University of South Bohemia in České Budějovice

Faculty of Science

and

Johannes Kepler University in Linz

Faculty of Engineering and Natural Sciences

Comparison of fast single-molecule experiments and simulation of translocation through the SecY translocon.

Bachelor Thesis

Volodymyr Tsybulskyi

Supervisor: Mgr. Tomáš Fessl, Ph.D.
Co-supervisor: RNDr. Marek Scholz, Ph.D.
České Budějovice 2018

Tsybulskyi, V., 2018: Comparison of fast single-molecule experiments and simulation of translocation through the SecY translocon. Bc. Thesis, in English. – 35 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic and Faculty of Engineering and Natural Sciences, Johannes Kepler University, Linz, Austria.

Annotation:
I have developed a computational package using novel Monte Carlo 1D approach to test current hypotheses of SecYEG translocation against available experimental data. I have also introduced new correlative approach using Protein Structure Networks to process data from molecular dynamics, which will improve success rate in design of future fluorescence experiments in the Laboratory of Single Molecule Fluorescence.

Keywords: protein translocation, simulation, SecYEG, 1D Brownian model, molecular motors

Affirmation:
I hereby declare that I have worked on my bachelor‘s thesis independently and used only the sources listed in the bibliography.

I hereby declare that, in accordance with Article 47b of Act No. 111/1998 in the valid wording, I agree with the publication of my bachelor thesis, in full to be kept in the Faculty of Science archive, in electronic form in publicly accessible part of the STAG database operated by the University of South Bohemia in České Budějovice accessible through its web pages. Further, I agree to the electronic publication of the comments of my supervisor and thesis opponents and the record of the proceedings and results of the thesis defense in accordance with aforementioned Act No. 111/1998. I also agree to the comparison of the text of my thesis with the Theses.cz thesis database operated by the National Registry of University Theses and a plagerism detection system.

In České Budějovice 18.4.2018

Signature .....

Acknowledgements:

I want to thank my supervisor Tomáš Fessl for providing me with data and scientific ideas. I would also like to thank him for supporting and guiding me during bachelor project. I also thank to the entire bioinformatics office at JKU, particularly Sepp Hochreiter and Birgit Hauer, for their assistance during my study. I also want to thank to Institute of Chemistry and Biochemistry in JCU for providing computers for simulation and analysis of data.

Contents

1. Introduction

Protein secretion is often realized by the ubiquitous Sec machinery. In all prokaryotes, the driving force for protein transduction comes from ATP hydrolysis by the cytosolic protein SecA, together with the proton motive force (PMF). Here we focus on SecYEG complex from Gramm negative bacteria, which mediates the selective transport of proteins across the bacterial inner membrane. This channel is also responsible for integration of membrane proteins into the inner membrane. Detailed mechanism of this type of translocation remains unclear. Until now, there are at least three main models trying to explain the mechanism: Brownian ratchet model, SecA dimerization model and power-stroke model.

In this study, we want to develop a computational framework based on Monte Carlo simulation in 1D and use this framework to test several working models of translocation. This simplified simulation would give us a simple and versatile tool to compare experimental data with simulation.

Translocon size and duration of translocation (tens of seconds) make atomistic molecular dynamic simulation rather challenging and it would require computational resources which are not available yet.

Specifically, we focused on physical properties of translocated protein and SecYEG pore to obtain a more precise estimate of mechanism of protein translocation through SecYEG translocon.

1.1. Motivation

Gram-negative bacteria can be found anywhere around the world, in all possible environments that support life. The gram-negative bacteria include model organisms, like *Escherichia coli*, and various pathogenic bacteria, like *Pseudomonas aeruginosa*, *Neisseria gonorrhoeae*, *Chlamydia trachomatis*, *and Yersinia pestis*. Model organisms are widely used for scientific research of biological processes. Pathogenic bacteria are a microorganism that can cause disease. [1] Although most of the bacteria are nor harmful some are often beneficial for human health, some are pathogenic. Nowadays fewer than 100 species cause infectious illness in humans. [2]

A major virulence mechanism in bacterial infection is linked to extracellular secretion of proteins. All proteins destined to pass to surrounding space around gram-negative bacteria have to pass double membrane during their long path through the bacterial cell envelope. This involves passing through inner and outer membrane. There are four secretion paths related to this translocation in Gramm negative bacteria.

A typical E.Coli cell contains about 3 million proteins.[3] To achieve a doubling time of 20 min, and given that approximately 20% of the protein mass of an E. coli cell is located in the cell envelope.[4] Most of them are transported via SecYEG. Amount of proteins nears to 30 000 per minute. From previous research we can state that majority of these proteins goes through Sec machinery complexes.

One of the main elements of Sec machinery is SecYEG. It is a hetero-trimeric membrane protein complex. In prokaryotes, it is located in inner membrane. In eukaryotes, the analog Sec61abg is situated in the membrane of ER (endoplasmic reticulum). SecYEG forms a protein channel from cytoplasm to periplasm. Due to this, it is often called the translocon. Translocation requires ATP hydrolysis, which provides energy for protein transduction. In living cells, proton motive force also helps translocation. [5]

1.2. SecYEG structure

SecYEG is viable part of a translocon complex. It consists of three membrane proteins: SecE, SecG, and SecY. SecY is the biggest protein, which plays an irreplaceable role in protein translocation. Translocon structure was first described in high resolution in 2002 on complex from the *Methanococcus jannaschii* (SecYEβ). [6]

Core of translocon is formed by 10 transmembrane (TM) helixes. They have crab-claw-like structure, and are divided into 2 types of domains: (i) N- terminal membrane domains (TMs 1-5) and C-terminal membrane domains (TMs 6-10). These TMs form a substrate-transducing pore.

To avoid passage of unwanted small molecules, or leakage of ions through translocon, SecYEG seals its pore by a central constriction ring formed by six hydrophobic residues and helical plug. This arrangement is extremely important to keep a membrane potential and proton-motive force.

Between N- and C- terminal membrane domains (more precisely between TMs 2/3 and 7/8) is a lateral gate (LG). LG is important for recognition of signal peptide. N-terminal part of the pre-protein, which helps translocon to sort translocating substrates. It decides, whether the protein will be rejected by translocon, or translocated either to periplasm or directly to inner membrane. After successful translocation, signal peptide gets cleaved off by signal peptidase. [5]

N- and C- terminal membrane domains are connected by a hinge region between TMs 5 and 6.[6] TM2 and TM7 are outermost parts of this domains and distance between them increase with ATP hydrolysis. See figure 1 for structural details.

Protein translocation through SecYEG is driven by cytoplasmic ATPase SecA.[7][8][9] SecA consist of 2 RecA-like nucleotide binding domains. They are NBD1 and NBD2 and together they form the nucleotide-binding site(NBS).[10] A “two-helix-finger”(2HF) in binding position between SecY and SecA, and pre-protein cross-linking domain (PPXD), which forms contact with the substrate. [11][12]

SecY and SecA has tight interaction and 2HF from SecA structure sticking into SecYEG channel and SecY loop between TM6 and TM7 immersed inside SecA. Large movement of the PPXD form SecA activate the ATPase, and clamp the substrate in place during opening of the SecYEG lateral gate. [13][14]

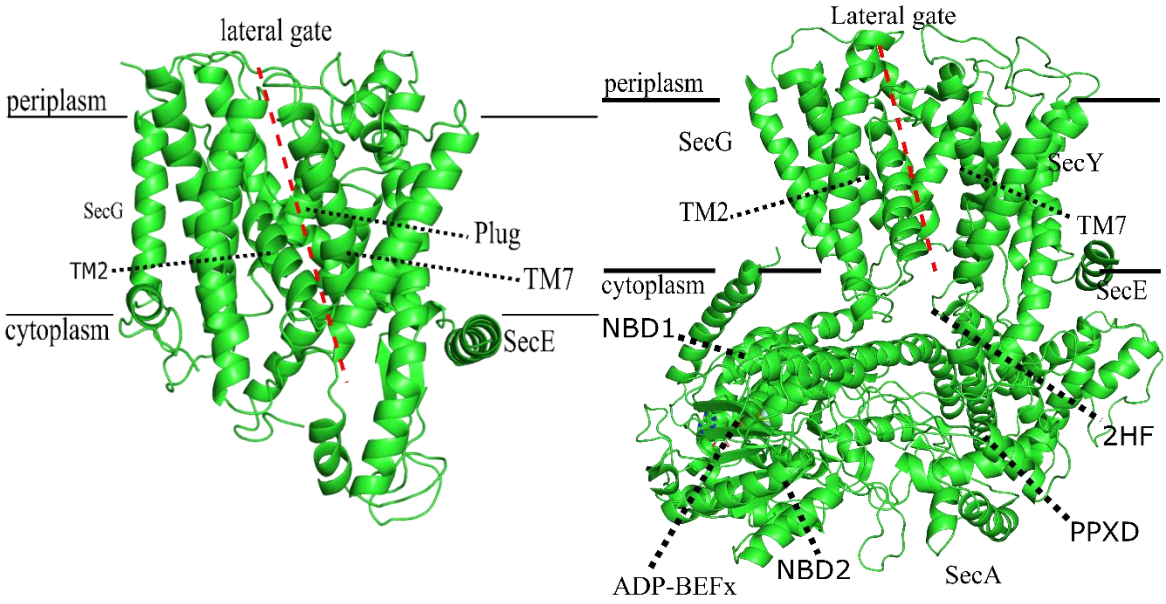

Figure 1. Left: structure of Methanococcus jannaschii complex (PDB: 1RHZ). See location of lateral gate relatively to membrane, and plug inside pore. Right side: Thermotoga maritima complex (PDB code 3DIN) with SecA attached, see change in state of lateral gate, 2HF, ADP inside SecA and PPXD and NBS domains.

1.3. Current models of translocation

Bacteria do not have ATP in the periplasmic region and therefor translocon has to produce energy for translocation from ATP on cytoplasmic side. There are three main theories describing this event. They are called power-stroke model, SecA dimerization model and Brownian ratchet model. [15][16][17][18][19][20][21] They can be divided in two main groups: power stroke models, with sequential steps based on ATP hydrolysis, and diffusion based models, where the ATP hydrolysis cycle acts on a ratchet to give it directionality. All of these models will be briefly described in following sections.

