Hydrogen bonding and conformational coupling of the SecA protein motor

A Dissertation

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Physics

to the Department of Physics of Freie Universität Berlin

by Stefan Milenkovic

Berlin, 2017

Supervisor: Prof. Dr. Ana-Nicoleta Bondar

Second examiner: Prof. Dr. Joachim Heberle

Date of defense: 07.11.2017

Abstract

SecA ATP-ase is the essential part of the bacterial secretion machinery, which employs the energy of the ATP binding and hydrolysis to push newly synthesized proteins (preproteins) across the cytoplasmic membrane. Due to its importance, SecA has been extensively studied by various experimental techniques. Still, the exact mechanism of SecA's chemo-mechanical coupling that leads to a successful completion of the translocation process remains unrevealed. The underlying reason for this could be a very complex nature of the translocation process, which involves many factors that are mutually coupled . X-ray studies, for instance, can capture SecA in a certain conformation and NMR could assign dynamical properties to the SecA's structural elements. The analysis of SecA's structures obtained in this way, can give predictions of possibly important amino acid residues for SecA function. Later on, these predictions can be confirmed or denied by mutagenesis studies. However, SecA has several hundreds of amino acid residues and a systematic mutagenesis check could demand a significant amount of work and time. Comined together, aforementioned methods provided valuable insights of the SecA function, but failed to answer the question of the mechanism behind aforementioned chemo-mechanical coupling. A promising approach to this problem could be molecular dynamics (MD) simulations. The MD simulations are capable of animating the static SecA structures obtained by the X-ray crystallography and providing information of SecA's dynamics that could be used for better understanding and explanation of the SecA's function. Further more, the MD simulations give per-atom description of the processes that occur in SecA thus, allowing an easy manipulation of the structure (such are ligand removal or binding, mutations etc.) which could cover many questions that remained open after the experimental studies. Another reason for employing the MD simulations is that the computational studies of the SecA were rarely performed. Moreover, this thesis

brought the first publication about the SecA that was solely based on computations or, to be more precise, on MD simulation to our knowledge.

MD simulations were the key asset in this thesis and they were performed on the SecA originating from three different organisms. They were used to probe different perturbations such as mutations, nucleotide removal or signal peptide binding. The trajectories obtained by the MD simulations have been a subject of a robust analysis which yielded several very important results.

A thorough analysis of the nucleotide binding pocket in the ADP-bound *B. subtilis* and *T. maritima* SecA led to the discovery of the water bridging of the amino acids D207 and E208, which are parts of the evolutionary conserved DEAD motif (*T.maritima* D252/E253). This observation was experimentally confirmed in the high-resolution crystal structure of *T. maritima* that was published in 2015 completely independent from this research. Additionally, the dependence of water behaviour in the nucleotide binding pocket on the nucleotide binding state suggested the importance of the H-bonding networking for SecA's function. The observed complexity of the H-bonding networking, initially localized in the nucleotide binding pocket, promptly spread out to the rest of the SecA showing how extensive the H-bonding networks in a protein could be. In order to get an answer how important these networks are, a fast and efficient way of characterizing the H-bonding networks, named H-bonding map, was developed. Indeed, H-bonding maps dissected complex hydrogen bonding dynamics in SecA marking the SecA's helical scaffold domain (HSD) as a central figure in already mentioned H-bonding networks. This observation led to the hypothesis that H-bonding networks play a crucial role in allosteric changes in SecA.

Another important aspect in this thesis was the interaction of the SecA with other elements of the translocation machinery. Therefore, a role of the ligand presence on SecA's dynamics was extensively studied with the special focus on the ADP and the signal peptide presence. The reasons for such peculiar interests are condensed in following: *i)* ATP binding/ADP release induce conformational changes in the SecA and promote the preprotein translocation and *ii)* binding of the preprotein behaves as an

allosteric activator of the SecA activity and prevents the uncontrolled ATP hydrolysis. To uncover the possible allosteric effects of the preprotein on SecA's dynamics complex of the SecA and the preprotein's signal sequence (SP) was simulated.

The principal component analysis (PCA) was chosen to be a tool for the dissection of important modes of motion and couplings established between the different functional domains of the SecA. The PCA findings were confirmed by the independent calculations of the linear mutual information (LMI), a measure of the correlation between two different atoms. These investigations indicated that a long-distance coupling is established between the SecA's domains in presence of the SP. Furthermore, the regions of highly correlated residues from the different domains correspond to the results of the H-bonding maps, confirming that H-bonding networks possibly serve as the allosteric transducers.

In the summary, the results obtained in this thesis led to the conclusion that instead of one specific amino acid residue serving as the catalytic base proposed by the previous experimental work, the water-bridge between two DEAD amino acid residues, D207 and E208, exists. Moreover, it was noticed that the water-bridge presence is nucleotide dependent and highly sensitive to D207N mutation, which is proven to abolish catalytic activity of SecA. H-bonding map, a novel method for visualization of the H-bonding dynamics showed complex interconnections between functional domains of SecA achieved via the H-bonds where helical scaffold domain was identified as the central part of such interactions. Finally, research was expanded to the complex relation between the SecA and SP, a short tripartite sequence which is found in the N-terminus of every preprotein and plays very important, but not fully understood role, in translocation process. The results of the analysis indicated the establishment of the long-distance coupling between different domains of SecA which seems to be achieved via already identified networks of H-bonds.

Since the ATP hydrolysis is one of the most common energy sources in biological systems and signal peptides are universal segment efficiently recognized by many different proteins, these results give a possible guidelines for analysis of similar systems and problems.

First Thesis Supervisor: Prof.Dr. Ana-Nicoleta Bondar

Second Thesis Supervisor: Prof.Dr. Joachim Heberle

Zusammenfassung

SecA ATPasen ist der essentielle Teil der bakteriellen Sektretionsmaschinerie, die die Energie der ATP-Bindung und Hydrolyse nutzt, um frisch synthetisierte Proteine aus Zyklomembran herauszutransportieren. Aufgrund seiner Bedeutung wurde SecA aufs intensivste mit unterschiedliche experimentellen Methoden untersucht. Dennoch konnte bislang der exakte Mechanismus des Translokationsprozesses nicht verstanden werden. Die Translokation beinhaltet mehrere aneinander gekoppelte Faktoren und die Komplexität des Prozesses könnte die Ursache für den Mangel einer Beschreibung sein. Molekulardynamik-Simulationen (MD-Simulationen) ermöglichen es, die statischen Röntgenkristallstrukturen von SecA zu animieren und Informationen über die Dynamik von SecA zu liefern, und scheinen ein guter Ansatz zur Lösung des Problems zu sein. Ein weiterer Grund für die Nutzung von MD-Simulationen ist, dass nur selten Computerstudien von SecA durchgeführt wurden. Dieser Promotion entsprang eine Erstpublikation über SecA, die komplett auf MD-Simulationen basierte. Diese MD-Simulationen bilden die Basis für diese Dissertation und wurden für SecA aus drei verschiedenen Organismen durchgeführt. Es wurden verschiedene Einflüsse, wie z.B. Mutationen, Entfernung von Nukleotiden oder Signalpeptid-Bindung, getestet. Die Trajektorien, die aus MD-Simulationen stammten, wurden einer gründlichen Analyse ausgesetzt, welche mehrere wichtige Resultate zur Folge hatte.

Eine sorgfältige Analyse der Nukleotidbindungsregion in B. subtilis und T. maritima mit gebundenem ADP führte zur Entdeckung, dass die Aminosäuren D207 und E208, beide Teil des stark konservierten DEAD-Motivs, durch Wassermoleküle miteinander verbunden sind. Unabhängig von der hier präsentierten Forschungsarbeit wurde diese Beobachtung im Jahr 2015 mit der Publikation der hoch auflösenden Kristallstruktur von T. maritima experimentell bestätigt. Ferner legte die Abhängigkeit des Verhaltens des Wassers in der Nukleotidbindungsregion vom Bindungszustands des Nukleotids Nahe, dass das H-Brückennetzwerk eine wichtige Rolle bei der Funk-

tion von SecA spielt. Die beobachtete Komplexität des H-Brückennetzwerk, das sich ursprünglich in der Nukleotidbindungsregion befand, weitete sich sehr schnell auf den Rest von SecA aus und zeigte wie ausgedehnt das H-Brückennetzwerk in einem Protein sein kann. Um eine Antwort dafär zu bekommen, wie wichtig diese Netzwerke sind, wurden H-Brückenkarten entwickelt, bei denen es sich um ein schnelles und effizientes Mittel handelt, mit dem H-Brückennetzwerk charakterisiert werden können. H-Brückenkarten offenbaren komplexe H-Brücken-Dynamiken in SecA, welche zur Annahme führten, dass H-Brückennetzwerke eine entscheidende Rolle bei allosterischen Änderungen in SecA spielen. Ein weiterer wichtiger Aspekt in dieser Dissertation war die Interaktion von SecA mit anderen Elementen der Translokationsmaschinerie. Die Einfluss des Vorhandenseins des Liganden auf die Dynamik von SecA wurde intensiv untersucht. Hierbei wurde besonderer Fokus auf ADP und die Präsenz des Signalpeptids gelegt. Das Interesse für diese beiden Punkte liegt darin begründet, dass: (I) die Bindung von ATP / die Freigabe von ADP zu konformationellen Änderungen in SecA führt und die Translokation des Präkursor-Proteins unterstützt und dass (II) das Binden des Präkursor-Proteins als allosterischer Aktivator für die Aktivität von SecA fungiert und unkontrollierter ATP-Hydrolyse vorbeugt. Um mögliche allosterische Effekte des Präkursor-Proteins auf die Dynamik von SecA aufzuzeigen, wurde eine Komplex bestehend aus SecA und Präkursor-Protein simuliert. Hauptkomponentenanalyse, engl. principal component analysis (PCA), wurde als Werkzeug gewählt, um einerseits die wichtigsten Moden der Bewegung unter verschiedenen Konditionen und andererseits die Kupplung zwischen unterschiedlichen funktionellen Proteindomänen in SecA zu identifizieren. Die Resultate der PCA wurden durch unabhängige Berechnungen der linearen gegenseitigen Information, engl. linear mutual information (LMI) unterstützt. LMI misst die Korrelation zwischen zwei verschiedenen Atomen. Diese beiden Untersuchungen, PCA und LMI, deuteten darauf hin, dass bei Vorhandensein von Signalpeptiden die Kopplung der Domänen von SecA über grösse Distanzen möglich ist. Die Regionen stark korrelierter Aminosäurereste von unterschiedlichen Domänen stimmen mit den Ergebnissen der H-Brückenkarten äberein und bestätigen, dass H-Brückennetzwerke mäglicherweise als allosterische Umwandler fungieren. Zusam-

menfassend lässt sich sagen, dass die in dieser Doktorarbeit vorgestellten Ergebnisse zu dem Schluss führten, dass entgegen dem Vorschlag vorangegangener Experimente nicht ein spezifischer Aminosäurerest als katalytische Grundlage dient, sondern eine Wasserbrücke zwischen zwei DEAD-Aminosäuren, D207 und E208, vorhanden ist. Ferner wurde festgestellt, dass die Präsenz von Wasserbrücken abhängig vom Nukleotid ist und sehr empfindlich auf die Mutation D207N reagiert, die zur Einstellung der katalytischen Aktivität von SecA führt. H-Brückenkarten, ein neues Mittel zur Visualisierung von H-Bräckendynamik, zeigen komplexe Verbindungen zwischen funktionellen Domänen von SecA, die über H-Brücken ermöglicht werden bei denen die helical scaffold-Domäne eine zentrale Rolle spielt. Im letzten Schritt wurde die Forschungsarbeit auf die komplexe Beziehung zwischen SecA und Signalpeptid ausgeweitet. Das Signalpeptid ist ein eine kurze dreiteilige Sequenz von Aminosäuren, die im N-Terminus jedes Präkursor-Proteins zu finden ist und eine sehr wichtige, wenn auch nicht vollständig verstandene Rolle beim Translokationprozess spielt. Die Ergebnisse der Analyse weisen auf die Bildung von Kopplungen zwischen unterschiedlichen Domänen von SecA über weite Distanzen hin. Diese Kopplungen scheinen über bereits identifizierte H-Brückennetzwerke realisiert zu werden. In dieser Dissertation wird ein möglicher Leitfaden zur Analyse ähnlicher Systeme und Probleme präsentiert, da ATP-Hydrolyse eine der häufigsten Energiequellen biologischer System ist und es sich bei Signalpeptiden um universelle Segmente handelt, die von vielen verschiedenen Proteinen erkannt werden.

Parts of this work were already published

del Val, C., Royuela-Flor, J., Milenkovic, S. and Bondar, A.N., 2014. Channelrhodopsins: a bioinformatics perspective. Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1837(5), pp.643-655.

Milenkovic, S. and Bondar, A.N., 2016. Mechanism of conformational coupling in SecA: key role of hydrogen-bonding networks and water interactions. Biochimica et Biophysica Acta (BBA)-Biomembranes, 1858(2), pp.374-385.

Milenkovic, S. and Bondar, A.N., 2018. Motions of the SecA protein motor bound to signal peptide: Insights from molecular dynamics simulations. Biochimica et Biophysica Acta (BBA)-Biomembranes, 1860(2), pp.416-427.

Acknowledgments

I would like to thank all of the people that supported me during my PhD.

Firstly, I wish to thank my supervisor, Prof. Ana-Nicoleta Bondar for the chance to step into very exciting world of computational biophysics. I would also like to thank Prof. Joachim Heberle from whose comments and views expressed during joint-group seminars and his group's seminars I learned a lot.

Also, I would like to thank all of the colleagues with whom I spent excellent 4 and a half years sharing some of the most beautiful and hardest moments of my life. Special thanks goes to Suliman, Federico and Christian with whom I was spending most of my FU time. Big thanks goes to Luka who printed this thesis and showed that he can be useful from time to time. Thanks guys!

I want to show separate gratitude to Suliman, our group's language expert, who translated the abstract of this thesis to German. Special gratitude goes to Sebastian Lorch, as well. His script was used to generate one of the Figures in this thesis.

The biggest gratitude goes to my family, my son, my wife, my sister, and my mother. Without their support this wouldn't be achievable. However, this thesis I dedicate to my father, the person who showed me the first tricks in mathematics, chess and everything else I can remember. He passed away in the meantime and I wish that he could had a chance to see this thesis completed.

My PhD was supported by a scholarship from Marie Curie International Reintegration Award IRG−276920. Implementation of data analysis tools was supported in part by the NanoScale Focus Area of the Freie Universität Berlin. Computations performed were supported by the North-German Supercomputing Alliance HLRN bec00063, the ZEDAT cluster of the Freie Universität Berlin, local cluster of the Department of Physics, and by the computing resources of the AG-Bondar laboratory. I would like to acknowledge SFB1078 whose doctoral school allowed me to attend to several very helpful workshops.

List of Abbreviations

ADP	adenosine-5'-diphosphate
ATP	adenosine-5'-triphosphate
CCM	cross correlation matrix
HSD	helical scaffold domain
HWD	helical wing domain
LMI	linear mutual information
MD	molecular dynamics
MI	muitual information
MM	molecular mechanics
PBC	periodic boundary conditions
PBD	pre-protein binding domain
PCA	principal component analysis
PME	Particle mesh Ewald
RMSD	root mean square deviation
RMSF	root mean square fluctuation
SP	signal peptide
QM	quantum mechanics

Contents

Chapter 1

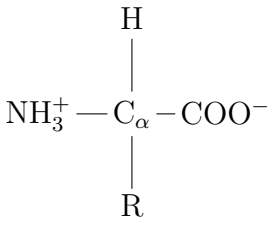
Proteins: Introduction

The name 'protein' comes from Old Greek word "$\pi\rho\omega\tau\omega\sigma$" (protos) which means "the first, the most important". The term "protein" was first used by Berzilius in 1838 in order to underline the striking importance of this group of molecules. Two centuries later, with full certainty, one can say that better name couldn't be adopted since the proteins are the most essential molecules for every living organism that ever existed on Earth. Their participation in the existence of life varies from simple constitutive roles to the initiators and regulators of the crucial processes that govern all life forms known by now. The main difference between proteins and other simpler molecules can be put in one word : **function**. Indeed, various proteins are "polished" during billions of years of evolution to perform specific functions in the most efficient way. They are not made to be beautiful or logical but, functional. Since the first proteins were solved (hemoglobin [1] and myoglobin [2]) along with its structure, protein's function was identified. Diversity of functions covered by the proteins is truly enormous: they cover transport processes, storing processes, catalytic activities, etc. The importance of studying such systems doesn't need to be emphasized.

The proteins can be divided into following classes: *i)* globular, *ii)* fibrous and *iii)* membrane proteins and they can very much vary in their size and shape but on average, each protein consists of several hundreds residues which are known as **amino acids residues** and their role and structure is going the be explained in the following subsection.

1.1 Amino acids

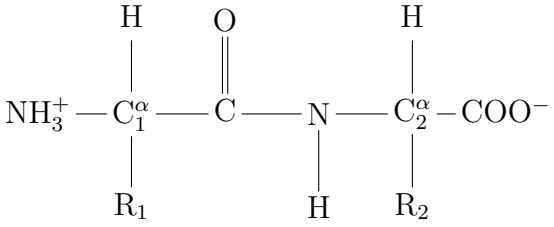
Proteins are composed of the amino acid residues and each protein is uniquely defined by the sequence of amino acids that are constituting it. This sequence is also known as the **primary structure** of the protein. In the Chem. Figure 1, the general structure of the amino acid is displayed: each amino acid consists of the central α carbon atom (C^{α}) around which a tetrahedron is formed by a protonated amino group (NH_3^+), a deprotonated carboxyl group, a hydrogen atom, and a specific sidechain usually denoted as R group.



$$\mathrm{NH_3^+ - C_\alpha - COO^-} \tag{1.1}$$

Chem. Figure 1. *A general formula for an amino acid at neutral pH.*

This dipolar form is most probable when the pH value of the solution is neutral and equals 7. Different protonation states of the amino and carbonyle groups are possible and depend on the given pH value [3].

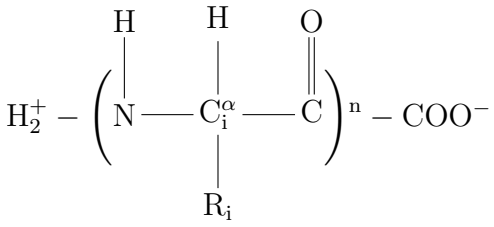


$$\mathrm{NH_3^+ - C_1^\alpha - C - N - C_2^\alpha - COO^-} \tag{1.2}$$

Chem. Figure 2. *A chemical formula for joining two amino acid residues via a peptide bond.*

In nature, a total of 20 amino acids can be found. Amino acids are usually denoted by using a three letter coding or single letter coding, and they are divided into groups: charged polar (ASP, GLU, HIS, LYS, ARG), uncharged polar (SER, THR, CYS, ASN, GLN) and nonpolar (ALA, VAL, LEU, ILE, PHE, PRO, MET, GLY, TRP, TYR) amino acids.

When many amino acids join together on a covalent basis, a polypeptide is formed. In equation 1.2 we see a formula for forming an arbitrary dipeptide. The existence of a pattern, where blocks around C_α atoms are repeating, is observable. By induction, the following general formula of the polypeptide(protein) of an arbitrary length can be written:



$$H_2^+ - \left(N(H) - C_i^\alpha(H)(R_i) - C(=O) \right)^n - COO^- \tag{1.3}$$

Chem. Figure 3. *A general formula of the polypeptide.*

1.2 H-bonds and Important Protein Parameters

Although the proteins are formed when a sequence of amino acid residues is covalently bonded, a very important type of interaction largely exploited by biomolecules is called **H-bond.** Since its formal definition by Linus Pauling in 1928 [4], the exact nature of the H-bond is under debate. The wide spread interpretation suggests that the H-bond is a weak electrostatic interaction which manifests as the attractive force between a hydrogen atom, covalently bound to an electronegative atom (known as donor,) and another electronegative atom (known as acceptor). The original work of Pauling suggests, based on quantum mechanical properties of the hydrogen atom, that hydrogen can engage solely in a single covalent bond, whereas any other bond formed by the same hydrogen atom has to be a form of electrostatic interaction between positively charge of the hydrogen and negative charge (a lone pair) of an electronegative atom originating from another molecule. The properties of H-bonds were calculated based only on electrostatics (e.g. [5]) yielding energy of 6 kcal/mol for a single bond, which seemed to be a reasonable value. This pure electrostatic interpretation was questioned promptly. The IR spectroscopy experiments detect a shift upon formation of the bond and this shift has to have its origin in the redistribution

of charge. Also, at the distances typical for H-bonds, a closed shell repulsion effects cannot be neglected. After including the repulsion term to electrostaticaly calculated H-bonding energy, previously obtained, reasonable value is lost. This indicated that the electrostatic energy is actually a contribution to the total H-bonding energy. In order to estimate all of the contributions to the H-bonding, several empirical methods were developed resulting in a conclusion that at longer distances, a prevalent contribution to the H-bond energy comes from the electrostatic interaction, whereas at closer distance quantum effects cannot be neglected and H-bond becomes partially covalent. A great summary of the methods that treat H-bonds as well as the general principles on this matter are given in the "Theory of the hydrogen bond" written by Kollman and Allen 1971 [6].

From the previous discussion and accompanying literature, it can be concluded that the question of what H-bond is, is rather complex to be answered in a simple way. However, the criteria about the actual existence of an H-bond is not so complicated. There are two parameters that determine whether or not two electronegative atoms form an H-bond and give an insight into H-bond's strength: *i)* distance between donor and acceptor (or hydrogen-acceptor distance) and *ii)* co-linearity of donor, acceptor and hydrogen (a.k.a angle). The energy of the single H-bond in proteins is usually estimated to only several kcal/mol [7], but this energy could be significantly increased by formation of the H-bonding networks [8] (please see the following section).

Another two very important parameters are denoted as ϕ and ψ, dihedral[1] angles, that describe rotation around two backbone bonds, {N-C_α} and {C_α-C}=O. These two angles display the conformational tendencies of the proteins by using a so-called Ramachandran plots [9]. By employing these plots, one can easily access the information about proteins **secondary structure**, a term that is going to be explained in detail in the following section.

[1]for definition of dihedral angles see Appendix

1.3 Protein Structural Elements

The protein secondary structure is term used to describe the 3D outlook of the protein sequence(primary structure). In general, a 3D form of the folded sequence can take many shapes, which are known as secondary structural elements and they are characterized by their $\{\phi,\psi\}$ values. The list of the common secondary structural elements with their relevant characteristics is given bellow.

1.3.1 α-helix

In a regular α-helix, a backbone hydrogen bonds are organized in a way that the backbone carbonyl {C=O} oxygen of the i-th residue forms a H-bond with the backbone NH group of $(i+4)$-th residue. The energy provided by this network of H-bonds is essential for the stabilization of an α-helix. Ramachandran plots of the α-helical structures are centered around $\{\phi,\psi\} = \{-60°, -50°\}$. Average number of the amino acids per turn in the regular α-helix is 3.6. Depending on various factors helix could be distorted. This distortion could be manifested locally in a form of a kink or globally, when entire helix tends to be curved to some extent. A great example of such

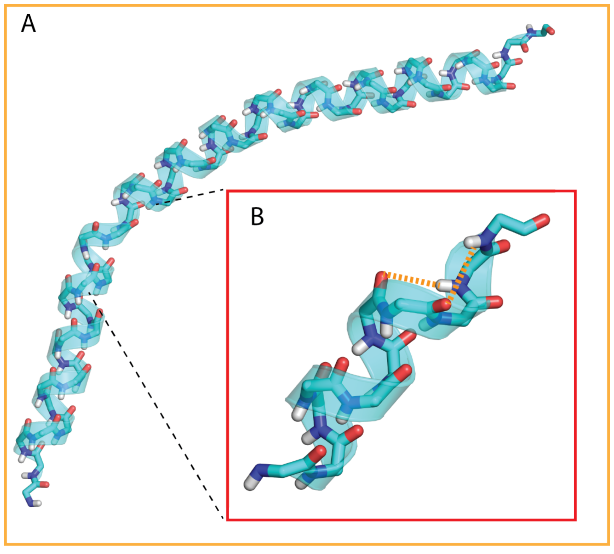


Figure 1-1: Example of the α-helical secondary structure element. (A) Long α helix from SecA ATPase's domain HSD [10]. The curvature due to a small kink(B) is observed. The kink appears as a consequence of a disturbance in the α-helical H-bonding network where two H-bonds are broken as indicated by orange dashed lines.

distortions is given in the Figure 1-1 where the long helix that belongs to SecA ATP-ase [10] is shown. Another two frequent types of α-helical structures are 3_{10} and π helices with characteristic $\{\phi, \psi\}$ values at $\{-50°, -25°\}$ and $\{-60°, -70°\}$ respectively. With 3 residues per turn and aa_i-aa_{i+3} (aa -amino acid) H-bonding pattern, 3_{10}-helices tend to be tighter than the regular α-helices. On the other hand, π-helices are a looser variations and their H-bonding pattern is given with aa_i-aa_{i+5}.

The striking biological importance of the α-helical structure, or the secondary structure in general, comes in sight when the protein translocation/insertion into the membrane is considered. The energetic cost of inserting a polypeptide chain that is stabilized by H-bonding networks formed in backbone is dramatically reduced [11,12] with the free energy estimated to 1.15 kcal/mol per residue[2] [14].

1.3.2 β-sheet

A very frequent secondary structure element is also a β-sheet. β-sheets are formed by accumulation of the β-strands which can occur in a parallel and anti-parallel manner. β-strands are a special form of helices that have only two amino acid residues per turn. Two strands are linked via H-bonds of the polar backbone atoms. Typical $\{\phi, \psi\}$ values are $\{-120°, 115°\}$ and $\{-140°, -135°\}$ for parallel and anti-parallel β-sheets, respectively. An example of the anti-parallel β-sheet is shown in Figure 1-2. This anti-parallel β-sheet also belongs to the SecA-ATPase and connects the PBD and NBD1, a SecA's functional domains [10].

1.3.3 Turns and loops

Another important secondary structure elements are turns and loops. Turns are observed in regions where orientation is sharply reversed (e.g. junctions of two anti-parallel β-strands). Loops appear mostly as short fragments that connect different secondary structure elements.[3]

[2] A detailed and very useful overview of the helical insertion into membrane and its energetics can be found in a review paper by White and von Hejine [13].

