UNIVERSITA' DEGLI STUDI DI ROMA

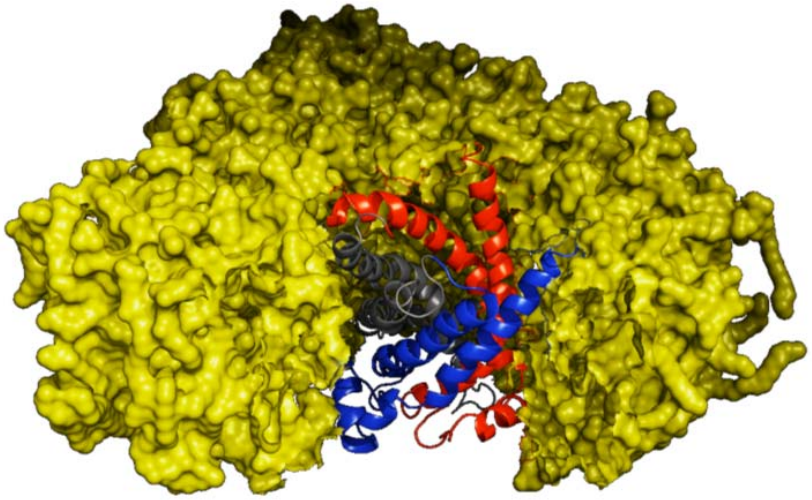
TorVergata

FACOLTÀ DI SCIENZE MATEMATICHE FISICHE E NATURALI

DOTTORATO DI RICERCA IN BIOLOGIA CELLULARE E MOLECOLARE

XXII CICLO

Molecular dynamics and docking simulations of the ADP/ATP mitochondrial carrier: structural-dynamical insights for the inactivation of pathological mutants and detection of potential ATP binding sites.



Docente Guida: **Falconi Mattia**

Coordinatore: **Cesareni Gianni**

Dottorando: **Di Marino Daniele**

– Index

Chapter 1
– Introduction

1.1 The mitochondrion

The mitochondrion is a membrane-enclosed organelle found in most eukaryotic cells. These organelles range from 0.5–10 micrometers (μm) in diameter. Mitochondria are sometimes described as "cellular power plants" because they generate most of the cell's supply of adenosine triphosphate (ATP), used as a source of the chemical energy. In addition to supplying cellular energy, mitochondria are involved in a range of other processes, such as signalling, cellular differentiation, cell death, as well as the control of the cell cycle and cell growth (McBride et al., 2006). Mitochondria have been implicated in several human diseases, including mitochondrial disorders (Gardner and Boles, 2005) and cardiac dysfunction, (Lesnefsky et al., 2001) and may play a role in the aging process. Several characteristics make

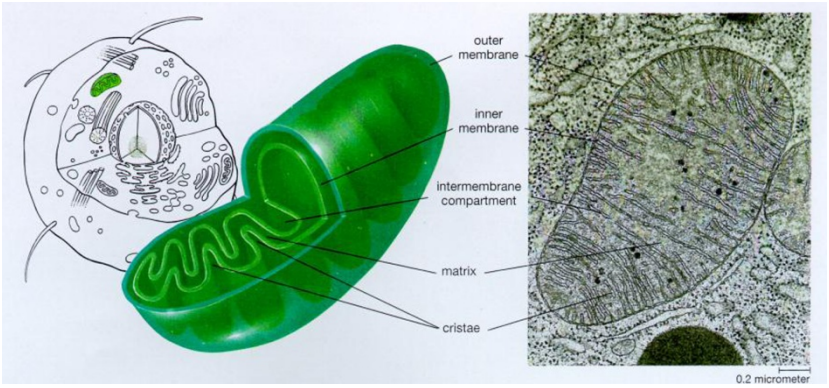


Figure 1. Schematization of the mitochondrion. In the green representation and in the photography at the transmission electron microscope are evident the two membranes, the compartments and the ridges.

mitochondria unique. The number of mitochondria in a cell varies widely by organism and tissue type. The organelle is composed of compartments that carry out specialized functions. These compartments or regions include the

outer membrane, the intermembrane space, the inner membrane, and the cristae and matrix (Fig. 1) .

1.2 The mitochondrion structure

A mitochondrion contains outer and inner membranes composed of phospholipids bilayer and proteins. The two membranes, however, have different properties. Because of this double-membrane organization, there are five distinct compartments within the mitochondrion. There is the outer mitochondrial membrane, the intermembrane space (the space between the outer and inner membranes), the inner mitochondrial membrane, the cristae space and the matrix.

1.2.1 The outer mitochondrial membrane

The outer mitochondrial membrane, which encloses the entire organelle, has a proteins-to-phospholipids ratio similar to that of the eukaryotic plasma membrane (about 1:1 by weight). It contains large numbers of integral proteins called porins. These porins form channels that allow molecules 5000 Daltons or less in molecular weight to freely diffuse from one side of the membrane to the other. Larger proteins can enter the mitochondrion if a signalling sequence at their N-terminus binds to a large multi-subunit protein called translocase of the outer membrane, which then actively moves them across the membrane (Herrmann and Neupert, 2000). Disruption of the outer membrane permits proteins in the intermembrane space to leak into the cytosol, leading to certain cell death (Chipuk et al., 2006). The mitochondrial outer membrane can associate with the ER membrane, in a structure called MAM (mitochondria-associated ER-membrane). This is important in ER-mitochondria calcium signalling and involved in the transfer of lipids between the ER and mitochondria (Hayashi et al., 2009).

1.2.2 The mitochondrial intermembrane space

The intermembrane space is basically the space between the outer membrane and the inner membrane. Because the outer membrane is freely permeable to small molecules, the concentrations of small molecules such as ions and sugars in the intermembrane space is the same as the cytosol (Alberts

et al., 1994). However, as large proteins must have a specific signalling sequence to be transported across the outer membrane, the protein composition of this space is different than the protein composition of the cytosol. One protein that is localized to the intermembrane space in this way is cytochrome c.

1.2.3 The inner mitochondrial membrane

The inner mitochondrial membrane contains proteins with five types of functions:

1. Those that perform the redox reactions of oxidative phosphorylation.

2. ATP synthase, which generates ATP in the matrix.

3. Specific transport proteins that regulate metabolite passage into and out of the matrix.

4. Protein import machinery.

5. Mitochondria fusion and fission protein.

It contains more than 100 different polypeptides, and has a very high proteins-to-phospholipids ratio (more than 3:1 by weight, which is about 1 protein for 15 phospholipids). The inner membrane is home to around 1/5 of the total protein in a mitochondrion (Alberts et al., 1994). In addition, the inner membrane is rich in an unusual phospholipids, cardiolipin. This phospholipid was originally discovered in beef hearts in 1942, and is usually characteristic of mitochondrial and bacterial plasma membranes (McMillin and Dowhan, 2002). Cardiolipin contains four fatty acids rather than two and may help to make the inner membrane impermeable (Alberts et al., 1994). Unlike the outer membrane, the inner membrane does not contain porins and is highly impermeable to all molecules. Almost all ions and molecules require special membrane transporters to enter or exit the matrix. Proteins are ferried into the matrix via the translocase of the inner membrane (TIM) complex or via Oxa1(Herrmann and Neupert, 2000). In addition, there is a membrane potential across the inner membrane formed by the action of the enzymes of the electron transport chain.

1.2.4 The mitochondrial cristae

The inner mitochondrial membrane is compartmentalized into numerous cristae, which expand the surface area of the inner mitochondrial membrane, enhancing its ability to produce ATP. These are not simple random folds but rather invaginations of the inner membrane, which can affect overall chemiosmotic function (Mannella, 2006). In typical liver mitochondria, for example, the surface area, including cristae, is about five times that of the outer membrane. Mitochondria of cells that have greater demand for ATP, such as muscle cells, contain more cristae than typical liver mitochondria. These folds are studded with small round bodies known as F1 particles or oxysomes.

1.2.5 The mitochondrial matrix

The matrix is the space enclosed by the inner membrane. It contains about 2/3 of the total protein in a mitochondrion (Alberts et al., 1994). The matrix is important in the production of ATP with the aid of the ATP synthase contained in the inner membrane. The matrix contains a highly-concentrated mixture of hundreds of enzymes, special mitochondrial ribosomes, tRNA, and several copies of the mitochondrial DNA genome.
Mitochondria have their own genetic material, and the machinery to manufacture their own RNAs and proteins. A published human mitochondrial DNA sequence revealed 16,569 base pairs encoding 37 total genes: 22 tRNA, 2 rRNA, and 13 peptide genes (Anderson et al., 1981). The 13 mitochondrial peptides in humans are integrated into the inner mitochondrial membrane, along with proteins encoded by genes that reside in the host cell's nucleus.

Chapter 2
– Mitochondrial carrier family

2.1 General features

Mitochondrial carriers are integral membrane proteins located in the inner mitochondrial membrane whose role is to selectively transport metabolites through the lipid bilayer. Human mitochondrial carriers are coded by nuclear genes called SLC25 and after their transduction they are translocated into the mitochondrial membrane (Palmieri, 2004). The role of the proteins belonging to this family is to help the communication between the mitochondrial matrix and the inner membrane space, by promoting the translocation of several substrates. This communication is essential for the progress of physiological processes carried out by reactions in both matrix and cytoplasm compartments (Palmieri, 2004).
Genetic or transductional defects of mitochondrial carriers are associated to the onset of genetic diseases such as Stanley syndrome, HHH syndrome, omocitrulline syndrome, Amish microencephaly, and can lead to physiologic disorders. The importance of the role of these proteins in cellular metabolism encouraged the study of their functioning at molecular level and of their structure-function relationship, even if many problems derive from the fact that membrane proteins are not easy to be solubilised and purified. Since the early 60s few laboratories started to hypothesize the presence of carriers in the inner mitochondrial membrane, able to catalyze the translocation of metabolites between cytosol and matrix. This hypothesis was finally confirmed at the end of the 60s, demonstrating the presence of carries, able to transport a series of substrates (Palmieri et al., 1996). Further, kinetic studies have been applied to understand the transport mechanism. Only in the 80s some mitochondrial carriers have been purified and reconstituted in artificial liposomes in order to study their activity.
Instead of early studies, when these proteins where extracted from mitochondria limiting the possibility of their study, the progresses in the field of genetics have enabled the production of recombinant transporters, better permitting their characterization and function analysis (Fiermonte et al., 1993). Nowadays, intensive studies are aimed to determine the 3D-structure of these proteins and to understand the molecular mechanism of transport (Palmieri et la., 2004).

2.1.1 Tissues distribution and metabolic function

Functional studies on mitochondria showed the presence of 16 translocation systems, not considering the isoforms, for the translocation of metabolites between cytosol and matrix, involved in oxidative phosphorilation, reduction equivalents transfer, metabolic pathways such as Krebs cycle, urea cycle, fatty acids oxidation, gluconeogenesis, amino acids degradation, proteins and nucleic acids synthesis (Table 1). Apart from these primary functions some carriers play an important role in the regulation and maintenance of cytosol/matrix equilibrium, for example in the levels of phosphorilation and redox potential. Mitochondrial carriers are able to transport a wide variety of substrates, from H^+ to bigger molecules like ATP. Most substrates are anions, even if there are few cations and zwitterions (Palmieri, 2004). The functional role of these proteins is strictly related to their distribution and tissue localization. Some of them are ubiquitous but others are highly specific, indicating a direct involvement in the tissue functions (Palmieri, 2004). ADP/ATP and phosphate transporters, involved in the oxidative phosphorilation process, play a crucial role for the life of a cell, and for this reason they can be found in all the cells that carry mitochondria, like those involved in the transport of reduction equivalents or substrates for the oxidative metabolism in the mitochondrial membrane. PYC transport piruvate from the cytosol to the matrix during the gluconeogenesis process, while CIC exchange citrate with tricarboxilic acids. Oxoglutarate transporter OGC, together with the aspartate/glutamate carrier, is fundamental in transferring reducing equivalents through the mitochondrial membrane.
Other transporters, like the carnitine/acilcarnitine carrier CAC abundant in scheletric and cardiac muscle where it promotes the ingress of fatty acids in mitochondria for their oxidative metabolism, are tissue specific. Dicarboxilic acids carrier DIC is involved in the gluconeogenesis process and for this reason is above all found in the liver, as ORC, which exchange citrulline and ornitine (Palmieri, 2004) and GC that take part at the urea cycle. UCP1 has been found in the brown adipose tissue (Aquila et al., 1985), where it transport protons wasting the proton gradient produced by the oxidative phosphorilation to produce heat. Its isoform UCP2 is however ubiquitous while UCP3, and UCP4 and 5 are specific for muscle, and brain tissues respectively.

Table 1
Main carriers, substrates and metabolic role

Transporter	***Main substrates***	***Metabolic process***
AAC	ADP, ATP	Oxidative phosphorilation
PiC	Phosphate	Oxidative phosphorilation
CIC	Citrate, isocitrate, cis-aconitate, malate	Fatty acids and steroids biosynthesis
OGC	Oxoglutarate, malate	malate/aspartate carrier, gluconeogenesis, metabolism
DIC	Phosphate, malate, succinate, solfate, tiosolfate	Krebs cycle, gluconeogenesis
PYC	Piruvate, chetonic bodies	Krebs cycle, gluconeogenesis
AGC	Aspartate, glutamate, cysteine-sulfinate	malate/aspartate carrier, gluconeogenesis, urea synthesis, cysteine catabolism
GC	Glutamate	Urea synthesis
ORC	Ornitine, citrulline, lysine, histidine, arginine	Urea cycle
UCP	H^+	Termogenesis
FAD carrier	FAD	Reduction equivalents for the tran sport of electrons
ATP-Mg/phospha te carrier	ATP-Mg, phosphate	Content of matricial adenin
CAC	Carnitine, acilcarnitine	Fatty acids oxidation

2.1.2 Transport mechanisms

Mitochondrial carriers act through two kind of transport mechanism: simport and antiport. In the first case the transport of a metabolite is coupled to that of a proton in the same direction, while in the second case the carrier exchanges two metabolites. Most carriers transport negatively charged substrates, even if a little part of them is able to translocate positively charged molecules also. The exchange of metabolites with a different net charge can cause an electrostatic gradient between the two sides of the membrane. For this behaviour carriers have been classified as electrogenic or electroneutral

(Palmieri, 2004). The ADP/ATP carrier, the UCP protein and the aspartate/glutamate carrier belong to the electrogenic transporters family.
In the electroneutral carriers, the balancing of the charge during the transport can be achieved either via simport of H^+ or antiport of OH^- (as in the case of phosphate, glutamate and Piruvate carriers), or through anions/cations exchange (as in the case of glutamate/malate or ornitine/lysine carriers). Mitochondrial carriers use for the transport the driving force derived from the concentration gradient of solutes and from the electrochemical potential generated from the respiratory chain.

2.1.3 Primary structure

In 1982 the sequence of a mitochondrial carrier, the ADP/ATP mitochondrial carrier (AAC), has been solved for the first time (Aquila et al., 1982). Further, other transporters have been sequenced, among which those of the UCP protein (Aquila et al., 1985), PYC (Runswick et al., 1987), OGC (Runswick et al., 1990) CIC (Kaplan et al., 1993) e CAC (Indiveri et al., 1997). The analysis of the 300 amino acids sequence of the AAC protein has shown a trifold structure formed by three homologous domains each one 100 amino acids long. Since all the sequences have shown the same structural scheme, it has been proposed that these proteins belong to the same family, the mitochondrial carrier family (MCF). The typical trifold structure probably derive from a double tandem duplication of a common ancestral gene.
Another common feature for all the mitochondrial carriers is the presence of a conserved sequence (MCS), P-h-(D/E)-x-h-(K/R) (P=Proline, D=Aspartic acid, E=Glutamic acid, K=Lysine, R=Arginine, h=Hydrophobic, x=Any amino acid), found at the C-terminal portion of the odd-numbered helices. Mitochondrial carriers belonging to the MCF are all 300 amino acids long and have an basic isoelectric point (Kaplan, 2001).

2.1.4 Topology

Based on the trifold structure and on the hydropathy analysis, it has been proposed that each 100 amino acids domain is formed by two transmembrane segments (TMS) folded in α-helix, giving rise to a monomer of six helices (Saraste and Walker, 1982; Aquila et al., 1985; Ranswick et al., 1987) (Fig. 2). The helices belonging to each domain are connected by an hydrophilic loop,

longer than those loops connecting the different domains (Fig. 2). In the whole protein indeed three long intra-domain loops and two short inter-domain loops are present (Fig. 2). All the mitochondrial carriers follow this topological scheme (see also Fig. 4).
The relevance of this model has been object of intensive studies. Some aspects have been experimentally proved by incubating the carrier with reagents impermeable to the membrane, such as specific antibodies, proteases and chemical labelling agents. Immunological studies of the CIC protein showed that both the N-terminal and C-terminal regions are exposed on the cytoplasmic side of the inner mitochondrial membrane, and also indicating an even number of helices as predicted by the model (Capobianco et al., 1995).

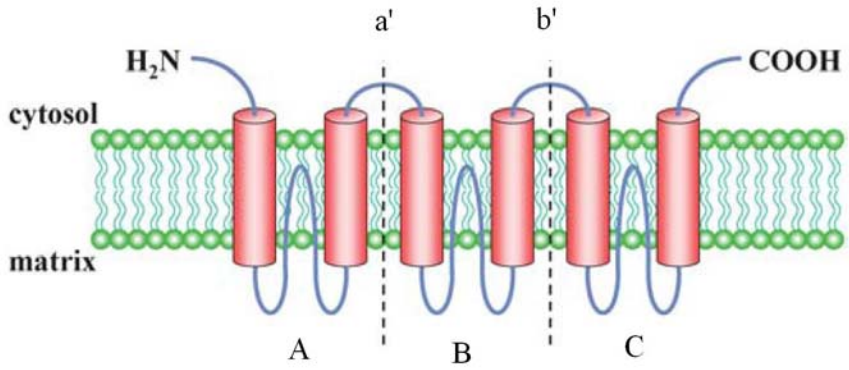


Figure 2. Topological model of mitochondrial carrier (Palmieri, 2004).

In order to study the UCP protein, fusion proteins to select specific antibodies for some fragments of this protein have been used. These studies have led to the determination of the orientation of the N-terminal portion of the transmembrane segments I, II, III and IV, and the orientation of both the N- and C-terminal regions of segment VI (Miroux et al., 1992; 1993). For the PYC transporter it has been proved the cytosolic exposure of N- and C-terminal regions and of the a' and b' loops. Further, it has been verified that loop B is exposed toward the mitochondrial matrix (Ferriera et al., 1990; Capobianco et al., 1991, Palmieri et al., 1993).
The largest amount of information derived from the resolution of the 3D structure of the ADP/ATP carrier through cryo-electron microscopy and X-ray diffraction (Kunji and Hardy 2003; Pebay-Peyorula et al., 2003). These two groups have chosen this carriers among all the others not only for its

metabolic role, but also for its abundance in mitochondria, promoting its isolation and purification.

2.2 The ADP/ATP mitochondrial carrier

ADP/ATP mitochondrial carrier (AAC) permits the electrogenic exchange of a cytosolic ADP^{3-} with an ATP^{4-} synthesized in the mitochondrial matrix by the ATP-syntase. The expression of the three isoforms of the carrier is regulated based on the energetic needs of the cell and on its degree of differentiation. AAC has been the first transporter belonging to the MCF for which the primary sequence has been obtained (Aquila et al., 1982).
It is known that this protein can bind two inhibitors carboxyatractyloside (CATR) and bongkrekic acid (BA), both acting as strong poisons by blocking nucleotide translocation (Vignais et al., 1962; Henderson and Lardy, 1970). These two compounds have been extensively used to characterize different conformational states of the carrier. The CATR stabilizes the c-state, in which the protein is opened towards the inter-membrane space and closed to the matrix side, while the BA stabilizes the opposite conformation, called m-state, that is closed towards the inter-membrane space and opened to the matrix side (Klingenberg, 2008). In absence of the two inhibitors the carrier undergoes a rapid transition between the cytosolic and matricial state, which presumably reflects the same rapid behavior in vitro.
Cryo-electronic microscopy of bi-dimensional crystals has proved the validity of the modeled proposed by Runswick et al (1990). The protein shows a ternary pseudo-symmetry, that reflects the trifold structure evidenced by the primary sequence alignment. The electronic density map is coherent with the hypothesis of a transmembrane domain formed by a six α-helices bent bundle with respect to the membrane plane (Kunji et al., 2003).
Finally, the AAC has been crystallized in the presence of the inhibitor CATR (Pebay-Peyroula et al., 2003), which is known to bind to the carrier from the inter-membrane space (IMS), hampering ADP^{3-} import. The structure was solved to 2.2 Å resolution (Pebay-Peyroula et al., 2003), revealing for the first time the overall fold of MCF carriers and highlighting the role played by the MCF motif in the folding.

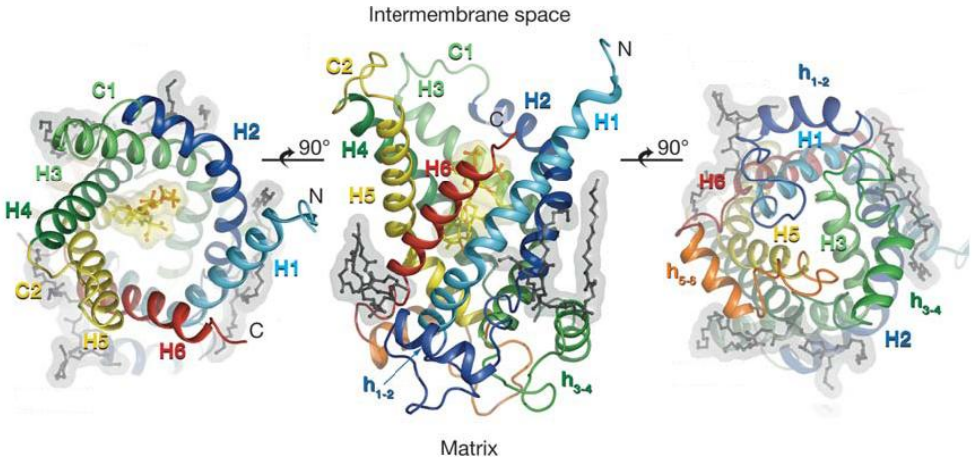


Figure 3. Three-dimensional structure of the AAC protein obtained through X-ray crystallography. The protein has been solved with its natural inhibitor CATR. Few molecules of cardiolipines are also present bound to the protein. The number of the helices and of the segments connecting them are defined.

The crystallography has evidenced that the six helices form a thick transmembrane domain with a cavity opened toward the inter-membrane space, where ADP^{3-} should be bound. The proline residue of the MCS motif present on the three odd-numbered helices causes a kink on these helices on the matricial side. The six helices create a bundle, interacting each other via hydrogen bonds and salt bridges that happen also between the residues of the MCS. At the centre of the structure from the cytosolic side there is a solvent accessible cavity with a V-shape. The long intra-domain loops on the matrix side are partially folded in α-helix parallel to the membrane space, that close the accessibility of the carrier from this side. These helices are called h_{1-2}, h_{3-4} and h_{5-6}, based on the transmembrane helices that they connect (i.e. helices 1 and 2) (Fig. 3).

In the X-ray structure CATR is located near residues belonging to helices H2, H4 and H6 at the bottom of the cytoplasmic side cavity, over the salt bridges network that keep the matrix side close. In this structure we can also find three molecules of cardiolipine, located in the groove formed by the matrix loops. The strong interaction between these molecules and the carrier could give us some clue about their role in stabilizing the protein into the membrane (Nury et al., 2006).