1.3.1. Brownian ratchet model

One of the most widely accepted models of these days relies on diffusion of translocated pre-protein together with SecA driven opening and closing of SecYEG pore. This model has multiple advantages over other theories. It could be potentially more energy efficient, due to decreased requirements on ATP hydrolysis. Diffusion of protein through pore occurs in one dimension, due to limited width of transduction channel. A novel Brownian ratchet model for protein translocation, described by Allen et al. in 2016, closely resembles turn still mechanism. In this model, protein diffuses through the SecYEG pore of varying diameter. This opening and closing of pore is driven by SecA in a two-way communication fashion. ATP hydrolysis in SecA triggers opening of pore, thus creating bigger channel for bulky amino acids to pass through. In closed state in ADP bound state or without any nucleotide, the channel stays in closed states and bulky residues, such as tryptophan cannot pass (see figure X). What gives this model directionality of translocation is interaction of SecA 2HF with bulky residues, which triggers ATP cycle and subsequently pore opening.

This model includes initiation and six further steps of translocation. Initiation includes recognition of signal peptide and its initial insertion through lateral gate in open state. Then in first step of translocation, pore is in open state and substrate can move forward and backwards with even probability. In second step, after ATP hydrolysis (in ADP bound state), pore closes and when bulky residue arrives directly to SecYEG entry or exit to the channel, it cannot pass. The important distinction is that when the bulky residue arrives at beginning of SecYEG pore it trigger double helix finger (2HF) in SecA, which cause next round of ATP hydrolysis and thus opening of pore. Bulky residue can now pass through the channel, due to increase in tunnel width, and diffusion continues until pore closes in ADP bound state again.

Diffusion of protein continues until new trigger reaction occurs and protein finally fully enters periplasm. In reverse, if a bulky residue enters periplasmic region after ATP hydrolysis cycle, it cannot diffuse back to the cytoplasmic region, because there is not bulky residue on another side, which can trigger next round of ATP hydrolysis. After whole sequence passes to periplasm, signal sequence gets cleaved by signal peptidase. [20] This turn still like behavior gives this model directionality and ATP efficiency.

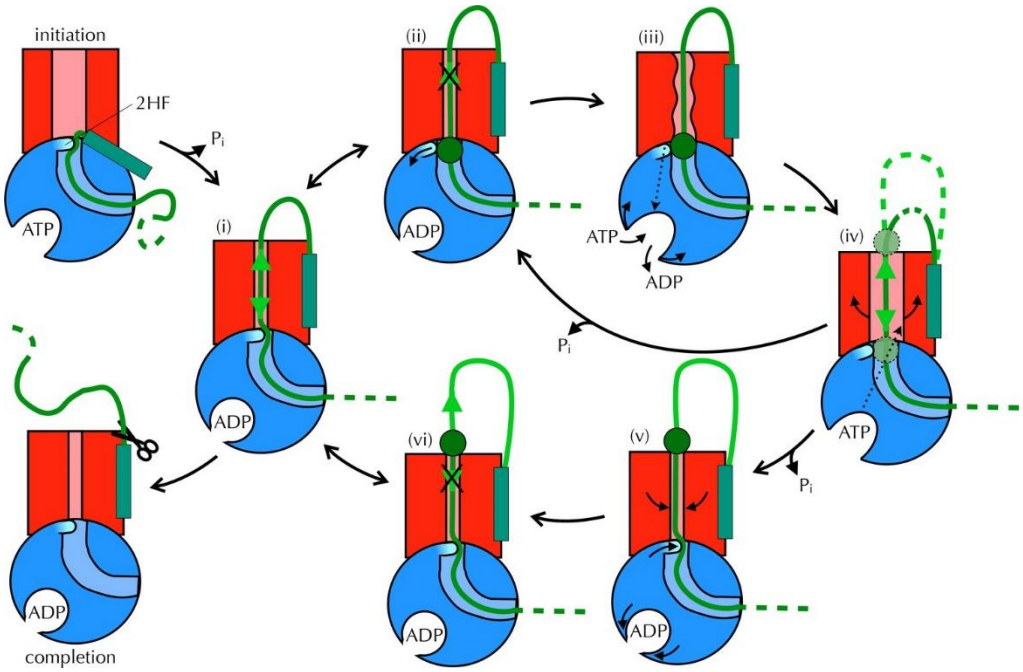

Figure 2. Graphical representation of Brownian Ratchet model and ATP driven opening and closing of pore. SecYEG is shown in red (lateral gate in light red), SecA in blue (substrate channel in light blue and 2HF in cyan and indicated), and substrate in green (with the signal sequence as a turquoise rectangle). Firstly, the initiation process (top left) caused structure changes between SecY and pre-protein, and put signal sequence to membrane near to secY complex. There are six important steps in translocation: initially, the substrate can diffuse backward or forwards (i) and finally bulky residue (green circle) arrive directly to SecYEG entry (ii) or exit (vi) of the channel. The bulky residue triggers the double helix finger (2HF), (iii) which causes opening of lateral gate (iv). Due to increasing radius of the tunnel, substrate with bulky residues diffuses through. Diffusion of protein continues until new trigger reaction occurs and protein finally fully enters periplasm. If bulky residue enters periplasmic region after the pore closes, it cannot diffuse back to the cytoplasmic region, (vi) because there is not bulky residue on another side, which can react reaction trigger. Due to this turn still mechanism substrate ratcheted to the periplasm region. After whole protein sequence passes through translocation complex, the signal sequence gets cleaved off and it remains near to SecYEG complex. [20]

1.3.2. Power-stroke model

For a long time, the main theory of protein translocation via Sec secretion pathway was the power-stroke model. First power stroke models were based on experiment using low concentration of ATP. [15][16] Researchers were able to detect translocation intermediates. Authors came to conclusion that translocation occurs in steps, each step requiring ATP hydrolysis. This model has several issues. One of them is theoretical step size. The step size of 16 nm, derived from known length of protein and measured number of ATP turnovers per one translocation would require extremely large rearrangements of SecA to push the substrate protein through pore. [22]

Other theories use some diffusion variable for translocation, but this would not create specific bands with fixed length on SDS-PAGE. Another experiment refuting power stroke model showed that changing of substrate sequence affects translocation rate in living bacteria.

Another important experiment supporting power-stroke model was based on protease essays.

The experiment showed that a large fragment of SecA becomes protease resistant when bound to SecYEG and pre-protein, but only in the presence of ATP or AMPPNP (non-hydrolysable analog of ATP), not ADP. [23]

This experiment was interpreted as an insertion of a 2HF domain of SecA bound to substrate deep into the pore. Thus, repetitive ATP driven conformational change would step-wise push substrate protein into the pore. Nevertheless, with wide range of all possible sequences that must be translocated, it is not simple to imagine how a single binding site could recognize all of them. Another experiment partly refuting this model showed, that the tip of the 2HF can be cross-linked to the channel of SecY without losing the capability to translocate. [24]

Different type of power-stroke model hypothesizes that translocation process might be driven by quaternary interactions between different SecA proteins. [23] SecA can adopt many different oligomeric states. It is clearly a dimer in solution, also many different dimer interfaces have been observed and studied. [25][26][27]

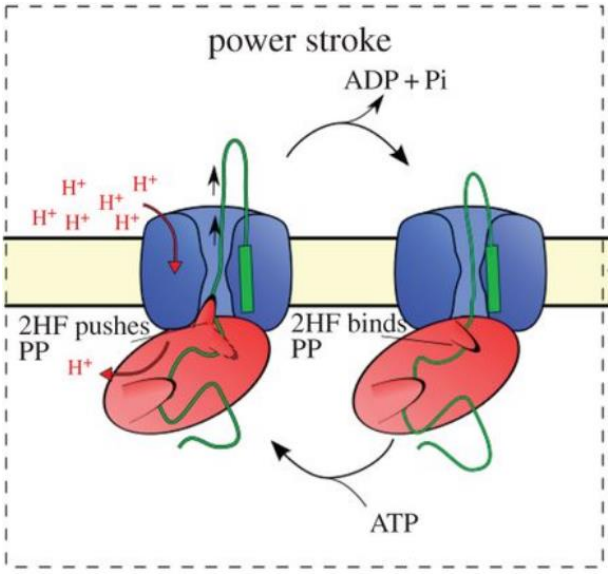

Figure 3. Power stroke model relies on 2HF pushing substrate through pore.

1.3.3. SecA dimerization model

The oligomeric state of SecA represents the controversial theme in this field study. It can adopt several different states: dimer in free solution and monomer while during the association with SecYEG, it forms yet again a number of different dimers. [17][18][19][28][29] The number of forms differs from experiment configurations.[30] Based on these observations, several possible models were proposed. First one is based on the idea of several SecA dimer configurations (with wide open, part-open and closed interface), transition between those is again activated by ATP hydrolysis. [31] Second hypothesis is based on alternate changing of state between monomer and dimer. [32] All of them are based on pushing process acting on the substrate against SecY pore.

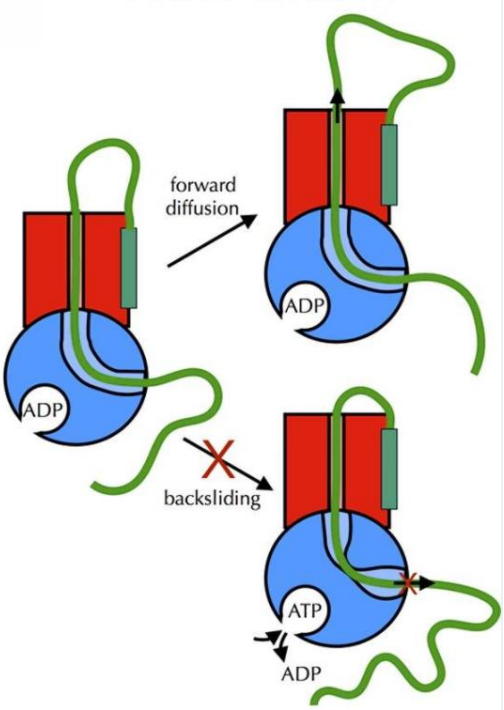

Figure 4. SecA dimerization model graphic representation. SecA allow forward diffusion of protein during ATP hydrolysis cycle and blocked backsliding by changing of SecA structure, while new ATP hydrolysis cycle begin.[20]

1.3.4. Other models

Several, more complicated, models propose some type of combination between pure power-stroke and monomer-dimer conversion. As the example of 'reciprocating piston', which again requires an extensive change of SecA state for translocation. Another support for this idea comes from experiment showing, that SecYEG has several structures, where SecA goes partially inside the translocon. This step could again push the substrate through pore. [33] Another proposed model requires different stoichiometries of SecYEG:SecA for each substrate type. [34] Other model relies on cytoplasmic chaperone SecB. Nevertheless, this model is by default not universal, because not all gram-negative bacteria has SecB. [34]

1.4. Molecular motors

Molecular motors drive all motion at the cellular level and are often viewed as microscopical biological machines. Most of them move in one direction through complex biological structures. The directionality of the motion is one of their most important aspects and it is often explained through a concept of ratchets (see below).