[3] Although, on the protein surface exposed to solvent long loops can be found, as it is shown in case of SecA's HWD

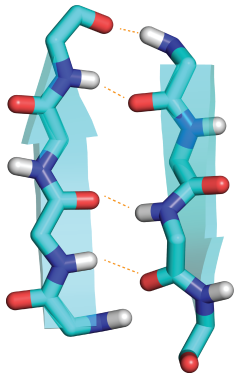


Figure 1-2: Anti-parallel β-sheet from the SecA's PBD [10]. The network of the H-bonds can be observed and it's formed by alternation of C=O:::H-N and N-H:::O=C. This alternation gives a unique, planar look to β-sheets.

Above listed secondary structure elements are explained because of their general importance and frequent occurrence. All of them participate in SecA and they are a good illustration of the how H-bonds are significant for the proteins. This significance of the H-bonding networks for the protein secondary structure was underlined due to a great focus which is going to be given them in this thesis.

Chapter 2

SecA ATPase and protein secretion

In bacteria, 30% of the newly synthesized proteins must be translocated across the cytoplasmic membrane. Machinery known as translocase, governs the entire process using the chemical energy obtained from ATP binding, ATP hydrolysis, and PMF[1] [15,16]. Depending on process being co- or post-translational, the preprotein is delivered to translocase via SRP(signal recognition particle) or SecB, a chaperone that recognizes the newly synthesized protein before it folds [17, 18]. Translocase consists of SecA- motor protein, which provides energy for preprotein secretion by employing above-mentioned ATP binding and hydrolysis, [18–20] and SecYEG, a heterotrimeric channel which can interact in some cases with SecDF, another membrane protein whose exact role in the secretion process remains unclear up to date [18]. This complex procedure is achieved by several independent proteins and demands high level of synchronization from the participants where, for instance, catalytic activity of SecA is coupled to the binding of preprotein [21, 22]. Moreover the nature of conversion of the chemical energy, which makes the biggest contribution to the translocation, to mechanical motion (chemo-mechanical coupling) remains a mystery. These reasons, combined with SecA being potential target for fighting bacterial diseases [23–25], make SecA extremely hot research topic, both experimentally and theoretically.

[1] See Appendix

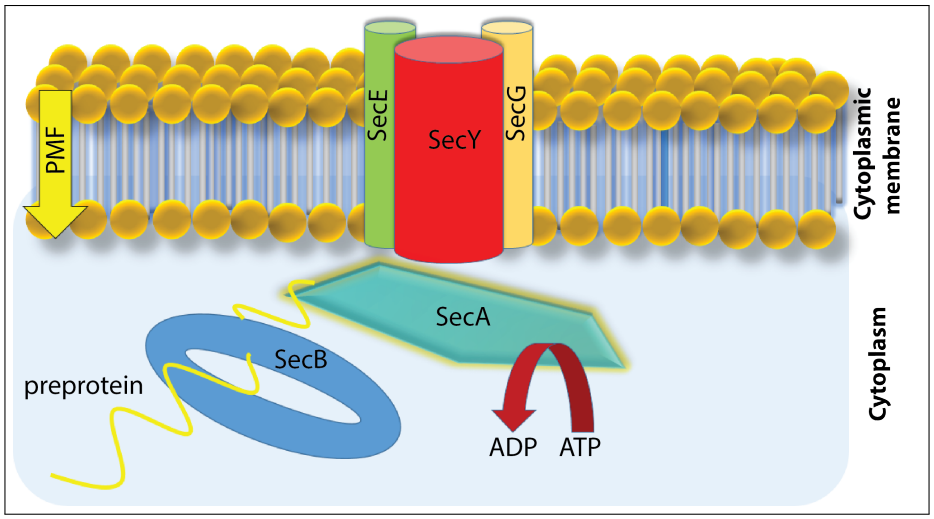


Figure 2-1: The post-translational secretory machinery without SecDF. It consists of SecYEG, which is inserted into inner membrane, SecA, a motor protein and SecB, a chaperone that recognizes the newly synthesized preprotein.

2.1 SecA

SecA is a soluble protein that plays a crucial role during process of translocation in bacteria by providing chemo-mechanical coupling which pushes the unfolded preprotein outside the cytoplasmic membrane. Architecture of the SecA is well conserved and it was first depicted by various biochemical experiments [26–29][2] and later confirmed with the crystal structures originating from *Bacillus subtilis*, *Thermotoga maritima* ,*Thermus thermophilus* etc. [10, 30–37]. This architecture consists of essentially five domains, nucleotide binding domains 1(NBD1) and 2(NBD2), preprotein binding domain(PBD), helical-scaffold domain(HSD), and helical-wing domain(HWD).

NBD1 exclusively binds the ATP molecule in DExD motif region whilst the NBD2 regulates the catalytic activity of the SecA via several interactions established with nucleotide itself and DExD region of NBD1 [38–40]. The NBD2 is also known as the intra-molecular regulator of ATPase(IRA2). Together, these two domains constitute a DEAD-motor, the minimal part of SecA that is capable of binding and hydrolyzing ATP molecule [41]. The origin of the name lies in the presence of DExD sequence, a motif found also in the DNA and RNA helicases [42]. PBD, or, as may be found in

[2]For a great review about SecA and translocon in general please see [18, 19]

the literature, preprotein cross-linking domain(PPXD) (e.g. [34]), emerges from the NBD1 and is validated to be the structural element that binds the newly synthesized preproteins [22, 43]. The position of this functional domain relative to the HWD and NBD2 characterizes the conformation of SecA, which could be open or closed [34]. For the transition between mentioned conformations, PBD must undertake a large rotation (60°) [34] which can be amplified even more when SecA is found in complex with SecYEG [10]. This marks the PBD as the most flexible domain of SecA. Helical scaffold domain (HSD) consists of three α-helical segments. The longest one spans trough entire SecA and is thought to mediate interactions among the other domains [44]. The latter two form two-helix finger which is assumed to perform mechanical work in pushing preprotein trough the membrane [45]. The functional role of helical-wing domain(HWD) is not completely clear but from the work performed in this PhD thesis [40], it can be concluded that it is one of the most intrinsically flexible domains in SecA and could serve as a recognition site to the positively charged amino termini of preprotein's signal peptide. In addition to this five functional domains of SecA, some species could also have one (*T. maritima, E. coli*) or two (*T. thermophilus*) insertions located in already defined functional domains. One of them, a variable subdomain also known as VAR domain, located in the NBD2 of the *E. coli* and the *T. therophilus* SecA was examined in a greater detail [46]. In that work, authors suggest that VAR subdomain is not essential for the SecA's function. However, in a specific organism (*E. coli*), it can play a modulating role in regulation of ATPasal activity of SecA by accelerating the ADP release step. In this thesis [40] it was shown that VAR subdomain has increased flexibility in comparison to NBD2 from which VAR emerges. VAR also engages in several H-bonding interactions with both, NBD1 and NBD2 which could play a role in ADP release acceleration.

Above described architecture is shown in Figure 2A, along with the overlap of the different conformational states of SecA in order to indicate flexibility of specific domains depicted in Figure 2B. The structure shown belongs to the *B.subtilis* SecA with the PDB:ID 3JV2. We note the significant rotational freedom of PBD which is detected in various crystal structures(80° in total). The crystal structure of another

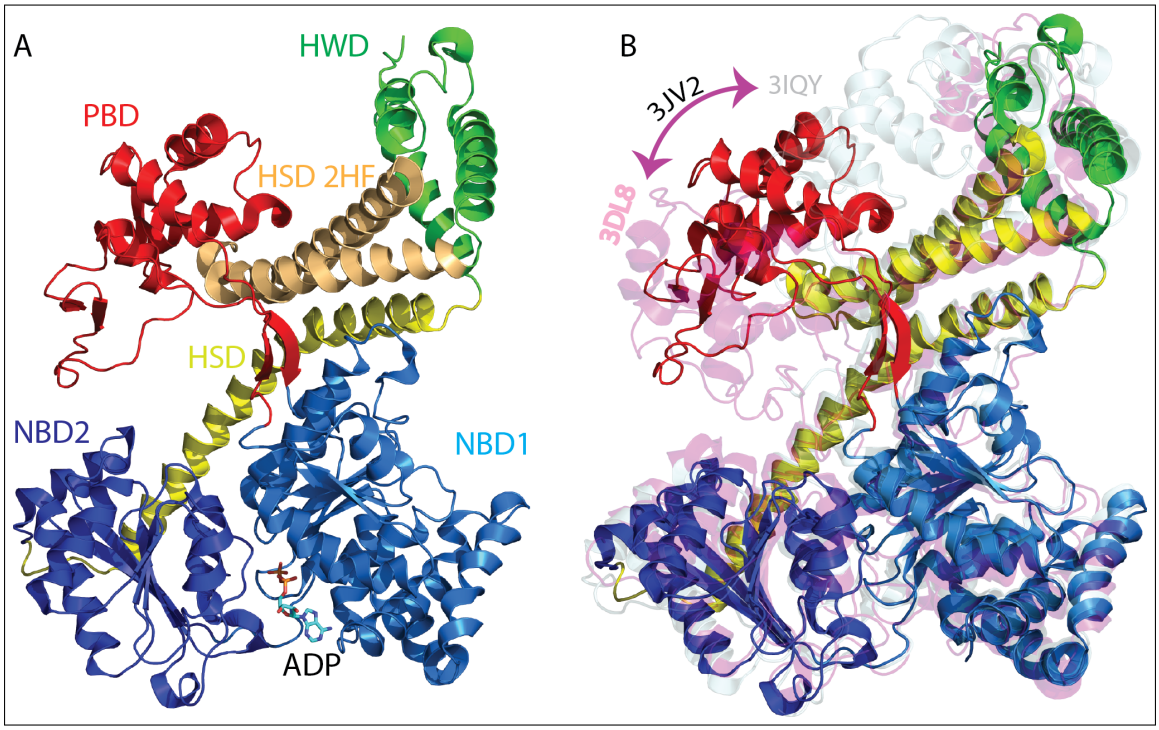


Figure 2-2: Architecture and conformational flexibility of SecA. (A) *B. subtilis* SecA [10] colored per domain:NBD1 (light blue), NBD2 (blue), PBD (red), HSD (yellow), HWD (green). Please, note that HSD's two-helix finger is in this panel colored light-orange. If not specified otherwise, HSD will be represented in an integral way, colored yellow. ADP molecule is shown as licorice and colored per atom (carbon(cyan), oxygen(red), nitrogen(blue), phosphorus(gold); hydrogen atoms are omitted). (B) Structural alignment of the SecA proteins originating from *B.sutbtilis* with PDB IDs 3DL8(transparent red), 3JV2(colored per domain) and 3IQY (transparent white) [10, 30, 47]. These structures are captured during crystallization in different conformational states. Overlap shown in this figure precisely illustrates the rigid body rotation sampled by the PBD.

B. subtilis SecA with the PDB:ID 3DL8 [47] is a very good example of the open conformational state which is characterized by the large grove betwe/en PBD, HSD and HWD also known as a cleft. This cleft is generated by the rigid-body rotation of the PBD from the HWD towards the NBD2(Figure 2B). In contrast, a closed conformation of the SecA has the PBD and HWD grouped together leaving no space in between. An example of the closed conformation is the *B. subtilis* SecA with the PDB:ID 3IQY [30]. Overall, overlap of these three different SecA structures summarize our knowledge of the domain dynamics and as mentioned before, marks the PBD and HWD as the two most flexible domains.

2.2 SecA and ATP-hydrolysis

The main energy source for the preprotein translocation comes from ATP binding and hydrolysis. ATP hydrolysis is an exoergonic reaction where the ATP is hydrolyzed by one water molecule yielding the ADP and inorganic phosphate, P_i as shown in equation 2.1

$$ATP + H_2O => ADP + P_i \tag{2.1}$$

with a range of energies between 11 and 16 kcal/mol depending on the conditions [48]. However in bacteria, the conditions are far from ideal and many parameters have to be accounted for. For instance, ATP structure is stabilized by the presence of the divalent cations, such as Mg^{2+}, and they usually form a complex in proteins. Also, the concentrations of the reaction participants are very much different [48]. These difficulties rise a barrier for a reaction to occur but still, the energy that is provided by the single reaction is very fruitful and is widely used by proteins called ATPases. These proteins can catalyze the ATP molecule by reducing those barriers, in a reaction known as the protein-mediated ATP-hydrolysis. During this reaction, the inorganic phosphate can form a covalent bond with a particular amino-acid residue, which becomes phosphorelated, or it is being released to the bulk.

Various computational studies have been performed in order to explain better the ATP-catalysis [49–54]. This studies concludde that in the proteins, interactions

established with the ATP·Mg^{2+} complex can reduce this barrier by storing water molecules at the nucleotide binding pocket. This water molecules can initiate the hydrolysis by performing an "attack" on the ATP molecule when one is found in vicinity. Proteins can also reduce this barriers by direct interactions with ATP. For example, so-called arginine fingers can help during cleavage of ATP's γ-phosphate. In work of Strajbl et al [54], it is shown that the initial barrier of 37 kcal/mol in solution is being significantly reduced in the protein.

When it comes to SecA, ATP-hydrolysis dependent secretion of preproteins can be summarized as follows: Binding of the ATP to SecA already provides the energy for the initiation of the translocation after which 2,5 kDa of the preprotein is being inserted into the SecYEG. ATP hydrolysis(ADP release) causes detachment of the preprotein from the SecA which inserts again 2,5 kDa of the preprotein into the SecYEG giving in total 5 kDa per cycle (20-30 residues) [16,20]. In this model, together with the preprotein, SecA is being inserted into the SecYEG and de-inserted upon ATP hydrolysis. However, a channel that is formed in the interior of SecYEG is too narrow to host big domains of SecA, which questions this model [55]. The alternative followed shortly, suggesting that SecA functions completely in the cytoplasm and delivers the preprotein to the SecYEG channel with two-helix finger motion [45, 47]. However, the extent and nature of the motions are under debate since the experiments which immobilize two-helix finger still recorded translocation [56].

From the previous paragraph, it can be concluded that the major conformation changes in the SecA occur upon changing of the SecA's nucleotide binding state (APO=>ATP-bound=>ADP-bound=>APO) but the mechanism of these changes is unknown. However, this mechanism can be accessed via MD simulations.

SecA binds the ATP molecule at the NBD1/NBD2 interface together with the Mg^{2+} ion [38,57]. Although it was believed in the existence of two ATP binding sites based on mutagenesis studies [38], it was shown that there is only one binding pocket for the ATP molecule located in the NBD1 while the residues of NBD2 (also known as IRA2) regulate the activity of catalysis itself [39]. Catalytic activity is regulated also by the two-helix finger domain (also known as IRA1) which is part of the HSD [58],

but it is shown that for the efficient binding and hydrolysis of the ATP, SecA needs only the DEAD motor domains NBD1 and NBD2 [41].

The key role in catalytic process is frequently associated to one of the first two DEAD motif amino acid residues. Mutational experiments clearly indicated that the mutation of both, D207 and E208 (*B.subtilis* numbering) inhibits the ATP binding and preprotein translocation [38], and up to recently, crystallographic structures assigned to one of them a storage of the catalytic water molecule via formation of the H-bond with the water molecule. After the hydrolysis process, ADP leaves the SecA. There is no sign of the phosphorelated amino acid residues after the catalysis suggesting that the inorganic phosphate is released directly to the bulk. A closer inspection of the nucleotide binding pocket in the *B.subtilis* SecA [10] reveals a complex network of protein-nucleotide and protein-protein interactions which is shown in Figure 3. We anticipate also a cluster of water molecules in the vicinity of the ADP molecule.

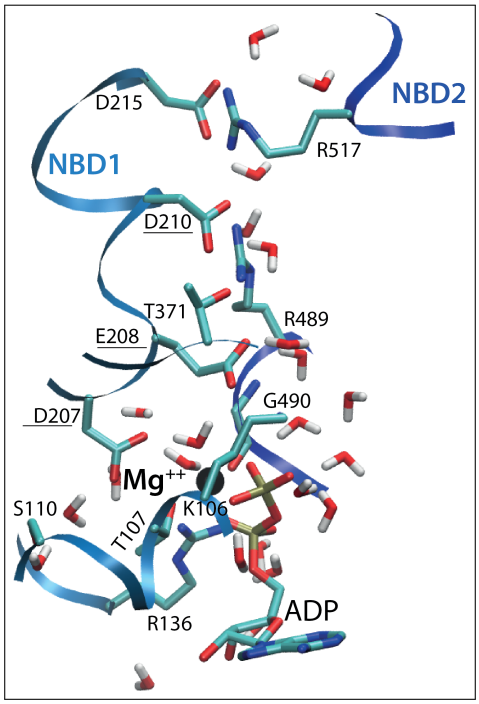


Figure 2-3: Nucleotide binding pocket of the *B. subtilis* with ADP and Mg^{2+} ion bound. The nucleotide binding pocket is located on the NBD1/NBD2 interface which are colored light blue and dark blue, respectively. The pocket is filled with water molecules and consumed by numerous H-bonding interactions formed by protein, ADP and water.

2.3 SecA and Preprotein

Although the SecA provides the chemo-mechanical coupling needed for the translocation process, the main protagonist of the process is a pre-secreted protein, also known as the preprotein. After synthesis, the preprotein is urgently recognized by the SecB, a chaperone protein, that delivers it to the SecA, a motor protein. The molecular details of how SecA interacts with the preprotein are poorly understood, but it is known that SecA recognizes the N-terminal segment of the secretory protein, the signal peptide. The signal sequences of different secretory proteins destined for the Sec pathway share common physical chemical properties: They have several positively-charged groups at the N terminus, an intermediate region that is hydrophobic, and a C terminal region that contains small, polar groups [59–61] (Figure 4B). Analyses of the signal peptide sequences of secretory proteins from various organisms indicated significant variation of the amino acid residues [62]. However, this variation depends on the region: C-terminus does not vary much, neither in length nor in sequence. In contrast the N-terminus varies significantly in both parameters but keeps positive net charge on the average value of +1.73. The governing principle behind variation in hydrophobic core is overall hydrophobicity [62].

The location of charged groups along the signal peptide sequence appears to be of paramount importance for the preprotein translocation. Experiments on prolipoprotein from the *E.coli* indicated that gradual replacing of N-termini positive charge from +2 to -1 led to marked decrease in the lipoprotein production, where the -1 charge drastically reduced the translocation of prolipoprotein [63]. Experiments done in the same fashion confirmed that translocation is most efficient when amino terminus has at least net charge of +1 [64, 65]. Moreover, mutagenesis studies that mutated the amino acid residues without changing the net charge (e.g. LYS to ARG) left the translocation rate unaffected [66]. Underlying reason for such high charge sensitivity may be the strict determinacy of important electrostatic interactions between signal peptide N-terminus and N-terminus of exported preprotein [67] or disruption of potentially functionally important dipolar character of the signal sequence [62]. With

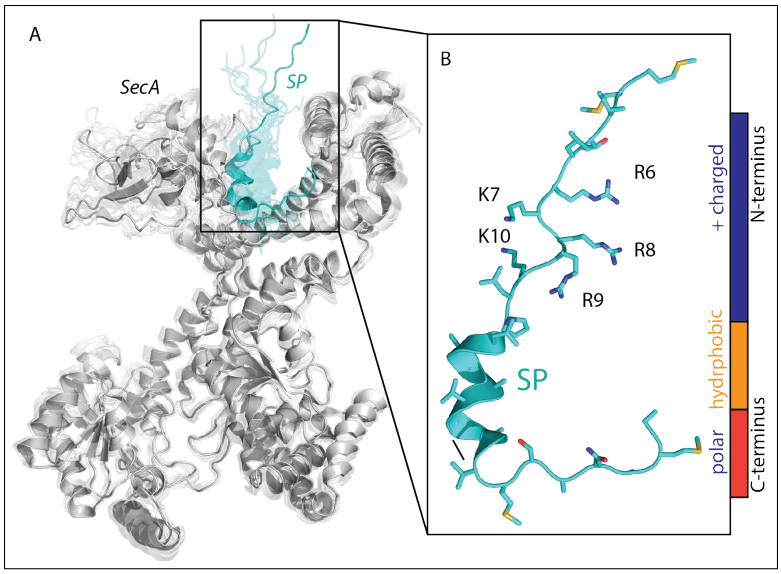


Figure 2-4: Signal peptide and SecA. (A) Overlapped NMR ensemble of the *E. coli* [31] SecA (PDB ID: 2VDA) solved with the KRR-LamB signal peptide located in the cleft formed by PBD,HWD, and HSD. (B) Enlarged structure of the SP taken from the first snapshot of the NMR ensemble from the panel A. We see a clear tripartite structure schematically represented by the color bar on the right.

SecA being rich in both positively and negatively charged amino acid residues, a dipolar character of the signal sequence indeed may play a significant role.

Even more strictly than the N-terminus, the hydrophobic core of the signal peptide displays high sensitivity to any perturbation imposed. The deletion mutation that removes seven residues form the hydrophobic region of the signal peptide completely prevented the secretion [68]. Alteration of hydrophobic composition affects translocation rates and it was shown, by interchange of hydrophobic residues, that the net hydrophobicity is crucial for efficient translocation [69]. Together, the positively charged N-terminus and the hydrophobic core interdependently contribute to the signal peptide function [70]. This is clearly visible if the SecA-SP complex [31]: In this NMR studies of the of E. coli SecA in the presence of a LamB signal peptide is observed. It can be seen that the signal peptide binds in a large groove generated by PBD, two-helix finger of the HSD, and HWD as shown in Figure 5. The nature of the groove is largely hydrophobic but there are many polar groups residing in its vicinity. Results of that study suggest that the HWD of SecA recognizes positively charged signal peptide sequence while the hydrophobic core of the signal peptide interacts

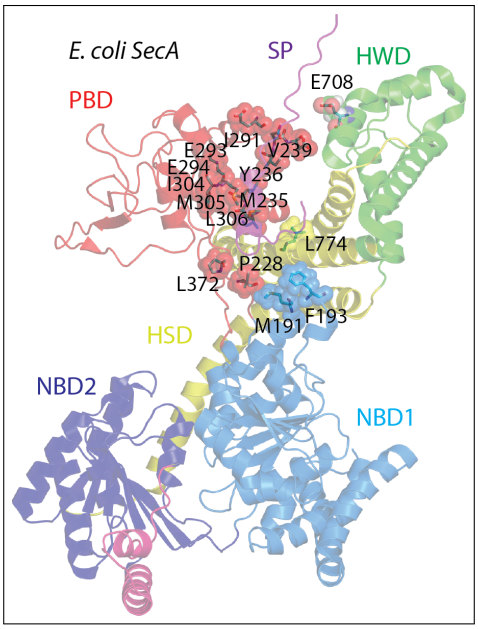


Figure 2-5: *E. coli* SecA/SP complex colored per domain. NBD1(light blue), NBD2(blue), PBD(red), HSD(yellow), HWD(green), VAR(magenta) and SP(violet). The SecA's amino acid residues that have been identified to engage into interactions with the SP are explicitly labeled.

with the PBD amino acid residues Ile225, Met235, Val239, Ile291, Met292, Ile304, Met305, Leu306, Val310 and Leu372, and Leu774, Met810 and Met814 from HSD.

In addition to their role in targeting the newly synthesized peptide to the Sec pathway, signal peptides (SP) are thought to function as allosteric activators of the translocase [21]. In general, allosteric regulation assumes binding of a regulator substrate to a site that is different from the active site of the protein. This binding is then coupled to the active site and it can promote proteins activity (allosteric activators) or reduce it (allosteric inhibitors). There are two main models of allosteric regulation: concerted [71] and sequential [72] model.

The allosteric activation proposed in SecA is assumed to work in three steps: Triggering, that brings translocase to a lower activation energy; Trapping, which secures binding of the preprotein to the translocation machinery; and Secretion, that promotes the translocation of already trapped preprotein in cycles [16,20]. Moreover, a signaling between the PBD, after the preprotein binding, and catalytic site at the NBD1-NBD2 was observed. This signaling is managed by the salt bridge named

Gate1 [22].

The above-mentioned aspect of the allostery in the SecA-SP relation was investigated with a special focus in this thesis. Since binding of the SP assumes conformational or at least changes in dynamics that promote ATPasal activity of SecA MD simulations could be a useful tool of tracking these changes on atomic scale. This atomic scale description could possible shed the light on the mechanism of the coupling between two distal sites, ATP binding site and SP binding site.

2.4 SecA and SecYEG

The purpose of the SecA's work is concentrated in its interaction with a heterotrimeric membrane channel, known as SecYEG, which can transclocate newly secreted proteins across the cytoplasmic membrane [57, 73]. Experiments show that the translocation process is achieved with SecA inserting two-helix finger part of the HSD which engages in interaction with the preprotein [45, 47]. The binding of SecA to the SecYEG is also found to be crucial for the above mentioned allosteric coupling between the preprotein binding site and nucleotide binding site. It is thought that the SecA-SecYEG binding event synchronizes the ATP cycles and the preprotein binding-release events. This leads to the preprotein translocation divided in steps [16, 20, 26]. Although of great importance, SecA-SecYEG interactions appear very difficult to investigate. MD simulations could possibly reveal nature of interactions established between SecA and SecYEG but the structures available at the time when this thesis began were not promising candidates for MD investigations [47]. X-ray resolutions of those structures were 4.5Å (PDB ID:3DIN) and 7.5Å (PDB ID:3DL8) with multiple strings of missing amino-acid residues. In Figure 6 we can see SecA(*B.subtilis*)-SecYEG(*A.aeloicus*) complex [47]. The resolution of this x-ray structure is 7.5Å. In the panel 6B, partial penetration of the two-helix finger into the SecY is indicated.