2.2.1 Sequence analysis and substrate specificity

The presence of a 3D structure of a carrier belonging to the MCF led us to apply computational techniques to understand the role of this protein and also the relationships between all the members of this family. Sequence alignment of sequences coming from several different species enabled us to evidence important conserved residues, mainly found in the transmembrane segments. Residues facing the cavity are mainly polar or charged, the latter ones being maybe involved in the substrate binding, in the translocation mechanism and in the maintenance of the structure of the carrier.
All the carriers transport metabolites structurally related, suggesting the existence of a common binding site and a similar transport mechanism. Recently, computational techniques and experimental evidences have been exploited to deduce the localization of a hypothetical binding site in the c-state of mitochondrial carriers (Robinson and Kunji, 2006). Carriers have been classified in three main subfamilies on the basis of the similarity of the functional groups of the substrate they bind: keto acids, amino acids and adenine containing substrates transporters. Homology modelling technique has been used to obtain the 3D structure of 19 carriers, starting from the crystal structure of the bovine ADP/ATP carrier, referred to as btAAc1, as a template (Robinson and Kunji, 2006). In order to rend this study easier all the sequences have been numbered based on the btAAC1 numeration. From the analysis it has been shown that all the sequences share a conserved region at the level of the so called contact point II (residue 182 for btAAC1), where the keto acids, amino acids and adenine transporters carry the motif R-[DE], R-[GHNT] and G-[IVLM] respectively. Figure 4 shows some carriers for all the three families, and for each one the residues of contact point II are depicted. It has been postulated that the residues of contact point II discern between the three classes of substrates, while contact point I (residue 79 for btAAC1) could interact with the variable group of the substrate rending to the binding highly specific. Mitochondrial carriers can transport a series of structurally related substrates since contact point II can accept groups with similar chemical characteristics, but the variability of the lateral group is tolerated based on residues of contact point I. Residue 279 of btAAC1, belonging to contact point III, is a well conserved arginine among all the family. Even if this residue doesn't confer specificity for the substrate binding it could act as a third important contact point in the binding site. Indeed, residues 79, 182 and 279 of btAAC1 are in equivalent structural position, due to the three-fold symmetry of the carrier. Residues forming the common binding site in these transporters

are conserved in the human orthologous ones (Robinson and Kunji, 2006). All the human carriers probably act with a similar mechanism of action.

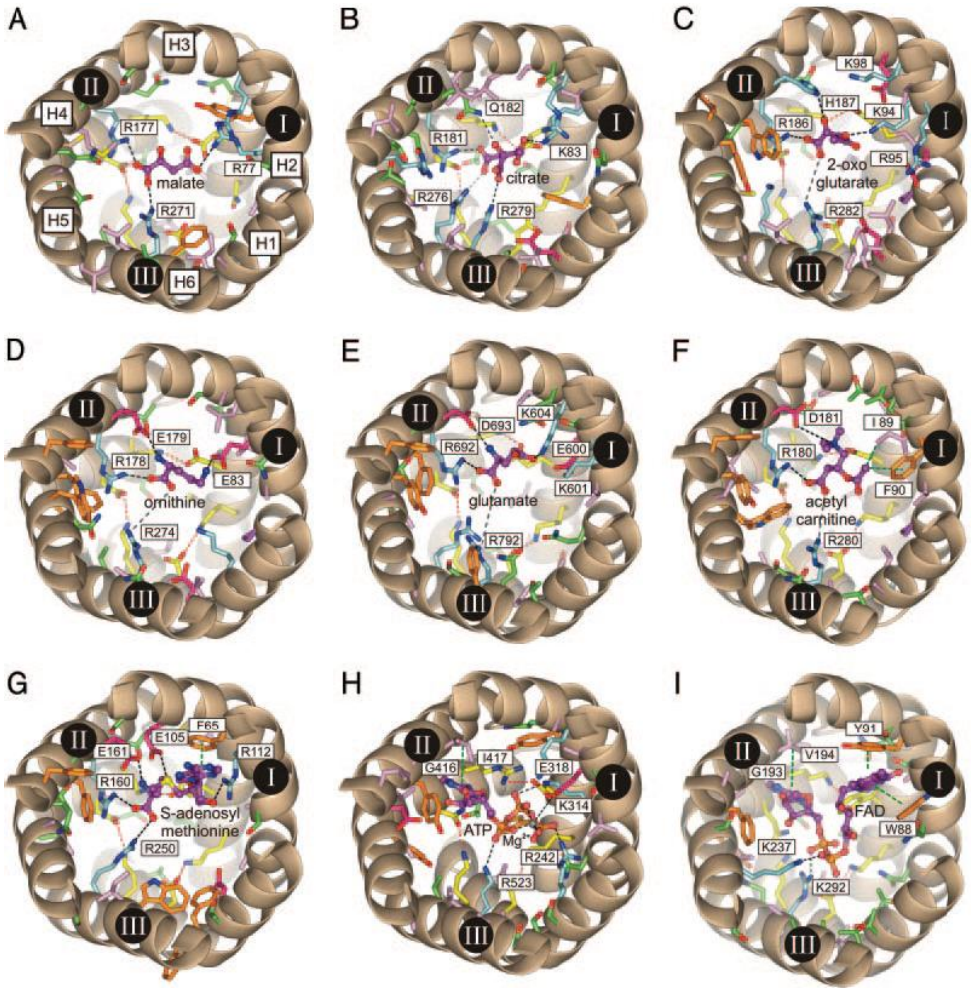


Figure 4. View from the cytosolic side of the transmembrane region of some carriers for which the 3D structure has been obtained with Homology modelling. (A) DIC (B) CIC (C) OGC (D) ORC (E) GC (F) CAC (G) Adenosil-methionine carrier (H) ATP-Mg/Phosphate carrier (I) FAD carrier. (Robinson & Kunji, 2006).

2.2.2 Transport mechanism

Kinetics studies have demonstrated that, with the only exception of the carnithine transporter, all the carriers act with a sequential simultaneous mechanism, that implies the formation of a ternary complex where the carrier is bound to two molecules of substrate (Palmieri, 2004). The binding of the substrate happens with a rapid and random equilibrium, in which it is not expected that the substrates binds in a defined order the external or the internal side (Kaplan, 2001). The carnithine carrier instead follows a ping-pong mechanism.

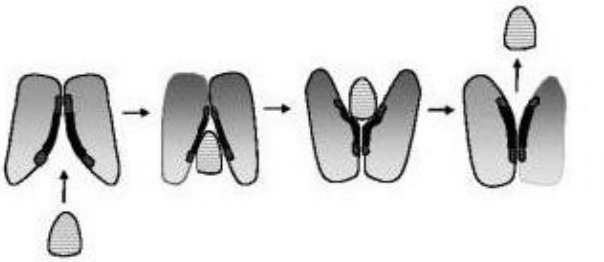


Figure 5. Transport mechanism following the SBCGP model. (Klingenberg, 2005).

The solutes transport across the biological membranes is a catalytic process, comparable to an enzymatic reaction (Klingenberg, 2005). Both processes are facilitated by proteins that guarantee selectivity and catalytic acceleration. The main difference between these two processes is that for the transport reaction the protein, instead of the substrate during the enzymatic reaction, undergoes a destabilization. Because of the absence of a chemical reaction, the existence of a transition state, similar to that found in the enzymatic reaction, during the transport has been ignored for a long time. At the beginning a mechanism in which the exchange between the cytoplasmic state (c-state) and the matricial state (m-state) followed the substrate translocation was proposed (Fig. 5). The "Single Binding Centre Gated Pore" (SBCGP) model was

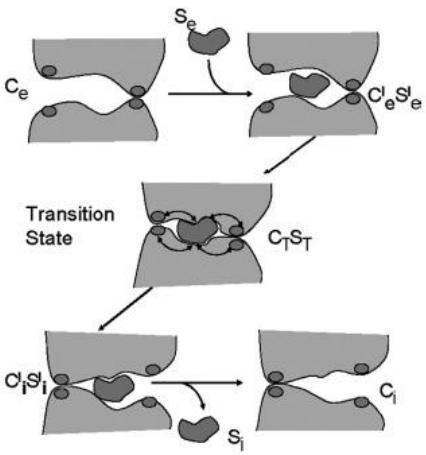


Figure 6. Transport mechanism model that implies a transition state. (Klingenberg, 2005).

proposed for the first time for the ADP/ATP carrier (Klingenberg, 1989). The binding of the substrate in the specific site, located at the centre of the protein, implies a conformational change of the protein that allows the closure of the entry site and the opening of the exit. In the SBCGP model, however, the catalytic aspects and the energetic balance of the carrier-substrate interaction and of the transporter changes are undervalued.

A more realistic analysis of the catalysis can be obtained by introducing a transition state and an unstable protein-substrate interaction (Klingenber, 2005; Fig. 6). At the beginning, the binding site could be only partially able to accommodate the substrate. Differently for what happens with the enzymes, in this case the substrate in the transition state remains unchanged while the conformation of the protein binding site is transformed to be adapted at the substrate, with a simultaneous closure of both sides of the carrier. The binding energy released in the transition state is funnelled in the structure to induce a conformational change that transform the binding site, opening the exit site.

It has been proposed that the binding of the substrate at helices H2, H4 and H6 allows helices H1, H3 and H5 to rotate and move, and so that the binding of the substrate act as a hinge around which the helices rotate in order to permit the opening of the carrier in the matrix side (Robinson and Kunji, 2006). The localization of the substrate into the carriers suggests that the protein-substrate interaction and the carrier opening are coupled thanks to the existence of charged groups that perturb the salt bridges causing their rearrangement that lead to a new conformation of the carrier that permits the translocation of the substrate.

The crystallographic structure of the AAC further suggests that the modifications in the bent angle of odd-numbered helices could induce a big conformational change, causing a reorientation of the anphipathic loops (Nury et al., 2006). Cross-linking and labelling experiments conducted on AAC have permitted to hypothesize the involvement of the loops exposed toward the matrix side in the transport process. In particular, it has been proposed that the M1 loop plays a crucial role thanks to its positive charge and the possibility to attract the negative charged moiety of the substrate (Hashimoto et al., 1999). Recently, it has been also shown that the M2 loop undergoes conformational rotations, probably due to the transport process (Dahout-Gonzalez et al., 2005). Further, recent studies suggest that the carriers are active in a dimeric form. If this is true, the transport mechanism should involve two translocations routes, each one in each monomer (Nury et al., 2006).

2.2.3 Quaternary structure

The transport mechanism can be accomplished by a dimeric structure in which each monomer carries a binding site and is able to translocate the substrate. Indeed, for many carriers experimental evidences exist in favour of the hypothesis that the MC family carriers act as dimeric units. Cross-linking studies conducted on UCP (Klingenberg et al., 1989) and OGC (Bisaccia et al., 1996a; 1996b) indicate the formation of a dimeric structure, due to the presence of disulphuric bonds. Studying CIC the dimeric shape has been confirmed through exclusion chromatography and native electrophoresis with charge periodicity (Kotaria et al., 1999). Interestingly the Cys-less mutant of this transporter is in a dimeric form, indicating that the disulphuric bonds are not necessary for the formation and the stabilization of the dimeric state. Further, covalent dimmers of the yeast AAC, in which the C-terminal of one monomer was fused with the N-terminal of the other, resulted to be functional (Hatanaka et al., 1999; Trézéguet et al., 2000). From these experiments the authors suggest that helix H1 of one monomer could interact with helix H6 of the other one.
Recently, the bi-dimensional characterization of the ADP/ATP carrier using cryo-electron microscopy has evidenced that the molecules in the crystals are organized as symmetric omodimers (Kunji et al., 2003). However, the authors were not able to demonstrate whether this conformation was induced by the crystallographic conditions or it reflects the multimeric nature of the carrier.

Chapter 3
– Aim of the Thesis

3.1 Introduction

In this thesis two different aspects of the mitochondrial ADP/ATP carrier have been investigated. In the first case, the different structural/dynamic characteristics of two mutants have been analyzed in relation with wild type using classical Molecular Dynamics simulation as a tool. In detail, the wild type protein, the single A114P pathological mutant and the A114P/V181M double mutant that is able to restore the normal transport activity of the protein, have been simulated for 20 ns each (Fig. 7).

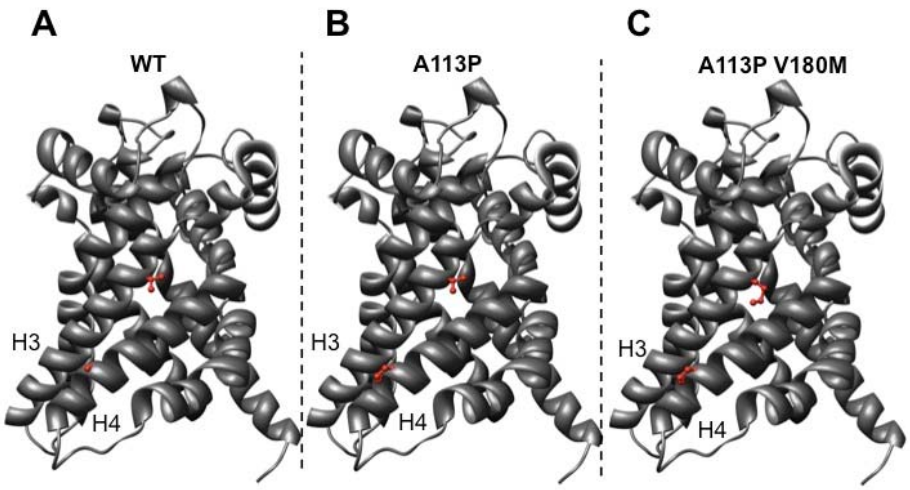


Figure 7. Ribbon representation of the wild type protein **(A)**, of the A113P single mutant **(B)** and A113P/V180M double mutant **(C)**. The residues involved in the mutations are represented in red ball and stick

In the second case, through the application of molecular dynamics simulation and molecular docking, possible interaction sites of the ATP^{4-} substrate on the matrix region of the carrier have been identified. In particular, both the X-Ray structure and the structures extracted from molecular dynamics simulation through cluster analysis have been used to perform the docking experiment (Fig. 8).

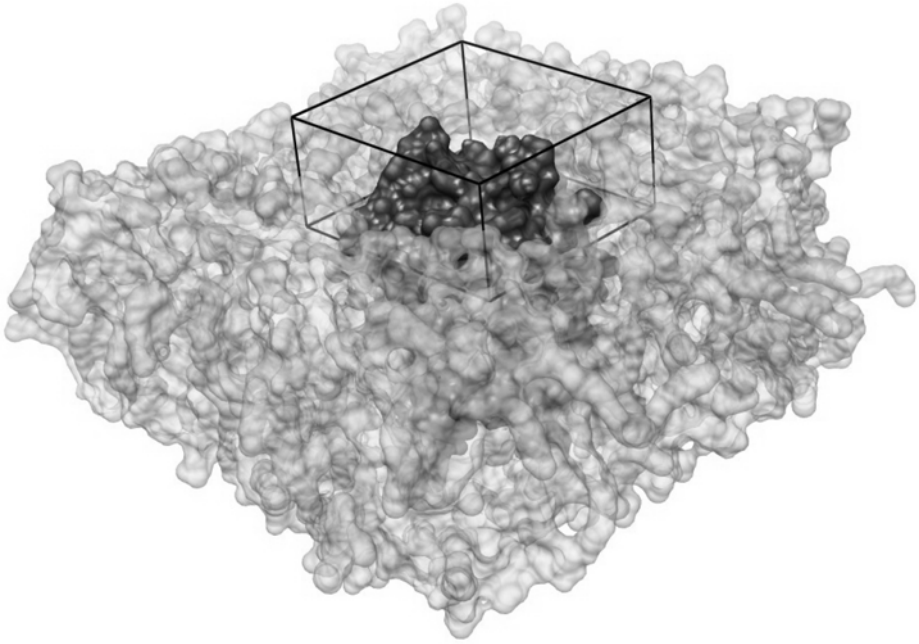


Figure 8. Snapshot of molecular dynamics simulation representing as a surface the protein in black colour, immersed in the lipid bilayer in gray colour. The black box defines the region used to carry out the molecular docking on the matrix side.

3.2 Simulations of the mutants ADP/ATP mitochondrial carrier

Several human single point ADP/ATP carrier mutations, such as Val289Met, Leu98Pro, Asp104Gly, and Ala114Pro, related to specific pathologies such as the external ophtalmoplegia (adPEO), have been functionally characterized (Kaukone et al., 2000, De Marcos et al., 2002, De Marcos et al., 2005, Palmieri et al., 2008).

The single mutation Ala114Pro, identified in patients affected by the Ad-Peo disease, has been found to affect the respiratory chain producing rearrangements on the mitochondrial DNA, suggesting to cause an impairment in the substrate translocation (Kaukonen et al. 2000).
The structural/functional effect of this mutation on the mitochondrial carrier has not been clearly established, and different hypotheses were reported in literature since this mutation was discovered (Kaukone et al., 2000, De Marcos et al., 2002, De Marcos et al., 2005, Fontanesi et al., 2004, Palmieri et al., 2008).
Working with the human gene expressed in yeast De Marcos et al. (De Marcos et al., 2002) infer that this mutation plays an important role during the post-transcriptional modification of the protein, justifying the impaired ADP/ATP translocation with the absence of the carrier in the inner mitochondrial membrane. On the other hand, other experiments carried out by Fontanesi et al. on the yeast humanized gene (Fontanesi et al., 2004) do not confirm this hypothesis demonstrating that the mutated protein is located in the inner mitochondrial membrane, although with a minor concentration if compared to the wild type carrier. Moreover, through specific transport assays, they indicate that this mutant shows a lower affinity for ADP with respect to ATP, if compared to the wild type protein, and suggest that the Ala114Pro mutation is involved in the impairment of the transport by changing the affinities for the substrates (Fontanesi et al., 2004). In a further work De Marcos et al. (De Marcos et al., 2005) have analyzed the effect of the additional mutation Val181Met. The double mutant Ala114Pro/Val181Met shows substrates exchange rate comparable with that of the wild type, suggesting that the second mutation is able to suppress the effect of the first, restoring the activity of the carrier (De Marcos et al., 2005). These authors, explaining the structural effect of the two mutations, propose that Ala114 may be involved in tight interactions with lipids that, lost due to the mutation in proline, are partially restored by the second mutation (De Marcos et al., 2005). Recently Wang et al. (Wang et al., 2008) point out that the Ala128Pro mutation in yeast, corresponding to the Ala114Pro mutation in the human protein, occurs modifying the activity of a crucial structural-dynamical gating region of the carrier. These authors provide direct evidence that this mutant is incorporated in the inner mitochondrial membrane and uncouples the respiration (Wang et al., 2008).
In this ambiguous scenario I tried to rationalize the structural effects of these two mutations using the molecular dynamics (MD) simulation. Recently we have applied such high resolution approach to investigate the ADP/ATP

carrier in presence and absence of CATR (Falconi et al., 2006). In this paper has been shown the presence of two different conformations stabilized by specific networks of intra- and inter-repeats salt bridges (Falconi et al., 2006).
In this thesis I present three 20 ns classical MD simulations of the bovine wild type ADP/ATP transporter, the sole resolved mitochondrial carrier structure having a high homology with the human enzyme, and of the two mutants. Using the MD atomistic approach, we hypothesize how the Ala114Pro mutation could affect the structure-function relationship of the carrier and how the Val181Met mutation could restore the impaired transport. The conformational mechanism that occur has been analyzed and the results provide a reasonable cause and effect hypothesis that can be useful in the elucidation of a confusing experimental background.

3.3 Searching of the ATP interaction sites on the ATP/ADP mitochondrial carrier matrix region

Identification of the binding sites between the ATP^{4-} molecule and the carrier matrix side may add a further step in the description of the ADP/ATP transport.
At present the only carrier 3D structure, solved at high resolution, is in complex with the CATR inhibitor, and thus not prone to transport ligands from the matrix lumen. However, Duyckaerts et al. showed via kinetic experiments that ATP^{4-} can bind the binary complex ADP^{3-}-carrier from the matrix side of the protein, forming a ADP^{3-}-carrier- ATP^{4-} ternary complex (Duyckaerts et al., 1980). Furthermore recent studies have shown that these carriers can also work as monomers (Bamber et al., 2007; Nury et al., 2008), findings that prompt for new models for the transport process.
Thus, in the case of a monomeric carrier loaded externally with ADP^{3-}, an internal binding site for ATP^{4-} has to be contemplated. So it can be supposed that even if the carrier cannot transport ATP^{4-} while bound to CATR and fixed in an abortive c'-state, it could still be able to recognize this substrate. Moreover, it has been shown that fluorescein derivatives as eosin Y, that are structurally related to adenine nucleotides, can recognize the ADP/ATP carrier form the matrix side, although they cannot be translocated into the cavity (Majima et al., 1998). Binding of eosin Y is displaced by ADP^{3-}or ATP^{4-} that prevents the thiol-reactive eosin-5-maleimide molecule to bind on Cys159, suggesting that ATP^{4-} binds in proximity of this residue, through a mainly hydrophobic interaction (Majima et al., 1994; 1998), suggesting that this

region can be fundamental for the binding on the matrix side as also proposed by our present docking results.
The 3D structure of the carrier (Pebay-Peyroula et al., 2003) can then be taken as an initial model to verify its ability to bind the ATP^{4-} molecule from the matrix lumen. Moreover it has been shown by a 10 ns MD simulation carried out in absence of CATR, that the matrix region of the protein undergoes significant changes, in particular in the salt bridge networks and in the appearance of matrix cavities not found in the crystal and so likely sampling conformations more prone to provide ATP^{4-} binding sites (Falconi et al., 2006). In this work, to extend the available conformations of the protein, a 20 ns MD simulation has been carried out starting from the X-ray structure of the bovine mitochondrial ADP/ATP carrier (Pebay-Peyroula et al., 2003), embedded in a double lipid bilayer, depleted of the CATR inhibitor and solvated in water. The trajectory has been clustered in families and each representative structure of these families has been used to carry out several docking runs with the ATP^{4-} molecule. Docking results have also been clustered and possible binding sites have been recognized and mapped.
The interacting residues have been identified, along with the specific interaction type, and have been compared with those identified in the crystal structure suggesting implications for multiple ATP^{4-} binding sites related to the AAC three-partite structure.

Chapter 4
– Methods

4.1 Introduction to the molecular dynamics simulation

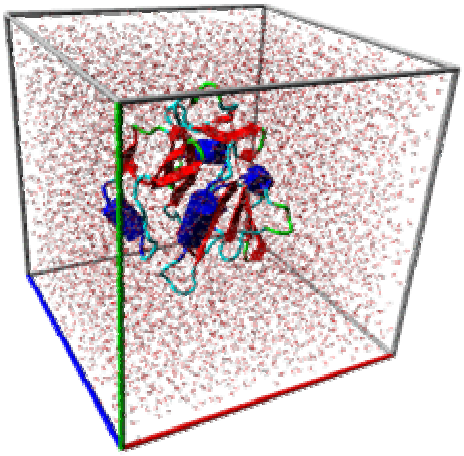


Figure 9. Representation of a globular protein into the solvated simulation box.

Molecular dynamics (MD) has provided many insights concerning the internal motions of macromolecules since the first protein was studied over 25 years ago. In general, the three-dimensional structure of a macro molecule, obtained through X-Ray or NMR methods, is inserted into a simulation box appropriately solvated (Fig. 9).

