature and pressure (NPT) an isothermal-isobaric *ensemble* is sampled, in that case equation 2.29 is altered to equation 2.32 in which P is the pressure and V the volume [271]. Equation 2.33 shows the new partition function that accounts with the temperature and pressure.

$$P_i = \frac{e^{-\beta(E_i + PV_i)}}{\Delta(P,T)} \quad 2.32$$

$$\Delta(\mathrm{P},\mathrm{T}) = \sum_i e^{-\beta(E_i + PV_i)} \quad 2.33$$

Other ensembles exist with other kinds of thermodynamic boundaries[272]. It is important to refer that these expressions sample equilibrium ensembles. Nonetheless, in some cases, it might be desirable to bias the state distribution to enhance the sampling of some states of interest. In

this work, non-equilibrium techniques such as steered molecular dynamics, slow-growth or approaches as umbrella sampling were used to do so.

2.5-Free energy calculations

Free energy is a thermodynamic quantity that indicates the position and direction of a system towards the equilibrium state, i.e, it indicates what will be the spontaneous direction of change experienced by the system to reach an equilibrium state, and how far it is from it[247]. It is a state function, depending only on the initial and final states and it is related with both enthalpy and entropy[247]. In biological systems, free energy governs the direction of processes, such as the direction of chemical reactions, molecular transport across concentration gradients, ligand binding, molecular partition in membranes, among others. In a system at *NPT ensemble*, the free energy is called the Gibbs free energy (G), while at *NVT* conditions, it is called the Helmholtz free energy (A). For the sake of simplicity, the following expressions will only refer to the *NVT* ensemble. Using statistical mechanics knowledge, we can express the free energy as a function of the partition function (Q)(equation 2.34):

$$A = -\frac{1}{\beta} lnQ \qquad 2.34$$

A refers to an absolute free energy, but in practice it is impossible to calculate absolute free energies, because in most cases is impossible to evaluate the complete partition function for a molecular system. Therefore, most molecular processes have to be described by relative free-energies (equation 2.35), in which two states are sampled and the free-energy difference is calculated, using a ratio of two states:

$$\Delta A = A_1 - A_0 = -\frac{1}{\beta} ln \frac{Q_1}{Q_0} \qquad 2.35$$

The different states can be different forms of a protein, such as a mutant and a wild-type or different values in a macroscopic variable such as the temperature. In fact, most methods for free-energy calculations are dedicated to sample the ratio of states, rather than sampling two partition functions[273].

The free-energy perturbation methodology, also called exponential averaging (equation 2.36) is one of the simplest techniques to calculate free-energy differences. A free-energy difference is given by the difference between the potential energy functions of two states (in this case 0 and 1)[274]. ΔA is estimated from a simulation of only one of the states. This method requires enough sampling to lead to an accurate estimate of the free energy.

$$\Delta A = -\frac{1}{\beta} ln\langle exp - \beta[U_1(x) - U_0(x)]\rangle_0 \quad 2.36$$

Despite its simplicity, it is often difficult to achieve convergency with this method, as it shows poor phase space overlap[275]. Nonetheless, it may be useful if the two states are close to each other or if the end state is a sub-state of the first one[276].

In thermodynamic integration (TI), an external parameter λ is coupled to an order parameter or to the Hamiltonian H (which is a function that represents the total energy of a system). In this approach, the derivative of the free energy ΔA with respect to λ is calculated, and ΔA is obtained by integration (equation 2.37):

$$\Delta A = \int_0^1 \left\langle \frac{dH(\lambda)}{d\lambda} \right\rangle_\lambda d\,\lambda \quad 2.37$$

The λ parameter varies between 0 and 1, in which 0 refers to the initial state, and 1 is the final one. Generally, the transition between both states is executed in discrete intervals, using a staging approach, in which several simulations are run, each with a different λ [273]. Nevertheless, a variation of this methodology was used in this work, in which λ changes in a slow and continuous form, through the course of one simulation only: the slow-growth methodology [277,278]. In this work, slow-growth was used only in a qualitative way to generate initial conformations for the equilibrium simulations of ABC transporters in the post-hydrolysis state. Slow-growth implies the usage of non-equilibrium simulations and is often susceptible to hysteresis and energy drift [279]. However, some approaches were suggested in order to mitigate these problems[280]. Very often, thermodynamic cycles are used in conjunction with thermodynamic integration to calculate free-energy differences. A thermodynamic cycle is a series of reversible transformations in which an initial state is restored, and the total energy is zero. Most times, the transformations performed do not have a direct physical correspondence, but they can be performed theoretically, such as annihilations or the creation of chemical groups.

Finally, free energy differences can also be regarded as a ratio between probability density functions (equation 2.38).

$$\Delta A = -\frac{1}{\beta} ln \frac{P_1}{P_0} \quad 2.38$$

These probability density functions are generally histograms of properties measured over the course of several simulations. The histograms can refer to any property such as energy or a

reaction coordinate. Each bin of the histogram embodies the number of visits to a state. It is easy to make an analogy to the partition function, in which P_1 can be the probability of any state and P_0 the range of all the visited states. Therefore, it becomes important to ensure good sampling. This strategy is commonly used in histogram-based methods to obtain free-energy landscapes and potentials of mean force, such as umbrella sampling, metadynamics and others.

2.5.1-Potential of mean force (PMF)

The potential of mean force (PMF) consists on the free-energy projected over a specific subset of the system coordinates and integrated over all the other coordinates. In this way, the free-energy is mapped on a hyper-surface as a function of any order parameter, such as a reaction coordinate[281]. Unlike free-energy (ΔG), a PMF is not a state function. In a similar spirit to equation 2.38, a PMF along a reaction coordinate ξ is given by equation 2.39[282]:

$$PMF = -\frac{1}{\beta} lnP(\xi) + constant \qquad 2.39$$

In which the probability $P(\xi)$ is obtained by equation 2.40, ξ is the reaction coordinate in function of the coordinates r. $P(\xi)$ is the probability of finding a system in a certain range of the reaction coordinate, U is the internal energy, N the number of particles and β is the inverse of the thermodynamic temperature ($1/k_B T$).

$$P(\xi) = \frac{\int \delta[\xi(r) - \xi] exp\,[(-\beta U) d^N r]}{\int exp\,[(-\beta U) d^N r]} \qquad 2.40$$

The PMF can also be given by the integration of the forces along the reaction coordinate ξ over a certain range (equation 2.41). The force can be given by the ensemble average of the derivative of the potential energy with respect to the reaction coordinate ξ.

$$PMF = \int_{\xi_a}^{\xi_b} \left\langle \vec{F} \right\rangle_{\xi} d\xi = \int_{\xi_a}^{\xi_b} \left\langle \frac{\partial V}{\partial \xi} \right\rangle_{\xi} \mathrm{d}\,\xi \qquad 2.41$$

The study of the PMF over a reaction coordinate is often useful to study ligand dissociation, molecular transport, conformational changes, among other phenomena. In this work, the computation of a PMF profile was used to study the transport of maltose in different states of the maltose importer.

2.5.2-Umbrella sampling

In order to obtain a PMF profile, it is necessary to have extensive sampling along the reaction coordinate ξ, allowing the calculation of the correct probabilities. In most cases, a plain molecular dynamics simulation is not enough to attain adequate sampling of the entire range of the reaction coordinate. Even if the simulation is extended for a considerable amount of time, it is very likely that the system gets stuck in a local energy minimum (of the reaction coordinate landscape) and it may never be possible to overcome energetic barriers and sample the entire span of ξ. Therefore, enhanced sampling techniques are needed to cover the entire range of the reaction coordinate. One of these techniques is umbrella sampling[283].

The umbrella sampling technique uses two strategies: a stratification strategy in which the reaction coordinate is divided in strata, also known as windows, and an importance sampling strategy in which the sampling is biased in each window, in order to increase the number of conformations at that point of the reaction coordinate. Afterwards, it is necessary to remove the bias in each window using an appropriate reweighing method. Generally, a biasing potential is used to enforce sampling in the desired region. Harmonic potentials are often used (equation 2.42) in which K is the force constant, ξ the instantaneous value of the reaction coordinate and ξ_i the equilibrium value of the reaction coordinate[284].

$$w(\xi) = \frac{1}{2}K(\xi - \xi_i)^2 \qquad 2.42$$

Figure 2.7 describes the steps necessary to obtain a free-energy profile using the umbrella sampling approach: first it is necessary to generate conformations along a reaction coordinate. Afterwards, umbrella sampling simulations are done in each conformation (generally using an harmonic potential), which will allow to obtain probability distributions of the molecule across the reaction coordinate. These distributions are then mapped on a grid and are reweighted and free-energy is then calculated.

The unbiased PMF will be given by equation 2.43, in which $w(\xi)$ is the constraining potential:

$$PMF = -\frac{1}{\beta} lnP(\xi) - w(\xi) + constant \qquad 2.43$$

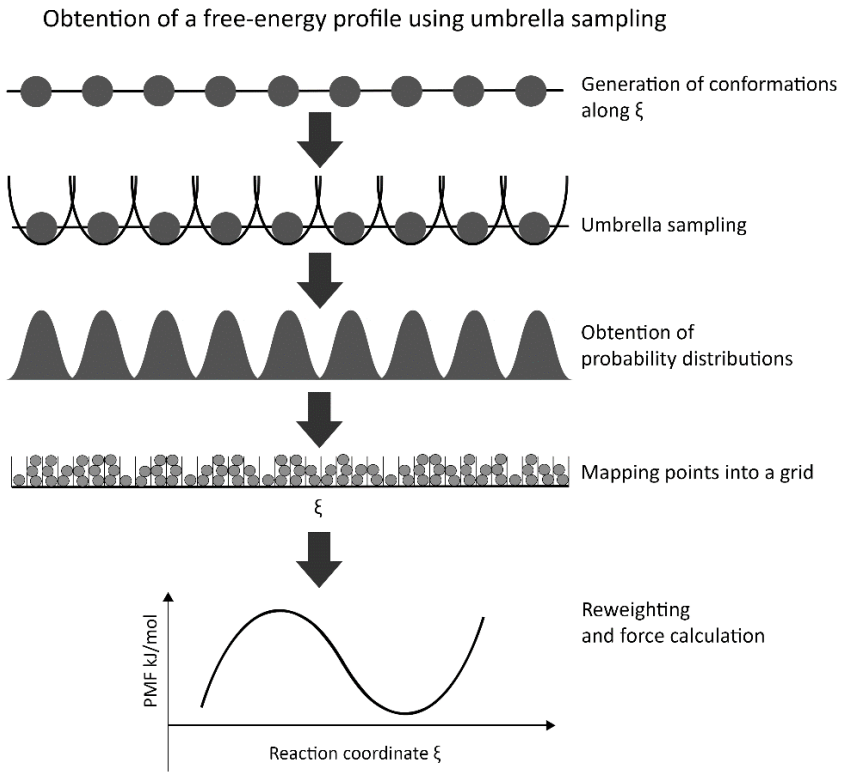


Figure 2.7- *Steps to obtain a PMF profile from umbrella sampling simulations. First, it is necessary to generate conformations in the desired range of the reaction coordinate. Afterwards, the reaction coordinate is divided in windows and in each window, a potential (generally harmonic) is applied in order to increase sampling in these regions. A probability distribution is obtained from each window. Later, these points are plotted into a grid and some method is used to reweight the points, calculate the force and obtain a PMF.*

2.5.3-Methods to obtain a PMF from umbrella sampling simulations

As previously described, when doing umbrella sampling, the values of an order parameter are mapped onto a grid (generally 1D or two 2D) and each bin of the grid indicates a value of the parameter for that reaction coordinate. The value of an order parameter in each bin is regarded as the number of visits (or conformations) of a system to that state. After that, it is necessary to proceed to the reweighting of all points in order to obtain the correct probabilities for each state.

The underlying assumptions for this procedure are that the partition function can be expressed as a sum of all the macroscopic states of a system (macrostates), and that each macrostate will be proportional to the probability to which these states will be visited in the simulated

ensemble. Upon the mapping of any order parameter into a grid, it is possible to build a total histogram (covering the full range of the order parameter) that reflects the probability with which macrostate is visited[273].

The reweighting approach consists in obtaining the probabilities of the microstates of an *ensemble* using a total histogram by combining and normalizing all the distributions sampled during umbrella sampling. The correct probability of the sampled values of ξ in each umbrella window is obtained by multiplying each distribution by a weighting factor – the microstate weight factors[273]. In this work, the method used to combine and reweight the umbrella windows to obtain a PMF profile was umbrella integration (UI)[285,286] . For sake of simplicity the following equations will refer to the NVT *ensemble*.

In the UI method, rather than calculating the PMF in a direct manner, based on equation 2.43 (obtaining ΔA), the $\frac{\partial A}{\partial \xi}$ is calculated, just like in thermodynamic integration (equation 2.44). For obtaining $\frac{\partial A}{\partial \xi}$, the sum of the unbiased free-energy derivative $\frac{\partial A_i^u}{\partial \xi}$ is calculated for all windows, that are weighted by $p_i(\xi)$ (equation 2.43). This calculation is done for each bin of the grid.

$$PMF = \int_{\xi_a}^{\xi_b} \frac{\partial A}{\partial \xi} d\,\xi \qquad 2.44$$

$$\frac{\partial A(\xi)}{\partial \xi}|_{\xi_{\text{bin}}} = \sum_{i}^{windows} p_i(\xi_{bin}) \left(\frac{\partial A_i^u(\xi)}{\partial \xi}\right)_{\xi_{bin}} \qquad 2.45$$

For each window, $\frac{\partial A_i^u}{\partial \xi}$ is given by $P_i^b(\xi)$ which is the biased distribution of ξ obtained by umbrella sampling, approximated by a normal distribution. As a consequence, $\frac{\partial A_i^u}{\partial \xi}$ is given by equation 2.46, in which $\left(\sigma_i^b\right)^2$ is the variance of each umbrella window, $\bar{\xi}_i^b$ is the corresponding mean of ξ in each window and K is the force constant used.

$$\frac{\partial A_i^u}{\partial \xi} = \frac{1}{\beta} \frac{\xi - \bar{\xi}_i^b}{\left(\sigma_i^b\right)^2} - K(\xi - \xi_i) \qquad 2.46$$

The normalized weights are given by (equation 2.47):

$$p_i(\xi) = \frac{N_i P_i^b(\xi)}{\sum_i N_i \, P_i^b(\xi)} \qquad 2.47$$

Another popular method for the obtention of PMF profiles is the weighted histogram analysis method (WHAM)[287]. In the WHAM method, the calculation of the PMF is based on the equation 2.43, using directly the unbiased distributions of ξ. The most important differences between the two methods are the weights given to the points and the treatment of the biased distributions: in UI they are approximated to normal distributions, while in WHAM they are not. A more detailed explanation of the differences between the two methods can be found in Kästner *et al*[284].

2.6-Comparative modelling

The starting point for a molecular dynamics simulation study is a molecular structure, that is generally conveyed as a text file with three-dimensional coordinates, along with other descriptors. The molecular structures are obtained by experimental techniques such as x-ray crystallography, cryogenic electron microscopy (cryo-EM) and nuclear magnetic resonance (NMR). These structures and experimental data are generally deposited in databases such as the Protein Data Bank (PDB) and the electron microscopy databank (EMDB).

Nonetheless, it happens that a desired structure is not available in any database, very often due to technical difficulties in its obtention, such as problems in expression and purification and others. In those cases, comparative modelling presents itself as a solution.

Comparative modelling is an empirical method that uses the structure of a known protein (that will serve as template) with a similar sequence to create a model of the desired protein. This is possible because proteins with similar sequences have similar structures [288]. A good template protein should have at least 40% sequence identity with the desired target [289].

In this work, the software package MODELLER was used to create homology models[289]. MODELLER performs comparative modelling by satisfaction of spatial restraints.

The first step in creating a homology model is to perform an alignment between the sequences of the protein to be modelled and one or several template structures, that show a high degree of sequence identity with the target, as previously mentioned. Afterwards, a series of spatial restraints are calculated using the template 3D structures. These spatial restraints are conditional probability density functions whose functional form was derived from a statistical analysis of a database of more than 400 protein structures[290]. The conditional probability density functions reflect correlations between protein features that are relevant for protein structure such as mainchain and sidechain dihedrals, Cα-Cα distances, residue types, among others. Afterwards, the spatial restraints and energy terms based on the CHARMM forcefield[291] are combined into a final function: the objective function. The energy terms allow to enforce a correct stereochemistry. The objective function is optimized and minimized using the variable target function method[292] which applies conjugate gradients and molecular dynamics with simulated annealing[293].

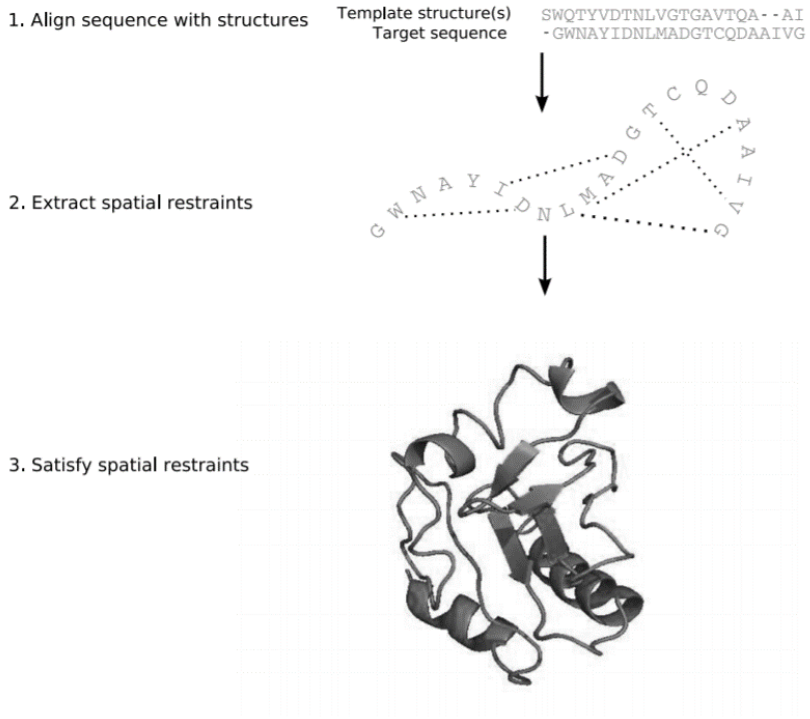


Figure 2.8- *Steps in the construction of a homology model, adapted from the MODELLER manual version 9v6. First it is necessary to align the sequence of the target with the template structures, then the spatial restraints are extracted from the template structures and finally, a model obeying these restraints is built and optimized.*

2.7-Determination of protonation states of protein residues

Of the existing aminoacids that compose protein structures, some (lysine, histidine, arginine, glutamate and aspartate) possess ionizable sidechains, whose protonation state varies with the surrounding chemical environment and pH. The protonation state of these residues can strongly influence structure and function. Therefore, knowledge about the protein protonation states becomes of extreme importance.

The protonation state of a residue within a protein can be different from its state in solution because they suffer the influence of the chemical environment created by the remaining residues of the protein. For instance, in water, a charged group is stabilized by the moving water dipoles around it, and it has a higher tendency to remain charged. In a protein environment, in general, there is a limited reorganization of the surrounding groups and charges are less stabilized. Additionally, the protonation states can also be affected directly by other charged groups, through long and short-ranged electrostatic interactions.

In this work we have used the PETIT package to perform protonation calculations along with Monte Carlo sampling of tautomers[294]. The electrostatic energies between protanatable groups were determined using MEAD[295].

The free-energy of protonation of a titrable site can be described by a thermodynamic cycle linking the deprotonated and protonated state, such as the one in figure 2.9.

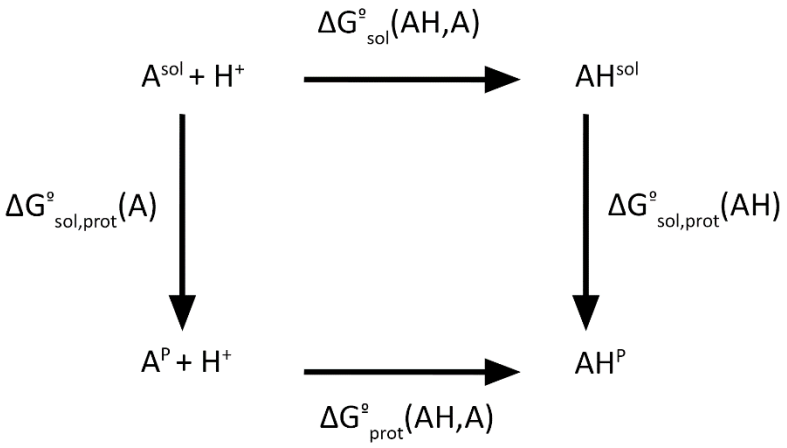


Figure 2.9- *Thermodynamic cycle used for calculating the protonation free energy of a protonatable group in a protein. The top part of the cycle refers to the calculation of the protonation free-energy in a model compound in solution, while the left and right sides account for the free-energy difference related with the free energy difference corresponding to the transfer of the deprotonated and protonated groups from the solution to the protein.*

Considering a single protonatable site, its free-energy of protonation is given by equation 2.48, in which n_i and n_j are the proton occupancy of the sites i and j , pK_i^{int} is the intrinsic pK_a of

the site i, i.e, the pK_a of the protonatable site when all other sites are kept neutral. z_i^0 and z_j^0 are the charges of the sites i and j and W_{ij} is the electrostatic energy between the two sites i and j, when i is different from j [296].

$$\Delta G_{\mathrm{P}}^{0} = -2.3RT \sum_{i} n_i pK_i^{int} + \frac{1}{2} \sum_{i} \sum_{j<i} \left(n_i n_j + n_i z_j^0 + n_j z_i^0\right) W_{ij} \qquad 2.48$$

The protonation states are encoded in a vector $n = (n_1, n_2, n_3, \ldots, n_i)$ in which n_i is 0 if the site is neutral or 1 if it is charged. A protein with *N* ionizable sites can have 2^N protonation states, given that they can either protonated or deprotonated. W_{ij} can be written as the sum of all combinations of the protonation states between the two sites:

$$W_{ij} = w_{ij}(1,1) - w_{ij}(1,0) - w_{ji}(0,1) + w_{ij}(0,0) \qquad 2.49$$

The pK_i^{int} is given by equation 2.50, in which $pK_{mod,i}$ is the pK_a for that site in solution obtained from experimental titrations of model compounds, $\Delta G_{sol,prot}(A)$ is the free energy difference corresponding to the transfer of the deprotonated group from the solution to the protein, while $\Delta G_{sol,prot}(HA)$ is the free-energy difference corresponding to the transfer of the protonated group from solution to the protein.

$$pK_i^{int} = pK_{mod,i} + \frac{\Delta \mathrm{G}_{\mathrm{sol,prot}}(\mathrm{A}) \; - \; \Delta \mathrm{G}_{\mathrm{sol,prot}}(\mathrm{AH})}{2.303 \mathrm{k_B T}} \qquad 2.50$$

With the values of ΔG_p^0 we can finally obtain the probability of a given protein protonation state m exists at a certain pH is given by equation 2.51.

$$p(m) = \frac{e^{-2.3z(m)pH - \Delta G/k_B T}}{\sum_{m'} e^{-2.3z(m')pH - \Delta G/k_B T}} \qquad 2.51$$

The free-energies $W_{ij,}$ $\Delta \mathrm{G}_{\mathrm{sol,prot}}(\mathrm{A})$ and $\Delta \mathrm{G}_{\mathrm{sol,prot}}(\mathrm{AH})$ can be calculated from the electrostatic potentials of each form by solving the linearized version of the Poisson-Boltzmann equation (equation 2.52) in which ϵ is the dielectric constant, k^2 is the Debye length, $\rho(r)$ is the charge density and $\phi(r)$ is the electrostatic potential. k is given by equation 2.53 in which e is the charge of the group r, I is the ionic strength, k_B is the Boltzmann constant and T the temperature[297].

$$\nabla.[\epsilon(r)\nabla\phi(r)] - \epsilon(r)k^2(r)\phi(r) = -4\pi\rho(r) \qquad 2.52$$

$$k = \begin{cases} \left(\frac{8\pi e^2 I}{\epsilon k_B T}\right)^{1/2} & \text{If } r \text{ is in the counterion region} \\ 0 & \text{Otherwise} \end{cases} \qquad 2.53$$

The Poisson-Boltzmann equation originates from Coulomb's law, but accounts for the inhomogeneity of the medium and the presence of ions and other charges in the environment. The Poisson-Boltzmann equation implies a macroscopic description of the system in which the behaviour of the system is described in average values, rather than instantaneous values, like in a microscopic description such as molecular mechanics. Therefore, it belongs to the group of continuum electrostatic (CE) methods[298].

In a CE model the nature of the chemical environment is embodied in a dielectric constant. The dielectric constant reflects the polarizability of the medium. The polarizability is the capability of inducing dipoles on other atoms. The aqueous medium shows a high dielectric value ($\varepsilon_s \approx 80$) not only due to the electronic polarizability of the water molecules, but also mainly to the dipolar rotation that arises from the high mobility of the water molecules. In contrast, a protein structure is more rigid, and most part of the polarizability comes electronic effects from the existing dipoles. Hence, a much lower dielectric value is attributed to the protein ($\varepsilon_P \approx 2$-20)[299].

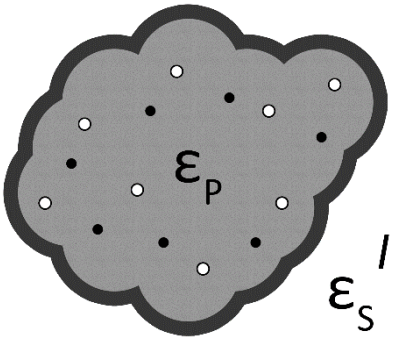


Figure 2.10- *Continuum electrostatics model of a protein in solution. The protein is modelled with a dielectric constant ϵ_P, while the solvent is modelled with a dielectric ϵ_S and ionic strength of I. The white dots on the protein structure are the actual titrable sites, while the black dots are the background charges. The dark gray area surrounding the protein is the ion exclusion region.*

In addition, the CE model also considers the charge density ρ, which consists in the partial atomic charges of the titrable sites and other atoms. Also, it is necessary to consider the presence of background charges and distinguish them from the titrable sites in the protein. The counterions in solution are modelled by the ionic strength I. An ion exclusion layer surrounding

the protein is also included to model the finite size of ions, blocking their effect at a distance lower than their atomic radius[299].

In practice, in order to estimate the electrostatic potential, the system is modelled as a 3D grid and the electrostatic potential is calculated at each point of the grid by solving the linear Poisson-Boltzmann equation applying a finite difference method. The electrostatic potential $\phi(r_i)$ is then used to obtain the electrostatic energy W, which as seen in equation 2.54 is dependent of the set of point charges $(q_1, q_2, \dots q_n)$ used to model each atom and its respective positions $(r_1, r_2, \dots, r_n)$. In a CE model the electrostatic energy can also be seen as the reversible work necessary to bring all charges from zero to their values. This electrostatic energy could be equivalent to a free-energy, because the configurational freedom of the system has been implicitly averaged in all the parameters. However, that is not the case because it would imply that the electrostatic free energy of a neutral protein would be zero. In addition, W does not take into account the hydrophobic effect, i.e, the apolar interaction of the protein with the solvent. Therefore, W is only the electrostatic contribution to the free-energy change[299].

$$W = \frac{1}{2}\sum_i q_i \phi(r_i) \qquad 2.54$$

Continuum models handle the protein structure as a rigid model, in which the sidechain flexibility cannot be modelled. Nonetheless, the protonatable sites often show multiple possible proton positions and orientations.

In order to overcome this problem, the different proton positions and orientations (tautomers) are modelled as separate entities. Each tautomeric site is converted in multiple pseudosites, corresponding to the individual tautomers, which are then titrated the methodology previously explained[300].

The calculation of the protonation probabilities of all the titratable sites, including the tautomeric ones, is done using Monte Carlo simulations in which a Metropolis criterion is used to determine the acceptance of the state[300]. In these calculations a titration curve is obtained and the midpoint pK_a is obtained. The midpoint pK_a is the pH value in which the residue has 50% probability to be protonated.

Chapter 3

Effect of the F508del mutation on the nucleotide binding domains of CFTR

This work was published in the following paper:

Abreu, B.; Lopes, E. F.; Oliveira, A. S. F.; Soares, C. M. The mutation F508del disturbs the dynamics of the nucleotide binding domains of CFTR before and after ATP hydrolysis, *Proteins Struct. Funct. Bioinforma.* **2019**, 88, 113

Contributions of the author of the present thesis to this work:

In this work I participated in the experimental design, performed all simulations and analyses, and wrote the manuscript.

3.1-Summary

The cystic fibrosis transmembrane conductance regulator (CFTR) channel is an ion channel responsible for chloride transport in epithelia and it belongs to the class of ABC transporters. The deletion of phenylalanine 508 (F508del) in CFTR is the most common mutation responsible for cystic fibrosis. Little is known about the effect of the mutation in the isolated nucleotide binding domains (NBDs), on dimer dynamics, ATP hydrolysis and even on nucleotide binding. Using molecular dynamics simulations of the human CFTR NBD dimer, we showed that F508del increases, in the pre-hydrolysis state, the inter-motif distance in both ATP binding sites (ABP) when ATP is bound. Additionally, a decrease in the number of catalytically competent conformations was observed in the presence of F508del. We used the subtraction technique to study the first 300ps after ATP hydrolysis in the catalytic competent site and found that the F508del dimer evidences lower conformational changes than the wild type. Using longer simulation times, the magnitude of the conformational changes in both forms increases. Nonetheless, the F508del dimer shows lower C-α RMSD values in comparison to the wild-type, on the F508del loop, on the residues surrounding the catalytic site and the portion of NBD2 adjacent to ABP1. These results provide evidence that F508del interferes with the NBD dynamics before and after ATP hydrolysis. These findings shed a new light on the effect of F508del on NBD dynamics and reveal a novel mechanism for the influence of F508del on CFTR.

3.2-Introduction

Cystic fibrosis is a genetic disease characterised by the formation of viscous secretions in epithelial tissues, leading to obstruction, inflammation, and infections[225]. It is caused by mutations in the gene that encodes for the cystic fibrosis transmembrane conductance regulator (CFTR)[301–303]. The CFTR channel is mainly responsible for the transport of Cl^- across the membrane, and it also regulates other channels[205]. For instance, it is responsible for the inhibition of sodium channels in epithelia and it is important for the production of an acidic environment in macrophages, allowing bactericidal activity[206,207]. This channel is a single-chain protein with 1480 residues that belongs to the class of ATP-binding cassette (ABC) transporters[110]. Like other ABC transporters, CFTR possesses two transmembrane (TM) domains and two nucleotide binding domains (NBD) which are responsible for ATP binding and hydrolysis (figure 3.1A). In the case of CFTR, there is also a regulatory domain, which can be phosphorylated by protein kinase A (PKA). The phosphorylation of this regulatory domain is essential for activity. In CFTR, the two NBDs are not identical, and they dimerize in a “head-to-tail” fashion to bind

ATP, forming the ATP binding pockets (ABP) (figure 3.1B). One pocket is catalytically competent and able to hydrolyse ATP (ATP binding pocket 2 - ABP2), having a canonical Walker A sequence (GSTGAGKTS), Walker B (LYLLDS) and the conserved ABC or signature sequence (LSGGQ)[214]. The other pocket is a degenerate site (ABP1-ATP binding pocket 1), which is unable to hydrolyse ATP and where some conserved residues of the structural motifs have been replaced, such as in the ABC motif (LSHGH), in the Walker A (GRTGSGKS) and the Walker B (KILLDE) [213,304]. CFTR possesses two exclusive segments, the regulatory insertion (RI) and the regulatory extension (RE)[208]. The RI is a loop in NBD1, which comprises residues 402 to 436, located between the first two β-sheets of NBD1. The RE is part of the C-terminal of NBD1, which spans approximately residues 638 to 672 and is considered to be part of the N-terminal of the R domain[209,210]. Both RI and RE contain additional phosphorylation sites[305].

Although there are various mutations that interfere in CFTR synthesis or function, the most common mutation in the Caucasian population is the deletion of phenylalanine 508 (F508del)[207]. This mutation causes flawed protein folding resulting in protein retention in the endoplasmic reticulum and its degradation by the proteosome. However, a small portion is able to reach the membrane, but displays an impaired function[227,228,306]. Additionally, the mutant CFTR present on the plasma membrane has a decreased half-life residency[229].

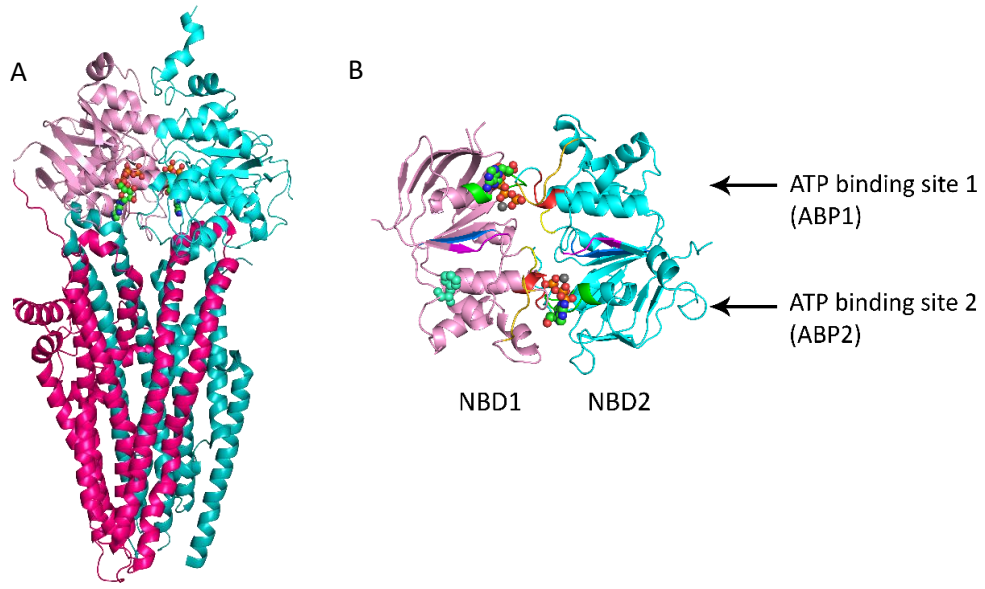


Figure 3.1- *A- Representation of full-length zebrafish CFTR (PDB code 5W81) with ATP bound. The upper part faces the intracellular space. The TMDs are colored in dark pink and teal, the NBD1 in pink and the NBD2 in cyan. The NBDs of this structure were used as template for our model. B- Bottom view of the simulated NBDs. The motifs are colored in the following way: the ABC motif in green, the Walker A motif in red, the Walker B in purple, the X-loop in gold, the D-loop in yellow and the Q-loop in dark blue. F508 is represented in light green spheres.*

X-ray structures of the F508del NBD1 reveal that this mutation causes small local changes in the conformation of the loop where it was supposed to be inserted[210]. However, biochemical studies report that this mutation also affects the gating and function of the full CFTR protein[307]. Some studies show that F508del impairs domain communication between NBD1 and the transmembrane domains[187,230].

X-ray studies on isolated NBD1 have reported that the impact of F508del was confined to increased fluctuations of the neighbouring residues 509 to 511[209]. Nonetheless, molecular simulation studies have revealed that F508del NBD1 in solution displays larger fluctuations and has tendency to form misfolded intermediates prone to aggregation[193,194]. Despite this information, little is known on the impact of F508del on the dynamics of a complete NBD dimer and on the conformational changes induced by ATP hydrolysis. Despite the detrimental impact of F508del in CFTR function, this protein is still able to have ATPase activity, which means it hydrolyses ATP[308,309].

Structures of the full length CFTR channel have only recently been resolved using cryo-EM. The available structures are from human, zebrafish and chicken CFTR[110,310,311,312]. The human CFTR structure is in a dephosphorylated form without nucleotides (PDB code 5UAK)[110]. There is also a zebrafish CFTR structure without nucleotides, with the R domain docked inside the intracellular vestibule (PDB code 5UAR)[310]. This structure is very similar to the previously referred human structure, although in the latter case the R domain was not modelled into the structure, despite the existing electron density. There is another zebrafish CFTR structure with the catalytic glutamate mutated, which possesses ATP in the NBDs and a phosphorylated R domain, allowing the NBDs to dimerize (PDB code 5W81)[311]. This latter structure was used as a template for creating a model of the human CFTR NBDs utilized in this work.

At the time of completion of this manuscript, a cryo-EM structure of the human phosphorylated CFTR was released (PDB code 6MSM)[107]. This structure corresponds to an open channel with an associated NBD dimer. Zebrafish and human CFTR share 55% sequence identity [311], which is considered a high degree of sequence similarity. The released human structure and the zebrafish structure used in this work (PDB codes 6MSM and 5W81) are also structurally similar [107]. Looking at the NBD dimer alone, which is our focus here, the zebrafish and human NBD sequences show an even higher identity of 64 %, and the C-α RMSD between the human NBDs obtained by cryo-EM and our model is 0.15 nm (for 439 aligned residues). These values are within typical C-α RMSD fluctuations obtained from MD simulations, evidencing the near equivalence of both experimental structures and, consequently, of our model.

In this work, the main goal is to investigate the effect of the F508del mutation on the closed NBD dimer before and upon ATP hydrolysis. To this end, molecular dynamics simulations of the

wild-type and mutant NBD dimers were performed in a closed state. First, the effect of F508del CFTR was studied on the NBDs dynamics before ATP hydrolysis. Afterwards, ATP hydrolysis was simulated in both mutant and wild-type forms. The impact of F508del on dimer dynamics, nucleotide binding and on the conformational changes after ATP hydrolysis was analysed.

3.3-Methods

3.3.1-System setup

A model of the CFTR NBDs was built by comparative modelling using MODELLER version 9.6[289]. The structure of the zebrafish CFTR in a closed NBD conformation was used as template (PBD code: 5W81)[311] for both mutant and wild-type systems. Although the full-length protein structure is available, our studies only contemplate the NBD dimer. This is mostly due for reasons of computational efficiency, since the NBD dimer is much smaller than the full-length protein. Nonetheless, the NBDs are considered to be essential for channel function in CFTR [112,313], they contain the machinery for ATP hydrolysis and are the structural site of the F508del mutation. Additionally, it has been shown that the features of the NBDs are key for the modulation of the CFTR channel. A study of Scott-Ward *et al* demonstrated that chimeric CFTR proteins with murine NBDs and human TMDs have different gating profiles than the fully human CFTR channel, displaying prolonged channel opening [314].

The different replicates for mutant and wild-type states were created generating different models using different initial seeds in MODELLER. The E1371Q mutation was reverted. Our model comprises residues 388 to 404, 437 to 636 and 1203 to 1445. The RI domain was not modelled. The C-terminal of residue 404 and the N-terminal of residue 437 were set as neutral. The protonation states of the protonatable residues of the protein were determined at pH 7.0 by performing Poisson-Boltzmann calculations using the package MEAD version 2.2.9 [315] and Metropolis Monte Carlo sampling of protonation states using PETIT [300]. According to the results of these calculations, all lysines and arginines were found to be protonated, all glutamates and aspartates were set as deprotonated. Histidine residues 484, 620 and 1402 were set as charged. Histidines 609, 1375 and 1350 were set to be neutral, protonated on Nε. Histidine 1348 is also neutral, but protonated on Nδ. The protonation states of the nucleotides ATP, ADP and the phosphate ion were the same used in Oliveira *et al* [161,171,316] and Damas *et al* [317], corresponding to a charge of -4, -3 and -2 respectively. The protonation states were determined using the model with the lowest value of the objective function of MODELLER, for each state (mutant or wild-type).

3.3.2-Setup of the molecular dynamics (MD) simulations

The MD simulations were performed using the GROMACS package version 5.0.7[318] and the force field used was GROMOS 54A7[256]. The nucleotides and phosphate ion parameters used were the same as used in Oliveira *et al* [161,316] and Damas *et al* [317]. The integration time step used was 2 fs. The systems were simulated using periodic boundary conditions and at constant temperature and pressure. Ions were added in order to neutralize the system. The temperature was set to 310K using a velocity-rescale heat bath[268], with a coupling constant of 0.1 ps with two coupling groups: one for the protein-nucleotides complex and another for the solvent and ions. The pressure was kept around 1 atm by isotropic coupling with a Parrinello-Rahman bath with a coupling constant of 2.0 ps and a compressibility of 4.6×10^{-5} bar^{-1} [270]. A cut-off of 1.0 nm was used in the calculation of Van der Waals interactions. Electrostatic interactions were treated using the particle mesh Ewald (PME)[261,319] method using a real-space cut-off of 1.0 nm. All neighbour lists were updated every 10 steps. All bonds were constrained to their equilibrium lengths using the LINCS[259] algorithm, except for the water molecules in which the SETTLE algorithm[258] was used to constrain its bonds. The SPC model[320] for water was used. The ATP-bound systems, either mutant or wild-type, were first solvated in a dodecahedral box and sodium ions were added in order to achieve electroneutrality. Afterwards, an energy minimization procedure was performed in two steps: the first step with position restraints of 1000 kJ mol^{-1} nm^{-1} on the all heavy atoms of the protein and ligands, followed by a second step with position restraints of 1000 kJ mol^{-1} nm^{-1} applied only on C-α atoms. The equilibration process was initiated with 100 ps of molecular dynamics at constant volume with position restraints of 1000 kJ mol^{-1} on all heavy atoms of the protein and ligands. Next, another 500 ps of simulation were performed at constant temperature and pressure with position restraints of 1000 kJ mol^{-1} on all C-α atoms of the protein. Finally, 500 ps of simulation were performed without any position restraints. During equilibration, temperature and pressure were controlled using the Berendsen thermostat and barostat, with coupling constants of 0.1 ps and 1 ps, respectively. Initially, 50 replicates of 150 ns each were made for the wild-type and another 50 replicates for the mutant. Nonetheless, when generating the ADP state, we noticed that not all replicates were suited to perform ATP hydrolysis at 20 ns, the intended time for performing hydrolysis. Therefore, more replicates of the ATP state were created to obtain 50 replicates of the ADP state, allowing reliable statistics to compare ATP and ADP states.

The average C-α RMSD along time showed that the last 130 ns correspond to the equilibrated portion of the simulation. Therefore, only this portion of the trajectory was used for analysis.

3.3.3-ATP hydrolysis using the slow-growth method

ATP hydrolysis was performed as previously described [161,171,316,317] in which a transformation from ATP to ADP and phosphate is performed. In order to select conformations suited for ATP hydrolysis, several analyses were carried using MDAnalysis[321,322]. The system Hamiltonian was coupled to a lambda parameter that varies from 0 to 1. This allows a smooth conversion of the molecules, with the aim of analysing the effect of the perturbation on the protein. Nonetheless, in this case we are not interested in calculating the free energy associated with this transformation. The phosphate molecule was created from the positions of the γ phosphate of ATP as described in Damas *et al* [317].This conversion was made during 5 ps using a time step of 0.0005 ps. In the case of the subtraction technique, this conversion was made for 6 ps using a time step of 0.0006 ps. This change in settings was necessary to allow the direct comparison with ATP simulations in which trajectories were written to disk every 2 ps. The neighbour list was updated every five steps. After the slow-growth procedure, an extra positive charge is added to neutralize the system. This charge is added by simply replacing a water molecule far away from the protein. This allows to avoid inconsistencies due to the PME algorithm.

This procedure was done to generate initial conformations for the ADP state in order to study short and long-term effects of ATP hydrolysis. 50 replicates of the resulting ADP state were simulated during 150 ns, for both mutant and wild-type. ATP hydrolysis was simulated after 20 ns of the beginning of the simulation on the ATP state. These simulations were performed using the same conditions as described above for the ATP state.

The change of ATP to ADP was only performed on the catalytically competent site (ABP2).

3.3.4-Subtraction technique

In this work, the subtraction technique (figure 3.2) was used to study the short-term conformational changes triggered by ATP hydrolysis. In the subtraction technique, a mechanical perturbation is induced on the system using non-equilibrium techniques[323]. The simulations of both perturbed and unperturbed states are run during a given time simultaneously and the desired property is obtained by calculating the difference between the perturbed and unperturbed states. This allows the study of short-term changes, because the fluctuations between perturbed and unperturbed are highly correlated and can be cancelled, allowing to study the response to the perturbation. The perturbation has to be introduced in a high number of conformations of the unperturbed system in order to average the differences found in a statistically reliable manner. In this case the perturbation caused in the system was the conversion of ATP to ADP and IP using slow-growth, as previously described. In this work, frames

were extracted every 20 ns from the ATP 150 ns simulations. A total of 230 frames were selected to proceed for hydrolysis in each state (mutant and wild-type), and each frame was simulated during 300 ps after ATP hydrolysis. To study the effect of hydrolysis, the C-α RMSD per residue was calculated between each conformation at time *t* after hydrolysis (x_{ADP}) and the respective frame on the pre-hydrolysis on *t+6* ps time (x_{ATP}) (equation 3.1). In the end, the C-α RMSD is averaged by the number of frames (*N*) in which hydrolysis was performed (in this case 230). Fitting was done to the entire NBD dimer.

$$C_\alpha RMSD = \frac{\sum \sqrt{(x_{ADP} - x_{ATP})^2}}{N} \qquad 3.1$$

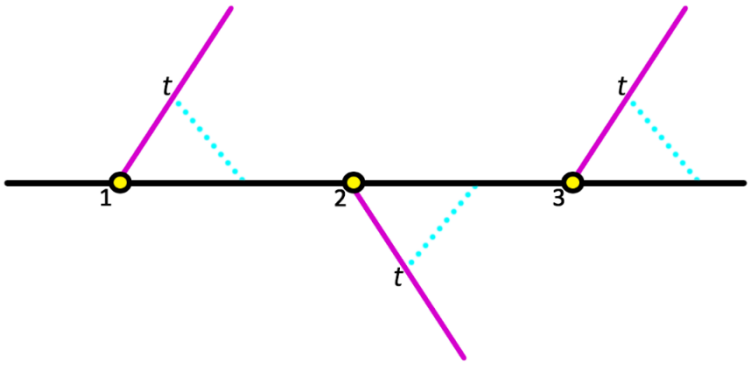


Figure 3.2- *Scheme explaining the concept of the subtraction technique (adapted from Ciccotti et al [323]). The black line represents an initial MD simulation (in this case the ATP state), in which at times 1, 2 and 3 a perturbation is applied (as represented by the yellow spheres). In this work, the perturbation was applied using slow-growth to create the phosphate group and delete the gamma phosphate from ATP. The purple lines represent the new simulations of the perturbed state (the ADP state). The broken lines represent comparable times (from instant t to instant t+n) in which the desired property is calculated as the difference of the values between the perturbed and unperturbed states.*

3.3.5-Data analysis

Analyses were made using the tools included in the GROMACS 5.0.7 package[318]. For the analysis of the long simulations, either in the ATP or ADP states, the first 20 ns were discarded.

The atomic deviations (C-α RMSD values) were calculated comparing the ADP and ATP states within each replicate and using average C-α positions only. The displayed values correspond to the average for all replicates. Fitting was done to residues of each β subdomain of each NBD, in order to eliminate rigid-body movements between NBDs. For NBD1 the residues considered were 452-457, 488-493, 568-572 (wild-type, 567-571 mutant) and 599-603 (wild-type, 598-602 mutant) and for NBD2 residues 1211-1213, 1231-1233,1239-1243, 1287-1290, 1367-1370, 1397-1399 and 1412-1415. In this analysis, only the last 130 ns of each replicate were used.

The secondary structure was determined using the DSSP program[324]. The elements of secondary structure considered were α-helix, β-sheet, β-bridge, turn and 3_{10} helix, according to the DSSP classification.

The C-α RMSD between the existing NBD1 structures and the NBD1 of our model was calculated using MDAnalysis[321,322].

The probability densities maps were obtained by estimating densities in a two-dimensional grid with a spacing of 0.01 nm^2 using a gaussian kernel estimator[325].

The error bars correspond to the 95% confidence interval obtained by bootstrapping[326], unless stated otherwise.

PyMOL 2.0 was used to visualize trajectories and produce the pictures presented[327].

3.4-Results and discussion

3.4.1-Molecular modelling of the closed NBD dimer of CFTR

A model of the human closed nucleotide binding domains of CFTR was built using homology modelling. The template used was the structure of zebrafish CFTR (PDB code 5W81) with the NBDs in a closed conformation. The human and zebrafish CFTR have 55% identity, and their unphosphorylated structures are very similar[311]. In order to create different replicates for MD simulation, different initial homology models were created, rather than attributing different velocities to several conformations prior to the production simulation. In figure 3.3 it is possible to see the average of the models of the ATP state, for both mutant and wild-type. The figure clearly shows that the main chain conformation of these two average conformations is superimposable, except for the zone of the F508del loop (inset).

The average C-α RMSD, of both wild-type and mutant, is stable during 150 ns of simulation and achieves a final value around 0.44 nm (figure 3.8). After ADP hydrolysis, the same behaviour is observed, with the average C-α RMSD achieving an average value around 0.35 nm (figure 3.9). The percentage of secondary structure retained in comparison with the initial structure was around 90% (figures 3.10 and 3.11). This means that there were no significant distortions of the protein structure in comparison with the initial conformation from which the simulations started. Hence, our model is stable and adequate for long simulations.

Tables 3.1 and 3.2 in supplementary information, show the C-α RMSD between the NBD1 of our model and the crystallographic structures available for this domain, for the wild-type and F508del forms, respectively. All values obtained are in the range of 0.1-0.18 nm, indicating a high structural similarity between our model and these structures. Moreover, these values of C-α RMSD on F508del NBD1 indicate that the mutant form was correctly modelled.

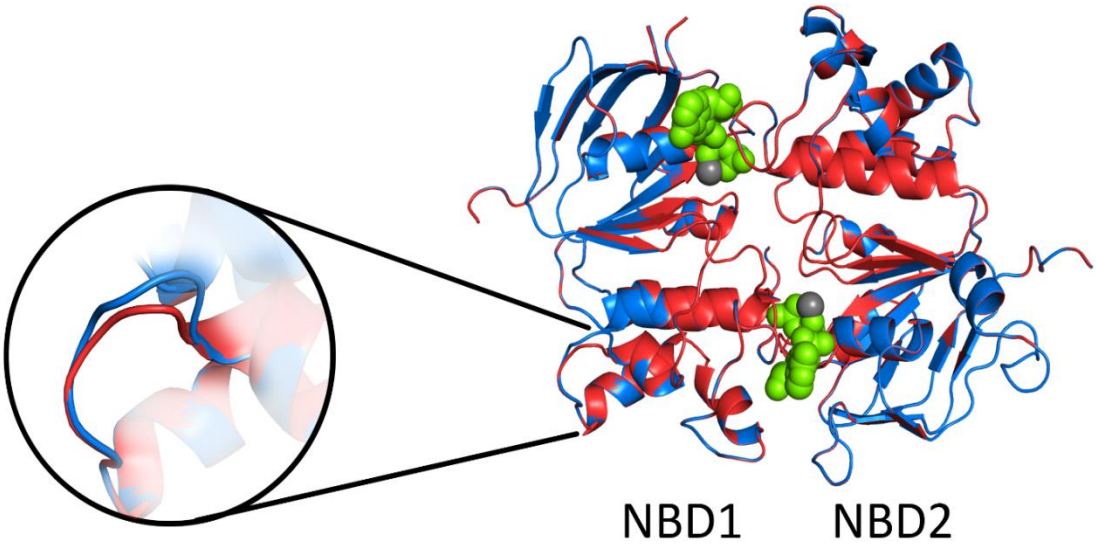


Figure 3.3- *Overlay of the average structures of the initial models generated for the wild-type (in blue) and mutant (in red), respectively. ATP and magnesium are represented in spheres (the two ATP conformations are superimposable). ATP and magnesium are colored green and grey, respectively. The inset focuses on the zone of the F508 loop.*

3.4.2-Effect of F508del in the behaviour of the dimer in the pre-hydrolysis state

The binding of ATP is essential for the function of ABC transporters. In CFTR, ATP becomes the "molecular glue" that stabilizes the dimeric form of the NBDs, leading to the opening of the pore[111–113,215].

The catalytic pocket (ABP2) possesses the canonical ABC and Walker A motifs, allowing ATP hydrolysis. The degenerate pocket (ABP1) possesses degenerate ABC and Walker A motifs leading to absent catalytic activity[213]. Recently, Sorum *et al* have found that the presence of ATP in ABP1 slows the closure of CFTR channel, while ATP in ABP2 speeds the opening of the CFTR pore[328]. Considering that NBD behaviour influences channel gating, we started by investigating the behaviour of wild-type and F508del NBDs with ATP in both pockets.

3.4.2.1-Does F508del influence the opening of ATP binding pockets?

As previously mentioned, when ATP is bound to both pockets, the NBDs are closed, and the opening of the pore is accelerated. Electrophysiological data of F508del CFTR show a gating pattern characterized by long periods of closure and infrequent opening, with rapid transitions from the closed to open state [306,307,329]. Jih *et al* found that F508del decreases the stability of the closed NBD dimer[307]. The distance between Walker A and ABC motifs indicates the degree of opening of each ABP, which can be ultimately understood as a measure of the stability of NBD closed state. A visual inspection of the trajectories revealed that in most conformations, both pockets show a concerted behaviour, when ABP2 is more likely to open, ABP1 is more likely to close and vice-versa.

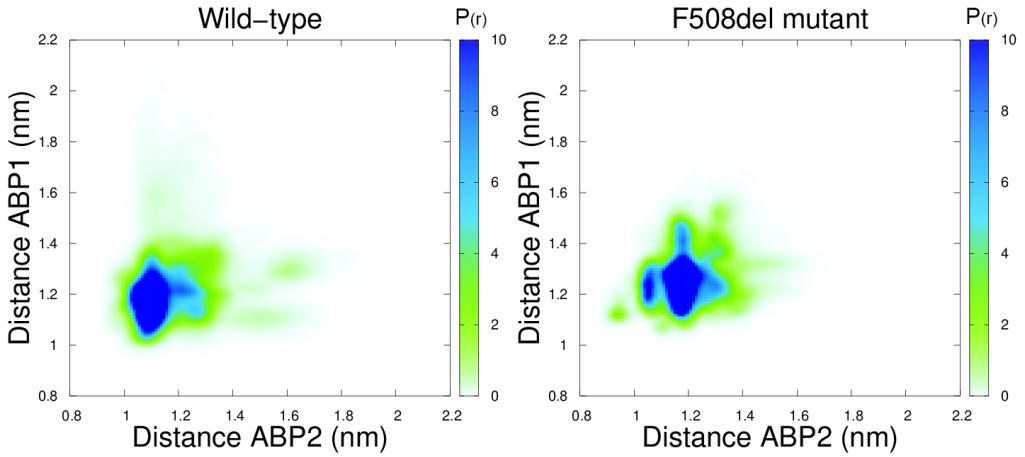


Figure 3.4- *Probability density ($P_{(r)}$) maps using distances between Walker A motif and the signature sequence on the ABP2 (x axis) and ABP1 (y axis). For ABP2, the distance was measured between the set of C-α atoms from residues K1250, S1251 and T1252 of Walker A and other set of C-α atoms comprising residues S548, G549, G550 and Q551 from the signature motif. In ABP1, the distance was measured using C-α atoms from residues L1346, S1347, H1348 and G1349 (ABC motif) and K464, T465, S466 from Walker A. The last 130 ns of the trajectories were used in this analysis. The blue colors mean higher probability density values, while the green ones (up to white) mean lower probability density values.*

Figure 3.4 shows the probability density profiles made using the opening distances of both ABP2 and ABP1 in mutant (right) and wild-type (left). These profiles show that both ABP1 and ABP2 have distinct opening behaviour, confirming our visual observations of the trajectories. Moreover, F508del and the wild-type sample different conformational spaces. As a consequence, it is possible to observe that F508del possesses higher probability density in the region from 1.2 nm to 1.4 nm of ABP2 distance. Most conformations with ABP2 distance larger or equal to 1.2 nm show a clear separation between the two NBDs in this pocket, and ATP was

mainly bound to the Walker A motif on NBD2, making very few contacts with the ABC motif on NBD1. Therefore, our simulations show an increased active site opening on F508del mutant upon ATP hydrolysis, in comparison to wild-type. Nonetheless, the wild-type displays small populations with distances of ABP2 around 1.6 nm.

Regarding ABP1, a visual inspection of trajectories shows that from 1.3 nm onwards this pocket is visibly open, with a clear separation between the two nucleotide binding domains with ATP remaining only bound to Walker A. F508del shows higher probability density at this distance and until 1.5 nm, while the wild-type shows higher density at values between 1 and 1.2 nm.

These results indicate that F508del affects the opening degree of ABP1 and ABP2 by increasing the distance between the ABC and Walker A motifs. Therefore, it can potentially explain the faster ligand exchange times in F508del CFTR observed by Jih *et al*[307], assuming that a more open probability of this pocket can lead to a shorter residence time of the nucleotide. Electrophysiological and biochemical assays have concluded that a dissociated NBD dimer implies a closed channel gate. On the other hand, when the channel gate is open, the NBD dimer is associated, with ATP bound in both pockets. ATP hydrolysis leads to a second open channel state and increases the closing probability[112,113,215,328]. An increased pocket opening in F508del, as seen in our results, destabilizes the associated dimer, and consequently, the open channel state, leading to an increased closed time and to the flickering opening characteristic of F508del CFTR[226,307,329].

Despite the different profiles of pocket opening, the ATP binding mode in both pockets remains almost unaltered in the F508del mutant as seen in the figures 3.12 and 3.13 in the supplementary information.

ABP1 and ABP2 display different distance profiles. However, the recent cryo-EM structure of human CFTR in the inward-facing conformation shows that the inter-motif distance of both ATP binding pockets is similar. Our simulations sample a different state of the ensemble in which the inter-motif distance in the two pockets is different. Although we may be biased by our starting structure, it is quite likely this state exists, because in other transporters with degenerated pockets, such as TM287/288, in which EPR studies found that in the presence of AMP-PNP-Mg, ABP2 is closed, while ABP1 is open[330]. On the other hand, the density maps (figure 3.4) show that states with similar ABP1 and ABP2 distances are also sampled. Therefore, our ensembles include conformations in which the NBD dimer is similar to recent human phosphorylated CFTR structure[107].

3.4.2.2-Does F508del influence the catalytic competence of the NBD conformations?

In order for catalysis to occur, the active site needs to be in a catalytically competent conformation. One of our questions concerns the effect of F508del in the observed percentage of catalytically competent conformations. But first, we must define which of these conformations are. The mechanism of ATP hydrolysis in CFTR is not precisely known, but studies of ATP hydrolysis on the MalK dimer and on the B_{12} vitamin importer BtuCD conclude that the presence of a water molecule between ATP and the catalytic glutamate is essential for ATP hydrolysis, regardless of the proposed mechanism[41,46,47,49]. Jones *et al* also defined geometric criteria for assessing hydrolysis-competent conformations, based on Sav1866 MD simulations, such as distances and angles between the catalytic residues and the water molecule[33]. Here, we define the following criteria for assessing catalytically competent conformations: a water molecule must be coordinated both by E1371 and ATP, the catalytic residues E1371 and H1402 must also be oriented towards ATP, and there must be contact between ATP and the active site motifs. To ensure contact between ATP and the active site motifs, only conformations with the ABP2 closed (with the distance between the Walker A and ABC motifs lower than 1.2 nm) were selected.

As explained in the methods section, we started by simulating 50 replicates of 150 ns for both F508del and wild-type in the pre-hydrolysis state. In order to generate the post-hydrolysis state, we extracted conformations at 20 ns, from each replicate and performed ATP hydrolysis as previously described in the methods section. Hydrolysis was done at 20 ns, the end of the equilibration phase (see figure 3.9). Nonetheless, an inspection of the conformations prior to hydrolysis, revealed that only 48% of the wild-type structures and 20% of the mutant structures fulfilled the requirements for ATP hydrolysis. These results indicate that F508del affects the existence of catalytically competent conformations in CFTR. In order to increase sampling for the ADP state, more ATP replicates were done, to obtain reliable statistics, as explained in the methods section.

3.4.3-Post-hydrolysis behaviour of mutant and wild-type CFTR NBDs

One of the direct consequences of ATP hydrolysis in ABC transporters is the opening of the active site in which hydrolysis happens, leading to a partial dimer state[161,163,171,216,317]. In CFTR, mutations that abolish ATP hydrolysis lead to an increase of the duration of the open channel state[109,330]. Additionally, the presence of orthovanadate and ADP, as well as nonhydrolyzable ATP analogues delay channel closure[331,332]. These experimental evidences suggest that ATP hydrolysis leads to channel closure and confirm that hydrolysis has functional consequences on

channel gating. Hence, it becomes relevant to study the possible changes triggered by ATP hydrolysis as well as the influence of F508del.

3.4.3.1-Does F508del affect short-term conformational changes after ATP hydrolysis?

Using the subtraction technique, it is possible to study the short-term conformational changes triggered by ATP hydrolysis. This will allow us to characterize the dimer response to ATP hydrolysis in a time-dependent manner and establish a possible route of signal transmission.

A total of 230 conformations (extracted from 150 ns ATP simulations) were subjected to ATP hydrolysis. Figure 3.5 shows the average C-α RMSD deviation along time, between the ATP and ADP states mapped on the average post-hydrolysis protein structure. 2 ps after ATP hydrolysis small structural changes can be seen on the wild-type and F508del dimers, namely on Walker A and ABC motifs in ABP2 as well as on the NBD2 region near ABP1 that comprises the X-loop and Walker A. However, the magnitude of these changes is lower in the mutant. Remarkably, the F508/F508del loop displays similar deviations in either form, indicating that this segment is particularly affected by ATP hydrolysis. As time progresses, there is an overall increase of the deviations across the dimer, even on the rigid regions, such as the β-sheet domains. The regions near the ABP2 (including the active site motifs) and ABP1 register the largest increase in C-α RMSD.

It is possible to establish a possible path of signal transmission, from the active site to the D-loop, whose C-α RMSD also increases along time, and the NBD2 region near ABP1. This NBD2 region interfaces with the intracellular loop 2 (ICL2) and it potentially affects signal transmission to the transmembrane domains[187]. The D-loop is thought to be involved in interdomain communication in other ABC transporters such as TM287/288[29]. This is in agreement with other computational studies that concluded that ATP hydrolysis starts by causing rearrangements on the active site motifs that are transmitted to the helical domains and other regions[317].

The F508del dimer also exhibits an increase of the C-α RMSD, but never reaches the same values as its wild-type counterpart. Therefore, in the F508del dimer, conformational changes happen in a delayed timing, even though the same regions are affected.

Figures 3.14 and 3.15 show the ABP2 distance distributions before and after 2, 20, 50, 100 and 300 ps of ATP hydrolysis, in wild-type and F508del, respectively. It is possible to observe that the pre and post-hydrolysis distance distributions are very similar during time. Nonetheless, in the wild-type, there is a decrease of the main peak from 1.06 to 1.11 nm in the pre-hydrolysis state along with an increase of the populations of the bins from 1.15 to 1.25 nm. However, the occupation of these bins varies with time until 300 ps. F508del displays a decrease of the

populations in the bins that correspond to distances from 1.15 to 1.21 nm, with a concerted increase of the populations in bins 1.21 to 1.33 nm. Interestingly, these last populations were persistent throughout time, in contrast with the wild-type, where the occupation of the bins with larger distance values oscillates through time (figure 3.16).