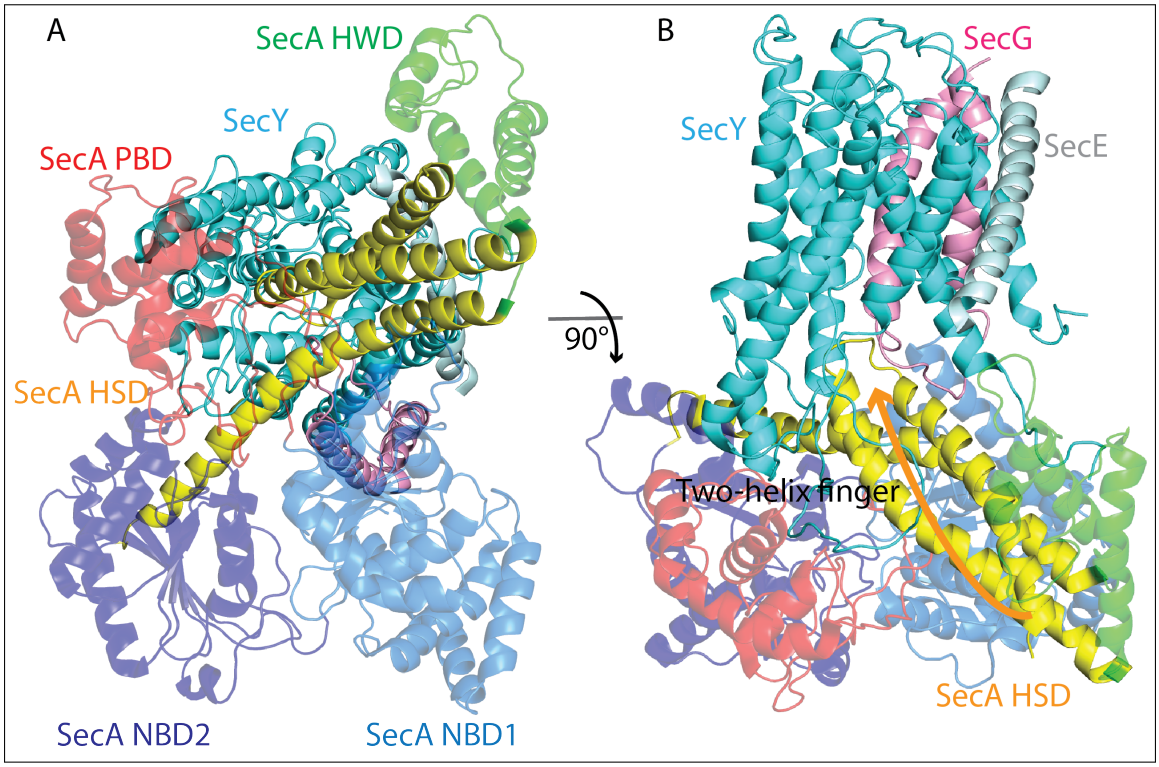


Figure 2-6: SecA/SecYEG ineteraction. (A)SecA/SecYEG complex [47]. The domains are colored by the scheme previously established in Figure 2-2. SecY is colored cyan. (B) SecA/SecYEG complex after 90° around the Y-axis. The proposed HSD's two-helix finger (2HF) insertion into the SecY is indicated with the orange arrow.

Chapter 3

Methods

3.1 Starting Coordinates

Xray and NMR. In order to animate the proteins by application of the desired force field, we need the starting positions of atoms the protein is constituted from. Although sequences of the proteins are widely accessible due to efficient genome-sequencing mechanisms present in recent years, a fold of the specific protein based on the sequence alone is theoretically not accessible in most cases. Prediction of the exact fold of the protein, or a 3D positions of the amino acid residues inside the protein and their atoms is a top priority question in computational biophysics today. This leaves us to the structures that are obtained by some of the experimental methods. Up to date, there are two dominant methods of the structure determination: X-ray and NMR crystallography. Successful application of the X-ray crystallography on the biological entities such as proteins started in the 1950's [1, 2]. Since then, this method resolved 110,000 structures which, in comparison to the structures resolved by the NMR (10,000), shows that the X-ray crystallography is established as the most dominant method in the field [74]. This is one of the reasons why the structures obtained by X-ray where confidently used in this thesis.

X-ray structures. The biologically relevant crystal structures are collected in the protein data bank [74], where one can access all the important data regarding the crystal structure of interest, concentrated in two files, a .pdb file(coordinates,

space group, etc.) and .fasta file (complete amino-acid sequence of the protein). Many times, the choice of the coordinates that one wants to use as the starting points for simulations is not unique. There could be several structures of the same protein, sometimes originating from different organisms [10,33], or proteins from the same organisms but obtained under different conditions [10,34]. This suggests that a choice has to be made. Before making the choice, there are two main parameters that need to be considered: The resolution of the crystal structure and the number of the amino acid residues missing. Crystal structure resolution is a parameter that is tightly connected with the quality of the crystallization process and shows how sure we are in the positions assigned to the atoms inside the protein. It is expressed in angstroms and quantifies how detailed a given protein is described. This means that, for example, a crystal structure with the resolution of 2.5 angstroms is more reliable then one with the resolution of 3.5 angstroms. Other important parameter, the number of missing amino-acid residues, is the factor that could impose a significant modeling problems. The missing amino-acid residues usually are located in the most flexible parts of the protein such are termini or interior loops. X-ray diffraction on this dynamic segments of the protein results in a blurred electron density map which could be non-conclusive leaving the gaps in the final structure. Smaller strings of missing amino-acid residues in the protein termini could be neglected during modelling because of the possibly small impact on the protein dynamics, but any missing residue in the middle of the protein has to be modeled in order to preserve a continuous chain of amino-acid residues. Also, every missing atom in the amino-acid residues that we include in our calculations has to be re-built.

NMR. Nucler Magnetic Resonance is one of the most promising techniques for determination of the protein structure. Briefly, the main advantage of NMR lies in fact that NMR can resemble the protein's dynamics. This provides an insight into various properties which can help the understanding of protein's function. A short, but yet very helpful summary of the method is given by one of it's fathers, K. WÃijthrich [75].

In this thesis total of seven different systems were designed using the X-ray struc-

tures of SecA from three different organisms (*B.subtilis, T.thermophilus, T.maritima*) [10, 33] and one NMR conformation of the signal peptide taken from the *E.coli* [31].

These structures, accompanied by the relevant parameters, are summarized in Table 1.

Organism	Type	PDB ID	Resolution	Missing amino acids
B. subtilis	X-ray	3JV2	2.5Å	ChainA:1-14,645-651,780
T. maritima	X-ray	3JUX	3.1Å	ChainA:1-3,817-822
T. thermophilus	X-ray	2IPC	2.8Å	ChainA: 940-997
E. coli	NMR	2VDA	Not aplicable	ChainA:1-8

3.2 Force field

Force field or *molecular mechanics* method is currently one of the most exploited techniques in the theoretical investigations of the bio-molecules (see [76–78] etc.). Force field is an explicitly classical method which treats molecules as sets of point-masses (atoms) with specific properties that are connected by springs (bonds). Although very rough and simplified, the representation of the molecules as a mechanistic bodies gives a good-enough description of the molecular properties and on-going processes if the implicit quantum mechanical part is derived correctly. However, the classical nature of such description, regardless of its implicit quantum properties, limits its application to the processes that do not involve events governed mainly by quantum mechanics such as breaking or forming of the covalent bonds, for example. By combining the QM calculations, a vast search for appropriate parameters trough the experimental databases, and self consistent minimization procedures, the force field potential is finally constructed.

3.2.1 Foundations

The force field doesn't contain quantum mechanics explicitly, but its foundations are deeply quantum mechanical. In quantum mechanics every system is fully described with a *wave function*. If that system is non-relativistic, its wave function can be

obtained by solving the Schrodinger's equation. The Schrodinger's equation is one of the most simple looking, but nevertheless most powerful assets in the world of physics. It can be formulated as

$$H\Psi_n = E_n\Psi_n \tag{3.1}$$

where H represents a Hamiltonian of the system and Ψ_n and E_n are eigenfunctions and eigenvalues of the system, respectively.

Many QM methods that can treat various biophysical systems were developed recently, but they all suffer from more or less the same weakness- they are computationally too expensive when it comes to huge bio-molecules such as proteins. In principle, these QM methods can be divided in two groups *i)* ab initio QM and *ii)* semi-empirical methods. Ab initio methods are focused on "strict" calculation of the Schrodinger's equation. This calculations are not completely exact: they lean on the Born-Oppenheimer approximation, while assuming that relativistic effects can be neglected, as mentioned before. For example, one of the most famous variations of the ab initio methods is DFT (density functional theory), originally developed by Hohenberg and Kohn [79] and has been successfully applied to a range of different problems in physics.

Semi-Empirical methods [80, 81], on the other hand, use the pre-obtained parameters for the atomic orbitals and energies. The key of such methods lies in parameterization (obtaining the parameters needed for calculation), which can be a very difficult task (Overlap integrals in SCC-DFTB, for instance [82]).

3.2.2 Force Field Properties

The practical implementation of the force field potential relies on three main principles: *i)* Thermodynamic hypothesis *ii)* additivity and *iii)* transferability of the effective energy potentials.

Thermodynamic hypothesis states that proteins are driven naturally to their native structure (a global free energy minimum). This is supported experimentally

for many different small globular proteins [83]. On the other hand, a theoretical study demonstrated that the probability for a protein to adopt its native state increases when temperature decreases [84]. Shortly, the thermodynamic hypothesis can be formulated as: **Sequence implies structure**. However, with an enormous configuration space available to average-sized proteins (For 10,000 atoms, in a simplest, mass-strings model $3N - 6$ degrees of freedom has to be accounted for, where N denotes the number of atoms.) it becomes clear that proteins cannot exploit configuration space randomly in order to find their native-state on the experimentally known time scale. This intuitive observations were summarized by the Levinthal in 1969 [85, 86] and are denoted as "Levinthal paradox". Levinthal concludes that predefined, specific folding pathways have to exist. This view was updated recently in a sense that there are many competing pathways that lead to the native fold of the protein [87].

Additivity in force filed potentials can be viewed as an assumption that the effective (one that matters) potential is a sum of the simple potentials: electrostatic potential, van der Waals potential, bonds, angle between two bonds and torsions of two different groups around their joining bond. The previously listed interactions can be intuitively divided into *bonded and non-bonded potentials*.

Bonded part of the force field is usually given by

$$E_{bonded} = \sum_{b_i} E_{bond}(b_i) + \sum_{\theta_i} E_{angle}(\theta_i) + \sum_{\tau_i} E_{tor}(\tau_i) \tag{3.2}$$

where functional forms are usually harmonic (proportional to squared displacements) and trigonometric. In general, there is a possibility for a cross-terms to exist (bond-bond or bond-angle interactions, for example), but some of them can be neglected in case of the large bio-molecules such as proteins. **Non-bonded part**, however, consists of the interactions achieved not vie "mechanical" strings. This interactions are present also between a non-neighboring amino acid residues and they depend on

the inter-atomic distances r_{ij}:

$$E_{non-bonded} = \sum_{i,j;i<j} E(r_{ij}) \tag{3.3}$$

Usually, non-bonded term is constituted from Coulomb and van der Waals interactions.

This intuitive separability can lead to some computational benefits. Since the non-bonded terms tend to change more slowly than the bonded terms with distance and time, different updating frequencies can be adopted during computation. This is known as *multiple timestep* technique [88–90] (See section about MD).

Transferabilty of the force field assumes that potentials can incorporate all of the experimental and QM derived data for representative structures and then be used to describe a larger biomolecular entities composed of these representative structures. This means that one can take that, for example, bonds between the same atoms should exhibit similar behaviour as parts of different structures. However, an environment dependency of geometric properties has to be accounted for. One of the ways, heavily exploited for big biomolecules, is to introduce 'atom types'. Atom types reflect the nature of molecular environment and are made to fit well experimental findings. Currently in CHARMM 22 force field (proteins) [91,92] and CHARMM 27 force field (nucleic acids) [93,94] there are 160 different atom types.

3.2.3 Force Field Form and parameterization

Summarizing the previous chapter, we can write the exact form of the force field used in this thesis. In the following equation CHARMM22 force field potential is given:

$$\begin{aligned} E = &\sum_{bonds} k_b(b-b_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} k_\phi[1+cos(n\phi-\delta)] + \\ &\sum_{impropers} k_\omega(\omega-\omega_0)^2 + \sum_{Urey-Bradley} k_u(u-u_0)^2 + \sum_{residues} u_{CMAP}(\Phi,\Psi) \\ &\sum_{nonbonded} \left[\epsilon_{ij} \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - \left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon r_{ij}} \right]. \end{aligned} \tag{3.4}$$

The first term considers the bond stretching with k_b being the force constant and $b - b_0$ is the displacement from the equilibrium position. The following term accounts for the angles formed by two neighboring bonds where, again, k_θ is the force constant and $\theta - \theta_0$ is displacement angle from the equilibrium. The third term considers dihedral (torsion) angles while the forth one takes into account impropers (out of plane bending) with k_ϕ and k_ω being the respective force constants. The fifth term, or Urey-Bradly term, quantifies a crossover interactions (angle bending). Finally, force field potential accounts for the non-bonded interactions which are comprised of the 12-6 Lennard-Jones potential and electrostatic Coulomb potential where q_i and q_j are partial charges of the atoms and r_{ij} is corresponding inter-atom distance.

As it can be seen from the previous equation, there is a number of parameters, such are force constants, and equilibrium values that have to be inserted before any calculation of the potential can be performed. Many of them can be obtained trough the experiments, but in some cases they have to be calculated. For instance, partial charges for a particular force field cannot be obtained from the experiments (literally, they can but they would have to be recalculated in order to be used in particular force field) since they are a resemblance of quantum mechanics in the classical systems. They have to be derived by performing the QM calculations in a manner that fits to already known experimental data (bond lengths, for example). The procedure of obtaining missing parameters for a specific force field is called *parameterizaton*. Good parameterization is essential for the practical implementation of the particular force field since it guaranties that calculations are going to resemble molecular interactions and geometries properly. Due to its complex form and implicit QM dependence, force field parameterization has to be done in a self-consistent manner [95]. Roughly, the force constants and the partial charges are being updated until the consistency is achieved. After the parameterization is completed, topology and parameter files are created. This files contain a complete set of information needed for performing any kind of the calculation that involves force field potential.

3.3 Setup

3.3.1 General Approach

Once the force field and coordinate set are chosen, one can begin with the preparation of the system for the simulation. The software of choice for the system preparation is CHARMM [77], which is compatible with the selection of the force field. The preparation of the system for md simulation goes as follows: Firstly, the topology and parameter files are read [91–94] with the amino-acid sequence of the protein. This loads a polypeptide chain which do not possess any relevant spatial characteristics of the protein but has the relevant force field parameters and topology assigned to the atoms. Sequentially, the coordinates obtained from the .pdb file are assigned to the loaded sequence of the amino-acid residues. This creates the protein conformation obtained in the X-ray structure. The following task is rebuilding of all of the atoms that are missing in the .pdb file and modeling of any interior sub-sequence of the amino-acid residues not obtained by the X-ray crystallography. Next in the general procedure of creating a biologically relevant environment for the MD simulations is to generate a water-box where the protein can reside. This represents now the unit cell of the MD simulation and it is multiplied infinitely in order to create periodic boundary conditions (PBC). Unit cell can contain an electric charge coming from the protein itself which could impose a rise of energy during the simulations. This may eventually lead to instabilities since a total system charge would become infinite (due to the nature of the PBC). One can deal with this issue by randomly replacing water molecules with neutralizing ions until total system charge is set to 0. After this, the system should be ready for the molecular dynamics simulation. Due to a small but still significant variations between different setups in this thesis, in Results Chapter, each setup is going the explained precisely.

3.4 Molecular Dynamics Simulations

"Molecular dynamics is clearly on the way to being a universal tool, as if it were the differential calculus." Sir John Royden Maddox, British science writer and Nature editor (1925-2009)

Molecular dynamics(MD) simulations represent a fine intersection between physics, chemistry, and biology. Due to its complexity, this junction has to be supplemented by mathematics and computer science. MD can implement valuable snapshots of structure obtained through experiments and animate them in order to access interesting, by experiment hardly accessible properties. For example, initially "frozen" X-ray structure of a protein can access many thermal states through MD simulation. By observing its evolution, and analyzing the data obtained, the motion of the protein could possibly connect to its function as done in pioneering work of Frauenfelder et al [96].

Regardless of their popularity and usefulness, MD simulations have their limitations. Some of those limitations come from the fact that MD simulations neglect the electronic motions but consider only the motions of the atom nuclei. This imply that they are not capable of capturing the changes in electronic structure caused by bond formation or breaking, polarization and metal ions binding. Another aspect, which is important in considering whether or not MD simulations should be used, is temperature. Conventional MD simulations do not describe well the low temperature systems. Reason for this inability is of quantum mechanical nature: On low temperatures system is close to its ground quantum state with the gaps between discrete quantum states being larger than the thermal fluctuations. This condition can be mathematically expressed as

$$\frac{h\nu}{k_B T} >> 1 \tag{3.5}$$

where h is Plank's constant, ν is frequency and $k_b T$ is thermal fluctuation of the system.

However, first of the two mentioned limitations can be addressed. In recent years,

hybrid methods (QM/MM) emerged. These methods divide the system into classical and quantum part. Classical part is treated with MD and quantum part is described with one of the QM methods [97]. This approach makes simulation of the chemical reactions that involve proton transfers possible (for example [98]).

The principal equation that lies behind the MD simulations is Newton's second law of motion

$$m\ddot{\vec{x}} = \vec{F}(\vec{x}) \tag{3.6}$$

where force is given by

$$\vec{F}(\vec{x}) = -\nabla V(\vec{x}) \tag{3.7}$$

where $V(\vec{x})$ is a force field potential, in MD simulations case. As it is clear from the previous equations, MD simulation results strongly depend on the force field accuracy. Another important condition for validity of the MD simulation is the protocol. In general, a MD simulation consists of three crucial parts: heating, equilibration, and production. **Heating** of the system is performed when velocities are assigned to the atoms which have coordinates obtained in a manner described in the section 3.3. The velocity distribution to the atoms is done a way that can mimic the desired temperature of the system according to

$$\langle E_k \rangle = \frac{3}{2} N k_b T. \tag{3.8}$$

After the heating has brought the system to desired temperature, an **equilibration** is needed. During the equilibration, the exchange between the potential and kinetic energy occurs. Convergence is achieved when these energy terms start to have a non-significant fluctuations around their average values. The final step in this procedure is so-called **production**. During the production run, trajectory of the system is being generated with the purpose of post-process analysis. This analysis of the bulk data obtained by the MD simulation serves to extract important protein features that are hardly accessible with the experiments or to confirm previous experimental findings.

Protocol employed. During MD runs the all atom CHARMM force field was used [77, 94], including the CMAP correction for the protein groups [92]. For the ADP molecule, CHARMM27 nucleic acid parameters were used [93, 94]. For the solvent, the TIP3P water model was used[1] [99]. The simulation runs were entirely run using NAMD [100]. SHAKE [101] was employed to constrain the bonds that involve hydrogen atoms. Electrostatic interactions are calculated with the smooth particle mesh Ewald summation [102, 103]. To achieve continuity of the derivatives, the switching 8-12 function was used. Temperature was kept constant by Langevin dynamics, whereas the pressure was maintained on the value of the 1 bar by using the Nose-Hoover Langevin piston [104, 105]. During the production runs, a multiple time-step integrator was used with 1fs, 2fs, and 4fs frequencies for the bonded, van der Waals and Coulomb interactions, respectively [88–90].

[1]The reason for choosing TIP3P among various water model available is the fact that TIP3P water was used during CHARMM parameterization

3.5 H-bonding maps

As a complete novel tool for quantification and visualization of the H-bonding dynamics in biomolecular systems, H-bonding maps were introduced. The H-bonding maps are implemented within VMD and use (A)acceptor-H(hydrogen) distance as criteria for detection of H-bond. As stated in the introduction there are two parameters that define a H-bond: A-D(donor) distance and A-D-H angle. However, these two parameters can be merged in one: A—H distance. This can be done due to the difference in electronegativity between H(positive) and D(negative), and pure geometric considerations of positions of H-atoms in amino acids[2]. Main reason for deciding to have one single parameter as the H-bonding criteria was computational speed[3] at the time when this tool was implemented for the first time. Nevertheless the reasoning for choosing only one parameter, A—H distance, as the criteria and choice of the value of 2.5 Å, as the cutoff distance, are given in the following paragraph.

Choice of parameters for detection of H-bonds is not trivial. One can choose either a geometric criteria, which usually involves the A—D distance and A-H-D angle [106], or an energetic criteria [107]. Due to its complexity, energetic criteria is not going to be considered here. The must of having distance and angle criteria met as necessary and sufficient conditions for H-bond existence lies in fact that if one uses only A-D distance, there is no guarantee that the hydrogen atom has an optimal orientation for forming the H-bond. To check that, one has to consult the A-H-D angle value. However, if A—H distance is considered, when the proper cutoff is chosen, one knows that hydrogen atom is indeed in a position to form an H-bond. Angle criteria in this case is not a necessity, since the "good" angle condition has to be met if hydrogen and acceptor are close enough. A thorough analysis of the H-bonding parameters is given by Buck and Karplus [108] where they show that a value of 2.5 Å for A—H distance is a very good choice for the H-bonding criteria in proteins which can be seen from the probability distributions in their work. If coupled with angle of 90° it covers well the

[2]The idea for this H-bond criteria comes from the discussions with my colleague Mr. Suliman Adam

[3]In VMD's scripting language, TCL, this really matters.

experimental finding on number of double H-bonds in proteins. They also show that the population of A-D contacts below 2.5Å is negligible. This allows me to make some geometric considerations that would show the choice of 2.5Å for A-H distance cutoff is sufficient for reliable detection of the H-bonds. If, hypothetically, a typical protein acceptor (oxygen atom) and a typical donor (nitrogen atom) are placed such that O—H-N distance is 2.5 Å and O-H-N angle is 90°, as in Buck-Karplus limit case, a simple application of the Pythagoras' theorem yields that the distance between donor and acceptor will be 2.7 Å. However, the sum of vdW radii of oxygen and nitrogen is larger than that, as shown in figure 3-1. In experiments this value ranges between 2.8 and 3.2, whereas in CHARMM force field [94], used in this thesis, depending on the exact atom types this sum takes a range of values, all larger than 3.5 Å. This suggests that the overlap between vdW spheres is significant in this limit case. This overlap

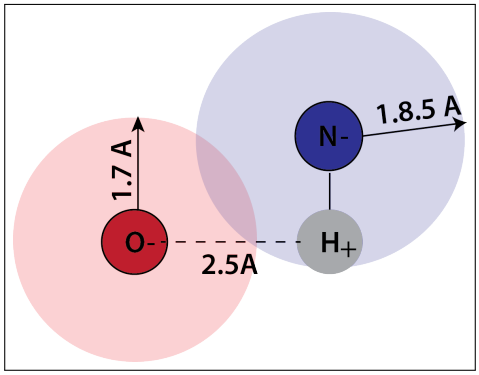


Figure 3-1: An illustration of H-bonding geometry in Buck-Karplus limit case. Overalap of vdW spheres suggests the unlikeliness of this event.

of vdW spheres would, together with Coulombic repulsion(both, acceptor and donor are electronegative), make this situation highly unlikely and non-sustainable. In the summary, these considerations suggest that using A—H distance as the only criteria would successfully cover all of the relevant situations for H-bond formation due to the vdW and Coulombic interactions. A table with the values of vdW radii is given bellow.

Source	O	N	H
Pauling [109]	1.4Å	1.5Å	1.2Å
Bondi [110]	1.52Å	1.55Å	1.2Å
Rowland [111]	1.58Å	1.64Å	1.1Å
CHARMM22 [91]	1.7Å	1.85Å	1.1Å

Finally, to create a H-bonding map, the following procedure is applied. Using the A—H cutoff distance of 2.5Å, trajectories are being scanned for possible H-bonding interactions. Each interaction is than scaled to represent a relative frequency of occurrence taking values between 0% and 100%. After the scaling, amatrix is created and plotted as a 2D map having as a coordinates H-bonding donors, on X-axis, and acceptors, on Y-axis. If the H-map are applied to single structure points on the map are colored the same and if H-bonding map is created for an ensemble, points are colored according to the respective H-bonding frequency. In this thesis H-bonding maps are always calculated for the last 30ns of a trajectory.

3.6 PCA and LMI

With 5 functional domains that are working together in order to facilitate preprotein transport by chemo-mechanical coupling, SecA is a very complex system to study even with the help of the MD simulations. In order to extract relevant and functionally important information, PCA (Principal Component Analysis) [112–114] and LMI (Liner Mutual Information) [115] were employed. **PCA** is a method that relies on diagonalizaion of a covariance (cross-correlation) matrix calculated for a given set of time-series and it was introduced by Karhunen in 1947 [112]. Its application to the proteins came much later with the computer power that enabled the simulations of large proteins and created a need for a method that can easily dissect important modes of protein motion [113, 114]termed essential dynamics. Indeed, essential dynamics of a protein, as named by Amadei and coworkers, provides an insight into the nature of motions sampled by protein during simulation. This is done by removing the noisy

modes of motion and focusing on a few modes that matter the most. In case of the protein dynamics, the time series used for PCA represent the positions of each atom that is taken into consideration. In this thesis only the Cα atoms are used as representatives of each amino acid residue when it comes to both, PCA and LMI analysis. Procedure of obtaining principal components applied to the proteins can be given as: *i)* The time series of Cα atoms were recorded during the simulation, after all of the recorded snapshots have been aligned. *ii)* The average positions of all the Cα atoms were determined. *iii)* These positions were then used as the reference values when displacement of each Cα atom was measured. *iv)* Radial vectors of Cα displacements are then employed to calculate pairwise correlation coefficient (Pearson coefficient) and construct a cross-correlation matrix C.

The cross-correlation matrix holds the information about correlated motions of the protein and takes values between -1, for fully anti-correlated, and 1, for fully correlated motions. A 0 denotes motions that do not show any correlation. After the C is produced, it has to be diagonalized. This diagonalization transforms a cross-correlation matrix to a matrix (P) that is constructed on set of completely uncorrelated eigenvectors (equation 3.9). After these vectors are orthogonalized and normalized, they are denoted as principal components. The first principal component is the one with the highest eigenvalue (variance), second one has the second highest eigenvalue etc.

$$P = \Lambda C \Lambda^T \tag{3.9}$$

LMI is a method to calculate the above mentioned cross correlation matrix and it was introduced by Lange in 2006 [115]. The conventional methods, such as Pearson[4] coefficient used in this thesis to perform PCA, suffer from the orientational dependence and tendency to underestimate the correlation between two atoms [115]. However, as presented in the work of Lange, it is possible to estimate a correlation as mutual information exchange between two atoms. This mutual information can be understood as a negative entropy and can have values from 0 to 1 for uncorre-

[4]For a definition, limits and comparison to LMI check [115]

lated and fully correlated atoms respectively. The correlation defined in this way is orientation-invariant but captures the non-linear motions that should be neglected. For that reason, LMI is defined as a strict lower boundary to the general MI and does not include non-linear correlations. PCA and LMI have been successfully applied to the biophysical systems as can be seen in [116–121]. Both of these methods are used within Bio3D [122] package of R [123] software. In this thesis, for the purpose of PCA and LMI calculations the last 100ns of a trajectory were used.