The power of MD lies in its capability to investigate biological motions that are often not accessible to experiments. Nowadays, the synergy between theory and experiments is necessary to progress in the understanding of different biological aspects. With the continuing advances in the methodology and the speed of computers, MD studies are progressively pushed to larger systems, greater conformational changes and longer scale times. The current situation points towards a promising role of MD in the years to come for the study of biochemical problems. There are many MD simulation packages available for the study of biomolecules. All the simulations presented here have been performed using the suite of program GROMACS (Berendsen et al., 1995; Lindahl et al., 2001).

4.2 MD principles

Using MD it is possible to simulate the time-evolution of a molecular system, which is represented classically considering a set of particles defined by their positions and momenta.
Starting from a coordinate set, taken for example from a crystal structure, and assigning velocities to each atom, typically from a Boltzmann distribution at a given temperature, successive coordinates and velocities are obtained by integrating the Newton's equation for the motion in each coordinate direction. In one dimension, the equation can be written as:

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i}$$

where mi and x_i are the mass and position of each atom, respectively and F_{xi} is the derivative of the potential according to a force field equation that is described in the next section. The result is a trajectory that shows how atomic positions and velocities evolve with time according to the influence of the remaining atoms in the system. Due to the large number of particles interacting with each other, the integration is performed numerically most commonly using the leap-frog algorithm (Van Gunsteren and Berendsen, 1990). The integration step is limited to the fastest motion in the system. Therefore, for atomistic simulations the step size is usually 1 fs, or 2 fs if restraining bond lengths, generally done using SHAKE (Van Gunsteren and Berendsen, 1977) and LINCS (Hess et al., 1997) algorithms.

4.2.1 Force fields

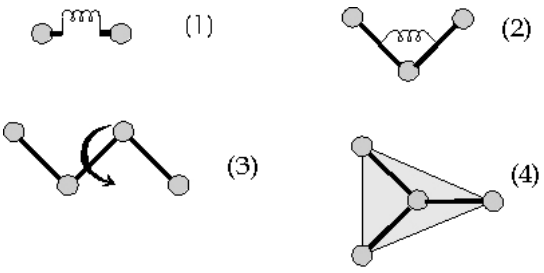


Figure 10. Schematic representation of the harmonic oscillators for the covalent bond. Bond stretching (1), angle bending (2), proper dihedral angle (3) and improper dihedral angle.

The calculation of the interaction energy within a classical description of a molecular system requires a force field. A force field is built up from two distinct components, a set of equations used to generate the potential energies (and their derivatives, the forces) and the parameters used in

these equations. Nowadays, there are four main force fields in common use for simulating biological macromolecules: AMBER (Pearlman et al., 1995), CHARMM (Brooks et al., 1983; MacKerell et al., 1998; Mackerell et al., 2004), GROMOS (Van Gunsteren and Berendsen, 1987) and OPLS (Jorgensen et al., 1996).

The intramolecular potential energy is typically represented by harmonic oscillators for bond stretching and angle bending, a Fourier series for each torsional angle, and Coulomb and Lennard–Jones (LJ) accounting for the interactions between atoms separated by three or more bonds (Fig. 10). The latter two terms, referred together as non-bonded terms, are evaluated between all atom pairs in the system to yield the intermolecular energy. Such force fields compute the energy as a sum of terms representing bond elongation, angle and dihedral deformation, and non-bonded interactions with the following general form:

$$E = \sum_{bonds} V^{str} + \sum_{angles} V^{bend} + \sum_{torsions} V^{tors} + \sum_{LJ} V^{LJ} + \sum_{Coulomb} V^{Coul}$$

where the three first summations correspond the bonded terms (that include atoms connected up to three consecutive bonds) and the last two refer to the non-bonded ones. All summations can be easily calculated from the coordinates of the system at a given time. For each pair of bonded atoms (i and j), the stretching term is computed as:

$$V_{ij}^{str}(r_{ij}) = k_{ij}^{r}(r_{ij} - r_{ij}^{o})^2$$

where k^{r}_{ij} is the stretching force constant and r_{ij} and r^{o}_{ij} are the distance between atoms and its equilibrium bond length, respectively. For every group of three bonded atoms (i, j and k), the angle term is described as:

$$V_{ijk}^{bend}(\theta_{ijk}) = k_{ijk}^{\theta}(\theta_{ijk} - \theta_{ijk}^{o})^2$$

where k^{θ}_{ijk} is the bending force constant, θ_{ijk} and θ^{o}_{ijk} are the angle between atoms and its equilibrium value, respectively. For every group of four bonded atoms (i, j, j, l), the dihedral term is often represented as a cosine expansion:

$$V_{ijkl}^{tors}(\phi_{ijkl}) = k_{ijkl}^{\phi}[1+\cos(n\phi_{ijkl} - \phi_{ijkl}^{o}))$$

where k^{φ}_{ijkl} is a dihedral constant affecting the barrier height, n is number of minima in a 360º rotation, and φ_{ijkl} and φ^{o}_{ijkl} are the dihedral angle and the equilibrium value according the biochemical convention (*trans* φ=180º, *cis* φ=0º and *gauche* φ=60º/300º), respectively. The first non-bonded term is often represented by a 6-12 LJ potential, a simple mathematical model (Lennard-Jones, 1931) that accounts for two distinct forces (an attractive and a repulsive) that neutral atoms and molecules are subject to:

$$V_{ij}^{LJ}(r_{ij}) = \left[\left(\frac{A_{ij}}{r_{ij}}\right)^{12} - \left(\frac{B_{ij}}{r_{ij}}\right)^{6}\right]$$

where A_{ij} and B_{ij} are parameters that depend from on each pair of atoms. The first term accounts for the attractive forces at long range (van der Waals or dispersion) and the other for the repulsive forces at short range, resulting from the overlap between electron orbitals. Finally, the last term in the force field equation is a Coulombic potential describing the electrostatic interactions:

$$V_{ij}^{Coul}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_o r_{ij}}$$

where q_i and q_j are the charges of atoms i and j, r_{ij} the relative distance between them and ε the vacuum permittivity.

The constants and parameters of the preceding equations need to be fed from biophysical experiments and/or quantum mechanics calculations which differ from one force field to another. The specific force fields parameters of the different components present in the systems studied –proteins, lipids, water molecules and ions are discussed in the following sections.

4.2.2 Proteins in MD

The parameters for the amino acids are well adjusted in the majority of force fields designed to compute biological macro-molecules. Therefore, once one has selected a force field (AMBER, CHARMM, GROMOS, OPL) to perform the simulation there are not many critical choices to be done except for non-standard residues.

4.2.3 Lipids in MD

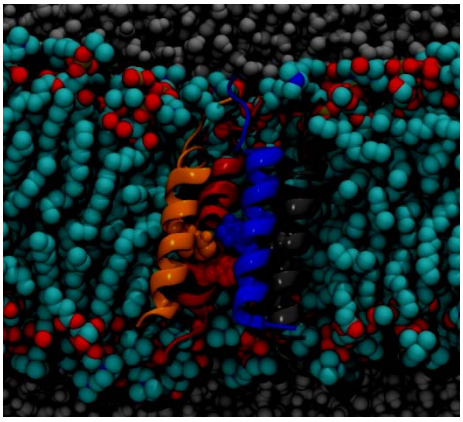


Figure 11. Representation of four α-helices immersed in a double lipid bilayer. The α-helices are represented in orange, red, blue and grey ribbons, while the lipid are represented in cyan and red spheres. The water molecules, that solvated the system, are represented in grey spheres.

For the simulation of lipid bilayer (Fig. 11) there are two largely used force fields: one that is part of the official CHARMM distribution (MacKerell et al., 1998) and the other introduced by (Berger et al., 1997), developed with parameters taken from OPLS and GROMOS for its use in the GROMACS package (Allen and tildesley, 1987; Berendsen et al., 1995; Lindahl et al., 2001). There is no experimental information indicating that one force field is substantially better than the other. Although both force fields are microscopic, CHARMM includes explicitly all lipid atoms in the system, whereas the one developed by Berger, does not account explicitly for the non-polar hydrogen atoms and CH2 and CH3 groups are treated as a single particle. This approximation is often referred as a *united-atom* force field.

Another approach (not explored in this thesis) is the use of the so-called coarse-grained models to perform mesoscopic simulations by grouping a bunch of atoms together (Venturoli et al., 2005; Bond and Sansom, 2006; Sperotto et al., 2006; Shih et al., 2006; Muller et al., 2006; Marrink et al., 2007). This allows to use a longer integration step, up to 40 fs, (Periole et al., 2007). Although with a lesser detailed description, they allow longer

simulation time scales as well as the possibility to deal with larger systems than with the atomistic approximation at a much less computational effort. In this regard, the recently developed MARTINI force field for mesoscopic simulations is very promising (Marrink et al., 2005; Periole et al., 2007). The lipid force field developed by Berger can be considered as a small coarse-grain model and accordingly, it permits using a time-step of 4 fs.

4.2.4 Water in MD

Many water models (Fig. 12) are available in the literature for an accurate representation of the liquid water (Wallqvist and Mountain, 1999; Guillot, 2002; Jorgensen and Tirado-rives, 2005). The most common ones are often distributed together with the force fields described above. These models have been parametrized to reproduce physical and thermodynamical properties such as the density, enthalpy of vaporization, radial distribution functions, energies of hydration or dipole moment. They can be classified by the number of points used to define the model (atoms plus dummy sites), whether the structure is rigid or flexible, and whether the model includes polarization effects or not. The simplest, and most popular, models for MD are TIP3P (Jorgensen et al., 1983) and SPC (Fig. 10) (Berendsen et al., 1987) series, which have three interaction sites, corresponding to the three atoms of the water molecule with rigid geometry. Each atom has a point charge assigned and the oxygen atom also gets Lennard-Jones parameters. The more complex 4-site or 5-site models such as TIP4P and TIP5P, respectively, place the negative charge on either a dummy atom placed near the oxygen along the bisector of the HOH angle or on two dummy atoms representing the lone pairs of the oxygen atom. These models improve the electrostatic distribution around the water molecule, though to a larger computational cost because of the larger number of electrostatic interactions to compute.

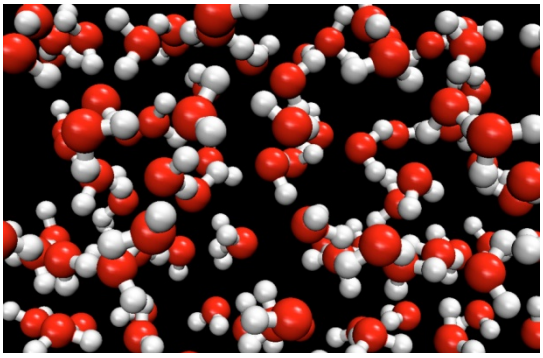


Figure 12. Ball and stick representation of the SPC model of water molecules. The atoms of oxygen and hydrogen are coloured in red and white respectively.

4.2.5 Ions in MD

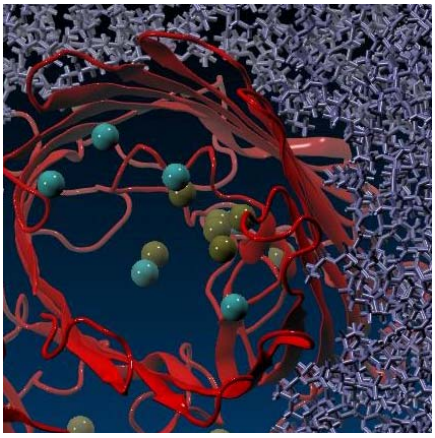


Figure 13. Ionic hydration of a Porin protein immersed in a double lipid bilayer. The protein is represented in red ribbon, the lipids in light grey stick. The ions are represented as a spheres coloured in light blue for Cl- ions while in green for Na+ ions.

Ion parameters for MD simulations are typically obtained by adjusting the LJ parameters to reproduce ionic hydration (Fig. 13) free energy and the radial density distributions. This strategy was first used for the determination of parameters for alkali (Åqvist, 1990) and alkaline earth (Åqvist, 1994) ions, but extended lately to other metals (Babu and Lim, 2006). Such sets of parameters describe a free ion without explicit consideration of its coordination capabilities although they can accurately reproduce the geometries of the first coordination shell. However, it is known that uncertainties of microscopic parameters, reflecting an incomplete experimental knowledge of the structural properties of ionic aqueous solutions at finite molality, translate into large differences in the computed radial distribution functions (Patra and Karttunen, 2004).

4.2.6 Long-range Interactions

The computation of the pair-wise non-bonded interactions (Fig. 14) is the most time-consuming part of a MD simulation as the evaluation of the forces scales quadratically with the number of atoms in the system. Accordingly, several approximations have been used in the last 20 years in order to reduce the computational effort needed (Loncharich and Brooks, 1989). One of the simplest approximations is to use a cut-off that defines the maximum distance between pairs of atoms before computing their energy of interaction, under the assumption that the force between atoms can be neglected when they are largely separated. Typical cutoffs used are about 1-2 nm. Since the LJ forces are short-range in nature, their contribution seems to be properly modeled with such cutoffs, although more accurate strategies have recently been reported (Lague et al., 2004; Klauda et al., 2007). However, dealing with the Coulombic term, the use of a cutoff is much more critical for the reliability of the simulations (Gilson, 1995; Cheatham et al., 1995). Accordingly, the treatment of long-range interactions has been an active field of research in the last years and is still a topic of considerable interest (Heinz and Hunenberger, 2005; Baumketner and Shea 2005). The most common alternative to truncation is to use the Ewald summation procedure (Allen and Tildesley, 1987) or the computationally more effective Particle Mesh Ewald or PME (Darden et al., 1993), where the long-ranged electrostatic interactions are calculated with fast Fourier transforms. These methods provide an exact solution for the electrostatic interactions for an infinite periodic system. The easiest way of simulating a periodic system is to treat it enclosed in a box and consider replicated boxes (to infinity) by rigid translation in all the three Cartesian directions, completely filling the space. This approximation is called periodic boundary

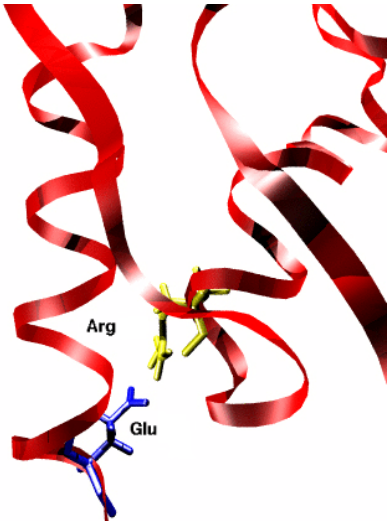


Figure 14. Representation of an electrostatic long-range interaction (salt bridge). The α-helices are represented in red ribbon while the amino acids as blue and yellow stick.

conditions or PBC.
The treatment used for electrostatics is an important issue for systems with abundance of charged groups such as nucleic acids (Cheatham et al., 1995; Louise-May et al., 1996; Beck et al., 2005), ions (Auffinger and Beveridge, 1995) or lipids (Venable et al., 2000; Tobias, 2001; Pandit and Barkovitz, 2002; Patra et al., 2003; Anézo et al., 2003; Patra and Karttunen 2004), but also for neutral systems like liquid water (Feller et al., 1996; Hess, 2002). In regard to lipid bilayers, the studies addressed the effects of using a cutoff on the bilayer structure, showing that the area per lipid is very sensitive to force field parameters, as well as to the way electrostatics are treated. Furthermore, the use of a cutoff together with some criteria to group atoms into neutral charge groups, may increase lipid order, resulting in thicker bilayers with smaller areas per lipid (Wohlert and Edholm, 2004).

4.2.7 Membrane Proteins Simulations

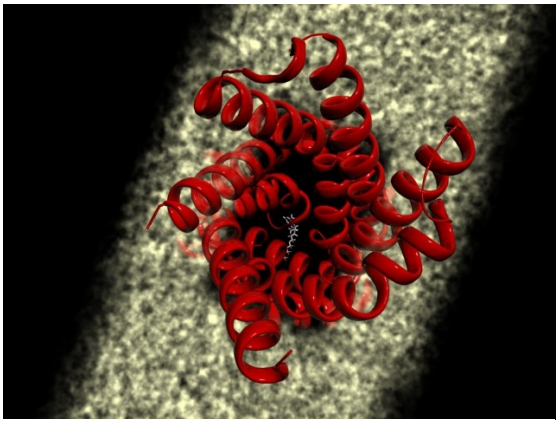


Figure 15. The figure shows the ADP/ATP mitochondrial carrier embedded in a lipid bilayer. The protein is represented in red ribbons and the membrane in yellow surface. At the bottom of the channel is represented the white stick structure of the ATP molecule.

Despite the importance of obtaining structures from X-ray crystals, a major drawback of this technique is that it does not permit to study membrane proteins in their natural environment. Thus, the development of other techniques that can provide structural information under more native conditions is of great interest. An emerging group of such methods are computer simulations of membrane proteins (Ash et al., 2004; Fanelli and Benedetti, 2005), and more specifically the use of MD with the proteins embedded on different model lipid bilayer (Fig. 15) (Tieleman et al., 2006). These approaches can provide useful information

about the structural features of a protein and about its dynamical behaviour that are not available from other experimental methods. Analogously to the difficulties associated to the experiments with membranes and membrane proteins, the necessity to mimic the hydrophobic environment provided by the membrane has been a major obstacle for the simulation of such proteins. One of the most important issues is the relevance of including explicitly the lipidic environment in the systems. Until recently, MD simulations of membrane proteins were carried out either using implicit models (Roux and Karplus, 1994; Strahs and Weinstein, 1997; Im et al., 2003) or hydrophobic solvents that mimicked the membrane environment instead (Åqvist, 2000; Deupi et al., 2004). These approximations have the advantage of being computationally cheaper than the simulations that include a more detailed description of the membrane. The increasing power of computers in the last years provided the opportunity to consider explicitly the lipid bilayer in the simulations. In this context, atomistic MD simulations of lipid bilayers represent a detailed microscopic picture of interactions and processes in biological membranes. However, the complexity of real membranes, containing many different proteins, lipids and other molecules, are still not presently affordable at this level of detail. For this reason, the study of model one-component pure hydrated lipid bilayers has been the focus of attention of a large amount of work published in the last years (Feller et al., 2002; Saiz ans Klein, 2002; Anézo et al., 2003). These studies, together with the effort in the characterization of lipid bilayers (Nagle and Tristram-Nagle, 2000), have been pivotal in order to extend the simulations to more complex membranes such as phospholipid mixtures (Gurtovenko et al., 2004; Leekumjorn and Sum, 2006) and/or the inclusion of other compounds such as cholesterol (Pitman et al., 2004; 2005) or ions (Sachs et al., 2004; Gurtovenko, 2005). Furthermore, the study of model peptides (Tieleman et al., 1999; Kandasamy and Larson, 2006) provided the necessary background for the simulations of membrane proteins. BR was the first integral membrane protein to be simulated embedded in a lipid bilayer (Edholm et al., 1995). However it was not until much later that the technique was more widely used. An updated review on the state of the art regarding common simulation setups and conditions for MD of lipid membranes with embedded or attached proteins can be found in two recent works (Ash et al., 2004; Sperotto et al., 2006). The conclusion of these studies is that the explicit inclusion of lipids is crucial for describing protein dynamics and to provide realistic simulations ranging over several tens or hundreds of nanoseconds and even reaching the microsecond (Martinez-Mayorga et al., 2006).

4.3 Introduction to the Molecular Docking Simulation

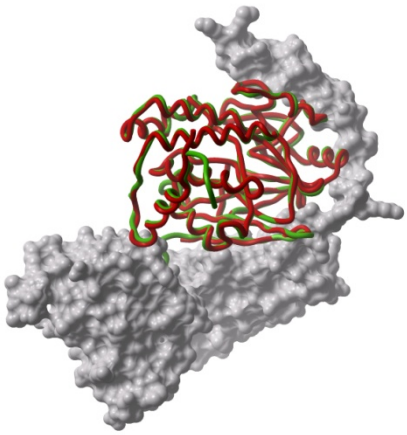


Figure 16. Protein-protein docking. The receptor is represented in white surface, while two conformations of the docked molecule are represented in red and green ribbon.

Molecular docking is a method to predict how two or more molecular structures, for example drug and enzyme or receptor and ligand, fit together (Fig. 16) (Lengauer and Rarey, 1996). In other words, the problem is like solving a 3D puzzle. In this 3D puzzle the geometric complementarity and the forces, that drive the interaction, are the two majors factors that drive the interaction between macro molecules.

Docking is frequently used to predict the binding orientation of small molecule drug candidates to their protein targets in order to in turn predict the affinity and activity of the small molecule. Hence docking plays an important role in the rational design of drugs (Kitchen et al., 2004). Given the biological and pharmaceutical significance of molecular docking, considerable efforts have been directed towards improving the methods used to predict docking.

4.3.1 Biological background

Molecular docking is used to predict the structure of the intermolecular complex formed between two or more molecules (Lengauer and Rarey, 1996). It's really important for the prediction of the correct structure of the protein-protein or protein-ligand complexes the study of the geometric complementarity between the surface of the molecules. In several molecular interactions the recognition specificity is indeed given by the high degree of the molecular surfaces complementarity. The different types of forces that come into play during molecular interactions are another important factor to take into account for the molecular docking (Vieth et al., 1998) . This means that we need to study the quality and quantity of forces between the interactive

particles. Depending on the computational method we may assign different weights to different kinds of forces. It is quite common to resort to certain simplifications, and some of the interacting forces are not used in the docking runs.
Often the forces, that play a crucial role in the molecular interactions, are divided into four categories:

- Forces with electrostatic origin
- Forces with electrodynamics origin
- Steric forces
- Solvent-related forces

Forces with electrostatic origin are due to the charges residing in the matter. The most common interactions are charge-charge, charge-dipole and dipole-dipole. These forces can be computed with the basic law of Coulomb. Dependencies on the distance are the following:

- charge-charge: 1/r
- charge-dipole: 1/r2
- dipole-dipole: 1/r3

In addition to purely electrostatic forces there exists also those with electro-dynamical background. The most widely known is probably the van der Waals interaction. Atoms, that are normally electrically neutral, may develop an induced dipole moment when an external electric field is applied. Van der Waals interaction is the force between the two induced dipoles, and it has a very short range. There are also forces between existing charges and induced dipoles.
Range dependences are the following:

- charge-induced dipole: 1/r4
- van der Waals: 1/r6

Solvent-related forces are due to the structural changes of the solvent. These structural changes are generated, when ions, colloids, proteins, etc. are added into the structure of solvent. For example, when water is acting as a solvent, one must take the polar nature of water molecules into account. Water molecules form hydrogen bonds, and for example the water mass around the

studied protein may turn into a highly organized structure. It is very hard to determine the solvent-related interactions, and their modelling depends very much on the way the actual solvent is modelled. Common thing to all these forces is the electromagnetic origin.

4.3.2 Molecular docking

The molecular docking technique presents several problems that can be identified as informatics problems. The search algorithm should create an optimum number of configurations that include the experimentally determined binding modes (Fig. 17) (Vieth et al., 1998). These configurations are evaluated using scoring functions to distinguish the experimental binding modes from all other modes explored through the searching algorithm (Martin, 1997). A rigorous searching algorithm would go through all possible binding modes between the two molecules. However, this is impractical due to the size of the search space. Consider a simple system comprised of a ligand with four rotable bonds and six rigid-body alignment parameters and a cubic active site measuring $10^3 Å^3$. The translational and rotational properties add up to six degrees of freedom. If the angles are considered in 10 degree increments and translational parameters on a 0.5. Å grid there are approximately 4x108 rigid body degrees of freedom to sample, corresponding to 6x1014 configurations to be searched. This would require approximately 2.000.000 years of computational time at a rate of 10 configurations per second. As a consequence only a small amount of the total conformational space can be sampled, and so a balance must be reached between the computational expense and the amount of the search space examined.