The most widely known group of molecular motors is also known as motor proteins. Most widely known are myosins, kinesins and dyneins. Myosins move along actin filaments and are responsible for muscle contraction. Kinesins and dyneins walk along microtubules, transporting cell's cargo or they assist in cell division. Exactly as in macro-world, there are several classes of molecular motors. Specifically protein translocation requires a linear form of molecular motor.

The microscopic environments of these molecular motors is close to thermal equilibrium. However, directed motion requires a source of non-equilibrium in order to comply with the second law of thermodynamics. To introduce non-equilibrium and bias to the system to get asymmetric potential and thus directionality in motion, some form of cost/fuel is usually required.

Most often it is ATP. ATP plays a role of fuel for all living cells and it stores chemical energy for directed motion. To transfer energy to molecular motor, ATP needs to get hydrolysed to ADP. During this process one phosphate group of ATP is removed by interaction with a water molecule, due to this ADP and hydrated inorganic phosphate P_i are created. Every ATP hydrolysis releases an energy of 14 k_BT. Where k_B is Boltzmann constant and T is temperature constant.

1.5. Brownian motion

Thermal equilibrium generates all sorts of motions in a system. Robert Brown described famous Brownian motion in 1827. He experimentally studied the zigzag motion of pollen grains suspended in water. [35] His system can be described by a stochastic equation for instantaneous velocity $v(t) = dx/dt$ of the particle of mass m:

$$m\frac{dv}{dt} = -\gamma v + \xi(t) \tag{1}$$

Where γ is the friction coefficient constant that describes the damping of the particle's motion from interaction with the immersed fluid, $\xi(t)$ is the fluctuating force form interaction with the fluid and modeled by a stationary Gaussian white noise with zero delta and average autocorrelation function:

$$\langle \xi(t) \rangle = 0$$
$$\langle \xi(t')\xi(t) \rangle = 2\Gamma\delta(t' - t) \quad (2)$$

Where coefficient Γ indicate the noise strength and relates to the temperature of the system and friction coefficient by applying to the condition of equipartition of energy

$$\frac{1}{2}m\left(v^2(\infty)\right) = \frac{1}{2}k_B T \quad (3)$$

has satisfied. In addition, there is a relation between the noise amplitude and the friction coefficient, which confirmed the relation between fluctuation and dissipation.

$$\Gamma = \gamma k_B T \quad (4)$$

As an addition, the average mean square displacement $\langle x^2(t) \rangle$ is proportional in the long time limit. Moreover, it can be described by

$$\langle x^2(t) \rangle \sim 2\left(\frac{k_B T}{\gamma}\right) t \, , where \; t \rightarrow \infty \quad (5)$$

This relation is valid for normal diffusion.

1.6. Brownian ratchet

Brownian ratchets are producing useful work, by rectifying microscopic fluctuations from their environment. The term "Brownian ratchet" consist of two elements required for the rectification of fluctuation to take place. First, one is fluctuating environment. This motion is often presented as random motion of small particles as the result of the multiple collisions with molecules around them. The second element is the presence of appropriate asymmetry in the system for the rectification of fluctuations. By which we can define a preferential direction of motion.

Based on these two conditions we can see one more important requirement for the Brownian ratchet implementation, system has to work out of thermal equilibrium, because of the second law of thermodynamics, which prevents the generation of directed motion out of unbiased fluctuations. [36]

1.6.1. Feynman ratchet model

Nobel laureate physicist Richard Feynman presented his famous thought experiment of Brownian ratchet model in a physics lecture at the California Institute of Technology on May 11, 1962. [37] Even though it was first analysed in 1912 by Polish physicist Marian Smoluchowski, therefor it is often called Feynman-Smoluchowski ratchet.

In this thought experiment, they created a special mechanism of mechanical rectifier. This machine was working at the same temperature and due to this, it violates the second law of thermodynamics. Therefor we can stand that microscopic analysis of this model is incorrect. Nevertheless, they pioneered and popularized the concept of molecular motors from the point of view of thermal and statistical physics as we know it today.

1.6.2. Stochastic ratchets

A similar concept to Brownian ratchet is a stochastic ratchet. It describes a process in which unoriented nonequilibrium fluctuations and a spatially anisotropic periodic potential conspire to produce directed motion. Again, transport in ratchet-like potentials provides a way for the transformation of nonequilibrium fluctuations into useful work. This potential is periodic but spatially anisotropic. Under the influence of the thermal fluctuations in the environment, the potential introduced to the particle randomly switches between two or more states. The output current generated by these systems is dependent on the spatial asymmetry of the potential, and importantly also on the statistics of the nonequilibrium fluctuations. [38]

2. Methods

2.1. Simulation Basics

Our simplified simulation of translocation is based on algorithms developed by M. G. Gauthier and G. W. Slater in 2008. Translocating polymer is here represented as a single Brownian particle, which can diffuse through pore in translocon complex. The simulation includes several different parameters influencing protein translocation. In parallel to calculation of diffusion properties of Brownian particle, we take into account also external driving force, internal entropy, and friction coefficients inside pore. [39]

To improve already published algorithms and to test turn still model of translocation, we added ATP hydrolysis cycle to the simulation. ATP hydrolysis by SecA triggers changes in conformation of pore and this opening and closing of pore significantly alters friction in pore, especially for large amino acid residues. To improve accuracy of calculation, we divided pore into segments of fixed length and varying radius. Using this approach, we can run calculation on different pre-proteins and test sequence dependent rate of translocation in turn still hypothesis introduced by Allen et al.

2.1.1. Integration of Stochastics simulation into Monte Carlo

The main idea of stochastic simulations requires understanding of a statistical model fully describing the system. Then we have to run this simulation multiple times and statistically analyze simulated stochastic events.

Imagine that we have a random variable X , with probability density function f(x) and we want to evaluate E(g(X)) for some function g(.). We already know that

$$E\big(g(X)\big) = \int_X g(x)f(x)\,dx \tag{6}$$

and was defined problem defined in integration. Integral might be straightforward for simple f(.) and g(.) computing. For solving we can realize $x_1,x_2,..x_n$ for X and random number also like series $g(x_1)$, $g(x_2)$,… g(X). And variance for random variable is finite by the laws of large numbers and for large n integral can be approximate as

$$E\big(g(x)\big) \cong \frac{1}{n}\sum_{i=1}^{n} g(x_i) \tag{7}$$

If they cannot simulate realization of X, they provide realizations $y_1, y_2, \ldots y_n$ of Y with same support as X with probability density function h(.) and they evaluate new formula

$$\begin{aligned} \big(g(X)\big) &= \int_X g(x)f(x)\,dx \\ &= \int_X \frac{g(x)f(x)}{h(x)}\,h(x)\,dx \end{aligned} \tag{8}$$

And approximate it into

$$E\big(g(X)\big) \cong \frac{1}{n}\sum_{i=1}^{n}\frac{g(y_i)f(y_i)}{h(y_i)} \tag{9}$$

And define that procedure as importance sampling, when the reasonable agreement between f(.) and h(.) exist.

For discrete-event simulation of time-evolution of biochemical events, we used exponential random quantities.

2.2. Forces in the simulation

Our simulation involves calculation of entropic force, external force, drag force and related constants. Before introduction of these forces, we need to explain related physical quantities. Firstly, we will speak about Kuhn length. In formulas, it is marked as l_K. Werner Kuhn created the Kuhn length. Kuhn segments of specific length represent it. They can be specified as freely joined segments between within each polymer. [40][41][42][43] In our simulation Kuhn length corresponds with a persistence length of the protein. The ideal state of wormlike chain and Kratky-Porod model will be with a bond length of 2 times persistence length. [44] Where persistence length is related to polymer local stiffness. Persistence length is usually in multiples of nanometers. [45][46][47]

Second used variable is the universal exponent. In formulas, it is written as γ. It is power-law correction parameter called the enhancement factor. Flory's scaling argument was derived from universal exponent with taking into account Flory theory, and estimates effects of the excluded volume on the average dimension of a polymer coil by mean field theory. It can be easily derived from equation $\nu = 3/(d + 2)$, where d is the dimensionality of the system. However, equation does not work for three-dimensional. In 3D this scaling value estimated to 0.588 [48]. Nevertheless, mostly in research scaling factor is presented in approximating version $\nu = 3/5$.