3.7 Software

Unit cells of the performed simulations were created using the CHARMM software [77]. Missing amino acid residues in *B. subtilis* SecA were modeled with MODELLER [124–127]. Missing hydrogen atoms were constructed within CHARMM by employing the HBUILD command [77]. **MD simulations** were performed using the NAMD software [100]. **Data analysis** was performed using the VMD [128], PyMol [129], R [123], Bio3D [122] and home-made scripts within these software packages. **Visualization** of the data is done by employing VMD, PyMol, Bio3D, and Corrplot [130]. For the **sequence alignment** Clustal2 was used. [131].

Chapter 4

Results

4.1 MD Trajectories

As it was declared in the introductory part of this thesis the main tool during the quest for a better understanding of processes that underlie SecA's function are MD simulations. During this research, a large set of the MD trajectories was produced. This trajectories were analyzed and results were put in the contest of the relevant biophysical questions related to SecA role during the translocation process. In the Table 2, the complete list of performed simulations is shown and in the following subsections, the setup of each one of them is explained.

4.1.1 Sim1

Our reference simulation in this research is *B.subtilis* SecA with ADP bound to it. The coordinates for this simulation are obtained from X-ray structure with the PDB ID 3JV2 [10]. This X-ray structure is resolved with the resolution of 2.5 angstroms. The SecA structure is obtained in a complex with ADP molecule, Mg^{2+} ion, and in the presence of a short, engineered, peptide which was intent to mimic a signal peptide. This peptide segment was not used in here-reported simulations for two reasons: first, the absence of the force field parameters for this engineered residues and second, and the most important, a suspicion about the proposed binding site which was based on

Table 4.1: Table of the simulations performed in this thesis.

Sim	Organism	Length(ns)	Protein	Nucleotide	SP	Nr. of atoms
Sim1	*B. subtilis*	198	wt	yes	no	199,740
Sim1'	-||-	99	wt	yes	no	-||-
Sim2	-||-	208	wt	no	no	199,921
Sim2'	-||-	100	wt	no	no	-||-
Sim2"	-||-	100	wt	no	no	-||-
Sim2r	-||-	3x20	wt	no	no	-||-
Sim3	-||-	99	D207N	yes	no	199,742
Sim3'	-||-	82	D207N	yes	no	-||-
Sim4	*T. maritima*	120	wt	yes	no	259,677
Sim4'	-||-	100	wt	yes	no	-||-
Sim5	*T. thermophilus*	101	wt	no	no	279,336
SimA	*B. subtilis*	230	wt	yes	no	210.903
SimB	*B. subtilis/E. coli*	245	wt	yes	yes	229.267
SimB'	*B. subtilis/E. coli*	150	wt	yes	yes	-||-

previous knowledge of the SecA-SP interactions [31]. Regarding the missing segments in 3JV2 X-ray structure, there are two stretches of the missing amino-acid residues, one at the N-terminus (14 residues) and one located in the HWD (6 residues). The first missing stretch was ignored because of a probably small impact on the protein dynamics due to the stretch's location (N-terminus of the protein which is located far from locations of interest in this thesis, ADP and SP binding site). The second stretch is actually a gap in the protein sequence which could not be neglected and must be modeled instead. This stretch of missing amino-acid residues was modeled using the MODELLER software [127]. After the structure was completed, heavy atoms, whose positions were unknown from the X-ray structure, were created using the CHARMM software. The missing hydrogen atoms were build also within CHARMM software by implementing the HBUILD command [77]. The X-ray structure contains a number of water molecules among which only those found in radius of 7Å from the ADP molecule (two hydration shells) were kept in order to ensure a proper hydration in the ADP binding region. Water box is created in such a way that it has not less then 20Å between the protein and the boundary. This results in the PBC with dimensions

161Å x 121Å x 101Å and 200,000 atoms and total charge of -14 which was neutralized by randomly choosing the water molecules and replacing them with the sodium ions. After this, system is prepared for the molecular dynamics simulations.

4.1.2 Sim2

The effect of the nucleotide presence on SecA's dynamics and conformation is still under debate [10, 34, 35]. What is known is that ATP binding is sufficient to initiate the translocation process [20]. In order to investigate the effects of the nucleotide on the SecA, the ADP was removed from the *B.subtilis* structure used in the Sim1. Unit cell was build in the same manner as in the Sim1.

4.1.3 Sim3

To explore effects of the D207N (*B. subtilis* numbering) mutation that is shown to abolish hydrolitic activity of the SecA, D207 was alchemically mutated to N207. This was performed by deleting all uncommon atoms between the aspartate and asparagine and replacing the amino-acid residue name in the PDB file as wellas in the sequence card. During loading of the parameters and topology of the structure, the final amino-acid at the position 207 would be N instead of D. The subsequent steps were performed according to the procedure established for Sim1.

4.1.4 Sim4

In recent years, it has been suggested that binding of the nucleotide is in a direct correlation with the orientation of the PBD relative to the HWD [34]. The X-ray structure resolved lately (PDB ID:3JUX) showed that the PBD can sample both, open and closed, conformational states regardless of the nucleotide presence. The X-ray structure of the *T. maritima* is solved with the PBD located in the vicinity of the HWD, in closed conformational state and in presence of the ADP. Simulating this particular protein could shed light on a potential differences between the open and closed conformational state with a special focus on the nucleotide binding site.

With a very few missing amino-acids, at N- and C-terminus, and with the X-ray resolution of 3.1 Å, the procedure of preparing this structure for MD simulations was straightforward. The setup of the unit cell is done in the same manner as in the previous simulations.

4.1.5 Sim5

In order to expand our understanding of the processes in SecA, a MD simulation using the X-ray structure of *T. themophilus* SecA was performed. This crystal structure is solved with the resolution of 2.8 angstroms. The missing amino-acids are concentrated at the C-terminus (40) and were neglected during the setup which was done by the protocol that was established.

4.1.6 SimA and SimB

As it was mentioned in the introductory paragraph, SecA engages into many interactions with various partners, but one of the most important interactions, if not the most important, is the SecA-SP. In order to reveal the effects of the SP binding on the SecA dynamics and check a possible change in allostery, a separate set of simulations without (SimA) and with signal peptide (SimB) was run. Although the SecA-SP complex originating from the *E.coli* [31] was available at the moment, it was of essence to employ the detailed knowledge about *B. subtilis* SecA, gathered and summarized in the first publication that came out of this thesis, to our further research. For that reason, we modeled the SecA-SP complex using the SecA coordinates from the *B. subtilis* (PDB ID:3JV2) and KRR-LamB SP from the *E. coli* SecA-SP complex (PDB ID:2VDA). **The key assumption that allowed combination of SecA and SP originating from a different organism lies in the previously mentioned universal properties of the SP** [62–65]. Indeed, upon analysis of model reported in this thesis was complete, an experimental study was published with exact same idea of combining SecA from *B. subtilis* and KRR-LamB SP [132]. In that work, SecA's affinity for the SP was measured and a position of the

SP binding identified suggesting that SecA binds SP with high affinity in the position predicted by modeling in this thesis.

4.2 General properties

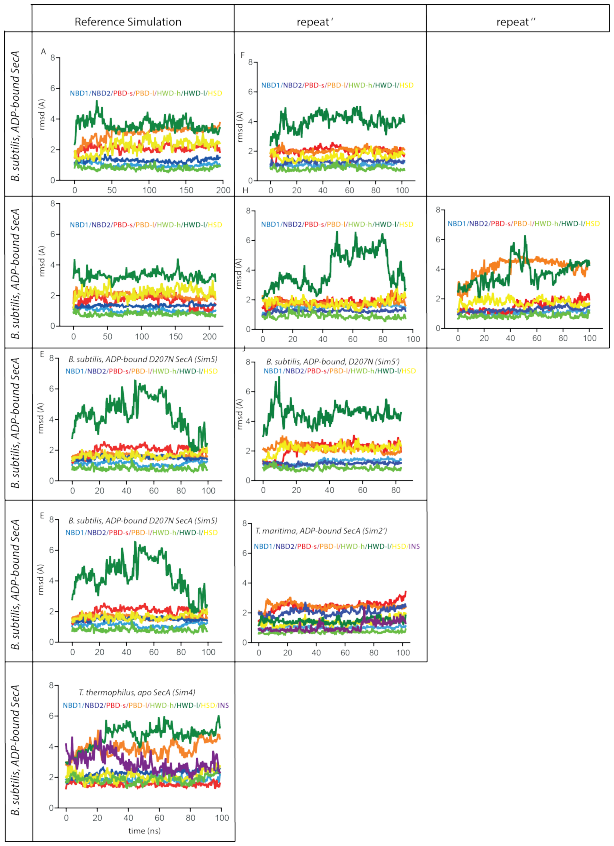


Figure 4-1: RMSD profiles of MD simulations decomposed to SecA's functional domains. It is notable that the functional domains are stable in general, apart from the HWD and PBD loop regions which can exhibit huge conformational changes.

The corner stone of performing the MD simulations is achieving convergence. To test the convergence of the long simulations RMSD(root mean square deviation) time series were recorded where conformation at the current frame is fitted to the initial set of coordinates. The RMSD profiles are good indicators of the protein convergence since they contain information on changes in secondary structure similarity. In the figure 4-1, the RMSD profiles of the simulations decomposed into domains are shown. The decomposition is made in order to dissect the SecA domains that ex-

hibit largest conformational changes from the ones that remain close to the starting coordinates. It is visible that the largest conformational changes belong to the HWD and PBD while the NBD1 and NBD2 conformation remains similar to their initial structure. Analyzing further the behaviour of the HWD's and PBD's RMSD profile, it becomes clear that these domains display the largest changes in conformation due to their long flexible loops. Overall, the RMSD profiles of the common domains in all of the simulations reach their plateau values and e+xhibit stable behavior. In *T. thermophilus* SecA, insertions also show enhanced conformational change which can be seen in Figure 4.1D.

To explore per amino-acid the differences in the flexibility of the various simulations, we calculate RMSF (root mean square flucutation) profiles for the last 30ns of each simulation (Figure 5). On the panels A, D and E of figure 4-2, the RMSF profiles of the SecA from the different organisms are shown. The largest differences are observed in PBD and HWD domains: in all of the simulations of *B. subtilis* SecA, the RMSF of the PBD does not manifest high flexibility while the most flexible domain is HWD. On the contrary in the *T. maritima* and *T. thermophilus* SecA, shown in panels D and E, the most flexible domain is PBD while the HWD has reduced dynamics in comparison to the *B. subtilis* SecA. This effect could be assigned to a different conformational state, open for *B. subtilis* SecA and closed for *T. maritima* and *T. thermophilus*, shown in Figures 4.2B and 4.2C, respectively. It appears that the configuration space available for the amino-acids of the HWD is reduced in the closed conformational state whereas the long loops of the PBD can explore more of the configuration space than in open conformational state.

The comparison of the RMSF profiles calculated for the variations of the *B. subtilis* SecA manifests different behavior only for HWD. Apart from being a good indicator of changes due to some perturbations or different origin, RMSF profiles could serve as a tool to verify how well the protein dynamics is reproduced in the repeat simulations. In Figure 4.2A, it is visible that the RMSF profiles of the repeated versions of the Sim1 and Sim2 are reproduced to a great extent with some small quantitative differences in the HWD for the Sim2, Sim2', and Sim2" simulations.

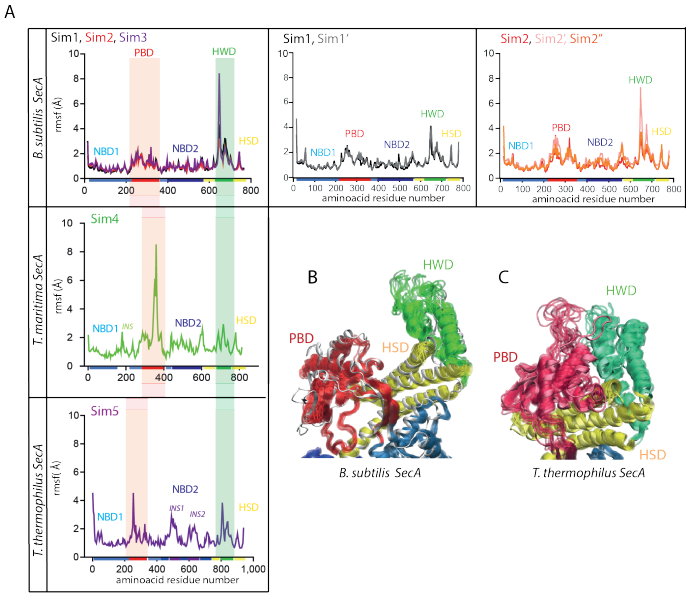


Figure 4-2: Flexibility of SecA domains. (A) Rmsf profiles of various SecA simulations performed in this thesis. The highest differences in flexibility between different organisms are in observed in the PBD and HWD regions. Significance of this differences is confirmed by repeat simulations which resemble original Rmsf profiles. (B-C) Molecular graphics of the 10 equavly spaced snapshots of the *B. subtilis* (open conformational state) and *T. thermophilus* SecA (closed conformational state). The relative position of the PBD has an impact to SecA's flexibility.

4.3 Water Occupancy in Nucleotide Binding Pocket

In the previous paragraph, it was shown that the simulations in this thesis are solid and well converged. Here, the specific properties of the nucleotide binding pocket are discussed. As mentioned in the introductory section, the fuel for the preprotein translocation activity comes from the ATP- binding, hydrolysis, and ADP release. For that reason, catalytic properties of SecA were studied extensively, but the mechanism of the chemo-mechanical coupling remains unclear. Moreover, the precise position of a water molecule essential for the hydrolysis reaction is still under debate. The mutagenesis studies [38] showed that ATP-hydrolysis is abolished when D207N or E208Q mutations were performed on SecA, whereas the X-ray experiments were usually assigning a hydrolytic water molecule to one of these two amino-acid residues. Together, these findings mark highly conserved DEAD motif as a main catalytic factor during the hydrolysis.

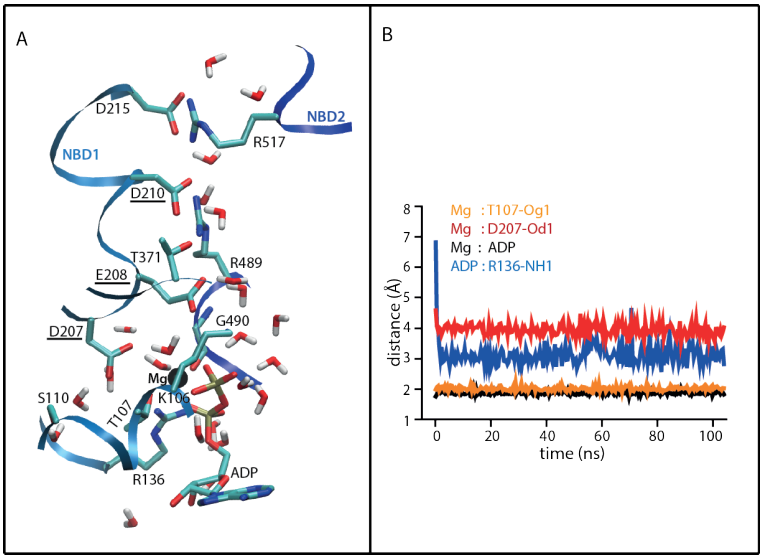


Figure 4-3: ADP behaviour in SecA. (A) Nucleotide binding pocket recorded at the end of the Sim1. The DEAD motif residues are underlined. We observe and extensive interaction scheme among protein, ADP-Mg^{2+} and water.(B) Interactions of the ADP-Mg^{2+} complex with the SecA. We observe that the ligand remains firmly attached to its partners from X-ray structure. The main difference is created by the prompt approach of the R136 to the β-phosphate of the ADP

In figure 4-3A, the snapshot of the *B. subtilis* SecA nucleotide binding pocket is presented. Molecular graphic depicts ADP-SecA interactions along with various

protein-protein and protein-water interactions. The most important ADP-SecA, as well as the Mg^{2+}-SecA, interactions can be seen in Figure 4.3B. It can be concluded that SecA binds firmly the ADP molecule and Mg^{2+} ion since the ADP-protein distances remain close to their X-ray values during the simulation.

If we reconsider the Figure 4-3A with a special focus on the water molecules, we see a complicated picture with no clear positioning of the water, especially in vicinity of the DEAD motif. For this reason, an average water occupancy in the nucleotide binding pocket was calculated by employing the VMD software. This result is used to create iso-occupancy surfaces which visualize the probability for a water molecule to be found in a particular part of space. Figure 4-4A-B displays the this iso-surfaces for

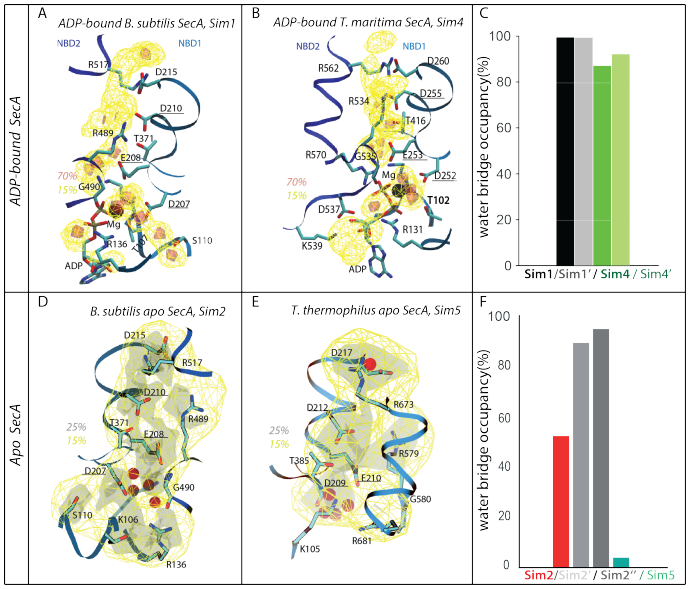


Figure 4-4: Water dynamics in SecA's nucleotide binding pocket. (A-B) Water occupancy iso-surfaces of the ADP-bound Sim1 and Sim4, respectively. A regions of the high occupancy(red surface) in the nucleotide binding pocket are detected. Some of the positions coincide with the x-ray water positions. In both organisms, a presence of water bridge between D207(252) and D208(253) is observed. (C) Histograms of the D-E water bridge in DEAD motif of the ADP bound SecA simulations. In all simulations, water bridge is present more than 80% of the time. (D-E) Water occupancy iso-surfaces of the APO Sim2 and Sim5, respectively. Oppositely to the ADP-bound simulations, the high occupancy regions in the APO simulations are not present. Moreover, the water seems to be equally spread over the binding pocket. (F) Histograms of the D-E water bridge in the DEAD motif. We see that in Sim2 and Sim5, indeed the water bridge frequency is reduced significantly. However, repeats of the Sim2 do not reflect this property but manifest high-frequency water bridge. This suggests that further examination is needed.

ADP-bound and APO SecA simulations. Even from a superficial visual examination, a difference between ADP-bound and APO systems can be drown: In ADP-bound SecA the water molecules occupancy in the nucleotide binding pocket has a discrete nature with regions that are occupied by water molecules for >70% of time. Some of this regions coincide with the positions of crystallographic water but, as a very interesting phenomena, the existence of a high-occupancy cite between D207(D252) and E208(E253) is noted. This water molecule seems to be shared by first two residues of the DEAD motif and stabilized via two H-bonds, forming a water bridge between D207 and E208. In case of *T. maritima* SecA, a formation of a double water bridge is possible. This means that D252 and E253 share two water molecules as observed in Sim4, repated *T. maritima* simulation [1].

On contrary, in APO systems displayed in Figure 4-4D-E no high occupancy regions are visible. The maximum occupancy detected in APO simulations is not higher than 25%. This leads to a conclusion that presence of the nucleotide alters the behavior of water in the nucleotide binding pocket of the SecA. Does this imply the lack of the bridge identified between the D207 and E208 in APO bound simulations? To answer this question a thorough analysis of the D207-E208 water-bridge was performed. The time series of the water bridge that can be formed by one water molecule (*B.subtilis* ADP-bound SecA) or ,in some instances, by two water molecules (*T. maritima* ADP-bound SecA) can be found in Appendix. The water bridge is defined as a simultaneous H-bonding of a two different amino-acid residues to the same water molecule (case of a single water bridge) or a simultaneous bonding of two amino acid residues with two different water molecules which are mutually H-bonded. In Figure 4.4C and 4.4F, the histograms of the D207-E208 water bridge frequencies are shown for ADP-bound and APO-simulations, respectively. The ADP-bound SecA simulations unequivocally show that first two amino-acids of the DEAD motif are bridged by water molecules with frequencies larger than 80%. However, in APO simulations situation is less conclusive. In Sim5, which X-ray structure is solved without the nucleotide, the water bridge is almost undetectable. In *B.subtilis* simulations where

[1] See Appendix.

ADP was removed, the water bridge frequency is significantly reduced only in Sim2, whereas in Sim2' and Sim2" there is no visible change relative to the ADP-bound simulations.

The computational hypothesis about a water molecule(s) that is "shared" between D207 and E208 was confirmed experimentally in work of Chen et al [37]. In this paper, a novel X-ray structure of the *T. maritima* SecA was presented. This structure contains a two water molecules in the almost exact positions predicted by the calculations performed in this thesis.

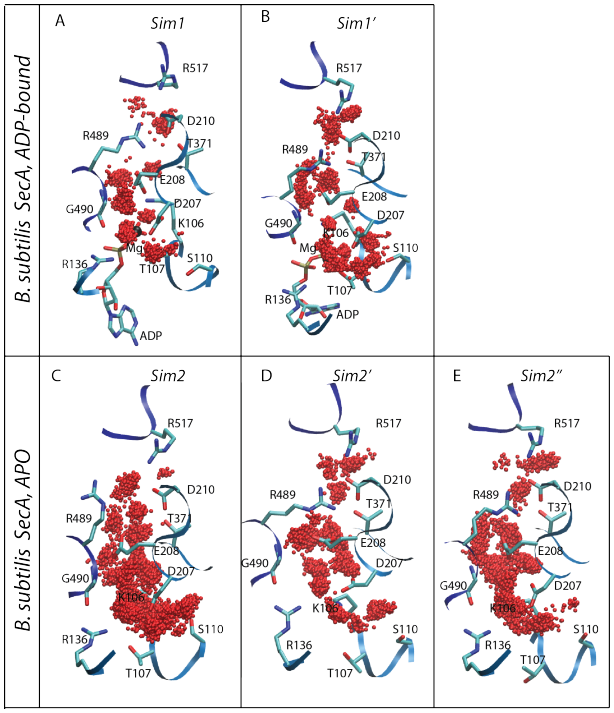


Figure 4-5: Overalap of the 1000 snapshots of the water molecules positions for *B.subtilis* ADP bound SecA (A-B) and apo SecA (C-E). Tt is visible that discrete distribution of preferable water sites observed in ADP bound SecA tends to be fully or partially abandoned in APO SecA. By visual inspection it can be concluded that the behaviour of water in the ADP binding pocket depends on the H-bond interactions achieved between ADP binding pocket amino-acid residues.

In order to enlighten this question, the water molecules positions of 1000 snapshots recorded each 10fs were overlapped. In Figure 4-5, these overlaps are shown for ADP bound *B. subtilis* SecA (A-B) and APO *B. subtilis* (C-E). In both ADP-bound simulations, a clear discretization of possible positions where water molecules can be found is visible. In APO simulations (C-E), this picture is more or less distorted

depending the on the simulation. Closer inspection suggests that origin of this distortion lies in the H-bonding network formed between NBD1 and NBD2 residues, mostly R489 and R517 and DEAD motif residues. When this residues interaction is stable and H-bonding network formed by them is solid, the water molecules can be found only in particular regions displayed as high occupancy sites in figure 4-4. When this is not the case, the water molecules flood the nucleotide binding pocket. Moreover, this finding, for the first time during this research, points towards a possibly great importance of H-bonding networks for the SecA function.

4.3.1 D207N mutation effect

The mutations D207N and E208Q prohibit the SecA's catalytic activity. As mentioned before, it was thoroughly debated which one of these residues stores a water molecule needed for the hydrolitic attack on the γ-phosphate of the ATP. In this thesis, it is proposed that both of the amino-acid residues store the precious water molecule in, what is called, water-bridge. Aiming to examine behaviour of water in case of D207 mutant that is simulated in this thesis, the same set of analysis that is used in previous paragraph was applied. In Figure 4.6A-B, the overlap of water molecules along the trajectories is shown for Sim1 and Sim3, respectively. In the Sim3, the water molecules are spread around N207 and there is no clear discrete positioning as in Sim1. This picture is reflected numerically in Figure 4-6C where the D(N)207-E208 water bridge frequency is shown. The water-bridging is almost completely absent in D207N simulations and this could be a possible reason for the prohibition of ATP-hydrolysis activity. This observation straightens the hypothesis of the water bridging. Indeed, if only one of the D207 and E208 stores the water molecule needed for the hydrolysis, some basal ATP activity in one of the mutants should be detected. Otherwise, one can speculate that the picture seen in Figure 4-6B is analogues to the situation when E208 is mutated and results into water-bridge absence as well.

If the behaviour of D207 and E208 in Sim1 is analyzed and compared to N207 and E208 in Sim3, seems that loss of the water bridge could affect relative orientation of

the DEAD motif residues. In Figure 4-6D, D/N207-Cγ:E208-Cδ distance profiles are shown. This distance tends to have 2Å higher values in D207N mutant (Sim3,Sim3') than in wild type (Sim1, Sim1') simulations.