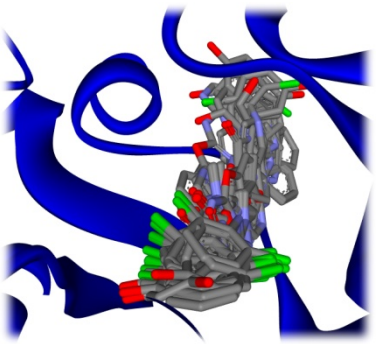


Figure 17. Protein-ligand docking. Particular of the binding site with all the conformation obtained for the ligand after the docking runs.

Some common searching algorithms include:

- Monte Carlo methods
- Genetic algorithms
- Fragment-based methods

- Point complementary methods
- Distance geometry methods
- Tabu searches
- Systematic searches

Current docking methods utilize the scoring functions in one of two ways. The first approach uses the full scoring function to rank a protein-ligand conformation. The system is then modified by the search algorithm, and the same scoring function is again applied to rank the new structure. In the alternative approach a two stage scoring function is used.

4.3.3 Monte Carlo methods

Figure 18. Representation of random distribution of data, representing the Monte Carlo sampling.

The Monte Carlo simulation method occupies a special place in the history of molecular modelling, as it was the technique used to perform the first computer simulation of a molecular system. The expression Monte Carlo simulation seems to be extremely general and many algorithms are called by that whenever they contain a stochastic process or some kind of random sampling (Metropolis and Ulam, 1949) (Fig. 18). In molecular docking the expression Monte Carlo usually means importance sampling or Metropolis method (Metropolis et al., 1953). The Metropolis method, which is actually a Markov chain Monte Carlo method, generates random moves to the system and then accepts or rejects the move based on a Boltzmann probability. The Monte Carlo methods play an important role in molecular docking but the variety of different kinds of algorithms is too large be considered here in detail.

Programs using MC methods include AutoDock (Morris et al., 2009), ProDock (Trosset and Sheraga, 2001), MCDOCK (Liu and Wang, 1999), QXP(Alisaraie et al., 2006) and Affinity.

4.3.4 Genetic algorithms

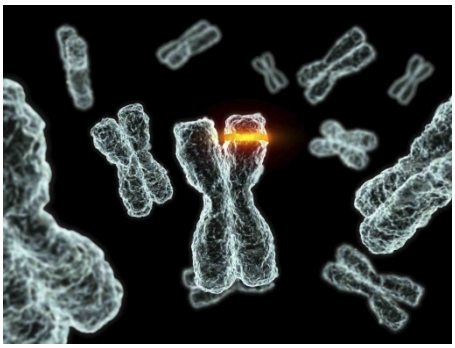


Figure 19. Genetic algorithm consider a population of data as a population of genes subjected to recombination.

Genetic algorithms and evolutionary programming are quite suitable for solving docking problems because of their usefulness in solving complex optimization problems (Morris et al., 1998). The essential idea of genetic algorithms is the evolution of a population of possible solutions via genetic operators (mutation, crossovers and migrations) to a final population, optimizing a predefined fitness function (Fig. 19). The process of applying genetic algorithms starts with encoding the variables, in this case the degrees of freedom, into a "genetic code", e.g. binary strings. Then a random initial population of solutions is created. Genetic operators are then applied to this population leading to a new population. This new population is then scored and ranked, and using "the survival of the fittest", their probabilities of getting to the next iteration round depends on their score. If the size of the population is kept constant, good solutions will occupy the population. It should be noted that genetic algorithms are well suitable for parallel computing. Some programs using GAs are GOLD (Jones et al., 1997), AutoDock (Morris et al., 2009), DIVALI (Clark and Ajay, 1995) and DARWIN (Taylor and Burnett, 2000).

4.3.5 Fragment-based methods

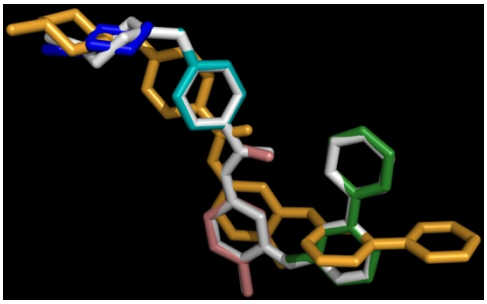


Figure 20. One ligand molecule is represented in wireframe orange. The other is depicted in several colour to show how it was generated starting from a big population. Each colour represent the portion of the ligand coming from a member of the population.

Fragment based methods can be described as dividing the ligand into separate portions or fragments, docking the fragments, and finally linking these fragments together (Fig. 20) (Congreve et al., 2008). These methods require subjective decisions on the importance of the various functional groups in the ligand, because a good choice of base fragment is essential for these methods. A poor choice can significantly affect the quality of the results. The base fragment must contain the predominant interactions with the receptor. Early algorithms required manual selection of base fragment, but this has been automated in newer implementations. Some well known programs using fragment based methods are FlexX (Kramer et al., 1999) and DOCK (Ewing and Kuntz, 1997).

4.3.6 Point complementary methods

These methods are based on evaluating the shape and/or chemical complementarity between interacting molecules. The interacting molecules are usually modelled in an easy way, for example using spheres or cubes as atoms. The ligand description is then rotated and translated to obtain maximum number of matches between ligand and protein surfaces, minus the number of volume overlaps. Additional constraints may be present, for example a demand for interacting surface normal to be approximately in opposite directions. Some algorithms use a 3D grid, which is placed over the protein and over the ligand. Each grid point is then labelled either open space or inside the ligand or protein. Then a correlation function is created and this function is optimized using rigid body translation and rotation. This often involves using traditional shape recognition algorithms like Fast Fourier Transform(FFT) with Fourier correlation theory. A high correlation score denotes good surface complementarity between the molecules. Because many

of the methods were originally created for protein-protein docking, the rigid body assumption is usually made.
This is a limitation in ligand-protein docking. However, some algorithms are addressed to ligand-protein docking and these allow some exibility. Examples of programs using point complementary methods are FTDOCK (Katchalski-Katzir et al., 1992), SANDOCK, FLOG (Miller et al., 1994) and the Soft Docking algorithm (Yoshimori et al., 2001).

4.3.7 Distance geometry methods

Many types of structural information can be expressed as intra- or inter-molecular distances. The distance geometry formalism allows these distances to be assembled and three-dimensional structures consistent with them to be calculated. The crucial feature is that it is not possible to arbitrarily assign values to the inter-atomic distances in a molecule and always obtain a low-energy conformation. Rather, the inter-atomic distances are closely interrelated and many combinations of distances are geometrically impossible. This enables fast sampling of the conformational space though not always resulting in good results. An example of a program using distance geometry in docking problem is DockIt (http://www.metaphorics.com/products/dockit.html).

4.3.8 Tabu searches

These methods are based on stochastic processes, in which new states are randomly generated from an initial state (referred to as the current solution). These new solutions are then scored and ranked in ascending order. The best new solution is then chosen as the new current solution and the same process is then repeated again. To avoid loops and ensure diversity of the current solution a tabu list is used. This list acts as a memory. It contains information about previous current solutions and a new solution is rejected if it reminds a previous solution too much. An example of docking algorithm using tabu search is PRO LEADS (Murray et al., 1999).

4.3.9 Systematic searches

These methods systematically go through all possible conformations and represent the brute force solution to the docking problem. All molecules are usually assumed to be rigid and interaction energy is evaluated from a force field model. Some constraints and restraints can be used to reduce the dimensionality of the problem.

4.4 Docking software

In addition to the existing large number of docking programs, there are also many molecular mechanics programs applicable to these problems. Despite the huge variety of available programs, no single program has been able to become recognized as a standard. Of course, there are some programs that are very widely used. Nevertheless it seems that the programs are not that easy to use and require some understanding of the underlying computational principles. This leads into situations, where people are using the same program they have been using before though better options could be available. It also seems that some of the existing programs are reaching a bit more mature state, since there seem to be an increasing number of commercial solutions available. Docking programs are usually sold in a package with other molecular design software. It should also be noted that the division made earlier is not very strict and many programs would fit into more than one category of methods. Tests have shown that there is not a significant difference in hit rates between different programs and they all produce false alarms. Because of this, combining different searching and scoring functions produces more reliable results. This has lead to the most successful docking programs usually being a collection of the methods described. It is also worth remembering that a molecular docking software is

Figure 21. View of a Desktop Linux operating system owning several bioinformatics tools.

only as good as its scoring function is. It does not help if we are able to create the right conformation not but able to recognize it. Probably the best known example of rational drug design has been the HIV-1 protease inhibitor. Starting with X-ray structures of HIV-1 protease, a group of scientists at DuPont Merck used docking and molecular design software to successfully design an inhibitor.

4.4.1 AutoDock

AutoDock (Morris et al., 2009) uses Monte Carlo simulated annealing (Metropolis et al., 1953) and Lamarckian genetic algorithm (Morris et al., 1998) to create a set of possible conformations. LGA is used as a global optimizer and energy minimization as a local search method. Possible orientations are evaluated with AMBER force field model (Pearlman et al., 1995) in conjunction with free energy scoring functions and a large set of protein-ligand complexes with known protein-ligand constants. The newest yet unreleased version 4 should contain side chain exibility. AutoDock has more informative web pages than its competitors and because of its free academic license, it is a good starting point when wondering into the world of molecular docking software.

4.5 Principal component analysis method

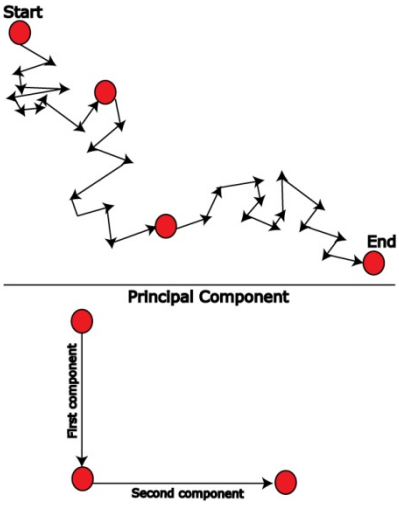


Figure 22. Principal Component Analysis. The trajectory in the upper panel is divided in the principal component describing the total motion.

Principal component analysis (PCA) is a useful statistical technique that has found application in fields such as recognition and image compression, and is a common technique for finding patterns in data of high dimensions. PCA involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components (Fig. 22). The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. PCA was invented in 1901 by Karl Pearson. Now it is mostly used as a tool in exploratory data analysis and for making predictive models (Loeffler et al., 2009; Zhang et al., 2009). PCA involves the calculation of the eigenvalue decomposition of a data covariance matrix or singular value decomposition of a data matrix, usually after mean centring the data for each attribute.

Covariance is a powerful measure in statistics since, unlike standard deviation or variance that are purely 1-dimensional, permits to calculate the relationship between more dimensions of the same data set. The formula to calculate the covariance is:

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

Typically covariance is always measured between 2 dimensions, for an *n*-dimensional data set, you can calculate *n!/(n-2)!*2* different covariance values.

A useful way to get all the possible covariance values between all the different dimensions is to calculate them all and put them in a matrix. The definition for the covariance matrix for a set of data with *n*-dimensions is:

$$C^{n \times n} = (c_{i,j}, c_{i,j} = \text{cov}(\text{Dim}_i, \text{Dim}_j))$$

Where $C^{n \times n}$ is a matrix with *n* rows and *n* columns, and Dim_x is the *x*th dimension. That means if you have an *n*-dimensional data set, then the matrix has *n* rows and columns (so is square) and each entry in the matrix is the result of calculating the covariance between two separate dimensions. Eg. the entry on row 2, column 3, is the covariance value calculated between the 2nd dimension and the 3rd dimension.
Each component of the matrix is defined by an eigenvector to which is associated an eigenvalue. In the case of molecular dynamics for example, each eigenvector describes each direction of the motion, so that we can decompose the total motion in all its components (Fig. 15). In this way we can consider only the eigenvectors that give the highest contribution to the total motion of a molecule, not considering the less informative ones (for example atoms vibration and bond stretching).

4.6 Cluster Analysis

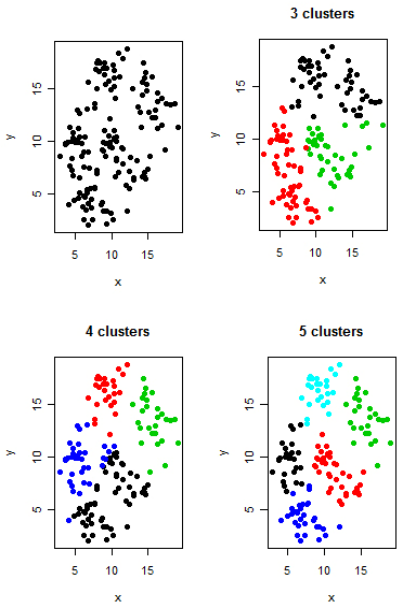


Figure 23. The data set represented in the first panel is divided in three different clusters, based on the parameter chosen.

The term cluster analysis encompasses a number of different algorithms and methods for grouping objects of similar kind into respective categories. Cluster analysis is an exploratory data analysis tool which aims at sorting different objects into groups in a way that the degree of association between two objects is maximal if they belong to the same group and minimal otherwise (Fig. 23). Given the above, cluster analysis can be used to discover structures in data without providing an explanation/interpretation. We deal with clustering in almost every aspect of daily life. For example, a group of diners sharing the same table in a restaurant may be regarded as a cluster of people. In food stores items of similar nature, such as different types of meat or vegetables are displayed in the same or nearby locations. There is a countless number of examples in which clustering plays an important role. For instance, biologists have to organize the different species of animals before a meaningful description of the differences between animals is possible. According to the modern system employed in biology, man belongs to the primates, the mammals, the amniotes, the vertebrates, and the animals. Note how in this classification, the higher the level of aggregation the less similar are the members in the respective class. Man has more in common with all other primates (e.g., apes) than it does with the more "distant" members of the mammals (e.g., dogs), etc. In short, whatever the nature of your business is, sooner or later you will run into a clustering problem of one form or another.

4.6.1 Statistical significance

The above discussion refers to clustering algorithms and do not mention anything about statistical significance testing. In fact, cluster analysis is not as much a typical statistical test as it is a "collection" of different algorithms that "put objects into clusters according to well defined similarity rules." Unlike many other statistical procedures, cluster analysis methods are mostly used when we do not have any a priori hypotheses, but are still in the exploratory phase of our research. In a sense, cluster analysis finds the "most significant solution possible." Therefore, statistical significance testing is really not appropriate here, even in cases when p-levels are reported (as in *k*-means clustering).

4.6.2 Areas of application

Clustering techniques have been applied to a wide variety of research problems. For example, in the field of medicine, clustering diseases, cures for diseases, or symptoms of diseases can lead to very useful taxonomies. In the field of psychiatry, the correct diagnosis of clusters of symptoms such as paranoia, schizophrenia, etc. is essential for successful therapy. In structural bioinformatics is really useful for example to identify similar tree dimensional structure in a large set of different proteins. In archaeology, researchers have attempted to establish taxonomies of stone tools, funeral objects, etc. by applying cluster analytic techniques. In general, whenever one needs to classify a "mountain" of information into manageable meaningful piles, cluster analysis is of great utility.

4.6.3 Applications in MD simulation

Through molecular dynamics simulations is possible to sample a large number of conformations of the protein of interest. The cluster analysis allows us to summarize the information contained into the trajectory through the choice of a suitable parameter. The parameters chosen for clustering can be different, for example: geometric, energetic, structural (i.e. conservation of secondary structure), and can be applied individually or all together. In the case of the work presented here, RMSD (Root Mean Square Deviation) was

used as a cluster parameter that allows to divide the MD sampled structures in different families on the basis of a geometric factor. The division in families allow us to uniquely identify the main structural changes that happened during the molecular dynamics simulation, and thus we can perform analysis directly on structures that generated the families. The structures that have identified the families have been used to carry out the docking runs.

4.7 Sequence alignment

4.7.1 Pairwise sequence alignment

In bioinformatics, a sequence alignment is a way of arranging the sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences.
Pairwise sequence alignment methods are used to find the best-matching piecewise (local) or global alignments of two query sequences.
The three primary methods of producing pairwise alignments are dot-matrix methods, dynamic programming, and word methods (Mount, 2004). One way of quantifying the utility of a given pairwise alignment is the 'maximum unique match', or the longest subsequence that occurs in both query sequence.
Needleman and Wunsch (1970) developed the classic dynamic programming algorithm that calculates the optimal alignment for a pair of sequences. Their algorithm determines, for every position of every possible combination of gaps, the maximum score between 1) inserting a gap in the first sequence 2) inserting a gap into the second sequence and 3) aligning the two characters.

4.7.2 Multiple sequence alignment

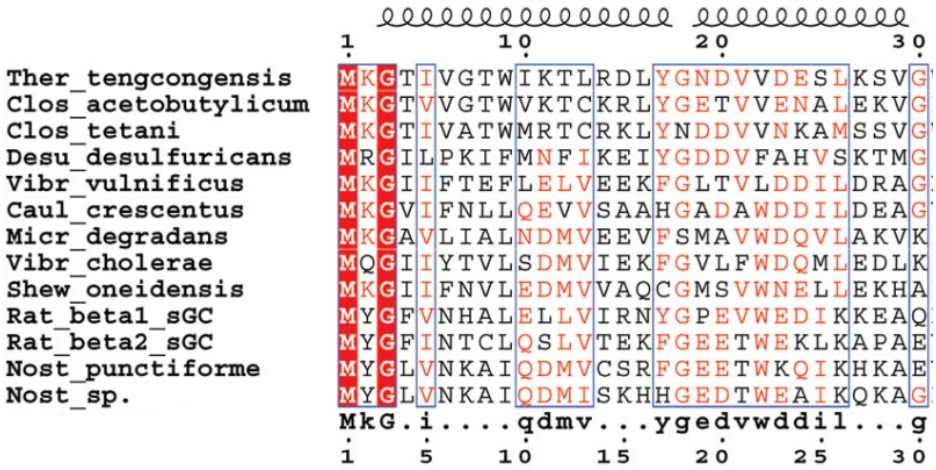


Figure 24. Representation of a multi-sequence alignment. Conserved residues are represented in red and underlined by a rectangular shape. On the top of the alignment the secondary structure is also represented.

A multiple sequence alignment (MSA) is a sequence alignment of three or more biological sequences, generally protein, DNA, or RNA (Lipman et al., 1989). In many cases, the input set of query sequences are assumed to have an evolutionary relationship by which they share a lineage and are descended from a common ancestor. From the resulting MSA, sequence homology can be inferred and phylogenetic analysis can be conducted to assess the sequences' shared evolutionary origins. Multiple sequence alignment is often used to assess sequence conservation of protein domains, tertiary and secondary structures, and even individual amino acids or nucleotides.
Most multiple sequence alignment programs use heuristic methods rather than global optimization because identifying the optimal alignment between more than a few sequences of moderate length is prohibitively computationally expensive.

4.7.3 Dynamic programming and computational complexity

A most direct method for producing an MSA uses the dynamic programming technique to identify the globally optimal alignment solution. For proteins, this method usually involves two sets of parameters: a gap penalty and a substitution matrix assigning scores or probabilities to the alignment of each possible pair of amino acids based on the similarity of the amino acids' chemical properties and the evolutionary probability of the mutation. For nucleotide sequences a similar gap penalty is used, but a much simpler substitution matrix, wherein only identical matches and mismatches are considered, is typical. The scores in the substitution matrix may be either all positive or a mix of positive and negative in the case of a global alignment, but must be both positive and negative, in the case of a local alignment. In the latter case it is essential that the average score be less than 0.
For n individual sequences, the naive method requires constructing the n-dimensional equivalent of the matrix formed in standard pairwise sequence alignment. The search space thus increases exponentially with increasing n and is also strongly dependent on sequence length. To find the global optimum for n sequences this way has been shown to be an NP-complete problem (Wang and Jiang, 1994) in 1989, Altschul (Altschul et al., 1989) introduced a practical method that uses pairwise alignments to constrain the n-dimensional search space. In this approach pairwise dynamic programming alignments are performed on each pair of sequences in the query set, and only the space near the n-dimensional intersection of these alignments is searched for the n-way alignment.

4.7.4 Progressive alignment construction

The most widely used approach to multiple sequence alignments uses a heuristic search known as progressive technique, that builds up a final MSA by combining pairwise alignments beginning with the most similar pair and progressing to the most distantly related. All progressive alignment methods require two stages: a first stage in which the relationships between the sequences are represented as a tree, called a guide tree, and a second step in which the MSA is built by adding the sequences sequentially to the growing MSA according to the guide tree. The initial guide tree is determined by an efficient clustering method such as neighbor- or UPGMA joining (Gascuel and Steel, 2006), and may use distances based on the number of identical two

letter sub-sequences (as in FASTA (Lipman, 1985) rather than a dynamic programming alignment).

Progressive alignments cannot be globally optimal. The primary problem is that when errors are made at any stage in growing the MSA, these errors are then propagated through to the final result. Performance is also particularly bad when all of the sequences in the set are rather distantly related. Most modern progressive methods modify their scoring function with a secondary weighting function that assigns scaling factors to individual members of the query set in a nonlinear fashion based on their phylogenetic distance from their nearest neighbours. This corrects for non-random selection of the sequences given to the alignment program.

Progressive alignment methods are efficient enough to implement on a large scale for many (100s to 1000s) sequences. Progressive alignment services are commonly available on publicly accessible web servers so users need not locally install the applications of interest. The most popular progressive alignment method has been the Clustal family (Higgins and Sharp, 1988), especially the weighted variant ClustalW (Thompson et al., 1994) to which access is provided by a large number of web portals. Different portals or implementations can vary in user interface and make different parameters accessible to the user. ClustalW is used extensively for phylogenetic tree construction, in spite of the author's explicit warnings that unedited alignments should not be used in such studies and as input for protein structure prediction by homology modelling.

Another common progressive alignment method called T-Coffee (Notredame et al., 2000) is slower than Clustal and its derivatives but generally produces more accurate alignments for distantly related sequence sets. T-Coffee calculates pairwise alignments by combining the direct alignment of the pair with indirect alignments that aligns each sequence of the pair to a third sequence. It uses the output from Clustal as well as another local alignment program LALIGN (Huang and Miller, 1991), which finds multiple regions of local alignment between two sequences. The resulting alignment and phylogenetic tree are used as a guide to produce new and more accurate weighting factors.