The subtraction technique data reveals that although ATP hydrolysis leads to higher C-α RMSD in the active site region of the wild-type, the ABP2 inter-motif distance suffers larger fluctuations. On the other hand, the F508del dimer displays lower C-α RMSD on the active site regions, but it shows a stable population of conformations with larger ABP2 distances, evidencing a higher rigidity of the dimer.

In this way, despite the short timescale, it can be speculated that F508del delays the transmission of energy after hydrolysis, which could account for the increased closed times observed in electrophysiology assays. Additionally, it is plausible that F508del affects the propagation of the conformational changes after hydrolysis, via allosteric mechanisms, in a similar way as photoswitchable systems, whose simulations, revealed a similar behaviour, in which, after an induced perturbation, there is an elastic response in the timespan of 0.1 ns [333].

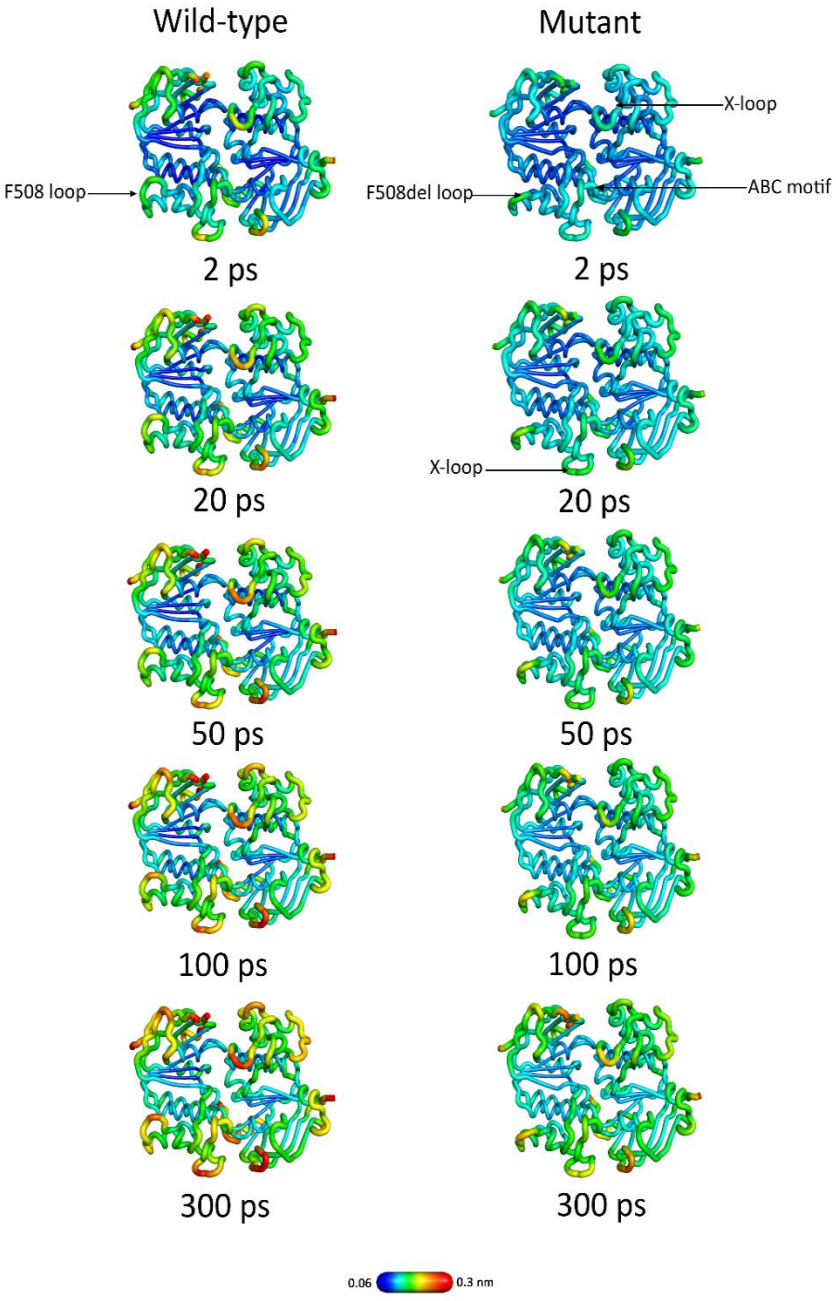


Figure 3.5- *Conformational changes at 2 ps, 20 ps, 50 ps, 100 ps and 300 ps after ATP hydrolysis. The average C-α RMSD deviation between the ATP and ADP state mapped on the average structure of the ADP state at the corresponding time. The protein is represented as a cartoon. The nucleotides were omitted for clarity. The scale below indicates the magnitude of the C-α RMSD changes observed, in which the dark blue corresponds to the minimum values observed, while the red corresponds to the maximum values observed.*

3.4.3.2-What is the effect of F508del on long term conformational changes after ATP hydrolysis?

Although the subtraction technique allowed a glimpse of the temporal evolution of the conformational changes after hydrolysis, its usage is restricted to smaller timescales. This happens because the random fluctuations in both perturbed and unperturbed states are highly correlated, and are cancelled, allowing an accurate study of the response of the system to the perturbation[323]. Nevertheless, most significant events for the function of proteins occur in a nanosecond to microsecond timescale. For this reason, we have also investigated the effect of F508del in the conformational changes on the 150 ns-long simulations. To do so, the C-α RMSD deviation per each individual residue was calculated between the average structure of each ADP replicate (using the last 130 ns) and the respective ATP simulation, the replicate from which the ADP state was generated (also using the last 130 ns). In the end, the C-α RMSD was averaged throughout all replicates and mapped on the average structures of the wild type and mutant (figure 3.6).

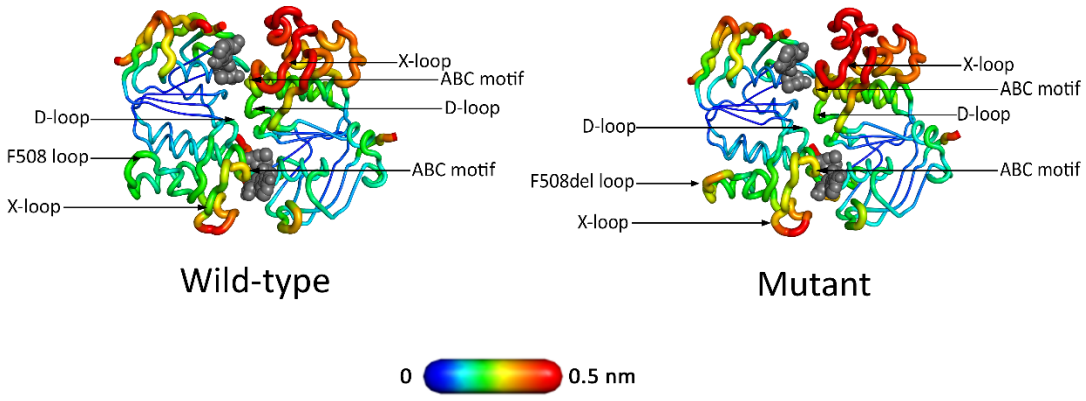


Figure 3.6- *Mapping of the C-α RMSD deviation between the ATP and ADP states on the average ADP structures for both mutant and wild-type. The C-α RMS deviation was calculated using the equilibrated part of the trajectory (the last 130 ns). The scale below indicates the magnitude of the C-α RMSD changes observed, in which the dark blue corresponds to the minimum values observed, while the red corresponds to the maximum values observed. ATP and Mg^{2+} are represented in grey spheres in the respective binding pockets.*

In both F508del and wild-type, significant conformational changes were observed in the active site, the NBD1 region near the active site, the D-loop and the region that comprises residues 1293 to 1350 that include the NBD2 X-loop and the signature sequence. The energy generated from ATP hydrolysis in the active site is spread out across the dimer reaching ABP1 and other structurally important regions, such as the X-loop and the helical domains. The X-loop is known for interacting with the coupling helices of the transmembrane domains [18] in ABC transporters, and the interfaces involving the X-loop and coupling loops 3 and 4 were found to

be important for CFTR activity [334].On the other hand, the helical domains are in contact with the intracellular loops that are associated with the transmembrane helices that control the ion flux. Additionally, the NBD2 region that comprises residues 1293 to 1350, interacts with the intracellular helix 2 and the transmembrane helices 5 and 4, as seen by the full-channel structure obtained by cryo-EM. Therefore, the transmission of the hydrolysis signal from the nucleotide binding domains to the transmembrane domains is also likely to be affected in F508del CFTR by this route.

In comparison with the time-scale of 300 ps after ATP hydrolysis, it is evident that the conformational changes in both mutant and wild-type have become more similar as the simulations progressed. Nonetheless, it is possible to observe that the conformational changes on the F508 loop and the X-loop and surrounding residues on ABP2 are particularly affected by F508del. The F508del loop experiences significant changes on the mutant, while on the wild-type there is lower difference between both states. These results confirm that this segment is particularly affected by hydrolysis.

Early NMR and mass spectrometry studies concluded that the absence of F508 does not cause major perturbations on the NBD1 structure, but only an increase of the fluctuations of F508 loop[209,210,335]. Nevertheless, our results show that dimer dynamics before and after ATP hydrolysis are different in the mutant, affecting the ABP inter-motif distance and conformational changes after ATP hydrolysis. In fact, Roxo-Rosa *et al* had already suggested that F508del might cause instability of the CFTR steady state, or disturb the transitions of the conformational cycle, due to the reduction of the open bursts and the increased duration of the closed state [184].

Other studies using cross-linking experiments and simulation, show that F508 interacts with ICL4 and the equivalent residues in NBD2 are important to interact with ICL2[187,336,337,338]. Therefore, other consequence of F508del is the interference with the interactions between the NBDs and the transmembrane domains. Still, our simulations show that F508del affects the pre and post hydrolysis behaviour of the NBD dimer, regardless of the presence of the transmembrane domains. Despite its impaired function, F508del CFTR is still able to hydrolyse ATP[308,309]. Our findings allow us to hypothesize that F508del can interfere with dimer function, and it can possibly affect communication with the transmembrane domains by mechanisms other than the interaction of F508 with ICL4.

3.4.3.3-Does F508del influence the opening of ABP2 and ABP1?

As previously mentioned, ATP hydrolysis leads to the closure of the CFTR channel. In many previous simulation studies on the NBDs of ABC transporters, ATP hydrolysis has led to the opening of the active site[161,163,171,316]. In the case of CFTR, ATP remains bound to ABP1 throughout the catalytic cycle[213]. This will result in a partial NBD dimer, in which ABP2 is open, while ABP1 remains closed. The opening of the active site also allows the exit of the reaction products and the entry of another ATP molecule to ABP2 and proceed to the next stage of the conformational cycle.

Figure 3.7 illustrates the distribution of distances between the signature motif and the Walker A in ABP2. ATP hydrolysis causes the opening of the active site in both wild-type (figure 3.7A) and F508del (figure 3.7B), as seen by the higher number of conformations that display a distance larger than 1.2 nm between the motifs and the decrease of the peak around 1.1 nm.

Comparing F508del with the wild-type dimer before and after hydrolysis, it is clear that the main differences can be seen in the ATP state (figure 3.7C), while the ADP state is similar in both forms (figure 3.7D). Although F508del and wild-type post-hydrolysis distributions are very similar, it is possible that larger differences appear on longer timescales.

When the NBDs are in an associated state, such as in the pre-hydrolysis state, the CFTR pore is open. On the other hand, ATP hydrolysis triggers the opening of ABP2, leading to closure of the transmembrane pore. Nonetheless, our results show that F508del has a reduced influence on the active site opening after hydrolysis. Yet, we can say that our results agree with experimental observations that report that F508del decreases the opening probability of the CFTR channel, increasing the closed time[306,329,307] concomitantly decreasing the opening probability[339], in fact we observe an increase of ABP2 opening in F508del, before hydrolysis, potentially leading to an increased probability of the pore being closed.

Despite the observation of opening of ABP2, no phosphate or ADP exit happened in the simulated timescale.

Concerning ABP1, it is known experimentally that ATP is bound to it for tens of seconds, but it is not hydrolysed [212,213]. Despite not being catalytically competent, it is believed that this site plays a significant part in gating energetics and commanding channel kinetics[214–216]. Tsai *et al* found that channel closure could be delayed when high-affinity analogues of ATP are bound to ABP1[216]. In addition, Csanády et al found that ABP1 has an allosteric role, because mutations and high-affinity analogues cause a decrease in the transition rate from a pre-hydrolytic state to the post-hydrolytic one[30].

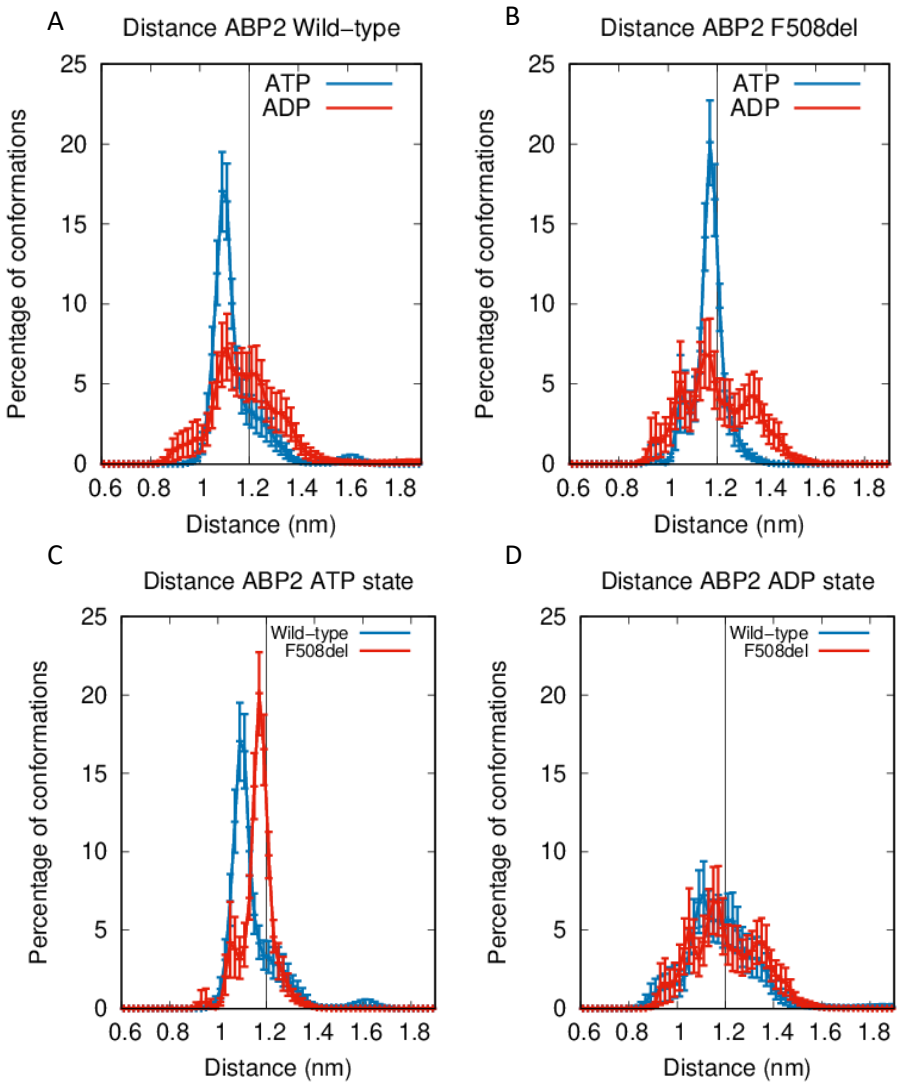


Figure 3.7- *Histogram of distances between Walker A motif and the signature sequence on the ABP2. A- Inter-motif distance of ABP2 in wild-type ATP and ADP. B-Inter-motif distance of ABP2 in mutant ATP and ADP. C-Inter-motif distance of ABP2 in the ATP state in mutant and wild-type. D-Inter-motif distance of ABP2 in the ADP state in F508del and wild-type. The distance was measured between the set of C-α atoms from residues K1250, S1251 and T1252 of Walker A and other set of C-α atoms comprising residues S548, G549, G550 and Q551 from the signature motif. The error bars correspond to the 95% confidence interval using bootstrapping. The replicates represented in the pre-hydrolysis state are only the ones suited for hydrolysis. These plots comprise the last 130 ns of the simulations. The ATP state only comprises the ATP simulations suited for hydrolysis at 20 ns (50 replicates). The upper panel compares the pre and post hydrolysis states for both F508del and wild-type, while the lower panel compares the F508del and the wild-type before and after ATP hydrolysis. The black line at 1.2 nm indicates the threshold at which the pocket is considered to be open, according to a visual inspection of the trajectories.*

The distribution of distances of ABP1 is represented in figure 3.17. The pre and post-hydrolysis distributions are very similar, indicating that ATP hydrolysis has little effect on the opening of this pocket and that the main differences are due to F508del.

Due to the differences observed in the inter-motif distance distributions of each ATP binding pocket, it becomes pertinent to question whether the nucleotide binding is also affected in each ABP. The data plotted on figures 3.18 to 3.20 show that the number of hydrogen bonds between the ligands (ATP, phosphate and ADP) with either Walker A or the ABC motif, in each ATP binding pocket is very similar in both mutant and wild-type. Experimental data using ATP analogues and pyrophosphate shows that F508del destabilize both full associated dimer and the partial dimer form (the dimer after hydrolysis with ABP2 open)[307]. This indicates that F508del also interferes with nucleotide interactions. Nonetheless, the experimental data was measured in the second time-scale and the timespan of our simulations is too short to see ligand dissociation.

3.5-Conclusions

In this work, molecular dynamics simulations of the wild-type and F508del CFTR isolated NBD dimer were performed to investigate the effect of F508del before and after ATP hydrolysis. We concluded that F508del interferes in the pre-hydrolysis dimer by increasing the opening distance of both ATP binding pockets. Additionally, a decrease in the number of catalytically competent conformations was observed in the F508del mutant. ATP hydrolysis leads to conformational changes on the entire NBD dimer, especially on the F508/F508del loop, the Walker A and ABC motifs in ABP2 as well as on the NBD2 region near ABP1 that comprises the X-loop and Walker A. On the first 300 ps after hydrolysis, the F508del dimer displays lower C-α RMSD. 150 ns after hydrolysis the conformational changes become similar between both forms, with marked differences in the above mentioned most affected regions.

These findings show that F508del alters the dynamics of the NBD dimer, shedding a new light on the mechanisms by which F508del impairs CFTR function, that are important for a further understanding about cystic fibrosis. Further studies are needed to understand the post-hydrolysis effect of F508del on the full-length CFTR.

3.6-Acknowledgments

This work was funded by Project LISBOA-01-0145-FEDER-007660 (Microbiologia Molecular, Estrutural e Celular) funded by FEDER funds through COMPETE2020 - Programa Operacional Competitividade e Internacionalização (POCI) and by national funds through FCT - Fundação para a Ciência e a Tecnologia. Bárbara Abreu acknowledges funding via a PhD scholarship (SFRH/BD/108002/2015) from FCT and funding from ITQB-NOVA via the grant INCENTIVO/EQB/LA0004/2014 ref. 010/BIS/2015.

3.7-Supplementary Information

Conformational drift for CFTR simulations

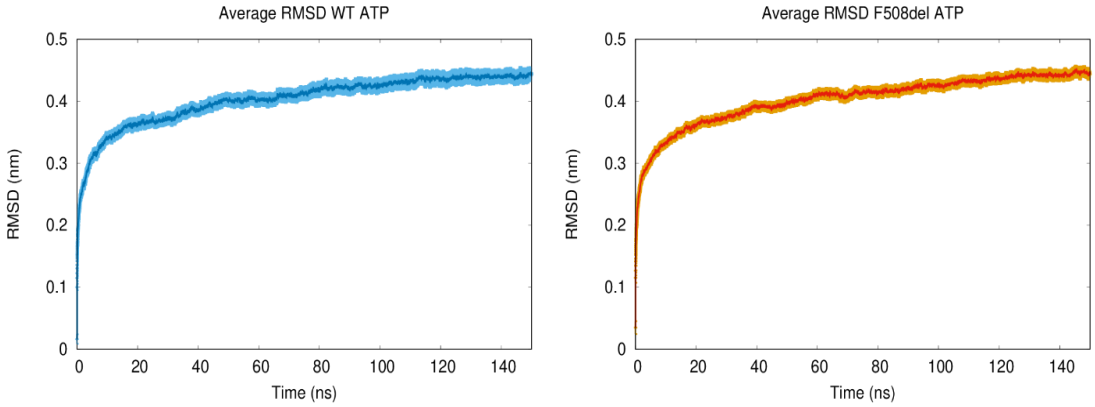


Figure 3.8- *Temporal evolution of average C-α RMSD for all ATP replicates (76 for wild-type, 88 for F508del). RMSD was calculated against the initial structure of each simulation. The error bars correspond to the standard deviation obtained by bootstrapping. The error bars are represented in light blue in wild-type ATP state, orange in the F508del ATP state. WT-Wild-type.*

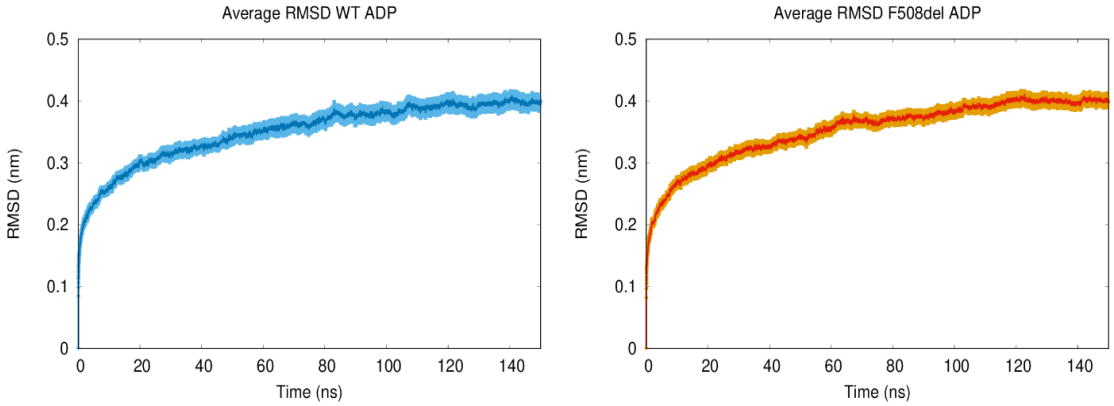


Figure 3.9- *Temporal evolution of average C-α RMSD for all ADP replicates of each state (50 replicates for each). RMSD was calculated against the initial structure of each simulation. The error bars correspond to the standard deviation obtained by bootstrapping. The error bars are represented in light blue in wild-type ADP state, orange in the F508del ADP state. WT-Wild-type.*

Structural stability of CFTR simulations

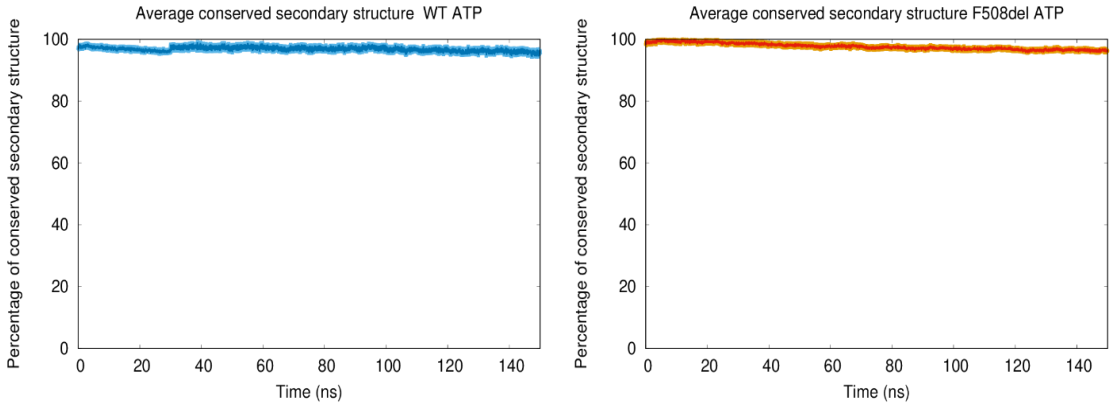


***Figure 3.10**- Percentage of retained secondary structure in all ATP replicates, comparing to the initial ATP structure. The values represented correspond to the average of all replicates for each state (76 for wild-type, 88 for F508del). The error bars correspond to the standard deviation obtained by bootstrapping. The error bars are represented in light blue in wild-type ATP state, orange in the F508del ATP state. WT-Wild-type.*

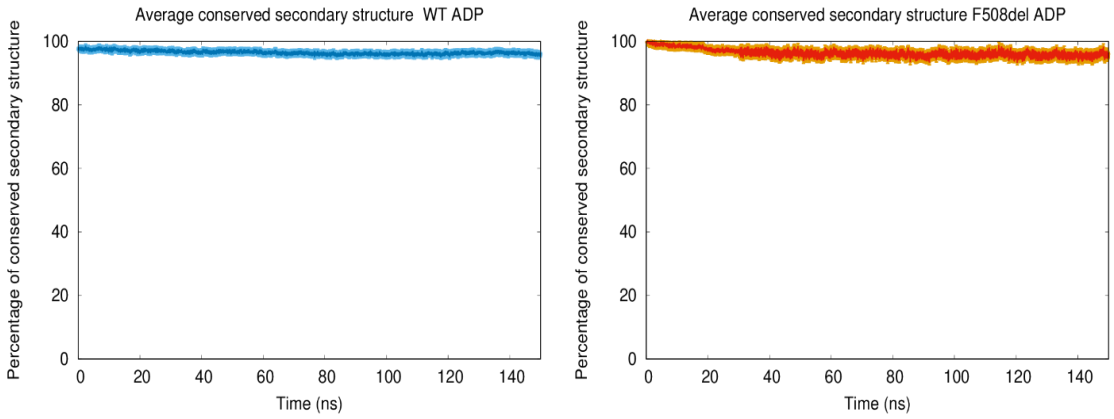


***Figure 3.11**- Percentage of retained secondary structure in all ADP replicates, comparing to the initial ATP structure. The values represented correspond to the average of all replicates for each state (50 replicates for each). The error bars correspond to the standard deviation obtained by bootstrapping. The error bars are represented in light blue in wild-type ADP state, orange in the F508del ADP state WT-Wild-type.*

Comparison of NBD1 x-ray structures with the NBD1 present in the used model

Table 3.1- *C-α RMSD of wild-type NBD1 x-ray structures against the average structure of the initial wild-type model. Residues 438 to 636 were used for C-α RMSD calculation.*

Structure (PDB code)	C-α RMSD (nm)
1Q3H	0.1195
2BBO	0.1857
1R0X	0.1204
2PZE	0.1093
1R0W	0.1210
1R10	0.1165

Table 3.2- *C-α RMSD of the NBD1 x-ray structures against the average structure of the initial F508del model. Residues 438 to 636 were used for C-α RMSD calculation in structures 3SI7,1XMJ,4WZ6 and 2PZF. Residues 438 to 541 and 546 to 636 were used for C-α RMSD calculation in structures 2BBT and 2BBS, due to missing residues in these structures. Residues 438 to 540 and 544 to 636 were for C-α RMS calculation in the structure 4WZ6 due to missing residues in this structure.*

Structure (PDB code)	C-α RMSD (nm)
3SI7	0.1302
1XMJ	0.1580
4WZ6	0.1341
2PZF	0.1001
2BBS	0.1377
2BBT	0.1429

Number of hydrogen bonds between ATP binding motifs and ATP on pre-hydrolysis CFTR simulations

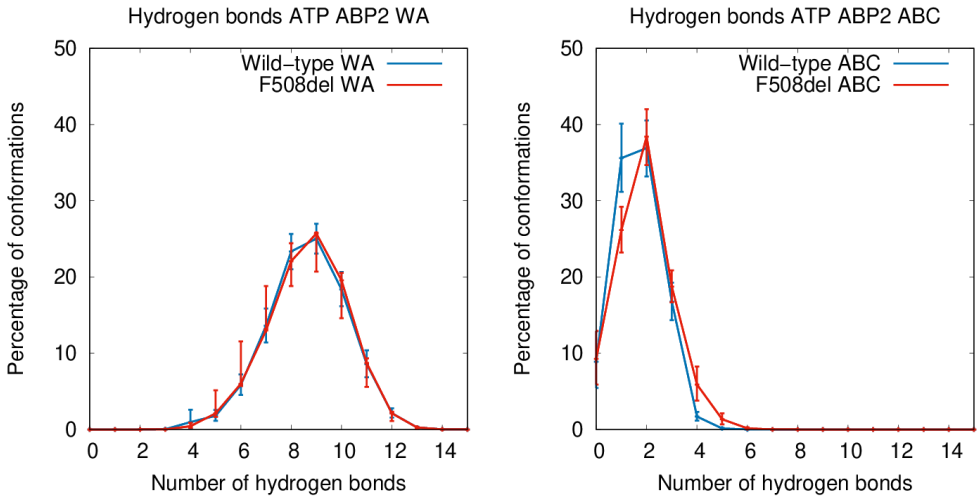


Figure 3.12- *Histograms depicting the number of hydrogen bonds between ATP with the Walker A motif (residues 1244 to 1252 on wild-type) and the ABC signature motif (residues 548 to 552 on wild-type) on ABP2. The values represented correspond to the average of all replicates for each state (76 for wild-type, 88 for F508del), using only the last 130 ns of each trajectory. The error bars correspond to the 95% confidence interval calculated by bootstrapping. WA- Walker A, ABC-ATP binding cassette.*

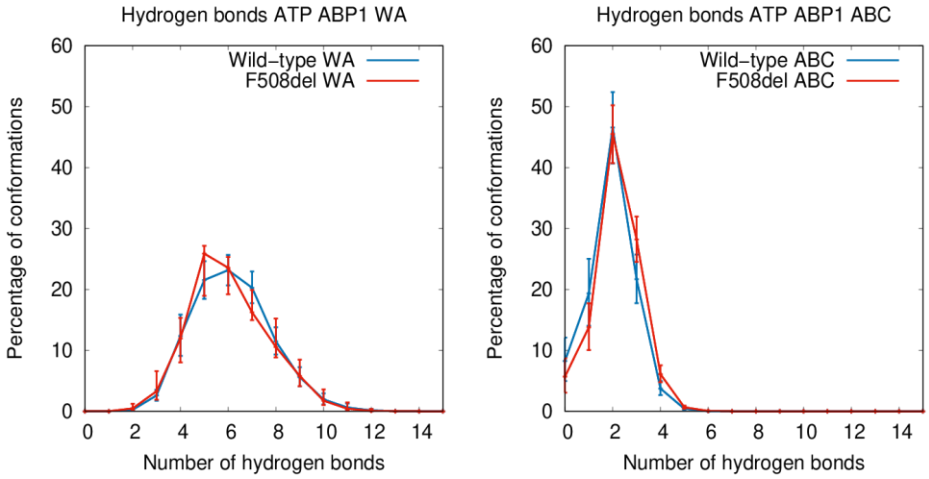


Figure 3.13- *Histograms depicting the number of hydrogen bonds between ATP with the Walker A motif (residues 458 to 466 on wild-type) and the ABC signature motif (residues 1346 to 1350 on wild-type) on ABP1. The values represented correspond to the average of all replicates for each state (76 for wild-type, 88 for F508del), using only the last 130 ns of each trajectory. The error bars correspond to the 95% confidence interval calculated by bootstrapping. WA- Walker A, ABC-ATP binding cassette.*

Intermotif distance of ABP2 on the post-hydrolysis CFTR simulations until 300 ps

Wild-type

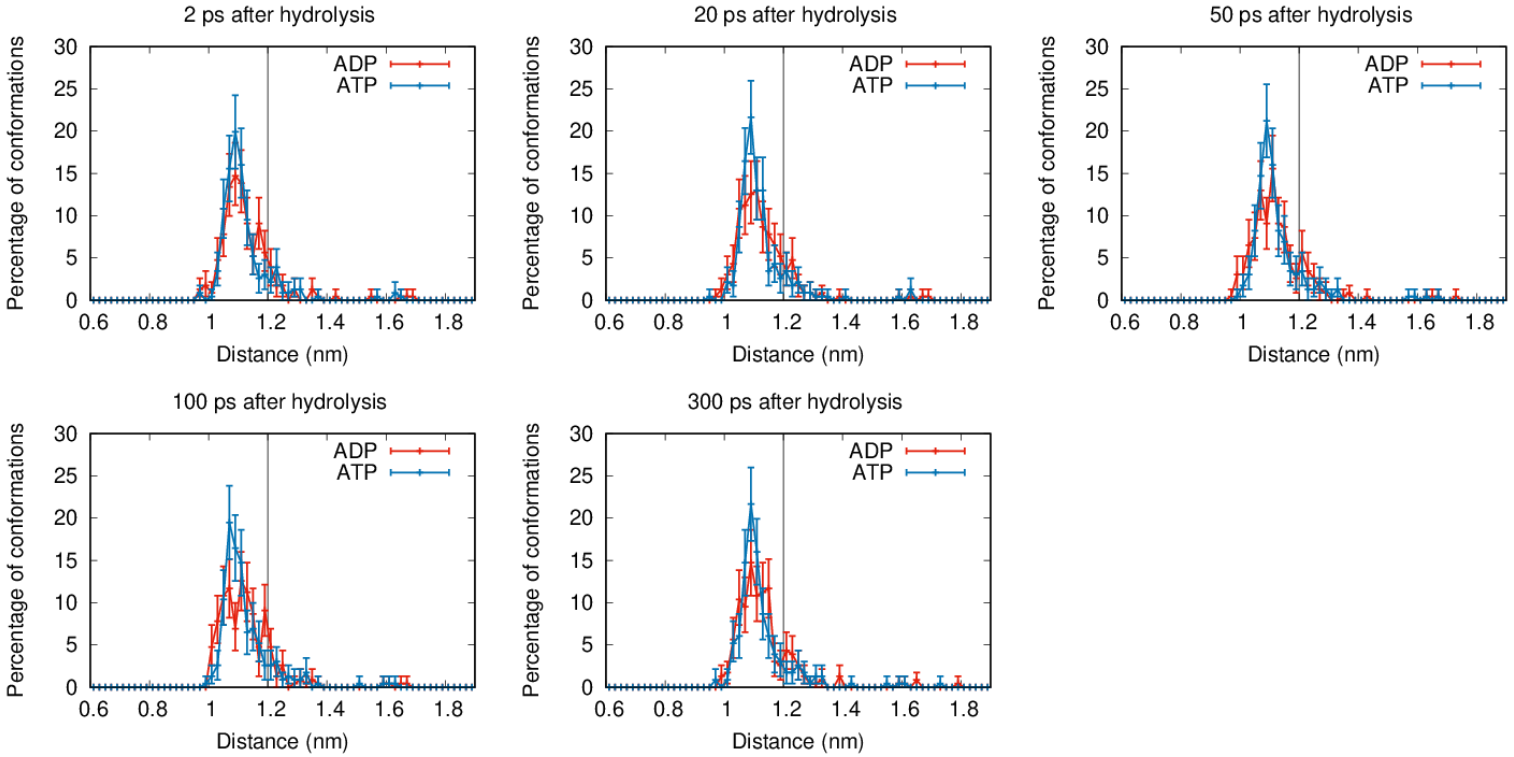


***Figure 3.14**- Histograms depicting the distance between Walker A motif and the signature sequence on the ABP2 after 2, 20, 50, 100 and 300 ps after hydrolysis on the wild-type. The ATP structures corresponds to the corresponding timeframes that can be compared using the subtraction technique. The distance was measured between the set of C-α atoms from residues K1250, S1251 and T1252 of Walker A and other set of C-α atoms comprising residues S548, G549, G550 and Q551 from the signature motif. The error bars correspond to the 95% confidence interval using bootstrapping. The black line at 1.2 nm indicates the threshold at which the pocket is considered to be open, according to a visual inspection of the trajectories, as described in the results section.*

F508del

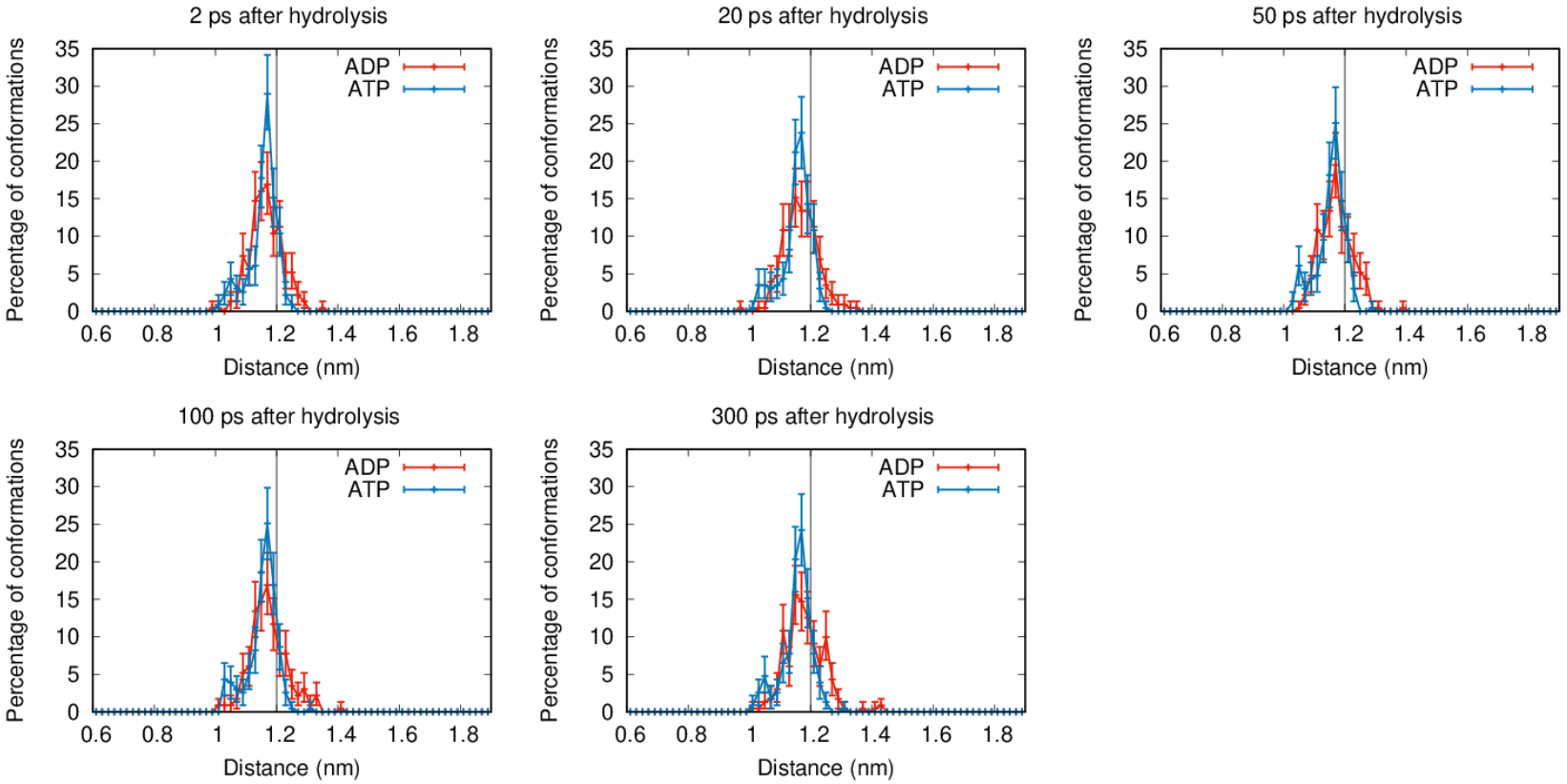


Figure 3.15- *Histograms depicting the distance between Walker A motif and the signature sequence on the ABP2 after 2, 10, 20, 50, 100 and 300 ps after hydrolysis on F508del. The ATP structures corresponds to the corresponding timeframes that can be compared using the subtraction technique. The distance was measured between the set of C-α atoms from residues K1250, S1251 and T1252 of Walker A and other set of C-α atoms comprising residues S548, G549, G550 and Q551 from the signature motif. The error bars correspond to the 95% confidence interval using bootstrapping. The black line at 1.2 nm indicates the threshold at which the pocket is considered to be open, according to a visual inspection of the trajectories, as described in the results section.*

Wild-type vs F508del ADP state

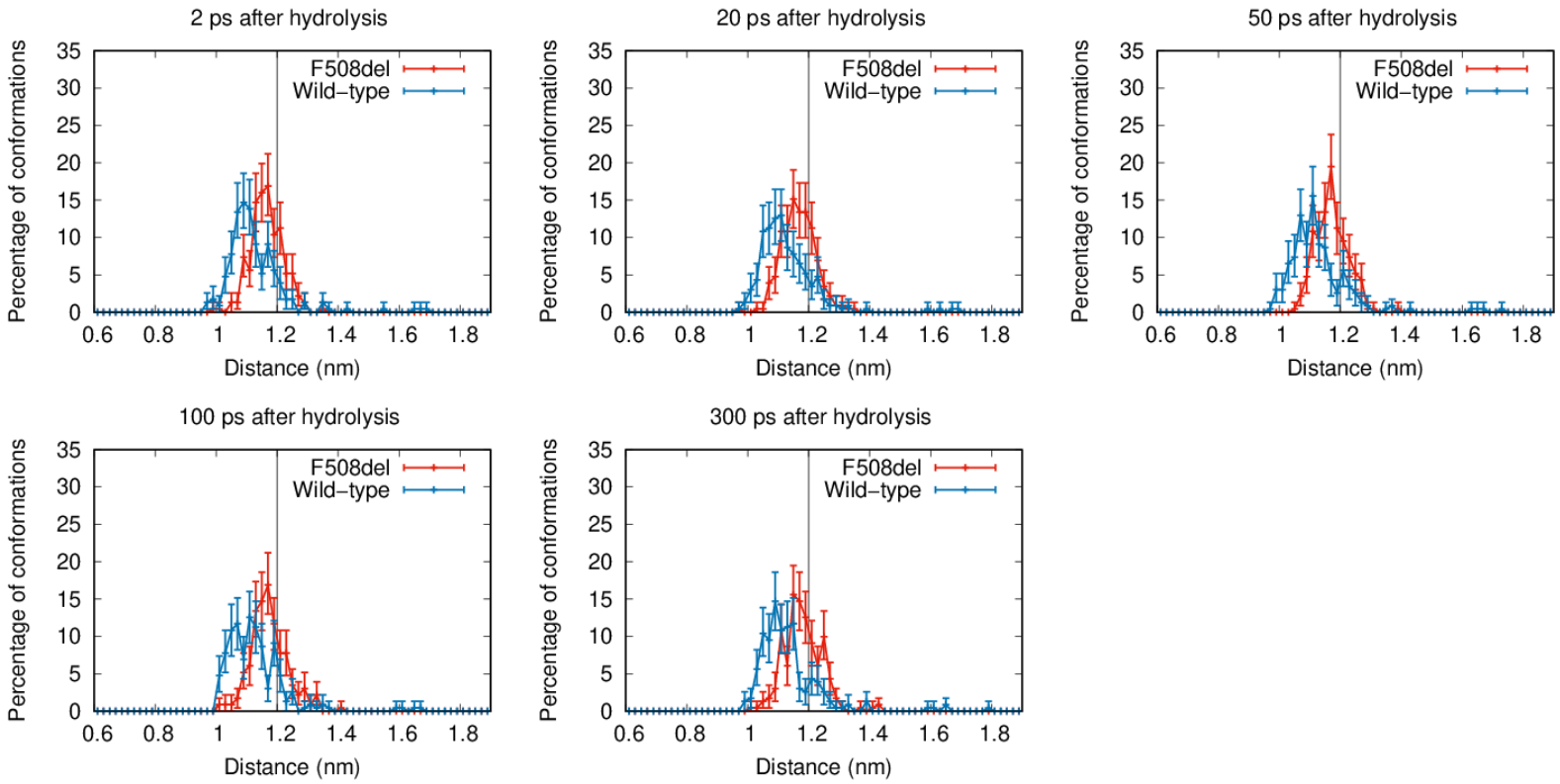


Figure 3.16- *Histograms depicting the distance between Walker A motif and the signature sequence on the ABP2 after 2, 20, 50, 100 and 300 ps after hydrolysis on the F508del mutant and wild-type. The distance was measured between the set of C-α atoms from residues K1250, S1251 and T1252 of Walker A and other set of C-α atoms comprising residues S548, G549, G550 and Q551 from the signature motif. The error bars correspond to the 95% confidence interval using bootstrapping. The black line at 1.2 nm indicates the threshold at which the pocket is considered to be open, according to a visual inspection of the trajectories, as described in the results section.*

Intermotif distance of ABP1 on the post-hydrolysis CFTR simulations

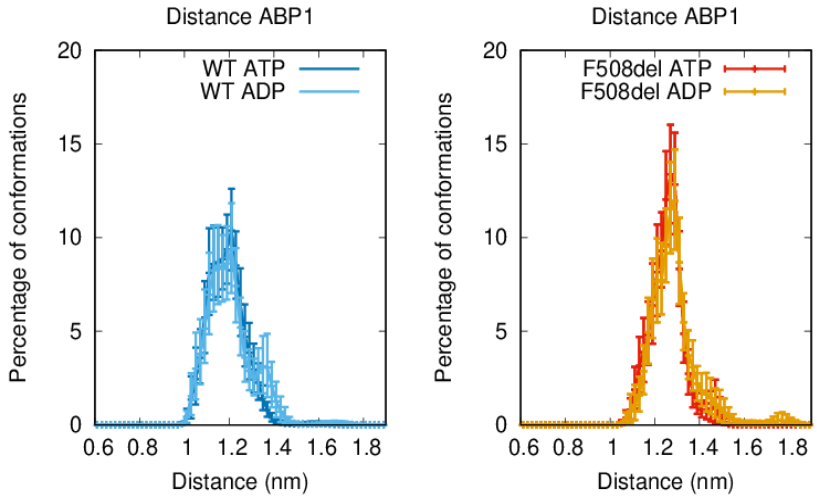


Figure 3.17- *Histogram of distances between Walker A motif and the signature sequence on the ABP1. The distance was measured between the set of C-α atoms from residues K464, T465 and S466 of Walker A and other set of C-α atoms comprising residues S1347, H1348, G1349 and H1350 from the signature motif. The error bars correspond to the 95% confidence interval using bootstrapping. The replicates represented in the pre-hydrolysis state are only the ones suited for hydrolysis. These plots were made from data from the last 130 ns of the simulations. The ATP state only comprises the ATP simulations suited for hydrolysis at 20 ns.*

Number of hydrogen bonds between ATP binding motifs and the nucleotides on post-hydrolysis CFTR simulations

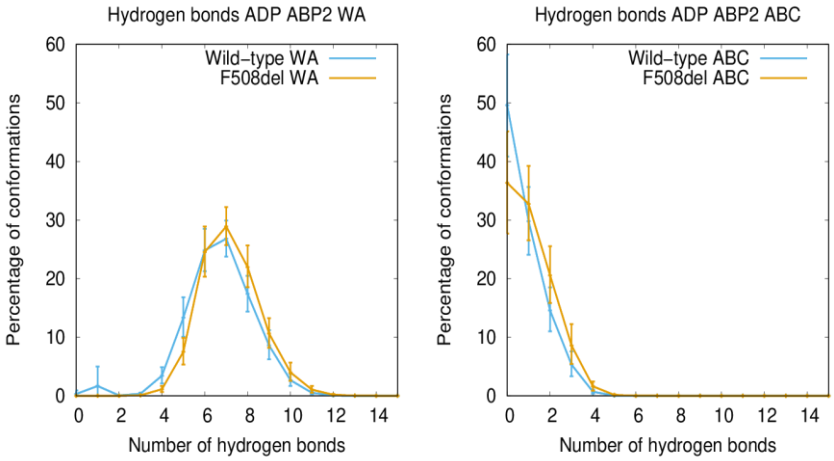


Figure 3.18- *Histograms depicting the number of hydrogen bonds between ADP with the Walker A motif (residues 1246 to 1350 on wild-type) and the ABC signature motif (residues 548 to 552 on wild-type) on ABP2. The error bars correspond to the 95% confidence interval calculated by bootstrapping. WA- Walker A, ABC-ATP binding cassette. These plots were made from data from the last 130 ns of the simulations. The ATP state only comprises the ATP simulations suited for hydrolysis at 20 ns.*

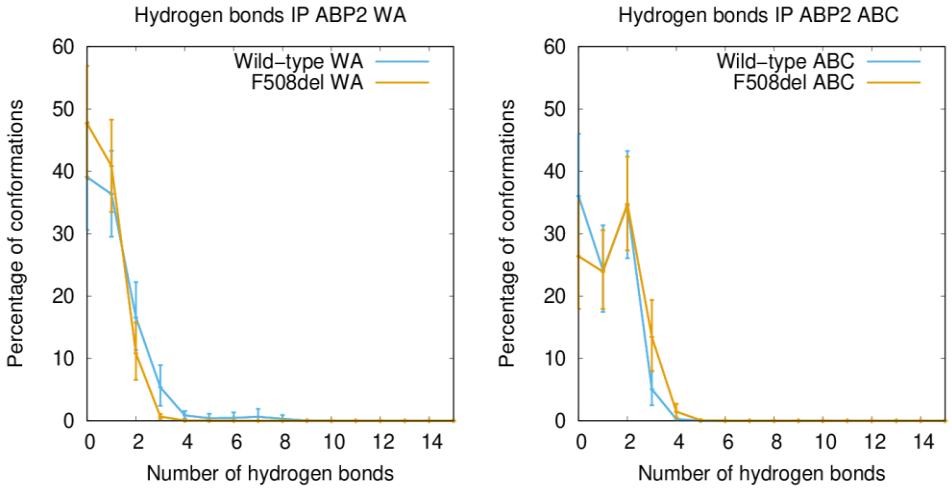


Figure 3.19- *Histograms depicting the number of hydrogen bonds between phosphate (IP) with the Walker A motif (residues 1246 to 1350 on wild-type) and the ABC signature motif (residues 548 to 552 on wild-type) on ABP2. The error bars correspond to the 95% confidence interval calculated by bootstrapping. WA- Walker A, ABC-ATP binding cassette. These plots were made from data from last 130 ns of the simulations. The ATP state only comprises the ATP simulations suited for hydrolysis at 20 ns.*

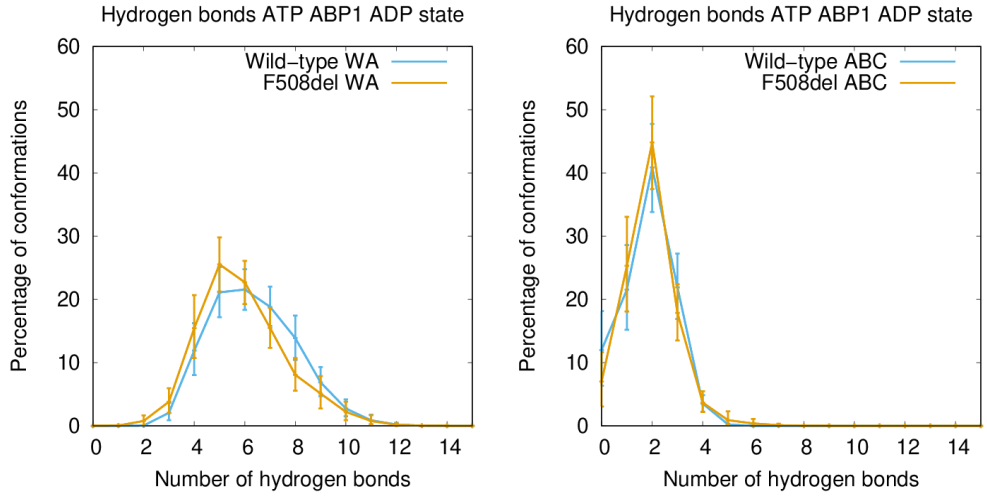


Figure 3.20- *Histograms depicting the number of hydrogen bonds between ATP with the Walker A motif (residues 458 to 466 on wild-type) and the ABC signature motif (residues 1346 to 1350 on wild-type) on ABP1. The error bars correspond to the 95% confidence interval calculated by bootstrapping. WA- Walker A, ABC-ATP binding cassette. These plots were made from data from the last 130 ns of the simulations. The ATP state only comprises the ATP simulations suited for hydrolysis at 20 ns.*

Chapter 4

Nucleotide binding modes during the reverse adenylate kinase reaction in MsbA.

This work was published in the following paper:

Kaur, H., **Abreu, B.**, Akhmetzyanov, D., Soares, C., Prisner, T.; Glaubitz, C., Unexplored nucleotide binding modes for the ABC exporter MsbA. *Journal of the Americal Chemical Society*, **140**, 14112

Contributions of the author of the present thesis to this work:

In this project I contributed with the molecular modelling results. I performed all simulations, analysed the results, and wrote the part of the manuscript corresponding to this portion of the work.

4.1-Summary

The ATP-binding cassette (ABC) transporter MsbA is an ATP-driven lipid-A flippase. It belongs to the ABC protein superfamily whose members are characterized by conserved motifs in their nucleotide binding domains (NBDs), which are responsible for ATP hydrolysis. Recently, it was found that MsbA could catalyze a reverse adenylate kinase (rAK)-like reaction in addition to ATP hydrolysis. Both reactions are connected and mediated by the same conserved NBD domains. Here, the structural foundations underlying the nucleotide binding to MsbA were therefore explored using a concerted approach based on conventional- and DNP-enhanced solid-state NMR, pulsed-EPR, and molecular dynamics (MD) simulations. MsbA reconstituted into lipid bilayers was trapped in various catalytic states corresponding to intermediates of the coupled ATPase-rAK mechanism. The analysis of nucleotide-binding dependent chemical shift changes, and the detection of through-space contacts between bound nucleotides and MsbA within these states provides evidence for an additional nucleotide-binding site in close proximity to the Q-loop and the His-Switch. By replacing Mg^{2+} with Mn^{2+} and employing pulsed EPR spectroscopy, evidence is provided that this newly found nucleotide binding site does not interfere with the coordination of the required metal ion. Molecular dynamics simulations of nucleotide and metal binding required for the coupled ATPase-rAK mechanism have been used to corroborate these experimental findings and provide additional insight into nucleotide location, orientation, and possible binding modes.

4.2-Introduction

Transport of substrates across the cell membrane is vital to all living organisms to maintain homeostasis. Integral membrane proteins belonging to various transporter families provide transmembrane pathways for a plethora of substrates, ranging from ions to macromolecules. Of particular importance are the ubiquitous ATP binding cassette (ABC) transporters, typically forming dimers, which contain substrate-specific transmembrane domains (TMDs) and highly conserved nucleotide-binding domains (NBDs) (figure 4.1a). They can be further subdivided into importers and exporters and form a part of the ABC superfamily. This family also contains soluble proteins without TMDs that are associated with diverse functions such as translation processes or chromosome segregation. All of these proteins use ATP hydrolysis mediated by their NBDs to carry out their respective functions. Each NBD comprises of the highly conserved A-loop, Walker A, Walker B, Q-loop, His switch, D-loop, and the signature motif, which form a cluster of interactions to coordinate nucleotide binding[1,5] (figure 4.1b).

MsbA is an ABC exporter of prokaryotic origin that translocates lipopolysaccharide (LPS) across the cell membrane[340,341]. Crystal structures of MsbA provide an insight into different conformational changes during the catalytic cycle[96]. The nucleotide-bound structures adopt an outward-open conformation, while an inward opening with different extents of NBD separations is observed for the apo state. Recently, cryo-EM structures of MsbA in nanodiscs with co-purified LPS have been reported[129]. Both nucleotide-free and nucleotide-bound structures differ perceivably from the crystal structures: nucleotide-free, LPS-bound MsbA adopts a closed apo state in which two NBDs are aligned in a head-to-tail fashion. The distance between the Walker A motif of NBD1 and signature motif of NBD2 is approximately two times smaller compared to the closed apo crystal structure. Furthermore, the vanadate trapped cryo-EM structure assumes an occluded state, which is different from the outward facing conformation observed under crystallization conditions. Despite these differences, all the structures represent progress toward understanding how LPS transport can possibly take place.

The NBD-catalyzed process of ATP hydrolysis is the key event for the functional mechanism of ABC proteins such as MsbA. However, we recently demonstrated by ^{31}P-MAS (magic angle spinning) NMR that MsbA is also able to catalyze a reverse adenylate kinase (rAK) like phosphoryl transfer reaction (2ADP→ATP+AMP) under low ATP conditions. This reaction is connected to the primary ATP hydrolysis (ATP → ADP+Pi)[59]. In this situation, ADP created from ATP via the MsbA catalyzed hydrolysis is partially regenerated into ATP by the rAK activity and hydrolyzed again. Such a coupled ATPase-rAK activity could be of importance during situations of ATP depletion in the cell. Its general relevance is emphasized by the fact that the rAK activity was also found for the ABC exporters LmrA and TmrAB for the ATP-gated ion channel CFTR (cystic fibrosis transmembrane conductance regulator), and for some soluble ABC proteins[60,61,342,343].

The observation that MsbA is able to catalyze two reactions in a cyclic manner has important structural consequences: two bound ADPs have to assume an appropriate binding mode for the phosphoryl transfer reaction to take place. But the canonical, high affinity binding sites needed for ATP hydrolysis are too far away from each other to fulfill this condition (figure 4.1b). Therefore, an additional site for nucleotide binding in between both canonical sites was proposed (figure 4.1b) based on mutational studies[59]. These data pointed in particular at Q424 within the Q-loop motif to play a role in mediating the rAK reaction, indicating its involvement in forming the additional nucleotide-binding site. In the following text, we will refer to this site as the ‘adenylate kinase (AK)-site. Furthermore, the protein was suggested to cycle through a number of intermediate states with differently bound nucleotides (figure 4.1c)[59]: ATP binding to apo MsbA creates the well-established pre-hydrolysis state (**1**), and the process of ATP hydrolysis leads to a high-energy transition state (HES), (**2**). States **(1)** and **(2)** can be emulated by using

ADP·BeF_3 and ADP·VO_4, respectively. Next, ADP is displaced by ATP from the high affinity site and binds to the AK-site (**3**), which can be emulated by trapping MsbA first with ADP·VO_4 followed by binding of ADP·βS.

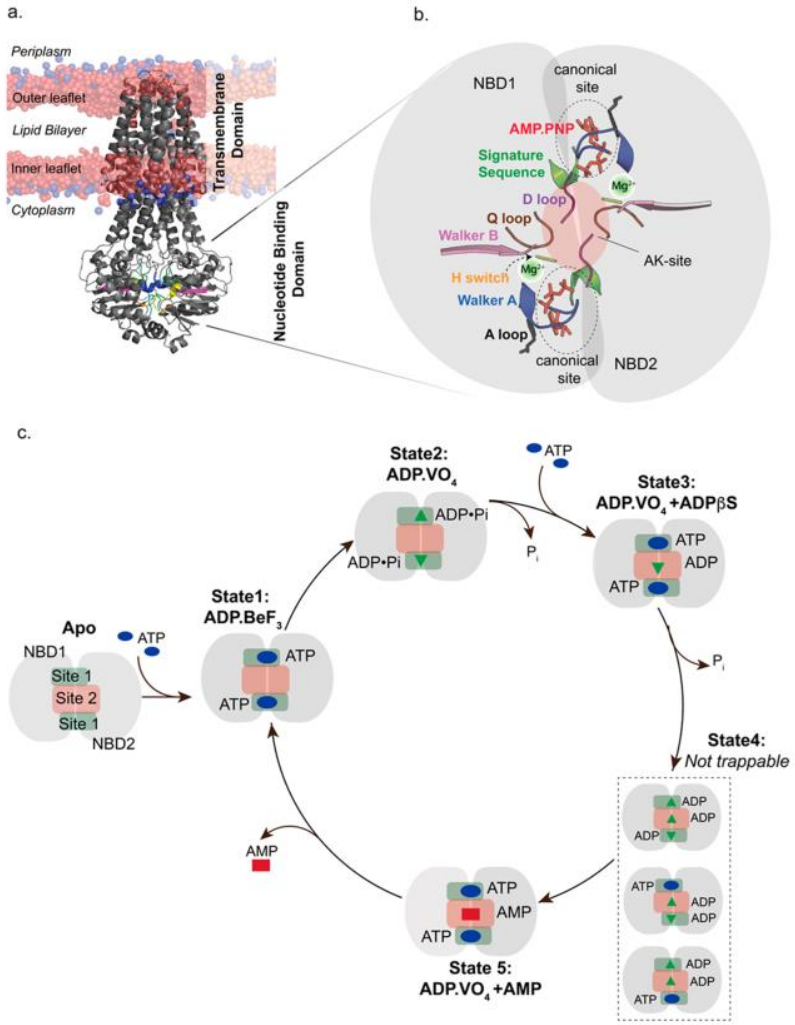


Figure 4.1- *ABC exporter MsbA. (a) MsbA within the lipid bilayer (cartoon based on PDB 3B60 using MemProtMD). (b) Cartoon representation of the MsbA NBD dimer, conserved motifs, canonical-, AK-, as well as metal-ion binding sites are highlighted. (c) MsbA cycles through a number of intermediate states needed for the coupled ATPase-rAK mechanism. These intermediates need to be stabilized for structural analysis: the prehydrolysis state (**1**) with ATP in the canonical sites can be emulated by ADP·BeF_3. The ATPase high-energy transition state (**2**) can be trapped by ADP·VO_4. ADP gets released, binds to the AK-site, while ATP binds again to the canonical site (**3**), which can be stabilized by ADP·VO_4 and ADP-βS. Subsequently, a state is created in which at least one ADP in the canonical site and one ADP in the AK-site must assume orientations suitable for the phosphoryl transfer reaction yielding ADP and AMP (**4**). This state cannot be trapped but will result in state (**5**), which can be emulated by ADP·VO_4 and AMP. The NBD cartoon is based on the MsbA structure PDB 3B60.*

A subsequent ATP-hydrolysis step leads to a state in which one or both canonical sites as well as the AK-site are occupied with ADP, so that two ADP molecules can bind in such a way that the phosphoryl transfer reaction can take place (**4**). As a result, a state containing ATP and AMP is

created (**5**), which can be emulated by ADP·VO_4 trapping followed by AMP binding. States **(1)**, **(2)**, **(3)**, and **(5)** have been successfully trapped, and bound nucleotides have been detected by ^{31}P cross-polarization MAS NMR experiments[59]. These data confirmed the possibility that more than two nucleotides can bind to the NBDs and that the AK-site indeed exists. However, a direct evidence for interactions of MsbA with nucleotides bound in the canonical site as well as in the AK-site is still missing. Here, we provide such data based on solid-state NMR spectroscopy.