Thirdly, we will describe Lattice coordination number (LCN). In formulas, it is written as ž It is the number of conformations one would get for an ideal chain grown on a network with a lattice parameter. [39]

Next one is Three-dimensional Flory exponent for self-avoiding chains. In formulas it is written as ν. is the most elementary version of the Flory theory. Based on a covering of volume effect "swells" the chain compared with the dimension it would take were it not for the excluded volume effect. [49]

The last one is constant lattice parameter (CLT). It is written as lattice model replaces the continuous motion of the Brownian walker by a series of discrete jumps on a 1-dimensional system. Constant lattice parameter is the length of this discrete jump. [39]

The entropy S is associated with the number of possible conformations that a chain (ideal or real) can take is given by S = ln Z. where Z is the number of distinct conformations, or random walks (RW or SAW), of N steps. This partition function scales like

$$Z = \check{z}^{N} N^{\gamma-1} \tag{10}$$

where ž is called the lattice coordination number (LCN) and γ a universal exponent that depends only on the number of dimensions in the system. For three-dimensional systems, a lattice coordinator number is equal to 6 on a cubic lattice, but at the same moment, its self-avoiding counterpart has it equal to ž ≈ 4.68. [50] Our equation is the enhancement factor and has multiple specific conditions, which can be calculated. In such a case, γ is equal to 0.5, 0.69(1) and 1 for ideal, self-avoiding and rod-like chains, respectively [51]. Based on this equation we can calculate entropic force for the different case of protein location. Entropic force influence on the polymer chain between each side of the channel, with corresponds to the number of confirmation that chain can take on each side and use Self-Avoiding Walk (SAW) for applying on our Monte Carlo simulation.

$$\epsilon'_S = \begin{cases} \frac{(1-\gamma)}{2N_c} - \frac{\ln(\check{z})}{2}\frac{\alpha}{l_k} & if\ N_c \geq 1, N_T = 0 \\ \frac{(1-\gamma)}{2N_c} - \frac{(1-\gamma)}{2N_T} & if\ N_c \geq 1, N_T \geq 1 \\ \frac{\ln(\check{z})}{2}\frac{\alpha}{l_k} - \frac{(1-\gamma)}{2N_T} & if\ N_c = 0, N_T \geq 1 \\ 0 & if\ N_c = 0, N_T = 0 \end{cases} \tag{11}$$

Where Nsc = number of monomers in the sub chain, γ= universal exponent, α= Constant lattice parameter, l_K = Kuhn length, Ž = the lattice coordination number (LCN).

For calculation of charge density per unit length, we use the Henderson-Hasselbalch equation, which defines a degree of dissociation, α:

$$\mathrm{pH} = \mathrm{pKa} + \log\frac{\alpha}{1-\alpha} \tag{12}$$

$$\alpha = \frac{1}{10^{\mathrm{pKa}-\mathrm{pH}} + 1} \tag{13}$$

α get to our information what amount of the ionizable groups are deprotonated. It can be from 0 to 100 %. After we have to get pK_a for each group of amino acid, which we will calculate. In our study, we use data from Lehninger Principles of Biochemistry book. [52] Based on this data we can get pK_1, pK_2, and pK_R. Each of amino acids has pK_1, pK_2, but R group not present in each. By deprotonation, each of that group changes their charge. 1st group change from neutral to the negative charge, the 2nd group from positive to neutral charge and R group neutral to a positive charge.

After calculating of the amount of ionizable groups for each pK we can finally construct charge of selected amino acid by the equation:

$$(-1) * \alpha_{pK_1} + (+1) * \alpha_{pK_2} + (-1) * \alpha_{pK_R} = charge \tag{14}$$

In charge of amino acid, we can define charge density for protein inside the pore. After calculating charges of amino acids we find charge density by dividing on constant lattice parameter.

We also calculate membrane potential energy. In formulas it is written as *E*. Based on Cell Biology By The Numbers we can purpose constant value for it. [53] Membrane potential in E.Coli bacteria in average equal to 100mV and membrane potential energy will be near equal to:

$$E \approx 4k_bT \tag{15}$$

Where k_B is Boltzmann constant and T is a constant temperature of simulation.

Force resulting from the application of an external electric field. Interactions are counted only with the monomers inside of pore channel.

$$E' = \lambda\alpha^2 E/k_BT$$
$$\epsilon_F = N_P E'/2 \tag{16}$$

Where λ = Charge density per unit length, α = Constant lattice parameter, E = External electric force(C), k_B = Boltzman's constant, T = Temperature constant, Np =number of monomers in pore. Temperature constant is the room temperature in Kelvins.
Before introduction of drag force we need to calculate a velocity of the system. The calculation is based on a radius of gyration.

$$v_{sc} = \left|\frac{dR_g}{dt}\right| \sim \alpha \left|\frac{dN_{sc}^{\nu}}{dt}\right| \sim N_{sc}^{\nu-1}\alpha \left|\frac{dN_{sc}}{dt}\right| \sim N_{sc}^{\nu-1} v_P \tag{17}$$

where R_g is a radius of gyration, α is a constant lattice, N_{sc} is number in subchain and ν is three-dimensional Flory exponent.
Chain information is highly affected by physical constraints like excluded volume effects in the real polymer chain. Based on this idea, real chain resembles more a self-avoiding walk (SAW) than RW, because of swelling effect. Using scaling variables for estimating of excluded volume effects on both the energy and entropy of the chain. Also, this radius cover radius in which protein chain can take all possible conformation in the unfolded state. [54] Flory was able to quantify the swelling due to long-range interactions in self-avoiding chains. [48] His theory produces scaling equation.

$$R_g = \alpha N^{\nu}_{sc} \tag{18}$$

Where α= Constant lattice parameter, N_{sc} = number of monomers in the sub chain, ν = three-dimensional Flory exponent for self-avoiding chains.
There are two ways of calculation of drag force. Using Zimm or Rouse. In our simulation, they take us same calculation result and we solve to use Zimm dynamics. Drag force reacts on Radius of Gyration on each side of a pore. During each step of random walk, protein resides in the relaxed state. In our simulation, we calculate drag force for each section of SecYEG translocon: cis-,trans-,polymer- section.

$$F_{drag} \sim \begin{cases} \xi_0 \frac{l_K}{\alpha} N'^{\nu}_{sc} v_{sc} \; Zimm\; Dynamics \\ \xi_0 \frac{l_K}{\alpha} N'_{sc} v_{sc} \; Rouse\; Dynamics \end{cases} \tag{19}$$

$$F'_{drag} \sim \begin{cases} (\frac{l_K}{\alpha})^{2(1-\nu)} F_{drag} \; Zimm\; Dynamics \\ (\frac{l_K}{\alpha})^{1-\nu} F_{drag} \quad Rouse\; Dynamics \end{cases} \tag{20}$$

$$N' = N\alpha / l_K$$

Where N' is updated by Kuhn length.

Brownian time is calculated by $t_B = \alpha^2/2D_0$, where D_0 is free-solution diffusion coefficient of a random walker. Free-solution diffusion calculation is presented by $D_0 = k_BT/\xi_0$, where k_B is Boltzmann constant and ξ_0 is friction coefficient per monomer. Brownian time is needed to be rescaled to account for frictional coefficients. And final formula of Brownian time will be given by:

$$t'_B = (\xi^{\star}_{sc} + \xi_P)t_B \tag{21}$$

Where $\xi^{\star}_{sc}$ is global effective friction coefficient and ξ_P is friction coefficient inside pore caused by substrate and diameter of the pore.

2.3. Coefficients used in our simulation

In our simulation, we use several coefficients. They are friction coefficient of one average monomer on surrounding fluid, free-solution diffusion coefficient, and one of the most important are global effective friction coefficient and friction coefficient of the polymer section.

Friction coefficient per monomer based on the average radius of amino acids in selected protein chain and friction caused by him in surrounding fluid. Based on using of the Brownian particle as monomer and water as surrounding fluid we can calculate it using Stokes' law equation. [55]

$$\xi_0 = 6\pi\mu r \tag{22}$$

Free-solution diffusion coefficient is diffusion coefficient caused by a particle or random walker, due to thermal fluctuations. In our simulation, it is a monomer and will be used in Brownian time calculation, described below. The friction coefficient ξ is related to the free-solution diffusion coefficient D through the Einstein relation. And based on free-solution diffusion coefficient we can calculate using ξ_0.

$$\xi_0 \tag{23}$$

where k_BT represents the energy of the thermal bath in the surrounding environment. [55]

Before going to global effective friction coefficient and Friction coefficient of the polymer section, we need to introduce the Ladenburg approximation. While molecule goes through the pore it's velocity is decreased by friction forces. [56] In our simulation correction of

these coefficients based on particle falling in the cylindrical tube. It was inspired by work of Storm et al in 2005. [57]

$$\Gamma_P \simeq 2.105\alpha/d \tag{24}$$

Global Effective Friction Coefficient is calculated from to the chain velocity. Where Γ_z is numerical constant related to Ladenburg approximation.

$$\xi_{sc}^{\bigstar} = \begin{cases} \Gamma_z \xi_0 (N_C^{2\nu-1} + N_T^{2\nu-1}) \\ \Gamma_R \xi_0 (N_C^{\nu} + N_T^{\nu}) \end{cases} \tag{25}$$

The friction coefficient of the polymer section is coefficient of friction, which occurs in polymer section of simulation. Calculation is based on friction coefficient of the sphere falling through a cylindrical tube (Ladenburg aproximation), and friction coefficient of individual amino acids.

$$\xi_P = (1 + \Gamma_P)\xi_0 N_P \tag{26}$$

$$\Gamma_P \simeq 2.105\alpha/d \tag{27}$$

In our simulation, we improved the initial algorithm. Slater considered the whole pore as cylindrical tube with one diameter, here we propose to use full channel profile. Based on data given from "MOLEonline 2.0" we get channel radiuses of pore. [58] We defined segments of pore for each state, which is calculated from the total length of tunnel divided by the constant lattice parameter. The new formula is:

$$\xi_P = \sum_{i}^{pore\ length} (1 + \Gamma_{Pi})\xi_{0i} \tag{28}$$

$$\Gamma_P \simeq 2.105\alpha/d_i \tag{29}$$

Finally, we get friction coefficients of each segment.

2.4. Realization of Random-Walk model

Our simulation of one-dimensional polymer translocation is based on lattice random walk, it also includes external force of arbitrary magnitude. [39] While most of the simulations were done with continuous Brownian walker, we replace it with lattice model. Lattice model consists of discrete jumps with defined length on a one-dimensional lattice. In our simulation, it is constant lattice parameter.

Based on external force and entropic force we can define the probability of movement in each direction.