Chapter 5
– Methods Application

5.1 ADP/ATP mitochondrial carrier simulations

Three simulations have been carried out with the GROMACS MD package version 3.3.1 (Berendsen et al., 1995), using a modified version of GROMOS96 force field, to take into account protein-lipid interaction (Berger et al., 1997). Initial coordinates were taken from the PDB file of the bovine ADP/ATP carrier 1OKC, the unique carrier having a 3D structure solved by X-ray diffraction (Pebey-Peyroula et al., 2003). Comparison of the sequences shows that Ala114 and Val181 of the human carrier correspond to Ala113 and Val180 in the bovine protein, so these latter two residues have been mutated using the SwissPDB-Viewer program (Guex and Petisch, 1997) to proline and methionine to build the Ala113Pro and Ala113Pro/Val180Met mutants, respectively. The initial positions of the residues in the N- and C-terminal regions (Ser1, Lys294, Lys295, Phe296, and Val297) were not determined by X-ray data and therefore were modeled using the SwissPDB-Viewer program (Guex and Petisch, 1997). The whole system consisted of the carrier molecule (without co-crystallized CATR), embedded in a palmitoyl-oleoyl-phosphatidyl-choline (POPC) bilayer containing 255 lipid molecules and immersed in a box filled by 16913 SPC water molecules. Insertion of the protein in the membrane bilayer has been performed as described in Falconi et al., 2006 (Falconi et al., 2006). 19 solvent molecules were replaced by Cl^- counter-ions to keep the system electrically neutral. The temperature was kept constant at 300 K by means of weak coupling the temperature to a Berendsen thermostat (Berendsen et al., 1984), using a 0.1 ps time constant. The pressure was kept constant at 1.0 bar using a semi-isotropic Rahman-Parrinello barostat with 0.1 ps time constant (Parrinello and Rahaman, 1981), where the pressure normal to the bilayer (Z-axis) was coupled independently of lateral X-Y membrane plane extension. Long-range electrostatic interactions were calculated using the particle-mesh Ewald summation method (Darden et al., 1993), once short-range were taken into account by using a cut-off method (1.0 nm). All bonds were constrained using the LINCS algorithm (Hess et al., 1997), allowing a time-step of 2.0 fs. The total simulations time was 20 ns. All coordinates were saved every 0.25 ps. The overall procedure was as follows: the initial structure was subjected to a cycle of energy minimization of the

solvent and POPC atom around the protein solute using the steepest descent algorithm with 100 kJ mol^{-1} nm^{-1} tolerance using a step of 0.01 nm and position restraints on the protein. A 100 ps restrained molecular dynamics at 300 K was then performed to equilibrate the water and POPC molecules around the protein, followed by a second 1000 steps cycle of steepest descent energy minimization on the whole system with no restraints. Temperature was raised to the final value (300 K) in a stepwise manner: a series of 50 ps MD runs were carried out at increasing temperature of 50, 100, 200, and 250 K, respectively.
Standard parameters of the trajectories (RMSD, RMSF) were firstly analyzed by using GROMACS available analysis tools (Lindahl et al., 2001) and the first 10 ns of simulation time were discarded to take into account system equilibration. Helix backbone RMSD from an ideal α-helix and time dependence of the α-helix length were examined using the g_helix GROMACS tool (Garcia, 1992). Principal components analysis (Essential Dynamics) was performed as previously described (Garcia, 1992). Briefly, higher frequency fluctuations have been filtered out by diagonalization of atomic positional fluctuations covariance matrix calculated during production run and the corresponding eigenvectors and eigenvalues have been used to describe large-amplitude motion.
The images were obtained using the programs VMD (Humphrey et al., 1996) and UCFS Chimera (Pettersen et al., 2004).

5.2 Clustering of MD structures

The MD trajectory of the wild type protein coming from the above mentioned experiment has been clustered through the g_cluster program, belonging to the GROMACS 3.3.3 package (Lindahl et al., 2001), using the Gromos algorithm (Daura et al., 1999), in order to perform a second experiment. The clustering procedure starts with the calculation of the relative RMSD between all conformations of the trajectory to build a N(N-1)/2 matrix made by the RMSD values, where the N is the number of values extracted from the 40.000 saved structures. The program defines the "neighbour configurations" and the "number of neighbours" for each configuration. The configuration with the highest number of neighbours is taken as "representative" for the first cluster, which will be composed by this structure together with all of its neighbours. Then the structures of this first cluster are removed from the pool, and the procedure is iterated. In

this way all the configurations are assigned to only one cluster. The number of clusters depends on the threshold value adopted and in this case, to obtain a maximum number of 34 clusters, a threshold value of 1.5 Å was used. The first 15, out of 34, clusters have been used for the analysis since they contain almost all the sampled conformations (i.e. about the 96% of the trajectory structures).

5.3 Docking and clustering of the ATP^{4-} molecule

Protein-ligand docking runs, using Lamarckian genetic algorithm (Morris et al., 1998), have been carried out through the Autodock 4.0 program (Morris et al., 2009). 750 independent runs have been performed on the X-ray structure and 50 independent runs have been carried out for each of 15 cluster representative configurations identified from the MD trajectory getting a total of 750 runs. For the sake of clearness the docking on the X-ray structure has been called X-docking, while those on the structures extracted from the molecular dynamics trajectory have been called MD-docking. The centre of a 38.3 x 42.0 x 19.5 Å grid has been placed over the geometric centre of the matrix portion of the carrier (data not shown). The dimensions and the position of the docking grid chosen covers the entire region of the protein protruding from the lipid bilayer. The complexes have been inspected through an in-house modified version the g_mindist program, belonging to the GROMACS 3.3.3 package (Lindahl et al., 2001). This analysis was done in order to better characterize the binding sites by selecting the protein atoms within a distance of 3.5 Å from any ATP^{4-} atom.
The clustering of the ATP^{4-}docked structures has been carried out using the same procedure described above for the MD structures clustering but using a threshold value of 1.2 Å, that generates 5 clusters for the X-docking and 10 clusters for the MD-docking. In the case of the X-docking the first cluster includes 98% of the docked complexes. In the case of MD-docking the first two clusters include 85% of the docked complexes.
To assess ATP^{4-} binding specificity, 750 docking runs were carried out also for the GTP^{4-} molecule on the X-ray structure and on the structures extracted from the MD simulation.

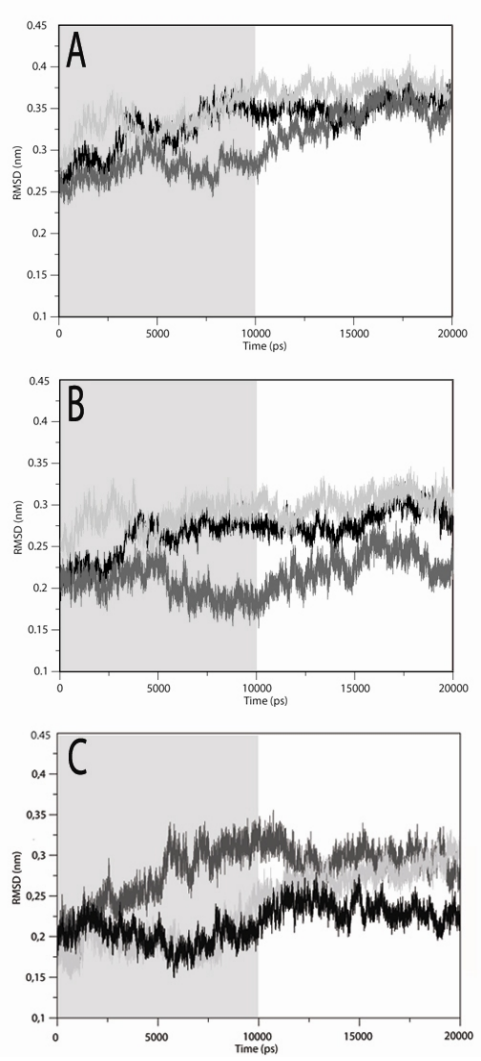
5.4 Multi sequences alignment

The ADP/ATP transporter sequences were extracted from SwissProt database (www.expasy.ch/sprot). Sequences from putative carriers were excluded, and also transporter sequences from plants that, due to their evolutionary distance and length variability, impede the correct execution of the multi-alignment algorithm. The multi-alignment of 33 sequences, shown in the Supplementary Data, has been performed using ClustalW program with default settings (www.ebi.ac.uk/clustalw/). When present, both isoform 1 and 2 of the same species were included.

Chapter 6
– Results

6.1 Results of the First Simulative System

6.1.1 RMSD analysis



The RMSD (Root Mean Square Deviation) of the backbone atoms of the wild type and the mutated proteins are reported in figure 25 A. The wild type and the double mutant reach a constant value after about 5 ns, whilst the single mutant only after 10 ns, hence all the analysis for the three systems have been carried out in the last 10 ns of simulation, i.e. from 10 to 20 ns.The RMSD calculated on the trans-membrane (TM) segments only (Fig. 25 B) shows lower values because the most fluctuating

Figure 25. Main chain RMSD of the entire protein **(A)**, of the trans-membrane helices **(B)**, and of the matrix region **(C)**. Black line represents the wild type simulation, the dark gray line represents the single mutant Ala113Pro and the light gray line represents the double mutant Ala113Pro/Val180Met. The grey box indicates the trajectories fractions that have not been included in the analyses.

part of the protein, i.e. the loops of the matrix region and even more the inter-membrane loops, that influence the overall RMSD value, are excluded from the calculation. The RMSD value of the Ala113Pro single mutant is lower than that of the wild type and double mutant (Fig. 25 B), indicating that the single mutation confers a higher rigidity to the TM segments, as can be seen from the RMSD calculated on the loop of the matrix region on the inter-membrane loops, which are more flexible in the

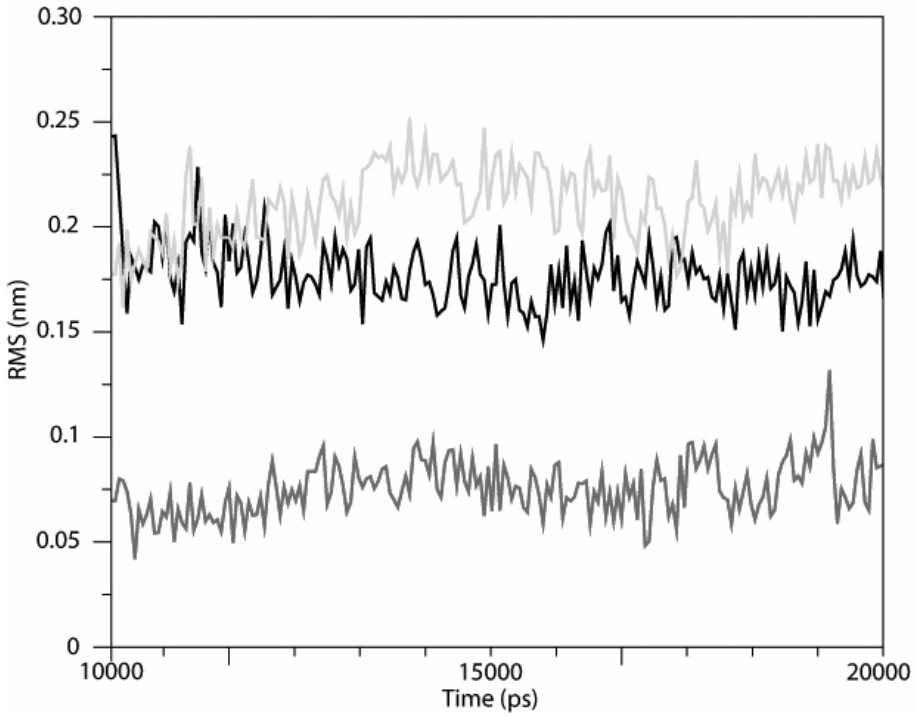


Figure 26. Mean square displacement from an ideal α-helix of the H3 helix backbone. Wild type, single and double mutant are represented by the black, dark gray and light grey lines, respectively.

single mutant with respect to the other two systems (Fig. 25 C). Moreover the deviation from an ideal helix calculated as a function of time for helix 3, where the mutation is located, indicates that in the single mutant this helix is more regular than in the other two systems (Fig. 26).

6.1.2 Principal Component Analysis

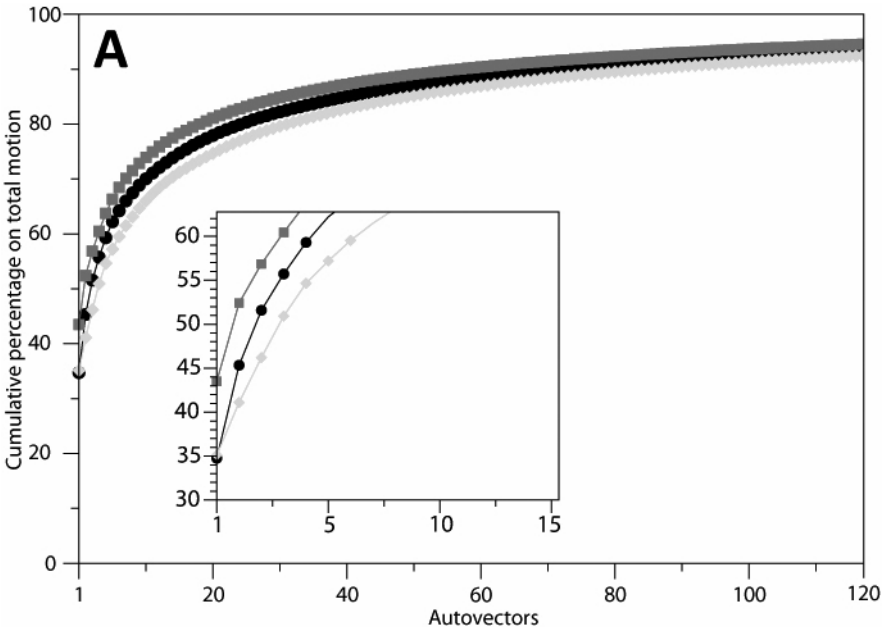


Figure 27. (A) Eigenvectors cumulative weight on total motion for: wild type (black circles), single mutant Ala113Pro (dark gray squares), and double mutant Ala113Pro/Val180Met (light gray diamonds).

The main dynamical features of the wild type, single and double mutants have been investigated performing a principal components analysis on the atomic positional fluctuations covariance matrix of the full Cα carbon atoms.

The analysis indicates that the fluctuations of the essential space described by the first 10 eigenvectors are about 70%, showing that they represent most of the protein motion (Fig. 27 A). In detail, the weight of the first eigenvector represents about 35% of global fluctuations for the wild type and double mutant simulations, and 43% of global fluctuations for the single mutant simulation (Fig. 26 A, inset). It is remarkable that the weight of the first eigenvector, almost identical for the wild type and double mutant proteins, it is different for the pathological single mutant that displays a preferential exchange for the ATP^{4-} molecule (Klingenberg, 2008).

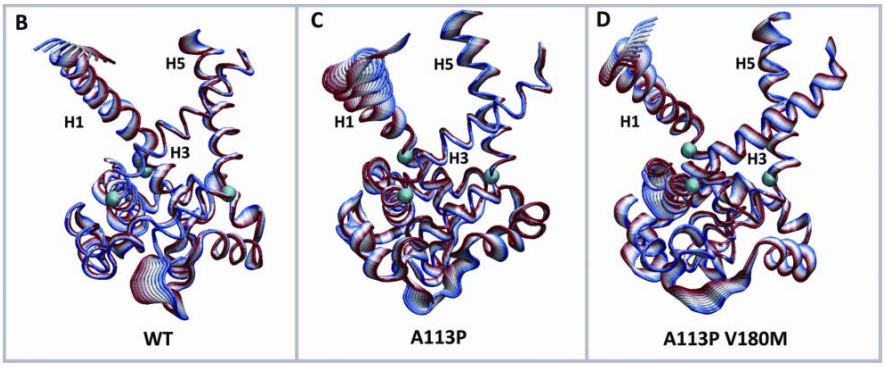


Figure 27. Projection of motion along the first eigenvector for the wild type **(B)**, single **(C)** and double **(D)** mutant. The amplitude of motion follows the colour scale from red to blue For sake of simplicity only the odd-numbered helices are represented.

The first eigenvector mainly describes the concerted motion of odd-numbered α-helices and matrix loops, suggesting that the trans-membrane and matrix regions are mechanically and functionally related although being well separated in the structure (Fiore et al., 2001; Berendsen et al., 1984) (Fig. 27 B, C, D).
Actually, the motions of the odd-numbered helices in association with the matrix loops have been indicated as crucial for the transmission of the global conformational changes associated with the transport mechanism (Fiore et al., 2001; Berendsen et al., 1984; Parrinello and Rahman, 1981). Proteins loops are the most flexible structural elements, hence it is expected to find the largest principal component highly dominated by their motion. The localization on the matrix loops of large component of the motion is consistent with our previous finding that these loops, in the simulation in the absence of the CATR inhibitor, achieve a flexible conformation characterized by a salt bridges network different from the one observed in the presence of the inhibitor (Pebey-Peyroula et al., 2003).

No relevant differences are observed among the three systems for the motions of the first and third carrier repeats filtered along the first

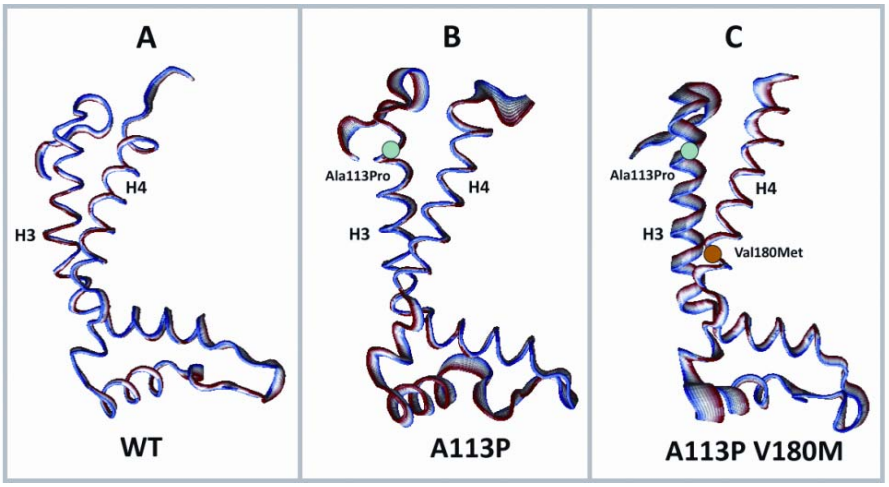


Figure 28. Projection of motion along the first eigenvector for the second repeat (H3-loop-H4) of the wild type **(A)**, single **(B)** and double **(C)** mutant. The amplitude of motion follows the colour scale from red to blue. The mutated residues Ala113Pro and Val180Met are represented by a cyan and orange sphere respectively.

eigenvector. On the other hand a main difference is observed in the motion along the first eigenvector for the second repeat (H3-loop-H4),
containing the mutations on helices H3 (Ala113Pro) and H4 (Val180Met) (Fig. 28 A , B, C).
The motion along the first eigenvector is higher in the matrix loop connecting helix H3 and H4 of the single and double mutants (Fig. 28 B and C, respectively), when compared to the wild type (Fig. 28 A). In the double mutant an increase of motion along the first eigenvector is also observed at the level of helix H3 (Fig. 28 C). It can be concluded that each of these mutations induces long distance effects: the Ala113Pro mutation, located on helix H3, generates the motion on the matrix loop (M2) (Fig. 28 B), while the Val180Met mutation, located on helix H4, in the presence of the other mutation, increases the motion of helix H3 (Fig. 28 C).

6.1.3 Analysis of the motions

Calculated correlated and anti-correlated motions, that a specific residue establishes with the overall remaining residues, have been added and the resultant sum is shown in figure 29. This simple parameter represents a

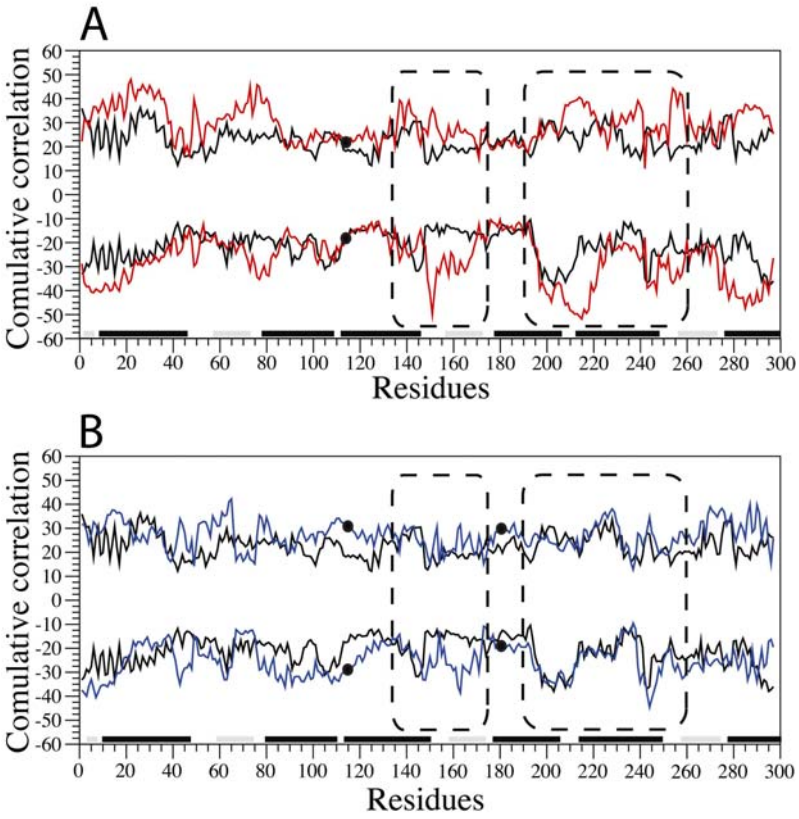


Figure 29. The cumulative correlation and anti-correlation for wild type and single mutant (black and red line respectively) panel **A**, and for wild type and double mutant (black and blue line respectively) panel **B**. The regions with major differences are highlighted by a dashed circle. The location of the single e double mutations are reported by a line with black dots. The black and grey boxes, reported at the bottom of the panels, represent the secondary structure of the protein (trans-membrane and matrix helices respectively).

quantitative measure that discriminates the similarity between the fluctuations directions of the residues in the trajectory.
Figure 29 A, showing the wild type and the single mutant, indicates that the single mutation causes a deep change in the profile of the cumulative correlation motions mostly localized in the region of loop M2 and on the helices 4 and 5, if compared with the wild type. On the contrary, in figure 29B, showing the wild type and the double mutant, the second mutation, localized on helix 4, restores cumulative correlations that are very similar to that observed in the wild type.