Spectroscopic approaches, in general, are very important for bridging the gap between biochemical and structural data. Solid-state NMR offers a great potential to probe membrane proteins with respect to their activity, structure, and dynamics directly within lipid bilayers. Applications range from highly specific, hypothesis-driven studies to full structure determination projects[344–350]. Sensitivity could be increased by orders of magnitude through dynamic nuclear polarization[351] by which novel questions or difficult proteins become now accessible[352–356] even within the cellular context[357]. As solid-state NMR does not rely on a molecular size limit, also large complexes such as ABC transporters could be investigated. However, an increase in size results in increasing spectral complexity, which needs to be reduced by well-tailored experimental approaches[358,359]. Examples include ^{31}P-MAS NMR on MsbA to probe nucleotide binding and its catalytic activity, DNP-enhanced MAS NMR to determine the structure of a peptide substrate in the binding site of a human ABC transporter complex[360] or to probe site resolved conformational dynamics in MsbA in different states[361] and paramagnetic NMR to obtain structural information about the nucleotide and cofactor binding[362]. Furthermore, a number of studies reported novel solid-state NMR approaches or applications on nucleotides, which might be also beneficial for the study of ABC transporter[363–365]. In addition, EPR spectroscopy in combination with site-directed mutations and spin labeling has been very extensively used[103,104,366,367]. The portfolio of methods was expanded recently toward high spin and hyperfine spectroscopy[363,368].

Here, the structural foundations underlying nucleotide binding to MsbA were explored using a concerted approach based on conventional- and DNP-enhanced solid-state (MAS) NMR, pulsed-EPR, and MD simulations. MsbA reconstituted into lipid bilayers was trapped in the states defined in figure 4.1c, which correspond to intermediates of the coupled ATPase-rAK cycle. The analysis of nucleotide-binding dependent chemical shift changes provides evidence for the postulated AK-site (i.e., the involvement of Q-loop and His-Switch). Direct contacts between MsbA residues and nucleotides residing in the canonical as well as in the AK-site could be detected for states (**1**), (**2**), and (**5**). While the former visualizes an expected interaction with the A-loop, the latter demonstrates the close proximity of the AK-site to the Q-loop. By replacing Mg^{2+} with Mn^{2+} and employing pulsed EPR spectroscopy, evidence is provided that occupying

both canonical- and AK-site(s) does not interfere with the coordination of the required metal ion. Finally, MD-simulations on the $MsbA^{nbds}$ were carried out for defining the nucleotide and metal binding modes in state (**5)** (figure 4.1c). The combination of MAS NMR, DNP, and pulsed Mn-EPR on full length MsbA within lipid bilayers with MD simulation on the NBDs provides first structural insight into the so far unexplored binding modes required for the coupled ATPase-rAK mechanism.

4.3-Materials and Methods

Protein expression, purification, and reconstitution into lipid bilayers was essentially carried out as described before and has also been described in methods in the supplementary information[129,369].

4.3.1-MAS NMR

Preparation [u-^{13}C-Val,$^{15}N_1$-His]-MsbA and [u-^{13}CAla, $^{15}N_1$-His]-MsbA: samples for the NCOCX experiments have been prepared by supplementing M9 media with $^{15}N_1$-His and u-^{13}C-Val or u-^{13}C-Ala, respectively. All other amino acids and glucose were added in an unlabeled form to the media.

Preparation of [^{13}C,^{15}N-His,^{13}C]-MsbA: samples have been prepared by supplementing M9 minimal media with ^{13}C,^{15}N-His (0.06 g*) along with $^{13}C_6$-glucose. For the apo- and BeF_3 trapped samples only, 15N1-His (0.06 g*) has been used. Unlabeled arginine (0.25 g*), cysteine (0.03 g*), lysine (0.25 g*), phenylalanine (0.08 g*), proline (0.06 g*), tyrosine (0.1 g*), tryptophane (0.03 g*,) and serine (0.125 g*) have been added to M9 media to make it richer in its components in order to increase the yield of isotope-labeled MsbA. The asterisk denotes quantities for 600 mL M9 minimal media.

The trapping of catalytic intermediate states of MsbA proteoliposomes has been described in previous papers and in the Supporting Information[129,361,369].

NMR Experiments.

The NCOCX spectra were recorded using a Bruker Avance 600 MHz spectrometer equipped with a triple resonance 4 mm e-free probe. All spectra are recorded on a 22−25 mg MsbA sample in a 4 mm rotor with a MAS spinning rate of 10 kHz at a nominal temperature of 250 K. ^{13}C- and ^{15}N- chemical shift referencing was carried out with respect to DSS through alanine (CO resonance corresponding to 179.85 ppm). For all experiments, 83.3 kHz decoupling using SPINAL-64 was applied during evolution and acquisition. ^{1}H- and ^{13}C-90°-pulses are set to 3 and 4.5 µs, respectively. For NCO(CX) experiments, a CP contact time of 1.0 ms for the first H−N CP

transfer step and 3.2 ms for the second NC−CP step was applied. 1D-NCOCX spectra were recorded with 144k acquisitions which took ∼23 h for acquisition. All samples were doped with 1 mM Gd-DOTA for faster acquisition[369] by shortening the recycle delay time to 1 s.

4.3.2-DNP-Enhanced MAS NMR.

Sample Preparation.

[U-^{13}C]-MsbA was prepared by expression in M9 minimal media supplemented with ^{13}C labeled glucose. Reconstituted sample was trapped as described above but with ^{15}N-labeled nucleotide. For doping with polarizing agent, eight aliquots of the resuspended reconstituted sample (∼17−20 mg) were incubated with 7−10 µL of 10% H_2O, 30% d^8-glycerol, 60% D_2O, and 20 mM AMUPol[370] at 4 °C for 3.5 h. This procedure provided an optimum between preserving a high trapping efficiency and even distribution of AMUPol[370] for maximum signal enhancement.

NMR Experiments.

DNP-enhanced MAS NMR spectra were recorded on a Bruker 400 DNP system consisting of a 400 MHz WB Avance II NMR spectrometer, a 263 GHz Gyrotron as microwave source, and a 3.2 mm HCN LT-MAS probe. All experiments were conducted with 10 kHz MAS and a microwave power of 11.5 W. During DNP experiments, the nominal temperature was kept at around 105 K. Chemical shift referencing for 13C and 15N was carried out indirectly with respect to DSS using the alanine CO resonance to 179.85 ppm. For all experiments, 100 kHz decoupling using SPINAL-64 was applied during acquisition.

Two-dimensional (2D) 15N−13C correlation spectra were acquired using the z-filtered TEDOR sequence[371]. Typical 1H, 13C, and 15N 90° pulse lengths were 2.5, 4, and 7.5 µs, respectively. The 1H-13C-CP (cross-polarization) contact time was 1000 µs, and a recycle delay of 2.5 s was applied. The TEDOR mixing time was set to 6.8 ms. The z-filter had a duration of 250 µs. All 2D-spectra were recorded with an acquisition time of 15 ms in the direct dimension and 20 increments of 100 µs each in the indirect dimension with 4k scans per increment. It took approximately 2.5 d in total to acquire one TEDOR spectrum. The 15N pulse carrier was set to 60 ppm, and the 13C pulse offset was set to 70 ppm. A Gaussian and a squared-cosine window function were applied to the direct and indirect dimension prior to Fourier transformation. The 2D TEDOR spectra were plotted with a factor of 1.1 between subsequent contour lines.

4.3.3-EPR Spectroscopy.

Sample Preparation

Unlabeled reconstituted MsbA was trapped with Mn·ADP·VO_4 and Mn·ADP·VO_4+AMP. The high-energy transition state that occurs during ATP hydrolysis can be emulated using vanadate as described above. Per 10 mg/mL (~150 μM) of reconstituted MsbA 40 mM ATP, 10 mM $MnCl_2$ and 10 mM orthovanadate were added. After vanadate trapping the samples, AMP was added in a 3:1 AMP:MsbA molar ratio. The trapped samples were washed thrice with 20 mM HEPES and then resuspended in 20 mM HEPES (pH 7.3) + 20% glycerol in order to remove free Mn^{2+} and nucleotide.

9.4 GHz Continuous Wave EPR Studies at Room Temperature

Continuous wave (CW) EPR experiments at 9.4 GHz frequencies were performed at Bruker Elexsys E500 spectrometer equipped with SHQE cavity. Measurements were performed with 20 μL sample volume using 12.63 mW microwave power, 0.5 mT modulation amplitude, with 100 kHz modulation frequency and 81.92 ms for both conversion time and time constant. The magnetic field sweep width was 0.4 T (with the center field of 0.335 T). The whole sweep time of a spectrum was 335.54 s. The receiver gain was set to 66 dB. One scan for each measurement was obtained.

263 GHz Pulsed EPR Studies at Low Temperatures

High-frequency pulsed EPR studies were performed at 263 GHz Bruker Elexsys E780 spectrometer equipped with a 9.4 T superconducting magnet (Ascend 400 MHz DNP), a quasi-optical front-end, and an arbitrary waveform generator, which was used for generating 8-step phase cycle needed for 5-pulse RIDME experiment. Five pulse RIDME sequence was used[372]. The time traces are obtained at the maximum of the lowest field hyperfine line. Pulse lengths for π/2/π pulses were 22/36 ns and 26/42 ns for MsbA·Mn·ADP·VO_4 and MsbA·Mn·ADP·VO_4+AMP, respectively. An interpulse delay between the first π/2 and π pulses of 400 ns was used. The initial delay time between the π pulse and the first π/2 pulse of the mixing block was 200 ns and incremented with a step of 8 ns and 150 points. The mixing time (time delay between two π/2 pulses in the mixing block) was 30 μs, which corresponds to T1e of Mn spins in both MsbA samples. The initial delay time between the second π/2 pulse of the mixing block and the last π pulse was 1.3 μs and was decremented with a step of 8 ns and 150 points. The shot repetition time was set to 500 μs with 100 shots per point. For MsbA·Mn·ADP·VO_4 and MsbA·Mn·ADP·VO_4+AMP, 700 and 500 scans with accumulation times of 14 and 10 h,

respectively, were recorded. Due to phase drift of our spectrometer, data were stored every 10 scans and subsequently phase corrected.

4.3.4-Molecular Dynamics Simulations.

System Setup.

The crystal structure of *Salmonella typhimurium* MsbA (PDB code 3B60) was used as a starting point for this work[96]. Only the nucleotide binding domains (NBDs) were simulated because it was assumed that the transmembrane domains were not relevant for the nucleotide interactions with the protein. The *S. typhimurium* NBDs share a 96.3% identity with *E. coli* NBDs, allowing the construction of a homology model of the nucleotide-binding domains (residues 325−581). In addition, parts of the transmembrane domains (109−121 and 209−226) were included as part of our system, because they are deeply buried in the nucleotide binding domains. The model was created using MODELLER[289]. The AMP-PNP residues present in *S. typhimurium* NBDs were replaced by ATPs, while the Mg^{2+} ion was inserted by alignment with the structure of the MalK dimer (PDB code 1L2T)[26]. The protonation states of the ionizable residues of the protein were determined at pH 7.0 as done in previous works[161,171,316,317]. Further details are provided as Supporting Information.

Simulation Setup.

The molecular dynamics (MD) simulations were performed with the GROMACS package version 5.0.7[318] using the GROMOS 54A7 force field[256]. The system with ATP on the active site was solvated in a dodecahedral box and 20 sodium ions were added in order to neutralize the system. The system was minimized, using position restraints of 1000 kJ mol^{-1} nm^{-1} on all non-hydrogen atoms, until the energy converged to machine precision, or the maximum number of steps was achieved or the maximum force on any atom was negative. This procedure was followed by a second minimization, with position restraints of 1000 kJ mol^{-1} nm^{-1} on C-α atoms. This second minimization was stopped when one of the above criteria was attained. The equilibration process was initiated with 100 ps of molecular dynamics at constant volume using a Berendsen thermostat with a coupling constant of 0.1 ps[267], using position restraints of 1000 kJ mol^{-1} nm^{-1} on all nonhydrogen atoms. Afterward, another 100 ps of simulation at constant temperature of 310 K and 1 bar pressure were done, using position restraints of 1000 kJ mol^{-1} nm^{-1} on the C-α atoms of the transmembrane segments in order to keep their position relative to the nucleotide-binding domains. A Berendsen coupling bath was used in this step to keep a constant temperature and pressure. Five replicates of the system with ATP on the active site were simulated during 200 ns, keeping the position restraints on the C-α atoms of the

transmembrane segments. On these production simulations, the velocity rescale[268] thermostat was used as well as the isotropic Parrinello−Rahman pressure coupling[270]. In all simulations, the integration step used was 2 fs. All bonds were constrained to their equilibrium length using the LINCS algorithm[259], except for water molecules in which the SETTLE algorithm was used[258]. The SPC model was used[373]. The van der Waals and electrostatic interactions were calculated up to 1.0 nm, and the neighbor list was updated every 10 steps. PME (Particle Mesh Ewald) was used to treat electrostatic interactions[319].

Simulations of Slow-Growth.

AMP was inserted in the MsbA NBD dimer using a slow-growth methodology, in which the molecule is grown in a given position, coupled to a lambda parameter that varies from 0 to 1. This growth procedure is made by increasing the van der Waals and electrostatic interactions of the molecule with the rest of the environment, allowing for smooth insertion and allowing the protein to adapt to AMP. This procedure is generally used for the calculation of free energies. This approach was chosen because the two NBD monomers were too close, and there was not enough free space to insert AMP using docking or a similar method. This was made several times in order to collect a statistically robust description of the phenomenon, and the different initial conformations for this process were extracted from the simulations with ATP on the active site (every 10 ns after the first 20 ns). A total of 95 conformations were used. The AMP molecules were initially placed in the dimer according to the experimental information, with the adenine moiety close to the Q-loop and the phosphate group oriented toward the histidine 537. AMP was inserted in each conformation by alignment with the conformation prior to minimization, in which the AMP molecules were inserted. Each slow-growth procedure was conducted during 1 ns, using a time step of 0.0005 ps. The neighbour list was updated every five steps. Extra position restraints were added to the C-α atoms of the residues that were more than 8 Å away from the AK-site, in order to stabilize the protein during the slow-growth process. After the slow-growth process, the output conformations remained bound near the Q-loop and the His-switch were selected. These conformations were further clustered according to their conformation on the AK pocket. The RMSD against the initial structure, previous to minimization, was the measure for clustering, and the Jarvis-Patrick method[374] was used as with a RMSD cutoff of 1 Å (see further parameters in the Supporting Information). All the slow-growth output conformations, prior to clustering, were used as starting points of MD simulations, in order to maximize sampling. These simulations were performed in the same conditions as previously described above.

4.4-Results

4.4.1-Chemical Shift Changes in MsbA during the Catalytic Cycle.

In order to obtain experimental evidence for the location of the AK-site within the conserved NBD motifs, a unique pair-labeling approach was used. For this purpose, all histidines in MsbA were ^{15}N- and selected preceding residues ^{13}C-labeled (see materials and methods). Two unique pairs, Val426-His427 in the previously proposed Q-loop (423-SQN**VH**-427) and Ala536-His537 (534-VI**AH**RLS-540) in the conserved His-switch, could be identified (figure 4.7a). An COCX magnetization transfer from ^{15}N in residue (i) to ^{13}C nuclei in residue (i-1) allowed the selective detection of either Val426 or Ala536, which then served as reporters for the catalytic cycle in the Q-loop and His-switch, respectively. Therefore, two samples, [U-^{13}C-Val,^{15}N1-His]-MsbA and [U-^{13}C-Ala,^{15}N1-His]-MsbA, were prepared. MsbA was reconstituted into DMPC/DMPA lipid bilayers. Such proteoliposomes were used for all experiments described in this paper. The resulting NCOCX spectra of both pairs for the apo-state, the high energy state (state (**2**): ADP·VO_4) and states with nucleotide binding to the AK-site (state (**3**): ADP.VO_4+ADPβS; state (**5**): ADP·VO_4+AMP) are shown in figure 4.2.

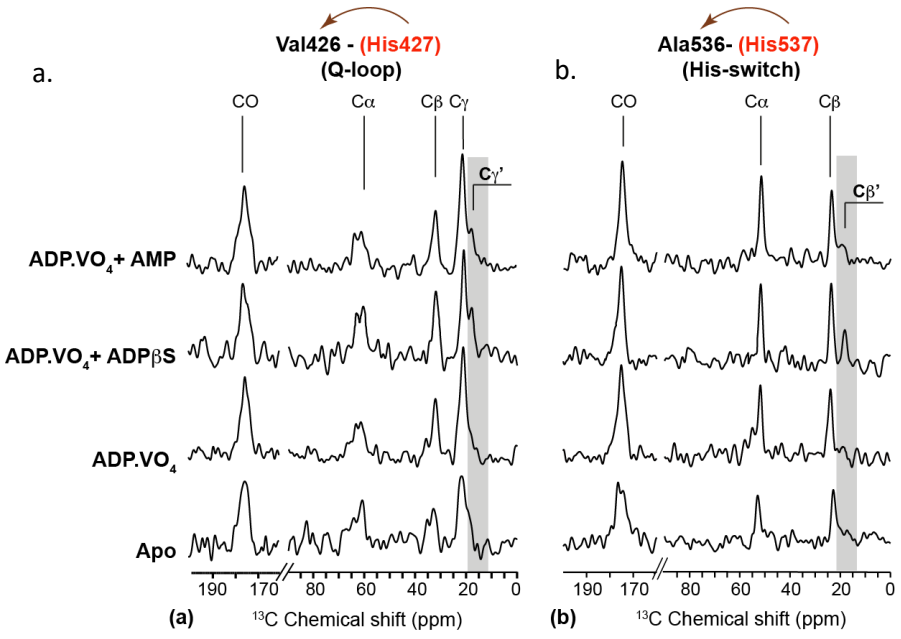


Figure 4.2- *NCOCX spectra of unique pair-labeled MsbA in different catalytic intermediate states at 600 MHz at 250 K.* ***(a)*** *[U-^{13}C-Val, ^{15}N1-His]-MsbA contains a unique pair ^{13}C-Val426-^{15}N-His427 in the Q-loop. A NCOCX experiment transfers magnetization from His427 to Val426, which can then be detected. An additional resonance, Cγ' shifted by -3 ppm, is observed in intermediate states (**3**) and (**5**) with additional nucleotide bound to the AK-site (ADP.VO_4+ADPβS, ADP.VO_4+AMP).* ***(b)*** *[U-^{13}C-Ala, ^{15}N1-His]-MsbA contains a unique pair ^{13}C-Ala536-^{15}N-His537 in the His-switch. As for Val426, an additional side chain resonance, Cβ' shifted by -4 ppm, is observed in states (**3**) and (**5**).*

For Val426, all carbons can be detected in all prepared states (figure 4.2a). A significant effect is observed for the Cγ resonance in states (**3**) (ADP·VO_4+ADPβS) and (**5**) (ADP·VO_4+AMP) (i.e., the states in which both the AK- as well as the canonical sites are supposed to be occupied): here, a second peak shifted by −3 ppm occurs (Cγ'). Similarly, all expected resonances of Ala536 can be detected (figure 4.2b), and an additional side chain resonance Cβ' shifted by −4 ppm is observed for states (**3**) and (**5**). These observations support the hypothesis that the Q-loop is indeed involved in the AK-site formation. Our findings also indicate a role for the His-switch in either binding or sensing of nucleotides in proximity to the AK-site. Notably, Ala536-Cβ and Val426-Cγ are not just shifted, but instead additional resonances Cβ' and Cγ' occur. One reason could be incomplete trapping or saturation of the AK-site: States (**3**) and (**5**) are prepared by creating first the ADP·VO4-trapped state followed by the addition of ADPβS or AMP, which could result in restricted accessibility or just one rather than two nucleotides binding at the same time.

Furthermore, to follow global changes during the coupled ATPase-rAK cycle, the unique-pair labeling scheme described above was expanded by uniform ^{13}C-labeling of MsbA together with ^{15}N-labeling of all histidines. In this way, nine ^{13}C–^{15}N residues pairs are created within MsbA (figure 4.7, panels a−c and table 4.1). These are uniformly distributed within the NBD and at the intracellular side of MsbA so that a characteristic fingerprint spectrum for different catalytic states can be obtained. The resulting 2D NCO spectra reveal that the backbone architecture of MsbA is different for each state (figure 4.8).

4.4.2-Direct Nucleotide Protein Contacts Using DNP Enhanced Solid State NMR.

Our data suggests an involvement of the Q-loop in the AK-site. In order to provide further structural evidence for this hypothesis, we probed direct nucleotide-MsbA contacts in different intermediate states of the ATPase-rAK cycle (figure 4.1c) by means of dipolar ^{13}C-^{15}N through-space spectroscopy. For this purpose, complexes of U-^{13}C-MsbA with ^{15}N-labeled nucleotides in differently trapped states were prepared. The NMR sensitivity was significantly improved by utilizing dynamic nuclear polarization. Sample doping by incubation with 20 mM AMUPol[370] for 3.5 h lead to a 25-fold signal enhancement for carbons (figure 4.3a; for ^{15}N, see figure 4.9a). Higher enhancements can be achieved by a longer incubation time, but this would have compromised the trapping efficiency, which is biochemically reversible[375].

On the basis of these conditions, DNP-enhanced ^{15}N-^{13}C TEDOR spectra of the canonical states (**1**) (U-^{13}C-MsbA+^{15}N-ADP·BeF_3) and (**2**) (U-^{13}C-MsbA+^{15}N-ADP·VO_4) were recorded (figure 4.4, panels b and c). In order to differentiate intramolecular natural abundance cross peaks from intermolecular nucleotide-protein contacts, a control experiment with ^{13}C-MsbA in complex with unlabeled nucleotide was acquired (figure 4.9b). Distinct N-CO, N-CA, as well as Arg-Nηε-

Cζ cross peaks arising from ~0.4% ^{15}N-natural abundance in ^{13}C labeled protein could be identified[354]. Any additional peaks observed from ^{13}C-MsbA in complex with ^{15}N-labeled nucleotide can therefore be assigned to MsbA-nucleotide contacts. Only weak correlations between ADP and carbons in the aromatic region could be detected in state (**1**) (figure 4.3b), but the same peaks occur more pronounced in state (**2**) (figure 4.3c). The ^{15}N-ADP resonances can be assigned based on their known chemical shift range[376]. The only aromatic residue close to the nucleotide in the canonical binding site is the well characterized A-loop Tyr351, which forms π−π stacking interactions with adenosine, as seen in the MsbA as well as in other prokaryotic ABC exporter crystal structures (see figure 4.3d)[96,377]. The strongest contact in figure 4.3c can therefore tentatively be assigned to Tyr351-Cδ and ADP-N9, while the other, weaker cross peak probably arises from Tyr351-Cζ and ADP-N7. The large differences in cross peak intensities within state (**2**) and between states (**1**) and (**2**) could be caused by different nitrogen−carbon distances. Furthermore, the trapping efficiencies between states (**1**) and (**2**) might differ.

After probing these canonical MsbA-nucleotide contacts, we sought to detect interactions with nucleotide(s) occupying the additional AK-site, which should occur in states (**3**), (**4**), and (**5**). State (**3**) could not be studied as ^{15}N-labeled ADP-βS is not available, and state (**4**) could not be trapped as it is a transition state of the phosphoryl transfer reaction (figure 4.1c). Hence, we focused on state (**5**) with ADP·VO_4+AMP.

In this case, U-^{13}C-MsbA was prepared in complex with unlabeled ADP·VO_4 and ^{15}N-AMP (figure 4.4a). A new correlation between ^{15}N-AMP-NH_2 at 74 ppm and a ^{13}C-carbonyl side chain at 178 ppm occurs in the TEDOR spectrum. On changing the ^{15}N-offset, an additional cross peak was obtained at a ^{13}C-chemical shift of 180 ppm with the N9 nitrogen of AMP at 172 ppm (figure 4.9c). A complementary sample was prepared with U-^{13}C-MsbA in complex with ^{15}N-ADP·VO_4 and unlabeled AMP. The TEDOR spectrum (figure 4.4b) shows the canonical contacts between ADP and A-loop Tyr as discussed above, while the correlation between AMP and MsbA is not observed as expected. Notably, the Q-loop in *E.coli* MsbA contains residues -S**QN**VHL-. Since both observed cross peaks correspond to a ^{13}C-chemical shift of a carbonyl side chain, they could originate from either ^{13}C-Glutamine-Cδ or ^{13}C-Asparagine Cγ within the Q-loop.

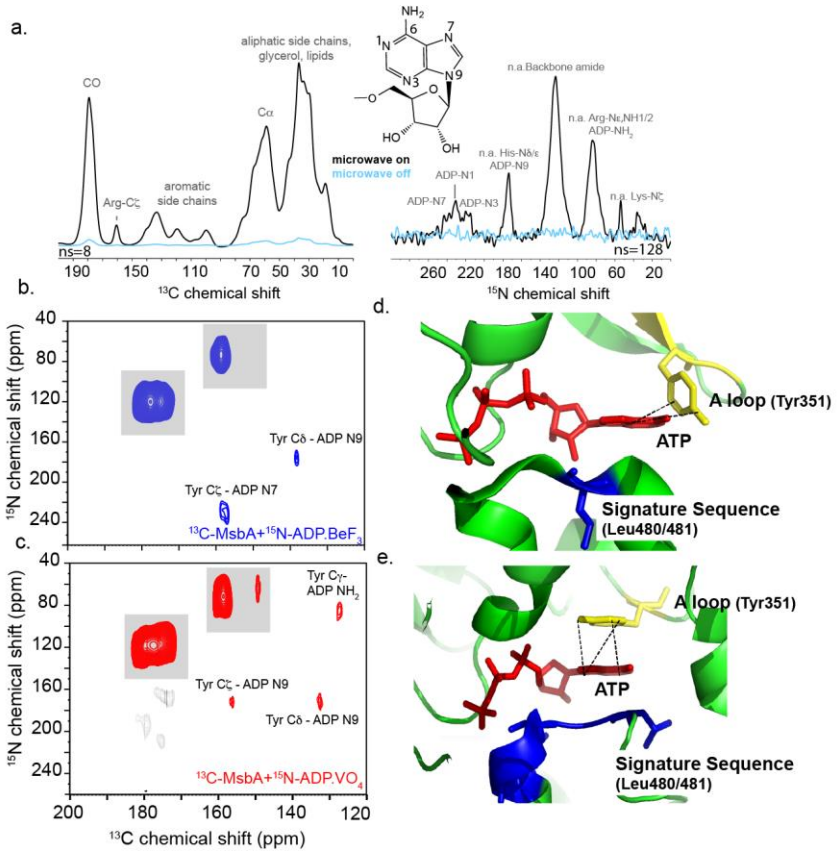


Figure 4.3- *DNP-enhanced TEDOR spectra on U-^{13}C-MsbA/^{15}N-nucleotide complexes in the pre-hydrolysis and in the catalytic transition state.* ***(a)*** *DNP-enhanced ^{13}C (left) and ^{15}N (right) CP-MAS NMR spectra. Spectra were recorded at 400MHz / 263GHz at 100K. Samples were incubated with 20 mM AMUPol in 10% H2O, 60% d8-glycerol, 30% D_2O for 3.5 h, which lead to a 25-fold signal enhancement for carbon.* ***(b)*** *^{15}N-^{13}C TEDOR spectrum of U-^{13}C-MsbA in complex with ^{15}N-ADP-BeFx (pre-hydrolysis, state 1 in Figure 1). Cross peaks between ADP nitrogens N7 (230 ppm) and N9 (174 ppm) with Tyr351 Cδ (138 ppm) and Cζ (156 ppm) are labeled.* ***(c)*** *U-^{13}C-MsbA in complex with ^{15}N-ADP.VO_4 (pre-hydrolysis, state 2 in Figure 1). Cross peaks between ADP nitrogens N9 (173ppm) and NH_2 (85.4 ppm) with Tyr351 Cδ (132.4ppm), Cζ (156ppm) and Cγ (127.2ppm) are labeled. The TEDOR spectra were recorded with a mixing time of 6.8 ms. Natural abundance cross peaks in (b) and (c) are highlighted in gray.* ***(d)*** *Illustrated contacts between the nucleotide and protein in a prehydrolysis state modeled according to the crystal structure of MalK with BeF_3 (pdb id: 3PUX)* ***(e)*** *A-loop Tyr 351 in a π-π stacking interaction with the bound nucleotide at the canonical binding site as depicted in the crystal structure of MsbA with AMP.PNP (pdb id: 3B60). Contacts observed in (b) and (c) are only illustrations of observed cross peaks shown in dotted black lines.*

In order to assign the Q-loop contact with AMP unambiguously, the single point mutation Q424A was introduced. It was found by using ^{31}P-CP MAS NMR experiments that this mutation reduces the ADP·VO_4 trapping efficiency (figure 4.4e). Therefore, U-^{13}C-MsbA and U-^{13}C-$MsbA_{Q424A}$ had to be prepared in complex with ^{15}N-AMP instead of ADP·VO_4+^{15}N-AMP. The TEDOR spectrum for the wild-type sample shows the same cross peak between ^{15}N-AMP-NH_2 at 74 ppm and a ^{13}C-carbonyl side chain at 178 ppm as described above (figure 4.4c), which

disappears upon introducing the Q424A mutation (figure 4.4d). This peak can be therefore assigned to a contact between AMP-NH_2 and Cγ of Q424 in the Q-loop.

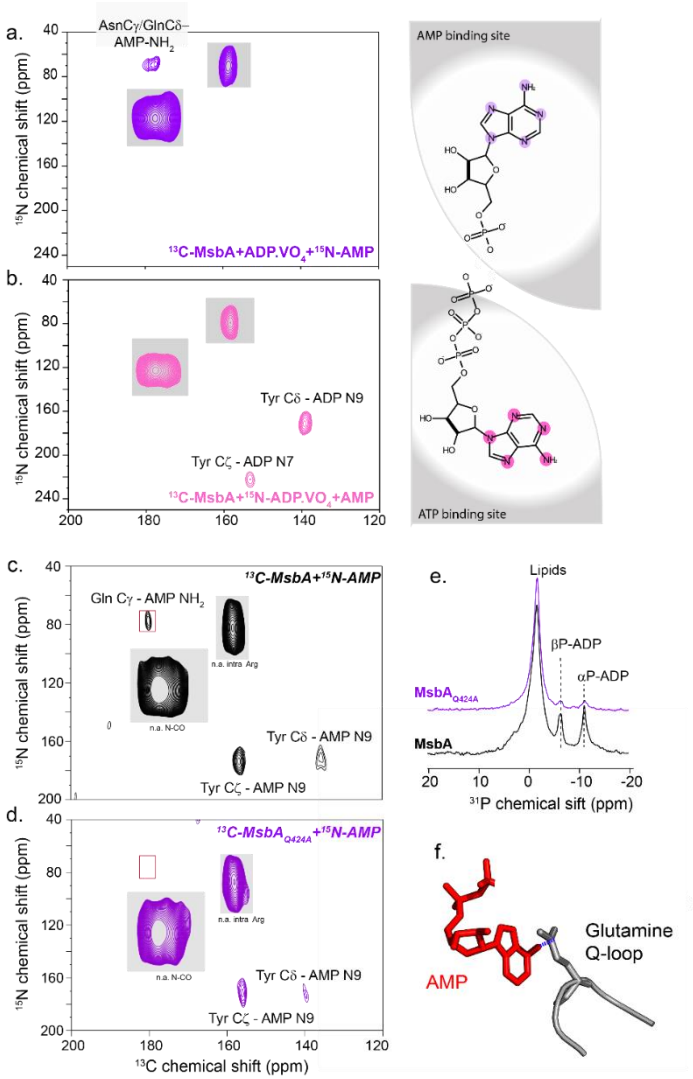


Figure 4.4- *DNP-enhanced TEDOR experiments on U-^{13}C-MsbA in complex with nucleotides occupying the canonical (ATP-site) and the AK-site (AMP-site).* ***(a)*** *U-^{13}C-MsbA with ADP.VO_4/^{15}N-AMP. The correlation between AMP-NH_2 and a carbonyl side chain can be tentatively assigned to Q-loop Q424 and/or N425.* ***(b)*** *U-^{13}C-MsbA with ^{15}N-ADP.VO_4/AMP (state 5 in Figure 1). Cross peaks between ADP and MsbA can be assigned to A-loop Tyr351.* ***(c)*** *U-^{13}C-MsbA with ^{15}N-AMP. In this state, AMP can occupy both the canonical as well as the AK-sites. AMP at the canonical most likely makes contact with Tyr351 so that the observed correlations can be tentatively assigned to AMP-N9 / Tyr351-Cζ and AMP-N9 / Tyr351-Cδ. In addition, a similar cross peak as in (b) between AMP-NH_2 and a carbonyl side chain is observed.* ***(d)*** *Effect of the Q424A mutation. U-^{13}C-$MsbA_{Q424A}$ with ^{15}N-AMP. The AMP-NH_2/carbonyl resonance disappears providing evidence that AMP in the AK-site is in close proximity to Q424.* ***(e)*** *^{31}P-CP MAS NMR spectra of MsbA-ADP.VO_4 and $MsbA_{Q424A}$-ADP.VO_4. The reduced α- and β-ADP peaks in $MsbA_{Q424A}$ demonstrate reduced Vi-trapping efficiency upon mutation.* ***(f)*** *Interaction between aromatic part of Ap5A and Q-loop side chain forming the AK-site as depicted in crystal structure of pfSMC with Ap5A (pdb id:3KTA).*

The additional cross peaks can be assigned tentatively to AMP-N9 and Tyr-Cζ most likely arising from contacts between AMP within the canonical binding site with the A-loop Tyr351. These data provide evidence that the nucleotide bound at the AK-site is located in direct contact with the Q-loop. It is important to point out that recording each of these TEDOR spectra required approximately 2.5 days of NMR time despite DNP enhancement, which means that these data could not have been obtained by conventional solid-state NMR.

4.4.3-Exploring the Properties of the Bound Metal Ion Using EPR Spectroscopy.

The binding of Mg^{2+} along with the nucleotide to MsbA is essential for the hydrolysis reactions to be catalyzed. The conserved Q-loop, Walker B, and His-switch have been proposed to play a role in coordination of the metal ion and the water molecule[1,368, 90,157,243,378]. Our DNP-enhanced MAS NMR data show that the Q-loop is directly involved in the additional AK-site. So the question arises whether nucleotide binding to the AK-site affects the coordination of Mg^{2+} binding. We have addressed this question by EPR spectroscopy on MsbA reconstituted into liposomes and trapped in state (**2**) (ADP·VO_4) and state (**5**) (ADP·VO_4+AMP). For this purpose, diamagnetic Mg^{2+} was replaced by paramagnetic Mn^{2+}, which has an electron spin of 5/2. The basal ATPase activity of MsbA is unaffected in the presence of Mn^{2+} as shown in figure 4.10a and as previously published. The amount of free Mn^{2+} in ADP·VO_4 trapped MsbA samples was minimized by optimizing the Mn:ATP ratio (figure 4.10b). The correct trapping of MsbA in both states has been verified by ^{31}P-CP MAS NMR figure 4.10c). First, we probed whether binding of AMP to the AK-site displaces Mn^{2+} (figure 4.5a). The 9.4 GHz, room temperature cw-EPR spectrum of Mn^{2+} in the presence of lipids consists of the $|+1/2> \rightarrow -|1/2>$ central electron transition, which splits into six lines due to the hyperfine coupling with the spin 5/2 nuclear spin. No signal for free Mn^{2+} can be detected due to binding to ATP (ATP:Mn^{2+} 4:1). This ratio was therefore used for preparing MsbA in the ADP·VO_4 trapped state (see figure 4.10 for optimization of sample conditions). The EPR spectrum of Mn^{2+} in complex with MsbA+ADP·VO_4 shows extreme line broadening. Upon addition of AMP, no signal of free Mn^{2+} could be detected, which suggests that it does not get displaced from its binding site. In addition, Hahn-echo detected fieldsweep spectra of ATP:Mn^{2+} (4:1) in lipids and ADP·VO4 in complex with MsbA with and without AMP were recorded at 263 GHz and 10 K (figure 4.5b). For ATP:Mn^{2+} , the characteristic hyperfine coupling sextet of the central transition is observed, which could not be detected under the conditions described before due to dynamic line broadening. The spectra of MsbA+ADP·VO_4 with and without AMP showed multiple features (see figure 4.11 for more details) and almost identical intensities (see table 4.3 for precision of signal intensities) and line shapes (figure 4.11b and figure 4.12), which supports the previous observation that Mn^{2+} is not

released. These findings are also supported by electron spin relaxation measurements of free and protein-bound Mn^{2+}. The electron relaxation in both MsbA+ADP·VO_4 and MsbA+ADP·VO_4+AMP corresponds to protein-bound rather than nucleotide-bound Mn^{2+} (figure 4.13). It can be therefore concluded that AMP-binding to the AK-site does not displace the bound metal-ion, which means that both binding sites do not overlap.

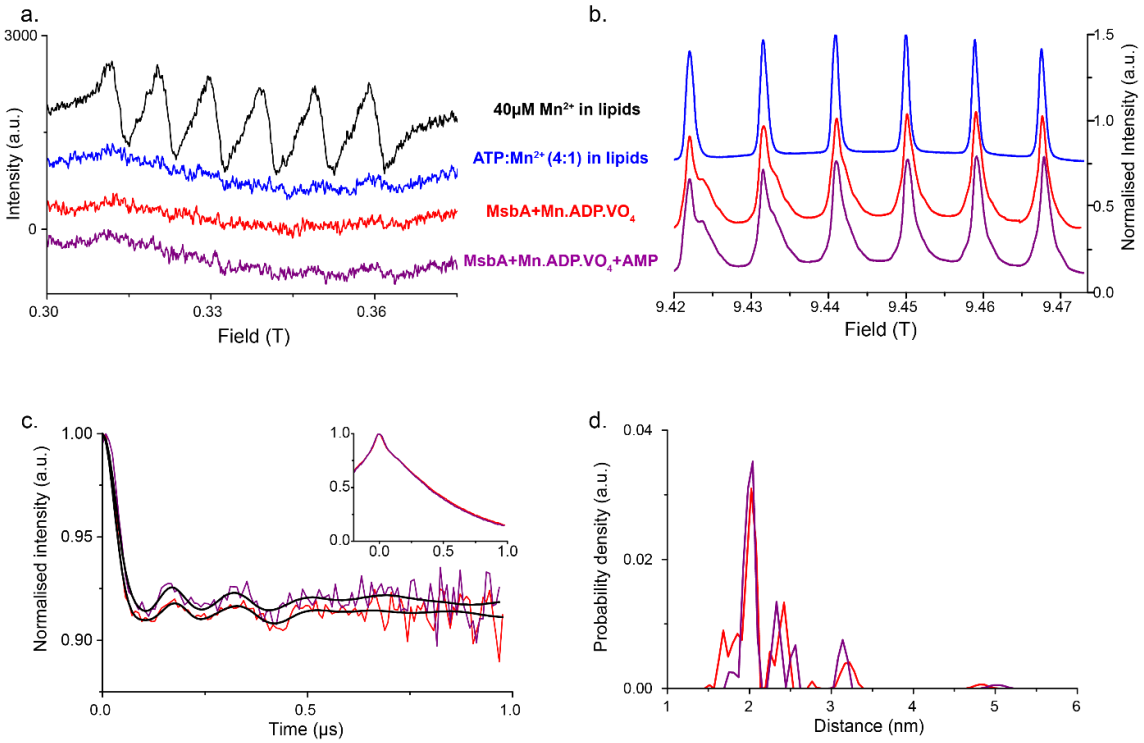


***Figure 4.5**- EPR spectroscopy on Mn^{2+} in complex with MsbA trapped in the ADP.VO_4 and ADP.VO_4+AMP states. **(a)** 9.4 GHz cw-EPR spectra at room temperature of Mn^{2+} in lipids (1:75, [Mn^{2+}] = 40μM), in complex with ATP in lipids (ATP:Mn 4:1) and in both nucleotide-bound MsbA states. The spectrum of non-bound Mn^{2+} consists of the central electron transition, which splits into a hyperfine coupling sextet. Complex formation with ATP (ATP:Mn of 4:1) causes severe line broadening indicating that all Mn^{2+} occurs in complex with ATP. In complex with MsbA in the ADP.VO_4 state, no free Mn^{2+} could be detected. Addition of AMP does not displace bound Mn^{2+} (see Figure B5 for sample optimization steps). **(b)** 263 GHz Hahn echo-detected field-swept EPR spectra at 10 K of ATP:Mn (4:1) and of MsbA in both trapped states. Here, ATP:Mn shows the characteristic hyperfine coupling sextet. The comparable signal intensity (see Table B4 for more details) and spectral lineshape of Mn^{2+} bound to MsbA+ADP.VO_4 with and without AMP shows that the metal cofactor has not been displaced by AMP (see also Figs. B7-B8). **(c)** Background-divided 263 GHz RIDME time traces obtained at a sample temperature of 10 K and back-calculated time traces (black) based on Tikhonov regularization. Inset shows the overlay of raw experimental time traces. **(d)** Distance distributions obtained by Tikhonov regularization with a regularization parameter of 0.1. Both the samples show comparable modulation depth and distance distribution obtained from RIDME time traces (for more details see figs. 4.9-4.11).*

Furthermore, 263 GHz relaxation-induced dipolar modulation enhancement (RIDME)[372,379] experiments were performed to probe the distance between Mn^{2+} centers in both MsbA trapped

states. The RIDME time traces for both the samples, obtained with the mixing time of 30 µs (corresponds to T_{1e} for the respective temperature), yielded modulation depths of about 8% (figure 4.5c and figures 4.14 and 4.15) and a most probable distance of about 2 nm, which is in good agreement with published data (figure 4.5d and figure 4.16)[362]. The low modulation depth can be explained by the ADP·VO4 trapping efficiency of 80−90%, which means that not every binding site in MsbA contains Mn^{2+} and ADP·VO_4. Furthermore, MsbA has been purified in a Mg-containing buffer, which could result in residual amounts of bound Mg^{2+} despite of washing and dilution steps. Only MsbA populations with both binding sites occupied by Mn^{2+} (Mn^{2+}/Mn^{2+}) will give rise to a RIDME signal, while all other populations (two sites occupied with Mn^{2+}/Mg^{2+}, Mg^{2+} /Mg^{2+} ; just one side occupied with Mg^{2+} or Mn^{2+}) will not contribute. Moreover, spectral overlap of multiple species reduces the modulation depth further. In a previous publication by some of the coauthors[380], a modulation depth of 37% was reported for 100% Mn−Mn pairs. The here observed modulation depth of just 8% could be rationalized by the reasons given above (the mixing times in both cases were chosen in accordance with T_{1e} relaxation times).

The reliability of the data analysis was verified by exploring variations of the Tikhonov regularization parameter and by assessing the quality of the background fit (figure 4.16−4.20). A detailed explanation of background validation and correction has been described before[380]. It is worth noting that the distance distributions do not reveal a considerable contribution from harmonics of the dipolar coupling frequency as observed in other RIDME measurements with high-spin systems[380–383]. Thus, the second and third harmonics of the dipolar coupling frequency are cut by the limited bandwidth of the π pulse. A cross peak pattern corresponding to the first (with 2 nm distance), second, and third harmonics would have a width of 13, 26, and 39 MHz, respectively. The π-pulses used here were 36 ns (MsbA+ADP·VO_4) and 42 ns (MsbA+ADP·VO_4+AMP), which are related to corresponding bandwidths of 22 and 19 MHz. Such an effect has been previously reported for RIDME measurements using high-spin Mn^{2+} centers[380].

4.4.4-Molecular Simulations on MsbA ^{nbds} to Explore Nucleotide Binding Modes.

Our experimental data described above clearly demonstrate so far unexplored nucleotide binding modes, which are required for the coupled ATPase-rAK mechanism. However, the molecular picture is still incomplete in terms of a full description of the nucleotide-MsbA interactions and in terms of nucleotide stoichiometry. As shown and discussed above, an additional ADP has to bind at the AK-site in proximity and with the correct orientation with respect to an ADP in the canonical site for the phosphoryl transfer reaction to take place. In principle, one extra ADP at a time could bind oriented towards the existing ADP in the canonical

site. However, it cannot be excluded that also two ADPs can bind to the AK-site reacting with the canonically bound ones. Previous ^{31}P-CP MAS NMR experiments on state (**3**) (ADP·VO_4+ADP-βS)[59] support the latter case, while the data in figure 4.2 hint toward the first situation. The existence of just one binding site was also suggested for the ATP-gated chloride channel CFTR, a member of the ABC superfamily that has been shown to be able to catalyse both ATPase and rAK reactions[342]. However, experimental uncertainties are too large in both cases preventing a confident statement about nucleotide stoichiometry in the AK-site. We therefore used MD simulations in order to explore under which conditions a stable nucleotide- MsbA configuration could be obtained. The MD simulations just focused on state (**5**) (ATP+AMP bound to MsbA) in order to link them to the experimental data described above.

A molecular model for the interactions between ATP, AMP, and MsbA was obtained by performing MD simulations on MsbA nucleotide binding domains ($MsbA^{nbds}$) containing the nucleotide ligands. Only the NBDs were simulated, assuming that transmembrane domains were not relevant for these interactions. First, $MsbA^{nbds}$ with two ATP molecules bound to their canonical sites were simulated over 200 ns. Simulations were carried out in 5-fold replicates in order to generate stable conformations for testing the insertion of one or two AMP molecules, using a slow-growth procedure, as described in the Material and Methods section. Structural stability during the time course of the simulation is illustrated by the RMSD evolution (figure 4.21).

For the binding of one AMP to the AK-site, two sets of simulations were carried out in which AMP was inserted in two opposite directions but facing one of the canonical ATP binding sites. Insertion was considered successful if, at the end of the process, AMP binding was in agreement with the NMR data with the Q-loop near the adenine group and the His switch somewhere close to AMP. In the cases where AMP insertion was not successful, either diffusion of AMP to the solvent mostly through the $MsbA^{nbds}$ interface took place or nonspecific binding was observed. Such nonspecific binding modes depicted an interaction between Q-loop and adenine, but the phosphate group was not oriented toward the canonical site. In addition, a reversed orientation with the adenine moiety toward the magnesium ion and the phosphate oriented toward the AK-site was also observed.

The success rate (in the slow-growth procedure) of single insertion of AMP was 13.7% in one active site and 15.8% on the other. These similar numbers reflect the symmetry of the homodimer and are not exactly the same due to limited sampling. Every time AMP insertion was successfully achieved, the resulting conformations were simulated for further 100 ns. The simulations showed that the insertion of AMP did not cause any significant conformational

change on the $MsbA^{nbds}$, indicating a stable structure in which the RMSD is constant during the course of the simulation (figures 4.22-4.24).

A closer analysis of the obtained models of single AMP insertion after the 100 ns simulation revealed that the adenine and ribose groups of AMP interact with the Q-loop establishing hydrogen bonds with the backbone carbonyl of Q424 and with the side chains of Q424 and N425. Additionally, AMP is pointed toward the ATPase site, being close to H537 as depicted in figure 4.6a. These observations are comparable with the data obtained from MAS NMR data (i.e., Q424 is essential for the binding of AMP in this pocket and with the chemical shift perturbation data which indicates that H537 is affected by AMP binding). Interestingly, in many cases AMP interacts with both Q-loop motifs. In addition, in some conformations where AMP was not fully oriented toward the active site, the adenine group of AMP started by interacting with the Q-loop motif and the entire molecule was further oriented toward the canonical ATPase site, indicating that the Q-loop might be essential to orient AMP at the AK-site toward the canonical site. Due to the proximity of the phosphate group to H537, it is tempting to speculate that H537 may help AMP coordinating with the magnesium ion. The AMP binding mode is also similar to the one observed in ABC enzyme *pf*SMC, in which Ap5A, an inhibitor resembling both AMP and ATP, also interacts with the Q-loop glutamine[61].

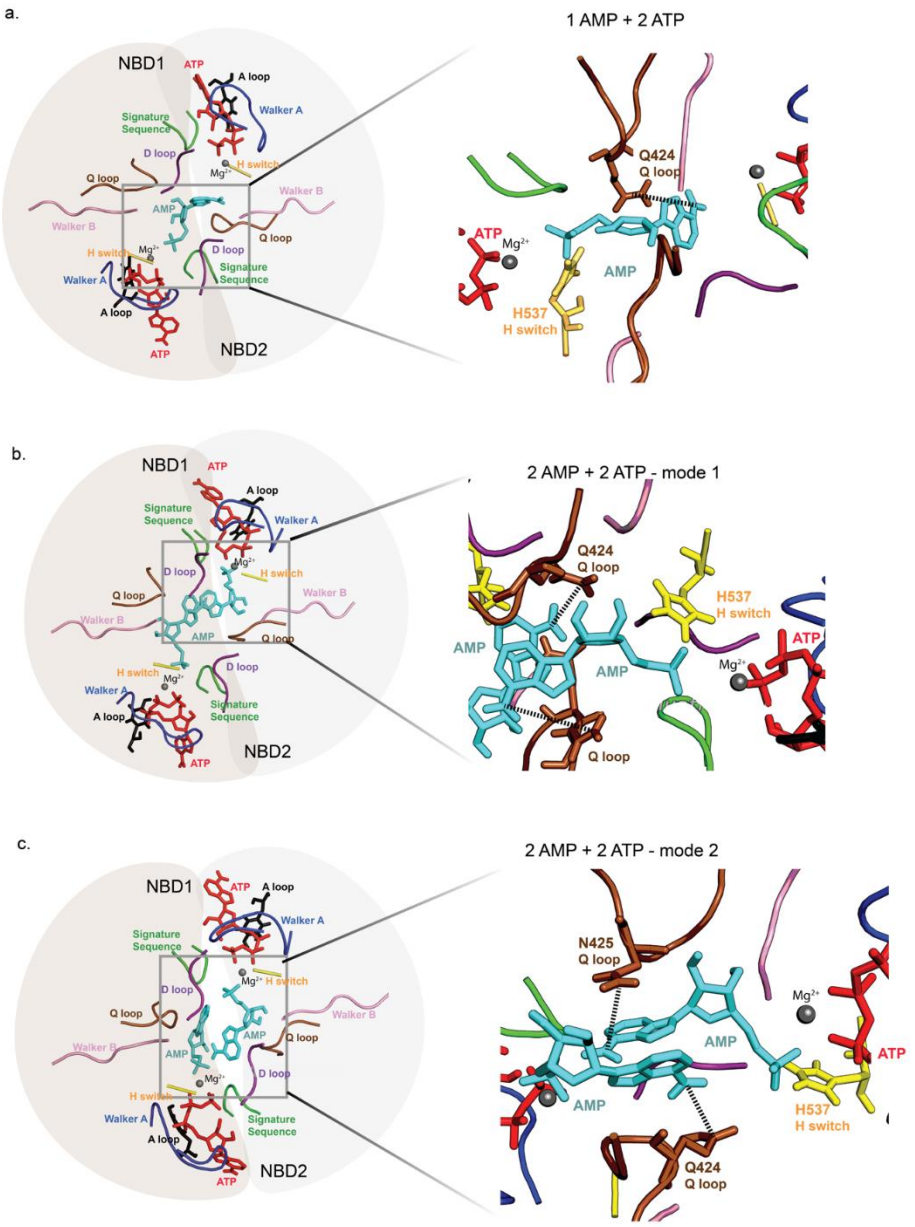


Figure 4.6- *Snapshots extracted from selected molecular dynamics simulations of the MsbA^{ndbs} dimer with two bound ATP molecules (depicted in red) at the canonical ATPase site and AMP molecule(s) (depicted in cyan) bound at the AK site formed by the Q-loop. The conserved motifs have been highlighted. The structures presented correspond to the conformations at the end of the 100 ns simulation after the insertion of AMP. Water molecules and the sidechains of the remaining residues were omitted for clarity.* ***(a)*** *In the case where two ATP molecules and one AMP is bound to the MsbA^{ndbs}, Q-loop (Q424A) side chain interaction with AMP-NH$_2$ group can be observed, as well as the proximity of the His-switch with the phosphates of the bound nucleotide.* ***(b)*** *In the case where two ATP molecule and two AMP molecules are bound to MsbA^{ndbs}, the conformation evidences a congested structure in which Q424 and N425 of the Q-loop form interactions with the bound AMP molecules. These interactions appear to keep the nucleotides in proper catalytic orientation since both AMP molecules are oriented towards the active site.* ***(c)*** *Another simulation where two AMP molecules and two ATP molecules are bound to MsbA^{ndbs} depicting similar interactions as observed by NMR.*

On the other hand, when two AMP molecules are simultaneously inserted in the two AK-sites using slow-growth, the success rate of insertion was 14.7%, which is similar to the one found in

the single insertions. However, the stability of the double occupied AK-sites is lower than in the single occupied cases described above, given that only a portion (65%) of the conformations displayed both AMP molecules after 100 ns simulations. In the other 35% of the cases, one of the AMP molecules diffused to the exterior of the protein, in stark contrast with the single AMP insertions, where all of the AMP molecules remained in the AK-site. This reinforces the idea that there might be only one binding site for one AMP molecule at a time. In most cases where the two molecules stay inside the protein, such as the one portrayed in figure 4.6b, one of the AMP molecules interacts with the Q-loop and H537 in the same way as previously described, while the other AMP molecule adopts a random conformation, even though both adenine groups can fit between both Q-loops. In many conformations, although the adenine groups of two AMP molecules were able to fit between the Q-loops, they did not establish as many hydrogen bonds to the Q-loop as those observed in the single insertions of AMP. Another binding mode observed is represented in figure 4.6c. One AMP is oriented toward the canonical site, establishing hydrogen bonds with the Q-loop and is found in proximity to the His switch. The second AMP is interacting with the NH2 group of the first AMP, while the base is still in contact with the Q-loop.

4.5-Discussion

4.5.1-Q-loop, a Conserved NBD Motif, Is Part of the AK-Site.

A unique-pair labeling scheme was used to probe the effect of trapping catalytic intermediates, as defined in figure 4.1, on the Q-loop and the His-switch. Occurrence of a new peak in states **(3)** and **(5)** (ADP·VO_4+ADPβS, ADP·VO_4+AMP) suggests that Q-loop as well as His-switch are integral parts of the AK-site. A direct evidence for the role of Q-loop is provided by DNP-enhanced MAS NMR. Dipolar ^{15}N–^{13}C through-space spectroscopy was used to visualize MsbA-nucleotide interactions. Specific cross peaks (figure 4.3, panels b and c) observed for state **(1)** (U-^{13}C-MsbA+^{15}N-ADP·BeF_3) and state **(2)** (U-^{13}C-MsbA+^{15}N-ADP·VO_4) are due to π–π stacking interactions between Tyr351 in the A-loop and the aromatic part of the nucleotide in the canonical ATP binding site as observed in the crystal structure. These data demonstrate the feasibility of probing direct nucleotide-protein contacts with the help of DNP-enhanced MAS NMR. In order to probe the AK-site, U-^{13}C MsbA trapped with ADP.VO_4+^{15}N-AMP (state **(5)**) was analyzed. A specific cross peak between NH_2 of ^{15}N-AMP and a ^{13}C-carbonyl from MsbA was detected, which could be unambiguously assigned as a contact with Cγ of Q424 based on site directed mutagenesis (figure 4.4). These experiments demonstrate that trapping with additional nucleotides indeed leads to occupation of an additional site other than the canonical ATPase

sites, which is in line with earlier experimental indications based on ^{31}P-real time MAS NMR and mutation studies[59].

It is interesting to note that a recent application of Mn^{2+} - hyperfine spectroscopy on cysteine-less ABC exporters in detergent revealed that ATP turnover for MsbA was not only much faster but also symmetric in contrast to BmrCD[368]. The application of EDNMR to MsbA in complex with ^{15}N-ATP revealed not only that ATP and ADP have different binding modes but also that bound Mn^{2+}:ADP is coordinated to a nitrogen atom. In this work, the authors state that the N atom in proximity to the Mn^{2+}:ADP complex could tentatively belong to the Q-loop of MsbA. These results provide further support toward a contact between Q-loop and nucleotide.

4.5.2-Metal Cofactor and Nucleotide Binding Sites Do Not overlap.

The Q-loop has been widely acknowledged to be involved in co-ordinating the metal cofactor (i.e., Mg^{2+})[41,368,372,379,384,385]. Therefore, diamagnetic Mg^{2+} was replaced with paramagnetic Mn^{2+} and states **(2)** and **(5)** were compared using pulsed Mn^{2+} EPR spectroscopy. The analysis of line shapes, spectral intensities, and relaxation parameters of Mn^{2+} in both the samples revealed that indeed binding of AMP to the AK-site does not displace the metal (figure 4.5, panels a and b, and figure 4.14). Therefore, the AK-site does not overlap with the metal binding site. Furthermore, dipolar EPR spectroscopy was carried out to probe the distance between both paramagnetic centers using RIDME. The acquired time traces for both the trapped states appeared similar, indicating comparable modulation depths and distance distributions in both states **(2)** and **(5)** (figure 4.5, panels c and d), proving that ion binding is independent of nucleotide binding at the AK-site. These distance distributions are analogous to what would be expected from the crystal structure and corroborate to the recently published data on another ABC exporter[362]. They also demonstrate that the RIDME experiment is highly suitable for measuring Mn^{2+}-Mn^{2+} dipolar couplings within an ABC transporter embedded in lipid bilayers.

4.5.3-Possible Modes of AMP Binding at the AK-Site.

Molecular simulations of $MsbA^{nbds}$ showed that either one or two AMP molecules can bind to $MsbA^{nbds}$ at the AK-site along with two ATP molecules bound at the canonical sites. In all these binding modes (i.e., either with one or two AMP molecules), interaction between adenine and the carbonyl side chain of Q-loop (Q424 or N425) could be detected. These simulations agree with the NMR data, showing that AMP interacts with the Q-loop at the AK-site. Additionally, the His-switch was found to be in close proximity to phosphate of the bound nucleotide supporting

the interpretation of the NMR data. The Q-loop seems to be important for the orientation of AMP toward the catalytic site.

Finally, based on the number of hydrogen bonds formed by the Q-loop with the nucleotide bound at the AK-site, it is highly likely that one rather than two AMP molecules bind simultaneously in a conformation ready for catalysis. These results also provide a valid explanation for the signal observed in the 1D-NCOCX measurements (figure 4.2). Only one binding site rather than two has also been hypothesized for CFTR[342,386]. However, it is difficult to exclude the possibility that two nucleotides at the AK-site and two at the canonical sites could bind in stable conformation suitable for the phosphoryl transfer reaction to take place, especially when the NBDs are more separated. Such a hypothesis would explain the low resolution (6.9 Å) of the ADP-bound cryo-EM structure as compared to apo-MsbA (4.2 Å)[129] because of the ADP-bound state could be more dynamic due to an ongoing rAK reaction.

4.5.4-Physiological Importance and Broad Relevance.

The rAK activity has been previously reported for atypical ABC proteins such as CFTR, *pf*SMC, and Rad50[5,60,61,342,63,386,387]. Our recent work showed that typical ABC exporters such as MsbA, LmrA, and TmrAB can catalyze a phosphoryl transfer rAK reaction, which is connected with ATP hydrolysis. It was proposed that ATP hydrolysis is the primary reaction. However, environmental factors such as the local nucleotide concentrations influence the rates of both reactions so that the protein could adjust its catalytic activity toward a coupled ATPase-rAK mechanism. In studies determining cellular concentration of nucleotides, a large diversity in the level of each nucleotide in each cell has been measured. It has been suggested that individual cell adopts different strategies under severe conditions, and this may benefit the whole population[388,389]. Following the same argument, it becomes important to consider the mechanisms that a protein might adopt in order to assist the cell to survive under stress conditions. The coupled cycle not only provides the possibility to explain the unnecessary ATP consumption during “basal activity” observed for many ABC transporters but also emphasizes the important roles that these proteins play in the cell. Absence or inactivity of MsbA causes lipid A accumulation in the inner membrane leading to lipid A induced toxicity and malformed outer membrane and thus resulting in cell death[390–392]. Under conditions that lead to cell stress, MsbA increases the chances of a cell survival since it is able to carry out LPS translocation by harnessing the energy from the coupled ATPase-rAK cycle. Additionally, there is increasing evidence that ADP can cause partial or complete closing of the NBDs[103,104]. Such a step could either lead to a conformational switch from an inward facing to an outward facing conformation like in Sav1866[377] or to an occluded state[104,393] in prokaryotic ABC exporters. Our data provides a probable explanation to these

phenomena. Of course, these conformational changes do depend on the individual protein, on nucleotide concentration, cofactors, and their accessibility to the protein. Additionally, we have investigated various binding modes of nucleotides at canonical and AK-sites by MD simulations. On the basis of these experimental and MD data, it can be speculated that the AK-site is probably a low affinity site, which might be formed transiently either as a result of locally increased ADP concentration or other unknown factors triggered *in vivo*.

4.6-Summary and conclusions

We have presented a concerted approach to investigate the nucleotide-protein complex of full length MsbA within lipid bilayers trapped in different catalytic states by combining high field and DNP-enhanced MAS NMR, Mn-EPR spectroscopy, and MD simulations. Our observations show that the Q-loop plays an essential role in nucleotide coordination at the AK-site. Furthermore, our data indicate an involvement of the His-switch located close to the phosphate groups of both nucleotides. DNP mediated signal enhancement enabled, for the first time, the spectroscopic identification of nucleotide-protein contacts such as those observed between the A-loop and ADP·VO_4 or between AMP and the Q-loop. The former data are in agreement with previous crystal structures[96,129], while the latter confirm our previous hypothesis[59]. Our work highlights the capability of DNP-mediated NMR to explore stable conformations as well as intermediate states that might be difficult to capture by cryo-EM or crystallization techniques. One further important aspect of nucleotide-protein interactions is the coordination of the required metal ion, which has been addressed here by Mn-EPR spectroscopy. Mn-EPR, especially at high field, is a rather attractive option for metal-binding proteins[394] such as ABC transporters since mechanistic insight could be obtained without mutations or spin labelling[362,368]. Here, the first application of RIDME to an ABC transporter demonstrates that dipolar couplings can be obtained with high sensitivity directly within liposomes, which will have implications for future studies on conformational changes of membrane proteins by EPR. The data reveal no significant effect on the protein-bound Mn^{2+} in different trapped states since the metal ion is important for both ATPase and reverse adenylate kinase reaction. The experimental NMR and EPR results were further substantiated by molecular dynamics simulations, which validated the role of Q-loop and His-switch in AK-site. Our study emphasizes the importance of advanced spectroscopic approaches such as DNP-enhanced solid-state NMR and pulsed EPR for obtaining insight into the catalytic mechanism of membrane proteins directly within the lipid bilayer. The description of novel nucleotide binding modes extends the paradigm of how ABC transporters seem to work, and it has an even broader impact if it applies to the whole ABC superfamily.

4.7-Acknowledgments

The work was funded by DFG/SFB 807 'Transport and communication across membranes' and the Cluster of Excellence Macromolecular Complexes Frankfurt (DFG EXC 115). The DNP experiments were enabled through an equipment grant to C.G. provided by DFG (GL 307/4-1). We thank Dr. Johanna Becker-Baldus for great technical support and for comments on the manuscript. Dr. Vasyl Denysenkov is gratefully acknowledged for technical support with the 263 GHz EPR spectrometer. This work was also supported by Project LISBOA-01-0145-FEDER-007660 (Microbiologia Molecular, Estrutural e Celular), by FEDER through COMPETE2020 - Programa Operacional Competitividade e Internacionalização (POCI), and by national funds through FCT - Fundação para a Ciência e a Tecnologia. B.A. acknowledges funding via a Ph.D. scholarship (SFRH/BD/108002/2015) from FCT.