Firstly, we define the probability of no move in Monte Carlo step and given by

$$s(\in) = \frac{\coth \in}{\in} - \operatorname{csch}^2 \in \tag{30}$$

Based on this probability we can calculate a positive and negative probability. In our simulation the formula is given by

$$p_{\pm}(\in) = \frac{1 - s(\in)}{1 + e^{\mp 2\in}} \tag{31}$$

And Monte Carlo time step will be calculated by

$$\Delta t(\in) = \frac{\big(1 - s(\in)\big)\tanh(\epsilon)}{\epsilon}\, t_B \tag{32}$$

Simulation calculates p_-, p_+ and s and picks a random number from Gaussian distribution centered on those values. Based on the result from random number generator we can say that chain either moves or stays in the same position and it will increase time translocation byΔt. The repetitive process repeats until the chain either escapes the pore to either a cis or trans side. As result, we assemble multiple trajectories and process them to get distribution of first passage times.

3.1. Pore dimensions in three SecYEG states

As we know from available crystal structures, SecYEG can adopt three different states, closed, part open and open.

Closed state is presented by Methanococcus jannaschii complex (PDB code 1RHZ). Part-open state is presented by Thermotoga maritima complex (PDB code 3DIN). Open state is presented also by Geobacillus thermodenitrificans complex (PDB code 5EUL). Allen et al. in 2016 evaluated these states by measurement of opening of lateral gate. In this work, we used web application "MOLEonline 2.0" to get exact dimensions of tunnels in each of three above mentioned states. [58] This can search for voids and tunnels specifically in transmembrane regions. It calculates several physico-chemical properties of tunnels divided into small segments. Main output we used in this study was channel radiuses. "Mole online" calculates tunnel radius in three ways, classical radius, free radius and Bradius. Radius is maximum possible radius of sphere, which fits in between all atoms of given structure. Free Radius is channel radius calculated purely from amino acids backbone. BRadius represent radius with additional RMSF (Root mean square fluctuation) calculated from B-factors of residues within individual layers.

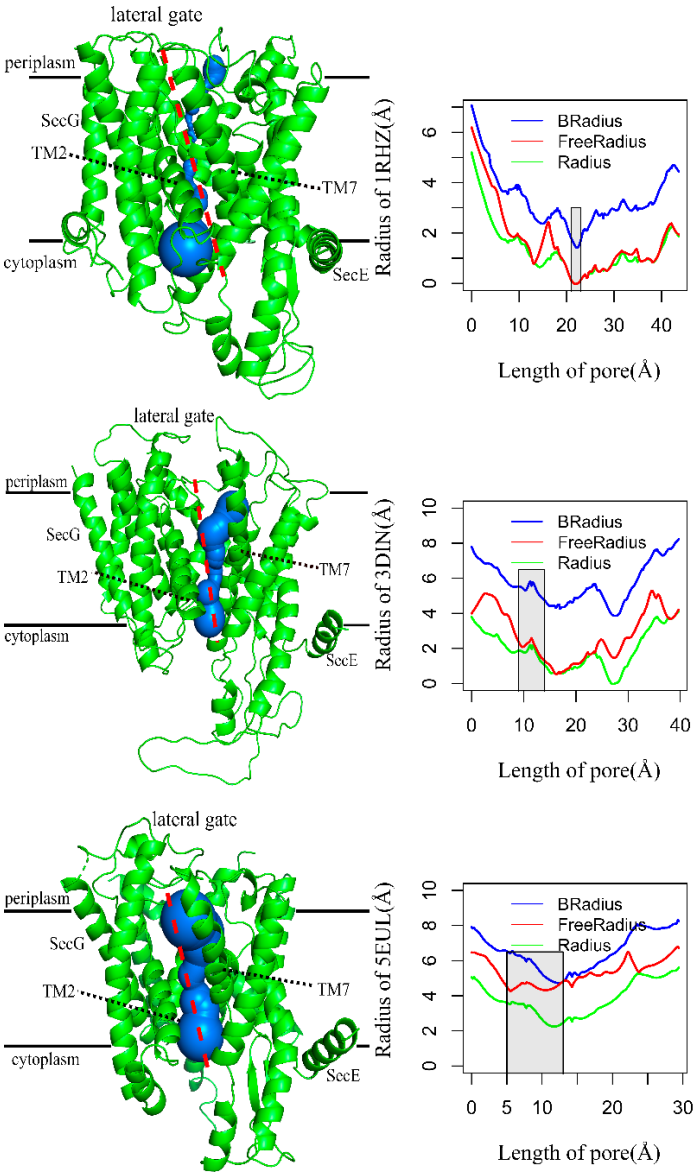

Figure 5. Repersentation of structure and possible tunnel inside Methanococcus jannaschii complex (PDB code: 1RHZ), Thermotoga maritima complex (PDB code: 3DIN) , Geobacillus thermodenitrificans complex (PDB code: 5EUL) SecYEG structure and graph with Radius, Free Radiuses and Bradiuses for each structure.

As result, we evaluated all three possible structures and derived different representations of tunnel radiuses for each states (see Figure 5). These one-dimensional representations of

entire tunnel surface of pore were later used to introduce friction in pore in our Monte Carlo simulation. Using this calculation, we can extend previously published description of dependence of SecYEG conformation on ATP hydrolysis in SecA.Data for local minimums described in Table 1.

Table 1. Local minima for SecYEG structures..

Structure name	Radius	Free Radius	Bradius
Methanococcus jannaschii complex (PDB code: 1RHZ)	0 Å	0 Å	1,416 Å
Thermotoga maritima complex (PDB code 3DIN)	0 Å	0.538 Å	3.858 Å
Geobacillus thermodenitrificans complex (PDB code 5EUL)	2.243 Å	4.254 Å	4.72

Monte Carlo simulation algorithm was created using Anaconda Software with Python 3.6 and multiple other packages: NumPy, Pandas, MpMath. [59][60][61][62] Code is available in supplement.

The software consists of initialization, and several subprograms, which are being updated in a while cycle. For detailed view see Figure 2.

The termination condition for the while cycle is either full translocation of protein to periplasmic site (first passage) or evaluation of 1.000.000 steps. At each 1000 steps, data are stored to a data log file. Example log file is available supplement.

More detailed description of simulation package is to be found in Figure 3. and Figure 4. These figures show, how two important variables are calculated, namely probability of movement, and Time of movement.

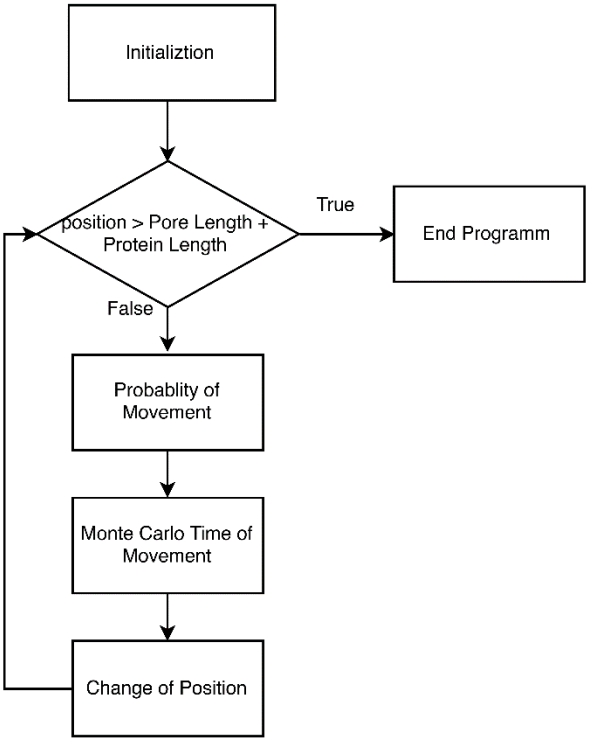

Figure 6. Simulation program flow chart. In the initialization part, program imports all important constants and information about the structure of pore. More precisely pore length and pore width in segments. While cycle calculates probability of step forward, back or no motion and it also calculates Monte Carlo time of each step until the condition is satisfied.

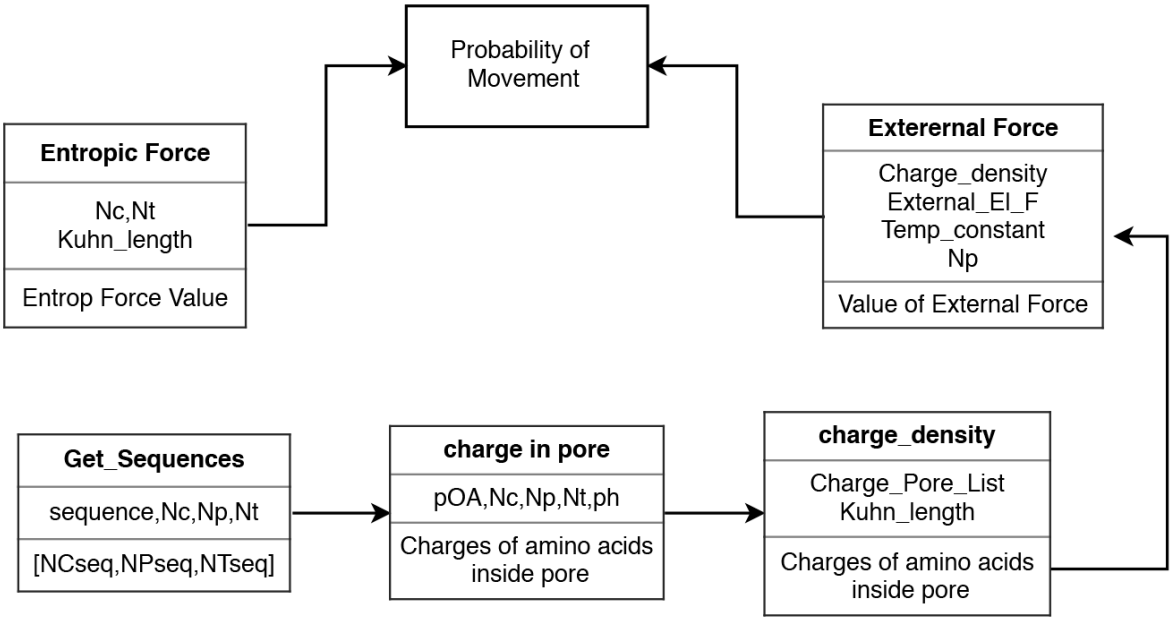

Figure 7. Probability of movement calculated from final energy. Probability of movement calculated by final energy consisting of a sum of External Force and Entropic Force. To get External Force, charge density of amino acids in pore was calculated by calculating their charge per pore width segment. In final charge density sum for each segment in pore. Entropic force is calculated from the amount of amino acids on cytoplasmic and periplasmic site.