6.1.4 Analysis of the binding site salt bridges

The X-ray structure of the carrier in presence of CATR shows that the inhibitor is connected to the protein through direct, or water mediated, hydrogen bonds and salt bridges networks (Berendsen et al., 1984). Among the involved amino acids residues an important role is played by Arg79 and Arg234 that establish two direct contacts with the inhibitor (Fig. 30 A) (Fiore et al., 2001; Berendsen et al., 1984).
Arg79 has been identified as an ADP/ATP binding site by either a static docking study (Napoli

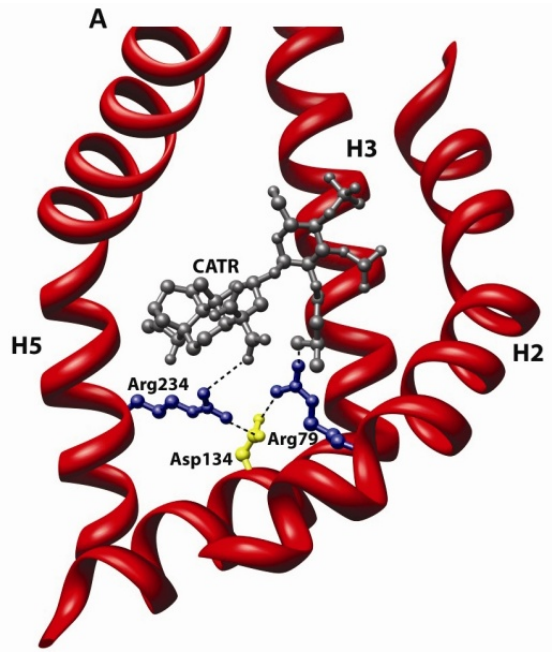
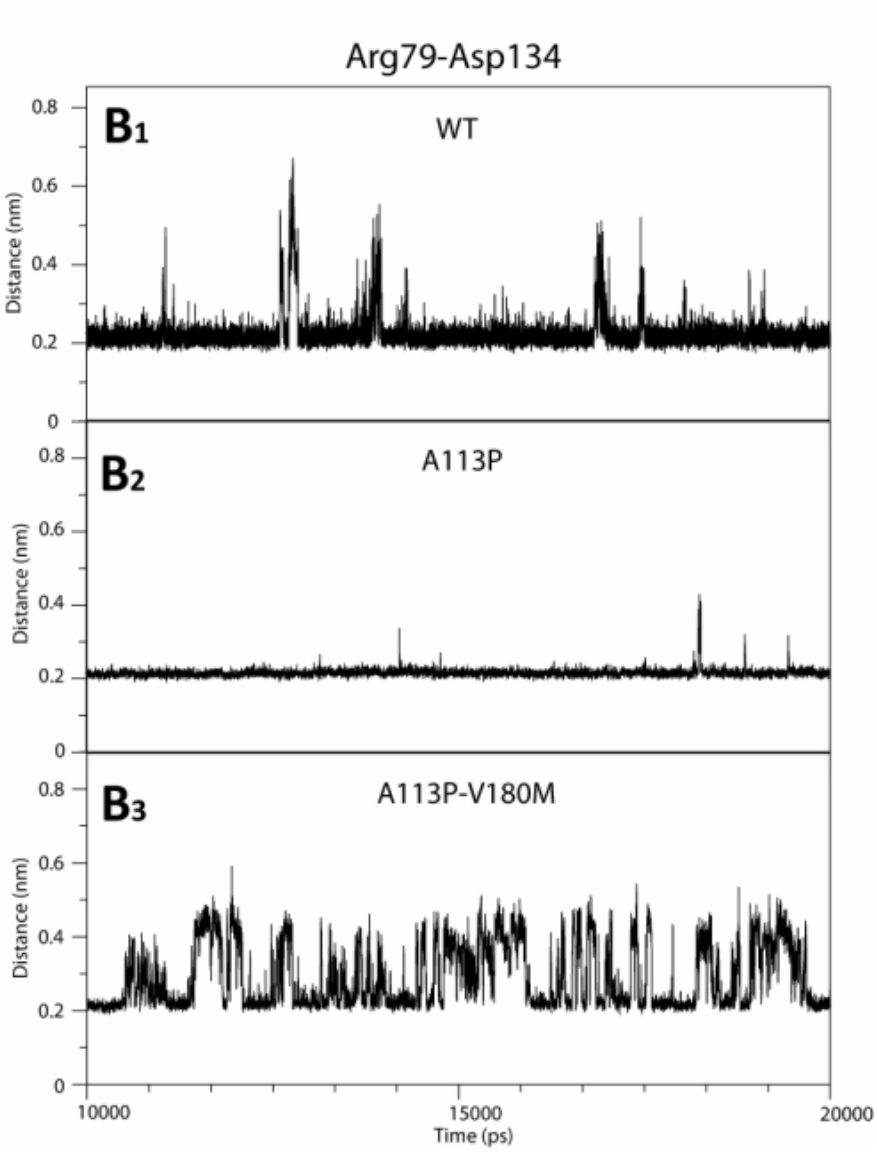


Figure 30. (A) Structure of the CATR binding site in the wild type carrier highlighting the Arg79, Asp134 and Arg234 cluster, forming the two salt bridges contacting the CATR inhibitor.

et al., 2001) and a very recent MD study of the carrier in presence of ADP^{3-} (Wang et al., 2009). In the wild type simulation Asp134 forms a salt bridge along all the trajectory with both Arg79 and Arg234, allowing the two arginines to gain a position able to directly contact the incoming inhibitor/substrate (Fig. 30 B1 and C1). In the single mutant the salt bridge Asp134-Arg79 is strong and stable for all the simulation time (Fig. 30 B2), whilst the Asp134-Arg234 interaction is almost completely lost (Fig. 30 C2). The double mutant has a behaviour similar to the wild type, maintaining both the Asp134-Arg79 and Asp134-Arg234 interactions (Fig. 30 B3 and C3). Note that Val180 is located at the border of second ADP/ATP contact point, as described by Robinson and Kunji (Napoli et al., 2001). Imposing a 4.0 Å threshold for the interaction existence, the Asp134-Arg79 and the Asp134-Arg234 salt bridges are present for 98% and 99% in the wild type, for 100% and 27% in the single mutant and for 82% and 87% in the double mutant of the total simulation time, respectively.

It is worthwhile noting that these interactions are the only significant differences between the wild type and the mutants among all the interactions occurring in the inhibitor/substrate binding pocket.


Arg79-Asp134
B1
WT
B2
A113P
B3
A113P-V180M
Distance (nm)
0
0.2
0.4
0.6
0.8
10000
15000
20000
Time (ps)

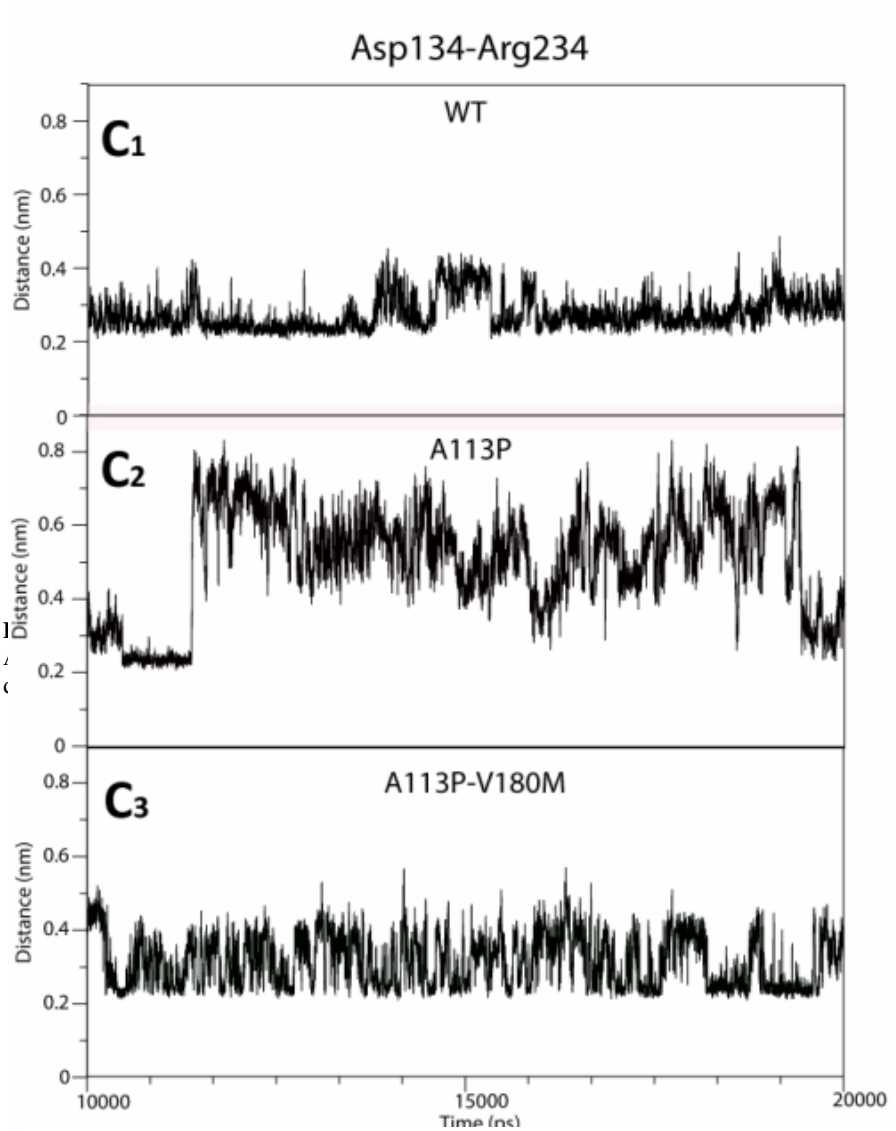


Figure 30. Time evolution of the distance between the centres of mass of residues Asp134 and Arg79 for the wild type **(B_1** and **C_1)**, the single **(B_2** and **C_2)** and the double mutant **(B_3** and **C_3)**.

6.2 Results of the second simulative system

6.2.1 Clustering of MD simulation

The root mean square deviation (RMSD) of the trajectory is depicted in figure 31. The A panel shows the RMSD calculated on the Cα atoms,

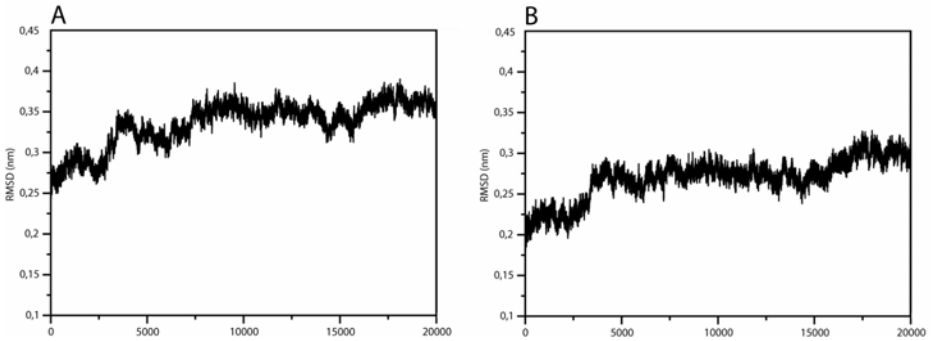


Figure 31. Cα atoms RMSD plotted as a function of time for the whole protein **(A)** and for the six trans-membrane helices (B).

while panel B reports the RMSD restricted to the six trans-membrane helices.

The latter ones reach a value lower than that found for the whole protein (Fig. 31 A) because of the higher intrinsic mobility of the loops and of the small matrix-facing helices, that being outside the lipid bilayer,

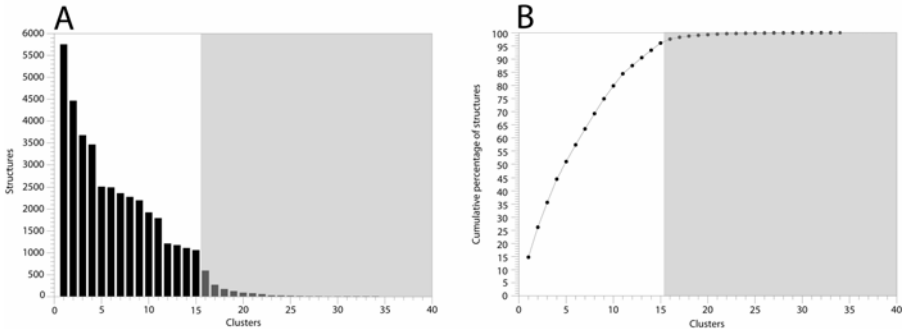


Figure 32. (A) Histogram of the clusters extracted from the analysis of the 20 ns MD trajectory. The gray box represents the lowest populated clusters not taken into account for the docking analysis. **(B)** Cumulative percentage of the number of structures belonging to each cluster. The gray box shows the cumulative percentage related to the clusters not taken into account for the docking analysis.

explore a larger conformational space. The ensemble of carrier conformations, generated from the MD simulation, was pruned and grouped in families of similar structure via a clustering analysis performed using the g_cluster (Lindahl et al., 2001) tool of the GROMACS package. The procedure is repeated, by iteration until all structures are assigned to a cluster. The structures representative of each cluster have been selected to carry out the docking with the ATP^{4-} molecule. Choosing a threshold of 1.5 Å, 34 clusters are identified and they are ordered following their population in figure 32 A. The first 15 clusters, that cover more than 95% of the sampled conformations (Fig. 32 B), have been selected for the docking analysis.

6.2.2 Docking of ATP^{4-} molecule

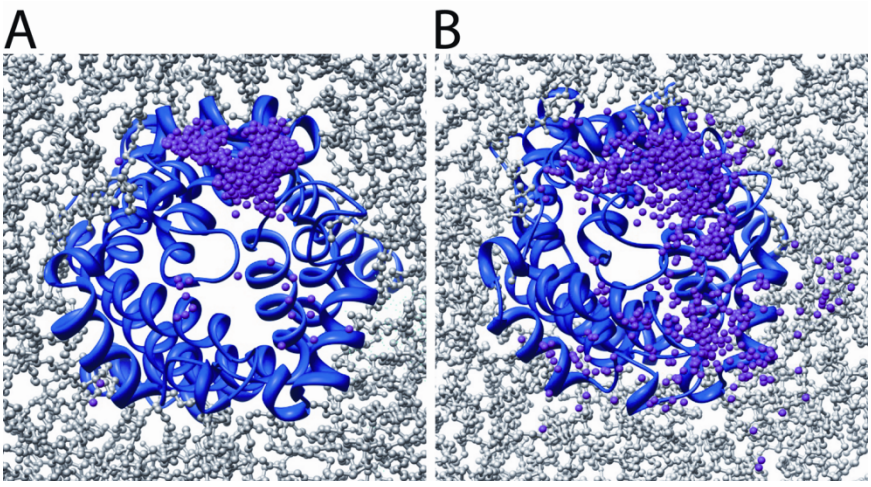


Figure 33. (A and **B)** Ribbon representation of the X-ray and MD average structure of the protein respectively, coloured in blue and immersed in a lipid bilayer, represented in gray ball and stick. In both cases the pink balls represent the spread of the centre of mass of 750 ATP^{4-} molecules docked on the X-ray structure **(A)** and of 750 ATP^{4-} molecules docked on the representative structures of the 15 clusters extracted from the MD simulation **(B)**.

Docking is a simulative method that predicts the preferred orientation of one molecule with respect to a second one, evaluating the strength of association using energetic and geometric scoring functions.

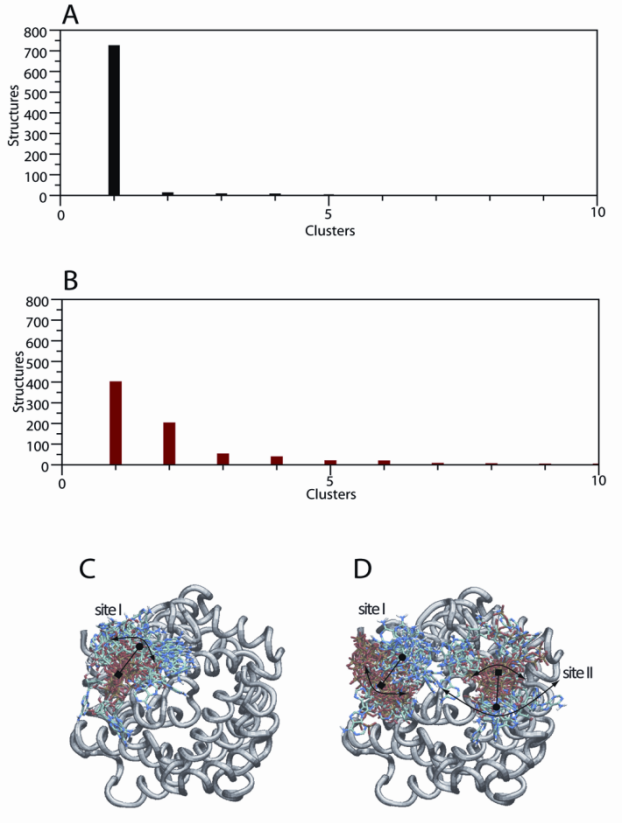


Figure 34. Histogram of the clusters of the 750 ATP^{4-} docked molecules on the X-ray structures **(A)** and on the 15 representative structures extracted from the 20 ns molecular dynamic simulations **(B)**. Localization on the protein (gray ribbon) of the ATP^{4-} molecule as observed in the X-docking **(C)** and in the MD-docking **(D)**. ATP^{4-} is schematically represent by a circle, a line and a square representing the base, the sugar and the phosphates respectively. The width of the arrows indicates the spread of the position.

The molecular docking between ATP4- and the carrier was accomplished using the Autodock program (36,37), and the whole matrix region of the protein, protruding from the phospholipid bilayer, has been selected in order to take into account the widest possible area of substrate interaction with the protein (data not shown). 750 docking runs have been carried out on the X-ray structure of the mitochondrial ADP/ATP carrier, and 50 docking runs have been carried out for each of the 15 structures, each one representative of the 15 clusters, again for a total of 750 complexes. The final result of the 750 runs of the X- and MD-docking is represented in figure 33 A and B respectively where the ATP^{4-} molecule is represented by its centre of mass. In the case of the X-docking (Fig. 33 A) a preferential interaction site is identified, whereas for the MD-docking (Fig. 33 B) the centres of mass of the docked

ATP^{4-} molecules are more spread out on the protein, covering approximately half of the matrix-side surface.
The structures of the bimolecular carrier-ATP^{4-} complexes resulting from X-docking and MD-docking have been clustered to select the preferred complexes. The results of the clustering of the X-docking runs (Fig. 34 A) show the presence of a unique family of ATP^{4-} configurations, having a unique preferential orientation, as represented by site I in figure 34 C where the position of phosphate moiety of the ATP^{4-} is indicated by a square, the sugar by a line and the base by a circle. In the case of MD-docking (Fig. 34 B) several clusters are identified, but two are the predominant ones, representing about 85% of the 750 ATP^{4-} docked complexes. The most representative one, including about 400 structures, identifies the same site observed in the X-docking, indicating that it is preserved during the dynamics and is represented by site I in figure 34 D. The other representative site is detected in about 200 structures and it is identified as site II in figure 34 D.
GTP^{4-} docking has been performed to evaluate the specificity of the ATP^{4-} binding sites found in the X- and MD-docking. In both the docking simulations the GTP^{4-} interact with the carrier through the phosphate groups but not through the base, strongly suggesting that the identified sites are specific for ATP^{4-}. Moreover in the MD docking GTP^{4-} is found spread over the entire matrix region, indicating the absence of specific interaction sites (data not shown). The binding free energy for ATP^{4-} and GTP^{4-} are very different, their average values for the X-docking runs being -8,13 and -2.52 Kcal/mol, respectively. For the MD-docking the binding free energy average values for ATP^{4-} are: -8.47 Kcal/mol for site I and -8.22 Kcal/mol for site II, while the average values for the docked GTP^{4-} molecule is -4.76 Kcal/mol, confirming the specificity of the binding sites for ATP^{4-}.

6.2.3 Binding sites analysis

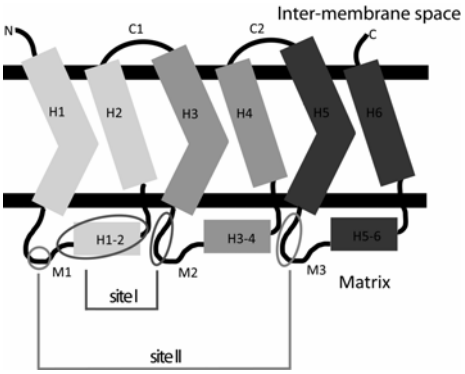


Figure 35. 2D topology view of the ADP/ATP carrier indicating the position of the protein regions (sites I and II) involved in the formation of the ATP^{4-} binding sites. The three carrier repeats are evidenced using a different grey scale.

Table 2
Percentage of contacts

A

		Phosphate	Sugar	Base
Sitel	Gln 43	1%	33%	**66%**
	Lys 48	**63%**	25%	12%
	Arg 59	**59%**	21%	20%
	Glu 63	**70%***	13%	17%
	Gly 145	1%	40%	**58%**
	Lys 146	**55%**	27%	18%
	Gln 150	1%	32%	**67%**

B

		Phosphate	Sugar	Base
Sitel	LYS 48	**58%**	20%	22%
	Gln 49	**45%**	25%	30%
	Arg 59	**53%**	30%	17%
	Glu 63	14%	**56%**	30%
	Gly 145	9%	40%	**51%**
	Lys 146	**46%**	29%	25%
	Gln 150	20%	23%	**58%**
Sitell	Lys 42	**55%**	22%	23%
	Gln 240	26%	28%	**46%**
	Arg 243	**50%**	22%	28%
	Asp 247	22%	**53%**	25%
	Met 249	22%	35%	**43%**

Using a modified version of the g_mindist program, belonging to the GROMACS 3.3.3 package (Lindahl et al., 2001), the contacts between all the protein's and ligand's atoms have been calculated every time they are at a distance lower than 3.5 Å and reported in Table 2 when observed in at least 10% of the complexes. In the X-docking runs the amino acids involved in the ATP^{4-} binding are Gln43, Lys48, Arg59, Glu63, located at the C-terminal region of the M1 loop and on the small helix H1-2, and Gly145, Lys146 and Gln150, located on the M2 loop (Fig. 35). Analysis of the percentage of interaction with the three portions of the substrate (Table 2 A) indicates that the positively charged amino acids (Lys48, Arg59, Lys146) are involved in the interaction with the negative phosphate groups, while the glutamines (Gln43 and Gln150) are involved with the base moiety. Interestingly, also Glu63, a negatively charged amino acid, is involved in the formation of the ATP^{4-} interaction site, being often at a distance lower than 3.5 Å from the phosphate moiety (Table 2 A and Fig. 36 A). The interaction between two molecules having the same charge is extremely disadvantageous, but the role of Glu63 is indirect. Its presence is required to stabilize Arg59 and Lys146 forming a salt bridges network and so providing the right

orientation to the positive charges to appropriately interact with the phosphate moiety. The last amino acid forming the ATP^{4-} interaction site is Gly145 that creates a cavity where the base moiety of the ATP^{4-} molecule can be accommodated.

The amino acids involved in the interaction with the ligand in the first site of the MD-docking are: Lys48, Gln49, Arg59, Glu63, Gly145, Lys146 and Gln150 (Table 2 B). They largely overlap with those found in site I for the X-docking, apart from Gln43 that in the X-docking interacts with the base moiety, now replaced by Gln49 that however interacts with the phosphate (Table 2,). A further difference is due to Glu63 that in the MD-docking interacts with the sugar moiety (Table 2 B and Fig. 36 B).

The MD-docking identifies a new interaction site, formed by: Lys42, Gln240, Arg243, Asp247, Met249, with Lys42 located on the M1 loop and the other four amino acids located on the M3 loop (Fig. 35). The amino acid composition of this site is similar to the one observed in site I, being composed by two positive amino acids (Lys42 and Arg243) interacting with the phosphate groups, a glutamine (Gln240) interacting with the base, a negatively

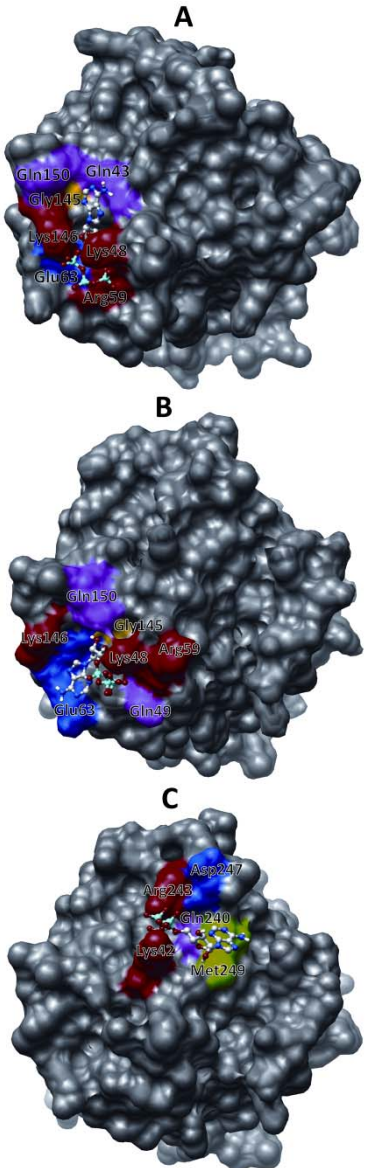


Figure 36. Representative carrier-ATP^{4-} complexes showing the site I identified in the X-docking **(A)** and the site I **(B)** and site II **(C)** identified in the MD-docking. Red and blue spots indicate the positive and negative amino acids charge, respectively. Magenta, yellow and green spots represent the Gln, Gly and Met residues.

charged amino acid (Asp247) interacting with the sugar moiety and a methionine (Met249) assisting Gln240 in the base interaction (Table 2 B, Fig. 35, 36 C).