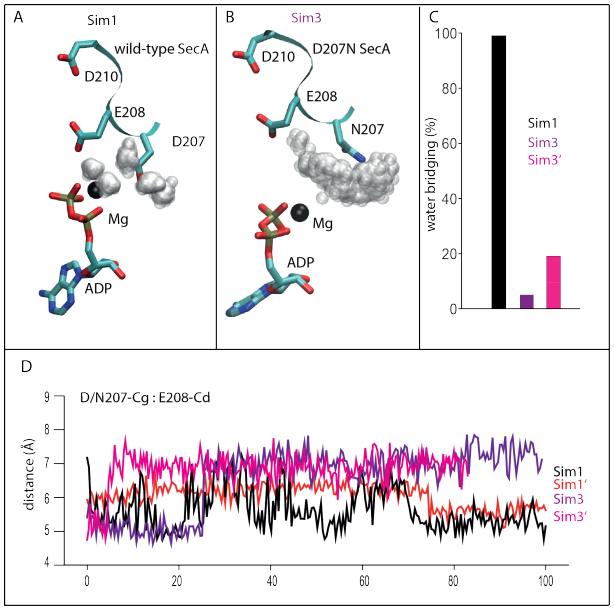


Figure 4-6: Effect of the D207N, the hydrolysis abolishing mutation, on the water dynamics in the ADP binding pocket. (A-B) Water distribution around D(N)207 in Sim1 and Sim3, respectively. (C) D(N)-E water bridge frequency in Sim1 (black) vs Sim3 (violet) and Sim3' (magenta). There is a reduction of the water bridge frequency after mutation. (D) Distance profiles of the D/N207-Cγ:E208-Cδ. A possible reason for the lack of the D-E water bridge lies in different rotameric behaviour of the N207 in comparison with D207 which is manifested through the change in the D/N207-Cγ:E208-Cδ distance.

4.4 H-bonding Maps

As shown in the section 4.3, the water dynamics in the nucleotide binding pocket responds to the removal of the ADP. How big this response is going to be, appears to depend on the change in the extensive H-bonding network. This H-bonding network starts at the nucleotide binding pocket and is formed by the NBD1 and NBD2 residues. In order to quantify possible importance of the H-bonding dynamics, the frequencies of the H-bonds formed between amino-acid's sidechains were recorded. In Figure 4-7, part of the H-bonding dynamics of the Sim1 is show. Only H-bonds with frequencies higher than 50% are represented. This is done by connecting the Cα atoms of the each H-bonding pair by color-coded lines. The amino acid residues, for which was experimentally shown to cause major effects on the SecA function[2], are depicted with blue spheres. As it is visible, all of the important amino-acid residues (explicitly labeled) are participating in H-bonding networks by forming some of the strongest H-bonds observed. This underlines the presumption of the importance that H-bonding dynamics has for the SecA function. The above-mentioned importance, imposed a difficult task of describing and quantifying the complex networks of H-bonds. There are several hundreds of unique H-bonds formed during trajectories in *B.subtilis* SecA and 200 sampled at every time frame. In order to deal with this problem, a novel method for visualization of the H-bonding networks was designed and implemented. This resulted in a publication in which some of the results presented in this thesis are summarized[3] [40]. The method is named H-bonding and it was explained in detail in Methods Chapter. Here the short recapitulation is given. The H-bonding map calculates the frequencies of all of the H-bonds formed during trajectory by the selected amino-acid residues. These frequencies are then displayed as a matrix and color-coded according to their value. The full H-bonding map has dimensions N x N where N is the number of amino acid residues. However, the interesting segments with the residues that form H-bonding networks can be easily extracted. Up to date, many tools in popular software package for calculation of the

[2]See Appendix
[3]See appendix

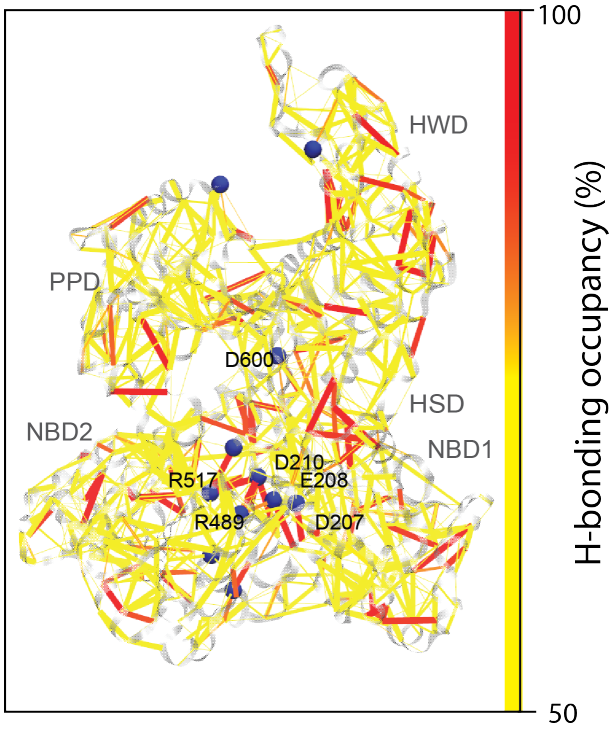


Figure 4-7: The complexity of the H-bonding networks in SecA ATPase. The amino-acid residues that engage into H-bonding interactions are connected with the lines color coded according to the strength of the H-bond. The experimentally important residues are shown as blue spheres. We observe that the residues located in the NBD1-NBD2 interface are creating a firm network of H-bonds.

H-bonds were implemented. However, the visualization of the obtained results was usually focusing on the number of H-bonds achieved between different domains. 2D representation of complete set of H-bonds gives the structural information on where the H-bonds are formed. If they are calculated on a trajectory segment, a color coding helps to understand better the dynamics of H-bonds detected. In Figure 4-8 a static example of the H-bonding map is illustrated. This map is calculated on a snapshot of SecA's trajectory depicting well the α-helical H-bonding networks close to diagonal. The inter-domain H-bonds far from the diagonal are also visible.

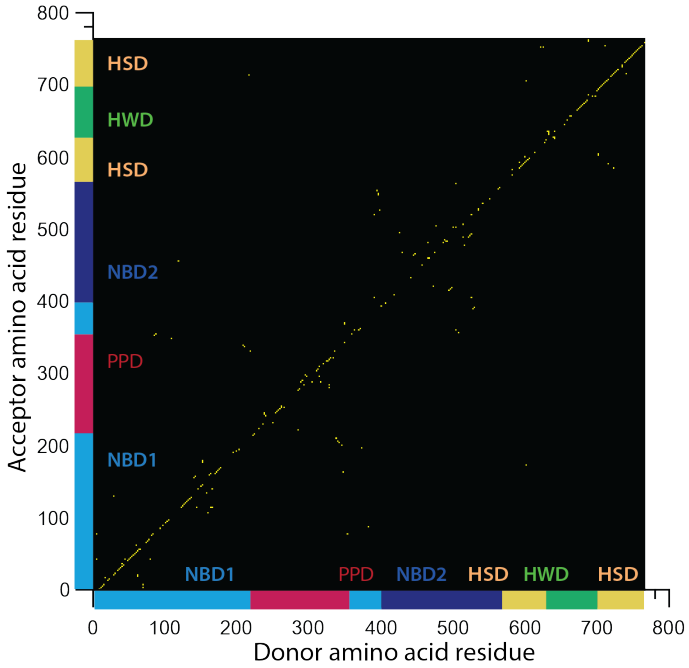


Figure 4-8: An example of H-bonding map calculated for entire protein. This H-bonding map was computed for one single snapshot of *B. subtilis* SecA. Each H-bond is represented as a yellow dot.

The main advantage of the matrix-like representation lies in the comparative analysis

footnoteH-bonding map simplicity and straight-forward implementation helped during analysis of several structures of Channelrhodopsin obtained with homology modeling showing the usefulness of such tool even for bio-informatics investigations [133]. . In the Figure 4-9 the H-bonding network that dominates the nucleotide binding domain is shown. Panels A-B are displaying the Sim1, ADP bound *B.subtilis* SecA

whereas panels C-D are showing Sim2, the *B. subtilis* APO SecA. The change in

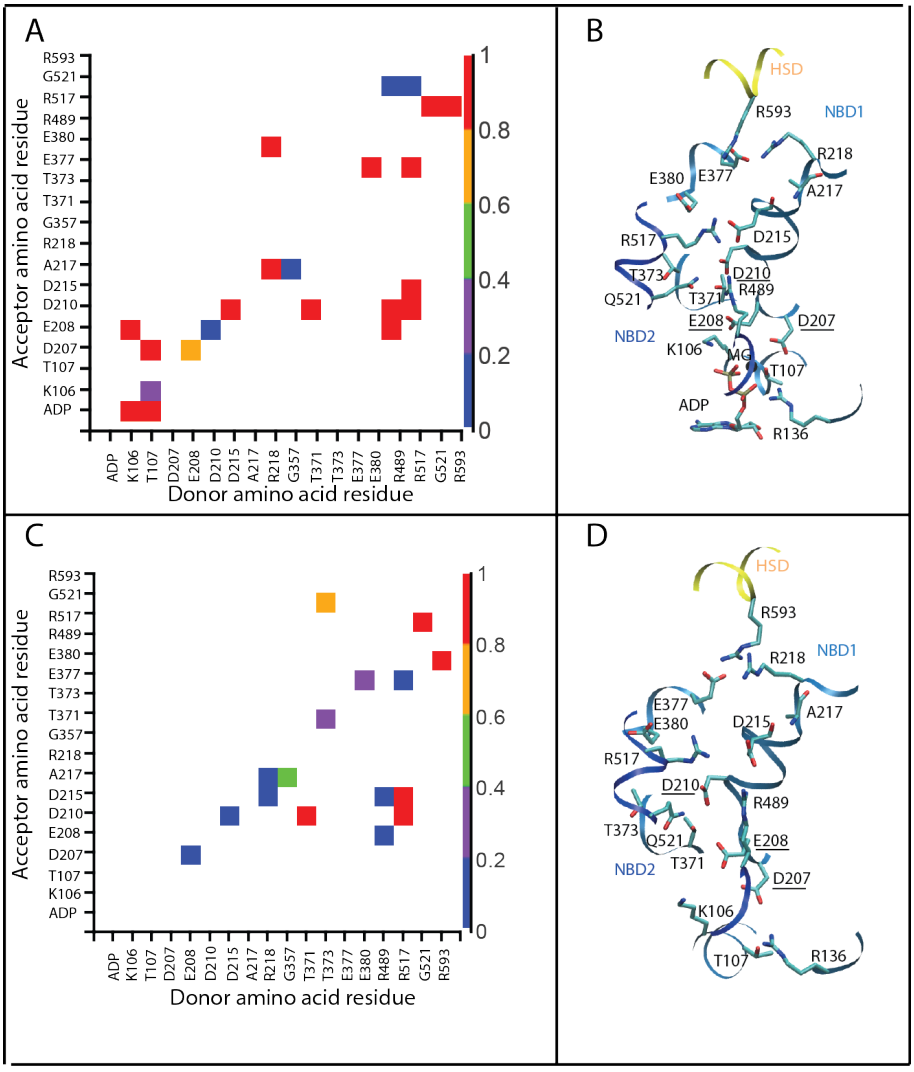


Figure 4-9: A comparison between the ADP and APO SecA's nucleotide binding pocket. (A-B) A hydrogen bonding map and molecular graphics of the ADP bound SecA. The complex network of H-bonds is captured by the H-bond maps. (C-D) A hydrogen bonding map and molecular graphics of the APO SecA's nucleotide binding pocket. The changes induced by the cleavage of the ADP are clearly visible: The H-bonds between K106/T107 and D207/E208 are lost. Also H-bonding cluster has lost several interactions close to HSD and the ones that remain are weaker than in ADP bound SecA.

H-bonding network that can be seen when panels B and D are compared is clearly notable when panels A and B are compared. Moreover the quantification of the change can be easily made: i) the H-bonds that are missing(appear) when ADP is removed can be identified and ii) the possible change in frequencies between the H-bonds sampled in both simulations can be quantified. It can be seen, that indeed,

ADP removal induces the change in H-bonding network formed on the interface of the NBD1 and NBD2. Moreover, these changes could be tracked up to HSD.

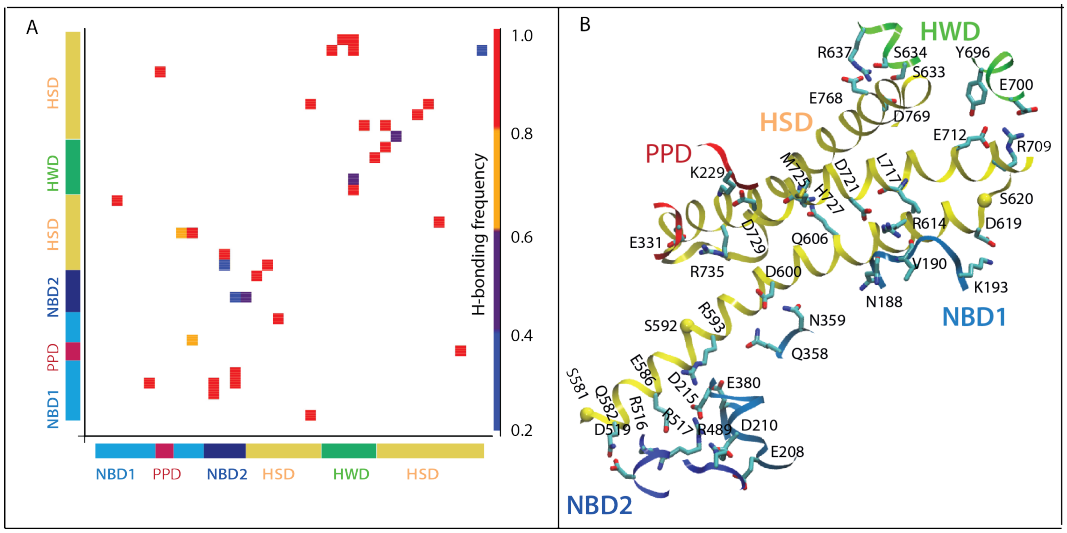


Figure 4-10: The H-bonding map and molecular graphics of the HSD interactions with other domains. It is observable that HSD forms a complex clusters of H-bonding interactions with the multiple domains. These interactions are especially interesting at the junction of NBD1, HSD's long helix, HWD and HSD's IRA1.

Since it is known from X-ray structures that HSD interacts with all of the domains of SecA, full H-bond map is reduced to the region that involves the HSD. In figure 4-10, the H-bonds of the HSD with all the other domains are shown. These H-bonds are grouped into several clusters that form H-bonding network of HSD. The clusters contain mostly the strong H-bonds with frequencies of occurrence above 80% and they are present in SecA from all different organisms studied here. One can make a hypothesis that these H-bonding clusters participate in long-distance coupling. Moreover, these results are reproducible and the repeated computations can be found in Appendix marking clustering of H-bonds on the interfaces of HSD and other domains as general feature of SecA.

4.5 Influence of Signal Peptide on SecA's dynamics

General properties. The interaction of the SecA with the preprotein is of crucial importance for the translocation. As stated in the introductory paragraph, preprotein binding promotes ATP hydrolysis and couples it to the translocation. In order to investigate this interaction and possible influence of the preprotein on SecA's dynamics, the SecA/SP complex was simulated. In Figure 4-11, the basic properties of the SecA/SP simulation are presented. In the panel A, the overlap of the initial structure (SecA-white, SP-pink) and the last 10 nanoseconds (1 snapshot/ns) of the MD simulation is shown. The SP seems to fit well the cleft predicted for its binding, although due to the steric clashes in the beginning it had to be removed outside of the the cleft. Rmsd profiles shown in the panel B indicate that SP is very flexible

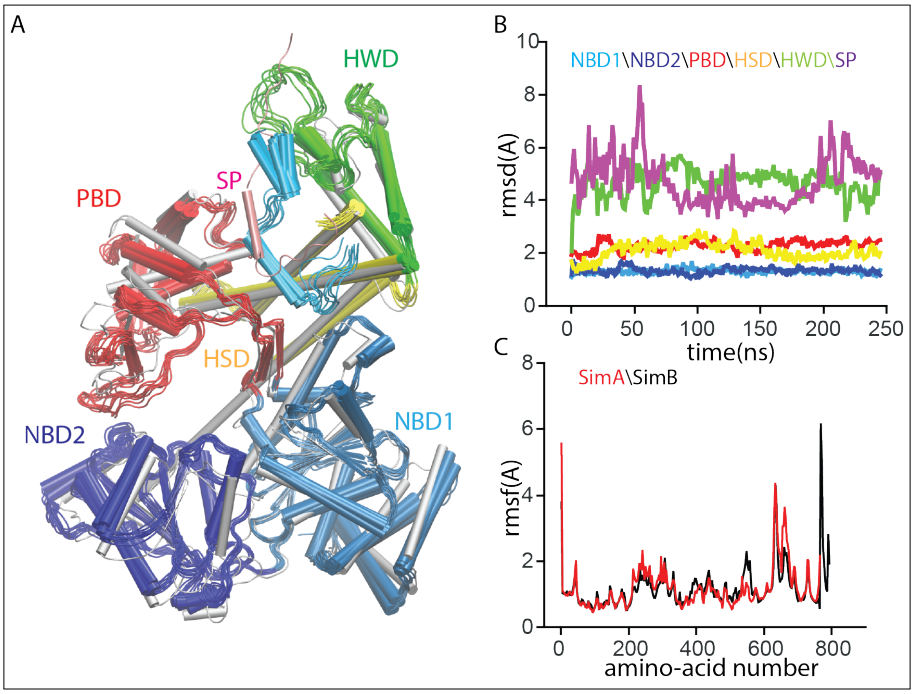


Figure 4-11: Dynamics of B. subtilis SecA in the presence of LamB SP. (A) Overlap of with 10 equally spaced snapshots from the last 10ns of SimB with the starting crystal structure of SecA (white). (B) Cα-rmsd profile computed from SimB for the various protein domains domains. (C) Rmsf profiles computed from SimA (red) and SimB (black). The SP contribution is visible

since the large amplitudes are observed during the trajectory. The panel C gives the rmsf comparison between SimA and SimB which suggests that the SP presence affects

the dynamics of the HWD and HSD. Indeed, when SP is present (black curve), the fluctuations of the HWD backbone are lowered.

Since the dynamics of the SP seems to induce a change to the SecA's dynamics, the behaviour of the SP should be investigated in a detailed way. In Figure 4-12, the behaviour of the SP in SimB was dissected. Rmsd profiles of the N-terminus, core and C-terminus show that most of the SP's dynamics comes from its termini. Analysis of the secondary structure elements shown in the panel B suggests that the SP indeed exhibits a behaviour typical for its well known tripartite structure. The hydrophobic core forms a short but stable α-helical structure as it is suggested in [31]. A shorter

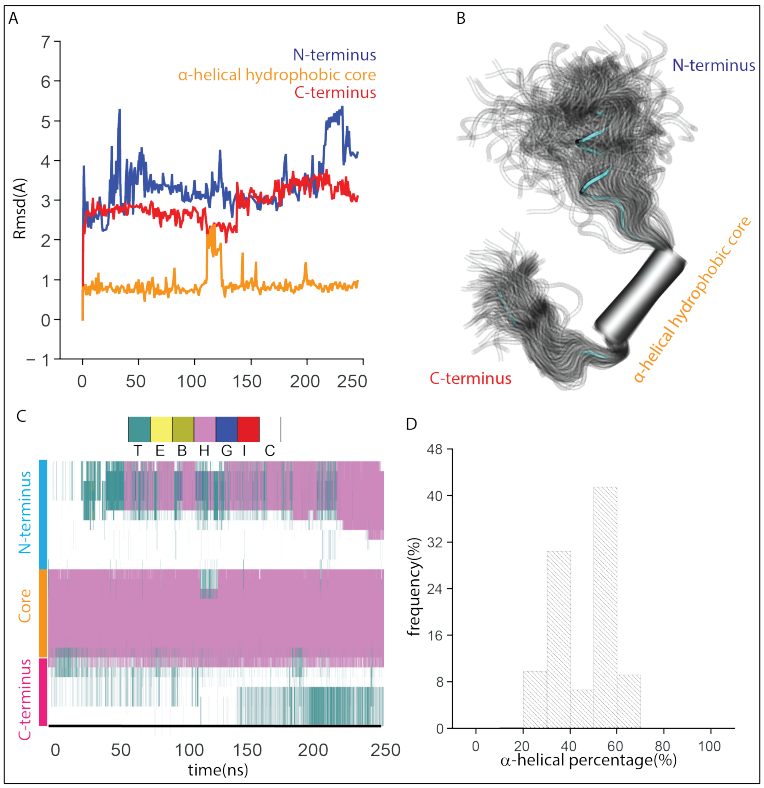


Figure 4-12: Dynamics of the LamB SP bound to B. subtilis SecA. (A) Cα rmsd profile of the SP computed from Sim2. (B) Overlap of 500 equally spaced structure snapshots of the SP taken from the last 100ns of SimB. Note that the N- and C-terminal segments are very dynamic, whereas the hydrophobic segment has a relatively small conformational dynamics and preserves a α-helical structure. (C) Secondary structure analysis of the SP from SimB. The analysis was performed using STRIDE [134] in VMD [128]. The structural elements are denoted as T- turn, E- extended, B- isolated bridge, H- helix, G- 3-10 helix, I- π-helix (D) Histogram showing the number of coordinate frames with certain percentage of α-helical structure of the SP.

and not very stable α-helical segment is observed also in the N-terminus of the SP and its origins are in the interactions in which the N-terminus engages with the SecA. This pattern of behaviour is observed in repeated version of SimB[4], SimB'.

PCA. As mentioned in the Methods chapter, PCA is a very convenient tool for reducing the complex dynamics of a protein to its essential part and for comparison between various setups preferably of the same protein. In this study, the main interest is on capturing the difference in SecA's dynamics caused by the SP presence. Therefore, the PCA was performed on the last 100ns of the SimA, SimB and SimB'. In Figure 4-13, the proportion of variance consumed by the first 20 principal components is shown. In all three simulations, the first 20 components consume 80% of the total motion. Moreover, the first three components take up to 50% of the total

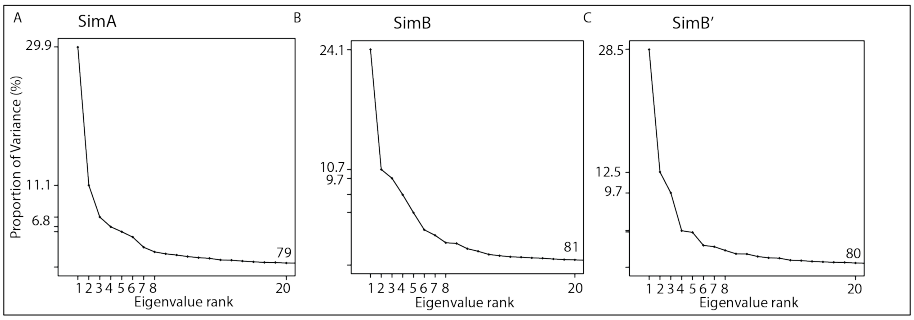


Figure 4-13: Eigenvalue ranks for the first 20 PCA modes of SecA from simulations without and with SP. (A-C) Percentage of contributions from the first 20 PCA modes to the total variance, as computed from simulations without SP (SimA, panel A) and with SP (SimB,SimB', panel B). Note that, for each Sim, the first three PCs contribute with 45%-51% of the total variance.

variance. Since the percentage of the variance rapidly drops with the principal component number, the complete motion of the SecA is limited to the essential subspace of 10 PCs. With the essential subspace set, one can project the protein's motion on the planes formed by two arbitrary chosen PCs. In the Figure 4-14, the projections on the planes formed by the largest (first) three PCs are given. The projections are color coded according to the timestep and they can give the hint about the convergence of a particular PC on a time scale that was accessed by the simulations. Figure 4-14 suggests that PC1 and PC3 are converged enough for being examined in greater

[4]See Appendix

detail. However, PC2 in SimA and SimB behaves as a drift and it represents motions that are not accessible on the time scale of the simulations performed. This is also confirmed by calculating the auto-correlation function as shown in dix.

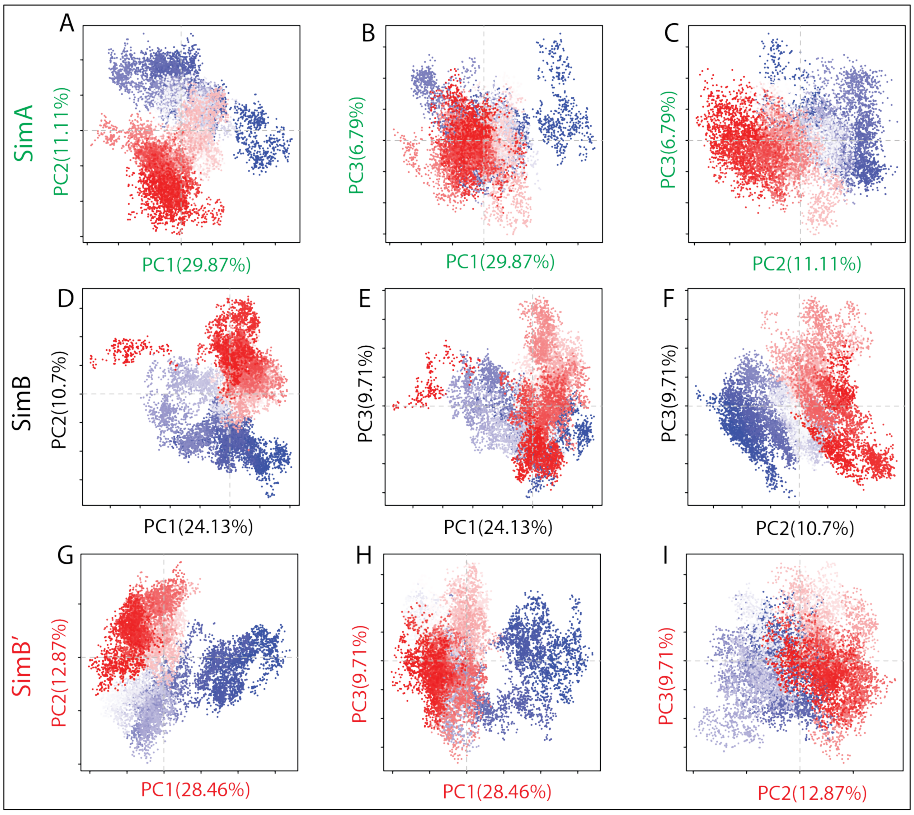


Figure 4-14: Evaluating convergence of PCA analyses in SimA-SimB' by means of projections between selected PC modes computed from SimA, SimB and SimB'. For each mode, we indicate the percentage of total variance as illustrated in Figure S3. (A-C) Projection of protein motions on the PC1/PC2 plane (panel A), on the PC1/PC3 plane (panel B), and on the PC2/PC3 plane (panel C) as computed from SimB. (D-F) Projection of protein motions on the PC1/PC2 plane (panel D), the PC1/PC3 plane (panel E), and on the PC2/PC3 plane (panel F) as computed from SimB'. (G-I) Projection of protein motions on the PC1/PC2 plane (panel G), the PC1/PC3 plane (panel H), and on the PC2/PC3 plane (panel I) as computed from SimA. In SimA and SimB, projections on the PC2 axis indicate a drift. Absence of such a drift in SimB'is likely due to the shorter simulation time sampled.