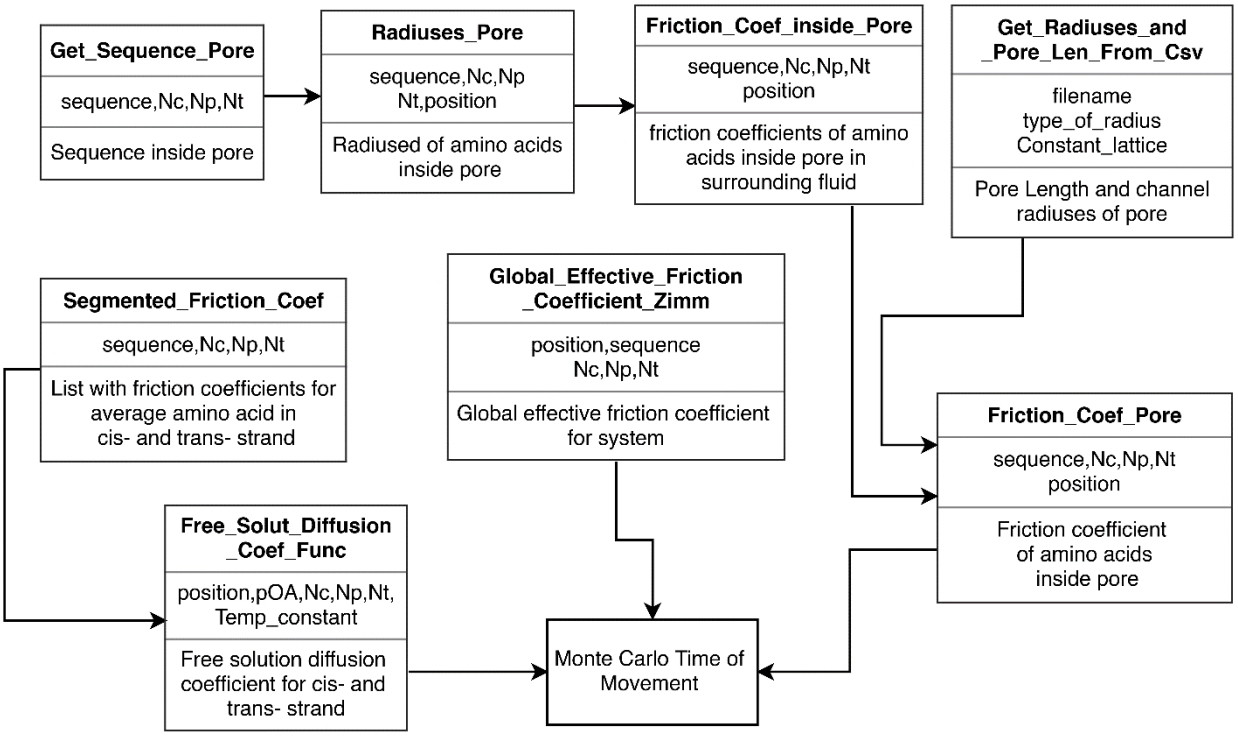

Figure 8. Calculation of Monte Carlo time. This block of code uses probability of no movement given from probability of movement block. For calculating of free-solution diffusion coefficient we need to calculate friction coefficient of average amino acids in water in cis and trans site. Solution friction coefficient of each amino acid inside the pore was calculated by Friction_Coef_inside_Pore function. Friction_Coef_Pore calculate friction coefficient of amino acids with respect to pore width segments obtained from Get_Radiuses_and_Pore_Len_From_Csv function. As a result we get the sum of friction coefficient for each amino acid in pore. For global effective friction coefficient, we use friction coefficient of average amino acids in water in cis and trans site and return a global effective friction coefficient. At a final step Brownian time is calculated and time of Monte Carlo step is calculated.

At the last part of the while cycle, a random generator using Gaussian distribution centering on calculated probabilities helps to define the direction of movement. This new position is then added to the simulation and while cycle repeats until the successful end of translocation or until the set limit is reached.

3.3. Calculated translocation trajectories

Final output of our simulation is a trajectory of translocation of protein through a translocon pore. After assembling more trajectories, we could extract statistics of first passage times and their sequence (and model) dependence. Here we show few example trajectories for each state in Figure 5.

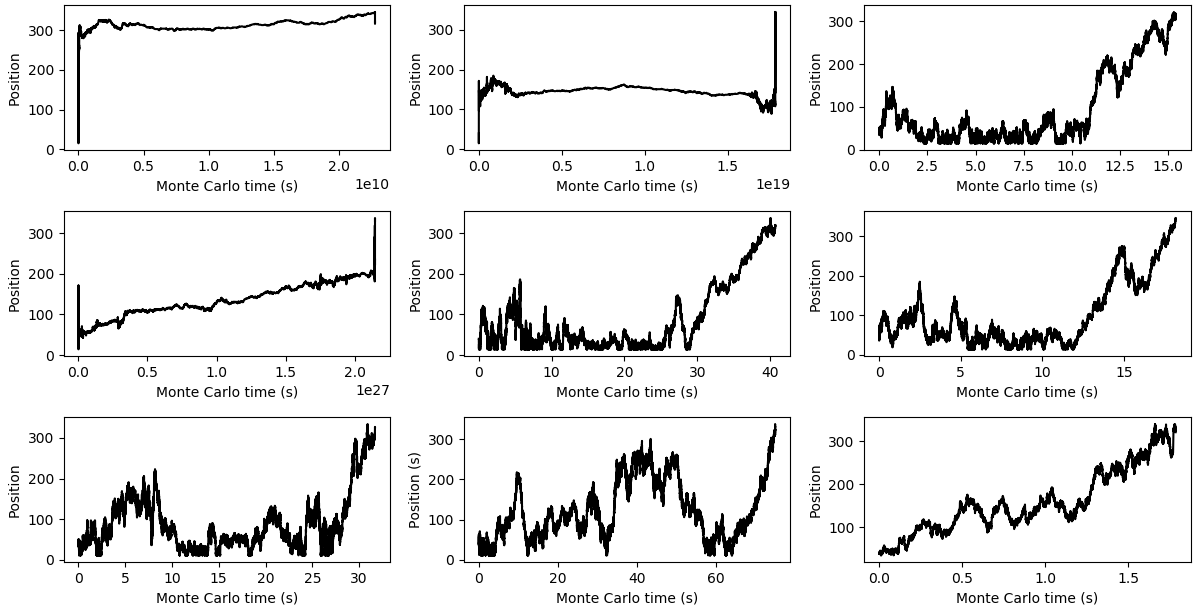

Figure 9. Example trajectories of translocation of protein through SecYEG complex using our program. Position is position from the beginning of the pore. In 1RHZ structure, translocation time calculated from normal radius became extremely long, similar behavior was seen for 3DIN structure. This behavior is more pronounced, when region rich on large amino acids reaches the pore.

Based on the data from our simulation, we identified 4 amino acids residues causing extreme increase in translocation time (see Figure 10.) These amino acids are tryptophan, phenylalanine, tyrosine, arginine, those with largest radiuses (see Figure 11.). For further information on these calculations and plots showing other quantities needed to obtain full trajectory see supplementary figures n. iii and iv.

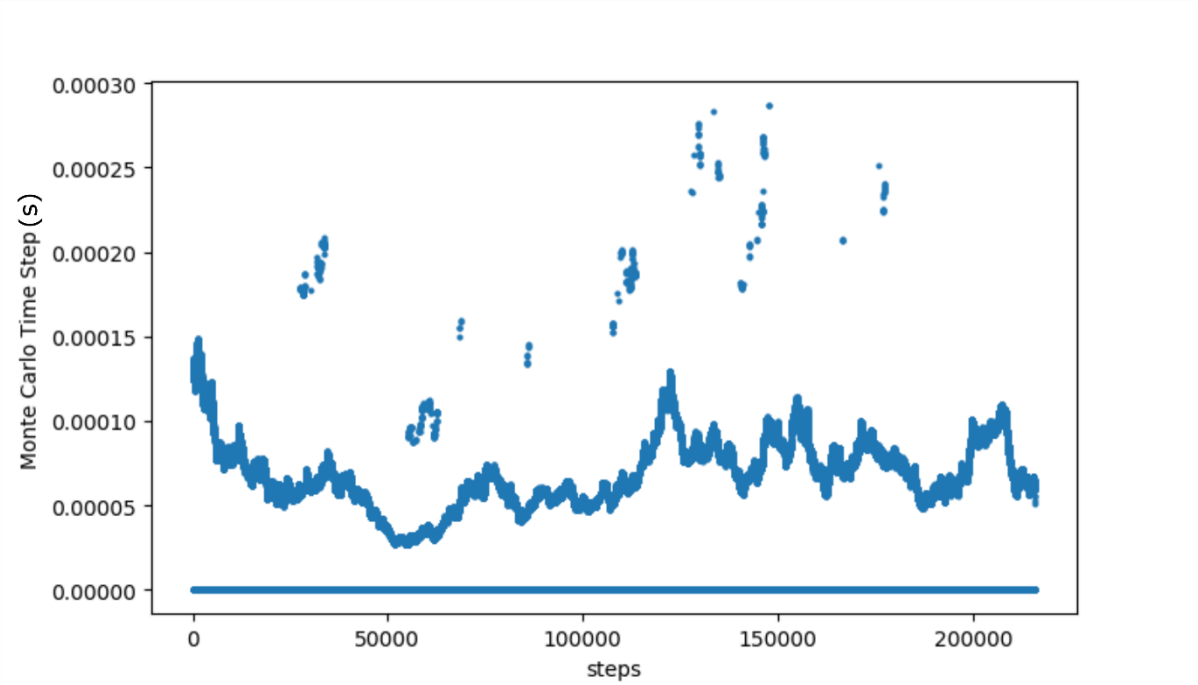

Figure 10. Monte Carlo time steps during the translocation. The increase in time, when large amino-acids get to pore is apparent.