For the two sites extracted from the MD-docking the orientation and the spread of the ligand over the protein has been correlated to the mobility of the interacting amino acids. Figure 35 shows the per-residue RMSF (root mean square fluctuations) of the whole protein calculated along the entire molecular dynamics trajectory, where the residues composing the first and second binding site are marked with bold and grey lines, respectively. Lys48, Lys146 and Arg59 have fluctuation values higher than the other interacting amino acids explaining the large spread of the phosphate groups observed in site I in Fig. 34 D. Similarly, the low spread observed for the base and the sugar moieties is explained by the low fluctuations of the amino acids interacting with them (Glu63, Gln150) (Fig. 37). In the case of site II, the spread of the phosphates is intermediate since Arg243 has a high fluctuation whilst Lys42 has a low fluctuation value (Fig. 36 D and Fig. 37). The spread, is higher for the base and sugar, which are contacted by amino acids (Gln240, Asp247, and Met249) characterized by large fluctuation values (Fig. 37). It is interesting to note that there is a spread in base position also in the rigid X-docking and this can be explained by the presence of two glutamine residues that can equally interact with the adenine ring (Fig. 34 A and Table 2 A).

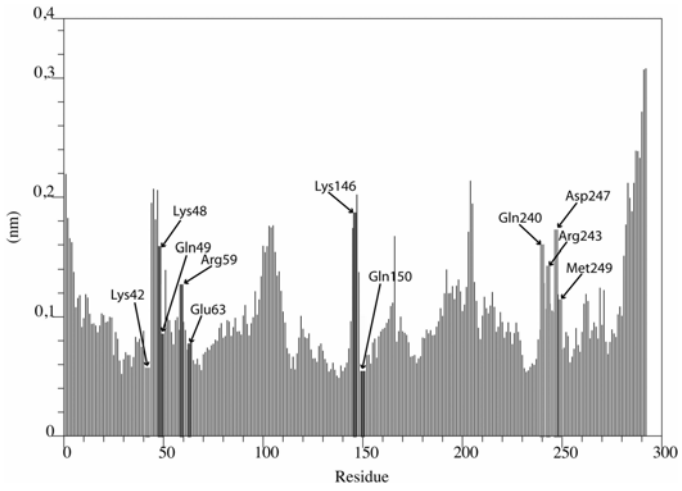


Figure 37. Average per residue root mean square fluctuations of the ADP/ATP carrier. The dark and light grey bars indicate the fluctuations of the residues belonging to site I and II, respectively. The black thin bars show the fluctuations of the residues not involved in the ATP recognition.

Chapter 7
– Conclusions

7.1 Conclusions of the mutants simulations of the ADP/ATP mitochondrial carrier

The detailed analysis of the simulations of the pathologic Ala113Pro single mutant and of the functional restoring double mutant, Ala113Pro/Val180Met, in comparison with the wild type, has allowed us to evaluate the structural-dynamical features introduced by these mutations that revert a pathologic mutant to the physiological function. The principal component analysis indicates that, in the three simulations, the first eigenvector is dominated by the coupled motion between the matrix loops and the odd-numbered helices, that may have a pivotal role in modulating the transport mechanism. The single mutation mainly affects the motion of the matrix M2 region, lowers the deviation of the TM segments from the starting structure (Fig. 25 B), and regularizes the H3 helix structure (Fig. 26).
At the same time, the single mutation causes a change in the correlated and anti-correlated protein motions (Fig. 29 A, B). The differences in the correlated and anti-correlated motions cause a drastic rearrangement of the global dynamic characteristics of the functional protein. In detail, the presence of a second proline, introduced by the single mutation on helix 3, causes, if compared with the native protein, a deep change in the anti-correlation motions of the M2 loop and helices 3 and 4 (Fig. 29 A). This change is minimized by the presence of the second mutation on helix 4, restoring a protein behaviour similar to that of the wild type (Fig. 29 B). The restoring of the correlated and anti-correlated motions is symptomatic to the re-establishment of the structural-dynamical characteristics of a functional carrier.
The increase in helix regularity upon the introduction of a proline is an almost unexpected result. However Pro113 is introduced by mutation at the N-terminal region of the H3 helix before the native Pro132. These two prolines enclose 19 residues and impose rigidity to this TM segment confining it in an ideal α-helix conformation (Fig. 26). The increase of H3 helix regularity is correlated to the strengthening of the inter-helical (H3-H2) Asp134-Arg79 salt bridge interaction that, constraining the carrier to a specific conformation, prevents the occurrence of the inter-helical (H3-H5) Asp134-Arg234

interaction observed in the wild type (Fig. 30 A, B_1-B_2-B_3 and C_1-C_2-C_3). Arg234 and Asp134 are part of the MCF motif and the importance of this salt bridge has been also predicted from revertant studies (30). The loss of the Asp134-Arg234 salt bridge restrains the protein motion and impedes the carrier to undergo the conformational transition necessary for the transport process. The introduction in the H4 helix of the Val180Met mutation restores the H3 helix native features, permitting to re-establish the electrostatic interaction of Asp134 with both Arg79 and Arg234 (Nelson et al., 1998) (Fig. 30 B_3 and C_3). Val180 residue has been proposed to be in proximity of an ADP/ATP contact point, important for the adenine ring hydrophobic interaction (Robinson and Kunji, 2006). Indeed, the second mutation, increasing the H3 helix flexibility, weakens the Arg79-Asp134 interaction and strength the Asp134-Arg234 salt bridge (Fig. 30 B_3 and C_3). The recovery of this interaction is crucial not only for the orientation of the substrate in the binding site, but also for the maintenance of the relative position between the three odd-numbered helices (Nury et al., 2006).
The hypothesis formulated by Wang et al. (Wang et al., 2008) is in line with the picture described by the MD analysis, in which the introduction of the helix-breaking proline of the single mutation may induce a loss of plasticity in the carrier structure. The drastic conformational changes occurred may impede the carrier gating function that consequently magnifies the proton-conducting activity intrinsically associated with the carrier (Wang et al., 2008). The presence of an internal flexibility is thus a crucial point for the establishment of the carrier functionality and this work shows that it can be lost or gained with single mutations able to alter long range communications along different protein regions.

7.2 Conclusions of the ATP interaction sites searching the on the ATP/ADP mitochondrial carrier matrix region

In this section my attention has been focused on the matrix carrier region located out of the membrane in order to describe an hypothetical binding site for the ATP^{4-} molecule. The analysis has been carried out both on the X-ray structure and on representative structures extracted from a 20 ns molecular dynamics simulation. This allowed us to relate the data obtained from the two systems in order to understand how the structural sampling due to MD simulation in the absence of the inhibitor is related to the retention or formation of different binding sites for the substrate. 750 docking runs carried

out on the crystal structure clearly show the presence of a unique ATP4- recognition site located on the H1-2 helix and on the M2 loop (Fig. 34 C and Fig. 35), in line with the experimental results, that identify hydrophobic residues flanking Cys159 as putative partners for the interaction with the substrate (Majima et al., 1998). This result indicates that the region that protrudes within the mitochondrial matrix has a preformed recognition site, in spite of the fact that the protein is in a conformation arranged for transport from the cytoplasmic side. Indeed, kinetic experiments show that the protein can form an ADP^{3-}-carrier- ATP^{4-} ternary complex during transport (Duyckaerts et al., 1980), suggesting that ATP^{4-} from the matrix side may still bind CATR-bound AAC, although it cannot be transported. Evidence of the existence of a specific binding site for ATP^{4-}, comes also from the comparison of the binding free energy obtained in the X-docking of ATP^{4-} and GTP^{4-}, respectively. Their values of -2.52 and -8,13 Kcal/mol, respectively, correspond to a difference in their dissociation constants of the order of 104, indicating that the carrier displays a site specific for ATP^{4-}. The ATP^{4-} interaction site is formed by three positively charged amino acids (Lys48, Arg59, Lys146), two polar amino acids (Gln43, Gln150), one negatively charged amino acid (Glu63) and a glycine (Gly145). These amino acids determine the binding of the substrate in a unique orientation where the phosphate groups are stabilized by the positively charged residues (Lys48, Arg59, Lys146) (Table 2). The positive residues rigidly anchor the phosphates groups, whilst the base moiety can change its position because it can equally interact with two glutamines (Gln43 and Gln150) located one in front of the other (Table 2 and Fig. 36 A).

This site (site I) is found also in the MD-docking runs, with the same global characteristics and a few important differences (Fig. 32 B). The amino acids composing it are the same as for the X-docking with the exception of Gln49 that replaces Gln43. Some differences are also observed in the 3D arrangement. In detail, in the MD-docking Gln49 is mainly interacting with the phosphates while Gln150 retains its interactions with the purine ring. The other and more notable difference involves Glu63 that is interacting with the sugar, because it is not involved anymore in the electrostatic network that is present in the crystal structure, where it forms an electrostatic triad with Arg59 and Lys146. Rearrangements of salt bridges have been proposed to be important in the transport mechanism (Nelson et al., 1998; Pebey-Peyroula et al., 2003; Nury et al., 2006) and it was proposed to be the triggering mechanism of substrate induced conformational change (Wang and Tajkhorshid, 2008; Dehez et al., 2008; Robinson et al., 2008). Indeed,

significant changes in the salt bridge networks and the formation of cavities at the level of the matrix region, have been assessed by MD simulation performed in absence of CATR (Falconi et al., 2006). As can be seen in a representative snapshot in figure 34 B, Glu63 is not bridging the same positive residues as before and it is now able to make a bond with a more suitable partner, like the sugar moiety. The different spread in the positioning of the ATP^{4-} molecules in the X- and in the MD-docking (Fig. 34 C and D) are due to the RMSF value of the amino acids interacting with the ligand. In site I Glu63 and Gln150, that interact with the sugar and the base moiety respectively, have RMSF value lower than Lys48, Arg59 and Lys146, that interact with the phosphate. It is worthwhile noticing that, in spite of the rearrangements of the residues and of the orientation dispersion of the substrate moieties, site I, found in the crystal structure, is maintained in the structures extracted from the MD simulation (Fig. 34 C and D).

In the MD-docking a new interacting site, site II, appears that is not mapped in X-docking (Fig. 34 D). The new site is composed by two positively charged amino acids (Lys432, Arg243), one polar amino acid (Gln240), one negatively charged amino acid (Asp247) and one hydrophobic amino acid (Met249). The physico-chemical characteristics of site II are almost the same as observed in site I, in fact the phosphates moiety of the substrate is stabilized by the positively charged amino acids, the basic moiety by the polar and hydrophobic residues and negatively charged amino acid stabilizes the sugar (Table 2). In site II there is a variability in the positioning of the phosphate groups that is even larger for the base, as shown by the bidirectional arrows in figure 34 D, whose length is proportional to the width of the family conformations, because of the large degree of fluctuations of the amino acids forming binding site II (Fig. 36).

The presence in MD-docking of two interacting sites located in two distinct matrix areas may be related to the tripartite structure of the protein. Experimental studies have shown that the substrates transport is 1:1 and so the carrier channel can be crossed by only one molecule at a time (Nury et al., 2006). This is due to the carrier assembly that doesn't permit to more than one molecule to pass through the channel at the same time because of steric hindrance. Recent studies have highlighted the symmetrical and asymmetrical conservation of residues in the mitochondrial carrier family (Robinson et al., 2008), and have identified three symmetrical substrate contact points in the bottom of the channel and three toward the inter-membrane space, each one located on a single repeat. Our MD and Docking studies show the occurrence of two binding sites for the ATP^{4-} substrate, each one spreading in a region

involving two repeats, residues composing site I being located on repeats 1 and 2, and those composing site II being located on repeats 3 and 1 (Fig. 35). These residues are highly conserved in a large number of sequences, taken from different species that are evolutionary far each from the other (data not shown), supporting the idea of the involvement of these residues in the formation of a conserved ATP^{4-} binding site.

Extending the symmetry considerations on the tripartite sequence and three-dimensional structure of the repeats, a third binding site could be predicted. Instead, in our approach this third binding site could not be detected. Although not supported by experimental data it could also be speculated that the recognition sites here identified could function as regulators of protein activity, as observed in the transport activity of the mitochondrial carrier UCP1, that is regulated by purine nucleotide binding sites on the matrix side (Arechaga et al., 2008).

In conclusion the coupling of molecular dynamics and protein-ligand docking, allowed us to propose for the first time the residues of the bovine mitochondrial ADP/ATP carrier binding the ATP^{4-} substrate in the matrix region (Fig. 34). The molecular docking applied on the c-state X-ray structure of the transporter (Pebey-Peyroula et al., 2003) led us to identify a unique binding site preformed, even if the structure of the protein has been crystallized in a conformation not ready to interact with the substrate in the matrix region. This site is also detected by the docking performed on the structures extracted from a 20 ns MD simulation carried out on the free carrier, where a second binding site also appears, with arrangement and amino acids physical chemical properties similar to the first one (Table 2 and Fig. 36). These findings open the question if the multiple sites act independently or via a concerted mechanism.

– References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (1994). Molecular Biology of the Cell. New York: Garland Publishing Inc.

Alisaraie, L., Haller, L.A., and Fels, G. (2006). A QXP-Based Multistep Docking Procedure for Accurate Prediction of Protein−Ligand Complexes. J. Chem. Inf. Model. *46*, 1174–1187.

Allen, M.P., and Tildesley, D.J. (1987). Computer simulations of liquids. Oxford University Press, Oxford, New York.

Altschul, S.F., Carroll, R.J., and Lipman, D.J. (1989). Weights for data related by a tree. J, Mol. Biol. *207*, 647-653.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., et al. (1981). Sequence and organization of the human mitochondrial genome. Nature. *290*, 457-465.

Anézo, C., de Vries, A.H., Holtje, H.D., Tieleman, D.P., and Marrink, S.J. (2003). Methodological issues in lipid bilayer simulations. J. Phys. Chem. B *107*, 9424-9433.

Aquila, H., Misra, D., Eulitz, M., and Klingenberg, M. (1982). Complete amino acid sequence of the ADP/ATP carrier from beef heart mitochondria. Physiol Chem. *363*, 345-349.

Aquila, H., Link, T.A., and Klingenberg, M. (1985). The uncoupling protein from brown fat mitochondria is related to the mitochondrial ADP/ATP carrier. Analysis of sequence homologies and folding of the protein in the membrane. EMBO. J. *4*, 2369-2376.

Åqvist, J. (1990). Ion Water Interaction Potentials Derived from Free-Energy Perturbation Simulations. J. of Phys. Chem. *94*, 8021-8024.

Åqvist J. (1994). Comment on Transferability of Ion Models. J. of Phys. Chem. *98*, 8253-8255.
Arechaga, I., Ledesma, A., and Rial, E. (2001). The mitochondrial uncoupling protein UCP1: a gated pore. IUBMB Life. *52*, 165-173.

Ash, W.L., Zlomislic, M.R., Oloo, E.O., and Tieleman, D.P. (2004). Computer simulations of membrane proteins. Biochim. Biophys. Acta. *1666*, 158-189.

Auffinger, P., and Beveridge, D.L. (1995). A Simple Test for Evaluating the Truncation Effects in Simulations of Systems Involving Charged Groups. Chemical Physics Letters. *234*, 413-415.

Babu, C.S., and Lim, C. (2006). Empirical force fields for biologically active divalent metal cations in water. J. Phys. Chem. A. *110*, 691-699.

Bamber, L., Harding, M., Monné, M., Slotboom, D.J., and Kunji, E.R. (2007). The yeast mitochondrial ADP/ATP carrier functions as a monomer in mitochondrial membranes. Proc. Natl. Acad. Sci. U S A. *104*, 10830-10834.

Baumketner, A., and Shea, J.E. (2005). The influence of different treatments of electrostatic interactions on the thermodynamics of folding of peptides. J. Phys. Chem. B *109*, 21322-21328.

Beck, D.A., Armen, R.S., Daggett, V. (2005). Cutoff size need not strongly influence molecular dynamics results for solvated polypeptides. Biochemistry *44*, 609-616.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Di Nol,a A., and Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. J. Chem. Phys. *81*, 3684-3690.

Berendsen, H.J.C., Grigera, J.R., and Straatsma, T.P. (1987). The Missing Term in Effective Pair Potentials. Journal of Physical Chemistry *91*, 6269-6271.

Berendsen, H.J.C., van der Spoel, D., and van Drunen, R. (1995). Gromacs – a Message-Passing Parallel Molecular-Dynamics Implementation. Comput. Phys. Commun. *91*, 43-56.

Berger, O., Edholm, O., and Jähnig, F. (1997). Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. Biophys. J. *72*, 2002-2013.

Bisaccia, F., Zar,a V., Capobianco, L., Iacobazzi, V., Mazzeo, M., and Palmieri, F. (1996)a. The formation of a disulfide crosslink between the two subunits demonstrates the dimeric structure of the mitochondrial oxoglutarate carrier. Biochim. Biophys. Acta *1292*, 281-288.

Bisaccia, F., Capobianco, L., Mazzeo, M., and Palmieri, F. (1996)b. The mitochondrial oxoglutarate carrier protein contains a disulfide bridge between intramembranous cysteines 221 and 224. FEBS. Lett. *392*, 54-58.

Bond, P.J., and Sansom, M.S. (2006). Insertion and assembly of membrane proteins via simulation. J. Am. Chem. Soc. *128*, 2697-2704.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. (1983). Charmm – a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J. Comput. Chem. *4*, 187-217.

Capobianco, L., Brandolin, G., and Palmieri, F. (1991). Transmembrane topography of the mitochondrial phosphate carrier explored by peptide-specific antibodies and enzymatic digestion. Biochemistry. *30*, 4963-4969.

Capobianco, L., Bisaccia, F., Michel, A., Sluse, F.E, and Palmieri, F. (1995). The N- and C-termini of the tricarboxylate carrier are exposed to the cytoplasmic side of the inner mitochondrial membrane. FEBS. Lett. *357*, 297-300.

Cheatham, T.E., Miller, J.L., Fox ,T., Darden, T.A., and Kollman P.A. (1995). Molecular- Dynamics Simulations on Solvated Biomolecular Systems - the Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. J. Am. Chem. Soc. *117*, 4193-4194.

Chillemi, G., Redinbo, M., Bruselles, A., and Desideri, A. (2004). Role of the linker domain and the 203-214 N-terminal residues in the human topoisomerase I DNA complex dynamics. Biophys. J. *87*, 4087-4097.

Chipuk, J.E., Bouchier-Hayes, L., Green, D.R. (2006). Mitochondrial outer membrane permeabilization during apoptosis: the innocent bystander scenario. Cell Death and Diff. *13*, 1396–1402.

Clark, K.P., and Ajay. (1995). Flexible ligand docking without parameter adjustment across four ligand-receptor complexes J. Comput. Chem. *16*, 1210-1226.

Congreve, M., Chessari, G., Dominic Tisi, D., and Woodhead, A.J. (2008). Recent Developments in Fragment-Based Drug Discovery. J. Med. Chem. *51*, 3661–3680.

Dahout-Gonzalez, C., Ramus, C., Dassa, E.P., Dianoux, A.C., and Brandolin, G. (2005). Conformation-dependent swinging of the matrix loop m2 of the mitochondrial *Saccaromyces cerevisiae* ADP/ATP carrier. Biochemistry. *44*, 16310-16320

Darden, T., York, D., and Pedersen, L. (1993). Particle Mesh Ewald – an N.Log(N) Method for Ewald Sums in Large Systems. J. Chem. Phys. *98*, 10089-10092.

Daura, X., van Gunsteren, W.F., and Mark, A.E. (1999). Folding-unfolding thermodynamics of a beta-heptapeptide from equilibrium simulations. Proteins. *34*, 269-280.

Dehez, F., Pebay-Peyroula, E., and Chipot, C. (2008). Binding of ADP in the mitochondrial ADP/ATP carrier is driven by an electrostatic funnel. J. Am. Chem. Soc. *130*, 12725-12733.

Deupi, X., Olivella, M., Govaerts, C., Ballesteros, J.A., Campillo, M., and Pardo, L. (2004). Ser and Thr residues modulate the conformation of pro-kinked transmembrane alpha-helices. Biophys. J. *86*, 105-115.

De Marcos Lousa, C., Trézéguet, V., Dianoux, A.C., Brandolin, G., and Lauquin, G.J. The human mitochondrial ADP/ATP carriers: kinetic properties and biogenesis of wild-type and mutant proteins in the yeast *S. cerevisiae*. (2002). Biochemistry. *41*, 14412-14420.

De Marcos Lousa, C., Trézéguet, V., David, C., Postis, V., Arnou, B., Pebay-Peyroula, E., Brandolin, G., and Lauquin, G.J. (2005). Valine 181 Is Critical for the Nucleotide Exchange Activity of Human Mitochondrial ADP/ATP Carriers in Yeast. Biochemistry. *44*, 4342-4348.

Duyckaerts, C., Sluse-Goffart, C.M., Fux, J., Sluse, F.E., and Liebecq, C. (1980). Kinetic mechanism of the exchanges catalysed by the adenine-nucleotide carrier. Eur. J. Biochem. *106,* 1-6.

Edholm, O., Berger, O., and Jähnig, F. (1995). Structure and Fluctuations of Bacteriorhodopsin in the Purple Membrane - a Molecular- Dynamics Study. J. Mol. Biol. *250*, 94-111.

Falconi, M., Gallimbeni, R., and Paci, E. (1996). Dimer asymmetry in superoxide dismutase studied by molecular dynamics simulation. J. Comput. Aided. Mol. Des. *10*, 490-498.

Falconi, M., Chillemi, G., Di Marino, D., D'Annessa, I., Morozzo della Rocca, B., Palmieri, L., and Desideri, A. (2006). Structural Dynamics of the Mitochondrial ADP/ATP Carrier Revealed by Molecular Dynamis Simulation Studies. Proteins. *65*, 681-691.

Fanelli, F., and De Benedetti, P.G. (2005). Computational modelling approaches to structure-function analysis of G protein-coupled receptors. Chem. Rev. *105*, 3297-351.

Feller, S.E., Pastor, R.W., Rojnuckarin, A., Bogusz, S., and Brooks, B.R. (1996). Effect of electrostatic force truncation on interfacial and transport properties of water. J. of Phys. Chem. *100*, 17011-17020.
Feller, S.E., Gawrisch, K., MacKerell, A.D. (2002). Polyunsaturated fatty acids in lipid bilayers: intrinsic and environmental contributions to their unique physical properties. J. Am. Chem. Soc. *124*, 318-326.

Ferriera, G.C., Pratt, R.D., and Pedersen, P.L. (1990). Mitochondrial proton/phosphate transporter. An antibody directed against the COOH terminus and proteolytic cleavage experiments provides new insights about its membrane topology. J. Biol. Chem. *265*, 21202-21206.

Fiermonte, G., Runswik, M.K., Walzer, J.E., and Palmieri, F. (1993). Abundant bacterial expression and reconstitution of an intrinsic membrane transport protein from bovine mitochondria. Biochem. J. *294*, 293-299.

Fiore, C., Arlot-Guilligay, D., Tre´ze´guet, V., Lauquin, G.J., and Brandolin, G. (2001). Fluorometric detection of ADP/ATP carrier deficiency in human muscle. Clin. Chim. Acta. *311*, 125–135.