4.8-Supplementary information

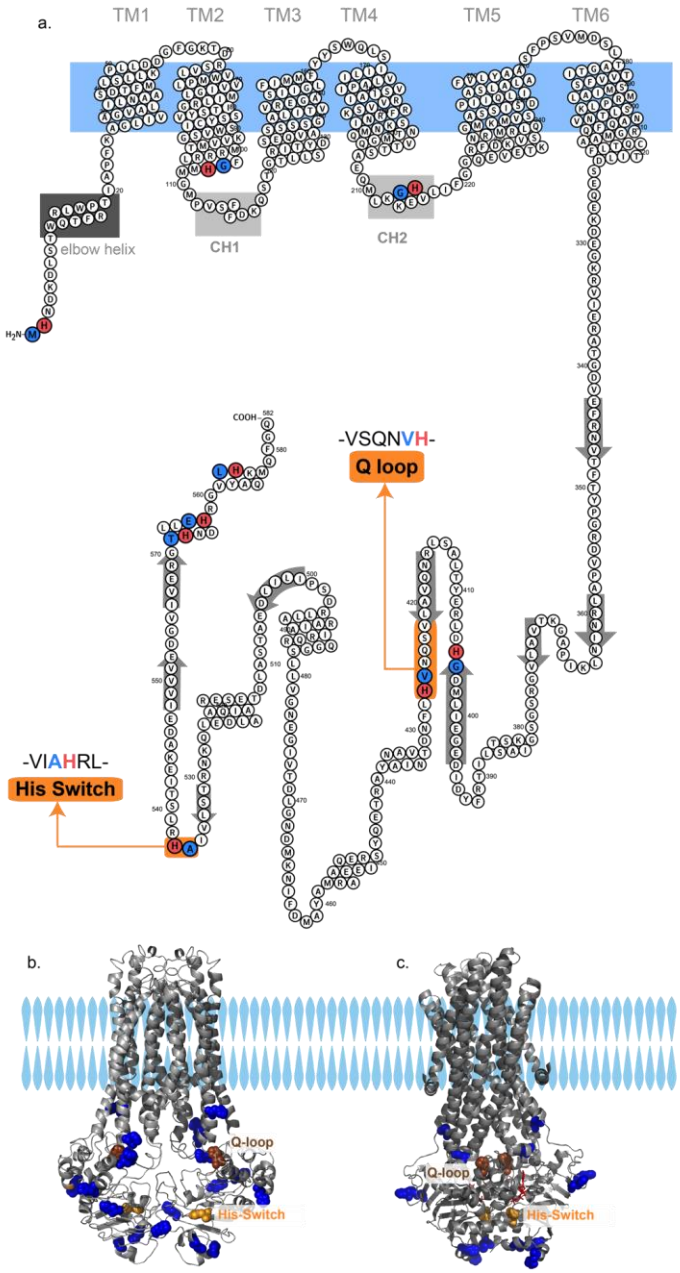


Figure 4.7- *Histidine-based pair labeling in MsbA as probe for structural and dynamic changes within the NBDs during the catalytic cycle. Magnetization transfer from ^{15}N-labelled histidines at position (i) to ^{13}C-labelled (i-1) residues reduces spectral complexity and simplifies resonance assignment. There are 21 histidines in the MsbA construct used here, which includes 10 in the His-tag and two in the linker sequence. The remaining 9 histidines can be used as probes in a pair labeling approach.* ***(a)*** *Location of residue pairs in the MsbA topology plot. Histidines are highlighted in blue and preceding residues marked in pink. Residue pairs are listed in table 4.1. Unique pairs occur in the Q-loop (V426-H427) and in the His-switch (A536-H537).* ***(b)*** *Location of histidines in the MsbA closed apo state (pdb 3B5W) and* ***(c)*** *in the closed AMP.PNP bound form (pdb 3B60). Histidines are distributed at the intracellular side of MsbA (blue spheres). Three pair-labeled MsbA samples were prepared: The V426-H427 pair in the Q-loop is created in [U-^{13}C-Val,^{15}N1-His]-MsbA, the A536-H537 pair in the His-switch is observed in [U-^{13}C-Ala,^{15}N1-His]-MsbA, while all nine pairs are simultaneously labeled in [U-^{13}C,^{15}N1-His]-MsbA. The catalytic intermediate states defined in figure 4.1 were prepared.*

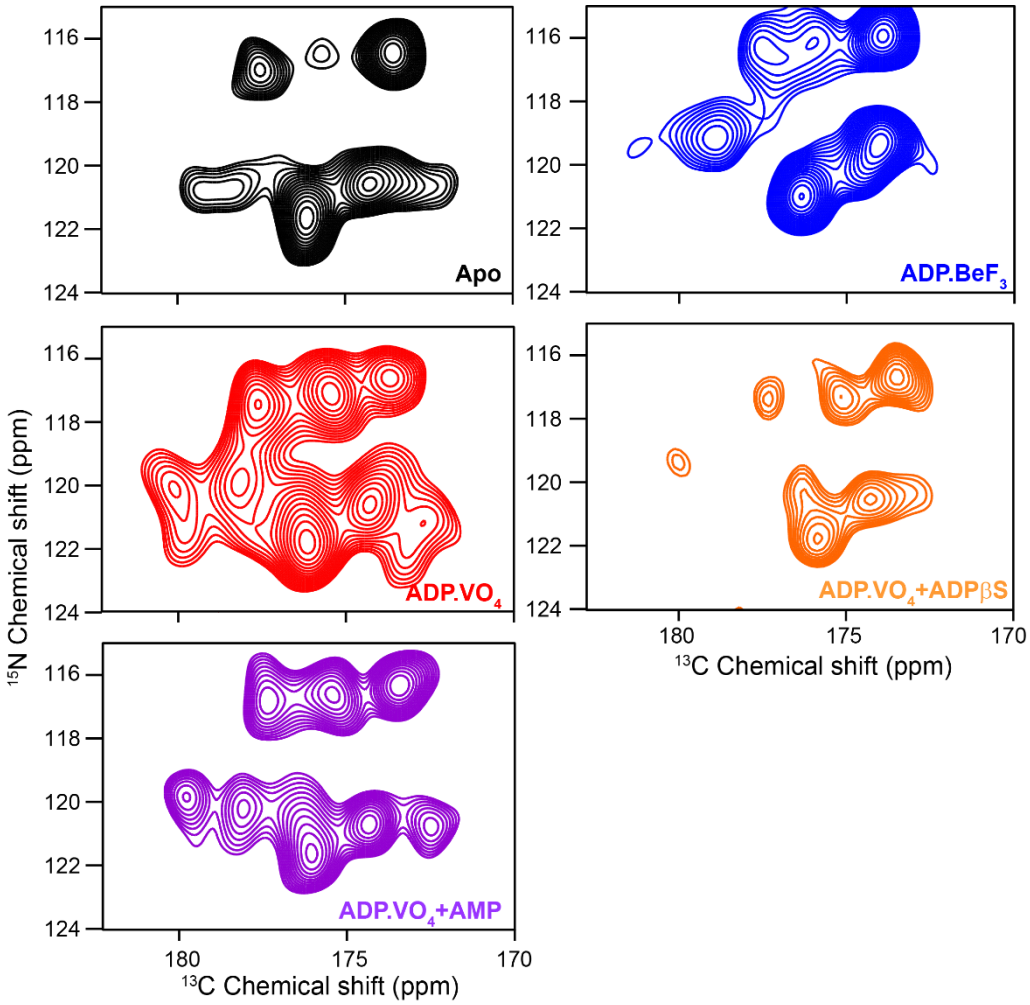


Figure 4.8- *2D-NCO spectra of [U-^{13}C, $^{15}N_1$-His]-MsbA in differently trapped states: MsbA apo state, prehydrolysis state (ADP.BeF_3), transition state (ADP.VO_4), state (**3**) (ADP.VO_4+ADPβS) and state (**5**) of the ATPase-AK cycle with ADP.VO_4+AMP. All spectra were recorded under identical conditions with the same amount of protein and processed in the same way. The labeling scheme used here should result in 9 cross peaks (table 4.1). Indeed 9 peaks could be identified for the ADP.VO_4. In all other states, the number of peaks is reduced. The different number of peaks and their altered intensities indicate changes in the backbone dynamics in these catalytic intermediate states. The ADP.VO_4-trapped states seem to be the most rigid one. Furthermore, none of the visible peaks overlap, which demonstrates changes in the backbone structure of MsbA. All measurements were carried out at spinning rate of 10 kHz at 250 K. An acquisition time of 9.2 ms in the direct and 10 ms in the indirect dimension was used with 100 increments and 3072 scans per increment. During data processing, the direct ^{13}C dimension was zero-filled to 4k complex data points followed by applying a Gaussian window function with 30 Hz linewidth. In the indirect ^{15}N dimension, 2k zero filling followed by a squared-cosine window function is used prior to Fourier transformation (for further details see M+M). The temperature for the NMR experiments was chosen, so that spectra resolution was preserved while stabilizing the trapping conditions for the time course of the experiment (250 K). A fresh sample was used for each state.*

Table 4.1: *Residue pairs in MsbA constisting of Residue(i-1)+His(i)*

	Pair sequence (residue number)	Pair	Secondary structure of the pair
1	**MH** (1-2)	M1-H2	Loop before elbow helix
2	F**GH**MM (106-107)	G106-H105	TM2 terminal - Helix
3	LK**GH**KE (213-214)	G213-H214	IH2 - Helix
4	D**GH**DLR (404-405)	G404-H405	βsheet-loop region after P-loop
5	SQN**VH**L (426-427)	V426-H427	**Q-loop**
6	VI**AH**RLS (536-537)	A536-H537	**His-switch loop**
7	RG**TH**ND (561-562)	T561-H562	Helix-loop region
8	LL**EH**RGV (567-568)	E567-H568	Helix-loop region
9	AQ**LH**KM (575-576)	L575-H576	Terminal helix in nbds of MsbA

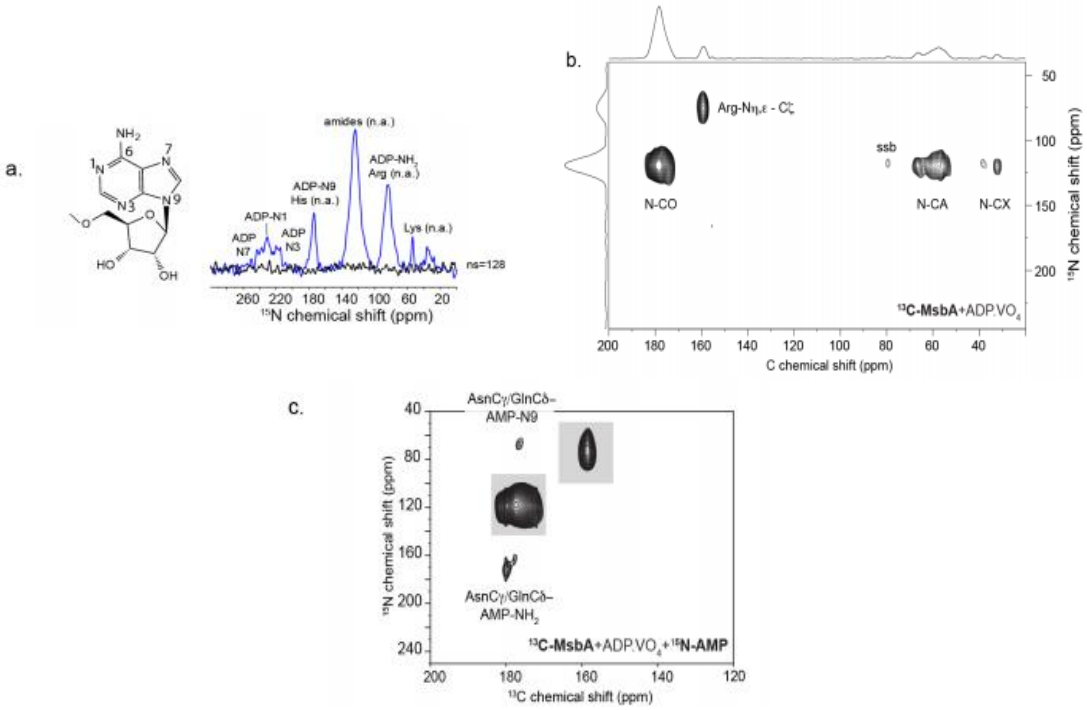


Figure 4.9- *DNP-enhanced TEDOR spectra on ^{13}C-MsbA in the pre-hydrolysis and in the catalytic transition state.* ***(a)*** *U-^{13}C-MsbA in complex with ^{15}N-ADP.VO$_4$ (state (**2**) in figure 4.1). Natural abundance contacts as well as contacts between N9 and NH$_3$ of ADP and carbons in Tyr351 are observed.* ***(b)*** *U-^{13}C-MsbA in complex with ^{15}N-ADP.BeF$_3$ (state (**1**) in figure 4.1). In addition to the natural abundance correlations, contacts between the N7 and N9 nitrogens in ADP and carbons in Tyr351 are observed (see figure 4.3 and main text). In addition, a contact tentatively assigned between ADP-N1/3 and Leu480/481 is observed.* ***(c)*** *U-^{13}C-MsbA in complex with ADP.BeF$_3$ (state (**1**) in figure 4.1). The observed cross peaks originate from intramolecular contacts between ^{13}C (100%) and natural abundance ^{15}N (0.4%). In particular, short-range contacts such as N-CO, N-CA, N-CB/CG etc. as well as intra-arginine correlations are observed.*

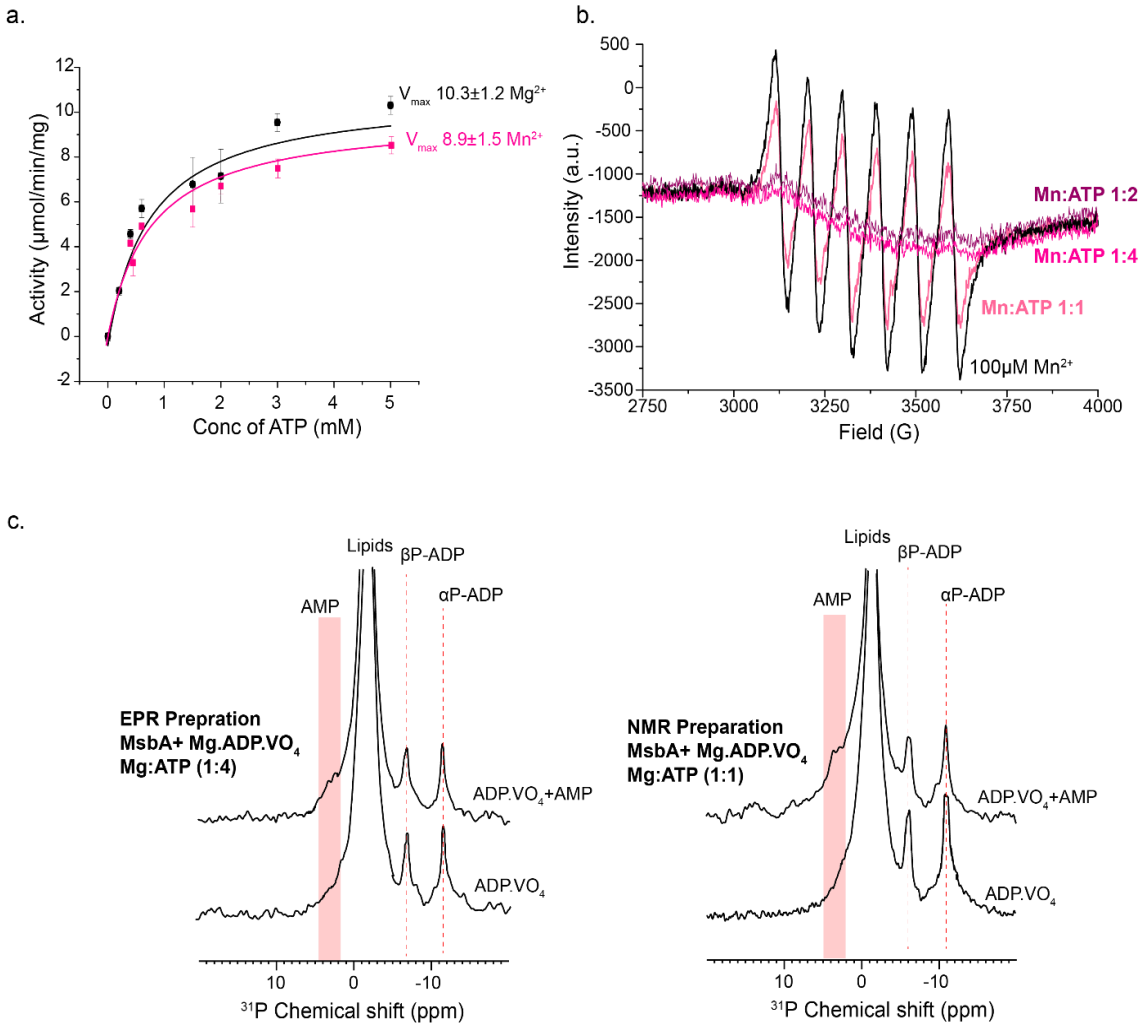


Figure 4.10- *Control experiments and sample optimization for Mn^{2+}-EPR on MsbA.* ***(a)*** *Basal ATPase activity of MsbA remains unaffected upon replacing Mg^{2+} by Mn^{2+}. The activity was determined using a colorimetric assays as described previously[59,369].* ***(b)*** *Optimizing the Mn:ATP ratio used for trapping MsbA in the ADP.VO_4 state. Sample preparation conditions were sought under which the occurrence of free Mn^{2+} in solution is minimized in order to simplify spectra analysis. Therefore, different Mn:ATP ratios were tested. Room temperature cw-EPR spectra at 9.4 GHz at different ratios show dramatic changes in the spectrum upon complex formation. The central EPR line of free Mn^{2+} in solution consists of a hyperfine-coupling sextet. The spectrum looks identical to Mn2+ in the presence of lipids (figure 4.5a). The EPR lines broaden significantly upon complex formation with ATP. With 2- and 4-fold excess ATP, no free Mn^{2+} could be detected anymore. Therefore, a Mn:ATP ratio of 1:4 was used for preparing MsbA in the ADP.VO_4 state. The corresponding spectrum in figure 4.5a shows that this sample does not contain free Mn^{2+}. Here, 20 mM HEPES buffer with 20% of glycerol was used. The concentration of Mn^{2+} was about 100 µM and kept constant for each sample.* ***(c)*** *^{31}P-CP-MAS NMR spectra depicting equivalent amounts of bound nucleotide when using Mg:ATP 1:4 (EPR sample preparation) and Mg:ATP 1:1 (original NMR sample preparation). Trapping protocols have been described in materials and methods. For this control, Mg^{2+} could not be replaced by Mn^{2+}, since paramagnetic quenching would have avoided the detection of bound nucleotides.*

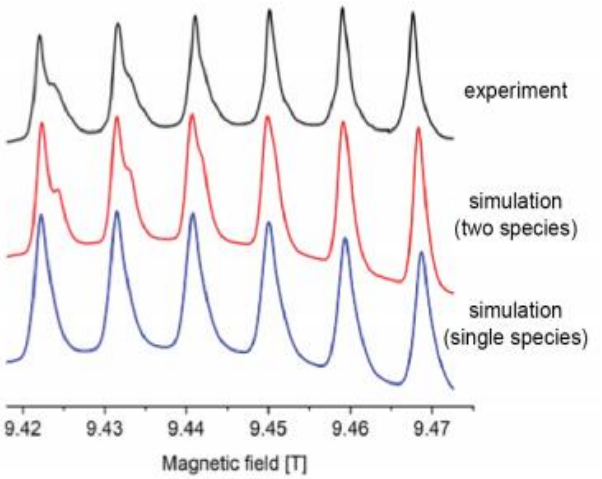


Figure 4.11- *Comparison between experimental and simulated Mn^{2+} EPR Spectra at 263 GHz. The experimental spectrum corresponds to figure 4.5b (MsbA+Mn.ADP.VO_4, 10 K). The simulations suggest that the shoulders observed in the experimental spectra are due to a second species and not caused by zero-field splitting. A possible explanation could be that the sample contains MsbA with one and two bound Mn^{2+}. Full matrix diagonalization using EasySpin was used. The parameters are summarized in table 4.2. Second order effects at 9.4 T and with a ZFS parameter D between 1.5 GHz and 1.8 GHz only contribute to line broadening, which is different for each hyperfine line, but do not lead to the observed shoulder. The simulations were based on a range of ZFS parameters (as indicated in the table below, because high-spin metal ions usually show a large statistical distribution in frozen solution.*

Table 4.2- *Parameters used for simulating the Mn^{2+} EPR spectra in figure 4.11*

	Species 1	Species 2	Single species
g-value*	1.99892	1.99866	1.99888
Hyperfine coupling* [MHz]	257	246.8	260.3
D [MHz]	-1500	-1500	-1800
E [MHz]	300	300	300
DStrain (D, E) [MHz]	(1125, 225)	(1200, 240)	(1440, 240)
HStrain [MHz]	20	20	20
Ratio	3 : 1 (species 1 : species 2)		-

Note, that the magnetic field was not calibrated. Field non-linearity is present. The microwave frequency is 264.252513 GHz.

The parameters presented in the table have qualitative character. They were used to demonstrate the difference between line shapes corresponding to one- and two-species cases.

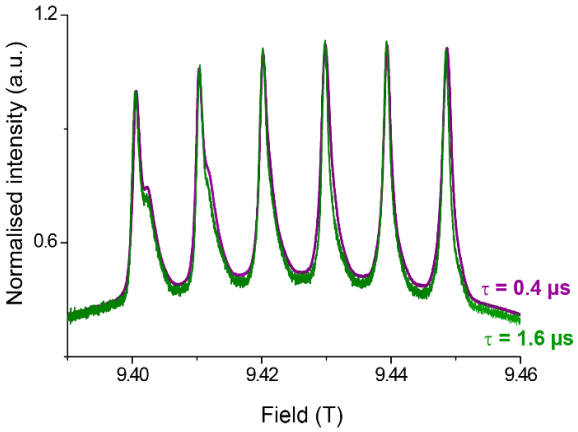


Figure 4.12- *263 GHz Hahn echo-detected field-swept EPR spectra of MsbA with Mn.ADP.VO_4 + AMP trapped MsbA at different Hahn echo delays. Different delay times (0.4 and 1.6 µs), did not cause any significant changes in the relative line intensities of the different species, which suggests similar electron spin relaxation characteristics (see also figure 4.7).*

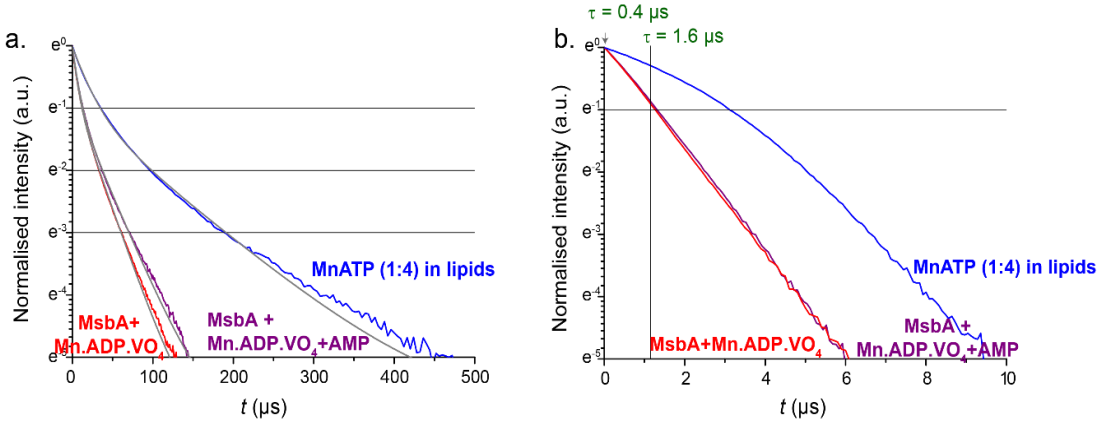


Figure 4.13- *263 GHz electron spin relaxation time measurements of Mn.ATP in the presence of lipids, MsbA with Mn.ADP.VO_4 and MsbA with Mn.ADP.VO_4+AMP recorded at 10 K.* ***(a)*** *Electron spin longitudinal relaxation time,* ***(b)*** *electron spin transverse relaxation time. In both cases, relaxation rates of Mn^{2+} in complex with MsbA significantly differ from the ones of Mn.ATP in lipid sample (see table 4.3). The longitudinal relaxation time measurements (a) were performed using a Picket Fence saturation recovery sequence with 28 π/2 pulses, where an interpulse delay was set roughly corresponding to $2T_{2e}$. Transverse relaxation time measurements are performed using Hahn-echo pulse sequence. The curves depicted in gray in the left figure are the fits with the biexponential function.*

***Table 4.3**: Electron spin relaxation characteristics (see figure 4.7)*

Sample	T_{1e}, µs*	T_{2e}, µs**
Mn.ADP.VO_4 trapped MsbA	30	1.3
Mn.ADP.VO_4 + AMP trapped MsbA	30	1.3
Mn.ATP in lipids	90	3.1

*T_{1e} is determined as τ_3 - τ_2, where τ_2 and τ_3 are the time constants after which the echo signal is decayed by a factor of e^2 and e^3, respectively. More detailed description of the procedure is given elsewhere[395].

**T_{2e} is determined as a time constant after which the echo signal is decayed by a factor of e.

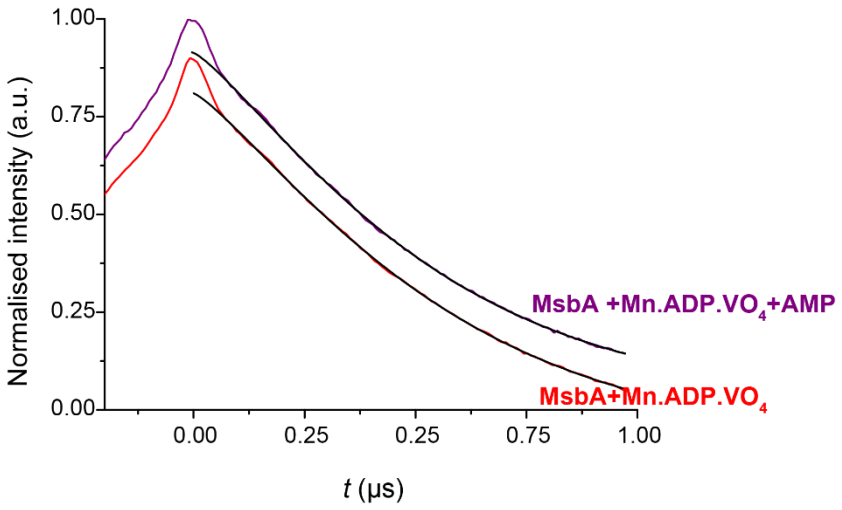


***Figure 4.14**- Raw experimental 263 GHz RIDME time traces and corresponding stretched exponential backgrounds for MsbA with Mn.ADP.VO_4 and Mn.ADP.VO_4 + AMP at 10 K. A five pulse RIDME sequence was used[372]. The time traces are obtained at the maximum of the lowest field hyperfine line.*

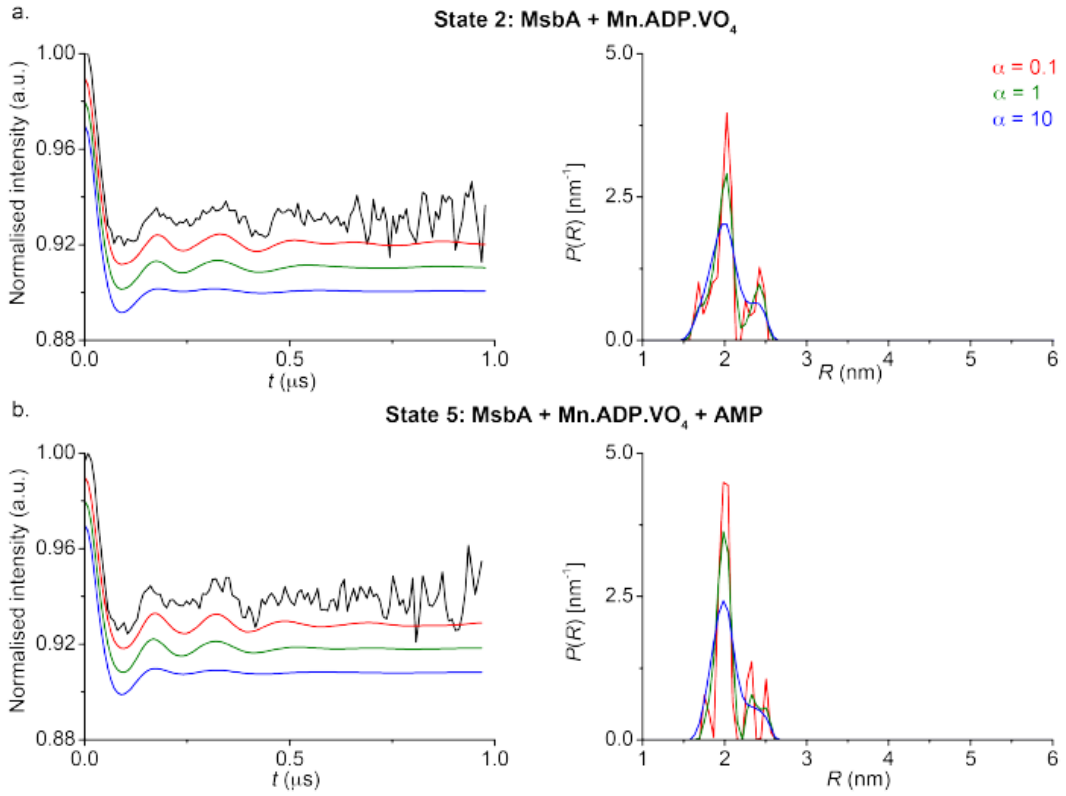


Figure 4.15- *The background-divided 263 GHz RIDME time traces obtained at sample temperatures of 10 K (depicted in black on the left) and Tikhonov regularization analysis performed with different regularization parameters (left: back-calculations, right: distance distributions) as given in the figure legend. (a) Mn.ADP.VO₄ trapped MsbA. (b) Mn.ADP.VO₄ + AMP trapped MsbA. The Tikhonov regularization is performed using DeerAnalysis2013 toolbox*[396] *using three different regularization parameters, 0.1, 1 and 10 (parameter α).*

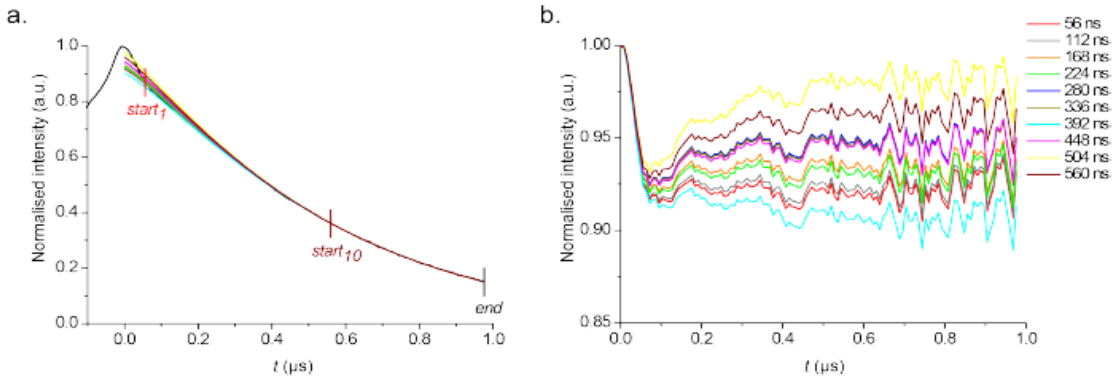


Figure 4.16- *(a) Experimental raw data of the RIDME time trace of MsbA in state (**2**) (MsbA.Mn.ADP.VO₄) and background functions obtained by a stretched exponential fit using 10 different initial positions. (b) Corresponding background-divided RIDME time traces.*

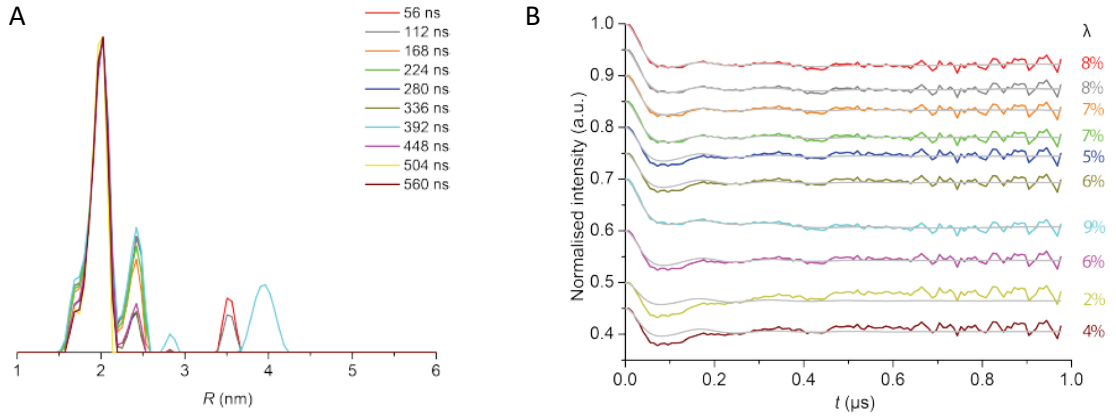


Figure 4.17- *A-Distance distributions obtained by Tikhonov regularization analysis (with regularization parameter of 1) of the background-divided RIDME time traces obtained on Mn.ADP.VO4 trapped MsbA (state (**2**)) shown in Figure B10. B- Background-divided RIDME time traces obtained on MsbA in state (**2**) (MsbA.Mn.ADP.VO$_4$, color code is consistent with figures 4.16 and 4.17) and backcalculations, based on Tikhonov regularization analysis (gray). The modulation depths are shown on the right.*

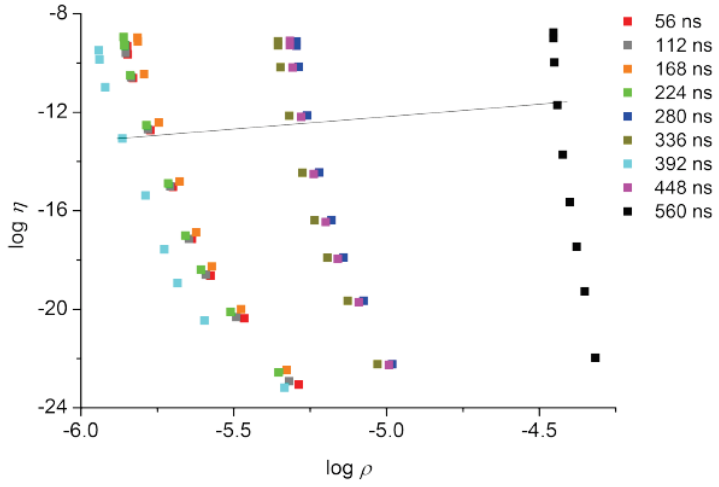


Figure 4.18- *L-curves corresponding to Tikhonov analysis of the background-divided RIDME time traces obtained on MsbA in state (**2**) (the trace with initial position of 504 ns is not considered in this analysis due to large deviation). The black line corresponds to regularization parameter 1. For the L-curve criterion, the axis of abscissas corresponds to the RMSD value between the background-divided experimental RIDME time trace and backcalculations $\rho = |S - KP_\alpha|^2$. The axis of ordinates quantifies the smoothness of distance distribution $\eta = \left|\frac{d^2}{dR^2}P_\alpha(R)\right|^2$ .S is the experimental time trace, K the kernel function and P_α denotes the distance distribution found for a given regularization parameter α.*

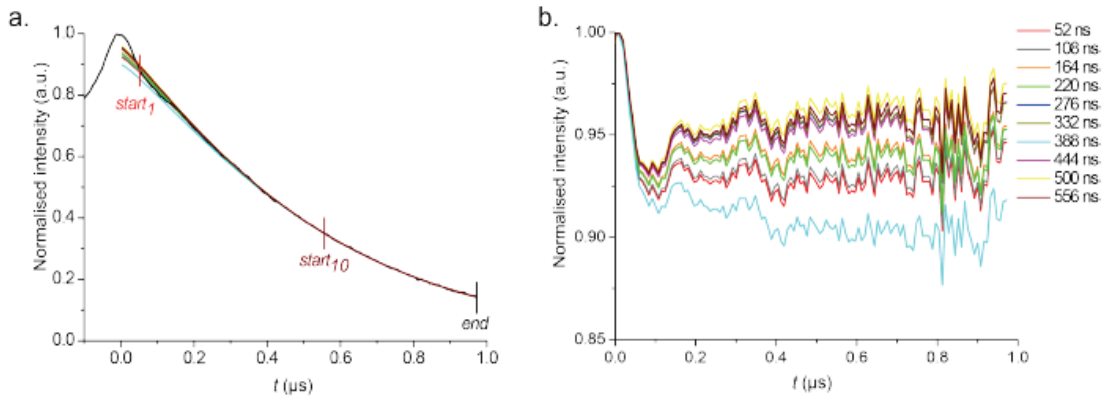


Figure 4.19- *(a) Experimental raw data of the RIDME time trace of MsbA in state (**5**) (MsbA.Mn.ADP.VO_4 + AMP) and background functions obtained by stretched exponential fit using 10 different initial positions. (b) Corresponding background-divided RIDME time traces.*

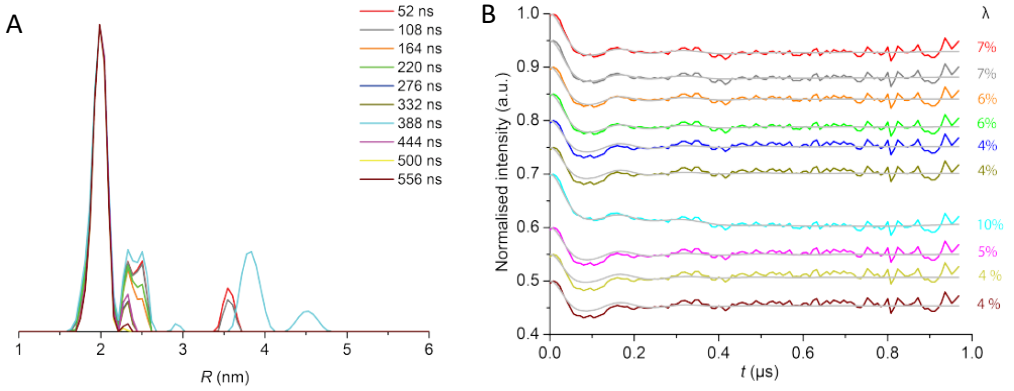


Figure 4.20- *A-Distance distributions obtained by Tikhonov regularization analysis (with regularization parameter of 1) of the background-divided RIDME time traces obtained for MsbA in state (**5**) (MsbA.Mn.ADP.VO_4 +AMP) as shown in figure 4.20a. B-Background-divided RIDME time traces obtained on MsbA in state (**5**) (MsbA.Mn.ADP.VO_4 + AMP, color code is consistent with figures 4.16 and 4.17) and backcalculations, based on Tikhonov regularization analysis (gray). The modulation depths are shown on the right.*

Table 4.4: Comparison of Mn^{2+}-EPR signal intensities based on 263 GHz EPR.

The EPR signal intensities, corresponding to Mn.ADP.VO_4 and Mn.ADP.VO_4 trapped MsbA, were compared based on the intensities of the electron spin echo signals measured on multiple days. Echo intensities adjusted to the same receiver gain and normalized to the inverted length of the π pulse. Both samples show the same signal intensity within the estimated error.

	Mn.ADP.VO_4		**Mn.ADP.VO_4+AMP**
1	**102**[1] 3th line, effectively 18dB gain, 18/32ns[(2)], 263.007118 GHz[(3)]	1	**127** 3th line effectively 18dB gain, 24/40ns, 262.968043 GHz
2	**119** 1st line, effectively 18dB gain, 16/28ns, 263.9477 GHz	2	**83** 1st line, effectively 18 dB gain, 14/26, 263.909673 GHz
3	**161** 1st line 18dB gain, 12/24, 264.252536 GHz	3	**132** 1st line 18dB gain, 12/24, 264.249573 GHz
4	**144** 1st line, 18 dB gain, 14/24ns, 264.169473 GHz	4	**137** 1st line 18dB gain, 14/26ns, 264.218353 GHz
Mean intensity.: 132±13		Mean intensity 120±12	

The EPR signal intensities, corresponding to Mn.ADP.VO4 and Mn.ADP.VO4 trapped MsbA, were compared based on the intensities of the electron spin echo signals measured on multiple days. Echo intensities adjusted to the same receiver gain and normalized to the inverted length of the π pulse. Both samples show the same signal intensity within the estimated error.

(1) Hahn-echo intensity (a.u.)

(2) π/2 and π pulse lengths

(3) Microwave frequency at which the echo intensity was measured

Protein expression, purification and reconstitution:

E. coli MsbA gene cloned into a pET-19b vector containing an N-terminal His10-tag connected via an 11 amino acid peptide linker has been used[99]. The plasmid was transformed into *E. coli* C43(DE3) cells for protein expression. For obtaining a high yield of MsbA, cells were grown initially in LB medium, collected by centrifugation and transferred to M9 minimal medium[397] . Expression was started by adding 10 ml of preculture to 1 liter of LB at 37 °C and 220 r.p.m. When the OD600nm reached 0.5–0.6, cells were harvested and then resuspended into 600 ml of M9 minimal medium. The cells were further incubated at 37 °C, 220 r.p.m. for 1 h for adaptation. The protein expression was induced with 1 mM IPTG at 20 °C and 260 r.p.m. once the OD at 600 nm has reached a value between 1.7 and 2.1. After 17 h of expression, the cells were harvested and resuspended in lysis buffer (10 mM Tris, 250 mM Sucrose, 150 mM NaCl and 2.5 mM $MgSO_4$, pH 7.5) with protease inhibitor, 0.5 mM dithiothreitol and DNase. Membranes were prepared by passing the resuspended cells through an I&L Biosystems high pressure cell disrupter at a pressure of 1.7–1.9 kbar (2–3 times), followed by centrifugation at 4,500 g (8,000 r.p.m., rotor F0850) for 10 min to remove cell debris followed by a final ultracentrifugation step at 223,000 g (55,000 r.p.m., rotor 70 Ti) for 1 h. Membranes were solubilized in buffer B (50 mM HEPES, 300 mM NaCl, 5 mM $MgCl_2$ and 10% Glycerol, pH 7.5) with 1.25% w/v n-Dodecyl-β-D-maltoside (DDM), 0.5 mM dithiothreitol and 10 mM imidazole at 4 °C overnight. The insoluble fraction was removed by ultracentrifugation at 223,000g (55,000 r.p.m., rotor 70 Ti) for 1 h. The remaining supernatant was loaded onto Ni-NTA prewashed with buffer B containing 50 mM imidazole. After 1.5–2 h of binding, elution was carried out using buffer B containing 0.015% w/v DDM and 400 mM imidazole. Typical yields of MsbA in 1L M9 minimal medium range between 10 – 13 mg and in 1L M9 minimal medium supplemented with amino acids range between 25 – 30 mg. Biochemical activity of each sample was determined using a phosphate based colorimetric assay as described before[59,369].

A lipid mixture of DMPC/DMPA (9:1) at molar lipid to protein ratios of 75:1 was used for reconstitution. Lipids were solubilized in $CHCl_3$:CH_3OH (2:1) and dried under a stream of nitrogen gas followed by vacuum rotor evaporation. The dried lipids were resuspended in buffer (50 mM HEPES and 50 mM NaCl, pH 7) and extruded 5 times through Whatman® Nucleopore™ Track-Etched membranes with pore diameters of 0.2 mm and 8 times with pore diameters of 0.1 mm. Liposomes were destabilized with 3-4 mM DDM. After drop-wise addition of protein in the DDM+liposome solution, the mixture was incubated at room temperature for 30 min. Detergent was removed by biobeads (80 mg/ml) that were first applied overnight at 4 °C and then again for 1 h at room temperature. The reconstituted protein was pelleted at 58,000 g (28,000 r.p.m., rotor 70 Ti) for 25 min. The supernatant buffer was decanted carefully. The pellet was then

resuspended in 700 ml of 20 mM HEPES. The pelleting and resuspension procedure was repeated three times. This process ensured removal of soluble components.

Trapping reconstituted MsbA in catalytic intermediate states:

As described in previous papers[59,361,369], each of the following reaction mixture was subject to freeze thaw cycles and then incubated at 37 °C for 20 min. The trapped samples were washed thrice with 20 mM HEPES (pH 7.3) to remove excess reagents before the final pelleting, which was carried out at 223,000 g for 3 h. The pellets were then transferred into the MAS rotor by centrifugation. **MsbA+ADP.BeF_3:** Beryllium fluoride is used as the phosphoryl analogue and has been used for trapping a prehydrolysis state in various NTPases[41,398]. Per 10 mg of reconstituted MsbA 10mM ATP, 10mM $MgCl_2$, 10mM $BeSO_4$ and 50 mM NaF were added. **MsbA+ADP.VO_4:** High-energy transition state that occurs during ATP hydrolysis can be emulated using vanadate[375,399]. Per 10 mg of reconstituted MsbA 10 mM ATP, 10 mM $MgCl_2$ and 3 mM orthovanadate were added. **MsbA+ADP.VO_4+AMP:** After vanadate trapping, AMP was added in the ratio of 3:1 mol/mol of AMP:MsbA. **MsbA+ADP.VO_4+ADPβS**: After vanadate trapping, ADPβS was added in the ratio of 6:1 mol/mol of ADPβS:MsbA. MsbA+AMP: To the 13C-MsbA proteoliposomes, AMP was added in the ratio of 6:1 mol/mol of AMP:MsbA.

Parameterization of AMP and ATP

The parameterization of ATP used in this work is the same as described in Oliveira *et al*[161]. The AMP has a *p*Ka of 6.2 in solution. Considering its proximity to the magnesium ion and the protein environment, AMP was set as deprotonated with a total charge of -2. The nucleotide charges (up to C5') were the ones from the GROMOS 54A7 force field. The phosphate partial charges were calculated, because the force field did not contain parameters for the deprotonated versions of these nucleotides. To this end, the approach used by Meagher et al was used[400], as well as the Hwang et al extended conformations in quantum chemical calculations. The partial charges for the phosphate group (along with a CH_3 group introduced in the phosphate chain to mimic the connection to ribose) were calculated using GAUSSIAN09[401] and RESP fitting[402]. The molecule was geometry optimized using RHF and the 6-31+G(d) basis set. Electrostatic potentials calculated on the final structure were used for fitting the partial charges with RESP fitting. In opposition to the usual procedure, polarization functions were used to account for highly charged molecule as done by Meagher *et al*[400].

Determination of protonation states

The protonation state of each residue in the protein has to be determined prior to MD simulation. The determination of the protonation states of the titrable residues of the protein and the ligand at pH 7 was done using a methodology previously reported[316,317], in which continuum electrostatics calculations and Monte Carlo sampling of the protonation states are used. The continuum electrostatics calculations were done using the MEAD package version 2.2.9. The electrostatic energy terms resulting of these calculations were necessary to sample the protonation states at pH 7, using the PETIT software[300]. These calculations were performed on the initial structure prior the simulation. According to the results of these calculations, all lysine residues were protonated and all glutamate and aspartate residues were deprotonated. Histidine 214 was set as neutral with the proton on Nδ, while histidines 399, 405, 537 and 562 were set as charged. Histidines 427, 569 and 576 were set as neutral but possessing a proton on Nε. The protonation state of ATP was also calculated and it was set as deprotonated on the γ phosphate. It has a charge of -4.

Clustering

The Jarvis-Patrick method[374] requires two parameters: the number of neighbors to be examined and the minimum number of neighbors in common. These parameters were tuned to the different systems in order to avoid excessive scattering of the conformations throughout the clusters.

Table 4.5- Table showing the parameters used for Jarvis-Patrick clustering in the simulated systems.

Simulated system	Number of neighbors examined	Number of neighbors in common
2AMP+2ATP	2	3
1AMP+2ATP (set 1)	8	3

Structural Behavior of the simulations with nucleotides

The structural behavior of the protein was monitored measuring the root mean square deviation (RMSD) in comparison with the initial structure, before minimization and equilibration, providing a measure of the structural drift relatively to the initial model prior to minimization and equilibration. RMSD was calculated using the Cα atoms only and fitting was done to the entire dimer.

Overall, the RMSD value increased and become more or less constant for all replicates. The RMSD values observed ranged between 0.3 to 0.4 nm due to a rigid body motion, in which there is a slight approximation of both NBDs.

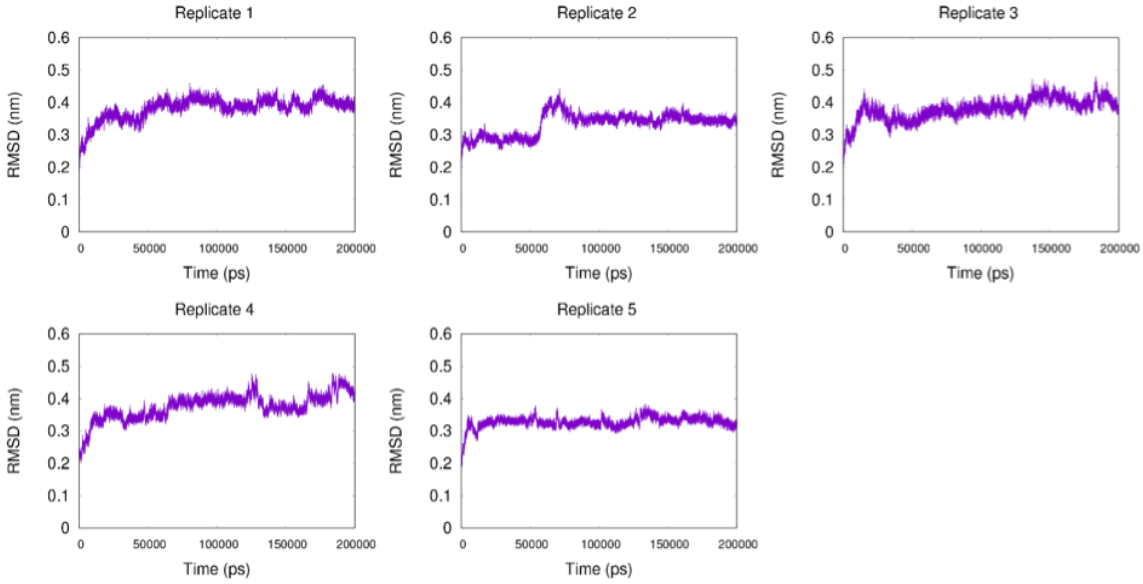


Figure 4.21- *Structural behavior of the MsbA^{nbds} dimer with bound ATP: RMSD evolution relative to the initial structure before minimization and equilibration. In replicate 2 the sudden rise of the RMSD is due to the increased mobility of the N-terminal of one of the NBDs.*

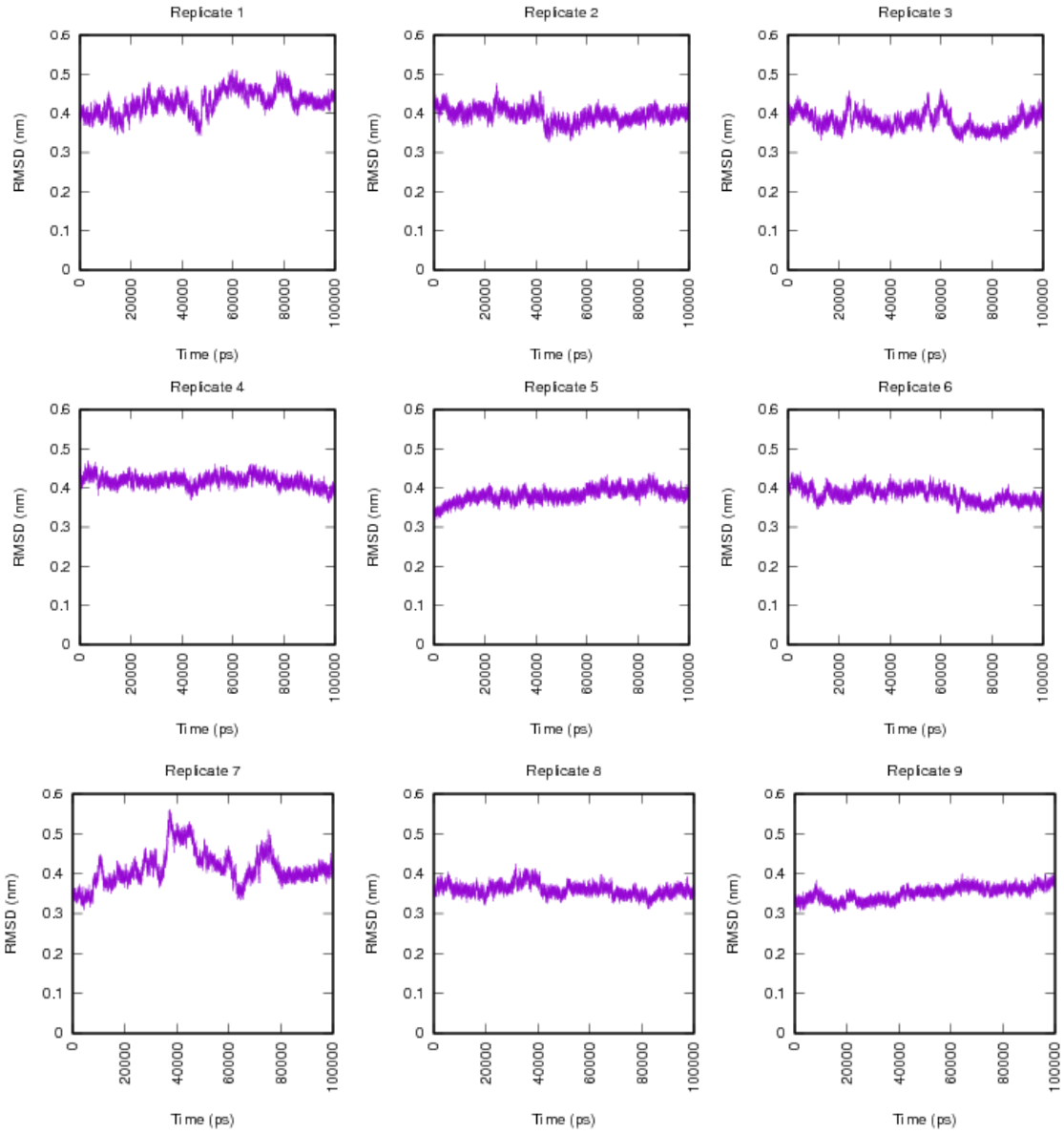


Figure 4.22- *Structural Behavior of the simulations of the MsbA^{nbds} dimer with 2 ATPs on the ATPase site and 2 AMPs on the AK-site: RMSD evolution relative to the initial structure before minimization and equilibration. The RMSD was calculated using the Cα atoms only and fitting was done to the entire dimer. In some replicates, the increments of the RMSD are due to the high flexibility of the terminal ends of the protein and are considered not to influence the binding of AMP and the desired properties.*

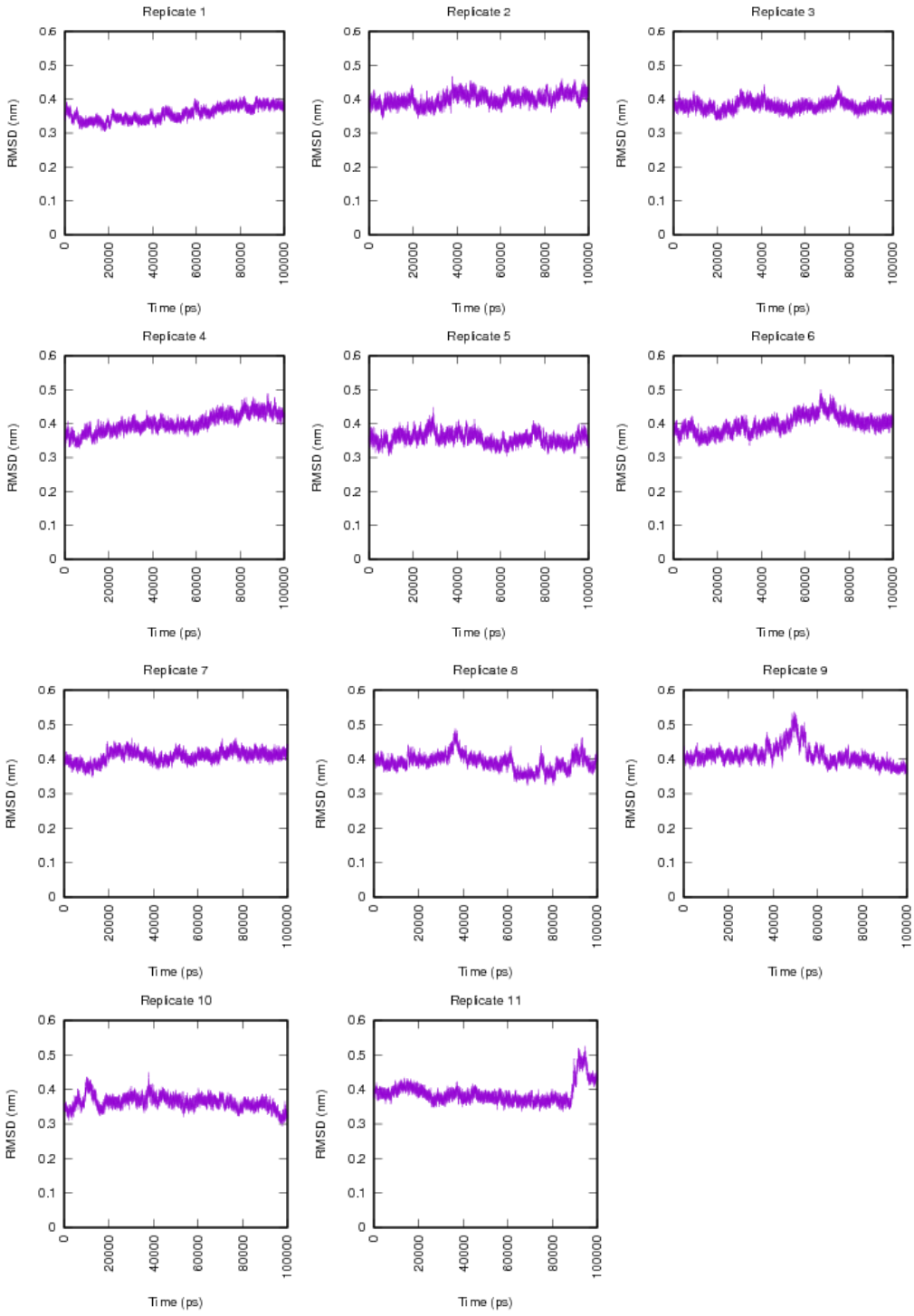


Figure 4.23- *Structural behavior of the simulations of the MsbA^{nbds} dimer with 2 ATP on the ATPase site and 1 AMP on the AK-site - set 1 of simulations: RMSD evolution relative to the initial structure before minimization and equilibration. The RMSD was calculated using the Cα atoms only and fitting was done to the entire dimer. In some replicates, the increments of the RMSD are due to the high flexibility of the terminal ends of the protein and are considered not to influence the binding of AMP and the desired properties.*

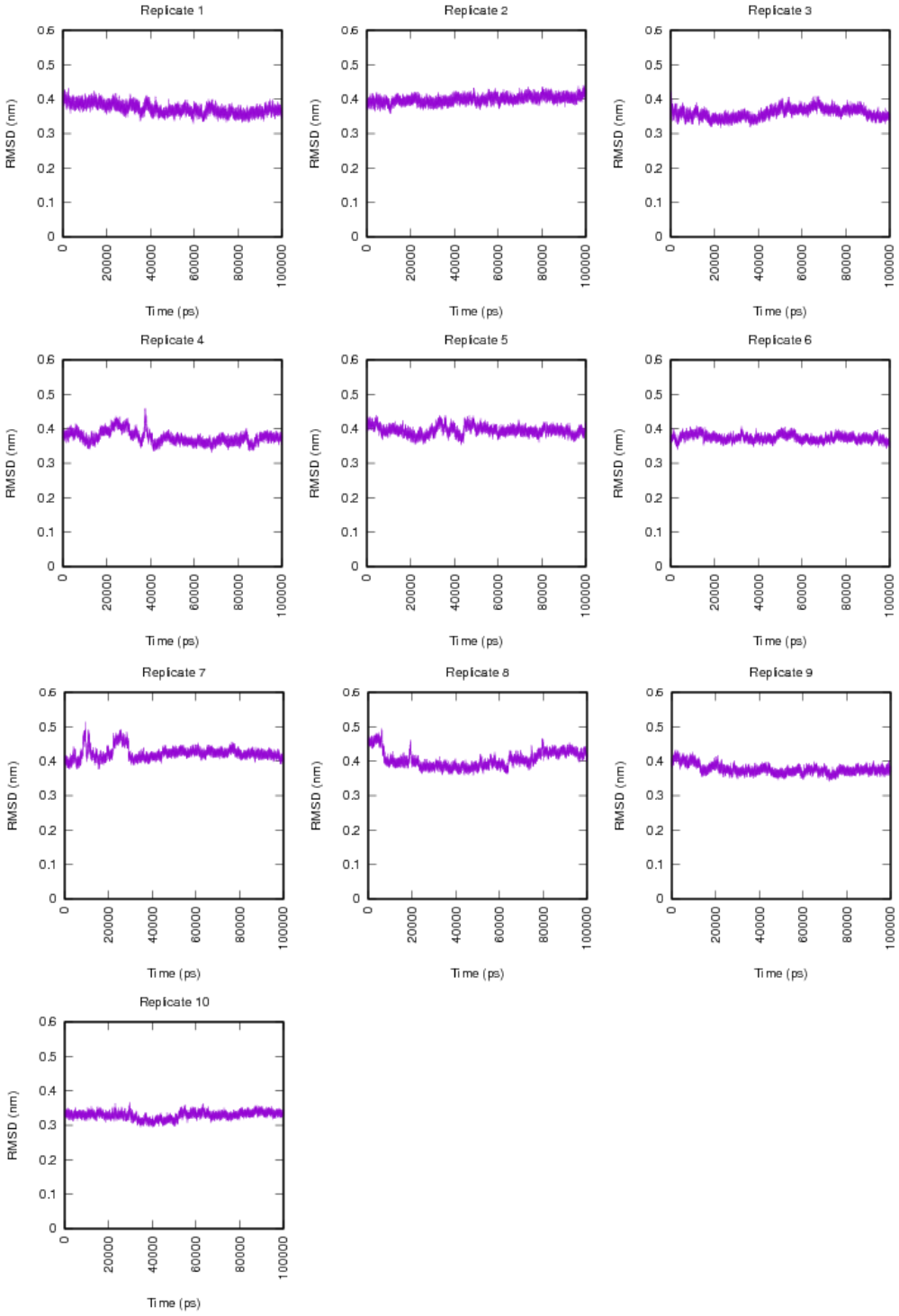


Figure 4.24- *Structural behavior of the simulations of the MsbA^{nbds} dimer with 2 ATP on the ATPase site and 1 AMP on the AK-site - set 2 of simulations: RMSD evolution relative to the initial structure before minimization and equilibration. The RMSD was calculated using the Cα atoms only and fitting was done to the entire dimer. In some replicates, the increments of the RMSD are due to the high flexibility of the terminal ends of the protein and are considered not to influence the binding of AMP and the desired properties.*

Chapter 5

ATP hydrolysis and nucleotide exit enhance maltose translocation in the $MalFGK_2E$ importer

This work was published in the following paper:

Abreu, B., Cruz, C., Oliveira, A. S., Soares, C., ATP hydrolysis and nucleotide exit enhance maltose translocation in the $MalFGK_2E$ importer, *Scientific Reports,* **2021, 11,** 10591

Contributions of the author of the present thesis to this work:

In this work I participated in the experimental design, performed all simulations and analyses, and wrote the manuscript.