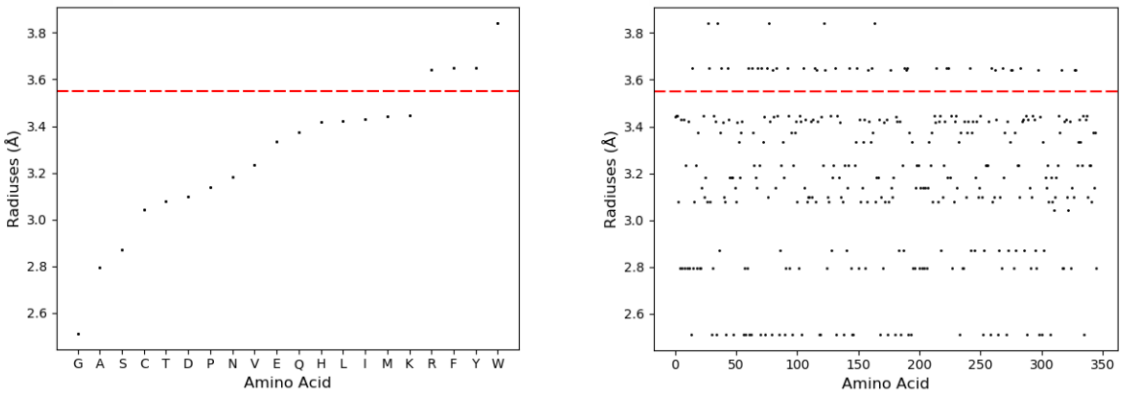

Figure 11. Left: radiuses of common amino acid residues, see residues above red line, which cause slowdown in pore. Right: Radius of amino acid residues in the proOMPA sequence. The separation line is at 3.55 Angstroms.

3.4. Structure analysis of 3DIN

The main idea of this thesis is to build a set of tools suitable to help with verification of functional models of translocation. For this we need not only a working hypothesis, but also a smart design of future experiments. We obtained molecular dynamics data used in (Allen at All.)[5] work for part-open state of 3DIN structure. Based on this we calculated Protein Structure Networks with Bio3D and performed structural cross-correlation analysis assisted with principal component analysis (for details see supplementary figures no i and no ii). [63][64][65]

At the next step was produced residue cross correlation analysis for whole simulation.(See Figure 12.) And with structure net.(See. Figure 13.)

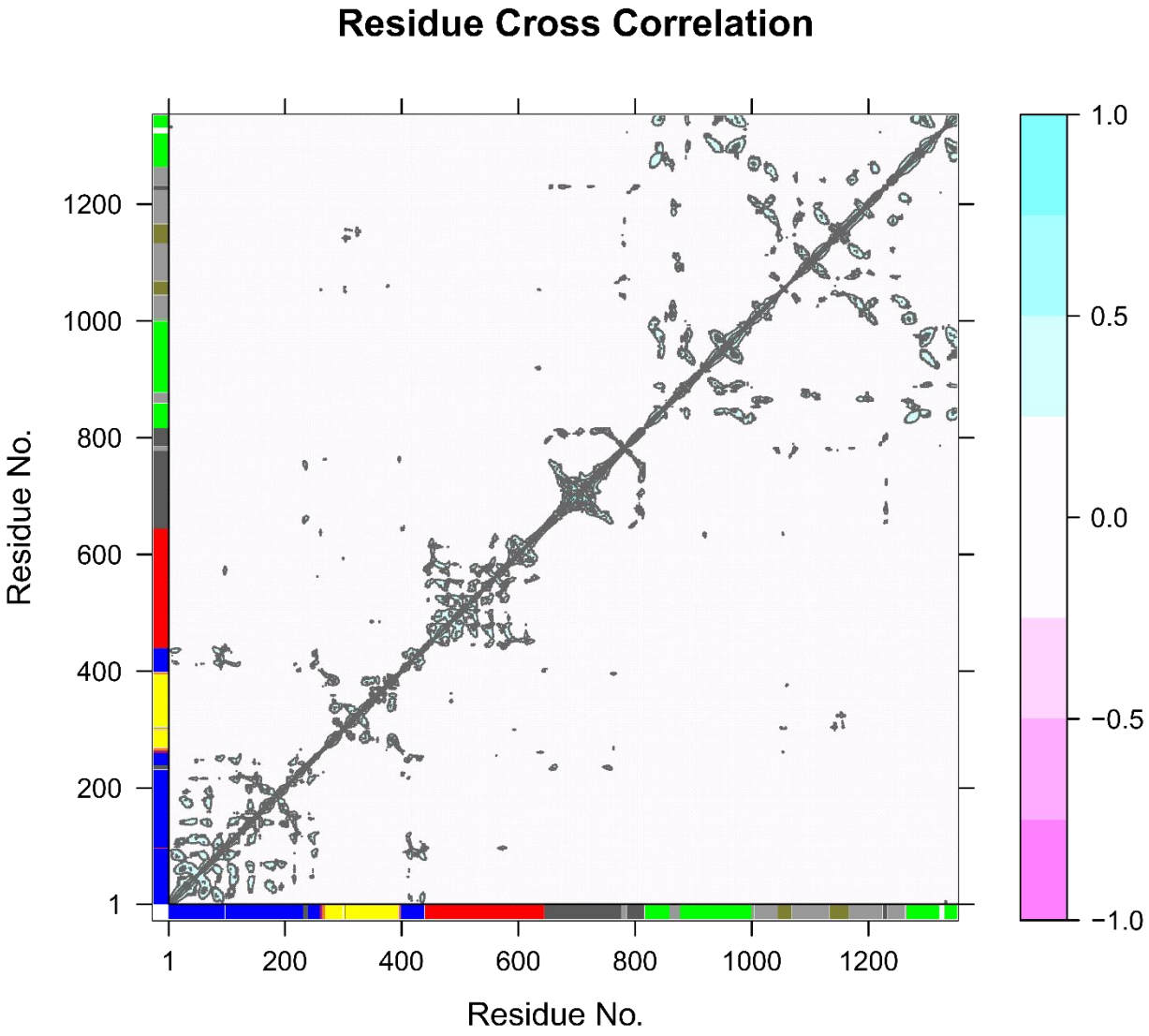

Figure 12. Residue Cross Correlation Matrix. See corresponding color scheme and structure in ***Figure 13****.*

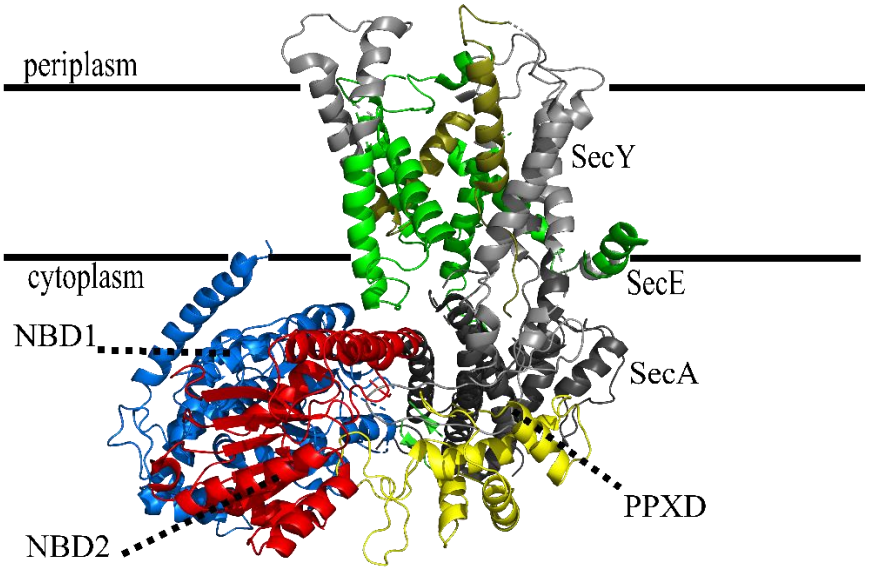

Figure 13. Net structure with domains from 1 to 9. Colors related to groups which used in residue cross correlation matrix. We can see division of NBD domain into 2 fluctuations groups NBD1 and NBD2. PPXD also divided into 2 segments. SecYEG was divided into 3 different group and SecG was not joined to any group.

4. Discussion

The next step will be running more simulation to get more data for statistical analysis. In our simulation, we estimated radius of amino acids from hard sphere approximation, better representation would probably yield results that are more accurate. Parallelization of code would improve running time of simulation.

5. Conclusion

We have successfully prepared a package to simulate translocation of substrate through SecYEG pore and test various models of protein translocation. Even though, the software package is ready, now we need to acquire more data to get reliable statistics and extract accurate representation of first passage time distributions.

Nevertheless, we have already identified residues in translocated sequences, which slow down translocation in closed state. They are tryptophan (W), tyrosine (Y), phenylalanine (F), arginine (R), and they slow down translocation time step 2 times for tyrosine, phenylalanine, arginine and 3 times for tryptophan.

In this work, we started with pOA sequence, but more sequences need to be tested and compared to experimental data. In fact it might be better to build artificial sequences, e.g. with repeated stretches of large and small amino acids to test turn the still model under

different circumstances. This could later lead to finding such sequences, which would be able to block SecYEG in particular states and thus weaken Gramm-negative bacteria to e.g. strength then the impact of antibiotics.

Even though, we have shown data testing turn still model based on Brownian model of translocation, this simulation package could be easily modified to test two remaining major theories, i.e. power stroke and SecA dimerization.

On top of development of Monte Carlo simulation in 1D to test already existing theories, we wanted to make a bridge between a simulation and future experiments. The usual scientific method involves testing a theory, its potential improvement and further experimental verification of this new theory by clever design of new experiments. To assist with this design, we have introduced Protein Structure Networks with Bio3D. [63][64][65] Calculation based on molecular dynamics which could help to find such parts of molecular motions which show correlated motion during simulation. After fluorescent labelling at these sites, novel florescence correlation experiments could be designed. Such potential labelling sites have been identified in this work and will be used in future work in the laboratory of single molecule fluorescence.