Although the motion represented by the PC2 cannot be properly addressed, the motions represented by the PC1 are more than indicative. By its definition, final result of PCA is set of mutually uncorrelated vectors, PCs. The amino acid residues that contribute to a single vector could be interpreted as being correlated. In 4-15 the parts of the protein that contribute the most to the largest PC, PC1, are shown. First

sight on the figure indicates that when SP is present, **the long distance coupling between HWD, SP and NBD2 occurs.**. However, this is not true for SimA, where the PC1 has significant contributions only from the HWD residues.

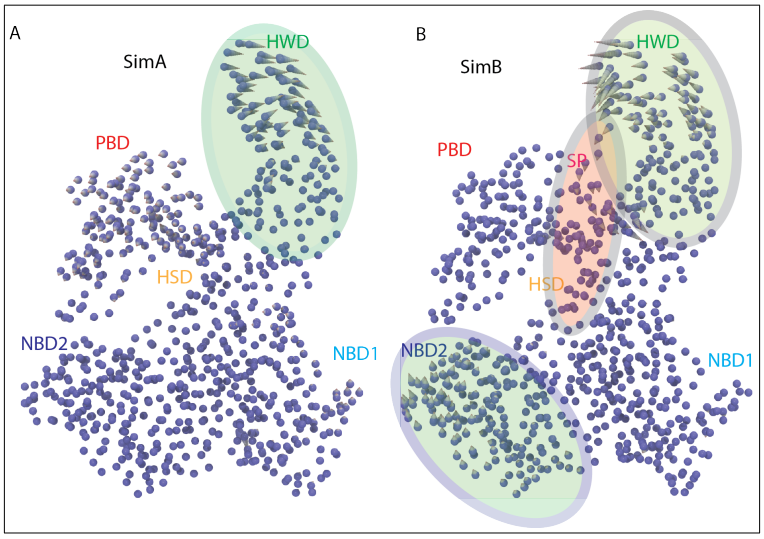


Figure 4-15: Essential motions of SecA. (A) Molecular graphics illustrating motions along PCA1 computed from simulations without the SP (SimA). Each sphere represents the Cα atom of an amino acid residue. Arrows indicate the direction of the motion; the length of the arrow indicates the amplitude of the atomic displacement. HWD groups have the largest-amplitude displacements in the first principal component. (B) Molecular graphics illustrating motions along PC1 in SimB. The largest amplitude motions are observed for the HWD, NBD2 and SP.

LMI. The long-distance coupling initiated by the presence of the SP seems promising and confirms what was proposed by previous experimental studies [21,22] but has to be confirmed and explained in detail. For that reason, LMI was employed as a more general and reliable method for estimating the correlation than standard Pearson coefficient used for calculating PC. In Figure 4-16 the significant (≥ 0.6) inter-domain LMI values in SimA are shown in a molecular graphics as a red lines that connect Cα atoms. The topology of these correlations resembles a picture the was shown to illustrate HSD's H-bonding interconnections with other domains. Indeed, when the H-bonding map is calculated and relevant molecular graphic drawn it becomes visible that high inter-domain LMI couplings match the inter-domain H-bonding clusters. The data presented in the previous figure suggest that H-bonding map can predict the inter-domain junctions that are highly correlated, and more importantly that

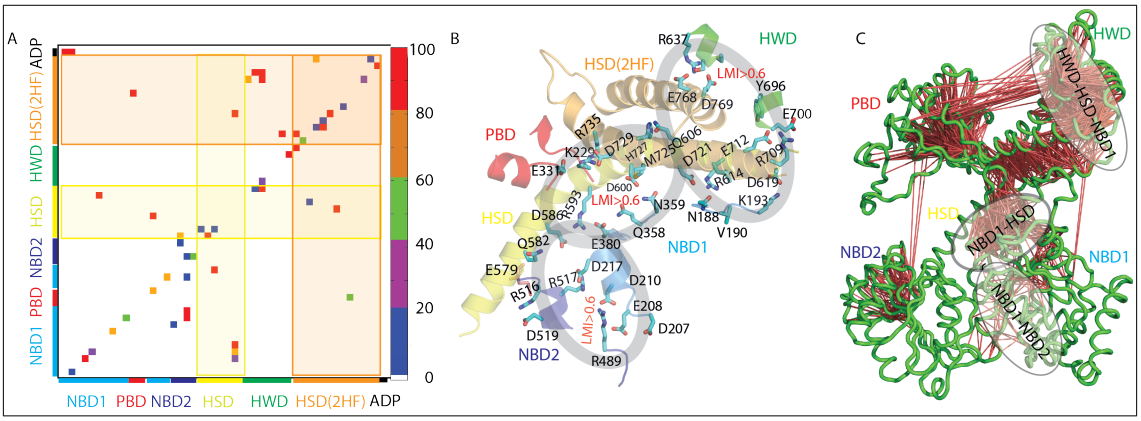


Figure 4-16: Inter-domain coupling of ADP-bound SecA without SP (SimA). (A-B) The HSD interacts with all other domains of SecA. (A) H-bond map identifies inter-domain H bonds that are present during the last 30 ns of the simulation. (B) Molecular graphics illustrating specific interactions between the HSD and other domains of ADP-bound SecA in the absence of the SP. Regions that are highly correlated are marked with gray. (C) LMI analysis of ADP-bound SecA. Red lines connect Cα atoms of amino acid residues with LMI values >0.6 for inter-domain interactions. There are high correlations of the HSD and other domains are located on places predicted by the H-bond map.

H-bonds can serve as a transducers of the correlation.

When it comes to the SimB (SecA/SP), the H-bonding map suggests that SP "plugs in" on the already established grid of H-bonds as it can be seen in Figure 4-17. SP interacts with the SecA via H-bonds at two interaction sites, IS1 and IS2. Both IS1 and IS2 are detected previously in SecA simulations without SP[5] [40] . IS1 consists of HSD/HWD residues, while IS2 involves HSD, HWD and NBD1. From these two sites, a path can be observed. This path leads trough the HSD's networks of H-bonds with other domains trough the GATE1 [21] all the way to the nucleotide binding pocket.

Above mentioned "plugging" may alter the already established correlation in the SecA. Indeed, in Figure 4-18, the molecular graphic with significant LMI values shows that a distant coupling between NBD2 and HWD is established. SP becomes a "center" of correlation and tightly correlates with the HWD, HSD and PBD. Panel B shows the ΔLMI for SimB and SimA from which can be concluded that NBD2(IRA2) couples to other parts of the SecA more intensively. This long distance coupling con-

[5]Notation IS1 and IS2 is introduced in this thesis.

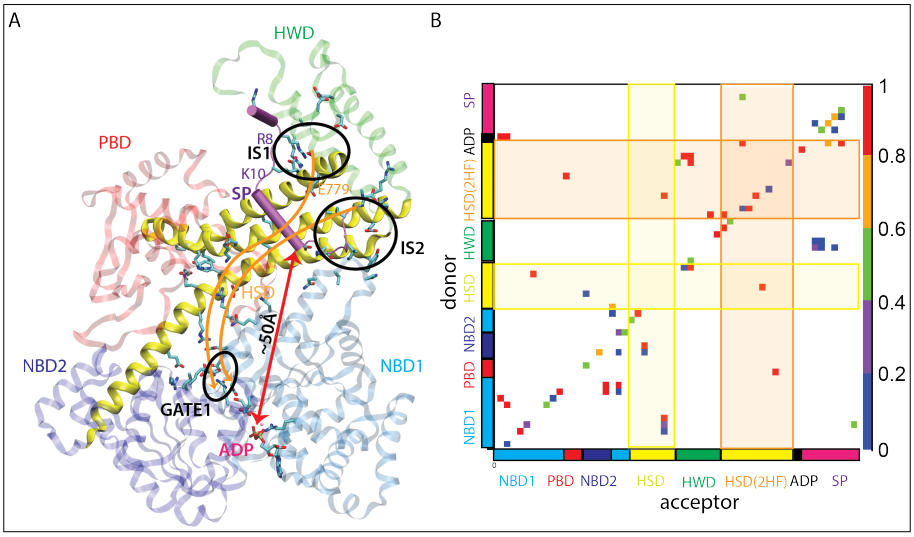


Figure 4-17: H bonding between SecA and the SP. (A) Molecular graphics illustrating H-bonds between the HSD and other SecA domains, and between the SP and SecA. Sites of SP-SecA H-bonding are denoted as IS1 and IS2. Gate1 stands for salt-bridge D215-R517, proposed as a SP's mechanism of control of SecA's catalytic site [21]. Orange arrows mark possible way of signaling from the SP binding site to the nucleotide binding site trough the GATE1. This signaling would be achieved via networks of H-bonds formed by HSD and other domains. (B) Two-dimensional H-bond corresponding to interactions depicted in panel A. Note that the SP H-bonds to groups involved in H-bond clusters involving the HSD,HWD and NBD1.

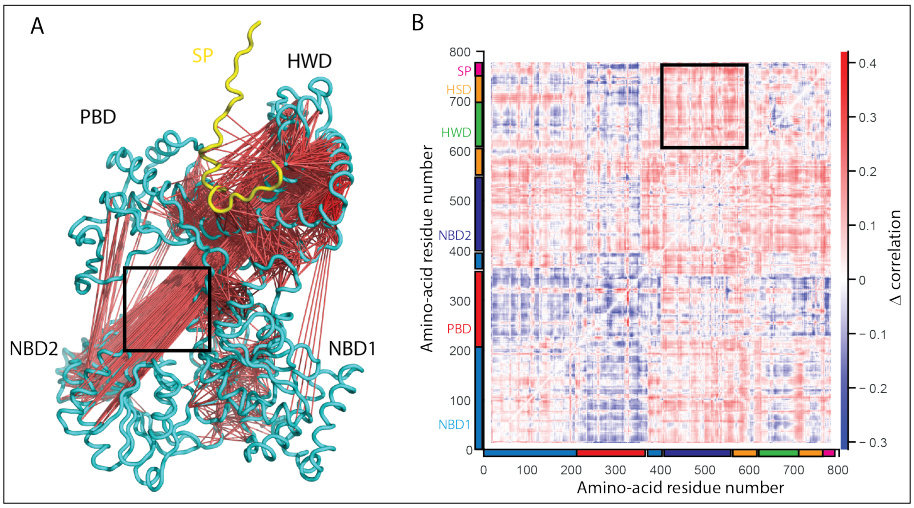


Figure 4-18: The SP alters inter-domain coupling in SecA. (A)Molecular graphics of SecA with bound SP (SimB) shown as ribbons; correlations with LMI values greater than 0.6 are shown as red lines connecting Cα atoms. (B)LMI difference matrix of SimB relative to SimA illustrating regions where change in correlation occurs. We consider that changes in correlated motions are significant when the difference LMI has an absolute value of at least 0.3. We observe significantly more correlations between the HWD and NBD2, and decreased correlations between the PBD and other SecA domains. The region delineated by the black square indicates the difference between HWD-NBD2 in SimA vs. SimB. This corresponding black square in the panel A shows the strong correlations that emerge after the SP binding.

firms the observations from the PCA which suggested that motions of the HWD and NBD2 are coupled. However, the mechanism of this coupling still has to be explained. In order to find the explanation of this coupling, inter-domain and intra-domain H-bonds formed among polar groups were investigated. Reason for that lies in fact that H-bonding clusters identified so far are mostly formed by polar residues, especially ones with long sidechains, such are Arg, Lys, Glu, Asp etc. Figure 4-19 shows the number of unique and average number of the inter-domain and intra-domain H-bonds. Upon SP binding, NBD2-PBD, NBD2-HSD and HWD-HSD numbers undergo the largest changes. NBD2 and PBD sample less unique, mutual, H-bonds although the average number of H-bonds sampled between them does not change significantly. This suggests that interactions between these two domains become less dynamics upon SP binding. Simliar to NBD2/PBD, HWD-HSD loose some of their unique contacts, while the average number of H-bonds, sampled among these domains, remains largely the same. Opposite holds for the NBD2-HSD interactions: The number

of the unique H-bonds sampled during simulation rises, as well as the average number of H-bonds formed between these two domains. These changes cumulatively could lead to the long distance coupling established between NBD2 and HWD, where HSD again appears as a mediator between different domains. Finally, these findings can be

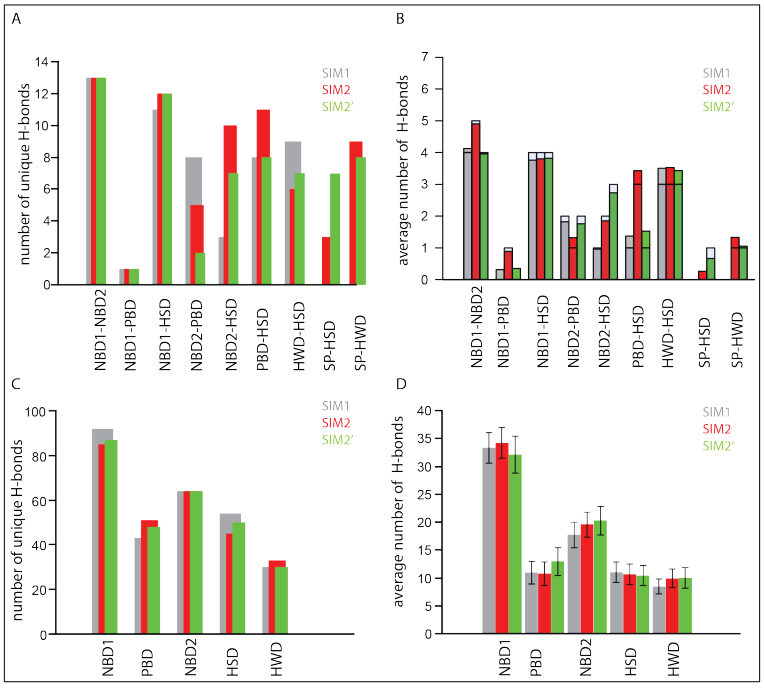


Figure 4-19: Inter- and intra-domain H-bonds in SimA(grey), SimB(red) and SimB'(green). The unique(A,C) and average(B,D) numbers are calculated to give an overview of H-bonding dynamics. NBD2-PBD number of unique H-bonds (A) is reduced in SecA-SP simulations (SimB,SimB') although respective average number doesn't change much (B). The biggest change occurs in NBD2-HSD interactions which become more intense and more dynamic.

summarized into following: Binding of SP indeed alters the communication between various SecA domains. Upon binding, HWD and NBD2 (which is on of the domains that regulate catalytic activity of SecA) establish a significant amount of correlation ($>$0.6) among them. Since these two domains are located on the opposite sites of SecA and do not have any direct contact, this could be a proof of long-distance coupling initiated after SP is bound to SecA. Mechanism that lies behind this coupling, at least partially, is set of H-bonding networks formed between different domains of SecA and HSD. These networks display change in their dynamics after the SP engages into H-bonding interactions with some of those, already established networks.

Chapter 5

Conclusion

The question of all questions that governed research in this thesis was a mechanism of chemo-mechanical coupling, or how chemical energy obtained during the ATP hydrolysis is used by the SecA for the mechanical work, which is conducted far from the catalytic site. With that idea, multiple computational techniques were applied on the SecA-ATPase motor protein. The main tool in this examination are MD simulations, which were used systematically to create a wide set of the SecA trajectories under various conditions. During the investigation of the nucleotide binding site, it was shown that the catalytic water molecule that is supposed to be used during the ATP hydrolysis is most likely "stored" in a form of a water-bridge between first two residues of the DEAD motif. This phenomenon was reported in 2015. In a novel X-ray structure [37] of *T. maritima* SecA, the D252-E253 water bridge was observed which confirmed the result of the research performed in this thesis 5-1. Also, the effect of D207N mutation that abolishes the ATPasal activity in the SecA was dissected. It was shown that this mutation changes the water dynamics in SecA's nucleotide binding pocket preventing the stable interaction of water molecules with DEAD motif residues observed in the wild type ADP-bound simulations.

As a novel tool for analysis of dynamics and visualization of complex H-bonding networks in proteins, the H-bonding maps were introduced. The H-bonding maps pointed out the HSD as a mediator domain, which engages in the formation of complex networks of the H-bonds with other SecA's domains. The H-bonding maps were

employed to track changes in SecA dynamics caused by ADP removal. This comparative analysis showed the disruption in the nucelotide binding site's H-bonding network upon ADP removal.

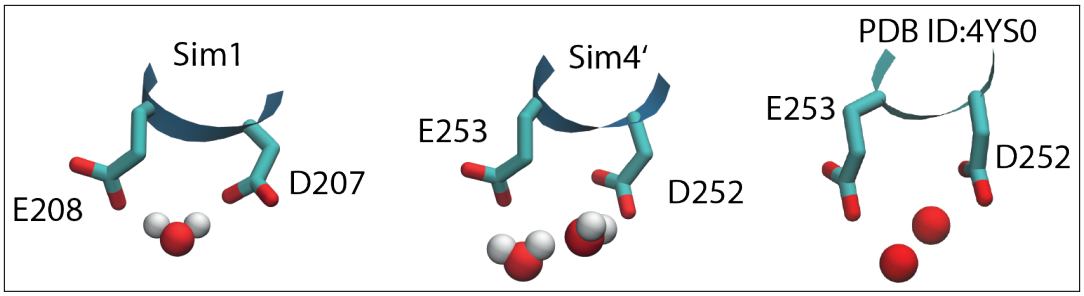


Figure 5-1: Comparison of computational predictions of water positioning in the nucletide bidning site with the experimentally observed water positions. (A-B) Water bridges observed in Sim1(*B. subtilis*) and Sim4(*T. maritima*), respectively. (C) Water bridge observed in X-ray structure (PDB ID:4YSP0) [37]. The water bridge formed by two water molecules detected in simulations is fatefully reproduced in X-ray structure.

Further, the effect of the SP binding to SecA was examined with a great care. The PCA and LMI were applied hoping to track the effects of this binding. This resulted in detection of the long distance coupling established between the NBD2, the HWD and the SP. Additionally, these results, combined with the previously installed H-bonding map tool, showed that the mechanism for such coupling seems to have its origin in the change of the inter-domain H-bonding networks to which SP attaches upon the binding. Worth mentioning is that the modeling performed in this thesis was once more experimentally confirmed when Zhang et al [132] created a functional chimera of *B. subtilis* SecA and KRR-LamB from *E. coli* in a similar manner as it was done here. The results of this part of the research are summarized in the second manuscript which should be submitted.

In the years to come, there is much about SecA and secretion in bacteria that has to be addressed. How several large proteins are coupled together to a single catalytic cite located in the SecA is a mystery. However, with an appropriate approach, this question can be answered. In my opinion, the question of chemo-mechanical coupling should be divided in two parts: *i)* How SecA deciphers various interactions established with SecB, preprotein, SecYEG, nuclotides etc. ? and *ii)* What are the free energies

required for the functional motion of particular domains that could be involved in the translocation process such as PBD, HSD, HWD? Theoretical studies like the one conducted in this thesis could help in both parts of this complicated problem. Moreover, I believe that the work done in this thesis sets a very solid ground for further investigations of long-distance couplings by means of allostery, which is the essence of the first part of the problem. The computational studies can also help addressing the second part of the problem by employing enhanced sampling techniques [37]. Together, this computational methods can represent a glue that could bind different, and sometimes contradictory, experimental results.

Finally, I think that, very soon, the secretion across the cytoplasmic membrane is going to be simulated from the ribosome to the preproteins final destination outside the bacteria with ease, giving us chance to truly understand wonderful nature of this process.

Appendix A

Publications

The work performed in this thesis is condensed in two manuscripts from which one is already published [40] and second one is at the time of writing in the submission stage[1] [135]. The first manuscript summarize water dynamics analysis and role of H-bonding networks on SecA dynamics. This manuscript introduces the concept of 2D H-bonding maps, a tool for analysis and visualization of the H-bonds in systems where H-bonds play a significant role.

The concept of H-bonding maps was applied successfully in the analysis of H-bonding patterns of various homology models of channelrhodopsin in work of del Val et al [133] where maps were used to detect and quantify changes in H-bonding in homology different models.

The second manuscript focuses mainly on effects of SP binding on SecA dynamics. The MD simulations reveal that the long distance coupling is established in the SecA upon binding of the SP. This coupling is detected by application of the PCA and LMI. The mechanism of this long distance coupling was, at least partially, revealed by using already developed H-bonding maps tool. This showed that, the coupling is established through changes in H-bonding dynamics on interfaces of SecA domains, NBD2/HSD and HSD/HWD pointing the HSD as an interaction transducer.

[1]The manuscript was accepted on the day of the defense and it is included in the publishing version of the thesis

Mechanism of conformational coupling in SecA:

Key role of hydrogen-bonding networks and water interactions

Stefan Milenkovic, Ana-Nicoleta Bondar

Biochimica et Biophysica Acta (BBA) - Biomembranes

Volume 1858, Issue 2, February 2016, Pages 374-385

https://doi.org/10.1016/j.bbamem.2015.11.010

Highlights

- We study the conformational dynamics of SecA monomers in water.
- We use hydrogen-bond maps to identify inter-domain interactions.
- SecA has dynamic H-bond networks.
- In ADP-bound SecA, water bridges carboxylate groups of the DEAD motif.
- A mutant with defective functioning has altered water dynamics

Motions of the SecA protein motor bound to signal peptide: Insights from molecular dynamics simulations

Stefan Milenkovic, Ana-Nicoleta Bondar

Biochimica et Biophysica Acta (BBA) - Biomembranes

Volume 1860, Issue 2, February 2018, Pages 416-427

https://doi.org/10.1016/j.bbamem.2017.11.004

Highlights

- We use simulations to derive a model of B. subtilis SecA bound to signal peptide.
- Two-dimensional hydrogen-bond maps identify inter-domain hydrogen bonds.
- Groups at an inter-domain interface engage in dynamic hydrogen bonds.
- Presence of the signal peptide associates with perturbed hydrogen bonds.
- The signal peptide influences domain motions and inter-domain correlations.

Appendix B

Relevant mutations

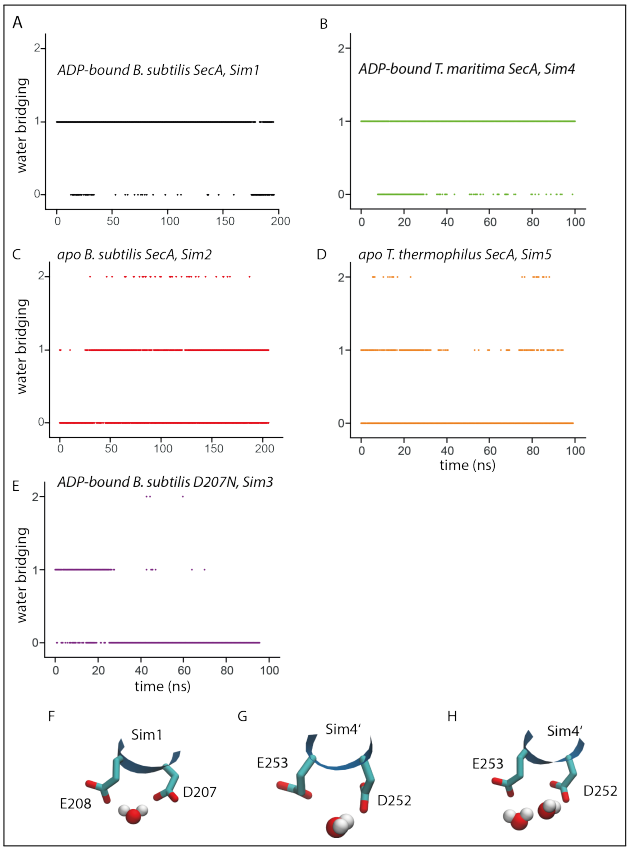
Table B.1: Illustration of selected site-directed mutagenesis effects. Unless specified with bs prefix, the mutations were studied in *E. coli* SecA. The '*B. subtilis* #'indicates the *B. subtilis* group that corresponds to that mutated in *E. coli*.

Mutation	***B. subtilis*#**	**Effect observed**
$_{bs}$K106N	K106	Binds ATP; ATPase activity reduced [136, 137]
K108R	K106	Defective ATPase activity [38]
T109N	T107	Increased ATPase activity [38]
D133N	E131	Reduced ATPase activity [138]
D133A	-||-	Does not affect ATPase activity [139]
D209N	D207	Inhibits ATP hydrolysis [38, 58, 140] Inhibits coupling to SecY [140];
E210Q	E208	Inhibits ATPase activity [38, 140]
E210D	-||-	Reduced ATP hydrolysis;Slow ADP release [139]
D217A	D215	Inhibits translocation of proOmpA [22]
Y326A/R/E	Y306	Defective release of preprotein [141]
E400R/R642E	E380/R593	Impairs translocation activity [44]
R509K	R489	Defective ATPase activity [38, 39]
-||-	-||-	Inhibits ATP binding [58]
R509Q	-||-	Defective ATPase activity [38]
T511N	T491	Increased ATPase activity [38]
D649	D600	Higher basal ATPase activity [142]
R656A	R607	Reduced affinity for ADP binding [143]
R566A	R517	Inhibits translocation of proOmpA [22]
R577K	R528	Defective ATPase activity [39]
R656C	R607	Reduced ADP binding at 4° C [143]

Appendix C

Data analysis and convergence tests

C.1 Water bridge time series



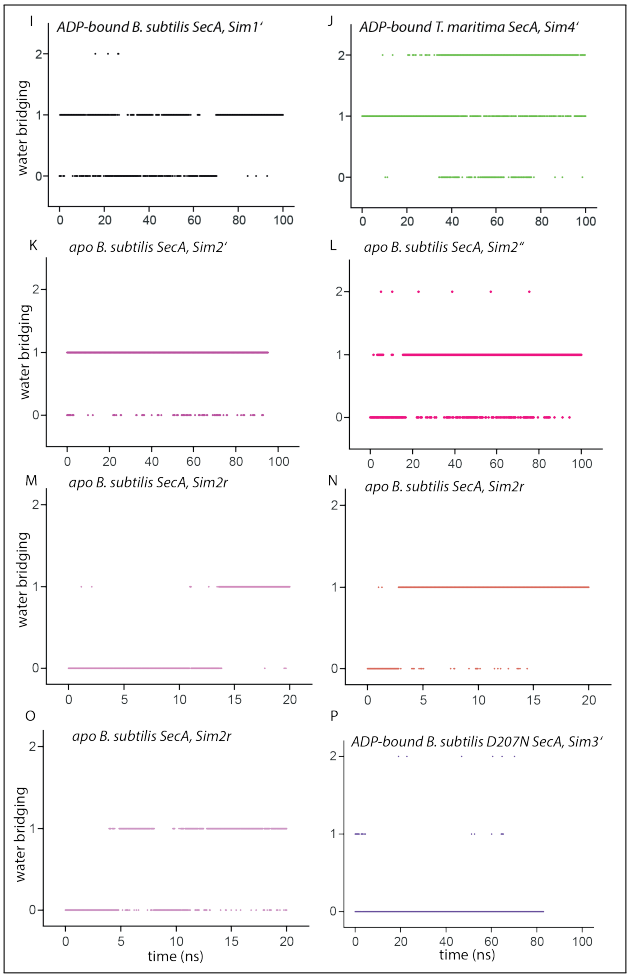


Figure C-1: Dynamics of the water-bridged structure of DEAD groups. On the vertical axis, values of 1 or 2 indicate water bridging via one or two water molecules, respectively; a value of zero indicates that the condition of bridging via one or two waters is not met.(A) B. subtilis ADP-bound SecA, Sim1; the data are the same as in Figure 5C in the main text. (B) T. maritima ADP-bound SecA, Sim4. (C&D) Apo SecA from B. subtilis (Sim2, panel C) and T. thermophilus (Sim5, panel D). (E) D207N mutant of B. subtilis ADP-bound SecA (Sim3). (F-H) Molecular graphics illustrating interaction geometries of water with the first two carboxylates of the DEAD group in ADP-bound SecA from B. subtilis (Sim1, panel F) and T. maritima (Sim4, panels G&H). (I-P) Water bridging monitored from repeat simulations on ADP-bound SecA from B. subtilis (Sim1', panel I), T. maritima (Sim4', panel J), apo B. subtilis SecA (Sim2', Sim2'', and Sim3r: panels K-O, respectively) and D207N mutant of B. subtilis SecA (Sim3', panel P) .