5.1-Summary

ATP binding cassette (ABC) transporters employ ATP hydrolysis to harness substrate translocation across membranes. The *Escherichia coli* MalFGK$_2$E maltose importer is an example of a type I ABC importer and a model system for this class of ABC transporters. The MalFGK$_2$E importer is responsible for the intake of malto-oligossacharides in *E.coli.* Despite being extensively studied, little is known about the effect of ATP hydrolysis and nucleotide exit on substrate transport. In this work, we studied this phenomenon using extensive molecular dynamics simulations (MD) along with potential of mean force calculations of maltose transport across the pore, in the pre-hydrolysis, post-hydrolysis and nucleotide-free states. We concluded that ATP hydrolysis and nucleotide exit trigger conformational changes that result in the decrease of energetic barriers to maltose translocation towards the cytoplasm, with a concomitant increase of the energy barrier in the periplasmic side of the pore, contributing for the irreversibility of the process. We also identified key residues that aid in positioning and orientation of maltose, as well as a novel binding pocket for maltose in MalG. Additionally, ATP hydrolysis leads to conformations similar to the nucleotide-free state. This study shows the contribution of ATP hydrolysis and nucleotide exit in the transport cycle, shedding light on ABC type I importer mechanisms.

5.2-Introduction

ATP binding cassette (ABC) proteins are one of the largest protein superfamilies, being ubiquitous in all domains of life. In common they all share two domains - the nucleotide binding domains (NBDs), that possess characteristic sequences that play a role in ATP binding and hydrolysis[11], allowing ABC proteins to fulfil their function. ABC transporters are a subclass of ABC proteins responsible for the translocation of various substrates across membranes, making use of ATP hydrolysis to harness transport. In addition to the NBDs, ABC transporters also contain transmembrane domains (TMDs) that bind and translocate the substrates[2,18]. In this way, these proteins can either act as importers or exporters, depending on directionality of the transport. ABC transporters play key roles in various processes such as drug excretion in bacteria and cancer cells, lipid export for membrane building and assembling, nutrient intake and even maintenance of transmembrane gradients[3].

ABC importers responsible for intake of molecules, such as nutrients and metals are exclusive of bacteria[2]. In addition to the NBDs and TMDs, they also have substrate binding proteins (SBPs) that scavenge the substrate from the extracellular medium and deliver it to the transmembrane complex. Most importers are associated with the intake of nutrients and some are even involved

in pathogenicity, such as the zinc importer ZnuABC present in *Brucella abortus* [70,403]. They can be divided in three subclasses: the type I, type II and type III importers, which differ in their structural features[18]). The *Escherichia coli* MalFGK$_2$E importer is a type I importer located in the inner membrane responsible for the transport of maltose and malto-oligossacharides from the periplasm. It is a heterodimer, constituted by five subunits: two copies of MalK, which constitute the NBD dimer, MalF and MalG are part of the TMDs and the substrate-binding protein – MalE[143]. figure 5.1 shows the structure of MalFGK$_2$E in the two distinct conformations used in this work. This importer is a model system for type I ABC importers, being extensively studied over the years. Substrate translocation in ABC transporters has been explained on basis on the alternating-access model, in which the transporter alternates between an outward-facing conformation and an inward-facing one[89]. Considering this model, in the outward-facing state, the TMDs are oriented in such a way that the transmembrane cavity is exposed towards the periplasm (figure 5.1a), while the inward-facing state contains the transmembrane cavity exposed to the cytoplasm (figure 5.1b). MalE delivers the substrate to the transmembrane complex in the outward-facing form.

The transport cycle of maltose transport starts with the delivery of maltose to the complex, mediated by MalE (figure 5.1c – step 1). This first step is disputed because there are two different models to explain substrate delivery: the first one defends that there is binding of loaded MalE to the complex leading to the outward-facing conformation[143]. The second model proposes that unliganded MalE binds to the complex, and along with ATP binding, generates the outward-facing conformation[404]. Afterwards, maltose binds to the MalE-MalFGK2 complex, via MalE (figure 5.1c -steps 2 and 3)[404]. Following maltose loading to the complex, ATP hydrolysis triggers its opening causing the conversion to the inward facing form and maltose release. In this chapter we will study the events happening on this latter stage (figure 5.1c- step 4).

Numerous experimental studies, along with simulation studies, have managed to characterize many of the molecular details of transport process, but the sequence of events that result from ATP hydrolysis and their corresponding molecular details, remains somewhat elusive. Nonetheless, EPR and cryo-EM data show that ATP hydrolysis results in a structure with a semi-open NBD dimer[147,148,405]. Moreover, kinetic studies proposed that release of phosphate is the rate limiting step and that maltose accelerates release of phosphate and ATP turnover[406]. Previous MD studies on the maltose MalFGK$_2$E have also versed several stages of the transporter function. For instance, it was discovered that MalE binding decreases the energetic barrier for MalFGK$_2$E complex closure, promoting the movement of the MalF and MalG helices towards the center[173]. In this way, the presence of MalE stabilizes the pre-translocation state. On the other hand, in the absence of MalE, the transmembrane helices of MalF arrange in a way that blocks

the translocation pathway. MalE binding also suppresses fluctuation of the P2-loop of MalF[407]. The effect of nucleotide binding to the NBDs was also simulated, as well as its energetic and mechanistic charcterization[162–164,407]. ATP binding in both pockets induces the closed form of the NBD dimer, while ATP hydrolysis triggers dimer opening[163]. Other simulation studies versed the communication between NBDs and TMDs[169,408] and concluded that the coupling helices and the Q-loops are essential for transmitting the impact of ATP hydrolysis from the NBDs to the transmembrane domains. Additionally, the mechanism of ATP hydrolysis was also investigated[47,49]. The authors suggested that ATP hydrolysis proceeds via a dissociative mechanism with a trigonal bipyramidal transition state. The conformational changes triggered by ATP hydrolysis in the complex were also studied, confirming the role of the coupling helices in transmitting the energy created by ATP hydrolysis[316]. So far, these studies characterized events that happen until ATP hydrolysis (included) and focused on the transformations suffered by the protein.

The present study aims to shed light on the process of substrate translocation. We investigate the consequences of ATP hydrolysis and nucleotide exit on the substrate transport, and on the molecular interactions between maltose and the protein complex, from the structural and energetic viewpoints. Regarding the full transport cycle, as portrayed in figure 5.1c, our study verses events on the last step of the process, when ATP is hydrolysed, and the substrate is released. To this end, we will use extensive molecular dynamics simulations of the maltose importer with maltose bound in the transmembrane domains with ATP, ADP.Pi and in the absence of nucleotides, henceforth designated as the Apo state, this latter one intended to mimic the state after hydrolysis and after nucleotide exit. The ATP and ADP.Pi states will be generated from the structure with an ATP analogue bound (figure 5.1a), while the Apo state will be created from the pre-translocation structure without nucleotides (figure 5.1b).

Even though X-ray structures containing the bound maltose (figure 5.1a and 5.1b) and even longer malto-oligassacharides[69] provide details about substrate binding, MD simulations allow a dynamic characterization of the binding and transport processes, along with the study of impact of ATP hydrolysis on the latter.

The main findings of our study are that ATP hydrolysis significantly lowers the energetic barriers for substrate translocation, and that nucleotide departure has a very similar effect as hydrolysis. Additionally, we have identified the protein conformational changes that are responsible for this effect. The most relevant conformational changes happen on key transmembrane helices, that result in increased substrate diffusion towards the intracellular medium. Potential of mean force (PMF) calculations confirmed that hydrolysis decreases the energetic barriers for the transport process and enhances its irreversibility.

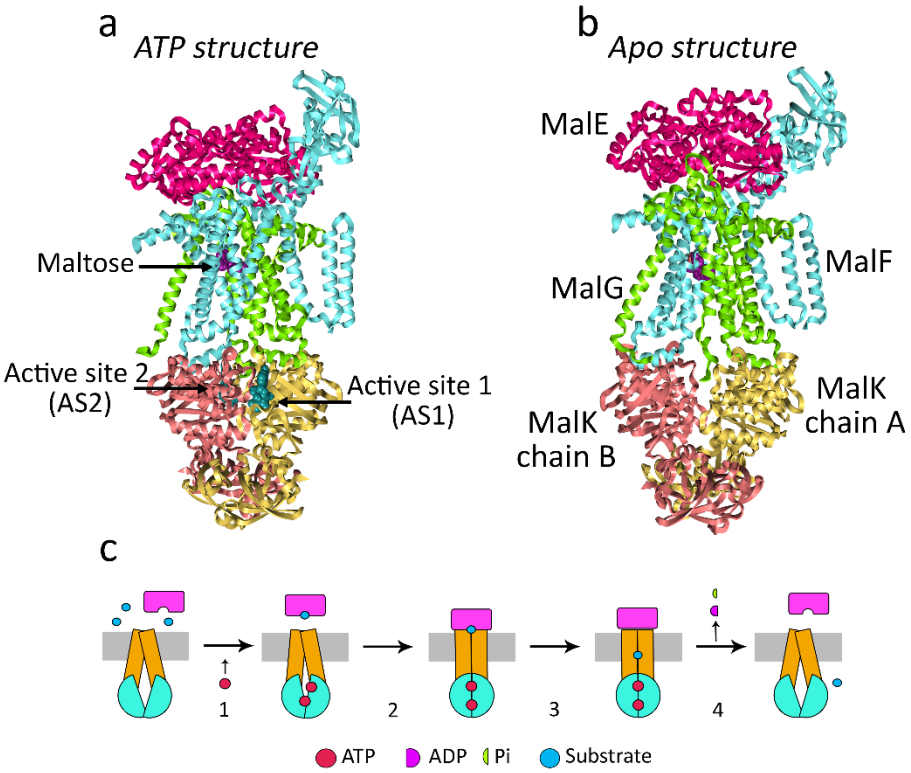


Figure 5.1- *Structures used in this work. (a)-Initial structure for the simulating the ATP state. This structure contains an ATP analogue bound in the NBDs. The post-hydrolysis state was generated from this structure. PBD code: 3RLF. (b)-Initial structure used to mimic the protein after nucleotide exit. This structure reflects the pre-translocation state and can be used to extrapolate about the state after nucleotide exit. PBD code: 3PV0. The MalE is represented in bright pink, MalF in cyan, MalG in green and MalK in yellow and salmon. ATP is represented in teal spheres and maltose in purple spheres. (c)- Transport cycle in the maltose importer. MalE is represented in purple, the TMDs in orange, the NBDs in cyan, ATP in red spheres, maltose as a blue sphere and ADP and Pi as purple and green semi-circles. 1-ATP binding and MalE loading with maltose.2-MalE binding to the complex. 3- Maltose diffusion in the complex. 4- ATP hydrolysis, release of the substrate and hydrolysis products and transporter reset.*

5.3-Methods

5.3.1-System setup

In order to achieve the goals previously stated in the introduction, we simulated three relevant states; i) the pre-hydrolysis state with ATP bound in the NBDs, to which the starting point was the pre-hydrolysis structure crystallized with ANP-PNP (PDB code: 3RLF)[41]; ii) The post-hydrolysis state was generated from this state as explained on section 2.3; iii) The Apo state, mimicking the nucleotide exit was simulated using the pre-translocation state (PDB code: 3PV0)[143]. We did this because there is no available X-ray structure explicitly reflecting the nucleotide exit state, but experimental EPR data shows that the post-hydrolysis state is highly similar to the pre-translocation structure [147,405,409]. The main similarities comprise the semi-open NBD dimer[147,405,409] and MalE binding to the complex[147,404]. In fact, experimental EPR and Cryo-EM show that MalE induces closure of the NBD dimer and stabilizes a semi-open NBD conformation in the absence of nucleotide, such as the one observed in this structure[147,405,148].

The missing segments were rebuilt using MODELLER version 9.6[290]. All structures were rebuilt in order to have the same number of residues, with the goal of increasing the structural similarity. In the pre-translocation structure (PDB code: 3PV0), the reconstructed segments were: residues 371 to 374 in MalE, residues 10 to 28, 242 to 248, 504 to 505 in MalF, and the segment 280 to 296 in MalG. Additionally, the dissulfide bridge between residues 69 and 337 in MalE, which was engineered for crystallisation, was reverted to the original serine residues in those positions. Additionally, the maltose bound to MalE was removed. In the pre-hydrolysis structure (PDB code: 3RLF), the reconstructed segments were the residues 371 to 374 in MalE, and residues 10 to 28, 242 to 248, 504 to 505 and 280 to 296 in MalF. The ANP-PNP molecule was replaced by ATP. Afterwards, each protein was embedded in a pre-equilibrated 480 POPC membrane. The optimal protein orientation was determined using LAMBADA[410] and the oriented protein was inserted in the membrane manually, removing lipid molecules that were within a 1.2 Å distance cut-off.

In order to obtain the protonation states taking into account the effect of membrane environment, first we build a temporary version of the system. The protein was inserted in the membrane as described in the above paragraph. The system was further solvated and neutralized. This system was minimized in two stages, using a maximum of 50000 steps each: the first, using position restraints of 1000 kJ/mol in all the heavy atoms, and the second stage with position restraints of the same magnitude in the protein-ligand complex heavy atoms only. An equilibration step was necessary to allow optimal membrane fitting to the protein. The

equilibration phase was also performed in two stages: the first contemplated the usage of position restraints of 1000 kJ/mol in the protein-ligand complex heavy atoms, at constant temperature and pressure during 500 ps. The Berendsen baths were used for controlling temperature and pressure, with coupling constants of 0.1 and 1 ps respectively. The second stage equilibration was done using position restraints of 10000 kJ/mol in the complex heavy atoms for 20 ns in the same conditions.

Finally, the protonation states of the protein were determined at pH 7.0 using the PETIT and MEAD packages[300,411] and a final system was built using this information and the above described protocol for membrane insertion. The final equilibration protocol is described in section 5.3.2. The detailed protonation states of the groups are described in tables 5.1 and 5.2.

5.3.2-Simulation setup

The equilibrium MD simulations were performed using GROMACS 5.0.7[257], along with the GROMOS 54A7 force field. The POPC parameters used were the ones derived by Poger at al[412] and the GROMOS53A6 CARBO parameters were used to model maltose[413]. The final protonation states of ATP and ADP were the same used in Oliveira *et al*[161,171,316] and Damas *et al*[414] corresponding to a charge of −4, −3, respectively. A charge of -1 was assigned to the phosphate ion. The integration time step used was 2 fs. The systems were simulated using periodic boundary conditions and at constant temperature and pressure. Ions were added in order to neutralize the system. The temperature was set to 303 K using a velocity-rescale heat bath[268], with a coupling constant of 0.1 ps with two coupling groups: one for the protein-nucleotides complex and another for the solvent and ions. The pressure was kept around 1 atm by semi-isotropic coupling with a Parrinello-Rahman bath with a coupling constant of 2.0 ps and a compressibility of 4.6×10^{-5} bar^{-1} [270]. A cut-off of 1.0 nm was used in the calculation of van der Waals interactions. Long range electrostatic interactions were treated using the particle mesh Ewald (PME)[261 319] method using a real-space cut-off of 1.0 nm. All neighbour lists were updated every 10 steps. All bonds were constrained to their equilibrium lengths using the LINCS[415] algorithm, except for the water molecules in which the SETTLE algorithm was used to constrain its bonds[416]. The SPC model for water was used[373]. After system building, its potential energy was minimized in three stages with a maximum of 50000 steps each: the first using position restraints of 1000 kJ/mol/nm on all the heavy atoms of the system, the second with position restraints of the same magnitude, but on the complex heavy atoms only, and the third with position restraints on the C-α atoms only. The equilibration stage was also done in three steps: the first one for 1 ns with position restraints of 1000 kJ/mol/nm on all the heavy atoms of the

system, the second one for 1 ns with position restraints on the heavy atoms of the protein and the last one with position restraints on the C-α atoms.

Nine replicates of the pre hydrolysis states were simulated for 300 ns, while the post hydrolysis state was simulated for 360 ns and the apo state for 410 ns. The first 40 ns of the pre hydrolysis state were discarded as equilibration time, as well as the first 100 ns for the post hydrolysis state and the first 150 ns of the apo state. The effective simulation time analysed was 260 ns.

5.3.3-Generation of the post-hydrolysis state

The post-hydrolysis state was generated extracting conformations from the ATP state at 20 ns. This timeframe was chosen to maximize the probability of collecting catalytically competent conformations, prior to the generation of the post-hydrolysis state. We have defined the following criteria to assess if a conformation is catalytically competent: the catalytic residues H192 and E159 must be oriented towards ATP, with a water molecule coordinated by both residues. Additionally, in order to ensure contact between ATP and the active site motifs, only conformation with an intermotif (ABC-Walker A) distance lower than 1.2 nm were considered.

Similarly to previous works[32,171,316,317], the post hydrolysis state was generated using the slow-growth method, making the transformation from ATP to ADP and phosphate coupled to a lambda parameter that varies from 0 to 1, allowing a smooth conversion of the molecules. In this case we are not interested in calculating the free energy associated with this transformation. Nonetheless, in the present case, the entire hydrolysis process was divided in two stages. The first step mimics the hydrolysis process, while the second mimics the catalytic residue regeneration. In the first step the conversion of ATP to ADP and phosphate is simulated, along with the annihilation of a proton in H192 and the creation of a proton in E159. In this way, the final state agrees with recently proposed catalytic mechanisms for ATP hydrolysis in the maltose importer[47]. In the second stage of hydrolysis there is the regeneration of the proton states of residues H192 and E159. The first step, that simulated the hydrolysis process was performed for 5 ps using a timestep of 0.0005 ps, while the regeneration step was done for 2 ps, using the same timestep. After both slow-growth procedures, an extra positive charge is added to neutralize the system, in the form of a sodium ion replacing a water molecule far away from the protein. This allows to avoid inconsistencies due to the PME algorithm. These simulations were performed using the same conditions as described above.

5.3.4-Analysis of equilibrium MD simulations

Analyses were done using the tools included in the GROMACS 5.0.7 package.

The harmonic ensemble similarity (D_{HES}) is a measure of the similarity between ensembles and was derived by Lindorff-Larsen *et al*[417]. We have used the implementation available in the ENCORE toolkit[418] included in the MDAnalysis package[321,322]. D_{HES} was calculated between the ADP and ATP states using the C-α positions of each individual residue. The trajectories were previously fitted to a common structure and the comparisons were made in a cross-replicate way, i.e, each replicate was compared to all other replicates, apart from itself. The displayed values correspond to the average for all comparisons for each residue. The pore radius was measured using HOLE, assigning the GROMOS atomic radii [419]. The secondary structure was determined using the DSSP program. The elements of secondary structure considered were α-helix, β-sheet, β-bridge, turn, and 3_{10} helix, according to the DSSP classification[324].

Principal component analysis (PCA) was performed using the atomic coordinates of the C-α atoms, obtained during the equilibrium MD simulations. We followed the protocol defined in Campos *et al*[325]. First, a central structure of the trajectory set is determined, being this structure the one that minimizes the total dispersion of the average of the squared RMSD. This central structure is then used for fitting the whole trajectory used for the PCA calculation, and as reference structure for the calculation of the covariance matrix. Afterwards, PCA of the covariance matrix is performed. The density maps were obtained using a two-dimensional grid with a spacing of 0.2 Å and a gaussian kernel estimator[325].

The diffusion maps of maltose were obtained by first extracting the center-of-mass coordinates using MDAnalysis[321,322] and the probability densities were estimated via a three-dimensional grid with a spacing of 0.05 nm^3 using a gaussian kernel estimator[325].

The 3D structure maps were obtained using PDBsum server[420] and the D_{HES} were mapped with the aid of PyMOL 2.0.

All the error bars presented in the histograms correspond to the 95% confidence interval obtained by bootstrapping[326].

PyMOL 2.0 was used to visualize trajectories and produce the pictures presented[327]. The plots were produced using the Matplotlib[421] package in Python and Gnuplot[422].

5.3.5-Potential of mean force (PMF) calculations

The PMF profiles were calculated using the umbrella sampling method. Pulling simulations were performed with the goal of generating initial conformations for umbrella sampling windows. GROMACS 2018.4[423] was used for pulling and umbrella sampling simulations, in the same conditions as the equilibrium MD, but this time setting an isotropic pressure coupling, to avoid fluctuations of the box in the zz' dimension that affect the reaction coordinate, which is also in the zz' direction. Pulling simulations were started from frames extracted from equilibrium MD simulations at several times. In these frames, the maltose molecules were at random positions along the pore as a result of their diffusion. In order to span the entire pore length, bidirectional pulling was made. In the direction towards the NBDs, the reaction coordinate used was the z-component of the distance between the center of mass of maltose and V442. The simulations were stopped as maltose reached the level of the coupling helices. An harmonic potential was applied to this reaction coordinate, with a force constant of 1000 kJ/mol/nm^2 and a constant velocity of 3.5 Å/ns was used in the pre and post hydrolysis states, while 2.5 Å/ns was used in the nucleotide-free state. The systems were simulated for a maximum of 50 ns. In the direction towards MalE, the reaction coordinate used was the z-component of the distance between the center of mass of maltose and MalE. In a similar way to the previous simulations, a harmonic potential was applied to this reaction coordinate, with a force constant of 1000 kJ/mol/nm^2. The conformations extracted from equilibrium MD simulations were simulated for a maximum of 20 ns. In the ATP simulation, the pulling velocity used was of 2.5 Å/ns. In the ADP simulation, 20 ns were simulated at the pulling velocity of 1.5 Å/ns, along with more 13 ns at the pulling velocity of 3.5 Å/ns. In the Apo state, 20 ns were simulated at the pulling velocity of 2.5 Å/ns, along with an extra 16 ns at the pulling velocity of 3.5 Å/ns. The extra step with increased velocity was necessary to overcome the steric hindrance caused by the different conformation of the transmembrane domains and reaching the periplasmic gate level.

For the umbrella sampling simulations, the PLUMED 2.5.1 plugin was used. The windows were initially spaced by 0.06 nm, but depending on the window behaviour, further windows were necessary in certain regions, while in other regions, windows were removed to avoid excessive overlap. The final list of windows used for each PMF can be found in supplementary information in tables 5.4 to 5.6. Harmonic restraints were used with a force constant of 500 kJ/mol/nm^2 or 800 kJ/mol/nm^2, depending on the difficulty of sampling the region. The restraint potential was applied to the z-component of the center of mass of maltose. The umbrella windows were simulated for 50 ns, in which the first 20 ns were discarded as the equilibration period. The PMF profiles were obtained using the Umbrella Integration method by Kästner et al[285,286]. We used the code created by Stroët et al[424]. The orientation angle of maltose was calculated using

MDAnalysis[321]. This angle is defined as the angle between the z-axis and the vector that links the most distant carbon atoms in maltose rings.

The probability densities maps for the analysing umbrella sampling simulations were obtained by estimating densities in a two-dimensional grid with a spacing of 0.1 hydrogen bonds/nm, 0.1 deg/nm and 0.1 Å/nm using a Gaussian Kernel estimator[325].

5.4-Results and discussion

5.4.1-Structural stability of MalFGK$_2$E

In this work we simulated the *E. coli* MalFGK$_2$E importer in different states of the transport cycle: the pre-hydrolysis state with ATP bound to the NBDs, the post-hydrolysis state with ADP and phosphate bound and an Apo (nucleotide-free) state.

In order to evaluate the structural stability of these simulations, the temporal evolution of the C-α RMSD and the percentage of retained secondary structure were monitored, along with a visual inspection of the trajectories. Regarding the ATP state, the C-α RMSD of the full complex reached a value of 0.35 nm at 40 ns, and a small drift was observed that increased the RMSD values up to 0.4 nm (figure 5.9). When observing the behaviour for each subunit it is possible to observe that the C-α RMSD of MalE and MalG stabilized around 0.2 nm (figures 5.11 and 5.12). The two MalK chains show higher values, around 0.25 nm and a stable evolution (figures 5.13 and 5.14). In contrast, the C-α RMSD of MalF shows a similar behaviour to the C-α of the full complex, reaching values around 0.3 nm and a similar drift behaviour. After a visual inspection of the trajectories, we observed that the P2 loop of MalF displays rigid-body motions, varying its position around MalE (figure 5.15). Another evidence that the complex has retained its structure throughout the simulations, is the evolution the percentage of retained native secondary structure, in comparison with the initial structure (figure 5.10), that remained stable around 95% from 40 ns onwards. Therefore, we have considered the first 40 ns of simulation as an equilibration period and were discarded. When looking at the post-hydrolysis state, a similar behaviour is observed, with the whole complex reaching C-α RMSD values around 0.35 nm, but without any significant drift (figure 5.9). MalE and MalG show again the lowest values, around 0.2 nm (figures 5.11 and 5.12), with a stable evolution. However, in this state, both MalK chains show higher C-α RMSD values (around 0.3 nm) than the MalF subunit (around 0.28 nm), and closer to the whole complex (figures 5.13 to 5.15). A visualization of the trajectories showed that significant conformational changes happen to MalK, such as the active site (AS) opening, rigid body motions back and forth of regulatory domains in the *xy* plane and increased flexibility

of the C-terminal segments. This resulted in a significantly higher RMSD in comparison with its ATP counterpart. During the production MD, the transmembrane domains have adopted a more relaxed conformation, that led to a pore radius decrease in the periplasmic side and increase at the cytoplasmic side (figure 5.16), allowing the study of maltose translocation. In a similar way to the pre-hydrolysis state, the first 100 ns of simulation were discarded. This state also retained approximately 95% of its secondary structure (figure 5.10). The nucleotide-free Apo state started from an inward-facing structure. Nonetheless, within the first 20 ns of simulation, a spontaneous approximation of the TMDs was observed leading to a structure more similar to the outward-facing state. A similar behaviour was reported by Weng *et al*[173]. This resulted in an increased time of equilibration, in which the first 150 ns were discarded. This closure was also reflected in pore radius, in which there was a radius increase of the periplasmic side, with a concomitant decrease on the cytoplasmic side (figure 5.16). The full complex C-α RMSD oscillates around 0.48 nm, stabilizing from 150 ns onwards (figure 5.9). On the other hand, MalE shows the lowest C-α RMSD, around 0.19 nm, while the remaining subunits show values centred in 0.3 nm, with some oscillations. The percentage of retained secondary structure is at least 96% (figure 5.10).

5.4.2-Structural effects of hydrolysis and nucleotide exit

5.4.2.1-On nucleotide binding domains

The nucleotide binding domains of every ABC transporter play a key role in controlling the protein function, being responsible for ATP binding and hydrolysis. The NBD dimer behaviour changes upon the presence or absence of nucleotide. Previous experimental data show that ATP binding to the NBD dimer induces the closed form[143,404,406,409]. Our simulations confirm these findings because the pre-hydrolysis state has always shown a closed NBD dimer. On the other hand, the presence of ADP increases the NBDs separation, and the complete absence of nucleotide leads to a further increase[143,147,409]. These motions on the NBDs are reflected throughout the entire complex and drive conformational changes that lead to substrate movement. Therefore, by studying the degree of separation between the NBD monomers, it is possible to assess the evolution of the conformational behaviour of the complex. We defined the active site 1 (AS1) as the catalytic pocket that contains the MalK chain B Walker A motif and the chain A ABC motif. The chain B is coupled to MalF, while chain A is coupled to MalG (figure 5.1a). The active site 2 (AS2) is composed by the chain A Walker A motif and by the chain B ABC motif (figure 5.1b).

Figure 5.2 shows the distance distributions between the Walker A and the ABC motifs (using C-α atoms only) in each pocket, in all states. This measure reflects the opening degree of the ATP binding pocket, allowing to assess the evolution of the conformational changes. Is possible to observe that both pockets show distinct behaviours, regardless of the state. For the active site 1 (AS1), the distributions of ATP and ADP states overlap, while the Apo state shows a much more disperse distribution ranging from 0.8 to 2.6 nm. Regarding the active site 2 (AS2) pocket, an opening increase was observed after hydrolysis as seen by the ATP and ADP distributions. In the Apo state the total distribution is clustered in two peaks, one of them overlapping with the ATP state in the region of 0.8 to 1.2 nm, while the other spans the distances between 1.2 and 1.8 nm. Despite the Apo distribution showing wider distributions, it is important to recall that this state started in a wide open pocket conformation and its time evolution showed that it has tendency to close.

The asymmetry in pocket behaviour can be explained by the different nature of the transmembrane domains MalF and MalG, which interact with the NBDs via the coupling helices. This behaviour is enhanced in our simulations upon hydrolysis. Asymmetry in the behaviour of the nucleotide binding pockets, was also observed in cryo-EM experiments, in the presence of ADP-phosphate[148].

Experimental studies have used EPR to evaluate the opening degree of the NBD dimer according to the nucleotide state [147,405,409]. In most cases, the measurements were done using pairs of residues in the helical domains, in contrast with our measurements using the C-α atoms of the ATP binding motifs. Additionally, these studies use ADP in the post-hydrolysis state rather than ADP.Pi, as we did. Therefore, the results on figure 5.2 are not directly comparable with the experimental results.

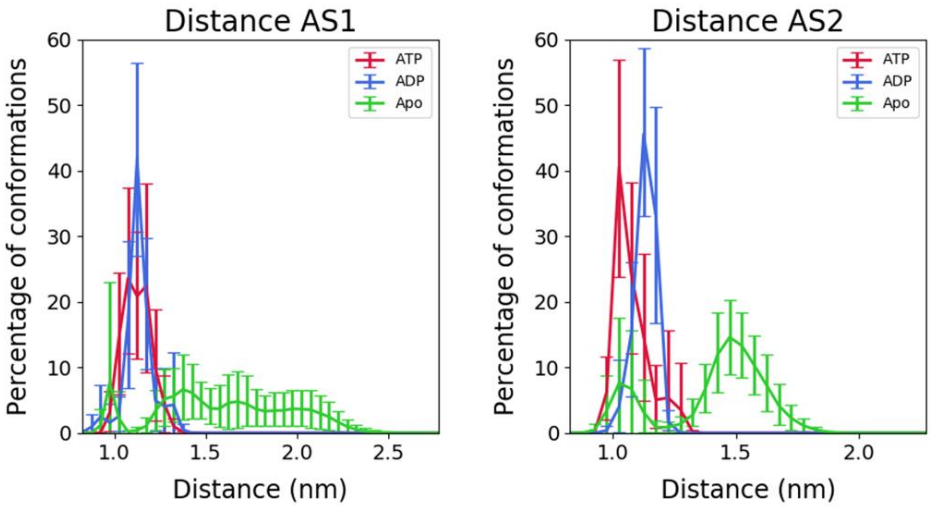


Figure 5.2- *Distance of active site opening in both ATP binding pockets. AS1- Active site 1, AS2-Active site 2. This distance is measured between the C-α atoms of Walker A and the ABC motifs of the active site. The ATP state is represented in red, the ADP in blue and the Apo state without nucleotide in green. The last 260 ns of each trajectory were used for analysis. The error bars correspond to the 95% confidence interval obtained by bootstrapping. The AS1 is formed by the MalK chain B Walker A motif and the chain A ABC motif, while the AS2 is formed by the chain A Walker A motif and by the chain B ABC motif.*

5.4.2.2.-On the transmembrane domains and MalE

In all ABC transporters, ATP hydrolysis is known to cause conformational changes, which result is different outcomes, namely for substrate release or to reset the transporter to the starting point after a transport cycle [57,103]. In the case of the maltose transporter, ATP hydrolysis triggers conformational changes that result in the reorientation of the transmembrane domains, creating an inward-facing state allowing the release of the substrate and MalE disengagement[404,409]. It is hypothesized that the closed, outward-facing form of the maltose importer is a short-lived high-energy catalytic intermediate[70]. In fact, upon MalE binding and NBD closing, ATP hydrolysis is highly stimulated [404–406].

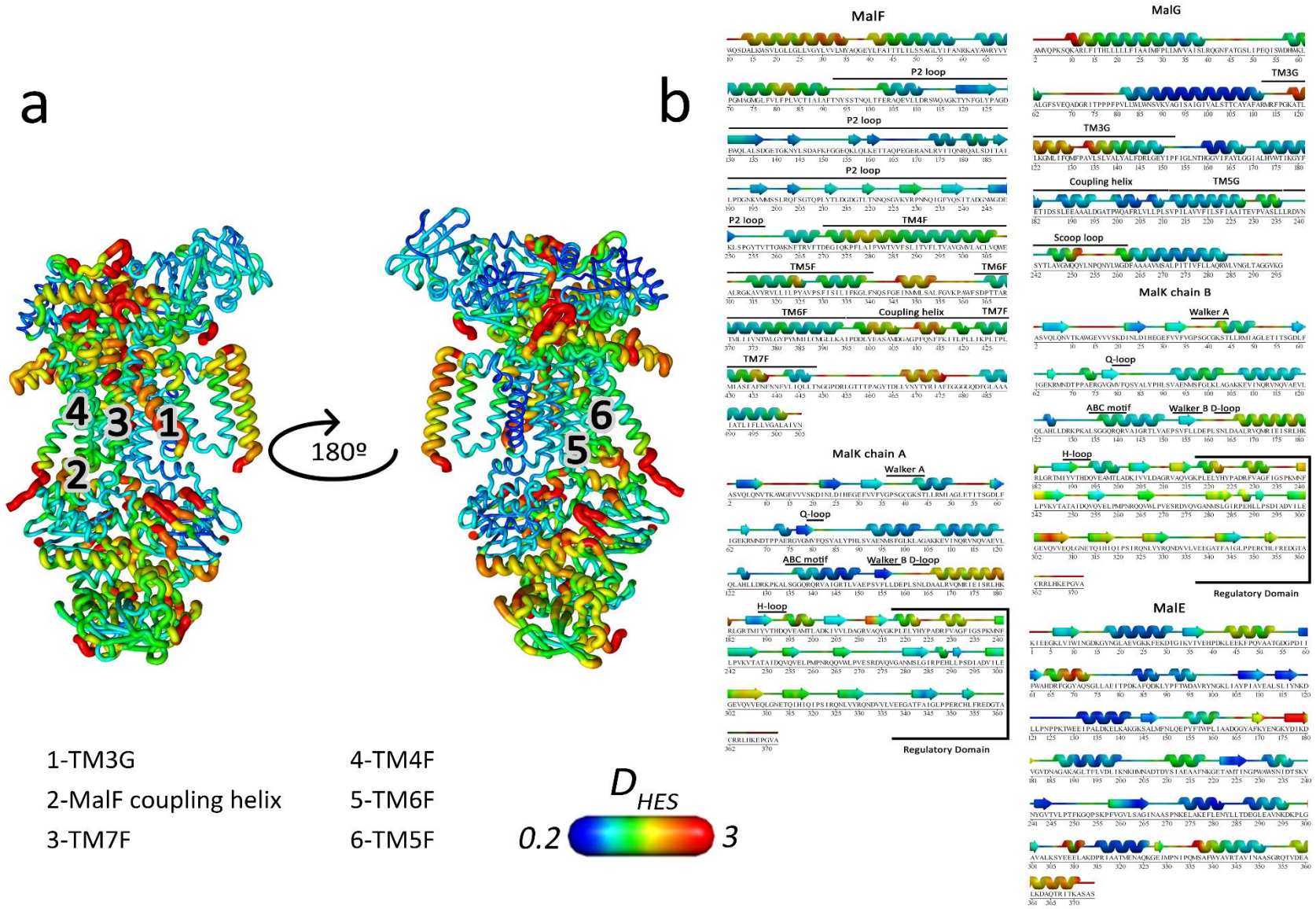

a
b
4
3
1
2
180º
6
5
1-TM3G
2-MalF coupling helix
3-TM7F
4-TM4F
5-TM6F
6-TM5F
D HES
0.2
3
MalF
P2 loop
TM4F
TM5F
TM6F
Coupling helix
TM7F
MalK chain A
Walker A
Q-loop
ABC motif
Walker B
D-loop
H-loop
Regulatory Domain
MalG
TM3G
Coupling helix
TM5G
Scoop loop
MalK chain B
MalE

Figure 5.3- *Mapping of the harmonic ensemble similarity (D_{HES}) between the ATP and ADP states. D_{HES} was calculated for each residue, using the C-α atoms. The last 50 ns of each trajectory were used for this analysis, and the values presented result from the average of cross-replicate comparisons. The scale below indicates the magnitude of the changes in D_{HES} observed, in which the dark blue corresponds to the minimum values observed, indicating increased similarity while the red corresponds to the maximum values observed, indicating increased dissimilarity. (a)- D_{HES} values mapped on the protein structure. The structure in the left represents the "front view" of the transporter, while the one on the right represents the "rear view", accomplished by a 180º rotation. Relevant regions are marked with numbers: 1-Transmembrane helix 3 MalG (TM3G), 2-MalF coupling helix, 3-Transmembrane helix 7 MalF (TM7F), 4- Transmembrane helix 4 MalF (TM4F), 5-Transmembrane helix 6 MalF (TM6F), 6-Transmembrane helix 5 (TM5F). (b)- Mapping of D_{HES} between the ATP and ADP states, on the sequence of each chain. The 3D structure map was obtained with PDBsum[69].*

In order to assess the structural impact of hydrolysis in the full-length complex, the harmonic ensemble similarity (D_{HES}) between ATP and ADP states was calculated for each residue, using the C-α atoms. Figure 5.3 shows D_{HES} calculated using the last 50 ns of the ATP and ADP simulations. ADP hydrolysis has led to significant differences in the distributions sampled by most residues throughout the complex. In the NBD dimer, there are key regions affected by hydrolysis, such as the D-loops, the Walker B motifs as well as the catalytic motifs, the Walker A and ABC sequences (figure 5.3). The α-helical domains also display significant changes, that are transmitted to regions of the regulatory domains nearby (figure 5.3a, marked with the number 2). Nonetheless, the MalK chain B is clearly more affected by ATP hydrolysis, showing higher D_{HES} values. Interestingly, this NBD contacts with the MalF coupling helix that displays significant conformational changes that are transmitted to the neighbouring helices 4, 5, 6 and 7 from MalF (TM4F, TM5F, TM6F and TM7F) (figure 5.3b, numbers 3 to 6 in the figure 5.3a). The P2 loop does not show significant changes. The conformational changes are transmitted to MalG through TM7F reaching transmembrane helix 3 in MalG (TM3G) (figure 5.3b, number 2 in figure 5.3a). This is remarkable since MalG transmembrane helices are fairly rigid in comparison with MalF. Nonetheless, the residues at the end of transmembrane helix 5 of MalG show high D_{HES} values. Furthermore, the MalG scoop loop that plays a role in maltose transfer from MalE to the transmembrane domains also shows high D_{HES} values (figure 5.3b). Other external loops from MalG and MalF that contact with MalE externally also show large variation (figure 5.3a). Regarding MalE, significant differences are observed, not only in the regions in direct contact with the transmembrane domains, but also in the C-lobe (residues 163-187 and 326 to 374 in figure 5.3b).

Two dimensional PCA analysis of the all the simulated states allowed us to obtain two main clusters of basins: one containing only Apo conformations and other with mixed ATP/ADP content (figures 5.17 and 5.18). Therefore, the Apo state can be fully separated from the ATP

and ADP states (figure 5.18). Regarding the cluster of basins belonging to the ATP and ADP states, these two states cannot be distinguished so easily, since they show overlap in some of the basins (figure 5.18), but in different degrees (table 5.3). Nevertheless, it is possible to identify characteristic basins of each state, which show a markedly ATP or ADP character, containing a larger proportion of conformations within that basin (table 5.3). By analysing the structures corresponding to the maximum probability on each of the most characteristic basins, it is possible to get a more detailed view on the conformational changes suffered by the most relevant regions above mentioned. Even though this analysis focuses on limited zones of the whole conformational space, the comparison between the characteristic structures of these two basins allows us to identify what is markedly different between the ATP and ADP states, and their difference from the Apo state. This is what is shown in figure 5.19. The Apo state shows a unique conformation that is clearly distinguishable from the ATP and ADP ones, in which the periplasmic portion of TM7F, moves to the center of the pore causing constriction in the region containing the periplasmic gate (figures 5.19B and 5.19F). On the other hand, TM6F moves away from the pore center (figure 5.19C), while the MalF coupling helix (figure 5.19A) and TM4F adopt a unique position different from the other states (figure 5.19D). Focusing now on the ATP and ADP states, we can see that ATP hydrolysis results on a shift of the MalF coupling helix (figure 5.19A) and TM6F moves towards the NBD direction (figure 5.19C), while TM7F is shifted in the opposite direction and it moves slightly to the exterior of the pore (figure 5.19B). On the other hand, TM4F shifts towards the MalE direction (figure 5.19D). ATP hydrolysis also prompts a displacement of the transmembrane helix 5F (TM5F) towards the periplasmic direction (figure 5.19E). Nonetheless, the most remarkable alterations happen in the bottom of the pore, where there is a major lateral displacement of TM3G (figure 5.19F). The TM5F and TM7F helices also suffer lateral displacements in this region (figure 5.19F), also causing pore constriction, similarly to the Apo state.

Wen *et al* suggested that the NBD-TMD coupling is done by a network of contacts that arises from the ABC motif, goes to the helical domains, followed by the Q-loop, ending on the coupling helices [172]. In fact, our results confirm these observations, by showing considerable differences in the ensembles sampled by the residues in this region.

The active site 2, that shows a larger degree of opening (see figure 5.2), is composed by the chain A Walker A motif and by the chain B ABC motif. Upon hydrolysis, the residues of the ABC motif show higher dissimilarities on the sampled ensembles, than the Walker A residues. Considering that the MalK chain B is connected with MalF and that MalF shows larger D_{HES} on key regions throughout the complex, it is possible to hypothesize a mechanism for signal transmission, from the NBDs to the remaining protein, in which the higher motion of the ABC

motif contributes for spreading the signal across MalF and the remaining complex. This stronger coupling with MalF is also corroborated by cryo-EM structures in which an asymmetric NBD dimer was observed, with MalF shifted from the center of the complex[148]. We have also observed novel conformational changes upon hydrolysis in the MalG subunit, such as the displacement of the transmembrane helix 3 (TM3G) at the periplasmic end of the pore, contributing for pore narrowing in that region.

In this way, ATP hydrolysis in the NBDs impacts the entire protein complex. Conformational changes in the transmembrane helices lead to variations in the pore properties, such as the pore radius and residue conformation.

Figure 5.4a shows the pore radius profile on the last 20 ns of simulation in all the simulated states. Overall, the transmembrane pore is constituted by a pear-shaped cavity, wider at the bottom (near the periplasmic gate) becoming increasingly narrowed as it progresses towards the cytoplasm (figure 5.4b).

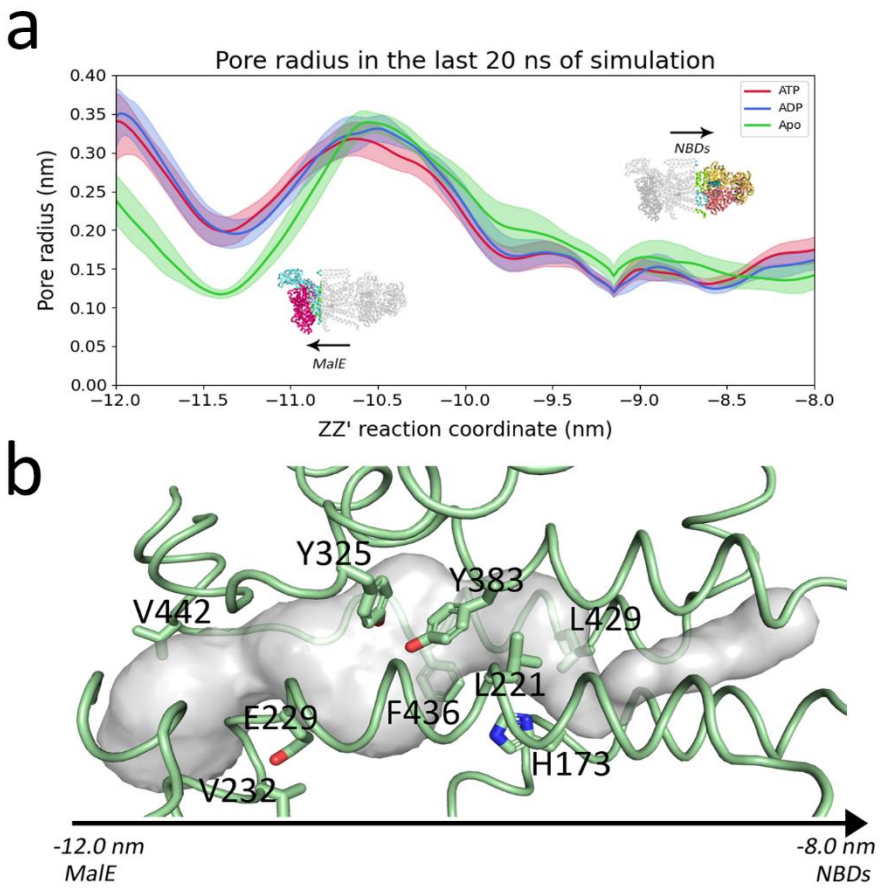


Figure 5.4- *Pore radius across the channel. (a)- Pore radius across the channel in the last 20 ns of simulation. The ATP state is represented in red, while the ADP one in blue. The observed values are the average for each state and the error bars represent the 95% confidence interval obtained by bootstrapping. Larger values of the zz' reaction coordinate indicate a path towards the intracellular medium, while lower values indicate the direction towards MalE and the periplasmic medium. (b)- Pore representation with relevant residues. The pore surface was generated using HOLE[420]. The residues are displayed in sticks, the pore volume as surface and the protein backbone as ribbon. The bottom of the pore heads towards the periplasmic direction.*

The region around -11.5 nm corresponds to the constriction point caused by the periplasmic gate residues (V442 and V232) and it is a local minimum in the ATP and ADP states and a global minimum in the Apo state. Downstream from this region, there is the MalE binding site. The Apo state is rather distinct of the other two states, mostly the part towards the periplasmic gate. These differences reflect the initial conformation of the protein in the Apo state, which has the transmembrane helices much more packed, leading to smaller radius. Despite the great similarity of the ATP and ADP profiles, hydrolysis causes a global change on the radius profile, leading to a further decrease of the radius in the MalE binding site (below -11.5 nm) and in its vicinity up until -11.0 nm. On the other hand, there is a concomitant increase of the radius above this region (figure 5.4). Nonetheless, the most striking difference upon ATP hydrolysis is the

increase of the pore radius that occurs in the -10.8 to -10.2 nm range, but there are still differences up until -9.5 nm, in which the error shading of the ADP state surpasses the ATP state in the upper limit. This region comprises the maltose binding site and it enclosed by E229 in its periplasmic end, and by Y383 and H137 in the cytoplasmic gate region. The aromatic residues Y325 and F436 are contained in this region. The minimum observed at -9.2 corresponds to the cytoplasmic gate residues (L429 and L221) and it is a main constriction point of the pore. The values upstream this cytoplasmic gate reflect the path until reaching the coupling helices. Overall, ATP hydrolysis leads to the compression of the pore near the periplasmic end, confirming the observations of the PCA analysis, in which TM3G, TM7F and TM5F move towards the pore center (figure 5.19F). A concomitant increase of the region containing the binding site (from -10.8 to -10.2 nm region) along with a marked increase of the region spanning the range from -10.2 nm to -9.5 nm, reflects the motions of the MalF coupling helix, TM7F, TM6F and TM5F on this portion of the pore (figure 5.19A, 5.19B, 5.19C and 5.19D).

The last 20 ns of simulation were chosen to represent this property, because it is the timeframe that allows to observe the maximum differences. The similarity of the Apo radius profile with the ATP and ADP profiles in the region towards the NBDs (mainly from -10.5 nm onwards) is due to the sudden closure observed in the beginning of the simulations (figure 5.16). This same behaviour was previously described by Weng *et al*[173], who performed metadynamics simulations on the Apo structure. Nevertheless, it should be noted that the displayed profiles are averages of multiple conformations and may not reflect the conformational diversity of each state.

Despite the several models for MalE action throughout the transport cycle[147,404,405,425,426], it can be speculated that the effect of ATP hydrolysis, along with the progressive conformational changes in MalE, results in an increasingly inward-facing state. This leads to an increase of the pore radius, facilitating substrate diffusion upwards and eventual exchange with the nucleotides. In fact, EPR data confirm that the nucleotide state in MalK deeply influences the coupling of P2-loop, with MalE eventually altering its conformation[426], and that in the presence of ADP, the MalE N-lobe adopts an unique conformation[147], while the MalE C-lobe is more disordered. This can potentially shape the TM helices into the inward-facing state in a concerted action with the NBDs. In addition, there are contradictory experimental data concerning the affinity of the open and closed MalE to the transmembrane complex [147,404].

Nonetheless, with MalE always bound, the unidirectionality of the transport is safeguarded. It may also be possible that ADP or phosphate exit are required to further enhance these transformations.

Regarding the behaviour of the pore residues, the most noteworthy residue is Y383, not only because of its position, located below the cytoplasmic gate, but also because its conformation

changes drastically upon the nucleotide state of the protein, as shown in figure 5.5. When measuring the χ_1 tyrosine dihedral, it is possible to differentiate four populations (figure 5.5a): a first population with χ_1 values between -180º and -150º. This population is most abundant before hydrolysis and corresponds to conformation in which the hydroxyphenyl ring adopts an orientation perpendicular to the pore axis, preventing maltose diffusion from this point upwards as seen in figure 5.5b. Additionally, the last population with χ_1 values between 150º and 180º also displays a similar behaviour and it is also most abundant on the ATP state. Upon ATP hydrolysis two other populations arise: one with χ_1 values between -100º and -50º, most abundant in the ADP state and other with χ_1 values between 50º and 100º, most abundant in the Apo, followed by the ADP state. This last population is more energetically accessible than the previous one, as seen by the higher percentage of conformations that occupies the same space, being even accessible by the ATP state. In these two populations Y383 adopts a conformation parallel to the pore axis, with the sidechain being away from the pore centre. This behaviour may be a consequence of the conformational changes described above, since Y383 is located in TM6F of MalF, which is one of the most affected helices, as seen before. Mutation of Y383 to a serine residue led to lower growth rates using maltose, but not maltoheptatose, leading to a possible change of the substrate specificity[427]. Therefore, Y383 is a key residue that controls substrate efflux and involved in substrate specificity. In this way, it can be considered a first gatekeeper before the periplasmic gate.

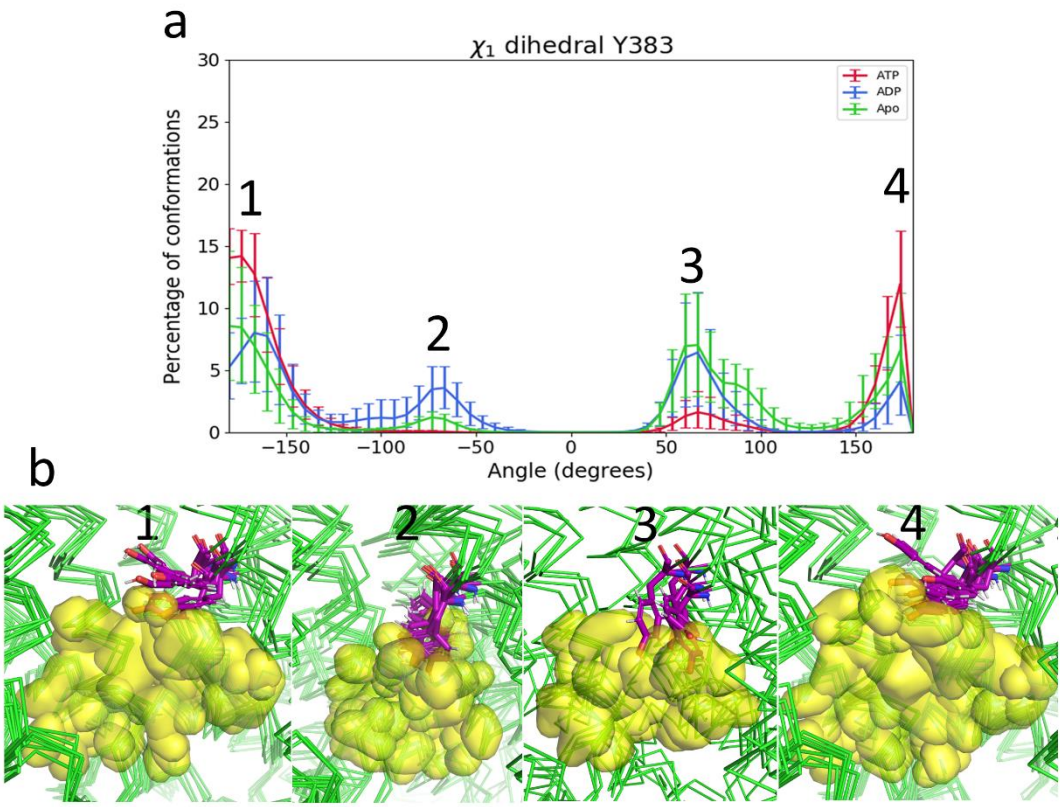


***Figure 5.5**-Behaviour of χ_1 dihedral values for Y383 for all states. (a)-Average χ_1 dihedral values for Y383 in each state. The χ_1 reflects the rotation of the first bond between the C-α atom and the aryl ring. ATP is represented in red, ADP in blue and the Apo state in green. The main populations observed are indicated with numbers from 1 to 4. The error bars represent the 95% confidence interval obtained by bootstrapping. (b)-Examples of representative conformations for each main population observed. The protein is coloured in green, and Y383 is represented in purple sticks and according to the atom type, while maltose as a density surface in yellow.*

5.4.2.3-Effect of hydrolysis and nucleotide exit on maltose diffusion and binding

As a consequence of the helical displacement upon hydrolysis and nucleotide exit, the pore volume is increased in some regions, which culminates in increased maltose vertical diffusion. Figure 5.6 shows the diffusion maps of the maltose center-of-mass for the three states. It is possible to observe two main binding sites. Prior to hydrolysis it is possible to observe that a large portion of the density is in the vicinity of MalG (at the left side of the picture). This binding spot is constituted by hydrophilic residues, such as the E229, Y166, S135, H173 and N129, along with hydrophilic mainchain groups. In addition to this spot, a significant amount of density is observed in the vicinity of MalF. In this spot, maltose is surrounded by a cluster of aromatic residues, formed by F436, Y325, Y383, along with sidechains from polar residues, mainly asparagine and serine residues. These interactions are constant in all the simulated states

(figures 5.20 and 5.21) and are also present in crystal structures that were the starting points of this work. The importance of these hydrophobic interactions is assessed by the fact that mutations in F436 and Y383 reduce transport activity by over 90% in comparison to the wild-type, and the mutation Y325S retains only 52% of the transport activity [427,428]. Figure 5.6 shows that this pocket in MalF becomes increasingly populated and expands upwards after hydrolysis and nucleotide exit. The pre-hydrolysis and pre-translocation X-ray structures show maltose bound to this pocket in MalF. However, our simulations revealed the existence of an additional pocket in MalG in the pre-hydrolysis state, with a more reduced occupancy in the post-hydrolysis and nucleotide-free states.

The existence of this additional pocket is corroborated by the crystal structure of MalFGK$_2$E with maltoheptatose bound in which at least one of the glucosyl units is pending towards MalG (in this structure, the other unit lacks electron density)[429].

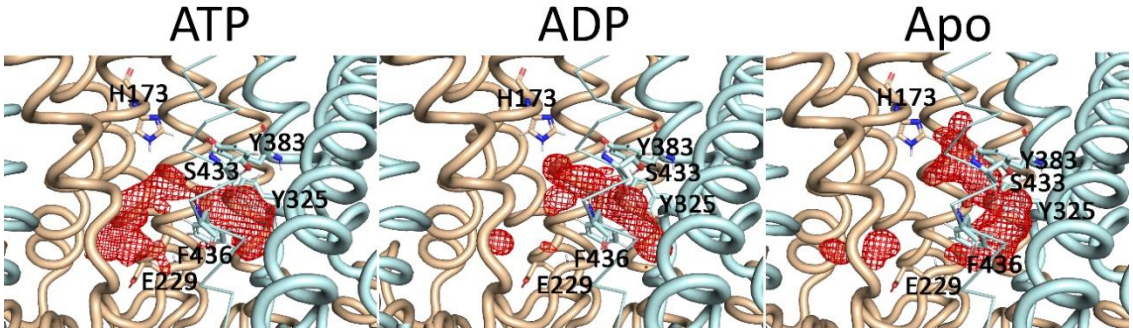


***Figure 5.6**-Maltose binding in the transmembrane domains. The isosurface was build using the center-of-mass positions of maltose. MalG is represented in beige, while MalF in light blue. Key residues that interact with maltose are represented in sticks. The red mesh represents the distribution of the maltose center-of-mass.*

Maltose seems to make a similar number of hydrogen bonds with MalF and MalG in either of the ATP and ADP states, displaying a slight preference for MalF in the Apo state (figure 5.22). The residues which make the most hydrogen bonds belong to MalF and are located within and nearby the cluster of aromatic residues previously mentioned - Y383, Y325, N376 and S433, along with H173 that belongs to MalG and is located upstream to this spot (figure 5.6 and figure 5.20). Maltose also interacts with E229 (figure 5.22), which is part of MalG and is located in the vicinity of the scoop loop towards the periplasmic side. Interestingly, this interaction is highly dependent of the nucleotide state, being most frequent in the ATP state, followed by the ADP and even lower in the absence of nucleotide. The movement of the transmembrane helices, which increases the pore volume and vertical diffusion, can explain these differences, namely changes in helices TM4F, TM5F, TM6F, TM3G and TM5G, which contain key residues that

influence substrate transport such as the periplasmic gate residues V442 and V232, the cytoplasmic gate residues L429 and L221, and Y383. Mutagenesis studies show that deletion of E229 leads to the failure in complex assembly and maltose transport[430]. It is possible to hypothesize that it may facilitate maltose diffusion from MalE to the transmembrane domains, or even prevent maltose diffusion in the opposite direction towards MalE.

Nonetheless, the most frequent interactions performed by maltose are with water molecules, consequence of the high solvation of the pore (figure 5.23).

5.4.3-Effect of hydrolysis and nucleotide exit on maltose translocation

The effect of ATP hydrolysis and nucleotide absence on maltose translocation was studied using both pulling and umbrella sampling MD simulations. Conformations were extracted from the equilibrium MD simulations at different times and pulling was performed in order to generate multiple maltose conformations across the channel. The initial positions of maltose used in each window are described in tables 5.4 to 5.6. Maltose was in various initial positions across the channel in the different frames extracted. Figure 5.7 shows the potential of mean force (PMF) profiles obtained for the several states. The lower reaction coordinate values point towards the periplasmic side of the protein, where MalE is bound, while the higher values point towards the cytoplasmic side, where the NBDs are located. The histograms used to make the PMF profiles, along with the convergence tests can be found in figure 5.24 to 5.32. The maltose conformations sampled spanned the lowest point possible towards the periplasmic side, up until reaching the vicinity of the NBDs. The high free-energy values, from 20 to 80 kJ/mol, possibly result not only from structural constraints, but also from instabilities generated by the pulling process, resulting in an artificial biasing of the energy values. We also recognize that our sampling is not perfect leading to an additional roughness of the profiles. Therefore, all the conclusions deducted henceforth will be of a strictly qualitative nature.

Overall, the profiles span around 3 nm, from -12 nm to ≈8.3 nm. The profiles show a local maximum around -12 nm to -11.5 nm, followed a global minimum in the region of -11 nm to -10 nm. This is followed by an increase of the free-energy towards the cytoplasmic side, until reaching the NBD level. The behaviour of these PMF profiles is inversely correlated with the pore radius behaviour, across states. Nonetheless, they are not tightly correlated, because the pore radius represented in figure 5.4a, is averaged over the simulation time, while the PMF profiles were obtained by extracting individual conformations from the MD simulations, upon which pulling, and umbrella sampling simulations were performed. Figure 5.33 shows the

average pore radius obtained in umbrella sampling simulations and it is more correlated with the PMF profiles.

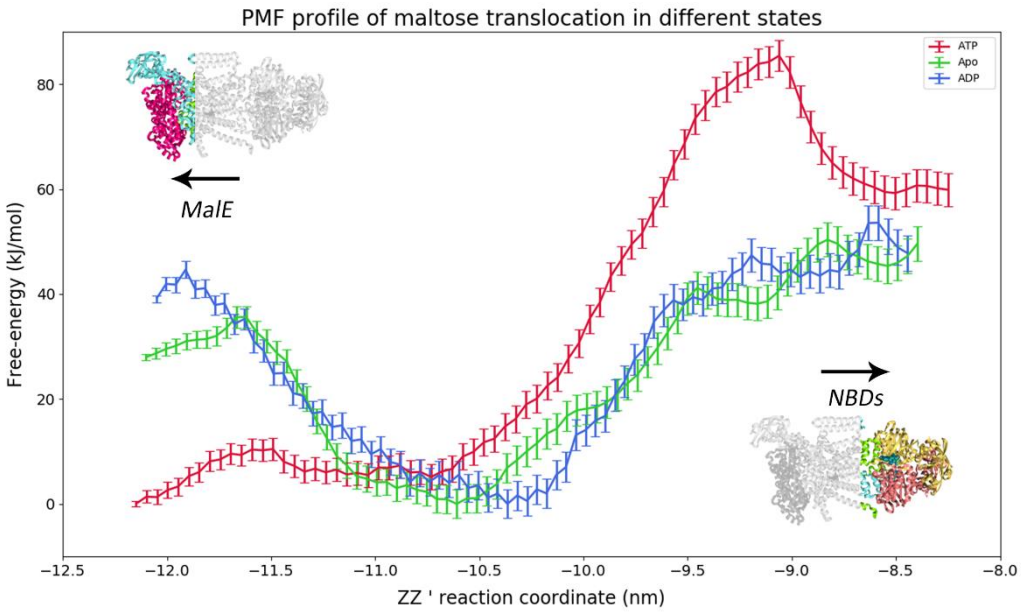


***Figure 5.7**- Potential of mean force (PMF) profiles for the ATP, ADP and Apo states. The ATP profile is represented in red, the ADP profile in dark blue and the Apo profile in green. The error bars correspond to a 95% confidence interval at each point, assuming that the energy values in each bin follow a normal distribution.*

The ATP PMF profile shows a local maximum at around -11.5 nm, corresponding narrowing of the pore in between the periplasmic gate valine residues: V442 and V232. At this point, maltose is in a predominantly vertical conformation, as calculated by the angle between the z-axis and the vector that links the most distant carbon atoms in the maltose rings (figure 5.34). The maltose conformation contributes to account for the high free-energy values near the periplasmic gate region. This region of the pore is considerably tighter as seen by the low pore radius shown in figure 5.4. Therefore, when maltose is in a horizontal conformation, its presence will not be energetically favourable in that portion of the channel. The vertical conformation will decrease the steric clashes with the protein residues. From -11.5 nm and under, maltose is below the periplasmic gate, interacting with the scoop loop from MalG, as well as some MalE residues, such as K46. The fact that maltose is able to adopt a horizontal conformation, as seen by the highly populated bins between 80º and 120º in Figure 5.35 (ATP state) is a consequence of the larger pore radius observed in this region. This leads to lower free-energy values prior to the local maximum. This local maximum is located downstream at -11.5 nm in the ADP and Apo profiles. This is a consequence of the different conformations assumed by the protein in the different states.