Fontanesi, F., Palmieri, L., Scarcia, P., Lodi, T., Donnini, C., Limongelli, A., Tiranti, V., Zeviani, M., Ferrero, I., and Viola, A.M. (2004). Mutations in AAC2, equivalent to human adPEO-associated ANT1 mutations, lead to defective oxidative phosphorylation in Saccharomyces cerevisiae and affect mitochondrial DNA stability**.** Hum. Mol. Genet. *13*, 923-934.

Garcia, A.E. (1992). Large-amplitude nonlinear motions in proteins. Phys. Rev. Lett. *68*, 2696–2699.

Gardner, A., and Boles, R.G. (2005). Is a Mitochondrial Psychiatry in the Future - A Review. Curr. Psychiatry Review *1*, 255–271.

Gilson, M.K. (1995). Theory of electrostatic interactions in macromolecules. Curr. Opin. Struct. Biol. *5*, 216-223.

Gascuel, O., and Steel, M. (2006). Neighbor-joining revealed. Mol. Biol. Evol. *23*, 1997-2000.

Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis. *18*, 2714-2723.

Guillot, B. (2002). A reappraisal of what we have learnt during three decades of computer simulations on water. Journal of Molecular Liquids. *101*, 219-260.

Gurtovenko, A.A., Patra, M., Karttunen, M., and Vattulainen, I. (2004). Cationic DMPC/DMTAP lipid bilayers: Molecular dynamics study. Biophys. J. *86*, 3461-3472.

Gurtovenko, A.A. (2005). Asymmetry of lipid bilayers induced by monovalent salt: Atomistic molecular-dynamics study. J. Chem. Phys. *122*,

Hatanaka, T., Hashimoto, M., Majima, E., Shinohara, Y., and Terada, H. (1999). Functional expression of the tandem-repeated homodimer of the mitochondrial ADP/ATP carrier in *Saccharomyces cerevisiae*. Biochem. Biophys. Res. Commun. *273*, 14269-14276.

Hayashi, T., Rizzuto, R., Hajnoczky, G., and Su, T.P. (2009). MAM: more than just a housekeeper. Trends. Cell Biol. *19*, 81–8.

Henderson, P.J., and Lardy, H.A.J. (1970). Bongkrekic acid. An inhibitor of the adenine nucleotide translocase of mitochondria. Biol. Chem. *245*, 1319–1326.

Hess, B., Bekker, H., Berendsen, H.J.C., and Fraaije, J.G.E.M. (1997). LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. *18*, 1463-1472.

Hess, B. (2002). Determining the shear viscosity of model liquids from molecular dynamics simulations. J. Chem. Phys. *116*, 209-217.

Heinz, T.N., and Hunenberger, P.H. (2005). Combining the lattice-sum and reaction-field approaches for evaluating long-range electrostatic interactions in molecular simulations. J. Chem.Phys. *123*, 34107.

Herrmann, J.M., and Neupert, W. (2000). Protein transport into mitochondria. Curr. Opin. Microbiol. *3,* 210–214.

Higgins, D.G., and Sharp, P.M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene. *73*, 237–244.

Huang, X., and Miller, W. (1991). A time-efficient, linear-space local similarity algorithm. Adv. Appl. Math. 12, 337-357.
Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD - Visual Molecular Dynamics. J. Mol. Graph. *14*, 33-38.

Im, W., Feig, M., and Brooks, C.L. (2003). An implicit membrane generalized born theory for the study of structure, stability, and interactions of membrane proteins. Biophys. J. *85*, 2900-2918.

Indiveri, C., Iacobazzi, V., Giangregorio, N., and Palmieri, F. (1997). The mitochondrial carnitine carrier protein: cDNA cloning, primary structure, and comparison with other mitochondrial transport proteins. Biochem. J. *321*, 713-719.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. J. Chem. Phys. *79*, 926-935.

Jorgensen, W.L., Maxwell, D.S., and TiradoRives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc. *118*, 11225-11236.

Jorgensen, W.L., and Tirado-Rives, J. (2005). Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. Proc. Nat. Acad. Sci. U S A. *102*, 6665-6670.

Kandasamy, S.K., and Larson, R.G. (2006). Molecular dynamics simulations of model trans-membrane peptides in lipid bilayers: A systematic investigation of hydrophobic mismatch. Biophys. J. *90*, 2326-2343.

Kaplan, R.S., Major, J.A., and Wood, D.O. (1993). The mitochondrial tricarboxylate transport protein: cDNA cloning, primary structure, and comparison with other mitochondrial transport proteins. J. Biol. Chem. *268*, 13682-13690.

Kaplan, R.S. (2001). Structure and function of mitochondrial anion transport proteins. J. Memb. Biol. *179*, 165-183.

Katchalski-Katzir, E., Shariv. I., Eisenstein, M., Friesem, A.A., Aflalo, C., and Vakser, I.A. (1992). Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. Proc. Natl. Acad. Sci. U S A. *89*, 2195-2199.

Kaukonen, J., Jukka, K.J., Tiranti, V., Kyttala, A., Zeviani, M., Comi, G.P., Keranen, S., Peltonen, L., and Suomalainen, A. (2000). Role of Adenine Nucleotide Translocator 1 in mtDNA Maintenance. Science. *289*, 782-785.

Kitchen, D.B., Decornez, H., Furr, J.R., and Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. Nature reviews. Drug discovery**.** *3*, 935–949.

Klauda, J.B., Wu, X., Pastor, R.W., and Brooks, B.R. (2007). Long-range Lennard-Jones and electrostatic interactions in interfaces: application of the isotropic periodic sum method. J Phys Chem B *111*, 4393-4400.

Klingenberg, M., and Appel, M. (1989). The uncoupling protein dimer can form disulfide cross-link between the mobile C-terminal SH groups. Eur. J. Biochem. *180*, 123-131.

Klingenberg, M. (2005). Ligand-protein interaction in biomembrane carriers. The induced transition fit of transport catalysis. Biochemistry. *44*, 8563-8570.
Klingenberg, M. (2008). The ADP and ATP transport in mitochondria and its carrier. Biochim. Biophys. Acta. *1778*, 1978-2021.

Kotaria, R., Mayor, J.A., Walters, D.E., and Kaplan, R.S. (1999). Oligomeric state of wild-type and cysteine-less yeast mitochondrial citrate transport proteins. J. Bioenerg. Biomembr. *31*, 543-549.

Kramer, B., Metz, G., Rarey, M., and Lengauer, T. (1999). Ligand docking and screening with FlexX. Med. Chem. Res. *9*, 463-478.

Kunji, R., Harding, M. (2003). Projection structure of the atractyloside-inhibited mitochondrial ADP/ATP carrier of *Saccaromyces cerevisiae*. J. Biol. Chem. *278*, 36985-36988.

Lague, P., Pastor, R.W., and Brooks, B.R. (2004). Pressure-based long-range correction for Lennard-Jones interactions in molecular dynamics simulations: Application to alkanes and interfaces. J. of Phys. Chem. B. *108*, 363-368.

Leekumjorn, S., and Sum, A.K. 2006. Molecular simulation study of structural and dynamic properties of mixed DPPC/DPPE bilayers. Biophys. J. *90*, 3951-3965.

Lengauer, T., and Rarey, M. (1996). Computational methods for biomolecular docking. Curr. Opin. Struct. Biol. *6*, 402–406.

Lennard-Jones, J.E. (1931). Cohesion. Proceedings of the Physical Society *43*, 461-482.

Lesnefsky, E.J., et al. (2001). Mitochondrial dysfuntion in cardiac disease ischemia-reperfusion, aging and heart failure. J. Mol. Cell. Cardiol. *33*, 1065–1089.

Lindahl, E., Hess, B., and van der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. J. Mol. Model. *7*, 306-317.

Lipman, D.J. (1985). Rapid and sensitive protein similarity searches. Science. *227*, 1435–144.
Lipman, D.J., Altschul, S.F., and Kececioglu, J.D. (1989). A tool for multiple sequence alignment. Proc. Natl. Acad. Sci. U S A. *86*, 4412–4415.

Liu, M., and Wang, S. (1999). MCDOCK: A Monte Carlo simulation approach to the molecular docking problem. Journal of Computer-Aided Molecular Design. *13*, 435-451.

Loncharich, R.J., and Brooks, B.R. (1989). The effects of truncating long-range forces on protein dynamics. Proteins. *6*, 32-45.

Loeffler, H.H., and Kitao, A. (2009). Collective dynamics of periplasmic glutamine binding protein upon domain closure. Biophys. J. *97*, 2541-2549.

Louise-May, S., Auffinger, P., and Westhof, E. (1996). Calculations of nucleic acid conformations. Curr. Opin. Struct. Biol. *6*, 289-298.

MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B. *102*, 3586-3616.

Mackerell, A.D., Feig, M., and Brooks, C.L. (2004). Extending the treatment of backbone energetic in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J. Comput. Chem. *25*, 1400-1415.

Majima, E., Shinohara, Y., Yamaguchi, N., Hong, Y.M., and Terada, H. (1994). Importance of loops of mitochondrial ADP/ATP carrier for its transport activity deduced from reactivities of its cysteine residues with the sulfhydryl reagent eosin-5-maleimide. Biochemistry. *33*, 9530-9536.

Majima, E., Yamaguchi, N., Chuman, H., Shinohara, Y., Ishida, M., Goto, S., and Terada, H. (1998). Binding of the fluorescein derivative eosin Y to the mitochondrial ADP/ATP carrier: characterization of the adenine nucleotide binding site. Biochemistry. *37*, 424-432.

Mannella, C.A. (2006). Structure and dynamics of the mitochondrial inner membrane cristae. *1763*, 542–548.

Marrink, S.J., Risselada, J., and Mark, A.E. (2005). Simulation of gel phase formation and melting in lipid bilayers using a coarse grained model. Chem. Phys. Lipids. *135*, 223-44.
Marrink, S.J., Risselada, H.J., Yefimov, S., Tieleman, D.P., and de Vries, A.H. (2007). The MARTINI force field: coarse grained model for biomolecular simulations. J. Phys. Chem. B.
111, 7812-7824.

Martin, Y. C. (1997). Challenges and prospects for computational aids to molecular diversity. Perspect. Drug Discovery Des. *7/8*, 159-172.

Martinez-Mayorga, K., Pitman, M.C., Grossfield, A., Feller, S.E., and Brown, M.F. (2006). Retinal counterion switch mechanism in vision evaluated by molecular simulations. J. Am. Chem. Soc. *128*, 16502-16503.

McBride, H.M., Neuspiel, M., and Wasiak, S. (2006). Mitochondria: more than just a powerhouse. Curr. Biol. *16*, 551-560.

McMillin, J.B., and Dowhan, W. (2002). Cardiolipin and apoptosis. Biochim. Biophys. Acta. *1585*, 97–107.

Metropolis, N., and Ulam, S. (1949). The Monte Carlo Method. Journal of the American Statistical Association**.** *44*, 335–341.

Metropolis, N., Rosembluth, A., Rosembluth, M., and Teller, A. (1953). Equation os state calculations by fast computing machines. J. Chem. Phys**.** *21*, 1087-1092.

Miller, M.D., Kearsley, S.K., Underwood, D.J., and Sheridan, R.E. (1994). FLOG: A system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. Journal of Computer-Aided Molecular Design. *8*, 153-174.

Miroux, B., Casteilla, L., Klaus, S., Raimbault, S., Gradin, S., Clement, J.M., Ricquier, D., and Bouillaud, F. (1992). Antibodies selected from whole antiserum by fusion proteins as tools for the study of the topology of mitochondrial membrane proteins. Evidence that the N-terminal extremity of the sixth α-helix of the uncoupling protein is facing the matrix. J. Biol. Chem. *267*, 13603-13609.

Miroux, B., Frossard, V., Raimbault, S., Ricquier, D., and Bouillaud, F. (1993). The topology of the brown adipose tissue mitochondrial uncoupling protein determined with antibodies against its antigenic sites revealed by a library of fusion proteins. EMBO. J. *12*, 3739-3745.

Mount, D.M. (2004). Bioinformatics: Sequence and Genome Analysis (2nd ed.). Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., Olson, A.J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem. *19*, 1639-1662.

Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., and Olson, A.J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J. Comput. Chem. *30*, 2785-2791.

Muller, M., Katsov, K., and Schick, M. (2006). Biological and synthetic membranes: What can be learned from a coarse-grained description? Physics Reports- Review Section of Physics Letters. *434*, 113-176.

Murray, C.W., Baxter, C.A., and Frenkel, A.D. (1999). The sensitivity of the results of molecular docking to induced fit effects: Application to thrombin, thermolysin and neuraminidase. Journal of Computer-Aided Molecular Design. *13*, 547–562.

Nagle, J.F., and Tristram-Nagle, S. (2000). Structure of lipid bilayers. Biochim. Biophys. Acta. *1469*, 159-195.

Napoli, L., Bordoni, A., Zeviani, M., Hadjigeorgiou, G.M., Sciacco, M., Tiranti, V., Terentiou, A., Moggio, M., Papadimitriou, A., Scarlato, G., and Comi, G.P. (2001). A novel missense adenine nucleotide translocator-1 gene mutation in a Greek adPEO family. Neurology. *57*, 2295–2298.

Needleman, S.B., and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. *48*, 443–453.

Nelson, D.R., Felix, C.M., and Swanson, J.M. (1998). Highly conserved charge-pair networks in the mitochondrial carrier family. J. Mol. Biol. *277*, 285-308.

Notredame, C., and Higgins, D.G. (1996). SAGA: sequence alignment by genetic algorithm. Nucleic Acids Res. *24*, 1515-1524.

Notredame, C., O'Brien, E.A., and Higgins, D.G. (1997). RAGA: RNA sequence alignment by genetic algorithm. Nucleic Acids Res. *25*, 4570-4580.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 302, 205–217.

Nury, H., Dahout-Gonzalez, C., Trézéguet, V., Lauquin, G.J., Brandoli,n G., and Pebay-Peyroula, E. (2006). Relations Between Structure and Function of the Mitochondrial ADP/ATP Carrier. Annu. Rev. Biochem. *75*, 713–741.

Nury, H., Manon, F., Arnou, B., le Maire, M., Pebay-Peyroula, E., and Ebel, C. (2008). Mitochondrial bovine ADP/ATP carrier in detergent is

predominantly monomeric but also forms multimeric species. Biochemistry. *47*, 12319-12331.

Palmieri, F., Bisaccia, F., Capobianco, L., Dolce, V., Fiermonte, G., Iacobazzi, V., and Zara, V. (1993). Transmembrane topology, genes and biogenesis of the mitochondrial phosphate and oxoglutarate carriers. J. Bioenerg. Biomembr. *25*, 493-501.

Palmieri, F., Bisaccia, F., Capobianco, L., Dolce, V., Fiermonte, G., Iacobazzi, V., Indiveri, C., and Palmieri, L. (1996). Mitochondrial metabolite transporters. Biochim. Biophys. Acta. *1275*, 127-132.

Palmieri, F. (2004). The mitochondrial transporter family (SLC25): physiological and pathological implications. Eur. J. Physiol. *447*, 689-709.

Palmieri, F. (2008). Diseases caused by defects of mitochondrial carriers: A review. Biochim. Biophys. Acta. *1777*, 564-578.

Pandit, S.A., and Berkowitz, M.L. (2002). Molecular dynamics simulation of dipalmitoylphosphatidylserine bilayer with Na+ counterions. Biophys. J. *82*, 1818-1827.

Parrinello, M., and Rahman, A. (1981). Polymorphic Transitions in Single-Crystals - a New Molecular- Dynamics Method. J. Appl. Phys. *52*, 7182-7190.

Patra, M., Karttunen, M., Hyvönen, M.T., Falck, E., Lindqvist, P., and Vattulainen, I. (2003). Molecular dynamics simulations of lipid bilayers: major artifacts due to truncating
electrostatic interactions. Biophys. J. *84*, 3636-3645.

Patra, M., and Karttunen, M. (2004). Systematic comparison of force fields for microscopic simulations of NaCl in aqueous solutions: diffusion, free energy of hydration, and structural properties. J. Comput. Chem. *25*, 678-689.

Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham, T.E., Debolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995). Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to

Simulate the Structural and Energetic Properties of Molecules. Comput. Phys. Commun. *91*, 1-41.

Pebay-Peyroula, E., Dahout-Gonzalez, C., Kahn, R., Trézéguet, V., Lauquin, G.J.M., and Brandolin, G. (2003). Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. Nature. *426*, 39–44.

Periole, X., Huber, T., Marrink, S.J., and Sakmar, T.P. (2007). G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. J. Am. Chem. Soc. *129*, 10126-10132.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera - A Visualization System for Exploratory Research and Analysis. J. Comput. Chem. *25*, 1605-1612.

Pitman, M.C., Suits, F., Mackerell, A.D., and Feller, S.E. (2004). Molecular-level organization of saturated and polyunsaturated fatty acids in a phosphatidylcholine bilayer containing cholesterol. Biochemistry *43*, 15318-15328.

Pitman, M.C., Grossfield, A., Suits, F., and Feller, S.E. (2005). Role of cholesterol and polyunsaturated chains in lipidprotein interactions: molecular dynamics simulation of rhodopsin in a realistic membrane environment. J. Am. Chem. Soc. *127*, 4576-4577.

Robinson, A.J., and Kunji, E.R. (2006). Mitochondrial carriers in the cytoplasmic state have a common substrate binding site. Proc. Natl. Acad. Sci. U S A. *103*, 2617-2622.

Robinson, A.J., Overy, C., and Kunji, E.R. (2008). The mechanism of transport by mitochondrial carriers based on analysis of symmetry. Proc. Natl. Acad. Sci. U S A. *105*, 17766-17771.

Roux, B., and Karplus, M. (1994). Molecular dynamics simulations of the gramicidin channel. Annu. Rev. Biophys. Biomol. Struct. *23*, 731-761.

Runswick, M.J., Powell, S.J., Nyren, P., and Walker, J.E. (1987). Sequence of the bovine mitochondrial phosphate carrier protein: structural relationship to ADP/ATP translocase and the brown fat mitochondria uncoupling protein. EMBO. J. *6*, 1367-1373.

Runswick, M.J., Walker, J.E., Bisaccia, F., Iacobazzi, V., and Palmieri, F. (1990). Sequence of the bovine 2-oxoglutarate/malate carrier protein: structural relationship to other mitochondrial transport proteins. Biochemistry. *29*, 11033-11040.

Sachs, J.N., Nanda, H., Petrache, H.I., Woolf, T.B. (2004). Changes in phosphatidylcholine headgroup tilt and water order induced by monovalent salts: Molecular dynamics simulations. Biophys. J. *86*, 3772-3782.

Sadowski, J., and Boström, J. (2006). MIMUMBA revisited : Torsion angle rules for conformer generation derived from X-ray structures. Journal of Chemical Information and Modeling. *46*, 2305-2309.

Saiz, L., and Klein, M.L. (2002). Computer simulation studies of model biological membranes. Accounts of Chemical Research *35*, 482- 489.
Saraste, M., and Walker, J.E. (1982). Internal sequence repeats and the path of polypeptide in mitochondrial ADP/ATP translocase. FEBS. Lett. *144*, 250-254.

Shih, A.Y., Arkhipov, A., Freddolino, P.L., and Schulten, K. (2006). Coarse grained protein-lipid model with application to lipoprotein particles. J. Phys. Chem. B. *110*, 3674-3684.

Sperotto, M.M., May, S., and Baumgaertner, A. (2006). Modelling of proteins in membranes. Chem. Phys. Lipids. *141*, 2-29.

Strahs, D., and Weinstein, H. (1997). Comparative modeling and molecular dynamics studies of the delta, kappa and mu opioid receptors. Protein. Eng. *10*, 1019-1038.

Taylor, J.S., and Burnett, R.M. (2000). DARWIN: a program for docking flexible molecules. Proteins. *41*, 173-191.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic. Acids. Res. 22, 4673–4680.

Tieleman, D.P., Sansom, M.S., Berendsen, H.J. (1999). Alamethicin helices in a bilayer and in solution: molecular dynamics simulations. Biophys. J. *76*, 40-49.

Tieleman, D.P., MacCallum, J.L., Ash, W.L., Kandt, C., Xu, Z.T., and Monticelli, L. (2006). Membrane protein simulations with a united-atom lipid and all-atom protein model: lipid-protein interactions, side chain transfer free energies and model proteins. J. Phys-Condens. Mat. *18*, S1221-S1234.

Tobias, D.J. (2001). Electrostatics calculations: recent methodological advances and applications to membranes. Curr. Opin. Struct. Biol. *11*, 253-261.

Trézéguet, V., Le Saux, A., David, C., Gourdet, C., Fiore, C., Dianoux, A.C., Brandolin, G., Lauquin, G.J.M. (2000). A covalent tandem dimer of the mitochondrial ADP/ATP carrier is functional in vivo. Biochim. Biophys. Acta. *1457*, 81-93.

Trosset, J.Y., and Scheraga, H.A. (1999). PRODOCK: Software Package for Protein Modeling and Docking. J. Comput. Chem. *20*, 412-427.

Van Gunsteren, W.F., and Berendsen, H.J.C. (1977). Algorithms for Macromolecular Dynamics and Constraint Dynamics. Molecular Physics. *34*, 1311-1327.

Van Gunsteren, W.F., and Berendsen, H.J.C. (1990). Computer-Simulation of Molecular-Dynamics - Methodology, Applications, and Perspectives in Chemistry. Angewandte Chemie-International Edition in English. *29*, 992-1023.

Venable, R.M., Brooks, B.R., and Pastor, R.W. (2000). Molecular dynamics simulations of gel (L-beta I) phase lipid bilayers in constant pressure and constant surface area ensembles. J. Chem. Phys. *112*, 4822-4832.

Venturoli, M., Smit, B., and Sperotto, M.M. (2005). Simulation studies of proteininduced bilayer deformations, and lipid-induced protein tilting, on a mesoscopic model for lipid bilayers with embedded proteins. Biophys. J. *88*, 1778-1798.

Vieth, M., Hirst, J.D., Kolinski, A., and Brooks, C.L.I. (1998). Assessing energy functions for flexible docking. J. Comp. Chem. *19*, 1612-1622.

Vignais, P.V., Vignais, P.M., and Stanislas, E. (1962). Action of potassium atractylate on oxidative phosphorylation in mitochondria and in submitochondrial particles. Biochim. Biophys. Acta. *60*, 284–300.

Wallqvist, A., and Mountain, R.D. (1999). Molecular models of water: Derivation and description. Reviews in Computational Chemistry. *13*, 183-247.

Wang, L., and Jiang, T. (1994). On the complexity of multiple sequence alignment. J. Comput. Biol. *1*, 337–348.

Wang, Y., and Tajkhorshid, E. (2008). Electrostatic funneling of substrate in mitochondrial inner membrane carriers. Proc. Natl. Acad. Sci. U S A. *105*, 9598-9603.

Wang X., Salinas K., Kucejova B., Chen X.J. (2008). Dominant membrane uncopling by mutant adenine nucleotide translocase in mitochondrial diseases. Hum. Mol. Genet. 24, 4036-4044.

Wohlert, J., and Edholm, O. (2004). The range and shielding of dipole-dipole interactions in phospholipid bilayers. Biophys. J. *87*, 2433-2445.

Yoshimori, A. Carlos, A., and Del Carpio, M. (2001). A Novel Protein Soft Docking Algorithm Based on Spectral Analysis of Shape Complementarity. Genome Informatics *12,* 354–355.

Zhang, Z., Pfaendtner, J., Grafmüller, A., and Voth, G.A. (2009). Defining coarse-grained representations of large biomolecules and biomolecular complexes from elastic network models. Biophys. J. *97*, 2327-2337.