C.2 H-bonding maps

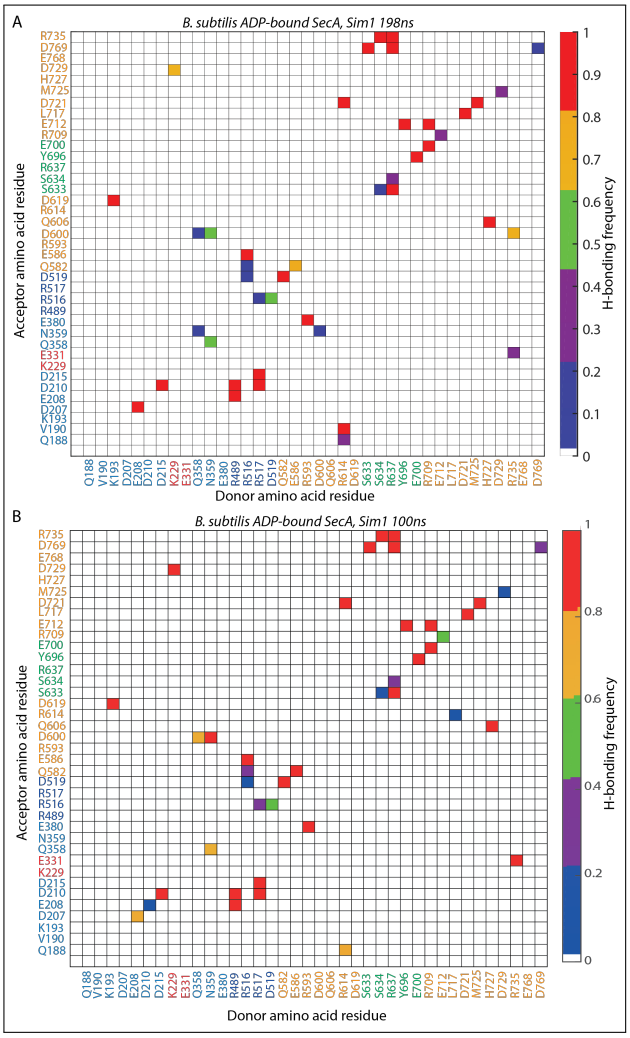


Figure C-2: Selected H bonds in B. subtilis ADP-bound SecA. (A-B) H bonds of selected interactions involving the HSD. (A) H-bond map computed from the last 30ns of Sim1. (B) H-bond map computed from Sim1 during the time interval 70-100ns.

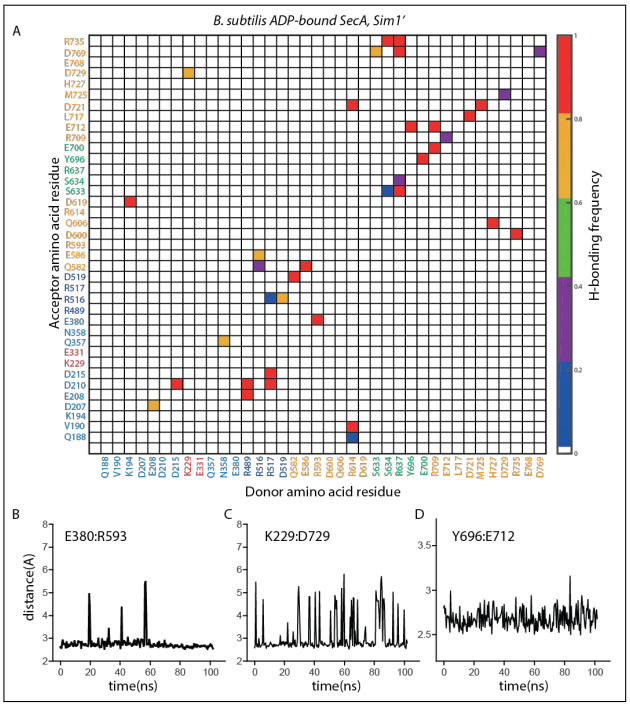


Figure C-3: Selected H bonds monitored from Sim1'. (A) H-bond map of selected inter-domain interactions in B. subtilis ADP-bound SecA computed from the last 30ns of Sim1'. We illustrate H bonds involving the HSD. (B&C) Minimum distances measured between E380 and R593 (panel B), and between K229 and D729 (panel C). (D) H bonding between Y696 and D712.

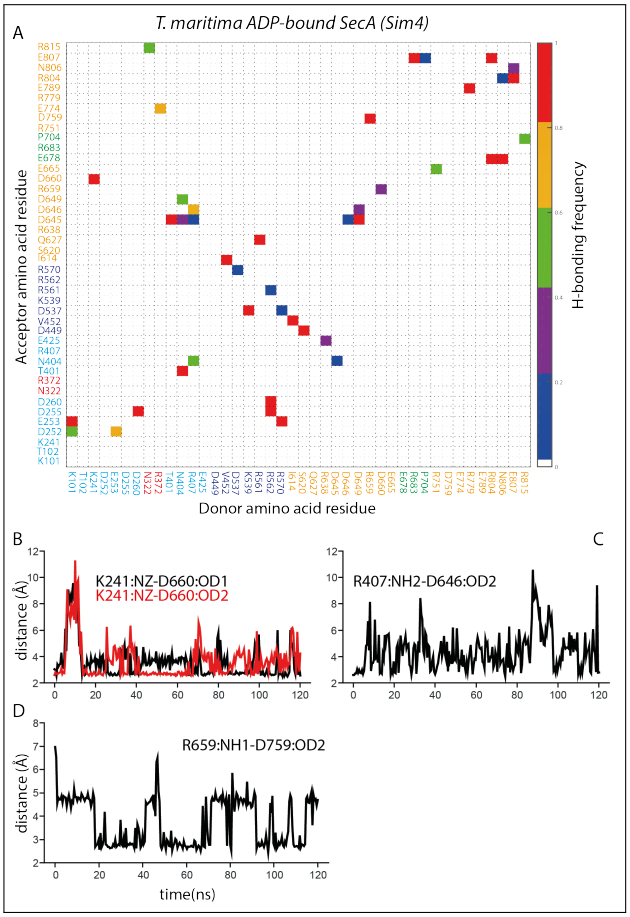


Figure C-4: Selected H bonds in T. maritima SecA. (A) H-bond map for selected inter-domain interactions computed from the last 30ns of Sim4. (B-D) Time series of selected interactions computed from Sim2.

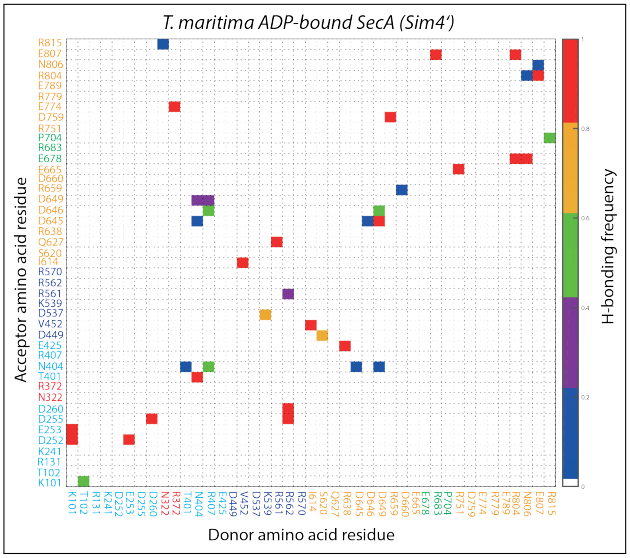


Figure C-5: H-bond map of selected interactions in T. maritima ADP-bound SecA computed from the last 30ns of Sim4'

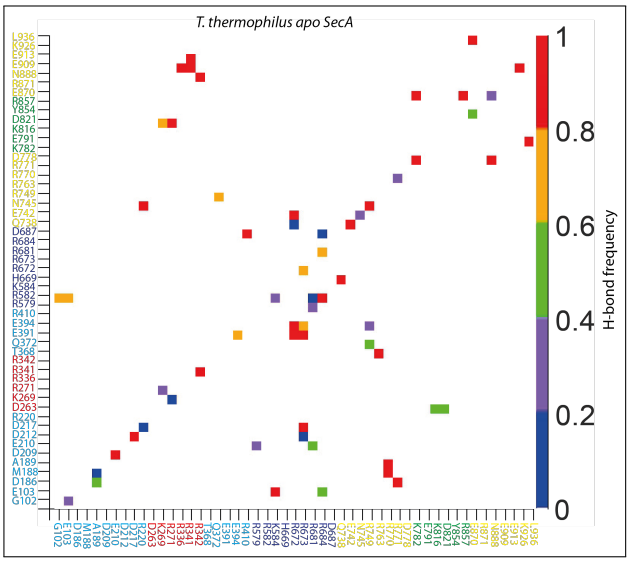


Figure C-6: H-bond map of selected inter-domain interactions computed from the last 30ns of T. thermophilus SecA (Sim5). The H-bond map was computed from the last 30ns of Sim5. We illustrate the H bonds of the HSD with other regions of SecA. A red dot corresponds to presence of an H-bonding interaction during 80-100% of the trajectory fragment used for analysis.

C.3 Secondary Structure Analysis

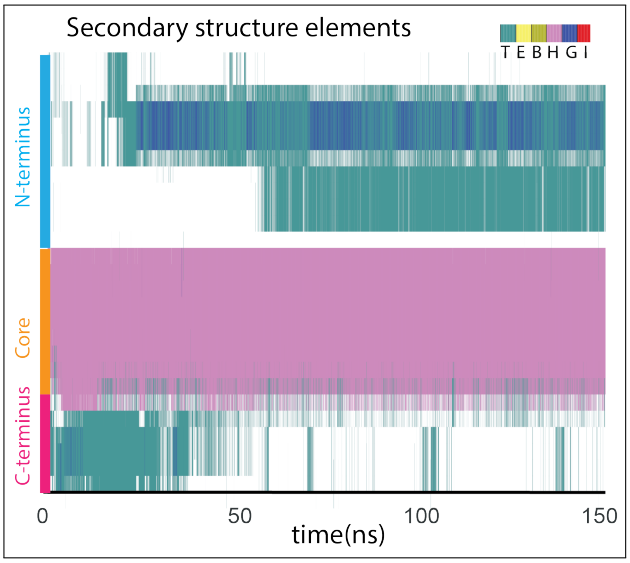


Figure C-7: Secondary structure analysis of the SP. The analysis was performed on SimB'using the same protocol as described for SimB in the main text. Note the presence of α-helical segment throughout SimB'in the hydrophobic part of the SecA. The secondary structure analysis was performed using STRIDE [134] in VMD [128]. The structural elements are denoted as T- turn, E- extended, B- isolated bridge, H- helix, G- 3-10 helix, I- π-helix

C.4 Convergence of SecA-SP interactions

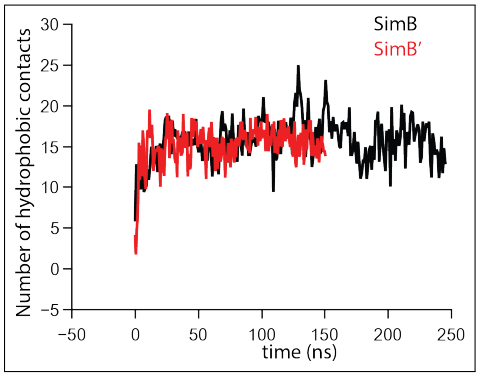


Figure C-8: Hydrophobic contacts between the SP and SecA. The number of hydrophobic contacts was computed as the number of the heavy-atom SecA/SP pairs of hydrophobic residues within a distance of 5Å. Black and red profiles give the number of hydrophobic contacts along SimB and SimB', respectively. In each Sim there are, on average, 15 hydrophobic contacts between the SP and SecA.

C.5 PCA convergence

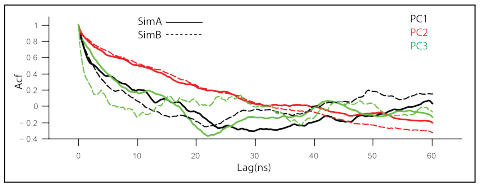


Figure C-9: Auto-correlation functions of selected PC modes computed from SimA and SimB. Auto-correlation functions (Acf) are plotted as a function of the correlation lag. Note that PC2 modes have slow convergence on the timescale of 100 ns, that is, during the sampling period these modes do not exhibit a repetitive pattern.

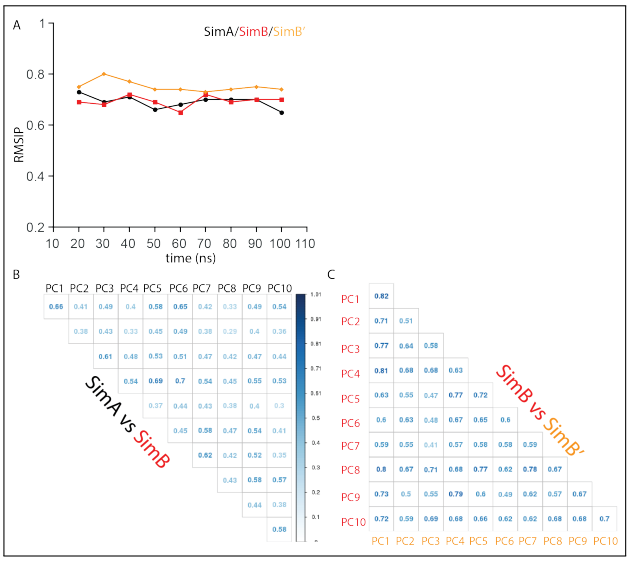


Figure C-10: Testing the convergence of the PCA analyses in SimA, SimB and SimB'. (A) RMSIP (root mean square inner product) as a function of the simulation time during last 100 ns of each Sim calculated for 10ns subsequent windows. Note that in all simulations RMSIP has steady behavior with plateau values that range between 0.7 and 0.75 which is a good indicator of the convergence [113]. Essential subspace overlap between SimA and SimB, and between SimB and SimB', as computed from the last 100 ns of each Sim. Note that the squared dot product of the PCs in SimA and SimB are smaller than the dot product of PCs in SimB vs. SimB', suggesting that the nature of motion in Sim2 is more similar to the SimB'than to SimA.

C.6 LMI convergence

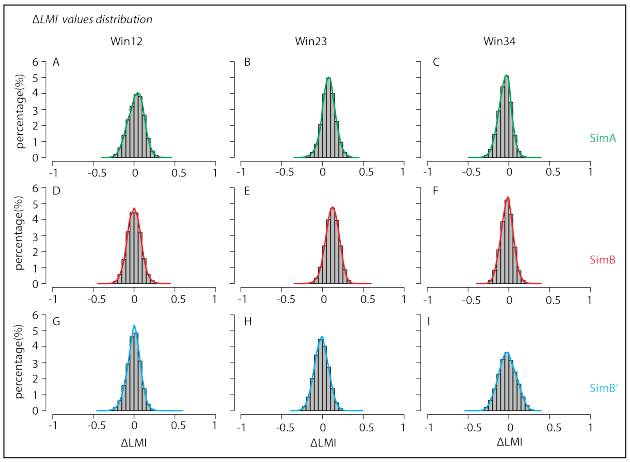


Figure C-11: LMI convergence tests. LMI difference value distributions calculated for 25ns windows of SimA (panels A-C), SimB (panels D-F), and SimB'(panels G-I). The LMI differences were calculate for each subsequent window pair, e.g., Win12 = Win1-Win2. It can be seen that the absolute values are largely âĽď 0.2, which suggests overall well-converged LMI calculations.

Appendix D

Dihedral angles

In proteins, large amount of different conformations is accessible only by rotating around covalent bonds. This rotational flexibility is described with a dihedral angle. Dihedral angle measures the relative rotation of 4 linked atoms in a molecule, i-j-k-l and is denoted as τ_{ijkl}. Dihedral angle is defined as the angle between normals to the planes defined by atoms *i-j-k* and atoms *j-k-l*. Suplemetanry angle to dihedral is denoted as torsion, $\tilde{\tau}_{ijkl}$ and it defines the angle between before mentioned two planes. The visual representation of dihedaral angle is given in figure D-1.

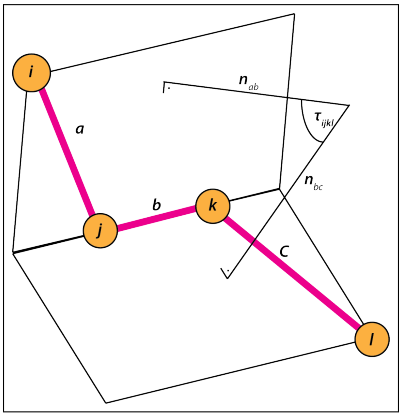


Figure D-1: Example of dihedral angle, given for an arbitrary set of four covalently bound atoms.

When the value of the dihedral angle is 0° the four atoms are co-planar and the atoms i and l are closest to each other. This conformation is denoted as *cis*. When the angle takes value of 180° all atoms are again co-planar, but now i and j are at

largest possible distance from each other. This conformation is denoted as *trans*.

When it comes to the computations or, more precise, force fields the interactions achieved via dihedral angles are desribed with $torsionalpotentials$. Torsional potentials are Incorporated into force fields because all the other interactions failed to provide decent predictions of rotational energy profiles observed when two groups rotate around a covalent bond even for the simplest molecules. The origin of the barriers observed during this rotations is not fully understood yet. It was thought that main contributions to this potential come from the repulsion of electron clouds of two rotating groups [144] whereas newer explanations suggest that torsional potential is dominated by quantum effect of resonance stabilization [145].

Appendix E

CMAP correction

Important step in improvement of the CHARMM force field was the CMAP correction [92]. CMAP correction deals with parametrization of the ψ and ϕ dihedral angles in protein backbone. This correction is performed with the goal to achieve better agreement of secondary structure in simulations with the experimental results. The QM calculations were done on alanine, glycine and proline dipeptides. The results obtained are then used to improve existing force field potential. This force field potential is then used for systematic MD study and compassion of results with the QM and crystallographic data. This showed that additional grid-based correction is needed in order to improve treatment of backbone in proteins. The necessity of this correction was demonstrated wit simulation lysozyme where previously detected loops in secondary structure which were not in agreement with experiments were eliminated [146].

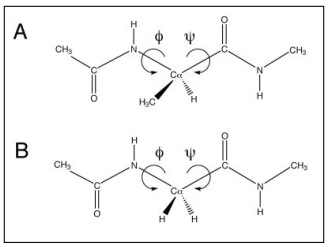


Figure E-1: Diagrams of the (A) alanine dipeptide and the (B) glycine dipeptides taken from [92]

Appendix F

PMF

The term PMF(proton-motive force) find its foundations back in 1961, in work done by Mitchell [147]. In this work he elaborates the chemiosmotic theory for which he was awarded with the Nobel Prize in Chemistry. His work, was focused on the synthesis of ATP and proposed that energy for this process comes from the proton transport across the membrane.

Protons are ions and in general movement of ions across the membrane depends on two parameters: *i*) Diffusion force that originates from the concentration gradient and *ii* electrostatic potential gradient, along which ions can flow. Together this two gradients five an electrochemical gradient. If the previous formulation is translated in terms of free energy the following equation can be written:

$$\Delta G = zF\Delta\phi + RTln\frac{[X^{z+}]_{IN}}{[X^{z+}]_{OUT}} \tag{F.1}$$

where ΔG is difference in the free energy for transport from cations from outside(OUT) of the cell towards the inside(IN), z is the charge number of the single cation X^{z+}, $\Delta\phi$ is electrostatic potential established between OUT and IN, $[X^{z+}]_{IN}$ and $[X^{z+}]_{OUT}$ are concentrations of cations inside and outside of the cell, respectively. F, R, T are Faraday constant, gas constant and temperature, respectively. This free energy difference is usually denoted as $\Delta\mu_{N^{z+}}$. For hydrogen ion, proton, the previous

equation takes form:

$$\Delta\mu_{H^+} = F\Delta\phi + RTln\frac{[H^+]_{IN}}{[H^{z+}]_{OUT}} \tag{F.2}$$

The proton-motive force is finally given with

$$\Delta p = -\frac{\Delta\mu_{H^+}}{F}. \tag{F.3}$$

PMF provides energy for many processes inside the cell, but most important process that is fueled by PMF is ATP synthesis, as mentioned in the beginning of this chapter. The energy that originates from spontaneous flow of protons across the membrane is being "stored" in a form of synthesized ATP molecules which are later used as energy source by ATPases (such as SecA).

Bibliography

[1] Hilary Muirhead and MF Perutz. Structure of hæmoglobin: A three-dimensional fourier synthesis of reduced human haemoglobin at 5.5 å resolution. *Nature*, 199(4894):633–638, 1963.

[2] John C Kendrew, G Bodo, Howard M Dintzis, RG Parrish, Harold Wyckoff, and David C Phillips. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662–666, 1958.

[3] Tamar Schlick. *Molecular modeling and simulation: an interdisciplinary guide: an interdisciplinary guide*, volume 21. Springer Science & Business Media, 2010.

[4] Linus Pauling. The shared-electron chemical bond. *Proceedings of the national academy of sciences*, 14(4):359–362, 1928.

[5] John Lennard-Jones and JA Pople. Molecular association in liquids. i. molecular association due to lone-pair electrons. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 205, pages 155–162. The Royal Society, 1951.

[6] Peter A Kollman and Leland C Allen. Theory of the hydrogen bond. *Chemical Reviews*, 72(3):283–303, 1972.

[7] Aaron R Dinner, Themis Lazaridis, and Martin Karplus. Understanding β-hairpin formation. *Proceedings of the National Academy of Sciences*, 96(16):9068–9073, 1999.

[8] Leonard J Prins, David N Reinhoudt, and Peter Timmerman. Noncovalent synthesis using hydrogen bonding. *Angewandte Chemie International Edition*, 40(13):2382–2426, 2001.

[9] Gopalasamudram Narayana Ramachandran, Chandrasekharan Ramakrishnan, and V Sasisekharan. Stereochemistry of polypeptide chain configurations. *Journal of molecular biology*, 7(1):95–99, 1963.

[10] Jochen Zimmer and Tom A Rapoport. Conformational flexibility and peptide interaction of the translocation atpase seca. *Journal of molecular biology*, 394(4):606–612, 2009.

[11] William C Wimley, Klaus Gawrisch, Trevor P Creamer, and Stephen H White. Direct measurement of salt-bridge solvation energies using a peptide model system: implications for protein stability. *Proceedings of the National Academy of Sciences*, 93(7):2985–2990, 1996.

[12] Alexey S Ladokhin and Stephen H White. Folding of amphipathic α-helices on membranes: energetics of helix formation by melittin. *Journal of molecular biology*, 285(4):1363–1369, 1999.

[13] Stephen H White and Gunnar von Heijne. How translocons select transmembrane helices. *Annu. Rev. Biophys.*, 37:23–42, 2008.

[14] Sajith Jayasinghe, Kalina Hristova, and Stephen H White. Energetics, stability, and prediction of transmembrane helices. *Journal of molecular biology*, 312(5):927–934, 2001.

[15] AJ Driessen. Precursor protein translocation by the escherichia coli translocase is directed by the protonmotive force. *The EMBO journal*, 11(3):847, 1992.

[16] Elmar Schiebel, Arnold JM Driessen, Franz-Ulrich Hartl, and William Wickner. $\delta\mu$ h+ and atp function at different steps of the catalytic cycle of preprotein translocase. *Cell*, 64(5):927–939, 1991.

[17] Peter Fekkes and Arnold JM Driessen. Protein targeting to the bacterial cytoplasmic membrane. *Microbiology and Molecular Biology Reviews*, 63(1):161–173, 1999.

[18] Arnold JM Driessen and Nico Nouwen. Protein translocation across the bacterial cytoplasmic membrane. *Annu. Rev. Biochem.*, 77:643–667, 2008.

[19] Katerina E Chatzi, Marios Frantzeskos Sardis, Anastassios Economou, and Spyridoula Karamanou. Seca-mediated targeting and translocation of secretory proteins. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1843(8):1466–1474, 2014.

[20] Jeroen PW van der Wolk, Janny G de Wit, and Arnold JM Driessen. The catalytic cycle of the escherichia coli seca atpase comprises two distinct preprotein translocation events. *The EMBO Journal*, 16(24):7297–7304, 1997.

[21] Giorgos Gouridis, Spyridoula Karamanou, Ioannis Gelis, Charalampos G Kalodimos, and Anastassios Economou. Signal peptides are allosteric activators of the protein translocase. *Nature*, 462(7271):363–367, 2009.