In fact, in all the simulated states, maltose tends to adopt a horizontal conformation near the periplasmic gate, being later reoriented to cross the periplasmic gate (figure 5.35). This location corresponds to a local maximum in all PMF profiles. Interestingly, this energy barrier is lower prior to ATP hydrolysis, suggesting increased diffusion probability towards MalE. Therefore, ATP hydrolysis enhances the irreversibility of the transport process, considering that the closure of the transmembrane helices on the periplasmic side decreases the probability of diffusion of maltose in this direction, while having the opposite effect in cytoplasmic direction. When maltose is between the periplasmic gate residues, it starts to interact with the E229 as well as other MalF residues, such as S329 and N440. Nonetheless, the visualization of the trajectories shows that E229 plays a key role in redirecting and reorienting maltose towards the maltose binding site on MalF, which constitutes the energy minimum observed in all the profiles. This binding site is the same described above and the one observed experimentally in the X-ray structures.

The global minimum observed in the PMF profiles, around -10.5 nm corresponds to the maltose binding site. The maltose binding site in MalF is delimited by three aromatic residues: Y383, Y325 and F436 as seen in Figure 5.8. Maltose establishes hydrophobic interactions with these residues as well as hydrogen bonds with the tyrosine hydroxyl groups. The residues in the pocket change conformation upon ATP hydrolysis and nucleotide exit (figure 5.8).

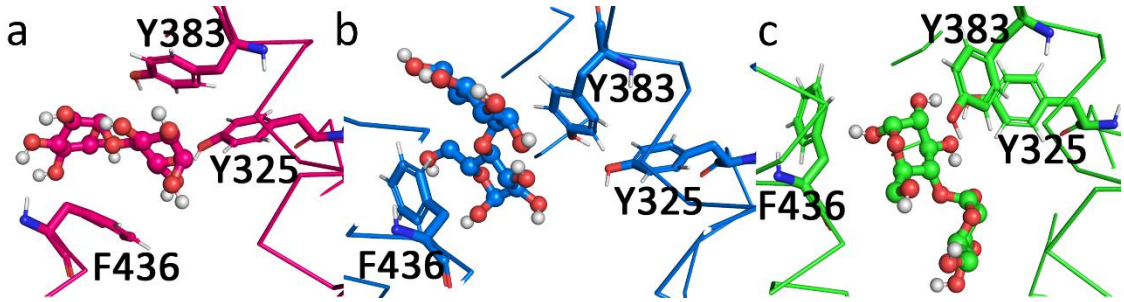


Figure 5.8- *Maltose binding pocket with maltose bound, surrounded by key residues. (a)-Maltose binding pocket in the ATP state, maltose is represented in sticks and spheres, while the residues are portrayed in sticks only. (b)-Maltose binding pocket in the ADP state, maltose is represented in sticks and spheres, while the residues are portrayed in sticks only. (c)- Maltose binding pocket in the Apo state, represented in blue and green, respectively. The relevant aminoacids are represented in sticks, while maltose is represented in spheres and sticks.*

The PMF profiles confirm the importance of Y383 as a gatekeeper. By changing its conformation upon hydrolysis, allows maltose to adopt a vertical conformation and removes any vertical constraints to diffusion, while the remaining interactions become weaker, such as the case of F436 and Y325 (figures 5.8b and 5.8c). This vertical conformation seems to be associated with lower energies, as a result of lower steric strains (figure 5.35), being able to cross

the pore more easily. The predominant conformation of maltose in the Apo and ADP profiles is in a vertical orientation (aligned with the pore axis) (figures 5.34 and 5.35).

E229 not only guides maltose towards the hydrophobic pocket, but it also guides maltose out of this pocket, as maltose leaves it by making hydrogen bonds with this acidic group. In the region of -9.5 nm to -9 nm, it is possible to find an energy maximum which corresponds to the cytoplasmic gate, and this maximum results not only from the presence of the bulky leucine residues, but also because it is the point of maximum constriction in the pore, i.e, minimal radius (figure 5.33). Closer to the cytoplasmic gate, H173 is also key for ensuring a proper vertical orientation, from the pore center up to the vicinity of the cytoplasmic gate. There are also additional interactions with T176. The vertical conformation of maltose allows it to cross the cytoplasmic gate and diffuse in a rectilinear way towards the NBD vicinity.

In contrast, in the ATP state, maltose reaches the cytoplasmic gate in a horizontal conformation. Due to the low radius in this region and above, maltose has the tendency to diffuse laterally towards the direction of TM5F and TM6F, reaching the MalF coupling helix. Interestingly, there is a cluster of hydrophobic residues in this region, mainly valine and leucine residues, creating a highly unfavourable environment for the residence of maltose in this place, resulting in a peak of free-energy from -9.2 nm to -9 nm, approximately. The subsequent decrease corresponds to the maltose diffusion towards the NBDs and to the middle of the pore.

After the cytoplasmic gate, the region between -9 and -8.5 nm is lined by hydrophobic residues, namely isoleucine, leucine, valine and tyrosine residues. Maltose is still able to perform a few hydrogen bonds with backbone atoms and a few polar residues (Y180, T176, T290), but the hydrophobic character of this region may well contribute to accelerate maltose exit towards the cytoplasm. The ADP profile also shows a local maximum at -8.5 nm that corresponds to an horizonal conformation of maltose, at the end of the TMDs and beneath the NBD dimer. All the PMF profiles end in the terminus of the TMDs, beneath the NBDs.

The ADP and Apo PMF profiles show substantially lower energies from the maltose binding pocket onwards, reflecting all conformational changes on the membrane helices and key residues, such as Y383, which facilitate translocation from that point onwards.

Interestingly, despite the pore constriction in the direction towards the NBDs, maltose is highly solvated, but the solvation degree slightly decreases upon hydrolysis and nucleotide exit (figure 5.35). Most hydrogen bonds at the bottom of the protein pore are performed with MalF, and as maltose diffuses upwards, the number of these interactions with MalF decreases. In the ATP state, this is accompanied by a decrease in the interactions with MalG, but this is compensated by a higher amount of hydrogen bonds with water molecules. On the other hand, in the ADP and Apo states there is an increase of the hydrogen bonds with MalG, while the solvation sphere

of maltose is smaller. As expected, all interactions decrease when reaching the cytoplasmic gate, due to the constriction in this region. In fact, the rise of the free-energy along the reaction coordinate is highly correlated with the progressive decrease of the pore radius (figure 5.4 and 5.33).

The free-energy differences observed in the PMF profiles are correlated with the pore radius (figure 5.7 and figure 5.33). In the periplasmic side of the profile (between -12.0 and -11.0 nm), the pore radius decreases upon hydrolysis and further decreases upon nucleotide exit, which coincides with an increase on free-energy observed in PMF profiles, meaning that the maltose has little tendency to remain in this portion of the pore. This is a consequence of the conformational changes observed in the transmembrane helices TM7F, TM5F, TM6F and TM3G that lead to the constriction of this region (figures 5.3 and 5.19). Nonetheless, in the binding site region the free-energy of translocation is similar in both ADP and Apo states, despite the ADP pore radius being similar to the ATP. We attribute this difference to the change in maltose orientation to the vertical orientation (figure 5.33). From the maltose binding site towards the NBDs, the free-energy of translocation becomes similar in both ADP and Apo states, reflecting the similarity in the pore radii of both states (figure 5.35). These alterations cause the constriction of the pore towards MalE while enlarging the pore in the opposite direction. The importance of the conformational changes in the transmembrane helices becomes evident since these key residues are located in the above referred helices. Y383 is located in TM7F, while F436 and L429 (cytoplasmic gate) and V442 (periplasmic gate) are located in TM6F. The residue V232 from the periplasmic gate belongs to TM5G, which was also affected by hydrolysis.

The PMF profiles, along with the structural information obtained in the umbrella sampling simulations, provide important clues on the transport mechanism. It is possible to observe that upon ATP hydrolysis and posterior nucleotide exit, the transport process becomes increasingly favourable from the energetic viewpoint. Nonetheless, the Apo PMF profile is quite similar to the ADP one, suggesting that hydrolysis is the main drive to lower the translocation energy, rather than nucleotide release, as seen by the similarity of PMFs along the reaction coordinate, but mainly at the top near the NBDs. In fact, this goes in agreement with the Cryo-EM data that shows that NBDs, in the absence of nucleotide, are mainly in a semi-closed state, while maltose and MalE are still present in the complex, suggesting that maltose exit is one of the last steps before transporter reset[11]. Further kinetic data[12] shows that phosphate release is the limiting step of the transport cycle and ADP exit can enhance NBD opening. The authors also suggest that the release of Pi is accelerated by the presence of maltose. Therefore, it is possible to assume that phosphate is released prior to maltose.

This decrease in free-energy upon hydrolysis and nucleotide release happens due to conformational changes triggered by both processes in the transmembrane helices. The closure of transmembrane cavity starting from the periplasmic side, with the concomitant opening of the pore on the cytoplasmic side, happens as a result of the motions in transmembrane helices TM7F, TM5F, TM6F and TM3G (figure 5.3). These concerted motions lead to an increase of the free-energy values in the periplasmic region (-12 to -10.5 nm) with the simultaneous decrease on the cytoplasmic side (from -10.5 nm onwards). This process somewhat resembles the peristaltic transport mechanism suggested for the type II importer BtuCD, in which the sequential closing of the periplasmic gate, followed by the opening of the cytoplasmic gate, creates a peristaltic movement, allowing substrate translocation[55]. However, the structures of type II importers are substantially more rigid, and the existing mechanisms reject allosteric coupling between the NBDs and SBP, which has already been proven to happen in the maltose importer by a plethora of experimental data[147,404–406,426].

Considering that our results show similar energetic behaviours for substrate translocation with ADP and in the absence of nucleotide, it is possible to question whether MalE has any influence on the transport process. Data from X-ray structures[143,429], EPR[147,405] and cryo-EM[148] experiments support the hypothesis that MalE remains bound throughout the entire conformational cycle. Furthermore, it is also hypothesized that MalE may have a role in the irreversibility of the transport, by stimulating diffusion towards the cytoplasmic direction, while decreasing towards the periplasmic one. Our simulations did not show any significant conformational changes that could indicate MalE unbinding. Additionally, different transport models[404] defend that MalE is bound in different binding states – open or closed throughout the cycle. Previous reports using cryo-EM structures of MalFGK$_2$E[148], obtained with maltose and MalE bound, showed that in 32% of the particles, MalE detached, raising the possibility that MalE unbinding precedes maltose release. Therefore, it is possible that the weakening of interactions of MalE with the membrane complex results in the increase of NBD opening, further decreasing the free-energy barrier, facilitating maltose exit.

Based in this information above presented, we suggest a possible model for MalFGK$_2$E function, in which ATP hydrolysis leads to NBD opening causing rearrangements of the transmembrane helices stimulating vertical diffusion. Yet, further diffusion towards the cytoplasm and eventual substrate release might require the rearrangement of MalE, resulting in the weakening of MalE interactions with the transmembrane complex and increasing the probability of diffusion outside the complex.

The mechanisms unveiled for the MalFGK$_2$E importer slightly resemble other sugar transporters, such as the ones that belong to the major facilitator superfamily (MFS). In a similar

way to ABC transporters, MFS proteins alternate between the inward and outward facing conformations, but not powered by ATP hydrolysis[431]. In the GLUT1 transporter, the conversion from the outward to the inward-facing state leads to rearrangements of the transmembrane helices and other loops. The pore closes from the extracellular side, leading to the diffusion of glucose from the sugar binding site to an intracellular gate. The resulting PMF of translocation shows an uphill behaviour towards the exit with a local minimum corresponding to a transient sugar binding site constituted by hydrophobic residues[432]. The GLUT4 transporter also possesses aromatic residues and a glucose binding site that aid in the reorientation and positioning of glucose in the pore, in order to interact with the appropriate polar groups. The main energetic barriers were related with hydrogen bond breaking/formation upon entry and exit, with small oscillations along the channel, mainly related with reorientation[433].

When comparing the MalFGK$_2$E importer mechanism with the sugar channel LamB located in the *E.coli* outer membrane, responsible for maltose intake to the periplasmic space, significant functional differences arise. In LamB, translocation is driven, not only by molecular interactions with hydrophobic residues (the so-called “greasy slide”), but also by hydrogen bonds with polar residues. The asymmetric distribution of residues in the pore also contributes to enhance substrate entry from the extracellular side[434], which allows substrate diffusion in a screw-like manner, interacting with the hydrophobic residues and polar residues, in which the latter compensate for the dehydration process [435]. In this way, it is possible to have a passive translocation process with little conformational changes in the protein. Translocation studies show a PMF profile with the main barriers located at the entry and exit of the channel that correspond to forming and breaking hydrogen bonds with the protein and the solvent[436]. In contrast, our simulations show that in the MalFGK$_2$E maltose importer, ATP hydrolysis triggers motion of the helices in order to stimulate upwards diffusion, but maltose is still reasonably solvated and seldom interacts with the hydrophobic residues along the pore.

5.5-Conclusions

In this work, we have performed molecular dynamics simulations of the *E.coli* MalFGK$_2$E importer in the pre and post hydrolysis state, as well as in the absence of nucleotide, with the goal of assessing the effects of ATP hydrolysis and nucleotide exit on the translocation of maltose.

We concluded that ATP hydrolysis triggers a series of conformational changes in the protein complex, starting by pocket opening of the active site, and spreading these conformational changes through the transmembrane domains reaching MalE. Hydrolysis affects critical transmembrane helices that conduct to pore constriction at the periplasmic side, while enlarging

the regions towards the cytoplasmic side, leading to increased maltose diffusion towards the NBDs. Y383 was identified as a novel gatekeeper prior to the periplasmic gate, and key to allow substrate diffusion. Additionally, a novel binding spot in MalG was found, with maximum occupation in the ATP state. The post-hydrolysis state showed similar properties to the state without nucleotides.

The PMF profiles of maltose translocation show that hydrolysis significantly lowers the energetic barriers of substrate diffusion towards the intracellular medium, with a concomitant increase of the energy in the opposite direction. Therefore, ATP hydrolysis considerably contributes for the irreversibility of the transport process.

Our data suggest that the maltose binding pocket may play a role in substrate reorientation from the periplasmic side onto the cytoplasmic side. Other key residues, such as E229, Y383 and H173 assist this task.

Interestingly, nucleotide exit does not lead to significant differences in maltose permeation free-energy profile, when comparing with the ADP state. In this way, it might be necessary a weakening in the MalE interaction with the rest of the complex in order to increase NBD separation and further TMDs transformations, leading to a further decrease of the energetic barriers. Further investigations are required to fully understand this phenomenon.

5.6-Acknowledgements

The authors would like to thank Prof. Syma Khalid, Dr. Firdaus Samsudin, Dr. Megan O'Mara and Hugo Macdermott-Opeskin for helpful discussions. This work was funded by Project LISBOA-01-0145-FEDER-007660 (Microbiologia Molecular, Estrutural e Celular) funded by FEDER funds through COMPETE2020—Programa Operacional Competitividade e Internacionalização (POCI) and by national funds through FCT— Fundação para a Ciência e a Tecnologia. B.A. acknowledges funding via a PhD scholarship (SFRH/BD/108002/2015) from FCT and funding from ITQB-NOVA via the grant INCENTIVO/EQB/LA0004/2014 ref. 010/BIS/2015. The authors also acknowledge the PRACE for awarding access to the MareNostrum cluster at Barcelona Supercomputing Center (BSC), Spain and to Piz Daint at CSCS, Switzerland for benchmarking and testing purposes.

5.7- Supplementary information

Table 5.1-*Protonation states of the histidine residues for the ADP and ATP states.*

MalE		
Residue	**Protonation**	**Charge**
H39	Protonated in both Nδ and Nε	+1
H64	Protonated in Nε	0
H203	Protonated in Nδ	0
MalG		
Residue	**Protonation**	**Charge**
H17	Protonated in Nε	0
H58	Protonated in both Nδ and Nε	+1
H159	Protonated in Nε	0
H173	Protonated in Nε	0
MalK		
Residue	**Protonation**	**Charge**
H27	Protonated in both Nδ and Nε	+1
H89	Protonated in Nε	0
H125	Protonated in Nε	0
H180	Protonated in Nε	0
H192	Protonated in both Nδ and Nε	+1
H223	Protonated in Nε	0
H289	Protonated in Nε	0
H317	Protonated in Nε	0
H353	Protonated in Nε	0
H366	Protonated in both Nδ and Nε	+1

Table 5.2- *Protonation states of the histidine residues for the Apo state.*

MalE		
Residue	**Protonation**	**Charge**
H39	Protonated in both Nδ and Nε	+1
H64	Protonated in Nε	0
H203	Protonated in Nδ	0
MalG		
Residue	**Protonation**	**Charge**
H17	Protonated in Nε	0
H58	Protonated in both Nδ and Nε	+1
H159	Protonated in Nε	0
H173	Protonated in Nε	0
MalK		
Residue	**Protonation**	**Charge**
H27	Protonated in both Nδ and Nε	+1
H89	Protonated in Nε	0
H125	Protonated in Nε	0
H180	Protonated in Nε	0
H192	Protonated in Nε	0
H223	Protonated in Nε	0
H289	Protonated in Nε	0
H317	Protonated in Nε	0
H353	Protonated in Nε	0
H366	Protonated in both Nδ and Nε	+1

Conformational drift of the equilibrium MD simulations

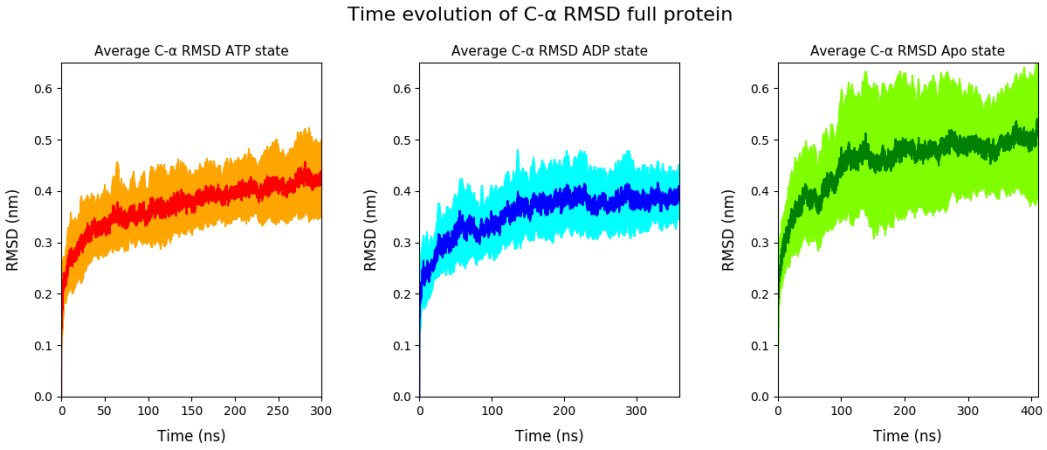


***Figure 5.9**- Temporal evolution of the average C-α RMSD for full MalFGK$_2$E complex, in the three simulated states: the ATP (pre-hydrolysis) state, the ADP (post-hydrolysis) state and the Apo (nucleotide-free) state. Each state comprises nine replicates. The C-α RMSD was calculated against the initial structure of each simulation. The error bars correspond to the standard deviation obtained by bootstrapping. The error bars are represented in orange, cyan and light green in the ATP, ADP and Apo states respectively. The length of the simulations is different in the three states, to ensure equilibrated simulations: the ATP state was simulated for 300 ns, the ADP for 360 ns and the Apo for 410 ns.*

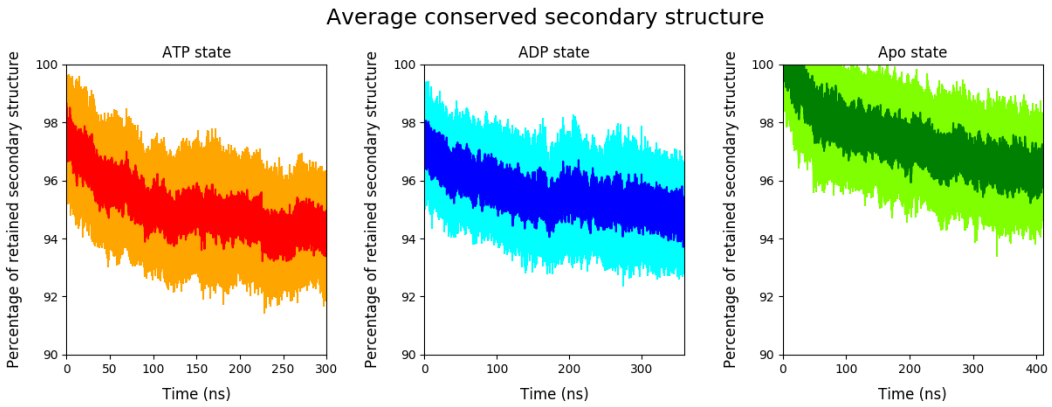


Figure 5.10- *Percentage of retained secondary structure for all the simulated states. Each state comprises nine replicates. The error bars correspond to the standard deviation obtained by bootstrapping. The error bars are represented in orange, cyan and light green in the ATP, ADP and Apo states respectively. The length of the simulations is different in the three states, to ensure equilibrated simulations: the ATP state was simulated for 300 ns, the ADP for 360 ns and the Apo for 410 ns.*

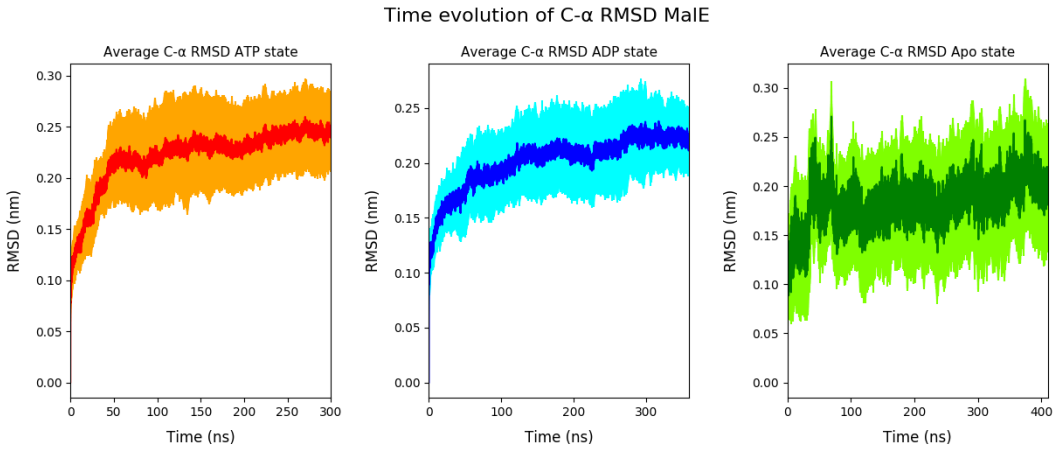


Figure 5.11- *Temporal evolution of the average C-α RMSD for MalE, in the three simulated states: the ATP (pre-hydrolysis) state, the ADP (post-hydrolysis) state and the Apo (nucleotide-free) state. Each state comprises nine replicates. The C- α RMSD was calculated against the initial structure of each simulation. The error bars correspond to the standard deviation obtained by bootstrapping. The error bars are represented in orange, cyan and light green in the ATP, ADP and Apo states respectively. The length of the simulations is different in the three states, to ensure equilibrated simulations: the ATP state was simulated for 300 ns, the ADP for 360 ns and the Apo for 410 ns.*

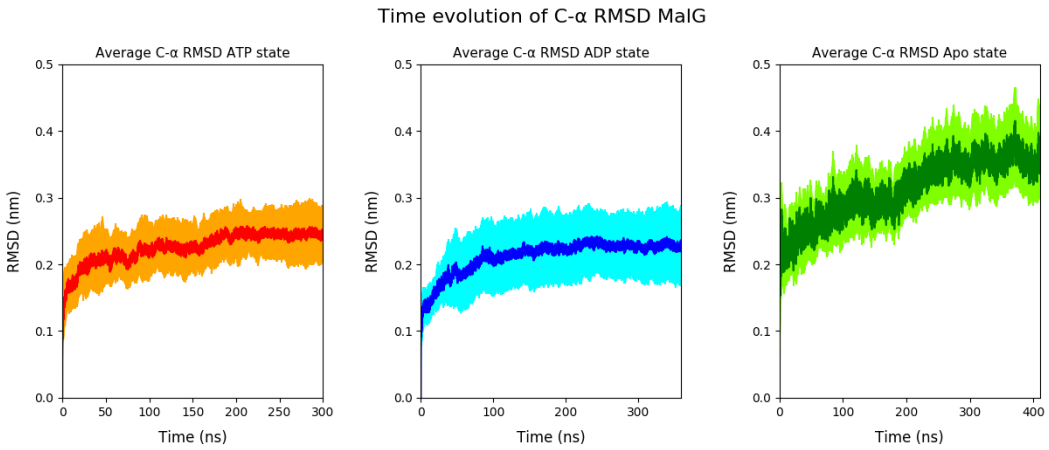


Figure 5.12- *Temporal evolution of the average C-α RMSD for MalG, in the three simulated states: the ATP (pre-hydrolysis) state, the ADP (post-hydrolysis) state and the Apo (nucleotide-free) state. Each state comprises nine replicates. The C-α RMSD was calculated against the initial structure of each simulation. The error bars correspond to the standard deviation obtained by bootstrapping. The error bars are represented in orange, cyan and light green in the ATP, ADP and Apo states respectively. The length of the simulations is different in the three states, to ensure equilibrated simulations: the ATP state was simulated for 300 ns, the ADP for 360 ns and the Apo for 410 ns.*

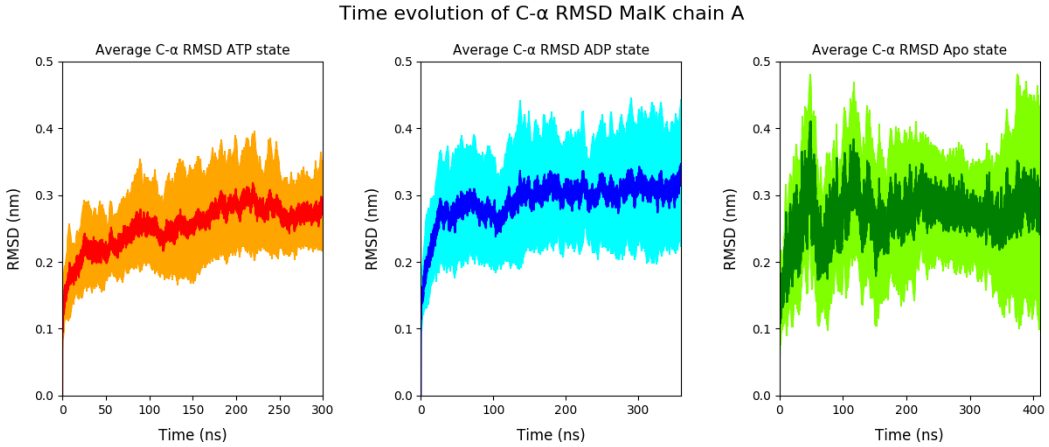


Figure 5.13- *Temporal evolution of the average C-α RMSD for MalK chain A, in the three simulated states: the ATP (pre-hydrolysis) state, the ADP (post-hydrolysis) state and the Apo (nucleotide-free) state. Each state comprises nine replicates. The C-α RMSD was calculated against the initial structure of each simulation. The error bars correspond to the standard deviation obtained by bootstrapping. The error bars are represented in orange, cyan and light green in the ATP, ADP and Apo states respectively. The length of the simulations is different in the three states, to ensure equilibrated simulations: the ATP state was simulated for 300 ns, the ADP for 360 ns and the Apo for 410 ns.*

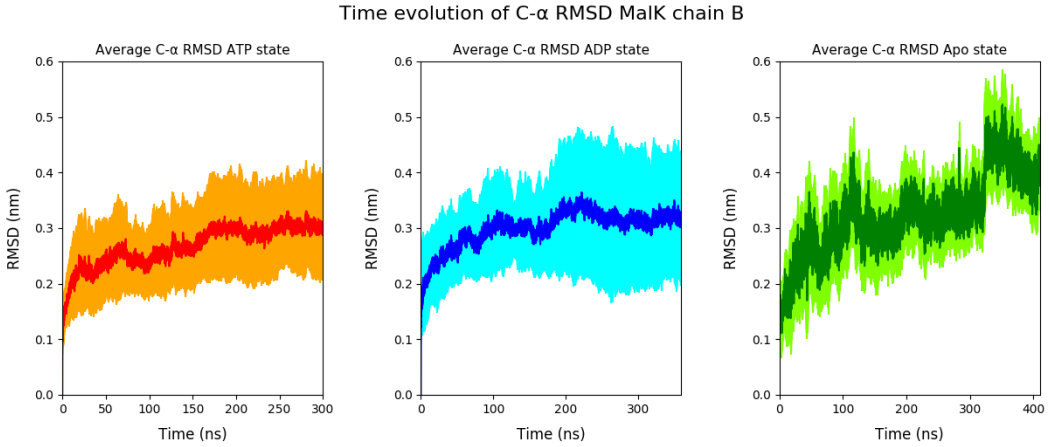


Figure 5.14- *Temporal evolution of the average C-α RMSD for MalK chain B, in the three simulated states: the ATP (pre-hydrolysis) state, the ADP (post-hydrolysis) state and the Apo (nucleotide-free) state. Each state comprises nine replicates. The C-α RMSD was calculated against the initial structure of each simulation. The error bars correspond to the standard deviation obtained by bootstrapping. The error bars are represented in orange, cyan and light green in the ATP, ADP and Apo states respectively. The length of the simulations is different in the three states, to ensure equilibrated simulations: the ATP state was simulated for 300 ns, the ADP for 360 ns and the Apo for 410 ns.*

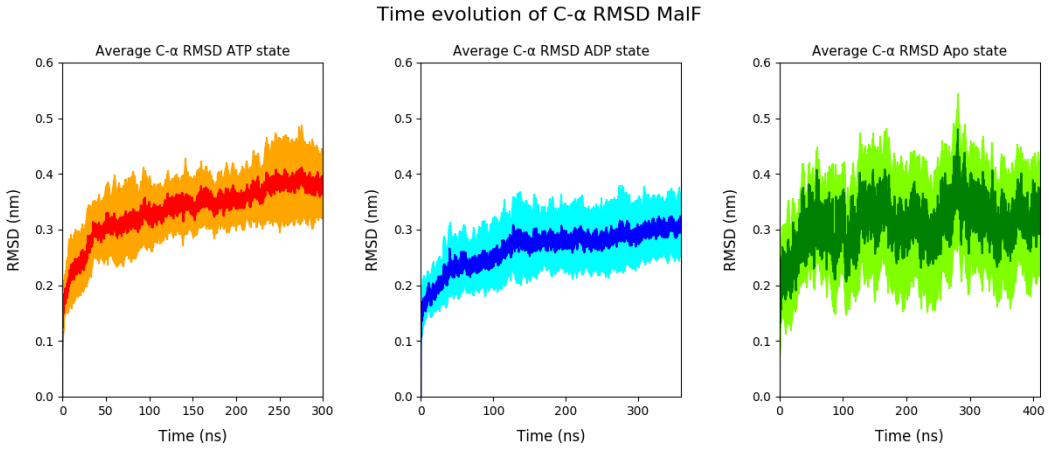


Figure 5.15- *Temporal evolution of the average C-α RMSD for MalF, in the three simulated states: the ATP (pre-hydrolysis) state, the ADP (post-hydrolysis) state and the Apo (nucleotide-free) state. Each state comprises nine replicates. The C-α RMSD was calculated against the initial structure of each simulation. The error bars correspond to the standard deviation obtained by bootstrapping. The error bars are represented in orange, cyan and light green in the ATP, ADP and Apo states respectively. The length of the simulations is different in the three states, to ensure equilibrated simulations: the ATP state was simulated for 300 ns, the ADP for 360 ns and the Apo for 410 ns.*

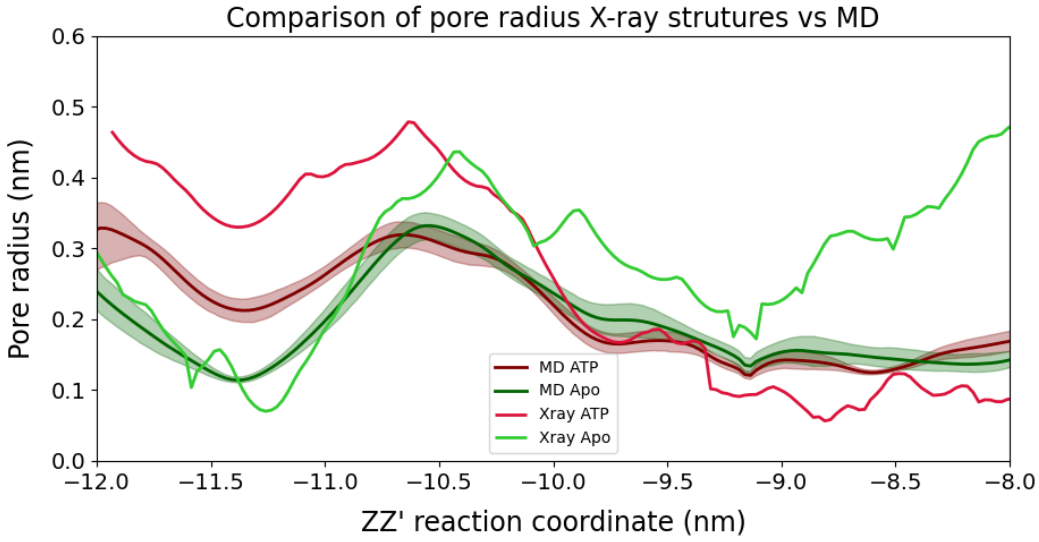


Figure 5.16-*Comparison of the pore radius between the X-ray structures used as starting points for this work and the MD simulations of the respective states, ATP and Apo. The radius of the starting structure of the ATP state is represented in red, while the radius of the x-ray structure for the Apo state is represented in green. The average radius from the ATP MD simulation is presented in dark red, while average radius from the Apo MD simulation is presented in dark green. The error shading is the 95% confidence interval obtained by bootstrapping.*

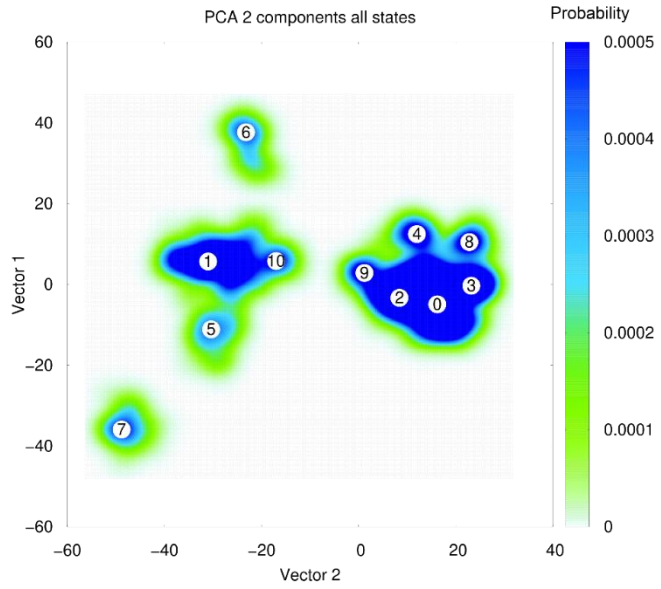


Figure 5.17- *Probability landscape obtained when performing PCA of all the simulated states (ATP, ADP and Apo). The probability is plotted in a gradient of colour from white (lower probability) to dark blue (higher probability). The white dots with numbers indicate the maxima of each population. The basins 1, 5, 6, 7 and 10 are exclusively occupied by the Apo state, while basins 0, 2, 3, 4, 8 and 9 are occupied by the ATP and ADP states.*

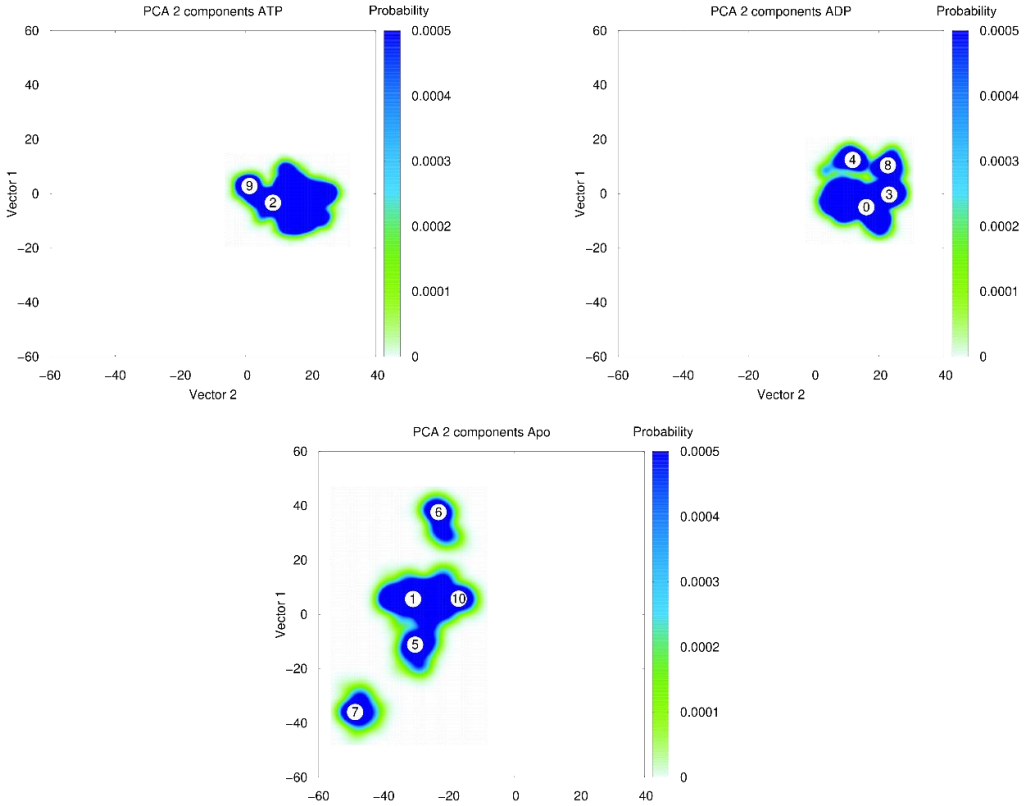
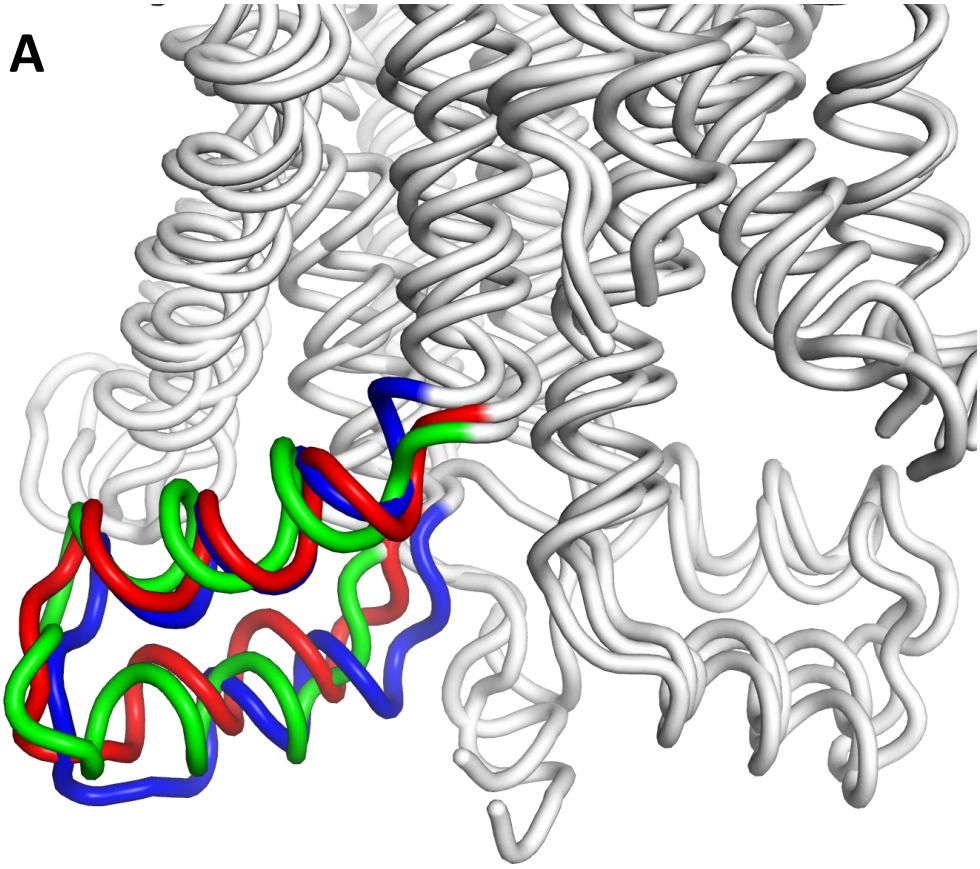


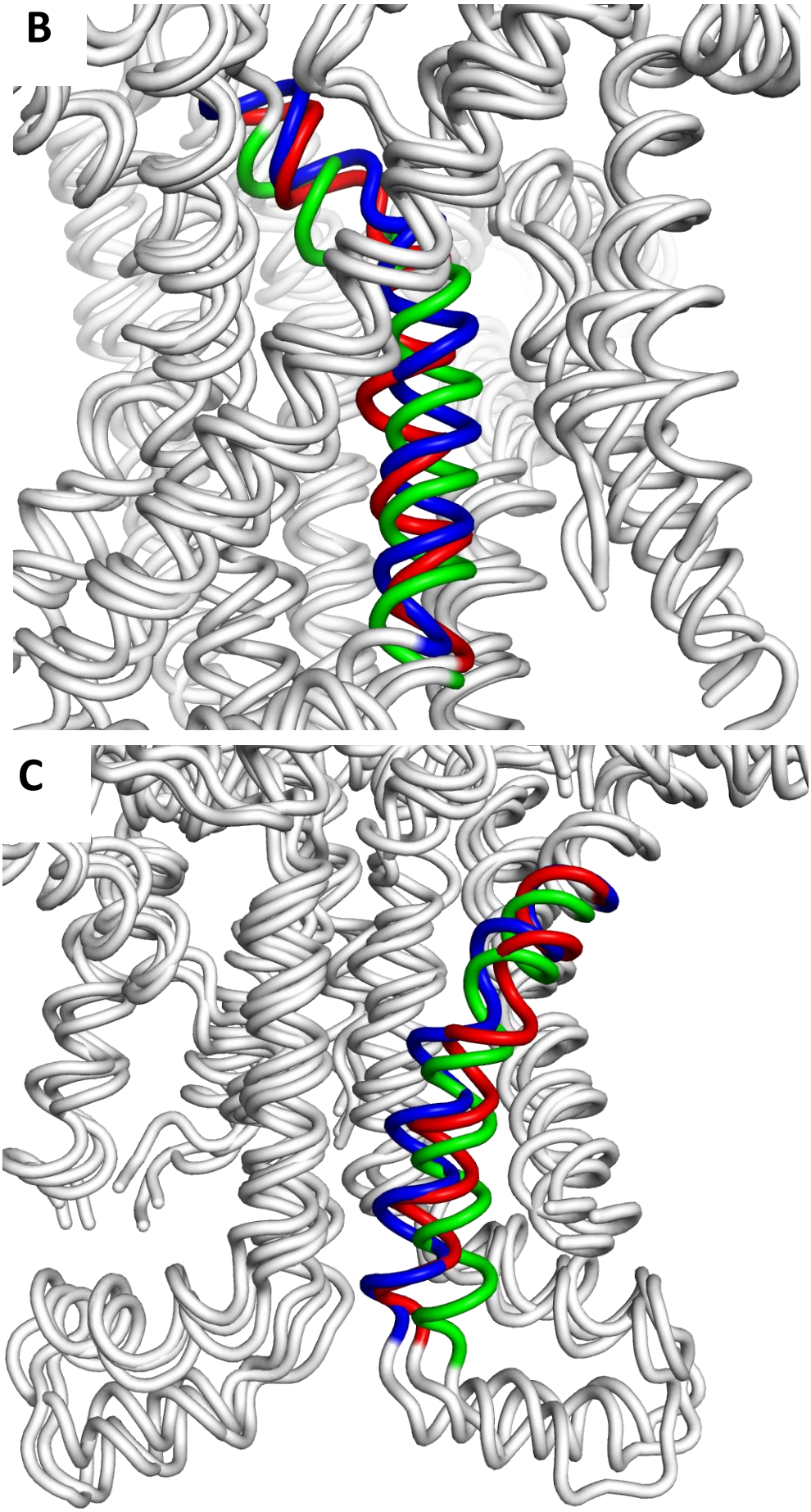
Figure 5.18- *Probability landscape of each simulated state, projected in the same space as the landscape in figure 5.17. The probability is plotted in a gradient of colour from white (lower probability) to dark blue (higher probability). The white dots with numbers indicate the maxima of each basin.*

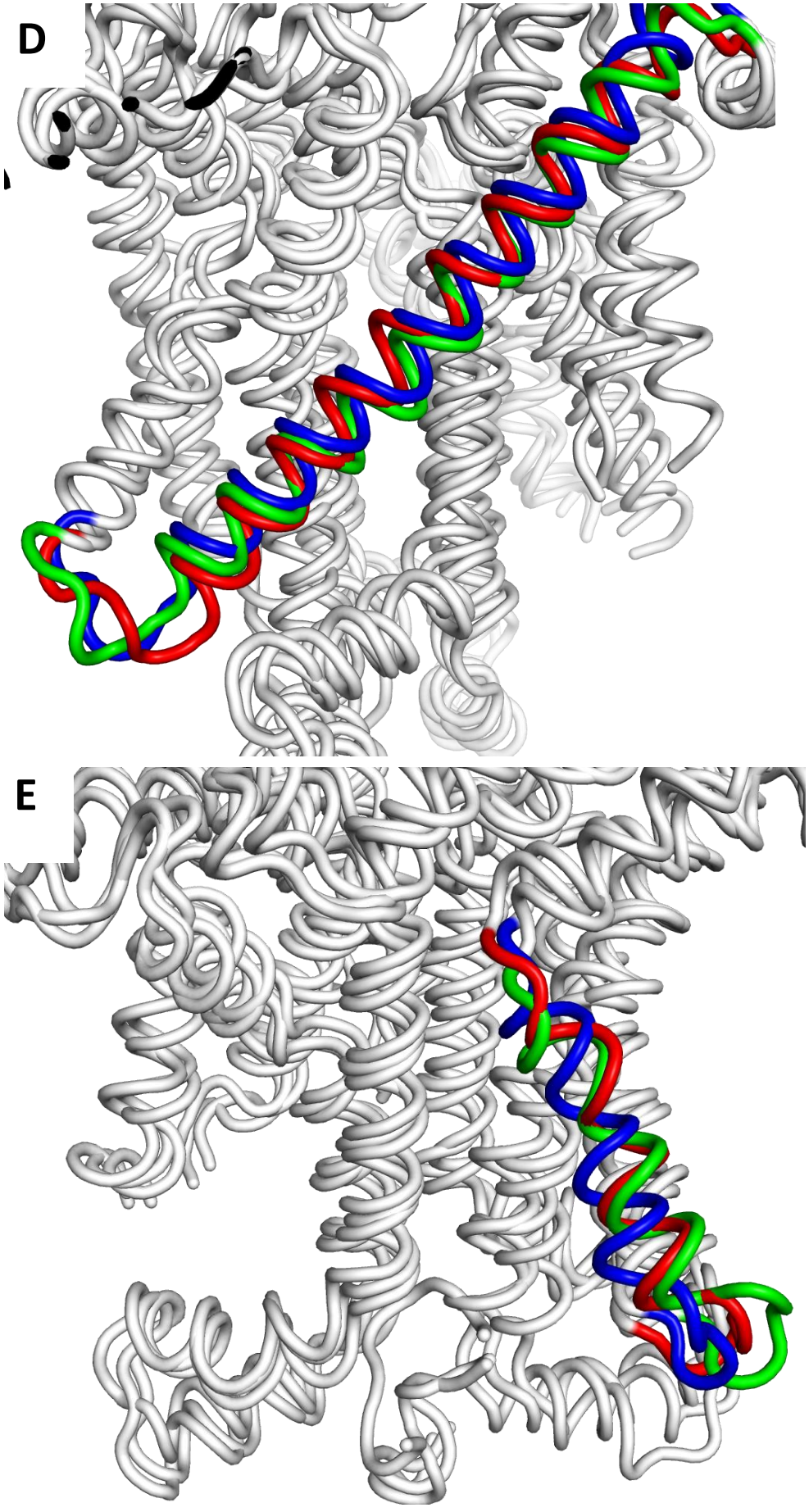
***Table 5.3**-Distribution of the conformations of three states: ATP, ADP and Apo per region. The lines coloured in green correspond to regions predominantly Apo, the lines coloured in blue correspond to regions with a predominant ADP character and the line in red corresponds to a region with a high ATP content.*

Basin	%ATP	%ADP	%Apo
0	66.84	45.33	0.00
1	0.00	0.00	59.11
2	15.42	18.56	0.00
3	9.16	14.85	0.00
4	1.43	10.43	0.00
5	0.00	0.00	11.55
6	0.00	0.00	11.11
7	0.00	0.00	11.11
8	0.07	10.23	0.00
9	7.09	0.61	0.00
10	0.00	0.00	7.12

Figure 5.19




B
C


D
E

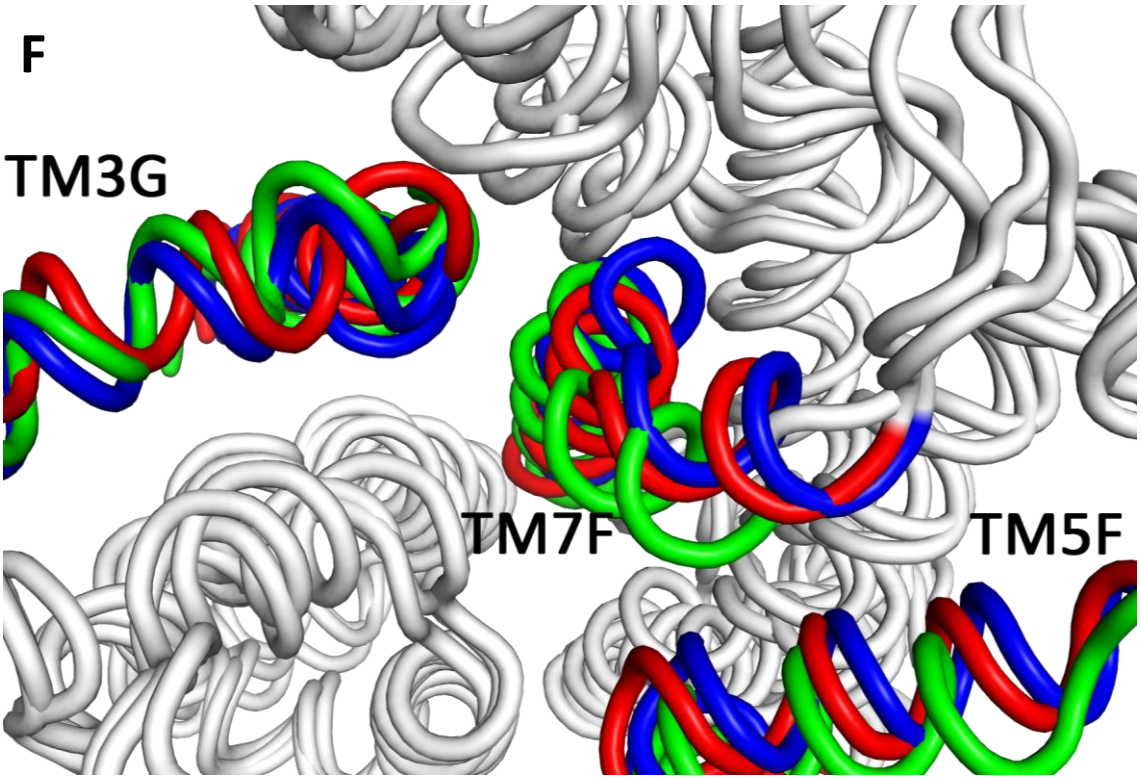


Figure 5.19- *Comparison of the ATP, ADP and Apo most probable structures. The conformations represented are the most probable structures of relevant basins, namely basins 8, 9 and 2, that correspond to the ADP, ATP and Apo states, respectively. The most relevant regions are coloured while the remaining protein is represented in white. The most probable ATP conformation is represented in red, while its ADP counterpart is represented in blue and the Apo conformation is represented in green. The orientation of the figures A to E is from the periplasmic side to the cytoplasmic one, while figure F is focusing the periplasmic side from the bottom. A - The MalF coupling helix. B- Transmembrane helix 7 of MalF (TM7F). C- Transmembrane helix 6 of MalF (TM6F). D- Transmembrane helix 4 of MalF (TM4F). E- Transmembrane helix 5 of MalF (TM5F). F- Bottom view of the pore highlighting the TM3G, TM7F and TM5F helices.*

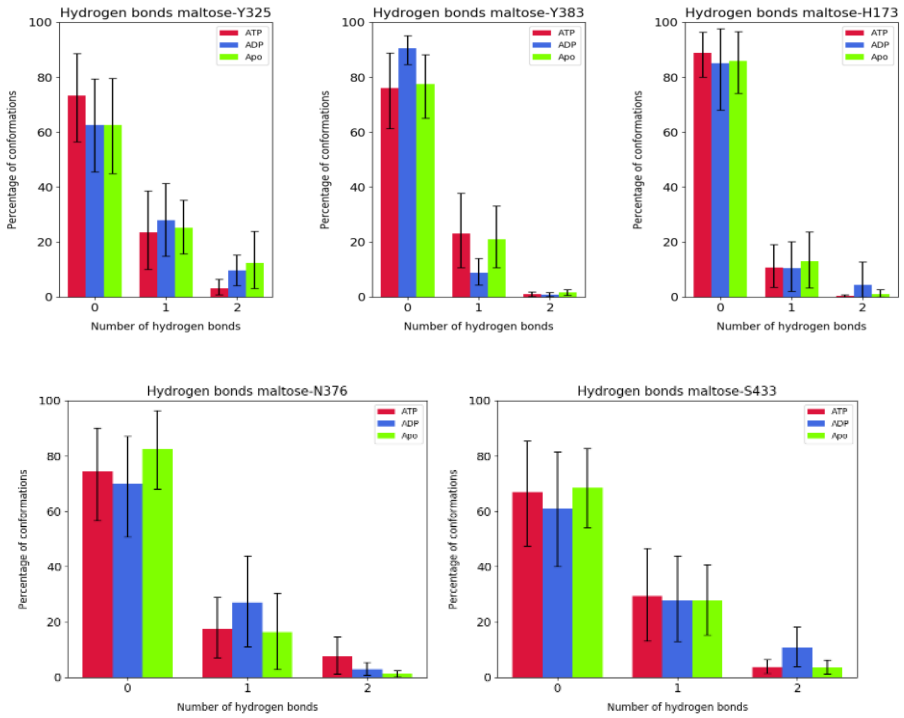


Figure 5.20*-Number of hydrogen bonds made between maltose and key protein residues Y325, Y383, H173, N376 and S433. ATP is represented in red, ADP in blue and the Apo state in green. The error bars represent the 95% confidence interval obtained by bootstrapping*.

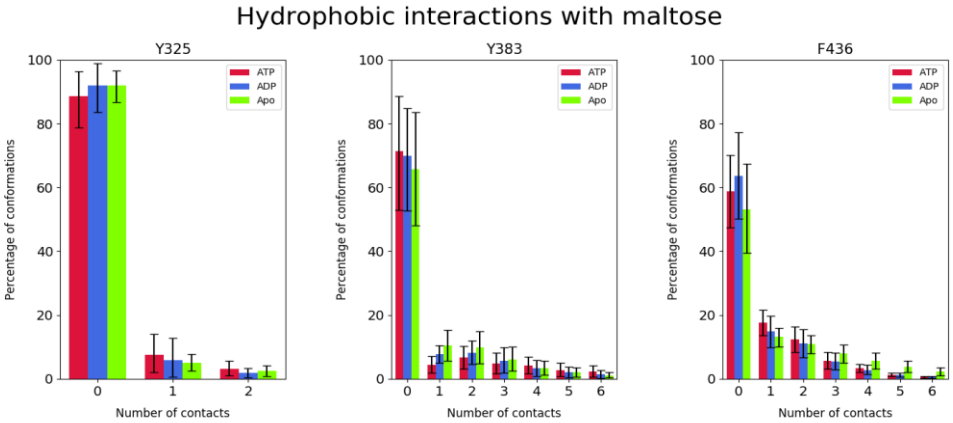


Figure 5.21*-Number of hydrophobic contacts made between the maltose rings and the aromatic sidechains of residues Y325, Y383 and F436. The maximum distance considered for a contact is 0.4 nm. ATP is represented in red, ADP in blue and the Apo state in green. The error bars represent the 95% confidence interval obtained by bootstrapping.*

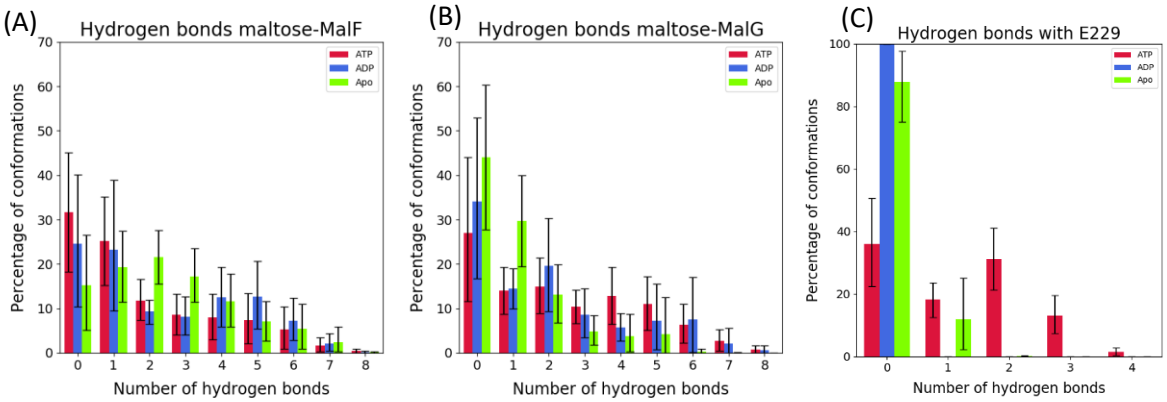


***Figure 5.22**-Number of hydrogen bonds made between maltose and protein residues. (A)-Hydrogen bonds between maltose and MalF. (B)- Hydrogen bonds between maltose and MalG. (C)-Hydrogen bonds between maltose and E229. ATP is represented in red, ADP in blue and the Apo state in green. The error bars represent the 95% confidence interval obtained by bootstrapping.*

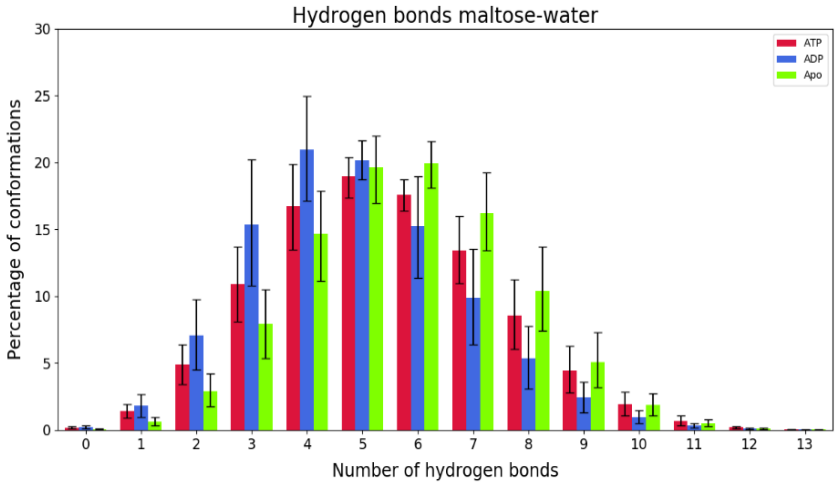


***Figure 5.23**-Number of hydrogen bonds made between maltose and water molecules within the pore. ATP is represented in red, ADP in blue and the Apo state in green. The error bars represent the 95% confidence interval obtained by bootstrapping.*

Table 5.4-*Windows used for umbrella sampling of the ATP PMF profile*

Initial position (nm)	Force constant (kJ/mol/nm^2)	Initial position (nm)	Force constant (kJ/mol/nm^2)
-12.1520	500	**-10.8662**	500
-11.0637	500	**-10.9405**	500
-11.2453	500	**-9.09825**	500
-11.4750	500	**-9.14214**	800
-11.5311	500	**-9.31624**	800
-11.8445	500	**-9.42136**	500
-11.9482	500	**-9.65039**	500
-10.0596	500	**-9.77419**	500
-10.1562	500	**-9.85904**	500
-10.2517	500	**-8.35064**	500
-10.4421	500	**-8.43524**	500
-10.5236	500	**-8.65938**	500
-10.7213	500	**-8.81653**	500

Table 5.5- *Windows used for sampling of the ADP PMF profile, in the ADP state*

Initial position (nm)	Force constant (kJ/mol/nm^2)	Initial position (nm)	Force constant (kJ/mol/nm^2)
-12.5292	500	**-10.2123**	500
-12.3985	500	**-10.0026**	500
-12.2780	500	**-9.86401**	500
-12.0389	500	**-9.64683**	500
-11.8563	800	**-9.45277**	500
-11.7915	500	**-9.39911**	500
-11.6772	500	**-9.33203**	500
-11.4938	800	**-9.10921**	500
-11.3291	500	**-9.04344**	500
-11.0986	500	**-8.80757**	500
-10.9470	500	**-8.69352**	500
-10.7803	800	**-8.55649**	800
-10.5825	500	**-8.43859**	800
-10.4495	500	**-8.25795**	500
-10.3328	500		

Table 5.6- *Windows used for sampling of the Apo PMF profile*

Initial position (nm)	Force constant (kJ/mol/nm^2)	Initial position (nm)	Force constant (kJ/mol/nm^2)
-12.0641	500	**-10.5113**	500
-12.3097	500	**-10.7496**	500
-12.4238	500	**-10.9657**	500
-11.1371	500	**-9.12126**	500
-11.2258	500	**-9.20689**	500
-11.4573	500	**-9.31658**	500
-11.6909	500	**-9.55786**	500
-11.8254	500	**-9.72485**	500
-11.9576	500	**-9.79942**	500
-10.0387	500	**-8.6486**	500
-10.1531	500	**-8.83127**	500
-10.3285	500	**-8.95105**	500
-10.3874	500		

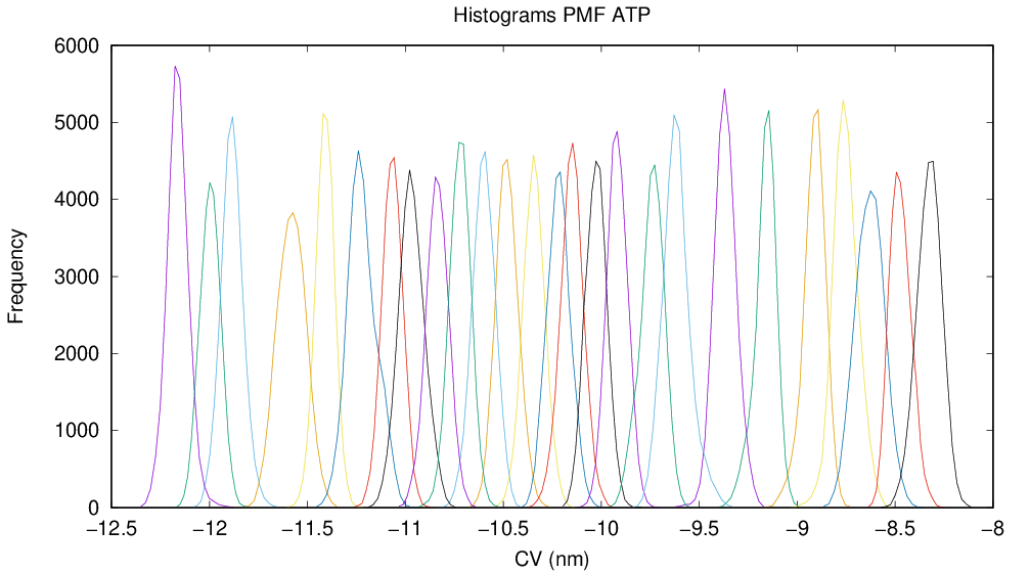


Figure 5.24*-Umbrella histograms obtained from each window in the ATP state. Each histogram represents the reaction coordinate distribution for each window and contains 30 ns of simulation.*

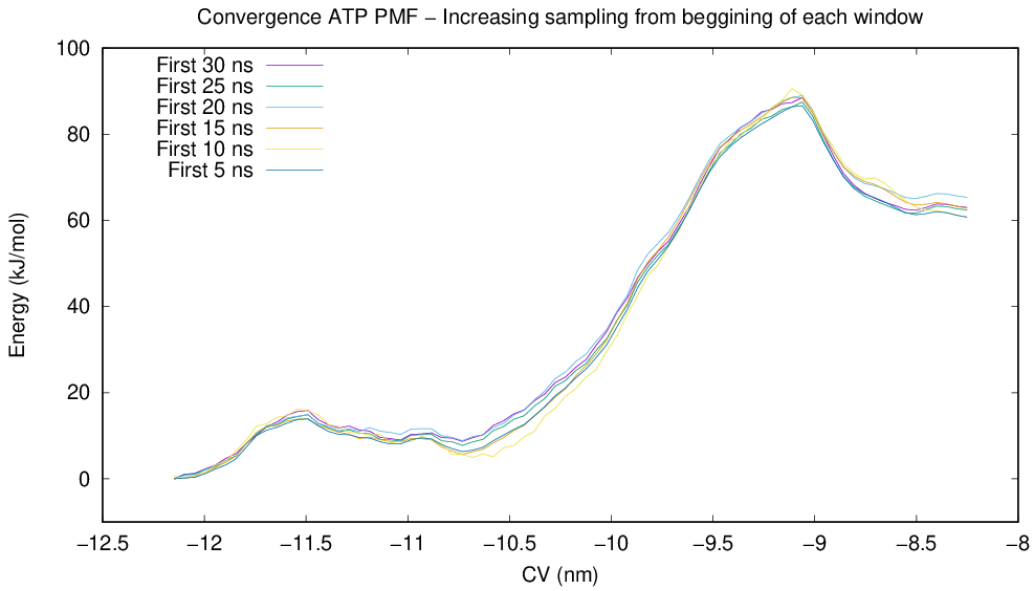


Figure 5.25- *Evaluation of ATP PMF convergence by increasingly adding 5 ns of sampling from the beginning of each umbrella window. The PMF using the first 5 ns of each window is represented in dark blue, using the first 10 ns of each window is represented in yellow, the first 15 ns of each window is represented in orange, the first 20 ns of each window is represented in cyan, the first 25 ns of each window is represented in green and using the whole 30 ns of each window is represented in purple.*

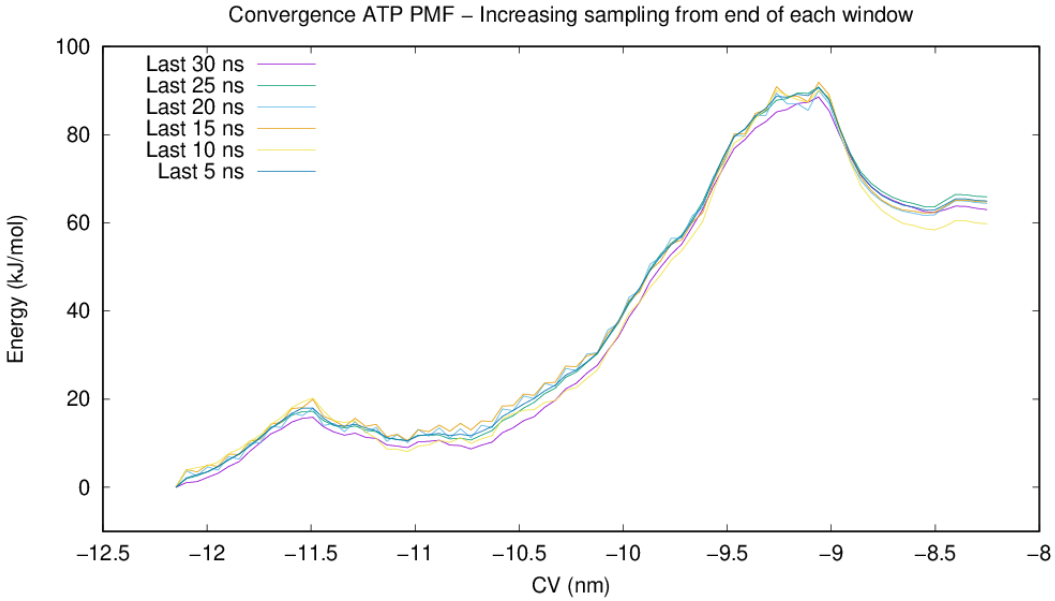


Figure 5.26- *Evaluation of ATP PMF convergence by increasingly adding 5 ns of sampling from the end of each umbrella window. The PMF using the last 5 ns of each window is represented in dark blue, using the last 10 ns of each window is represented in yellow, the last 15 ns of each window is represented in orange, the last20 ns of each window is represented in cyan, the last 25 ns of each window is represented in green and using the whole 30 ns of each window is represented in purple.*

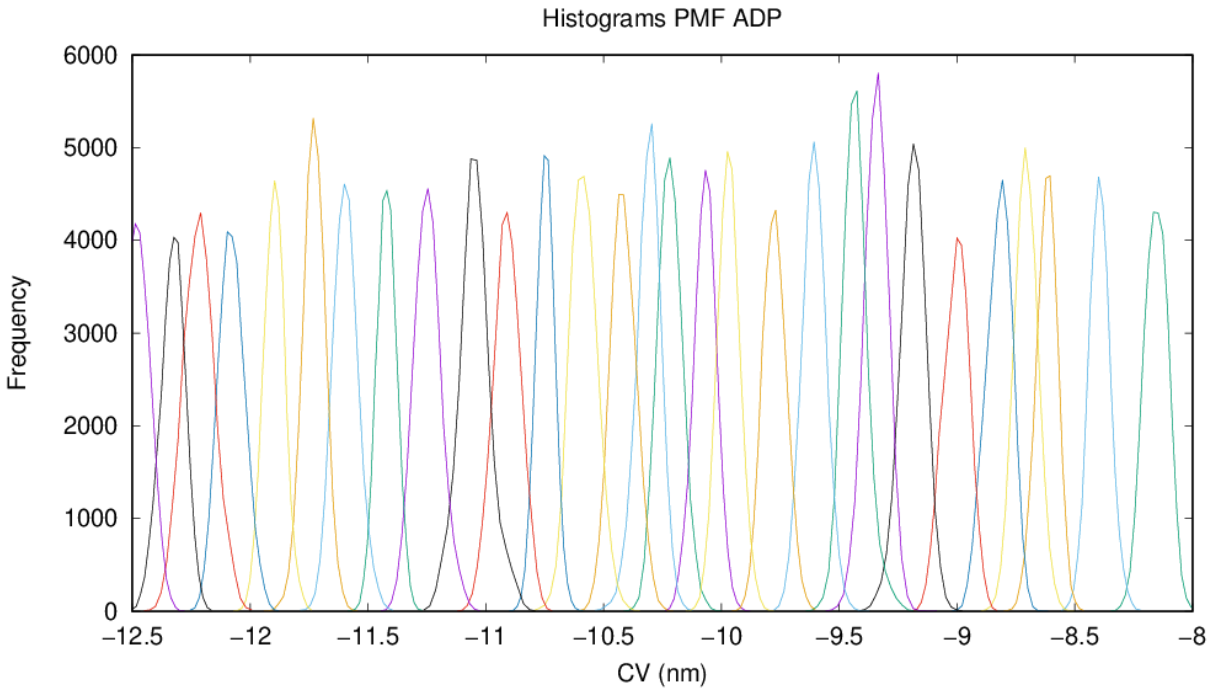


Figure 5.27- *Umbrella histograms obtained from each window in the ADP state, for the ADP-2 PMF profile. Each histogram represents the reaction coordinate distribution for each window and contains 30 ns of simulation.*

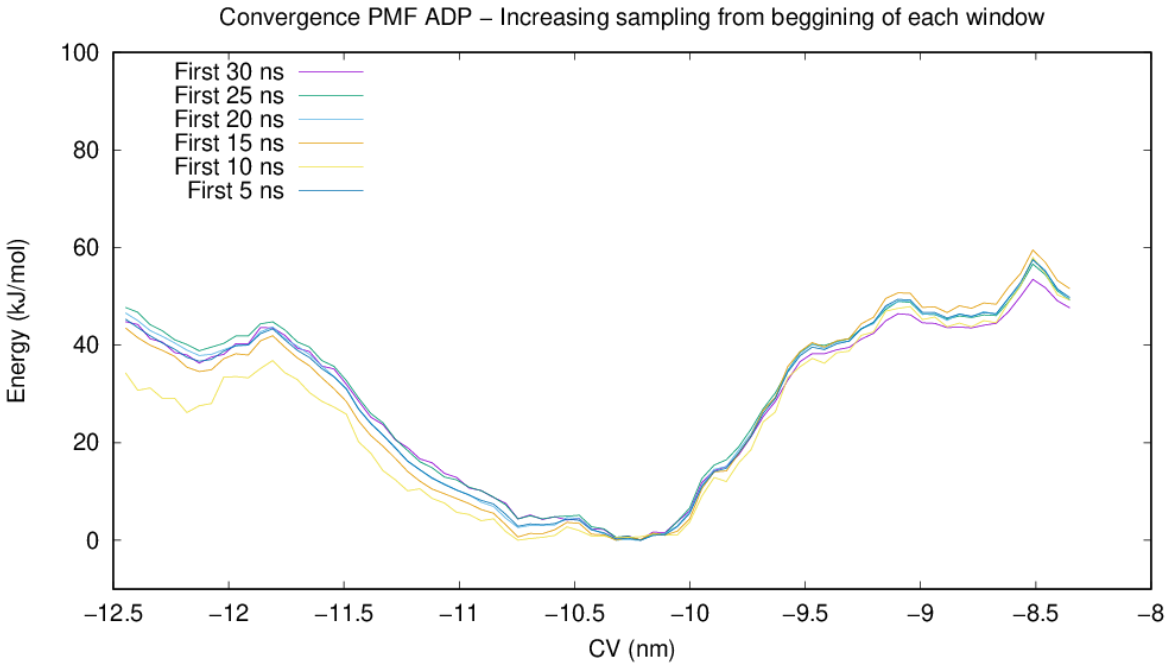


Figure 5.28*- Evaluation of ADP PMF convergence by increasingly adding 5 ns of sampling from the beginning of each umbrella window. The PMF using the first 5 ns of each window is represented in dark blue, using the first 10 ns of each window is represented in yellow, the first 15 ns of each window is represented in orange, the first 20 ns of each window is represented in cyan, the first 25 ns of each window is represented in green and using the whole 30 ns of each window is represented in purple.*

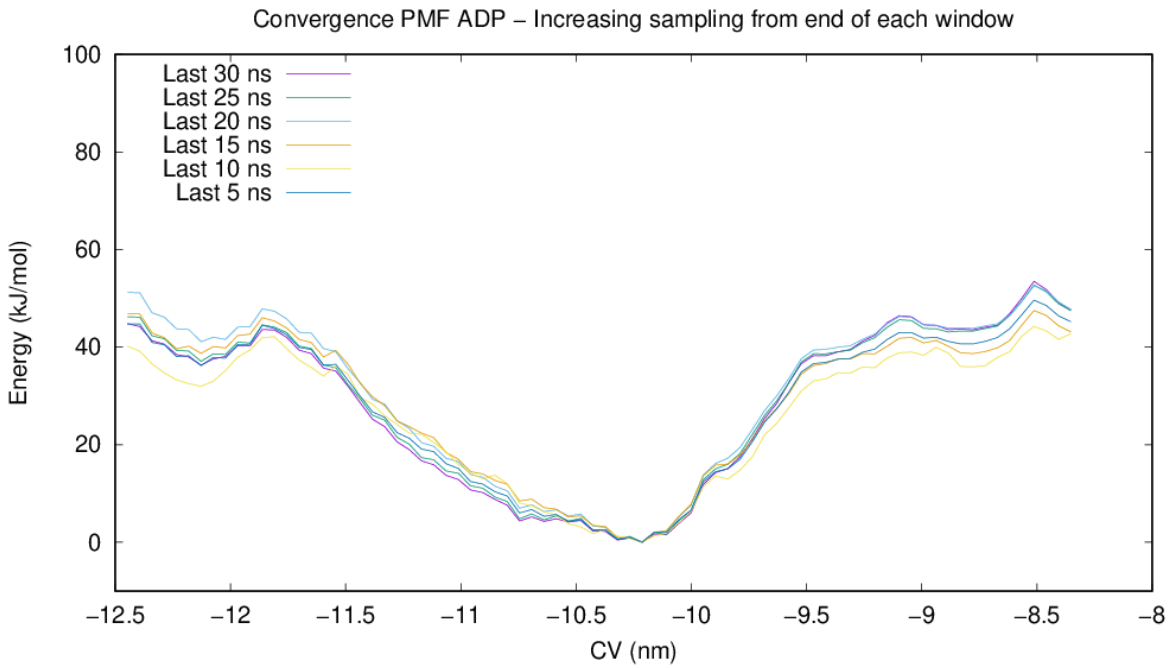


Figure 5.29- *Evaluation of the ADP PMF convergence by increasingly adding 5 ns of sampling from the end of each umbrella window. The PMF using the last 5 ns of each window is represented in dark blue, using the last 10 ns of each window is represented in yellow, the last 15 ns of each window is represented in orange, the last20 ns of each window is represented in cyan, the last 25 ns of each window is represented in green and using the whole 30 ns of each window is represented in purple.*

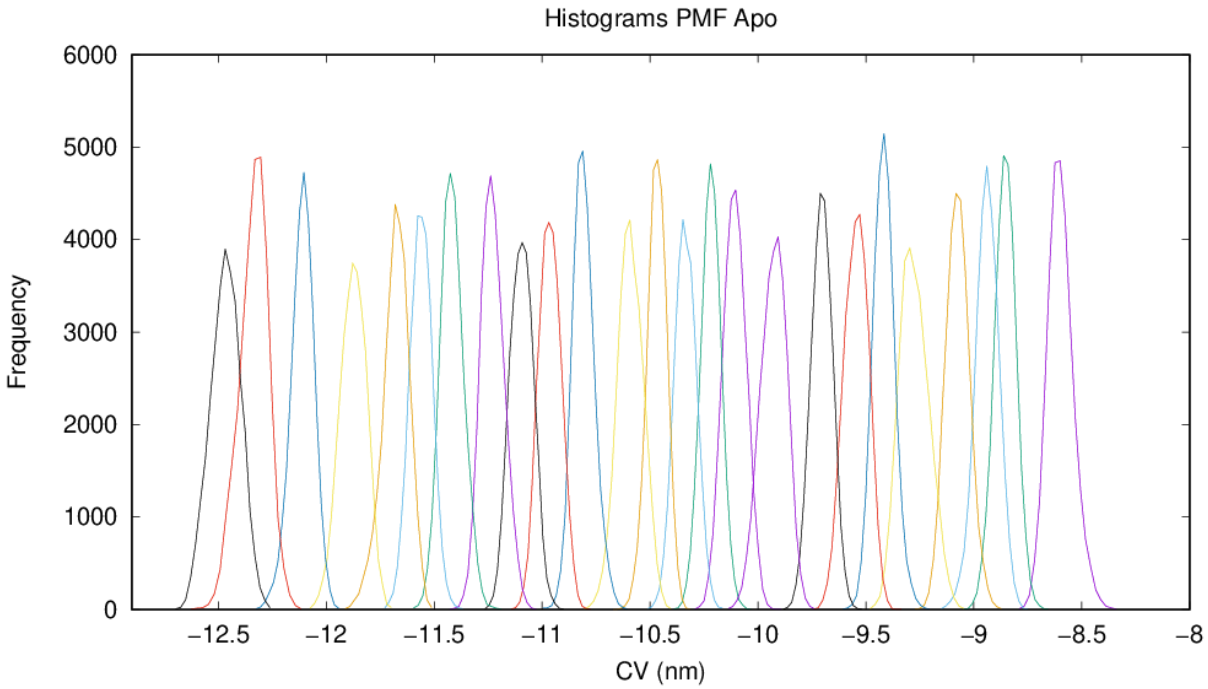


Figure 5.30- *Umbrella histograms obtained from each window in the Apo state. Each histogram represents the reaction coordinate distribution for each window and contains 30 ns of simulation.*

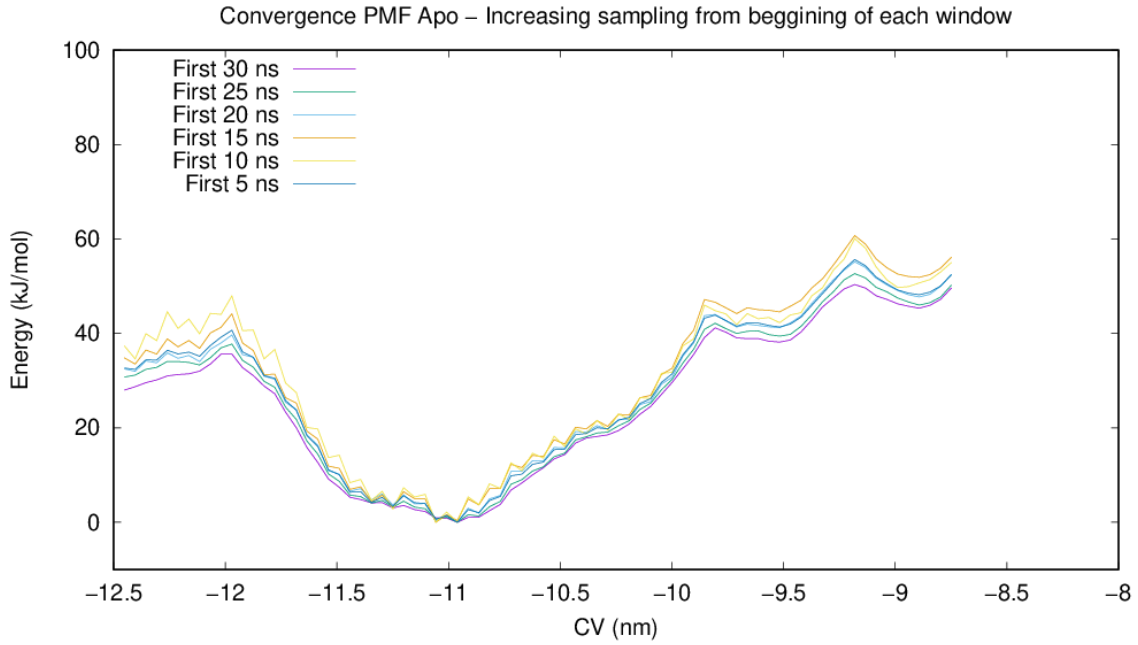


Figure 5.31- *Evaluation of Apo PMF convergence by increasingly adding 5 ns of sampling from the beginning of each umbrella window. The PMF using the first 5 ns of each window is represented in dark blue, using the first 10 ns of each window is represented in yellow, the first 15 ns of each window is represented in orange, the first 20 ns of each window is represented in cyan, the first 25 ns of each window is represented in green and using the whole 30 ns of each window is represented in purple.*

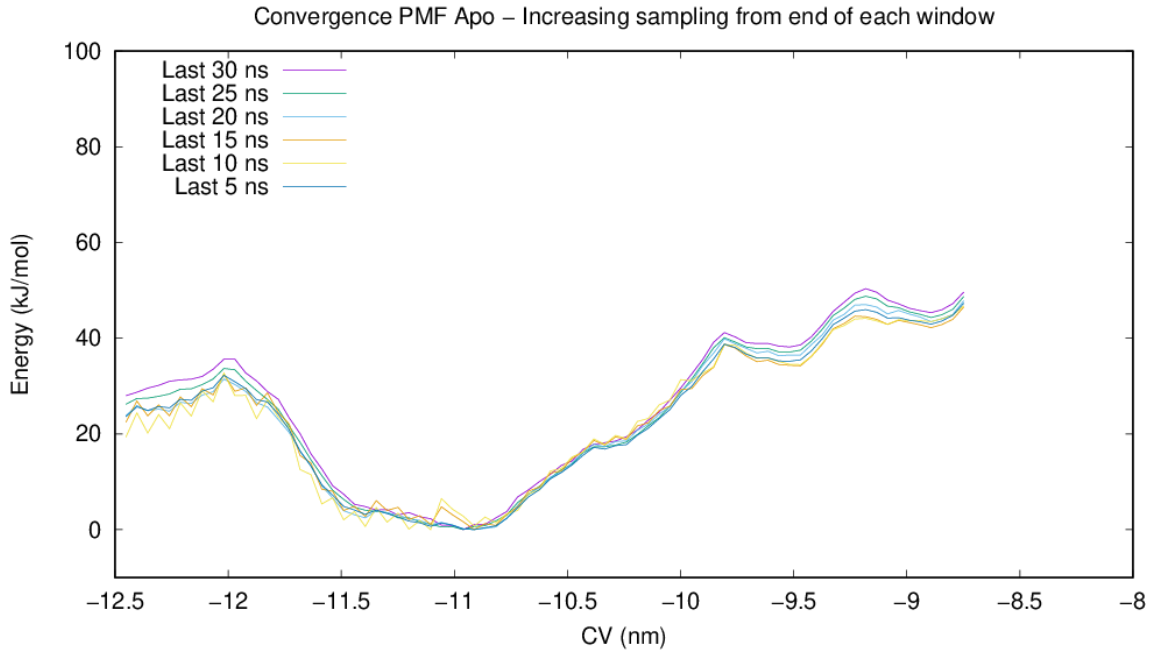


Figure 5.32-*Evaluation of the Apo PMF convergence by increasingly adding 5 ns of sampling from the end of each umbrella window. The PMF using the last 5 ns of each window is represented in dark blue, using the last 10 ns of each window is represented in yellow, the last 15 ns of each window is represented in orange, the last20 ns of each window is represented in cyan, the last 25 ns of each window is represented in green and using the whole 30 ns of each window is represented in purple.*

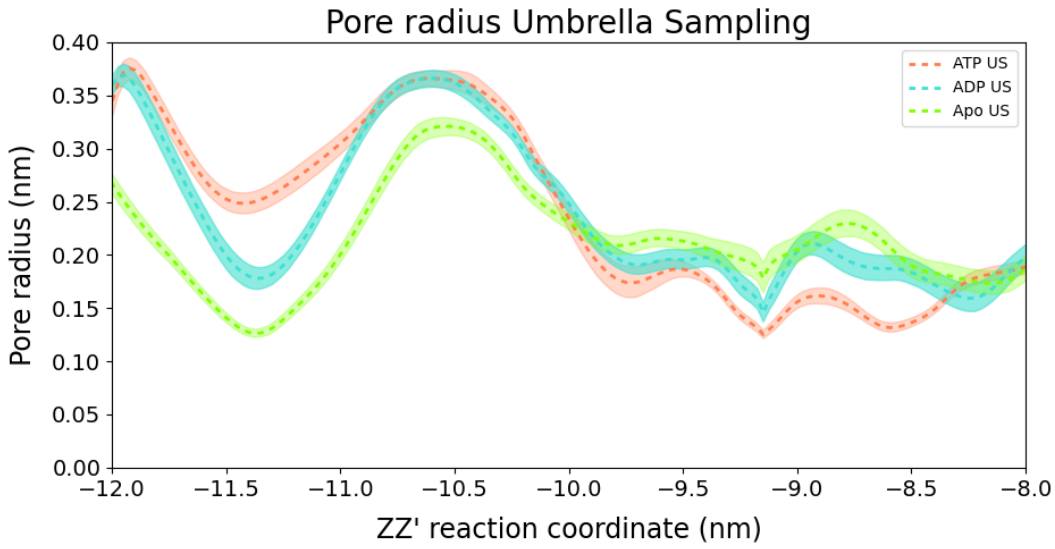


Figure 5.33-*Average pore radius during the umbrella sampling simulations for the three simulated states: ATP, ADP and Apo. The error shading corresponds to the 95% confidence interval obtained with bootstrapping.*

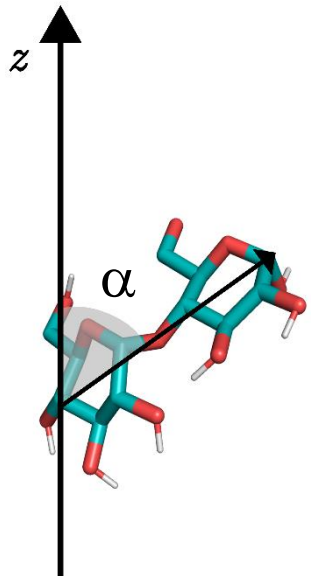


Figure 5.34- *Angle used to determine maltose orientation. The angle is calculated between the z-axis and the vector that links the most distant carbon atoms in the maltose rings, covering the entirety of the maltose molecule.*

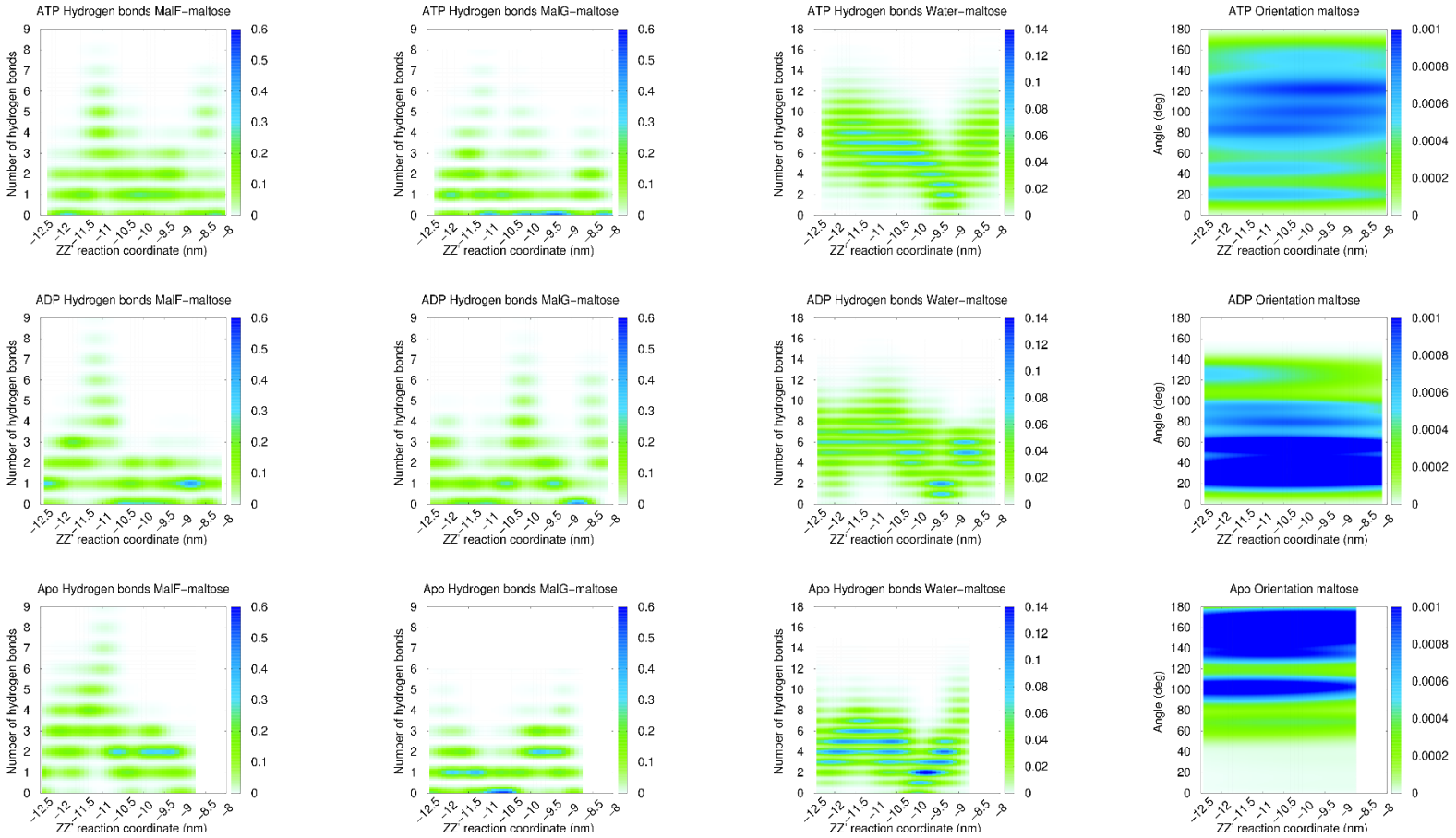


Figure 5.35*- 2D histograms showing the variation of several properties related with maltose in function of the reaction coordinate. These are: the number of hydrogen bonds performed by maltose with the MalF, MalG and water molecules. The orientation angle of maltose in the pore is also represented. It is defined as the angle between the z-axis and the vector that links the most distant carbon atoms in the maltose rings. Angles between 80º and 120º reflect horizontal conformations, while angles between 140º and 170º or 20º to 60º reflect vertical conformations.*

Chapter 6

Final Conclusions

ABC transporters are a group of proteins responsible for molecular transport across membranes. Members of this family play key roles, as for instance, in human health: the CFTR channel, whose malfunction is responsible for cystic fibrosis[207], the P-glycoprotein and ABCG2 proteins, involved in multidrug resistance[68], or the SUR1 receptor involved in diabetes[437]. Nonetheless, despite being extensively studied for over 40 years, numerous questions remain partially solved or even unsolved. These questions include the interplay between ATP hydrolysis and substrate binding and transport, the conformational details of the transport cycle or even the communication between domains. ATP stoichiometry during the cycle also caresses of explanations.

Despite having a common core constituted by two NBDs and TMDs, experimental and computational data point to a significant mechanistic diversity in ABC transporters, as seen in chapter 1. This is a consequence of the unique structural features of each transporter and eventually of the diversity of transported substrates. Therefore, it is necessary to proceed to the study of representative model systems of each transporter class, if not specific transporters.

Computational methods, such as the ones used in this thesis, can provide important knowledge at the atomic level on the dynamics of these systems, aiding in the interpretation of experimental data, as well as complementing the same.

In this thesis, it was proposed to unveil the structural determinants that underlie the function of ABC transporters, using computational approaches. To this end, three systems were studied: the CFTR NBDs, the MsbA NBDs and the *E.coli* $MalFGK_2E$ transporter. Therefore, it was possible to perform a thorough mechanistic investigation on ABC transporters, with the final goal of studying substrate translocation and its dependence on the ATP hydrolysis.

In the CFTR system, it was possible to determine the effect of the F508del mutation on the conformational changes triggered by ATP hydrolysis, while the molecular details concerning the adenylate kinase reaction were investigated in MsbA. Finally, the molecular details of substrate translocation were studied in the *E.coli* $MalFGK_2E$ transporter, which is a model system for type I ABC importers and despite being thoroughly studied, remained elusive, until now.

6.1- Effect of the F508del mutation in the CFTR nucleotide binding domains of CFTR before and upon ATP hydrolysis

The CFTR receptor is an one-of-a-kind ABC exporter involved in the efflux of chloride in epithelia contributing for maintaining the correct level of hydration of these tissues. The mutation F508del is the most common mutation in the Caucasian population. This mutation causes flawed protein folding resulting in protein retention in the endoplasmic reticulum and its degradation by the proteosome[438]. However, a small portion is able to reach the membrane, but

displays an impaired function and a decreased half-life residency[226–228]. The main goal of this project was to find whether F508del could influence the isolated nucleotide binding domains (NBDs), namely on dimer dynamics, ATP hydrolysis and even on nucleotide binding.

In order to study the effect of F508del on the CFTR NBDs, extensive molecular dynamics simulations of the wild-type and mutant NBD dimer were performed. In the presence of this mutation, a decrease in the number of catalytically competent conformations was observed, along with increased opening of the ATP binding sites in the pre-hydrolysis state. The increased pocket opening may affect the stability of ATP binding in this site. Additionally, using the subtraction technique, it was shown that F508del has the ability to disturb the conformational changes throughout the dimer, during the first 300 ps after ATP hydrolysis. Nonetheless, after a longer time, the conformational changes upon hydrolysis in the mutant become similar to the wild-type, except on the F508del loop, the residues surrounding the catalytic site and the portion of NBD2 adjacent to ABP1. Therefore, all these findings show that F508del affect the nucleotide binding domains of CFTR.

Further studies on this subject would pass by assessing the effect of F508del as well as the effect of ATP hydrolysis in the full transporter, eventually correlating the same with the channel permeability. In addition, various data suggest that CFTR activity is highly affected by lipid composition of the membrane. It would be interesting to access the effect of the membrane composition on the full length CFTR F508del and even check the effect of common CFTR modulators[439].

6.2- Exploring the nucleotide binding modes in the adenylate kinase cycle in MsbA

The MsbA transporter is responsible for the efflux of lipopolysaccharide in gram-negative bacteria[234]. In the presence of low ATP levels, MsbA showed the ability of performing a reverse adenylate kinase (rAK) reaction, that involves phosphate transfer from 2 ADP molecules to generate ATP and AMP[59]. The generated ATP would be later hydrolysed to ADP and phosphate. It is hypothesized that AK activity might be relevant in situations of ATP depletion in the cell[64]. Evidence of AK activity was also found in other transporters, such as CFTR, LmrA and TmrAB[61,62]. Previous experimental data suggested that the AK site was somewhere around the Q-loop, interacting with the adenine moiety of AMP while the H-loop eventually involved in phosphate interactions[59].

The main goal of this work was to investigate the nucleotide binding modes necessary for the adenylate kinase reaction to happen. In addition to the use of computational methods, this work also counted with data from solid-state NMR and EPR.

AMP binding was studied using a model of the MsbA NBDs and two AMP molecules were initially inserted between the NBD monomers, as hinted by the experimental data. Further MD simulations were performed for assessing the AMP binding modes and respective binding stability. The AMP binding site is located the Q-loops of each NBD. Regarding the AMP binding mode, it was found out that the adenine moiety interacted with the Q-loop, while the phosphate group was oriented towards the histidine residues from the H-loop.

In addition, it was shown that it is unlikely that two nucleotides bind at the same time in the AK site, because a single nucleotide needs to interact with both Q-loop motifs for a proper orientation towards the catalytic reaction centre.

Further work in this subject could eventually pass by the study of the catalytic mechanism of the adenylate kinase reaction and the evaluation of the conformational changes triggered by this reaction as well as further comparison with the ATPase reaction.

6.3-Effect of ATP hydrolysis and nucleotide exit on maltose translocation in the *E.coli* $MalFGK_2E$ importer

The $MalFGK_2E$ maltose importer is responsible for the intake of malto-oligossacharides in *E.coli*[69]. This system is an example of a type I ABC importer and a model system for this class of ABC transporters. Nonetheless, despite being extensively studied for more than 30 years, there is little information on the sequence of events posterior to ATP hydrolysis. Therefore, the main aim of this project was to evaluate the effect of ATP hydrolysis and nucleotide exit on maltose translocation within the $MalFGK_2E$ complex. In this way, exhaustive molecular dynamics (MD) simulations were performed of the $MalFGK_2E$ importer with ATP, ADP.IP and in the absence of nucleotides (Apo state).

It was concluded that ATP hydrolysis triggers a series of asymmetrical conformational changes throughout the complex that result in the decrease of energetic barriers to maltose translocation.

Considering the transmembrane domains, the most affected regions were the MalF coupling helix and the transmembrane helices TM4F, TM5F, TM6F and TM7F from MalF and helices TM3G and TM5G from MalG. The MalF coupling helix and TM6F suffer a shift towards the NBD direction. TM4F, TM5F and TM7F suffer a displacement towards MalE and TM5F moves away from the pore centre, near the NBDs. Nonetheless, TM5F and TM7F move laterally in the opposite direction at the bottom of the pore, causing constriction in this region. Interestingly TM3G, also suffers lateral displacements, but away from the pore center. Nucleotide exit causes similar effects to ATP hydrolysis in the same transmembrane helices, leading to pore constriction

in the periplasmic region, while TM5F and TM6F move further away from the pore center and MalF coupling helix and TM4F adopt a different conformation from the other states.

All these conformational changes resulted in modifications of the pore radius, in which a radius decrease was observed in the periplasmic region with a concomitant increase towards the cytoplasmic regions.

Regarding maltose binding, an extra binding spot in MalG was observed, with maximum occupation in the ATP state. This pocket was not observed in the crystal structures of $MalFGK_2E$. In this state, maltose also interacts the most with E229, reflecting its tendency to be located in the lower region of the pore.

The PMF profiles of maltose translocation show that the transformations triggered by ATP hydrolysis decrease greatly the energetic barriers to translocation towards the NBD, while simultaneously increasing the energetic barriers in the periplasmic side of the pore, stimulating the irreversibility of the process.

The structural analysis of the umbrella sampling simulations showed that E229 is key in reorienting and redirecting maltose from the periplasmic gate to the binding site and from the latter upwards. The maltose binding site, which is the same observed in the x-ray structure, corresponds to the energy minimum observed in the PMF profiles and might have the role of reorienting maltose towards the NBDs. E299 also aids in the exit of substrate and redirection towards the NBDs along with Y383 and H173. In this way, maltose is able to cross the cytoplasmic gate in a vertical orientation, which is the most energetically favourable.

Nucleotide exit does not lead to significant differences in free-energy of translocation comparing to the late post-hydrolysis state. In this way, it can be hypothesized that a weakening in the MalE interaction with the complex might be necessary to increase NBD separation and lead to further TMDs transformations, decreasing the remaining energetic barriers and facilitating maltose exit.

Further investigations on this topic can pass by understanding the role of MalE after hydrolysis. In this way, the post-hydrolysis state without MalE could be simulated and a PMF profile for maltose translocation could be obtained. Additionally, coarse-grain simulations could be used to simulate the progressive MalE unbinding and assess its effect on substrate translocation.

Overall, this thesis has contributed for a better understanding of the mechanisms underlying the function of ABC transporters. The three works here presented approached different aspects of the function of this class of proteins. In CFTR, F508del has an impact on the isolated NBD dimer, in addition to the previously suggested effect on the interdomain communication. Regarding MsbA, it was possible to shed light on the nucleotide binding modes related with the adenylate kinase reaction and contribute to the knowledge of the mechanism of this reaction. Concerning substrate translocation in $MalFGK_2E$, it was possible to discover the role of ATP hydrolysis and nucleotide exit on substrate translocation in this complex, suggesting a mechanism for substrate transport, which can be extrapolated to all the class I importers. Finally, it was shown that computational and theoretical methods can be highly valuable to address the molecular details related with the mechanism and function of ABC transporters.

Bibliography

1. Locher, K. P. Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat Struct Mol Biol* **23**, 487–493 (2016).

2. Rees, D. C., Johnson, E. & Lewinson, O. ABC transporters: the power to change. *Nat Rev Mol Cell Biol* **10**, 218–227 (2009).

3. *ABC Proteins - From Bacteria to Man*. (Academic Press, 2003).

4. Szöllősi, D., Rose-Sperling, D., Hellmich, U. A. & Stockner, T. Comparison of mechanistic transport cycle models of ABC exporters. *Biochim. Biophys. Acta - Biomembr.* **1860**, 818–832 (2018).

5. Hopfner, K. P. Invited review: Architectures and mechanisms of ATP binding cassette proteins. *Biopolymers* **105**, 492–504 (2016).

6. Saurin, W., Hofnung, M. & Dassa, E. Getting in or out: Early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *J. Mol. Evol.* **48**, 22–41 (1999).

7. Fitzpatrick, A. W. P. *et al.* Structure of the MacAB-TolC ABC-type tripartite multidrug efflux pump. *Nat. Microbiol.* **2**, (2017).

8. Qian, H. *et al.* Structure of the Human Lipid Exporter ABCA1. *Cell* **169**, 1228-1239.e10 (2017).

9. Dassa, E. & Bouige, P. The ABC of ABCs: A phylogenetic and functional classification of ABC systems in living organisms. *Res. Microbiol.* **152**, 211–229 (2001).

10. Rice, A. J., Park, A. & Pinkett, H. W. Diversity in ABC transporters: Type I, II and III importers. *Crit. Rev. Biochem. Mol. Biol.* **49**, 426–437 (2014).

11. *ABC transporters - 40 years on*. *ABC Transporters - 40 Years on* (Springer International Publishing, 2015). doi:10.1007/978-3-319-23476-2.

12. Barthelme, D. *et al.* Ribosome recycling depends on a mechanistic link between the FeS cluster domain and a conformational switch of the twin-ATPase ABCE1. *Proc. Natl. Acad. Sci.* **108**, 3228–3233 (2011).

13. Barthelme, D. *et al.* Structural organization of essential iron-sulfur clusters in the

evolutionarily highly conserved ATP-binding cassette protein ABCE1. *J. Biol. Chem.* **282**, 14598–14607 (2007).

14. Seifert, F. U., Lammens, K., Stoehr, G., Kessler, B. & Hopfner, K. Structural mechanism of ATP-dependent DNA binding and DNA end bridging by eukaryotic Rad50. *EMBO J.* **35**, 759–772 (2016).

15. Xiong, J., Feng, J., Yuan, D., Zhou, J. & Miao, W. Tracing the structural evolution of eukaryotic ATP binding cassette transporter superfamily. *Sci. Rep.* **5**, 1–15 (2015).

16. Dassa, E. Natural history of ABC systems: not only transporters. *Essays Biochem.* **50**, 19–42 (2011).

17. Wilkens, S. Structure and mechanism of ABC transporters. *F1000Prime Rep.* **7**, 1–9 (2015).

18. ter Beek, J., Guskov, A. & Slotboom, D. J. Structural diversity of ABC transporters. *J Gen Physiol* **143**, 419–435 (2014).

19. Grossmann, N. *et al.* Mechanistic determinants of the directionality and energetics of active export by a heterodimeric ABC transporter. *Nat. Commun.* **5**, 5419 (2014).

20. Vakkasoglu, A. S., Srikant, S. & Gaudet, R. D-helix influences dimerization of the ATP-binding cassette (ABC) transporter associated with antigen processing 1 (TAP1) nucleotide-binding domain. *PLoS One* **12**, e0178238 (2017).

21. Verdon, G., Albers, S. V., Dijkstra, B. W., Driessen, A. J. M. & Thunnissen, A. M. W. H. Crystal structures of the ATPase subunit of the glucose ABC transporter from Sulfolobus solfataricus: Nucleotide-free and nucleotide-bound conformations. *J. Mol. Biol.* **330**, 343–358 (2003).

22. Yuan, Y. R. *et al.* The crystal structure of the MJ0796 ATP-binding cassette. Implications for the structural consequences of ATP hydrolysis in the active site of an ABC transporter. *J. Biol. Chem.* **276**, 32313–32321 (2001).

23. Dalmas, O. *et al.* The Q-loop Disengages from the First Intracellular Loop during the Catalytic Cycle of the Multidrug ABC Transporter BmrA. *J. Biol. Chem.* **280**, 36857–36864 (2005).

24. Zaitseva, J., Jenewein, S., Jumpertz, T., Holland, I. B. & Schmitt, L. H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. *EMBO J.*

24, 1901–1910 (2005).

25. Kluth, M. *et al.* A mutation within the extended X loop abolished substrate-induced ATPase activity of the human liver ATP-binding cassette (ABC) transporter MDR3. *J. Biol. Chem.* **290**, 4896–4907 (2015).

26. Smith, P. C. *et al.* ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol Cell* **10**, 139–149 (2002).

27. Dawson, R. J. P. & Locher, K. P. Structure of the multidrug ABC transporter Sav1866 from Staphylococcus aureus in complex with AMP-PNP. *FEBS Lett.* **581**, 935–938 (2007).

28. Karpowich, N. K. & Wang, D. N. Assembly and mechanism of a group II ECF transporter. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 2534–2539 (2013).

29. Hohl, M., Briand, C., Grutter, M. G. & Seeger, M. A. Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. *Nat Struct Mol Biol* **19**, 395–402 (2012).

30. Csanady, L., Mihalyi, C., Szollosi, A., Torocsik, B. & Vergani, P. Conformational changes in the catalytically inactive nucleotide-binding site of CFTR. *J Gen Physiol* **142**, 61–73 (2013).

31. Hohl, M. *et al.* Structural basis for allosteric cross-talk between the asymmetric nucleotide binding sites of a heterodimeric ABC exporter. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11025–11030 (2014).

32. Abreu, B., Lopes, E. F., Oliveira, A. S. F. & Soares, C. M. F508del disturbs the dynamics of the nucleotide binding domains of CFTR before and after ATP hydrolysis. *Proteins Struct. Funct. Bioinforma.* 1–14 (2019) doi:10.1002/prot.25776.

33. Jones, P. M. & George, A. M. Role of the D-Loops in Allosteric Control of ATP Hydrolysis in an ABC Transporter. *J. Phys. Chem. A* **116**, 3004–3013 (2012).

34. Hunke, S., Mourez, M., Jéhanno, M., Dassa, E. & Schneider, E. ATP modulates subunit-subunit interactions in an ATP-binding cassette transporter (MalFGK2) determined by site-directed chemical cross-linking. *J. Biol. Chem.* **275**, 15526–15534 (2000).

35. Kreimer, D. I., Chai, K. P. & Ames, G. F. L. Nonequivalence of the nucleotide-binding subunits of an ABC transporter, the histidine permease, and conformational changes in the membrane complex. *Biochemistry* **39**, 14183–14195 (2000).

36. Diederichs, K. *et al.* Crystal structure of MalK, the ATPase subunit of the

trehalose/maltose ABC transporter of the archaeon Thermococcus litoralis. *EMBO J.* **19**, 5951–5961 (2000).

37. Dean, D. A., Reizer, J., Nikaido, H. & Saier, M. H. Regulation of the maltose transport system of Escherichia coli by the glucose-specific enzyme III of the phosphoenolpyruvate-sugar phosphotransferase system. Characterization of inducer exclusion-resistant mutants and reconstitution of inducer exclusion i. *J. Biol. Chem.* **265**, 21005–21010 (1990).

38. Orelle, C., Dalmas, O., Gros, P., Di Pietro, A. & Jault, J. M. The Conserved Glutamate Residue Adjacent to the Walker-B Motif is the Catalytic Base for ATP Hydrolysis in the ATP-binding Cassette Transporter BmrA. *J. Biol. Chem.* **278**, 47002–47008 (2003).

39. Moody, J. E., Millen, L. & Binns, D. Cooperative, ATP-dependent association of the nucleotide binding cassettes during the catalytic cycle of ATP-binding cassette transporters. *J. Biol. Chem.* **277**, 21111–21114 (2002).

40. Payen, L. F., Gao, M., Westlake, C. J., Cole, S. P. C. & Deeley, R. G. Role of carboxylate residues adjacent to the conserved core Walker B motifs in the catalytic cycle of multidrug resistance protein 1 (ABCC1). *J. Biol. Chem.* **278**, 38537–38547 (2003).

41. Oldham, M. L. & Chen, J. Snapshots of the maltose transporter during ATP hydrolysis. *Proc Natl Acad Sci U S A* **108**, 15152–15156 (2011).

42. Tombline, G. *et al.* Properties of P-glycoprotein with Mutations in the "Catalytic Carboxylate" Glutamate Residues. *J. Biol. Chem.* **279**, 46518–46526 (2004).

43. Urbatsch, I. L., Beaudet, L., Carrier, I. & Gros, P. Mutations in either nucleotide-binding site of P-glycoprotein (Mdr3) prevent vanadate trapping of nucleotide at both sites. *Biochemistry* **37**, 4592–4602 (1998).

44. Ernst, R., Koch, J., Horn, C., Tampé, R. & Schmitt, L. Engineering ATPase activity in the isolated ABC cassette of human TAP1. *J. Biol. Chem.* **281**, 27471–27480 (2006).

45. Zhou, Y., Ojeda-May, P. & Pu, J. H-loop histidine catalyzes ATP hydrolysis in the E. coli ABC-transporter HlyB. *Phys. Chem. Chem. Phys.* **15**, 15811–15815 (2013).

46. Prieß, M., Göddeke, H., Groenhof, G. & Schäfer, L. V. Molecular Mechanism of ATP Hydrolysis in an ABC Transporter. *ACS Cent. Sci.* (2018) doi:10.1021/acscentsci.8b00369.

47. Huang, W. & Liao, J.-L. Catalytic Mechanism of the Maltose Transporter Hydrolyzing ATP.

Biochemistry **55**, 224–231 (2016).

48. Herdendorf, T. J. & Nelson, S. W. Catalytic Mechanism of Bacteriophage T4 Rad50 ATP Hydrolysis. *Biochemistry* **53**, 5647–5660 (2014).

49. Hsu, W.-L., Furuta, T. & Sakurai, M. ATP Hydrolysis Mechanism in a Maltose Transporter Explored by QM/MM Metadynamics Simulation. *J. Phys. Chem. B* **120**, 11102–11112 (2016).

50. Patzlaff, J. S., Van der Heide, T. & Poolman, B. The ATP/substrate stoichiometry of the ATP-binding cassette (ABC) transporter OpuA. *J. Biol. Chem.* **278**, 29546–29551 (2003).

51. Muir, M., Williams, L. & Ferenci, T. Influence of transport energization on the growth yield of Escherichia coli. *J. Bacteriol.* **163**, 1237–1242 (1985).

52. Mimmack, M. L. *et al.* Energy coupling to periplasmic binding protein-dependent transport systems: Stoichiometry of ATP hydrolysis during transport in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 8257–8261 (1989).

53. Davidson, A. L. & Nikaido, H. Overproduction, solubilization, and reconstitution of the maltose transport system from Escherichia coli. *J. Biol. Chem.* **265**, 4254–4260 (1990).

54. Borths, E. L., Poolman, B., Hvorup, R. N., Locher, K. P. & Rees, D. C. In vitro functional characterization of BtuCD-F, the Escherichia coli ABC transporter for vitamin B12 uptake. *Biochemistry* **44**, 16301–16309 (2005).

55. Korkhov, V. M., Mireku, S. A. & Locher, K. P. Structure of AMP-PNP-bound vitamin B 12 transporter BtuCD-F. *Nature* **490**, 367–372 (2012).

56. Janas, E. *et al.* The ATP hydrolysis cycle of the nucleotide-binding domain of the mitochondrial ATP-binding cassette transporter Mdl1p. *J. Biol. Chem.* **278**, 26862–26869 (2003).

57. Verhalen, B. *et al.* Energy transduction and alternating access of the mammalian ABC transporter P-glycoprotein. *Nature* **543**, 738–741 (2017).

58. Randak, C. & Welsh, M. J. An Intrinsic Adenylate Kinase Activity Regulates Gating of the ABC Transporter CFTR. *Cell* **115**, 837–850 (2003).

59. Kaur, H. *et al.* Coupled ATPase-adenylate kinase activity in ABC transporters. *Nat. Commun.* **7**, 13864 (2016).

60. Bhaskara, V. *et al.* Rad50 Adenylate Kinase Activity Regulates DNA Tethering by Mre11/Rad50 Complexes. *Mol. Cell* **25**, 647–661 (2007).

61. Lammens, A. & Hopfner, K.-P. Structural Basis for Adenylate Kinase Activity in ABC ATPases. *J. Mol. Biol.* **401**, 265–273 (2010).

62. Randak, C. O. & Welsh, M. J. Adenylate kinase activity in ABC transporters. *J. Biol. Chem.* **280**, 34385–34388 (2005).

63. Randak, C. O., Ver Heul, A. R. & Welsh, M. J. Demonstration of Phosphoryl Group Transfer Indicates That the ATP-binding Cassette (ABC) Transporter Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Exhibits Adenylate Kinase Activity. *J. Biol. Chem.* **287**, 36105–36110 (2012).

64. Kaur, H. *et al.* Unexplored Nucleotide Binding Modes for the ABC Exporter MsbA. *J. Am. Chem. Soc.* (2018) doi:10.1021/jacs.8b06739.

65. Wang, Z., Hu, W. & Zheng, H. Pathogenic siderophore ABC importer YbtPQ adopts a surprising fold of exporter. *Sci. Adv.* **6**, 1–10 (2020).

66. Rempel, S. *et al.* A mycobacterial ABC transporter mediates the uptake of hydrophilic compounds. *Nature* **580**, 409–412 (2020).

67. Locher, K. P. Structure and mechanism of ATP-binding cassette transporters. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 239–245 (2009).

68. Wright, J., Muench, S. P., Goldman, A. & Baker, A. Substrate polyspecificity and conformational relevance in ABC transporters: New insights from structural studies. *Biochem. Soc. Trans.* **46**, 1475–1484 (2018).

69. Oldham, M. L., Chen, S. & Chen, J. Structural basis for substrate specificity in the Escherichia coli maltose transport system. *Proc Natl Acad Sci U S A* **110**, 18132–18137 (2013).

70. Lewinson, O. & Livnat-Levanon, N. Mechanism of Action of ABC Importers: Conservation, Divergence, and Physiological Adaptations. *J. Mol. Biol.* **429**, 606–619 (2017).

71. Kelly, D. J. & Thomas, G. H. The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. *FEMS Microbiol. Rev.* **25**, 405–424 (2001).

72. Armstrong, N. & Gouaux, E. Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor: Crystal structures of the GluR2 ligand binding core. *Neuron*

28, 165–181 (2000).

73. Misono, K. Natriuretic peptide receptor: Structure and signaling mechanism. *Natriuret. Pept. Physiol. Mol. Biol. Clin. Implic.* 119–136 (2014) doi:10.1007/978-1-4615-0927-1_3.

74. Poolman, B. & van der Heide, T. ABC transporters: one, two or four extracytoplasmic substrate-binding sites? *EMBO Rep.* **3**, 938–43 (2002).

75. Nguyen, P. T., Lai, J. Y., Lee, A. T., Kaiser, J. T. & Rees, D. C. Noncanonical role for the binding protein in substrate uptake by the MetNI methionine ATP Binding Cassette (ABC) transporter. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E10596–E10604 (2018).

76. Clore, G. M. Interplay between conformational selection and induced fit in multidomain protein-ligand binding probed by paramagnetic relaxation enhancement. *Biophys. Chem.* **186**, 3–12 (2014).

77. Seo, M. H., Park, J., Kim, E., Hohng, S. & Kim, H. S. Protein conformational dynamics dictate the binding affinity for a ligand. *Nat. Commun.* **5**, 1–7 (2014).

78. Kim, E. *et al.* A single-molecule dissection of ligand binding to a protein with intrinsic dynamics. *Nat. Chem. Biol.* **9**, 313–318 (2013).

79. Gouridis, G. *et al.* Conformational dynamics in substrate-binding domains influences transport in the ABC importer GlnPQ. *Nat. Struct. Mol. Biol.* **22**, 57–64 (2015).

80. Sharff, A. J., Rodseth, L. E., Spurlino, J. C. & Quiocho, F. A. Crystallographic Evidence of a Large Ligand-Induced Hinge-Twist Motion between the Two Domains of the Maltodextrin Binding Protein Involved in Active Transport and Chemotaxis. *Biochemistry* **31**, 10657–10663 (1992).

81. Felder, C. B., Graul, R. C., Lee, A. Y., Merkle, H. P. & Sadee, W. The venus flytrap of periplasmic binding proteins: An ancient protein module present in multiple drug receptors. *AAPS PharmSci* **1**, 1–20 (1999).

82. Berntsson, R. P. A., Smits, S. H. J., Schmitt, L., Slotboom, D. J. & Poolman, B. A structural classification of substrate-binding proteins. *FEBS Lett.* **584**, 2606–2617 (2010).

83. Majsnerowska, M. *et al.* Substrate-induced conformational changes in the S-component ThiT from an energy coupling factor transporter. *Structure* **21**, 861–867 (2013).

84. Rodionov, D. A., Gelfand, M. S., Todd, J. D., Curson, A. R. J. & Johnston, A. W. B. Computational reconstruction of iron- and manganese-responsive transcriptional

networks in α-proteobacteria. *PLoS Comput. Biol.* **2**, 1568–1585 (2006).

85. Finkenwirth, F. *et al.* ATP-dependent conformational changes trigger substrate capture and release by an ECF-type biotin transporter. *J. Biol. Chem.* **290**, 16929–16942 (2015).

86. Yu, Y. *et al.* Planar substrate-binding site dictates the specificity of ECF-type nickel/cobalt transporters. *Cell Res.* **24**, 267–277 (2014).

87. Zhang, P., Wang, J. & Shi, Y. Structure and mechanism of the S component of a bacterial ECF transporter. *Nature* **468**, 717–720 (2010).

88. Faustino, I. *et al.* Membrane mediated toppling mechanism of the folate energy coupling factor transporter. *Nat. Commun.* **11**, 1–9 (2020).

89. Jardezky, O. Simple Allosteric Model for Membrane Pumps. *Nature* **211**, 969–970 (1966).

90. Chen, J., Lu, G., Lin, J., Davidson, A. L. & Quiocho, F. A. A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol. Cell* **12**, 651–661 (2003).

91. Parcej, D. & Tampé, R. ABC proteins in antigen translocation and viral inhibition. *Nat. Chem. Biol.* **6**, 572–580 (2010).

92. Higgins, C. F. & Linton, K. J. The ATP switch model for ABC transporters. *Nat. Struct. Mol. Biol.* **11**, 918–926 (2004).

93. Senior, A. E., Al-Shawi, M. K. & Urbatsch, I. L. The catalytic cycle of P-glycoprotein. *FEBS Lett.* **377**, 285–289 (1995).

94. Jones, P. M. & George, A. M. Opening of the ADP-bound active site in the ABC transporter ATPase dimer: Evidence for a constant contact, alternating sites model for the catalytic cycle. *Proteins Struct. Funct. Bioinforma.* **75**, 387–396 (2009).

95. Sauna, Z. E. *et al.* Catalytic cycle of ATP hydrolysis by P-glycoprotein: Evidence for formation of the E·S reaction intermediate with ATP-γ-S, a nonhydrolyzable analogue of ATP. *Biochemistry* **46**, 13787–13799 (2007).

96. Ward, A., Reyes, C. L., Yu, J., Roth, C. B. & Chang, G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc. Natl. Acad. Sci.* **104**, 19005–19010 (2007).

97. Jin, M. S., Oldham, M. L., Zhang, Q. & Chen, J. Crystal structure of the multidrug transporter P-glycoprotein from Caenorhabditis elegans. *Nature* **490**, 566–569 (2012).

98. Shintre, C. A. *et al.* Structures of ABCB10, a human ATP-binding cassette transporter in

apo- and nucleotide-bound states. *Proc. Natl. Acad. Sci.* **110**, 9710 LP – 9715 (2013).

99. Borbat, P. P. *et al.* Conformational Motion of the ABC Transporter MsbA Induced by ATP Hydrolysis. *PLOS Biol.* **5**, e271 (2007).

100. Dong, J., Yang, G. & Mchaourab, H. S. Structural Basis of Energy Transduction in the Transport Cycle of MsbA. *Science (80-.).* **308**, 1023 LP – 1028 (2005).

101. Zou, P., Bortolus, M. & Mchaourab, H. S. Conformational Cycle of the ABC Transporter MsbA in Liposomes: Detailed Analysis Using Double Electron–Electron Resonance Spectroscopy. *J. Mol. Biol.* **393**, 586–597 (2009).

102. Moradi, M. & Tajkhorshid, E. Mechanistic picture for conformational transition of a membrane transporter at atomic resolution. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18916–18921 (2013).

103. Mishra, S. *et al.* Conformational dynamics of the nucleotide binding domains and the power stroke of a heterodimeric ABC transporter. *Elife* **3**, e02740 (2014).

104. Timachi, M. H. *et al.* Exploring conformational equilibria of a heterodimeric ABC transporter. *Elife* **6**, e20236 (2017).

105. Hofmann, S. *et al.* Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature* **571**, 580–583 (2019).

106. Stefan, E., Hofmann, S. & Tampé, R. A single power stroke by ATP binding drives substrate translocation in a heterodimeric ABC transporter. *Elife* **9**, 1–17 (2020).

107. Zhang, Z., Liu, F. & Chen, J. Molecular structure of the ATP-bound, phosphorylated human CFTR. *Proc. Natl. Acad. Sci.* **115**, 12757 LP – 12762 (2018).

108. Hwang, T. C. *et al.* Structural mechanisms of CFTR function and dysfunction. *J. Gen. Physiol.* **150**, 539–570 (2018).

109. Gunderson, K. L. & Kopito, R. R. Conformational states of CFTR associated with channel gating: The role of ATP binding and hydrolysis. *Cell* (1995) doi:10.1016/0092-8674(95)90310-0.

110. Liu, F. Y., Zhang, Z., Csanady, L., Gadsby, D. C. & Chen, J. Molecular Structure of the Human CFTR Ion Channel. *Cell* **169**, 85-95 e8 (2017).

111. Csanády, L., Vergani, P. & Gadsby, D. C. Strict coupling between CFTR's catalytic cycle and

gating of its Cl− ion pore revealed by distributions of open channel burst durations. *Proc. Natl. Acad. Sci.* **107**, 1241–1246 (2010).

112. Jih, K. Y. & Hwang, T. C. Nonequilibrium Gating of CFTR on an Equilibrium Theme. *Physiology* **27**, 351–361 (2012).

113. Zhang, J. & Hwang, T. C. Electrostatic tuning of the pre- and post-hydrolytic open states in CFTR. *J Gen Physiol* **149**, 355–372 (2017).

114. Yu, Y. C., Sohma, Y. & Hwang, T. C. On the mechanism of gating defects caused by the R117H mutation in cystic fibrosis transmembrane conductance regulator. *J. Physiol.* **594**, 3227–3244 (2016).

115. Venter, H., Shilling, R. A., Velamakanni, S., Balakrishnan, L. & Van Veen, H. W. An ABC transporter with a secondary-active multidrug translocator domain. *Nature* **426**, 866–870 (2003).

116. Agboh, K. *et al.* Powering the ABC multidrug exporter LmrA: How nucleotides embrace the ion-motive force. *Sci. Adv.* **4**, (2018).