[22] Spyridoula Karamanou, Giorgos Gouridis, Efrosyni Papanikou, Giorgos Sianidis, Ioannis Gelis, Dimitra Keramisanou, Eleftheria Vrontou, Charalampos G Kalodimos, and Anastassios Economou. Preprotein-controlled catalysis in the helicase motor of seca. *The EMBO journal*, 26(12):2904–2914, 2007.

[23] Nagaraju Akula, Pankaj Trivedi, Frank Q Han, and Nian Wang. Identification of small molecule inhibitors against seca of candidatus liberibacter asiaticus by structure based design. *European journal of medicinal chemistry*, 54:919–924, 2012.

[24] Jinshan Jin, Jianmei Cui, Arpana Sagwal Chaudhary, Ying-Hsin Hsieh, Krishna Damera, Hao Zhang, Hsiuchin Yang, Binghe Wang, and Phang C Tai. Evaluation of small molecule seca inhibitors against methicillin-resistant staphylococcus aureus. *Bioorganic & medicinal chemistry*, 23(21):7061–7068, 2015.

[25] Evelien De Waelheyns, Kenneth Segers, Marios Frantzeskos Sardis, Jozef Anné, Gerry AF Nicolaes, and Anastassios Economou. Identification of small-molecule inhibitors against seca by structure-based virtual ligand screening. *The Journal of antibiotics*, 68(11):666–673, 2015.

[26] Franz-Ulrich Hartl, Stewart Lecker, Elmar Schiebel, Joseph P Hendrick, and William Wickner. The binding cascade of secb to seca to secye mediates preprotein targeting to the e. coli plasma membrane. *Cell*, 63(2):269–279, 1990.

[27] Roland Lill, William Dowhan, and William Wickner. The atpase activity of seca is regulated by acidic phospholipids, secy, and the leader and mature domains of precursor proteins. *Cell*, 60(2):271–280, 1990.

[28] Lorna Brundage, Joseph P Hendrick, Elmar Schiebel, Arnold JM Driessen, and William Wickner. The purified e. coli integral membrane protein secye is sufficient for reconstitution of seca-dependent precursor protein translocation. *Cell*, 62(4):649–657, 1990.

[29] Karen Douville, Albert Price, Jerry Eichler, Anastassios Economou, and William Wickner. Secyeg and seca are the stoichiometric components of preprotein translocase. *Journal of Biological Chemistry*, 270(34):20106–20111, 1995.

[30] Dorothy M Kim, Haiyan Zheng, Yuanpeng J Huang, Gaetano T Montelione, and John F Hunt. Atpase active-site electrostatic interactions control the global conformation of the 100 kda seca translocase. *Journal of the American Chemical Society*, 135(8):2999–3010, 2013.

[31] Ioannis Gelis, Alexandre MJJ Bonvin, Dimitra Keramisanou, Marina Koukaki, Giorgos Gouridis, Spyridoula Karamanou, Anastassios Economou, and Charalampos G Kalodimos. Structural basis for signal-sequence recognition by the translocase motor seca as determined by nmr. *Cell*, 131(4):756–769, 2007.

[32] Yannis Papanikolau, Maria Papadovasilaki, Raimond BG Ravelli, Andrew A McCarthy, Stephen Cusack, Anastassios Economou, and Kyriacos Petratos.

Structure of dimeric seca, the escherichia coli preprotein translocase motor. *Journal of molecular biology*, 366(5):1545–1557, 2007.

[33] Dmitry G Vassylyev, Hiroyuki Mori, Marina N Vassylyeva, Tomoya Tsukazaki, Yoshiaki Kimura, Tahir H Tahirov, and Koreaki Ito. Crystal structure of the translocation atpase seca from thermus thermophilus reveals a parallel, head-to-head dimer. *Journal of molecular biology*, 364(3):248–258, 2006.

[34] Andrew R Osborne, William M Clemons, and Tom A Rapoport. A large conformational change of the translocation atpase seca. *Proceedings of the National Academy of Sciences of the United States of America*, 101(30):10937–10942, 2004.

[35] John F Hunt, Sevil Weinkauf, Lisa Henry, John J Fak, Paul McNicholas, Donald B Oliver, and Johann Deisenhofer. Nucleotide control of interdomain interactions in the conformational reaction cycle of seca. *Science*, 297(5589):2018–2026, 2002.

[36] Vivek Sharma, Arulandu Arockiasamy, Donald R Ronning, Christos G Savva, Andreas Holzenburg, Miriam Braunstein, William R Jacobs, and James C Sacchettini. Crystal structure of mycobacterium tuberculosis seca, a preprotein translocating atpase. *Proceedings of the National Academy of Sciences*, 100(5):2243–2248, 2003.

[37] Yu Chen, Benedikt W Bauer, Tom A Rapoport, and James C Gumbart. Conformational changes of the clamp of the protein translocation atpase seca. *Journal of molecular biology*, 427(14):2348–2359, 2015.

[38] Christine Mitchell and Donald Oliver. Two distinct atp-binding domains are needed to promote protein export by escherichia coli seca atpase. *Molecular microbiology*, 10(3):483–497, 1993.

[39] Georgios Sianidis, Spyridoula Karamanou, Eleftheria Vrontou, Kostantinos Boulias, Kostantinos Repanas, Nikos Kyrpides, Anastasia S Politou, and Anas-

tassios Economou. Cross-talk between catalytic and regulatory elements in a dead motor domain is essential for seca function. *The EMBO Journal*, 20(5):961–970, 2001.

[40] Stefan Milenkovic and Ana-Nicoleta Bondar. Mechanism of conformational coupling in seca: key role of hydrogen-bonding networks and water interactions. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1858(2):374–385, 2016.

[41] Stanley Nithianantham and Brian H Shilton. Analysis of the isolated seca dead motor suggests a mechanism for chemical–mechanical coupling. *Journal of molecular biology*, 383(2):380–389, 2008.

[42] Eugene V Koonin and Alexander E Gorbalenya. Autogenous translation regulation by escherichia coli atpase seca may be mediated by an intrinsic rna helicase activity of this protein. *FEBS letters*, 298(1):6–8, 1992.

[43] Eiichiro Kimura, Mitsuru Akita, Shin-ichi Matsuyama, and S Mizushima. Determination of a region in seca that interacts with presecretory proteins in escherichia coli. *Journal of Biological Chemistry*, 266(10):6600–6606, 1991.

[44] Hiroyuki Mori and Koreaki Ito. The long α-helix of seca is important for the atpase coupling of translocation. *Journal of Biological Chemistry*, 281(47):36249–36256, 2006.

[45] Karl J Erlandson, Stephanie BM Miller, Yunsun Nam, Andrew R Osborne, Jochen Zimmer, and Tom A Rapoport. A role for the two-helix finger of the seca atpase in protein translocation. *Nature*, 455(7215):984–987, 2008.

[46] Sanchaita Das, Lorry M Grady, Jennifer Michtavy, Yayan Zhou, Frederick M Cohan, Manju M Hingorani, and Donald B Oliver. The variable subdomain of escherichia coli seca functions to regulate seca atpase activity and adp release. *Journal of bacteriology*, 194(9):2205–2213, 2012.

[47] Jochen Zimmer, Yunsun Nam, and Tom A Rapoport. Structure of a complex of the atpase seca and the protein-translocation channel. *Nature*, 455(7215):936–943, 2008.

[48] Ron Milo, Paul Jorgensen, Uri Moran, Griffin Weber, and Michael Springer. BionumbersâĂŤthe database of key numbers in molecular and cell biology. *Nucleic acids research*, 38(suppl_1):D750–D753, 2009.

[49] Arieh Warshel. Computer simulations of enzyme catalysis: methods, progress, and insights. *Annual review of biophysics and biomolecular structure*, 32(1):425–443, 2003.

[50] Arieh Warshel and Jan Florián. Computer simulations of enzyme catalysis: finding out what has been optimized by evolution. *Proceedings of the National Academy of Sciences*, 95(11):5950–5955, 1998.

[51] Markus Dittrich, Shigehiko Hayashi, and Klaus Schulten. Atp hydrolysis in the β tp and β dp catalytic sites of f 1-atpase. *Biophysical journal*, 87(5):2954–2967, 2004.

[52] Markus Dittrich, Shigehiko Hayashi, and Klaus Schulten. On the mechanism of atp hydrolysis in f 1-atpase. *Biophysical journal*, 85(4):2253–2266, 2003.

[53] Markus Dittrich and Klaus Schulten. Zooming in on atp hydrolysis in f 1. *Journal of bioenergetics and biomembranes*, 37(6):441–444, 2005.

[54] Marek SÌĘtrajbl, Avital Shurki, and Arieh Warshel. Converting conformational changes to electrostatic energy in molecular motors: The energetics of atp synthase. *Proceedings of the National Academy of Sciences*, 100(25):14834–14839, 2003.

[55] Bert Van den Berg, William M Clemons, Ian Collinson, Yorgo Modis, Enno Hartmann, Stephen C Harrison, and Tom A Rapoport. X-ray structure of a protein-conducting channel. *Nature*, 427(6969):36–44, 2004.

[56] Sarah Whitehouse, Vicki AM Gold, Alice Robson, William J Allen, Richard B Sessions, and Ian Collinson. Mobility of the seca 2-helix-finger is not essential for polypeptide translocation via the secyeg complex. *J Cell Biol*, 199(6):919–929, 2012.

[57] Anastassios Economou, Joseph A Pogliano, Jonathan Beckwith, Donald B Oliver, and William Wickner. Seca membrane cycling at secyeg is driven by distinct atp binding and hydrolysis events and is regulated by secd and secf. *Cell*, 83(7):1171–1181, 1995.

[58] Spyridoula Karamanou, Eleftheria Vrontou, George Sianidis, Catherine Baud, Tilmann Roos, Andreas Kuhn, Anastasia S Politou, and Anastassios Economou. A molecular switch in seca protein couples atp hydrolysis to protein translocation. *Molecular microbiology*, 34(5):1133–1145, 1999.

[59] Gunnar HEIJNE. Patterns of amino acids near signal-sequence cleavage sites. *The FEBS Journal*, 133(1):17–21, 1983.

[60] Gunnar von Heijne. How signal sequences maintain cleavage specificity. *Journal of molecular biology*, 173(2):243–251, 1984.

[61] Gunnar von Heijne. Analysis of the distribution of charged residues in the n-terminal region of signal sequences: implications for protein export in prokaryotic and eukaryotic cells. *The EMBO journal*, 3(10):2315, 1984.

[62] Gunnar Von Heijne. Signal sequences: the limits of variation. *Journal of molecular biology*, 184(1):99–105, 1985.

[63] Sumiko Inouye, Xavier Soberon, Thomas Franceschini, Kenzo Nakamura, Keiichi Itakura, and Masayori Inouye. Role of positive charge on the amino-terminal region of the signal peptide in protein secretion across the membrane. *Proceedings of the National Academy of Sciences*, 79(11):3438–3441, 1982.

[64] George P Vlasuk, S Inouye, H Ito, K Itakura, and M Inouye. Effects of the complete removal of basic amino acid residues from the signal peptide on secretion

of lipoprotein in escherichia coli. *Journal of Biological Chemistry*, 258(11):7141–7148, 1983.

[65] John W Puziss, JD Fikes, and PJ Bassford. Analysis of mutational alterations in the hydrophilic segment of the maltose-binding protein signal peptide. *Journal of bacteriology*, 171(5):2303–2311, 1989.

[66] Mitsuru Akita, Shoichi Sasaki, Shin-iehi Matsuyama, and Shoji Mizushima. Seca interacts with secretory proteins by recognizing the positive charge at the amino terminus of the signal peptide in escherichia coli. *Journal of Biological Chemistry*, 265(14):8164–8169, 1990.

[67] S MacIntyre, ML Eschbach, and B Mutschler. Export incompatibility of n-terminal basic residues in a mature polypeptide of escherichia coli can be alleviated by optimising the signal peptide. *Molecular & general genetics: MGG*, 221(3):466–474, 1990.

[68] Vytas A Bankaitis, Beth A Rasmussen, and Philip J Bassford. Intragenic suppressor mutations that restore export of maltose binding protein with a truncated signal peptide. *Cell*, 37(1):243–252, 1984.

[69] Debra A Kendall, Suzanne K Doud, and Emil Thomas Kaiser. A comparative analysis of single-and multiple-residue substitutions in the alkaline phosphatase signal peptide. *Biopolymers*, 29(1):139–147, 1990.

[70] Jennifer W Izard, Sharyn L Rusch, and Debra A Kendall. The amino-terminal charge and core region hydrophobicity interdependently contribute to the function of signal sequences. *Journal of Biological Chemistry*, 271(35):21579–21582, 1996.

[71] Jacque Monod, Jeffries Wyman, and Jean-Pierre Changeux. On the nature of allosteric transitions: a plausible model. *Journal of molecular biology*, 12(1):88–118, 1965.

[72] DE Koshland Jr, G Nemethy, and D_ Filmer. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, 5(1):365–385, 1966.

[73] David JF du Plessis, Greetje Berrelkamp, Nico Nouwen, and Arnold JM Driessen. The lateral gate of secyeg opens during protein translocation. *Journal of Biological Chemistry*, 284(23):15805–15814, 2009.

[74] Protein Data Bank. Hm berman, j. westbrook, z. feng, g. gilliland, tn bhat, h. weissig, in shindyalov, pe bourne. *Nucleic Acids Res*, 28:235, 2000.

[75] Kurt Wüthrich. The way to nmr structures of proteins. *Nature Structural & Molecular Biology*, 8(11):923–925, 2001.

[76] FA Momany, R F_ McGuire, AW Burgess, and Harold A Scheraga. Energy parameters in polypeptides. vii. geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *The Journal of Physical Chemistry*, 79(22):2361–2381, 1975.

[77] Bernard R Brooks, Robert E Bruccoleri, Barry D Olafson, David J States, S a Swaminathan, and Martin Karplus. Charmm: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, 4(2):187–217, 1983.

[78] NL Allinger. Calculation of molecular structure and energy by force-field methods. *Advances in physical organic chemistry*, 13:1–82, 1976.

[79] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[80] Michael JS Dewar, Eve G Zoebisch, Eamonn F Healy, and James JP Stewart. Development and use of quantum mechanical molecular models. 76. am1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society*, 107(13):3902–3909, 1985.

[81] James JP Stewart. Optimization of parameters for semiempirical methods i. method. *Journal of Computational Chemistry*, 10(2):209–220, 1989.

[82] Marcus Elstner, Dirk Porezag, G Jungnickel, J Elsner, M Haugk, Th Frauenheim, Sandor Suhai, and Gotthard Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Physical Review B*, 58(11):7260, 1998.

[83] CB Anfinsen and HA Scheraga. Experimental and theoretical aspects of protein folding. *Advances in protein chemistry*, 29:205–300, 1975.

[84] EI Shakhnovich and AM Gutin. Implications of thermodynamics of protein folding for evolution of primary sequences. *Nature*, 346(6286):773, 1990.

[85] Cyrus Levinthal. How to fold graciously. *Mossbauer spectroscopy in biological systems*, 67:22–24, 1969.

[86] Cyrus Levinthal. Are there pathways for protein folding? *Journal de chimie physique*, 65:44–45, 1968.

[87] Peter G Wolynes. Folding funnels and energy landscapes of larger proteins within the capillarity approximation. *Proceedings of the National Academy of Sciences*, 94(12):6170–6175, 1997.

[88] Helmut Grubmüller, Helmut Heller, Andreas Windemuth, and Klaus Schulten. Generalized verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molecular Simulation*, 6(1-3):121–142, 1991.

[89] MBBJM Tuckerman, Bruce J Berne, and Glenn J Martyna. Reversible multiple time scale molecular dynamics. *The Journal of chemical physics*, 97(3):1990–2001, 1992.

[90] Darryl D Humphreys, Richard A Friesner, and Bruce J Berne. A multiple-time-step molecular dynamics algorithm for macromolecules. *The Journal of Physical Chemistry*, 98(27):6885–6892, 1994.

[91] Alex D MacKerell Jr, Donald Bashford, MLDR Bellott, Roland Leslie Dunbrack Jr, Jeffrey D Evanseck, Martin J Field, Stefan Fischer, Jiali Gao, H Guo, Sookhee Ha, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B*, 102(18):3586–3616, 1998.

[92] Alexander D MacKerell, Michael Feig, and Charles L Brooks. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of computational chemistry*, 25(11):1400–1415, 2004.

[93] Nicolas Foloppe and Alexander D MacKerell Jr. All-atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal of computational chemistry*, 21(2):86–104, 2000.

[94] Alexander D Mackerell and Nilesh K Banavali. All-atom empirical force field for nucleic acids: Ii. application to molecular dynamics simulations of dna and rna in solution. *Journal of Computational Chemistry*, 21(2):105–120, 2000.

[95] Shneior Lifson and Arieh Warshel. Consistent force field for calculations of conformations, vibrational spectra, and enthalpies of cycloalkane and n-alkane molecules. *The Journal of Chemical Physics*, 49(11):5116–5129, 1968.

[96] Hans Frauenfelder, Stephen G Sligar, and Peter G Wolynes. The energy landscapes and motions of proteins. *Urbana*, 51(61801):61801, 1991.

[97] Arieh Warshel and Michael Levitt. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of molecular biology*, 103(2):227–249, 1976.

[98] Ana-Nicoleta Bondar, Stefan Fischer, Jeremy C Smith, Marcus Elstner, and Sándor Suhai. Key role of electrostatic interactions in bacteriorhodopsin proton transfer. *Journal of the American Chemical Society*, 126(44):14668–14677, 2004.

[99] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[100] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26(16):1781–1802, 2005.

[101] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, 1977.

[102] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An nâŃĚ log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

[103] Ulrich Essmann, Lalith Perera, Max L Berkowitz, Tom Darden, Hsing Lee, and Lee G Pedersen. A smooth particle mesh ewald method. *The Journal of chemical physics*, 103(19):8577–8593, 1995.

[104] Glenn J Martyna, Douglas J Tobias, and Michael L Klein. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, 101(5):4177–4189, 1994.

[105] Scott E Feller, Yuhong Zhang, Richard W Pastor, and Bernard R Brooks. Constant pressure molecular dynamics simulation: the langevin piston method. *The Journal of chemical physics*, 103(11):4613–4621, 1995.

[106] Ian K McDonald and Janet M Thornton. Satisfying hydrogen bonding potential in proteins. *Journal of molecular biology*, 238(5):777–793, 1994.

[107] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983.

[108] Matthias Buck and Martin Karplus. Hydrogen bond energetics: a simulation and statistical analysis of n-methyl acetamide (nma), water, and human lysozyme. *The Journal of Physical Chemistry B*, 105(44):11000–11015, 2001.

[109] Linus Pauling. *The nature of the chemical bond and the structure of molecules and crystals: an introduction to modern structural chemistry*, volume 18. Cornell university press, 1960.

[110] A_ Bondi. van der waals volumes and radii. *The Journal of physical chemistry*, 68(3):441–451, 1964.

[111] R Scott Rowland and Robin Taylor. Intermolecular nonbonded contact distances in organic crystal structures: Comparison with distances expected from van der waals radii. *The Journal of Physical Chemistry*, 100(18):7384–7391, 1996.

[112] Kari Karhunen. *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*, volume 37. Universitat Helsinki, 1947.

[113] Andrea Amadei, Antonius Linssen, and Herman JC Berendsen. Essential dynamics of proteins. *Proteins: Structure, Function, and Bioinformatics*, 17(4):412–425, 1993.

[114] Bernard R Brooks, Dušanka Janežič, and Martin Karplus. Harmonic analysis of large systems. i. methodology. *Journal of computational chemistry*, 16(12):1522–1542, 1995.

[115] Oliver F Lange and Helmut Grubmüller. Generalized correlation for biomolecular dynamics. *Proteins: Structure, Function, and Bioinformatics*, 62(4):1053–1061, 2006.

[116] Antje Wolf and Karl N Kirschner. Principal component and clustering analysis on molecular dynamics data of the ribosomal l11· 23s subdomain. *Journal of molecular modeling*, pages 1–11, 2013.

[117] Hongfeng Lou and Robert I Cukier. Molecular dynamics of apo-adenylate kinase: a principal component analysis. *The Journal of Physical Chemistry B*, 110(25):12796–12808, 2006.

[118] Shozeb Haider, Gary N Parkinson, and Stephen Neidle. Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophysical journal*, 95(1):296–311, 2008.

[119] Paul M Gasper, Brian Fuglestad, Elizabeth A Komives, Phineus RL Markwick, and J Andrew McCammon. Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities. *Proceedings of the National Academy of Sciences*, 109(52):21216–21222, 2012.

[120] Guido Scarabelli and Barry J Grant. Mapping the structural and dynamical features of kinesin motor domains. *PLoS computational biology*, 9(11):e1003329, 2013.

[121] Anurag Sethi, John Eargle, Alexis A Black, and Zaida Luthey-Schulten. Dynamical networks in trna: protein complexes. *Proceedings of the National Academy of Sciences*, 106(16):6620–6625, 2009.

[122] Barry J Grant, Ana PC Rodrigues, Karim M ElSawy, J Andrew McCammon, and Leo SD Caves. Bio3d: an r package for the comparative analysis of protein structures. *Bioinformatics*, 22(21):2695–2696, 2006.

[123] R Core Team. R language definition. *Vienna, Austria: R foundation for statistical computing*, 2000.

[124] Andrej Šali and Tom L Blundell. Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology*, 234(3):779–815, 1993.

[125] Marc A Martí-Renom, Ashley C Stuart, András Fiser, Roberto Sánchez, Francisco Melo, and Andrej Šali. Comparative protein structure modeling of genes and genomes. *Annual review of biophysics and biomolecular structure*, 29(1):291–325, 2000.

[126] Benjamin Webb and Andrej Sali. Protein structure modeling with modeller. *Protein Structure Prediction*, pages 1–15, 2014.

[127] András Fiser, Richard Kinh Gian Do, et al. Modeling of loops in protein structures. *Protein science*, 9(9):1753–1773, 2000.

[128] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[129] LLC Schrodinger. The pymol molecular graphics system, version 1.7. 4, 2010. *There is no corresponding record for this reference.*

[130] Taiyun Wei and Viliam Simko. corrplot: Visualization of a correlation matrix. *R package version 0.73*, 230(231):11, 2013.

[131] Mark A Larkin, Gordon Blackshields, NP Brown, R Chenna, Paul A McGettigan, Hamish McWilliam, Franck Valentin, Iain M Wallace, Andreas Wilm, Rodrigo Lopez, et al. Clustal w and clustal x version 2.0. *bioinformatics*, 23(21):2947–2948, 2007.

[132] Qi Zhang, Yan Li, Rich Olson, Ishita Mukerji, and Donald Oliver. Conserved seca signal peptide-binding site revealed by engineered protein chimeras and foĬĹrster resonance energy transfer. *Biochemistry*, 55(9):1291–1300, 2016.

[133] Coral del Val, José Royuela-Flor, Stefan Milenkovic, and Ana-Nicoleta Bondar. Channelrhodopsins: a bioinformatics perspective. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1837(5):643–655, 2014.

[134] Dmitrij Frishman and Patrick Argos. Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics*, 23(4):566–579, 1995.

[135] Stefan Milenkovic and Ana-Nicoleta Bondar. Motions of the seca protein motor bound to signal peptide: Insights from molecular dynamics simulations. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1860(2):416–427, 2018.

[136] J Wolk, M Klose, E Breukink, RA Demel, B de Kruijff, R Freudl, and AJM Driessen. Characterization of a bacillus subtilis seca mutant protein deficient in translocation atpase and release from the membrane. *Molecular microbiology*, 8(1):31–42, 1993.

[137] Michael Klose, Karl-Ludwig Schimz, J Van der Wolk, AJ Driessen, and Roland Freudl. Lysine 106 of the putative catalytic atp-binding site of the bacillus subtilis seca protein is required for functional complementation of escherichia coli seca mutants in vivo. *Journal of Biological Chemistry*, 268(6):4504–4510, 1993.

[138] Ken Sato, Hiroyuki Mori, Masasuke Yoshida, and Shoji Mizushima. Characterization of a potential catalytic residue, asp-133, in the high affinity atp-binding site of escherichia coli seca, translocation atpase. *Journal of Biological Chemistry*, 271(29):17439–17444, 1996.

[139] Christopher R Zito, Edwin Antony, John F Hunt, Donald B Oliver, and Manju M Hingorani. Role of a conserved glutamate residue in the escherichia coli seca atpase mechanism. *Journal of Biological Chemistry*, 280(15):14611–14619, 2005.

[140] Alice Robson, Antonia EG Booth, Vicki AM Gold, Anthony R Clarke, and Ian Collinson. A large conformational change couples the atp binding site of seca to the secy protein channel. *Journal of molecular biology*, 374(4):965–976, 2007.

[141] Lauralynn Kourtz and Donald Oliver. Tyr-326 plays a critical role in controlling seca–preprotein interaction. *Molecular microbiology*, 37(6):1342–1356, 2000.

[142] Dimitra Keramisanou, Nikolaos Biris, Ioannis Gelis, Georgios Sianidis, Spyridoula Karamanou, Anastassios Economou, and Charalampos G Kalodimos. Disorder-order folding transitions underlie catalysis in the helicase motor of seca. *Nature structural & molecular biology*, 13(7):594–602, 2006.

[143] Marcel Schmidt, Haiyuan Ding, Visvanathan Ramamurthy, Ishita Mukerji, and Donald Oliver. Nucleotide binding activity of seca homodimer is conformationally regulated by temperature and altered byprld and azi mutations. *Journal of Biological Chemistry*, 275(20):15440–15448, 2000.

[144] Russell M Pitzer. The barrier to internal rotation in ethane. *Accounts of Chemical Research*, 16(6):207–210, 1983.

[145] Vojislava Pophristic and Lionel Goodman. Hyperconjugation not steric repulsion leads to the staggered structure of ethane. *Nature*, 411(6837):565–568, 2001.

[146] Matthias Buck, Sabine Bouguet-Bonnet, Richard W Pastor, and Alexander D MacKerell. Importance of the cmap correction to the charmm22 protein force field: dynamics of hen lysozyme. *Biophysical journal*, 90(4):L36–L38, 2006.

[147] Peter Mitchell. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature*, 191(4784):144–148, 1961.

Statement of Authenticity

Hiermit erkläre ich, dass ich die vorliegende schriftliche Arbeit in allen Teilen selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel (einschliesslich elektronischer Medien und Online-Ressourcen) verwendet habe.

Stefan Milenkovic
Berlin, July 2017