117. Singh, H. *et al.* ATP-dependent substrate transport by the ABC transporter MsbA is proton-coupled. *Nat. Commun.* **7**, (2016).

118. van Veen, H. W. *et al.* Energy coupling in ABC exporters. *Res. Microbiol.* (2019) doi:10.1016/j.resmic.2019.08.003.

119. Sharom, F. J. Complex Interplay between the P-Glycoprotein Multidrug Efflux Pump and the Membrane: Its Role in Modulating Protein Function. *Front Oncol* **4**, 41 (2014).

120. Ferreira, R. J., Ferreira, M.-J. U. & dos Santos, D. J. V. A. Molecular Docking Characterizes Substrate-Binding Sites and Efflux Modulation Mechanisms within P-Glycoprotein. *J. Chem. Inf. Model.* **53**, 1747–1760 (2013).

121. Pleban, K. *et al.* P-Glycoprotein Substrate Binding Domains Are Located at the Transmembrane Domain/Transmembrane Domain Interfaces: A Combined Photoaffinity Labeling-Protein Homology Modeling Approach. *Mol. Pharmacol.* **67**, 365 LP – 374 (2005).

122. Higgins, C. F. & Gottesman, M. M. Is the multidrug transporter a flippase? *Trends Biochem. Sci.* **17**, 18–21 (1992).

123. Romsicki, Y. & Sharom, F. J. Phospholipid flippase activity of the reconstituted P-

glycoprotein multidrug transporter. *Biochemistry* **40**, 6937–6947 (2001).

124. Ferreira, R. J., Ferreira, M. J. U. & dos Santos, D. J. V. A. Reversing cancer multidrug resistance: Insights into the efflux by ABC transports from in silico studies. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **5**, 27–55 (2015).

125. Manolaridis, I. *et al.* Cryo-EM structures of a human ABCG2 mutant trapped in ATP-bound and substrate-bound states. *Nature* **563**, 426–430 (2018).

126. Cleophas, M. C. *et al.* ABCG2 polymorphisms in gout: Insights into disease susceptibility and treatment approaches. *Pharmgenomics. Pers. Med.* **10**, 129–142 (2017).

127. Matsuo, H. *et al.* ABCG2 Dysfunction Increases the Risk of Renal Overload Hyperuricemia. *Nucleosides. Nucleotides Nucleic Acids* **33**, 266–274 (2014).

128. Ferreira, R. J., Bonito, C. A., Cordeiro, M. N. D. S., Ferreira, M. J. U. & Dos Santos, D. J. V. A. Structure-function relationships in ABCG2: Insights from molecular dynamics simulations and molecular docking studies. *Sci. Rep.* **7**, 1–17 (2017).

129. Mi, W. *et al.* Structural basis of MsbA-mediated lipopolysaccharide transport. *Nature* **549**, 233–237 (2017).

130. Perez, C. *et al.* Structure and mechanism of an active lipid-linked oligosaccharide flippase. *Nature* **524**, 433–438 (2015).

131. Tang, X. *et al.* Cryo-EM structures of lipopolysaccharide transporter LptB2FGC in lipopolysaccharide or AMP-PNP-bound states reveal its transport mechanism. *Nat. Commun.* **10**, 1–12 (2019).

132. Li, Y., Orlando, B. J. & Liao, M. Structural basis of lipopolysaccharide extraction by the LptB 2 FGC complex. *Nature* **567**, 486–490 (2019).

133. Owens, T. W. *et al.* Structural basis of unidirectional export of lipopolysaccharide to the cell surface. *Nature* **567**, 550–553 (2019).

134. Oldham, M. L., Grigorieff, N. & Chen, J. Structure of the transporter associated with antigen processing trapped by herpes simplex virus. *Elife* **5**, 1–16 (2016).

135. Oldham, M. L. *et al.* A mechanism of viral immune evasion revealed by cryo-EM analysis of the TAP transporter. *Nature* **529**, 537–540 (2016).

136. Bountra, K. *et al.* Structural basis for antibacterial peptide self-immunity by the bacterial

ABC transporter McjD. *EMBO J.* **36**, 3062–3079 (2017).

137. Gu, R. X. *et al.* Conformational Changes of the Antibacterial Peptide ATP Binding Cassette Transporter McjD Revealed by Molecular Dynamics Simulations. *Biochemistry* **54**, 5989–5998 (2015).

138. Lin, D. Y. W., Huang, S. & Chen, J. Crystal structures of a polypeptide processing and secretion transporter. *Nature* **523**, 425–430 (2015).

139. Farkas, B. *et al.* Discovering the chloride pathway in the CFTR channel. *Cell. Mol. Life Sci.* **77**, 765–778 (2020).

140. Csanády, L., Vergani, P. & Gadsby, D. C. Structure, Gating, and Regulation of the CFTR Anion Channel. *Physiol. Rev.* **99**, 707–738 (2018).

141. Liu, F. *et al.* Structural identification of a hotspot on CFTR for potentiation. *Science (80-.).* **364**, 1184–1188 (2019).

142. Davidson, A. L., Shuman, H. A. & Nikaido, H. Mechanism of maltose transport in Escherichia coli: Transmembrane signaling by periplasmic binding proteins. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 2360–2364 (1992).

143. Oldham, M. L. & Chen, J. Crystal Structure of the Maltose Transporter in a Pretranslocation Intermediate State. *Science (80-.).* **332**, 1202 LP – 1205 (2011).

144. Doeven, M. K., van den Bogaart, G., Krasnikov, V. & Poolman, B. Probing Receptor-Translocator Interactions in the Oligopeptide ABC Transporter by Fluorescence Correlation Spectroscopy. *Biophys. J.* **94**, 3956–3965 (2008).

145. Merino, G., Boos, W., Shuman, H. A. & Bohl, E. The Inhibition of Maltose Transport by the Unliganded Form of the Maltose-binding Protein ofEscherichia coli: Experimental Findings and Mathematical Treatment. *J. Theor. Biol.* **177**, 171–179 (1995).

146. Ames, G. F.-L., Liu, C. E., Joshi, A. K. & Nikaido, K. Liganded and Unliganded Receptors Interact with Equal Affinity with the Membrane Complex of Periplasmic Permeases, a Subfamily of Traffic ATPases. *J. Biol. Chem.* **271**, 14264–14270 (1996).

147. Böhm, S., Licht, A., Wuttge, S., Schneider, E. & Bordignon, E. Conformational plasticity of the type I maltose ABC importer. *Proc. Natl. Acad. Sci.* **110**, 5492–5497 (2013).

148. Fabre, L., Bao, H., Innes, J., Duong, F. & Rouiller, I. Negative Stain Single-particle EM of the Maltose Transporter in Nanodiscs Reveals Asymmetric Closure of MalK2 and Catalytic

Roles of ATP, MalE, and Maltose. *J. Biol. Chem.* **292**, 5457–5464 (2017).

149. Karpowich, N. K., Huang, H. H., Smith, P. C. & Hunt, J. F. Crystal structures of the BtuF periplasmic-binding protein for vitamin B12 suggest a functionally important reduction in protein mobility upon ligand binding. *J. Biol. Chem.* **278**, 8429–8434 (2003).

150. Clarke, T. E., Braun, V., Winkelmann, G., Tari, L. W. & Vogel, H. J. X-ray crystallographic structures of the Escherichia coli periplasmic protein FhuD bound to hydroxamate-type siderophores and the antibiotic albomycin. *J. Biol. Chem.* **277**, 13966–13972 (2002).

151. Hvorup, R. N. *et al.* Asymmetry in the Structure of the ABC Transporter-Binding Protein Complex BtuCD-BtuF. *Science (80-.).* **317**, 1387 LP – 1390 (2007).

152. Vigonsky, E., Ovcharenko, E. & Lewinson, O. Two molybdate/tungstate ABC transporters that interact very differently with their substrate binding proteins. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5440–5445 (2013).

153. Lewinson, O., Lee, A. T., Locher, K. P. & Rees, D. C. A distinct mechanism for the ABC transporter BtuCD–BtuF revealed by the dynamics of complex formation. *Nat. Struct. Mol. Biol.* **17**, 332–338 (2010).

154. Xu, K. *et al.* Crystal structure of a folate energy-coupling factor transporter from Lactobacillus brevis. *Nature* **497**, 268–271 (2013).

155. Zhang, M. *et al.* Structure of a pantothenate transporter and implications for ECF module sharing and energy coupling of group II ECF transporters. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18560–18565 (2014).

156. Wang, T. *et al.* Structure of a bacterial energy-coupling factor transporter. *Nature* **497**, 272–276 (2013).

157. Ter Beek, J., Duurkens, R. H., Erkens, G. B. & Slotboom, D. J. Quaternary structure and functional unit of Energy Coupling Factor (ECF)-type transporters. *J. Biol. Chem.* **286**, 5471–5475 (2011).

158. Jones, P. M. & George, A. M. Mechanism of ABC transporters: A molecular dynamics simulation of a well characterized nucleotide-binding subunit. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12639–12644 (2002).

159. Campbell, J. D., Deol, S. S., Ashcroft, F. M., Kerr, I. D. & Sansom, M. S. P. Nucleotide-dependent conformational changes in HisP: Molecular dynamics simulations of an ABC

transporter nucleotide-binding domain. *Biophys. J.* **87**, 3703–3715 (2004).

160. Jones, P. M. & George, A. M. Nucleotide-dependent allostery within the ABC transporter ATP-binding cassette: A computational study of the MJ0796 dimer. *J. Biol. Chem.* **282**, 22793–22803 (2007).

161. Oliveira, A. S. F., Baptista, A. M. & Soares, C. M. Insights into the Molecular Mechanism of an ABC Transporter: Conformational Changes in the NBD Dimer of MJ0796. *J. Phys. Chem. B* **114**, 5486–5496 (2010).

162. Oloo, E. O. & Tieleman, D. P. Conformational transitions induced by the binding of MgATP to the vitamin B 12 ATP-binding cassette (ABC) transporter BtuCD. *J. Biol. Chem.* **279**, 45013–45019 (2004).

163. Wen, P. C. & Tajkhorshid, E. Dimer Opening of the Nucleotide Binding Domains of ABC Transporters after ATP Hydrolysis. *Biophys J* **95**, 5100–5110 (2008).

164. Hsu, W. L., Furuta, T. & Sakurai, M. Analysis of the Free Energy Landscapes for the Opening-Closing Dynamics of the Maltose Transporter ATPase MalK2 Using Enhanced-Sampling Molecular Dynamics Simulation. *J. Phys. Chem. B* **119**, 9717–9725 (2015).

165. Hayashi, T., Chiba, S., Kaneta, Y., Furuta, T. & Sakurai, M. ATP-induced conformational changes of nucleotide-binding domains in an ABC transporter. Importance of the water-mediated entropic force. *J. Phys. Chem. B* **118**, 12612–12620 (2014).

166. Barreto-Ojeda, Estefania; Corradi, Valentina ; Gu, Ruo-Xu; Tieleman, P. _2018_JGP_150_417.pdf. *J. Gen. Physiol.* **150**, 417–429 (2018).

167. Campbell, J. D., Biggin, P. C., Baaden, M. & Sansom, M. S. P. Extending the Structure of an ABC Transporter to Atomic Resolution: Modeling and Simulation Studies of MsbA. *Biochemistry* **42**, 3666–3673 (2003).

168. Haubertin, D. Y., Madaoui, H., Sanson, A., Guérois, R. & Orlowski, S. Molecular Dynamics Simulations of E. coli MsbA Transmembrane Domain: Formation of a Semipore Structure. *Biophys. J.* **91**, 2517–2531 (2006).

169. Weng, J. W., Fan, K. N. & Wang, W. N. The conformational transition pathway of ATP binding cassette transporter MsbA revealed by atomistic simulations. *J. Biol. Chem.* **285**, 3053–3063 (2010).

170. Wen, P.-C., Padayatti, P., Zhang, Q. & Tajkhorshid, E. Spontaneous Phospholipid Binding

to the Bacterial Flippase MsbA. *Biophys. J.* **116**, 127a (2019).

171. Oliveira, A. S. F., Baptista, A. M. & Soares, C. M. Inter-domain Communication Mechanisms in an ABC Importer: A Molecular Dynamics Study of the MalFGK(2)E Complex. *Plos Comput. Biol.* **7**, (2011).

172. Wen, P. C. & Tajkhorshid, E. Conformational coupling of the nucleotide-binding and the transmembrane domains in ABC transporters. *Biophys. J.* **101**, 680–690 (2011).

173. Weng, J., Gu, S., Gao, X., Huang, X. & Wang, W. Maltose-binding protein effectively stabilizes the partially closed conformation of the ATP-binding cassette transporter MalFGK2. *Phys. Chem. Chem. Phys.* **19**, 9366–9373 (2017).

174. Weng, J., Fan, K. & Wang, W. The conformational transition pathways of ATP-binding cassette transporter BtuCD revealed by targeted molecular dynamics simulation. *PLoS One* **7**, (2012).

175. Pan, C., Weng, J. & Wang, W. ATP hydrolysis induced conformational changes in the Vitamin B12 Transporter BtuCD Revealed by MD simulations. *PLoS One* **11**, 1–19 (2016).

176. Prieß, M. & Schäfer, L. V. Release of Entropic Spring Reveals Conformational Coupling Mechanism in the ABC Transporter BtuCD-F. *Biophys. J.* **110**, 2407–2418 (2016).

177. Furuta, T., Sato, Y. & Sakurai, M. Structural Dynamics of the Heterodimeric ABC Transporter TM287/288 Induced by ATP and Substrate Binding. *Biochemistry* **55**, 6730–6738 (2016).

178. Göddeke, H. *et al.* Atomistic Mechanism of Large-Scale Conformational Transition in a Heterodimeric ABC Exporter. *J. Am. Chem. Soc.* **140**, 4543–4551 (2018).

179. Liu, H., Li, D., Li, Y. & Hou, T. Atomistic molecular dynamics simulations of ATP-binding cassette transporters. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **6**, 255–265 (2016).

180. Ferreira, R. J., Dos Santos, D. J. & Ferreira, M. J. U. P-glycoprotein and membrane roles in multidrug resistance. *Future Med. Chem.* **7**, 929–946 (2015).

181. Becker, J.-P., Depret, G., Van Bambeke, F., Tulkens, P. M. & Prévost, M. Molecular models of human P-glycoprotein in two different catalytic states. *BMC Struct. Biol.* **9**, 3 (2009).

182. O'Mara, M. L. & Mark, A. E. The effect of environment on the structure of a membrane protein: P-glycoprotein under physiological conditions. *J. Chem. Theory Comput.* **8**, 3964–3976 (2012).

183. Callebaut, I., Eudes, R., Mornon, J.-P. & Lehn, P. Nucleotide-binding domains of human cystic fibrosis transmembrane conductance regulator: detailed sequence analysis and three-dimensional modeling of the heterodimer. *Cell. Mol. Life Sci. C.* **61**, 230–242 (2004).

184. Roxo-Rosa, M. *et al.* Revertant mutants G550E and 4RK rescue cystic fibrosis mutants in the first nucleotide-binding domain of CFTR by different mechanisms. *Proc Natl Acad Sci U S A* **103**, 17891–17896 (2006).

185. Huang, S. Y., Bolser, D., Liu, H. Y., Hwang, T. C. & Zou, X. Molecular modeling of the heterodimer of human CFTR's nucleotide-binding domains using a protein-protein docking approach. *J. Mol. Graph. Model.* **27**, 822–828 (2009).

186. Mornon, J.-P., Lehn, P. & Callebaut, I. Atomic model of human cystic fibrosis transmembrane conductance regulator: Membrane-spanning domains and coupling interfaces. *Cell. Mol. Life Sci.* **65**, 2594–2612 (2008).

187. Serohijos, A. W. R. *et al.* Phenylalanine-508 mediates a cytoplasmic-membrane domain contact in the CFTR 3D structure crucial to assembly and channel function. *Proc Natl Acad Sci U S A* **105**, 3256–3261 (2008).

188. Mornon, J. P., Hoffmann, B., Jonic, S., Lehn, P. & Callebaut, I. Full-open and closed CFTR channels, with lateral tunnels from the cytoplasm and an alternative position of the F508 region, as revealed by molecular dynamics. *Cell. Mol. Life Sci.* **72**, 1377–1403 (2015).

189. Dalton, J., Kalid, O., Schushan, M., Ben-Tal, N. & Villà-Freixa, J. New Model of Cystic Fibrosis Transmembrane Conductance Regulator Proposes Active Channel-like Conformation. *J. Chem. Inf. Model.* **52**, 1842–1853 (2012).

190. Corradi, V., Gu, R. X., Vergani, P. & Tieleman, D. P. Structure of Transmembrane Helix 8 and Possible Membrane Defects in CFTR. *Biophys. J.* **114**, 1751–1754 (2018).

191. Strickland, K. M. *et al.* ATP-Dependent Signaling in Simulations of a Revised Model of Cystic Fibrosis Transmembrane Conductance Regulator (CFTR). *J. Phys. Chem. B* **123**, 3177–3188 (2019).

192. Bisignano, P. & Moran, O. Molecular dynamics analysis of the wild type and dF508 mutant structures of the human CFTR–nucleotide binding domain 1. *Biochimie* **92**, 51–57 (2010).

193. Estácio, S. G., Martiniano, H. F. M. C. & Faísca, P. F. N. Thermal unfolding simulations of NBD1 domain variants reveal structural motifs associated with the impaired folding of

F508del-CFTR. *Mol. Biosyst.* **12**, 2834–2848 (2016).

194. Zhenin, M., Noy, E. & Senderowitz, H. REMD Simulations Reveal the Dynamic Profile and Mechanism of Action of Deleterious, Rescuing, and Stabilizing Perturbations to NBD1 from CFTR. *J. Chem. Inf. Model.* **55**, 2349–2364 (2015).

195. Callebaut, I., Hoffmann, B., Lehn, P. & Mornon, J. P. Molecular modelling and molecular dynamics of CFTR. *Cell. Mol. Life Sci.* **74**, 3–22 (2017).

196. Moran, O., Galietta, L. J. V. & Zegarra-Moran, O. Binding site of activators of the cystic fibrosis transmembrane conductance regulator in the nucleotide binding domains. *Cell. Mol. Life Sci.* **62**, 446–460 (2005).

197. Kalid, O. *et al.* Small molecule correctors of F508del-CFTR discovered by structure-based virtual screening. *J. Comput. Aided. Mol. Des.* **24**, 971–991 (2010).

198. Odolczyk, N. *et al.* Discovery of novel potent ΔF508-CFTR correctors that target the nucleotide binding domain. *EMBO Mol. Med.* **5**, 1484–1501 (2013).

199. Norimatsu, Y. *et al.* Locating a plausible binding site for an open-channel blocker, GlyH-101, in the pore of the cystic fibrosis transmembrane conductance regulator. *Mol. Pharmacol.* **82**, 1042–1055 (2012).

200. Okiyoneda, T. *et al.* Mechanism-based corrector combination restores ΔF508-CFTR folding and function. *Nat. Chem. Biol.* **9**, 444–454 (2013).

201. Farinha, C. M. *et al.* Revertants, low temperature, and correctors reveal the mechanism of F508del-CFTR rescue by VX-809 and suggest multiple agents for full correction. *Chem. Biol.* **20**, 943–955 (2013).

202. Veit, G. *et al.* Some gating potentiators, including VX-770, diminish ΔF508-CFTR functional expression. *Sci. Transl. Med.* **6**, 246ra97 LP-246ra97 (2014).

203. Kim, Y. *et al.* Benzopyrimido-pyrrolo-oxazine-dione (R)-BPO-27 inhibits CFTR chloride channel gating by competition with ATP. *Mol. Pharmacol.* **88**, 689–696 (2015).

204. Cox, M. H., Kapoor, P., Briggs, D. A. & Kerr, I. D. Residues contributing to drug transport by ABCG2 are localised to multiple drug-binding pockets. *Biochem. J.* **475**, 1553–1567 (2018).

205. Quinton, P. M. Chloride Impermeability in Cystic-Fibrosis. *Nature* **301**, 421–422 (1983).

206. Di, A. *et al.* CFTR regulates phagosome acidification in macrophages and alters bactericidal activity. *Nat. Cell Biol.* **8**, 933-U52 (2006).

207. O'Sullivan, B. P. & Freedman, S. D. Cystic fibrosis. *Lancet* **373**, 1891–1904 (2009).

208. Hunt, J. F., Wang, C. & Ford, R. C. Cystic fibrosis transmembrane conductance regulator (ABCC7) structure. *Cold Spring Harb Perspect Med* **3**, a009514 (2013).

209. Lewis, H. A. *et al.* Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator. *Embo J.* **23**, 282–293 (2004).

210. Lewis, H. A. *et al.* Structure and dynamics of NBD1 from CFTR characterized using crystallography and hydrogen/deuterium exchange mass spectrometry. *J. Mol. Biol.* **396**, 406–430 (2010).

211. Liu, F. Y., Zhang, Z., Csanady, L., Gadsby, D. C. & Chen, J. Molecular Structure of the Human CFTR Ion Channel. *Cell* **169**, 85-95 e8 (2017).

212. Basso, C., Vergani, P., Nairn, A. C. & Gadsby, D. C. Prolonged nonhydrolytic interaction of nucleotide with CFTR's NH2-terminal nucleotide binding domain and its role in channel gating. *J. Gen. Physiol.* **122**, 333–348 (2003).

213. Aleksandrov, L., Aleksandrov, A. A., Chang, X. B. & Riordan, J. R. The first nucleotide binding domain of cystic fibrosis transmembrane conductance regulator is a site of stable nucleotide interaction, whereas the second is a site of rapid turnover. *J. Biol. Chem.* **277**, 15419–15425 (2002).

214. Berger, A. L., Ikuma, M. & Welsh, M. J. Normal gating of CFTR requires ATP binding to both nucleotide-binding domains and hydrolysis at the second nucleotide-binding domain. *Proc Natl Acad Sci U S A* **102**, 455–460 (2005).

215. Vergani, P., Lockless, S. W., Nairn, A. C. & Gadsby, D. C. CFTR channel opening by ATP-driven tight dimerization of its nucleotide-binding domains. *Nature* **433**, 876–880 (2005).

216. Tsai, M. F., Li, M. & Hwang, T. C. Stable ATP binding mediated by a partial NBD dimer of the CFTR chloride channel. *J Gen Physiol* **135**, 399–414 (2010).

217. Linsdell, P. Structural Changes Fundamental to Gating of the Cystic Fibrosis Transmembrane Conductance Regulator Anion Channel Pore BT - Protein Reviews: Volume 17. in (ed. Atassi, M. Z.) 13–32 (Springer Singapore, 2017). doi:10.1007/5584_2016_33.

218. Gao, X. & Hwang, T.-C. Localizing a gate in CFTR. *Proc. Natl. Acad. Sci.* **112**, 2461 LP – 2466 (2015).

219. El Hiani, Y. & Linsdell, P. Conformational changes opening and closing the CFTR chloride channel: Insights from cysteine scanning mutagenesis. *Biochem. Cell Biol.* **92**, 481–488 (2014).

220. Fatehi, M. & Linsdell, P. State-dependent Access of Anions to the Cystic Fibrosis Transmembrane Conductance Regulator Chloride Channel Pore. *J. Biol. Chem.* **283**, 6102–6109 (2008).

221. Cui, G. *et al.* Three charged amino acids in extracellular loop 1 are involved in maintaining the outer pore architecture of CFTR. *J. Gen. Physiol.* **144**, 159–179 (2014).

222. Sebastian, A. *et al.* Origin and evolution of the cystic fibrosis transmembrane regulator protein R domain. *Gene* **523**, 137–146 (2013).

223. Xie, J. *et al.* A Short Segment of the R Domain of Cystic Fibrosis Transmembrane Conductance Regulator Contains Channel Stimulatory and Inhibitory Activities That Are Separable by Sequence Modification. *J. Biol. Chem.* **277**, 23019–23027 (2002).

224. Vankeerberghen, A. *et al.* Characterization of 19 Disease-Associated Missense Mutations in the Regulatory Domain of the Cystic Fibrosis Transmembrane Conductance Regulator. *Hum. Mol. Genet.* **7**, 1761–1769 (1998).

225. Cutting, G. R. Cystic fibrosis genetics: from molecular understanding to clinical application. *Nat. Rev. Genet.* **16**, 45–56 (2015).

226. Dalemans, W. *et al.* Altered chloride ion channel kinetics associated with the ΔF508 cystic fibrosis mutation. *Nature* **354**, 526–528 (1991).

227. Rich, D. P. *et al.* Expression of cystic fibrosis transmembrane conductance regulator corrects defective chloride channel regulation in cystic fibrosis airway epithelial cells. *Nature* **347**, 358–363 (1990).

228. Sharma, M., Benharouga, M., Hu, W. & Lukacs, G. L. Conformational and temperature-sensitive stability defects of the delta F508 cystic fibrosis transmembrane conductance regulator in post-endoplasmic reticulum compartments. *J. Biol. Chem.* **276**, 8942–8950 (2001).

229. Swiatecka-Urban, A. *et al.* The short apical membrane half-life of rescued {Delta}F508-

cystic fibrosis transmembrane conductance regulator (CFTR) results from accelerated endocytosis of {Delta}F508-CFTR in polarized human airway epithelial cells. *J. Biol. Chem.* **280**, 36762–36772 (2005).

230. Aleksandrov, A. A. *et al.* Regulatory insertion removal restores maturation, stability and function of DeltaF508 CFTR. *J. Mol. Biol.* **401**, 194–210 (2010).

231. Ren, H. Y. *et al.* VX-809 corrects folding defects in cystic fibrosis transmembrane conductance regulator protein through action on membrane-spanning domain 1. *Mol. Biol. Cell* **24**, 3016–3024 (2013).

232. He, L. *et al.* Correctors of ΔF508 CFTR restore global conformational maturation without thermally stabilizing the mutant protein. *FASEB J.* **27**, 536–545 (2012).

233. Al-Awqati, Q. Alternative treatment for secretory diarrhea revealed in a new class of CFTR inhibitors. *J. Clin. Invest.* **110**, 1599–1601 (2002).

234. Voss, B. J. & Stephen Trent, M. LPS Transport: Flipping Out over MsbA. *Curr. Biol.* **28**, R30–R33 (2018).

235. May, J. M., Sherman, D. J., Simpson, B. W., Ruiz, N. & Kahne, D. Lipopolysaccharide transport to the cell surface: periplasmic transport and assembly into the outer membrane. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20150027 (2015).

236. Dong, H., Zhang, Z., Tang, X., Paterson, N. G. & Dong, C. Structural and functional insights into the lipopolysaccharide ABC transporter LptB2FG. *Nat. Commun.* **8**, 1–11 (2017).

237. Padayatti, P. S. *et al.* Structural Insights into the Lipid A Transport Pathway in MsbA. *Structure* **27**, 1114-1123.e3 (2019).

238. Ho, H. *et al.* Structural basis for dual-mode inhibition of the ABC transporter MsbA. *Nature* **557**, 196–201 (2018).

239. Schirmer, T., Keller, T. A., Wang, Y. F. & Rosenbusch, J. P. Structural basis for sugar translocation through maltoporin channels at 3.1 Å resolution. *Science (80-.).* **267**, 512–514 (1995).

240. Licht, A. *et al.* Structural and functional characterization of a maltose/maltodextrin ABC transporter comprising a single solute binding domain (MalE) fused to the transmembrane subunit MalF. *Res. Microbiol.* **170**, 1–12 (2019).

241. Quiocho, F. A., Spurlino, J. C. & Rodseth, L. E. Extensive features of tight oligosaccharide

binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure* **5**, 997–1015 (1997).

242. Mächtel, R., Narducci, A., Griffith, D. A., Cordes, T. & Orelle, C. An integrated transport mechanism of the maltose ABC importer. *Res. Microbiol.* **170**, 321–337 (2019).

243. Oldham, M. L., Khare, D., Quiocho, F. A., Davidson, A. L. & Chen, J. Crystal structure of a catalytic intermediate of the maltose transporter. *Nature* **450**, 515–521 (2007).

244. Chen, S., Oldham, M. L., Davidson, A. L. & Chen, J. Carbon catabolite repression of the maltose transporter revealed by X-ray crystallography. *Nature* **499**, 364–368 (2013).

245. Khare, D., Oldham, M. L., Orelle, C., Davidson, A. L. & Chen, J. Alternating Access in Maltose Transporter Mediated by Rigid-Body Rotations. *Mol. Cell* **33**, 528–536 (2009).

246. Leach, A. . *Molecular Modelling: Principles and Applications*. (Peason Education Limited, 2001).

247. Bergeton, P. R. *The Physical Basis of Biochemistry*. (Springer International Publishing, 2010).

248. Morrell, W. E. & Hildebrand, J. H. The distribution of molecules in a model liquid. *J. Chem. Phys.* **4**, 224–227 (1936).

249. Metropolis, N. Metropolis-Et-Al-1953. *The Journal of chemical physics* vol. 21 1087 (1953).

250. Levitt, M. & Warshel, A. Computer simulation of protein folding. *Nature* **253**, 694–698 (1975).

251. McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585–590 (1977).

252. Prasad, S. *et al.* Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living J. Comput. Mol. Sci.* **1**, 1–28 (2018).

253. Born, M. & Oppenheimer, R. J. On the quantum theory of molecules (English translation). *Ann. Phys.* **457**, 1–32 (1927).

254. Periole Xavier, Marrink. S.J. Biomolecular Simulations. *Methods Mol. Biol.* **924**, 533–565 (2013).

255. González, M. A. Force fields and molecular dynamics simulations. *Collect. SFN* **12**, 169–

200 (2011).

256. Schmid, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J* **40**, 843–856 (2011).

257. Abraham M.J., van der Spoel D., Lindahl E., H. B. and the G. development team. *GROMACS User Manual version 5.0.7*. (www.gromacs.org, 2015).

258. Miyamoto, S. & Kollman, P. A. Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* **13**, 952–962 (1992).

259. Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **4**, 116–122 (2008).

260. Allen, Tildesley, D.J., M. P. *Computer Simulation of Liquids*. (Oxford University Press, 1987).

261. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

262. Atkins, Peter; De Paula, J. *Physical Chemistry*. (Oxford University Press, 2006).

263. Frenkel, Daan ; Smit, B. *Understanding molecular simulation*. (Academic Press, 2002).

264. Hockney, R. W. POTENTIAL CALCULATION AND SOME APPLICATIONS. (1970).

265. Verlet, L. Computer 'Experiments' on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **159**, 98–103 (1967).

266. Born, M ; von Karman, T. Uber Schwingungen im Raumgittern. *Phys. Zeitschrift* **13**, 297 (1912).

267. Berendsen, H. J. C. C. *et al.* Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

268. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, (2007).

269. Harvey, S. C., Tan, R. K.-Z. & Cheatham, T. E. The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J. Comput. Chem.* **19**, 726–740 (2002).

270. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

271. Hill, T. L. *An introduction to statistical thermodynamics*. (Dover Publications, 1986).

272. Hünenberger, P. H. Thermostat algorithms for molecular dynamics simulations. *Adv. Polym. Sci.* **173**, 105–147 (2005).

273. *Free Energy Calculations - Theory and Applications in Chemistry and Biology*. (Springer International Publishing, 2007).

274. Zwanzig, R. W. High†ï¿½Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **22**, 1420–1426 (1954).

275. Shirts, M. R. & Pande, V. S. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J. Chem. Phys.* **122**, 144107 (2005).

276. Lu, N., Singh, J. K. & Kofke, D. A. Appropriate methods to combine forward and reverse free-energy perturbation averages. *J. Chem. Phys.* **118**, 2977–2984 (2003).

277. Postma, J. P. M., Berendsen, H. J. C. & Haak, J. R. Thermodynamics of Cavity Formation in Water. *Faraday Symp. Chem. Soc.* 55–67 (1982).

278. Straatsma, T. P., Berendsen, H. J. C. & Postma, J. P. M. Free energy of hydrophobic hydration: A molecular dynamics study of noble gases in water. *J. Chem. Phys.* **85**, 6720–6727 (1986).

279. Hermans, J. Simple analysis of noise and hysteresis in (slow-growth) free energy simulations. *J. Phys. Chem.* **95**, 9029–9032 (1991).

280. Hu, H., Yun, R. H. & Hermans, J. Reversibility of free energy simulations: Slow growth may have a unique advantage. (With a note on use of Ewald summation). *Mol. Simul.* **28**, 67–80 (2002).

281. Zuckerman, D. M. *Statistical physics of biomolecules: An introduction*. *Statistical Physics of Biomolecules: An Introduction* (2010). doi:10.1201/b18849.

282. Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **3**, 300–313 (1935).

283. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23**, 187–199 (1977).

284. Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 932–942 (2011).

285. Kästner, J. & Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: “Umbrella integration”. *J. Chem. Phys.* **123**, 144104 (2005).

286. Kästner, J. & Thiel, W. Analysis of the statistical error in umbrella sampling simulations by umbrella integration. *J. Chem. Phys.* **124**, 234106 (2006).

287. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).

288. Chothia, C. & Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**, 823–826 (1986).

289. Sali, A., Blundell, T. L., Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).

290. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).

291. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).

292. Braun, W. & Go, N. Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm. *J. Mol. Biol.* **186**, 611–626 (1985).

293. ŠAli, A. & Overington, J. P. Derivation of rules for comparative protein modeling from a database of protein structure alignments. *Protein Sci.* **3**, 1582–1596 (1994).

294. Baptista, A. M., Martel, P. J. & Soares, C. M. Simulation of Electron-Proton Coupling with a Monte Carlo Method: Application to Cytochrome c$_3$ Using Continuum Electrostatics. *Biophys. J.* **76**, 2978–2998 (1999).

295. Bashford, D. & Karplus, M. pKa’s of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* **29**, 10219–10225 (1990).

296. Baptista, A. M., Martel, P. J. & Petersen, S. B. Simulation of protein conformational freedom as a function of pH: constant-pH molecular dynamics using implicit titration. *Proteins Struct. Funct. Bioinforma.* **27**, 523–544 (1997).

297. Honig, B. & Nicholls, A. Classical electrostatics in biology and chemistry. *Science (80-.).* **268**, 1144 LP – 1149 (1995).

298. Gilson, M. K. Introduction to continuum electrostatics, with molecular applications. http://gilsonlab.umbi.umd.edu (2006).

299. Martel, P. J., Baptista, A. & Petersen, S. B. Protein Electrostatics. in (ed. El-Gewely, M. R. B. T.-B. A. R.) vol. 2 315–372 (Elsevier, 1996).

300. Baptista, A. M. & Soares, C. M. Some theoretical and computational aspects of the inclusion of proton isomerism in the protonation equilibrium of proteins. *J. Phys. Chem. B* **105**, 293–309 (2001).

301. Riordan, J. R. *et al.* Identification of the Cystic-Fibrosis Gene - Cloning and Characterization of Complementary-DNA. *Science (80-.).* **245**, 1066–1072 (1989).

302. Kerem, B. S. *et al.* Identification of the Cystic-Fibrosis Gene - Genetic-Analysis. *Science (80-.).* **245**, 1073–1080 (1989).

303. Rommens, J. M. *et al.* Identification of the Cystic-Fibrosis Gene - Chromosome Walking and Jumping. *Science (80-.).* **245**, 1059–1065 (1989).

304. Gadsby, D. C., Vergani, P. & Csanady, L. The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature* **440**, 477–483 (2006).

305. Csanady, L., Chan, K. W., Nairn, A. C. & Gadsby, D. C. Functional roles of nonconserved structural segments in CFTR's NH2-terminal nucleotide binding domain. *J Gen Physiol* **125**, 43–55 (2005).

306. Dalemans, W. *et al.* Altered chloride ion channel kinetics associated with the delta F508 cystic fibrosis mutation. *Nature* **354**, 526–528 (1991).

307. Jih, K. Y., Li, M., Hwang, T. C. & Bompadre, S. G. The most common cystic fibrosis-associated mutation destabilizes the dimeric state of the nucleotide-binding domains of CFTR. *J. Physiol.* **589**, 2719–2731 (2011).

308. Wellhauser, L. *et al.* A Small-Molecule Modulator Interacts Directly with Delta Phe508-CFTR to Modify Its ATPase Activity and Conformational Stability. *Mol. Pharmacol.* **75**, 1430–1438 (2009).

309. Chiaw, P. K., Wellhauser, L., Huan, L. J., Ramjeesingh, M. & Bear, C. E. A Chemical Corrector Modifies the Channel Function of F508del-CFTR. *Mol. Pharmacol.* **78**, 411–418 (2010).

310. Zhang, Z. & Chen, J. Atomic Structure of the Cystic Fibrosis Transmembrane Conductance

Regulator. *Cell* **167**, 1586-1597.e9 (2016).

311. Zhang, Z., Liu, F. & Chen, J. Conformational Changes of CFTR upon Phosphorylation and ATP Binding. *Cell* **170**, 483-491.e8 (2017).

312. Fay, J. F. *et al.* Cryo-EM Visualization of an Active High Open Probability CFTR Anion Channel. *Biochemistry* (2018) doi:10.1021/acs.biochem.8b00763.

313. Vergani, P., Lockless, S. W., Nairn, A. C. & Gadsby, D. C. Opening/closing of the CFTR Cl- channel is linked to dynamic dimerization/dissociation of its two nucleotide binding domains. *Biophys J* **88**, 612a-612a (2005).

314. Scott-Ward, T. S. *et al.* Chimeric constructs endow the human CFTR Cl- channel with the gating behavior of murine CFTR. *Proc Natl Acad Sci U S A* **104**, 16365–16370 (2007).

315. Bashford, D. & Gerwert, K. Electrostatic Calculations of the Pka Values of Ionizable Groups in Bacteriorhodopsin. *J. Mol. Biol.* **224**, 473–486 (1992).

316. Oliveira, A. S., Baptista, A. M. & Soares, C. M. Conformational changes induced by ATP-hydrolysis in an ABC transporter: A molecular dynamics study of the Sav1866 exporter. *Proteins-Structure Funct. Bioinforma.* **79**, 1977–1990 (2011).

317. Damas, J. M., Oliveira, A. S. F., Baptista, A. M. & Soares, C. M. Structural consequences of ATP hydrolysis on the ABC transporter NBD dimer: Molecular dynamics studies of HlyB. *Protein Sci.* **20**, 1220–1230 (2011).

318. Páll, S., Abraham, M. J., Kutzner, C., Hess, B. & Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. in *Solving Software Challenges for Exascale: International Conference on Exascale Applications and Software, EASC 2014, Stockholm, Sweden, April 2-3, 2014, Revised Selected Papers* (eds. Markidis, S. & Laure, E.) 3–27 (Springer International Publishing, 2015). doi:10.1007/978-3-319-15976-8_1.

319. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

320. Hermans, J., Berendsen, H. J. C., Vangunsteren, W. F. & Postma, J. P. M. A Consistent Empirical Potential for Water-Protein Interactions. *Biopolymers* **23**, 1513–1518 (1984).

321. Naveen, M., J., D. E., B., W. T. & Oliver, B. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).

322. Gowers, R. J. *et al.* MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. in *Proceedings of the 15th Python in Science Conference* (eds. Benthall, S. & Rostrup, S.) 98–105 (2016).

323. Ciccotti, G., Jacucci, G. & McDonald, I. R. "Thought-experiments" by molecular dynamics. *J. Stat. Phys.* **21**, 1–22 (1979).

324. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

325. Campos, S. R. & Baptista, A. M. Conformational analysis in a multidimensional energy landscape: study of an arginylglutamate repeat. *J Phys Chem B* **113**, 15989–16001 (2009).

326. Efron, B. 1977 Rietz Lecture - Bootstrap Methods - Another Look at the Jackknife. *Ann. Stat.* **7**, 1–26 (1979).

327. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).

328. Sorum, B., Töröcsik, B. & Csanády, L. Asymmetry of movements in CFTR's two ATP sites during pore opening serves their distinct functions. *Elife* **6**, e29013 (2017).

329. Cai, Z., Liu, J., Li, H. & Sheppard, D. N. Targeting F508del-CFTR to develop rational new therapies for cystic fibrosis. *Acta Pharmacol Sin* **32**, 693–701 (2011).

330. Ramjeesingh, M. *et al.* Walker Mutations Reveal Loose Relationship between Catalytic and Channel-Gating Activities of Purified CFTR (Cystic Fibrosis Transmembrane Conductance Regulator). *Biochemistry* **38**, 1463–1468 (1999).

331. Bompadre, S. G. *et al.* CFTR Gating II. *J. Gen. Physiol.* (2005) doi:10.1085/jgp.200409228.

332. Zeltwanger, S., Wang, F., Wang, G. T., Gillis, K. D. & Hwang, T. C. Gating of cystic fibrosis transmembrane conductance regulator chloride channels by adenosine triphosphate hydrolysis. Quantitative analysis of a cyclic gating scheme. *J. Gen. Physiol.* (1999) doi:10.1085/jgp.113.4.541.

333. Buchenberg, S., Sittel, F. & Stock, G. Time-resolved observation of protein allosteric communication. *Proc. Natl. Acad. Sci.* **114**, E6804 LP-E6811 (2017).

334. He, L. *et al.* Multiple membrane-cytoplasmic domain contacts in the cystic fibrosis transmembrane conductance regulator (CFTR) mediate regulation of channel gating. *J. Biol. Chem.* **283**, 26383–26390 (2008).

335. Lewis, H. A. *et al.* Impact of the Delta F508 mutation in first nucleotide-binding domain of human cystic fibrosis transmembrane conductance regulator on domain folding and structure. *J. Biol. Chem.* **280**, 1346–1353 (2005).

336. Das, J. *et al.* Transmembrane helical interactions in the CFTR channel pore. *PLoS Comput. Biol.* **13**, e1005594–e1005594 (2017).

337. Belmonte, L. & Moran, O. On the interactions between nucleotide binding domains and membrane spanning domains in cystic fibrosis transmembrane regulator: A molecular dynamic study. *Biochimie* **111**, 19–29 (2015).

338. Odera, M., Furuta, T., Sohma, Y. & Sakurai, M. Molecular dynamics simulation study on the structural instability of the most common cystic fibrosis-associated mutant ΔF508-CFTR. *Biophys. Physicobiology* **15**, 33–44 (2018).

339. Miki, H., Zhou, Z., Li, M., Hwang, T. C. & Bompadre, S. G. Potentiation of disease-associated cystic fibrosis transmembrane conductance regulator mutants by hydrolyzable ATP analogs. *J. Biol. Chem.* **285**, 19967–19975 (2010).

340. Doerrler, W. T., Reedy, M. C. & Raetz, C. R. H. An Escherichia coli Mutant Defective in Lipid Export . *J. Biol. Chem.* **276**, 11461–11464 (2001).

341. Doerrler, W. T. & Raetz, C. R. H. ATPase Activity of the MsbA Lipid Flippase of Escherichia coli . *J. Biol. Chem.* **277**, 36697–36705 (2002).

342. Randak, C. O., Dong, Q., Ver Heul, A. R., Elcock, A. H. & Welsh, M. J. ATP and AMP Mutually Influence Their Interaction with the ATP-binding Cassette (ABC) Adenylate Kinase Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) at Separate Binding Sites. *J. Biol. Chem.* **288**, 27692–27701 (2013).

343. Williams, R. S. & Tainer, J. A. Learning Our ABCs: Rad50 Directs MRN Repair Functions via Adenylate Kinase Activity from the Conserved ATP Binding Cassette. *Mol. Cell* **25**, 789–791 (2007).

344. Kaiser, A. *et al.* Unwinding of the C-Terminal Residues of Neuropeptide Y is critical for Y2 Receptor Binding and Activation. *Angew. Chemie Int. Ed.* **54**, 7446–7449 (2015).

345. Kaur, J. *et al.* Solid-state NMR analysis of the sodium pump Krokinobacter rhodopsin 2 and its H30A mutant. *J. Struct. Biol.* **206**, 55–65 (2019).

346. Lange, A. *et al.* Toxin-induced conformational changes in a potassium channel revealed

by solid-state NMR. *Nature* **440**, 959–962 (2006).

347. Mandala, V. S., Gelenter, M. D. & Hong, M. Transport-Relevant Protein Conformational Dynamics and Water Dynamics on Multiple Time Scales in an Archetypal Proton Channel: Insights from Solid-State NMR. *J. Am. Chem. Soc.* **140**, 1514–1524 (2018).

348. Schmidt, P., Thomas, L., Müller, P., Scheidt, H. A. & Huster, D. The G-Protein-Coupled Neuropeptide Y Receptor Type 2 is Highly Dynamic in Lipid Membranes as Revealed by Solid-State NMR Spectroscopy. *Chem. – A Eur. J.* **20**, 4986–4992 (2014).

349. Ullrich, S. J., Hellmich, U. A., Ullrich, S. & Glaubitz, C. Interfacial enzyme kinetics of a membrane bound kinase analyzed by real-time MAS-NMR. *Nat. Chem. Biol.* **7**, 263–270 (2011).

350. Zhao, Y. *et al.* Gating Mechanism of Aquaporin Z in Synthetic Bilayers and Native Membranes Revealed by Solid-State NMR Spectroscopy. *J. Am. Chem. Soc.* **140**, 7885–7895 (2018).

351. Lilly Thankamony, A. S., Wittmann, J. J., Kaushik, M. & Corzilius, B. Dynamic nuclear polarization for sensitivity enhancement in modern solid-state NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **102**–**103**, 120–195 (2017).

352. Becker-Baldus, J. *et al.* Enlightening the photoactive site of channelrhodopsin-2 by DNP-enhanced solid-state NMR spectroscopy. *Proc. Natl. Acad. Sci.* **112**, 9896 LP – 9901 (2015).

353. Joedicke, L. *et al.* The molecular basis of subtype selectivity of human kinin G-protein-coupled receptors. *Nat. Chem. Biol.* **14**, 284–290 (2018).

354. Maciejko, J. *et al.* Visualizing Specific Cross-Protomer Interactions in the Homo-Oligomeric Membrane Protein Proteorhodopsin by Dynamic-Nuclear-Polarization-Enhanced Solid-State NMR. *J. Am. Chem. Soc.* **137**, 9032–9043 (2015).

355. Ni, Q. Z. *et al.* Primary Transfer Step in the Light-Driven Ion Pump Bacteriorhodopsin: An Irreversible U-Turn Revealed by Dynamic Nuclear Polarization-Enhanced Magic Angle Spinning NMR. *J. Am. Chem. Soc.* **140**, 4085–4091 (2018).

356. Sergeyev, I. V, Itin, B., Rogawski, R., Day, L. A. & McDermott, A. E. Efficient assignment and NMR analysis of an intact virus using sequential side-chain correlations and DNP sensitization. *Proc. Natl. Acad. Sci.* **114**, 5171 LP – 5176 (2017).

357. Kaplan, M. *et al.* Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR. *Nat. Methods* **12**, 649–652 (2015).

358. Lange, V. *et al.* A MAS NMR Study of the Bacterial ABC Transporter ArtMP. *ChemBioChem* **11**, 547–555 (2010).

359. Kunert, B. *et al.* Efficient and stable reconstitution of the ABC transporter BmrA for solid-state NMR studies. *Frontiers in molecular biosciences* vol. 1 5 (2014).

360. Lehnert, E. *et al.* Antigenic Peptide Recognition on the Human ABC Transporter TAP Resolved by DNP-Enhanced Solid-State NMR Spectroscopy. *J. Am. Chem. Soc.* **138**, 13967–13974 (2016).

361. Spadaccini, R., Kaur, H., Becker-Baldus, J. & Glaubitz, C. The effect of drug binding on specific sites in transmembrane helices 4 and 6 of the ABC exporter MsbA studied by DNP-enhanced solid-state NMR. *Biochim. Biophys. Acta - Biomembr.* **1860**, 833–840 (2018).

362. Wiegand, T. *et al.* Solid-state NMR and EPR Spectroscopy of Mn2+-Substituted ATP-Fueled Protein Engines. *Angew. Chemie Int. Ed.* **56**, 3369–3373 (2017).

363. Wiegand, T. *et al.* Protein–nucleotide contacts in motor proteins detected by DNP-enhanced solid-state NMR. *J. Biomol. NMR* **69**, 157–164 (2017).

364. Marchanka, A., Simon, B. & Carlomagno, T. A Suite of Solid-State NMR Experiments for RNA Intranucleotide Resonance Assignment in a 21 kDa Protein–RNA Complex. *Angew. Chemie Int. Ed.* **52**, 9996–10001 (2013).

365. Middleton, D. A., Hughes, E. & Esmann, M. The Conformation of ATP within the Na,K-ATPase Nucleotide Site: A Statistically Constrained Analysis of REDOR Solid-State NMR Data. *Angew. Chemie Int. Ed.* **50**, 7041–7044 (2011).

366. Westfahl, K. M., Merten, J. A., Buchaklian, A. H. & Klug, C. S. Functionally Important ATP Binding and Hydrolysis Sites in Escherichia coli MsbA. *Biochemistry* **47**, 13878–13886 (2008).

367. Hellmich, U. A. *et al.* Probing the ATP hydrolysis cycle of the abc multidrug transporter LmrA by pulsed EPR spectroscopy. *J. Am. Chem. Soc.* **134**, 5857–5862 (2012).

368. Collauto, A., Mishra, S., Litvinov, A., Mchaourab, H. S. & Goldfarb, D. Direct Spectroscopic Detection of ATP Turnover Reveals Mechanistic Divergence of ABC Exporters. *Structure*

25, 1264-1274.e3 (2017).

369. Kaur, H. *et al.* The ABC exporter MsbA probed by solid state NMR – challenges and opportunities. *Biol. Chem.* **396**, 1135–1149 (2015).

370. Sauvée, C. *et al.* Highly Efficient, Water-Soluble Polarizing Agents for Dynamic Nuclear Polarization at High Frequency. *Angew. Chemie Int. Ed.* **52**, 10858–10861 (2013).

371. Jaroniec, C. P., Filip, C. & Griffin, R. G. 3D TEDOR NMR Experiments for the Simultaneous Measurement of Multiple Carbon−Nitrogen Distances in Uniformly 13C,15N-Labeled Solids. *J. Am. Chem. Soc.* **124**, 10728–10742 (2002).

372. Milikisyants, S., Scarpelli, F., Finiguerra, M. G., Ubbink, M. & Huber, M. A pulsed EPR method to determine distances between paramagnetic centers with strong spectral anisotropy and radicals: The dead-time free RIDME sequence. *J. Magn. Reson.* **201**, 48–56 (2009).

373. Smith, P. E. & Vangunsteren, W. F. Consistent Dielectric-Properties of the Simple Point-Charge and Extended Simple Point-Charge Water Models at 277 and 300 K. *J. Chem. Phys.* **100**, 3169–3174 (1994).

374. Jarvis, R. A. & Patrick, E. A. Clustering Using a Similarity Measure Based on Shared Near Neighbors. *IEEE Trans. Comput.* **C–22**, 1025–1034 (1973).

375. Eckford, P. D. W. & Sharom, F. J. Functional Characterization of Escherichia coli MsbA: INTERACTION WITH NUCLEOTIDES AND SUBSTRATES . *J. Biol. Chem.* **283**, 12840–12850 (2008).

376. Ulrich, E. L. *et al.* BioMagResBank. *Nucleic Acids Res.* **36**, D402–D408 (2007).

377. Dawson, R. J. P. & Locher, K. P. Structure of a bacterial multidrug ABC transporter. *Nature* **443**, 180–185 (2006).

378. Oldham, M. L., Davidson, A. L. & Chen, J. Structural insights into ABC transporter mechanism. *Curr. Opin. Struct. Biol.* **18**, 726–733 (2008).

379. Kulik, L. V, Dzuba, S. A., Grigoryev, I. A. & Tsvetkov, Y. D. Electron dipole–dipole interaction in ESEEM of nitroxide biradicals. *Chem. Phys. Lett.* **343**, 315–324 (2001).

380. Akhmetzyanov, D. *et al.* RIDME spectroscopy on high-spin Mn2+ centers. *Phys. Chem. Chem. Phys.* **18**, 30857–30866 (2016).

381. Collauto, A. *et al.* RIDME distance measurements using Gd(iii) tags with a narrow central transition. *Phys. Chem. Chem. Phys.* **18**, 19037–19049 (2016).

382. Keller, K. *et al.* EPR characterization of Mn(ii) complexes for distance determination with pulsed dipolar spectroscopy. *Phys. Chem. Chem. Phys.* **18**, 25120–25135 (2016).

383. Razzaghi, S. *et al.* RIDME Spectroscopy with Gd(III) Centers. *J. Phys. Chem. Lett.* **5**, 3970–3975 (2014).

384. Lu, G., Westbrooks, J. M., Davidson, A. L. & Chen, J. ATP hydrolysis is required to reset the ATP-binding cassette dimer into the resting-state conformation. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 17969 LP – 17974 (2005).

385. Oswald, C., Holland, I. B. & Schmitt, L. The motor domains of ABC-transporters. *Naunyn. Schmiedebergs. Arch. Pharmacol.* **372**, 385–399 (2006).

386. Dong, Q. *et al.* Mutating the Conserved Q-loop Glutamine 1291 Selectively Disrupts Adenylate Kinase-dependent Channel Gating of the ATP-binding Cassette (ABC) Adenylate Kinase Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) and Reduces Channel Function in Prima. *J. Biol. Chem.* **290**, 14140–14153 (2015).

387. Dong, Q., Randak, C. O. & Welsh, M. J. A Mutation in CFTR Modifies the Effects of the Adenylate Kinase Inhibitor Ap5A on Channel Gating. *Biophys. J.* **95**, 5178–5185 (2008).

388. Yaginuma, H. *et al.* Diversity in ATP concentrations in a single bacterial cell population revealed by quantitative single-cell imaging. *Sci. Rep.* **4**, 6522 (2014).

389. Berg, J., Hung, Y. P. & Yellen, G. A genetically encoded fluorescent reporter of ATP:ADP ratio. *Nat. Methods* **6**, 161–166 (2009).

390. Zhou, Z., White, K. A., Polissi, A., Georgopoulos, C. & Raetz, C. R. H. Function of Escherichia coli MsbA, an Essential ABC Family Transporter, in Lipid A and Phospholipid Biosynthesis . *J. Biol. Chem.* **273**, 12466–12475 (1998).

391. Doerrler, W. T., Gibbons, H. S. & Raetz, C. R. H. MsbA-dependent Translocation of Lipids across the Inner Membrane of Escherichia coli. *J. Biol. Chem.* **279**, 45102–45109 (2004).

392. Doerrler, W. T. Lipid trafficking to the outer membrane of Gram-negative bacteria. *Mol. Microbiol.* **60**, 542–552 (2006).

393. Choudhury, H. G. *et al.* Structure of an antibacterial peptide ATP-binding cassette transporter in a novel outward occluded state. *Proc. Natl. Acad. Sci.* **111**, 9145 LP – 9150

(2014).

394. Reed, G. H. & Markham, G. D. EPR of Mn(II) Complexes with Enzymes and Other Proteins BT - Biological Magnetic Resonance: Volume 6. in (eds. Berliner, L. J. & Reuben, J.) 73–142 (Springer US, 1984). doi:10.1007/978-1-4615-6546-8_3.

395. Mao, J. *et al.* Host–Guest Complexes as Water-Soluble High-Performance DNP Polarizing Agents. *J. Am. Chem. Soc.* **135**, 19275–19281 (2013).

396. Jeschke, G. *et al.* DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.* **30**, 473–498 (2006).

397. Marley, J., Lu, M. & Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR* **20**, 71–75 (2001).

398. Bigay, J., Deterre, P., Pfister, C. & Chabre, M. Fluoride complexes of aluminium or beryllium act on G-proteins as reversibly bound analogues of the gamma phosphate of GTP. *EMBO J.* **6**, 2907–2913 (1987).

399. Smith, R. L., Zinn, K. & Cantley, L. C. A study of the vanadate-trapped state of the (Na,K)-ATPase. Evidence against interacting nucleotide site models. *J. Biol. Chem.* **255**, 9852–9859 (1980).

400. Meagher, K. L., Redman, L. T. & Carlson, H. A. Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.* **24**, 1016–1025 (2003).

401. Hwang, M.-J., Chu, P.-Y., Chen, J.-C. & Chao, I. Conformational analysis of three pyrophosphate model species: Diphosphate, methyl diphosphate, and triphosphate. *J. Comput. Chem.* **20**, 1702–1715 (1999).

402. Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269–10280 (1993).

403. Tanaka, K. J., Song, S., Mason, K. & Pinkett, H. W. Selective substrate uptake: The role of ATP-binding cassette (ABC) importers in pathogenesis. *Biochim. Biophys. Acta - Biomembr.* 0–1 (2017) doi:10.1016/j.bbamem.2017.08.011.

404. Bao, H. & Duong, F. ATP Alone Triggers the Outward Facing Conformation of the Maltose ATP-binding Cassette Transporter. *J. Biol. Chem.* **288**, 3439–3448 (2013).

405. Alvarez, F. J. D. *et al.* Full engagement of liganded maltose-binding protein stabilizes a

semi-open ATP-binding cassette dimer in the maltose transporter. *Mol. Microbiol.* **98**, 878–894 (2015).

406. Bao, H., Dalal, K., Cytrynbaum, E. & Duong, F. Sequential Action of MalE and Maltose Allows Coupling ATP Hydrolysis to Translocation in the MalFGK2 Transporter. *J. Biol. Chem.* **290**, 25452–25460 (2015).

407. Hsu, W.-L., Furuta, T. & Sakurai, M. The mechanism of nucleotide-binding domain dimerization in the intact maltose transporter as studied by all-atom molecular dynamics simulations. *Proteins Struct. Funct. Bioinforma.* **86**, 237–247 (2018).

408. Su, J. G. *et al.* Analysis of conformational motions and related key residue interactions responsible for a specific function of proteins with elastic network model. *J. Biomol. Struct. Dyn.* **34**, 560–571 (2016).

409. Orelle, C., Ayvaz, T., Everly, R. M., Klug, C. S. & Davidson, A. L. Both maltose-binding protein and ATP are required for nucleotide-binding domain closure in the intact maltose ABC transporter. **105**, 12837–12842 (2008).

410. Schmidt, T. H. & Kandt, C. LAMBADA and InflateGRO2: Efficient Membrane Alignment and Insertion of Membrane Proteins for Molecular Dynamics Simulations. (2012) doi:10.1021/ci3000453.

411. Bashford, D. Macroscopic electrostatic models for protonation states in proteins. *Front. Biosci.* **9**, 1082–1099 (2004).

412. Poger, D. & Mark, A. E. On the Validation of Molecular Dynamics Simulations of Saturated and cis-Monounsaturated Phosphatidylcholine Lipid Bilayers: A Comparison with Experiment. *J. Chem. Theory Comput.* **6**, 325–336 (2010).

413. Hansen, H. S. & Hünenberger, P. H. A reoptimized GROMOS force field for hexopyranose-based carbohydrates accounting for the relative free energies of ring conformers, anomers, epimers, hydroxymethyl rotamers, and glycosidic linkage conformers. *J. Comput. Chem.* **32**, 998–1032 (2011).

414. Damas, J. M. *et al.* Predicting the Thermodynamics and Kinetics of Helix Formation in a Cyclic Peptide Model. *J. Chem. Theory Comput.* **9**, 5148–5157 (2013).

415. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).

416. Miyamoto, S. & Kollman, P. A. Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* **13**, 952–962 (1992).

417. Lindorff-Larsen, K. & Ferkinghoff-Borg, J. Similarity measures for protein ensembles. *PLoS One* **4**, (2009).

418. Tiberti, M., Papaleo, E., Bengtsen, T., Boomsma, W. & Lindorff-Larsen, K. ENCORE: Software for Quantitative Ensemble Comparison. *PLoS Comput. Biol.* **11**, 1–16 (2015).

419. Smart, O. S., Goodfellow, J. M. & Wallace, B. A. The pore dimensions of gramicidin A. *Biophys. J.* **65**, 2455–2460 (1993).

420. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S. & Thornton, J. M. PDBsum: Structural summaries of PDB entries. *Protein Sci.* **27**, 129–134 (2018).

421. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

422. Williams, T., Kelley, C. & many others. Gnuplot 4.6: an interactive plotting program. (2013).

423. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

424. Stroet, M. Umbrella integration: Initial version. (2016) doi:10.5281/ZENODO.164996.

425. Jacso, T., Schneider, E., Rupp, B. & Reif, B. Substrate-transport activation is mediated through the second periplasmic loop of the transmembrane protein MalF in the maltose transport complex of Escherichia coli. *J. Biol. Chem.* (2012) doi:10.1074/jbc.M112.340679.

426. Grote, M. *et al.* Transmembrane Signaling in the Maltose ABC Transporter MalFGK2E: Periplasmic MalF-P2 LOOP communicates substrate availability to the ATP-bound MalK dimer. *J. Biol. Chem.* **284**, 17521–17526 (2009).

427. Ehrle, R., Pick, C., Ulrich, R., Hofmann, E. & Ehrmann, M. Characterization of transmembrane domains 6, 7, and 8 of MalF by mutational analysis. *J. Bacteriol.* **178**, 2255 LP – 2262 (1996).

428. Steinke, A., Grau, S., Davidson, A., Hofmann, E. & Ehrmann, M. Characterization of Transmembrane Segments 3, 4, and 5 of MalF by Mutational Analysis. *J. Bacteriol.* **183**, 375 LP – 381 (2001).

429. Oldham, M. L., Chen, S. & Chen, J. Structural basis for substrate specificity in the Escherichia coli maltose transport system. *Proc Natl Acad Sci U S A* **110**, 18132–18137 (2013).

430. Nelson, B. D. & Traxler, B. Exploring the Role of Integral Membrane Proteins in ATP-Binding Cassette Transporters: Analysis of a Collection of MalG Insertion Mutants. *J. Bacteriol.* **180**, 2507 LP – 2514 (1998).

431. Quistgaard, E. M., Löw, C., Guettou, F. & Nordlund, P. Understanding transport by the major facilitator superfamily (MFS): Structures pave the way. *Nat. Rev. Mol. Cell Biol.* **17**, 123–132 (2016).

432. Park, M. Molecular Dynamics Simulations of the Human Glucose Transporter GLUT1. 1–18 (2015) doi:10.1371/journal.pone.0125361.

433. Sheena, A., Mohan, S. S., Haridas, N. P. A. & Anilkumar, G. Elucidation of the Glucose Transport Pathway in Glucose Transporter 4 via Steered Molecular Dynamics Simulations. **6**, (2011).

434. Ranquin, A. & Gelder, P. Van. Maltoporin : sugar for physics and biology. **155**, 611–616 (2004).

435. Danelon, C. Probing the Orientation of Reconstituted Maltoporin Channels at the Single-protein Level *. **278**, 35542–35551 (2003).

436. Dutzler, R., Schirmer, T., Karplus, M., Fischer, S. & Pasteur, L. Translocation Mechanism of Long Sugar Chains across the Maltoporin Membrane Channel. **10**, 1273–1284 (2002).

437. Aittoniemi, J. *et al.* Review. SUR1: a unique ATP-binding cassette protein that functions as an ion channel regulator. *Philos Trans R Soc L. B Biol Sci* **364**, 257–267 (2009).

438. O'Sullivan, B. & Freedman, S. Carrier screening for cystic fibrosis reply. *Lancet* **374**, 978 (2009).

439. Cottrill, K. A., Farinha, C. M. & McCarty, N. A. The bidirectional relationship between CFTR and lipids. *Commun. Biol.* **3**, 1–10 (2020).

The work presented in this thesis was financed by Fundação para a Ciência e a Tecnologia, through grant SFRH/BD/108002/2015.