6. References

1. Ryan, Kenneth J.; Ray, C. George; Ahmad, Nafees; Drew, W. Lawrence; Lagunoff, Michael; Pottinger, Paul; Reller, L. Barth; Sterling, Charles R. (2014). "Pathogenesis of Bacterial Infections". Sherris Medical Microbiology (6th ed.). New York: McGraw Hill Education. pp. 391–406. ISBN 978-0-07-181826-1.
2. Mcfall-Ngai, M. (2007). Care for the community. Nature, 445(7124), 153-153. doi:10.1038/445153a
3. Milo, R. (2013). What is the total number of protein molecules per cell volume? A call to rethink some published values. BioEssays, 35(12), 1050-1055. doi:10.1002/bies.201300066
4. Liu, J. K., O'Brien, E. J., Lerman, J. A., Zengler, K., Palsson, B. O., & Feist, A. M. (2014). Reconstruction and modeling protein translocation and compartmentalization in Escherichia coli at the genome-scale. BMC Systems Biology, 8(1). doi:10.1186/s12918-014-0110-6
5. Allen, W. J., Corey, R. A., Oatley, P., Sessions, R. B., Baldwin, S. A., Radford, S. E., . . . Collinson, I. (2016). Two-way communication between SecY and SecA suggests a Brownian ratchet mechanism for protein translocation. ELife, 5. doi:10.7554/elife.15598
6. Berg, B. V., Clemons, W. M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C., & Rapoport, T. A. (2003). X-ray structure of a protein-conducting channel. Nature, 427(6969), 36-44. doi:10.1038/nature02218
7. R.Lill, K.Cunningham, L.A. Brundage, K. Ito, D. Oliver, W Wickner, (1989), SecA protein hydrolyzes ATP and is an essential component of the protein translocation ATPase of Escherichia coli. PMCID: PMC400897
8. Brundage, L., Hendrick, J. P., Schiebel, E., Driessen, A. J., & Wickner, W. (1990). The purified E. coli integral membrane protein SecYE is sufficient for reconstitution of SecA-dependent precursor protein translocation. Cell, 62(4), 649-657. doi:10.1016/0092-8674(90)90111-q
9. Akimaru, J., Matsuyama, S., Tokuda, H., & Mizushima, S. (1991). Reconstitution of a protein translocation system containing purified SecY, SecE, and SecA from Escherichia coli. Proceedings of the National Academy of Sciences, 88(15), 6545-6549. doi:10.1073/pnas.88.15.6545

10. Hunt, J. F. (2002). Nucleotide Control of Interdomain Interactions in the Conformational Reaction Cycle of SecA. Science, 297(5589), 2018-2026. doi:10.1126/science.1074424
11. Zimmer, J., Nam, Y., & Rapoport, T. A. (2008). Structure of a complex of the ATPase SecA and the protein-translocation channel. Nature, 455(7215), 936-943. doi:10.1038/nature07335
12. Bauer, B. W., & Rapoport, T. A. (2009). Mapping polypeptide interactions of the SecA ATPase during translocation. Proceedings of the National Academy of Sciences, 106(49), 20800-20805. doi:10.1073/pnas.0910550106
13. Gold, V. M., Whitehouse, S., Robson, A., & Collinson, I. (2013). The dynamic action of SecA during the initiation of protein translocation. Biochemical Journal, 449(3), 695-705. doi:10.1042/bj20121314
14. Zimmer, J., Nam, Y., & Rapoport, T. A. (2008). Structure of a complex of the ATPase SecA and the protein-translocation channel. Nature, 455(7215), 936-943. doi:10.1038/nature07335
15. Schiebel, E., Driessen, A. J., Hartl, F., & Wickner, W. (1991). ΔμH and ATP function at different steps of the catalytic cycle of preprotein translocase. *Cell, 64*(5), 927-939. doi:10.1016/0092-8674(91)90317-r
16. Tani K, Shiozuka K, Tokuda H, Mizushima S. (1989) In vitro analysis of the process of translocation of OmpA across the Escherichia coli cytoplasmic membrane. A translocation intermediate accumulates transiently in the absence of the proton motive force. J. Biol. Chem. 264, 18 582–18 588. PMID: 2553715
17. Wowor A.J., Yan Y., Auclair S.M., Yu D., Zhang J., May E.R., (2014), Gross M.L., Kendall D.A., Cole J.L. Analysis of SecA dimerization in solution . Biochemistry. doi: 10.1021/bi500348p
18. Driessen, A. J. (1993). SecA, the peripheral subunit of the Escherichia coli precursor protein translocase, is functional as a dimer. Biochemistry, 32(48), 13190-13197. doi:10.1021/bi00211a030
19. Gold, V. A., Robson, A., Clarke, A. R., & Collinson, I. (2007). Allosteric Regulation of SecA. Journal of Biological Chemistry, 282(24), 17424-17432. doi:10.1074/jbc.m702066200

20. Allen, W. J., Corey, R. A., Oatley, P., Sessions, R. B., Baldwin, S. A., Radford, S. E., . . . Collinson, I. (2016). Two-way communication between SecY and SecA suggests a Brownian ratchet mechanism for protein translocation. ELife, 5. doi:10.7554/elife.15598
21. Ait-Haddou, R., & Herzog, W. (2003). Brownian Ratchet Models of Molecular Motors. Cell Biochemistry and Biophysics, 38(2), 191-214. doi:10.1385/cbb:38:2:191
22. Gold, V. M., Whitehouse, S., Robson, A., & Collinson, I. (2013). The dynamic action of SecA during the initiation of protein translocation. Biochemical Journal, 449(3), 695-705. doi:10.1042/bj20121314
23. Krek, W. (1994). Negative regulation of the growth-promoting transcription factor E2F-1 by a stably bound cyclin A-dependent protein kinase. Cell, 78(1), 161-172. doi:10.1016/0092-8674(94)90582-7
24. Whitehouse, S., Gold, V. A., Robson, A., Allen, W. J., Sessions, R. B., & Collinson, I. (2012). Mobility of the SecA 2-helix-finger is not essential for polypeptide translocation via the SecYEG complex. The Journal of Cell Biology, 199(6), 919-929. doi:10.1083/jcb.201205191
25. Gouridis, G., Karamanou, S., Sardis, M., Schärer, M., Capitani, G., & Economou, A. (2013). Quaternary Dynamics of the SecA Motor Drive Translocase Catalysis. Molecular Cell, 52(5), 655-666. doi:10.1016/j.molcel.2013.10.036
26. Kusters, I., & Driessen, A. J. (2011). SecA, a remarkable nanomachine. Cellular and Molecular Life Sciences, 68(12), 2053-2066. doi:10.1007/s00018-011-0681-y
27. Singh, R., Kraft, C., Jaiswal, R., Sejwal, K., Kasaragod, V. B., Kuper, J., . . . Bhushan, S. (2014). Cryo-electron Microscopic Structure of SecA Protein Bound to the 70S Ribosome. Journal of Biological Chemistry, 289(10), 7190-7199. doi:10.1074/jbc.m113.506634
28. Or, E. (2002). Dissociation of the dimeric SecA ATPase during protein translocation across the bacterial membrane. The EMBO Journal, 21(17), 4470-4479. doi:10.1093/emboj/cdf471
29. Or, E., Boyd, D., Gon, S., Beckwith, J., & Rapoport, T. (2004). The Bacterial ATPase SecA Functions as a Monomer in Protein Translocation. Journal of Biological Chemistry, 280(10), 9097-9105. doi:10.1074/jbc.m413947200

30. Gouridis, G., Karamanou, S., Gelis, I., Kalodimos, C. G., & Economou, A. (2009). Signal peptides are allosteric activators of the protein translocase. Nature, 462(7271), 363-367. doi:10.1038/nature08559
31. Sharma, V., Arockiasamy, A., Ronning, D. R., Savva, C. G., Holzenburg, A., Braunstein, M., . . . Sacchettini, J. C. (2003). Crystal structure of Mycobacterium tuberculosis SecA, a preprotein translocating ATPase. Proceedings of the National Academy of Sciences, 100(5), 2243-2248. doi:10.1073/pnas.0538077100
32. Benach, J., Chou, Y., Fak, J. J., Itkin, A., Nicolae, D. D., Smith, P. C., . . . Hunt, J. F. (2002). Phospholipid-induced Monomerization and Signal-peptide-induced Oligomerization of SecA. Journal of Biological Chemistry, 278(6), 3628-3638. doi:10.1074/jbc.m205992200
33. Kusters, I., & Driessen, A. J. (2011). SecA, a remarkable nanomachine. Cellular and Molecular Life Sciences, 68(12), 2053-2066. doi:10.1007/s00018-011-0681-y
34. Mao, C., Cheadle, C. E., Hardy, S. J., Lilly, A. A., Suo, Y., Gari, R. R., . . . Randall, L. L. (2013). Stoichiometry of SecYEG in the active translocase of Escherichia coli varies with precursor species. Proceedings of the National Academy of Sciences, 110(29), 11815-11820. doi:10.1073/pnas.1303289110
35. Newell, L. C. (1923). Robert Brown and the Discovery of the Brownian Movement. Industrial & Engineering Chemistry, 15(12), 1279-1281. doi:10.1021/ie50168a029
36. Cubero, D., & Renzoni, F. (2016). Brownian ratchets: From statistical physics to bio and nano-motors. Cambridge, United Kingdom: Cambridge University Press.
37. Feynman, R. P., Leighton, R. B., & Sands, M. (1963). The Feynman lectures on physics, vol. 1. Mass.: Addison-Wesley.
38. Doering, C. R., Horsthemke, W., & Riordan, J. (1994). Nonequilibrium fluctuation-induced transport. Physical Review Letters, 72(19), 2984-2987. doi:10.1103/physrevlett.72.2984
39. Gauthier, M. G., & Slater, G. W. (2008). A Monte Carlo algorithm to study polymer translocation through nanopores. I. Theory and numerical approach. The Journal of Chemical Physics, 128(6), 065103. doi:10.1063/1.2826339
40. Flory, P. J. (1953). Principles of polymer chemistry. Ithaca: Cornell University. ISBN 0-8014-0134-8
41. Flory, P.J. (1969). Statistical Mechanics of Chain Molecules. NY, Interscience. ISBN 1-56990-019-1

42. Rubinstein, M., & Colby, R. H. (2003). Polymer physics. Oxford: Oxford University Press. ISBN 0-19-852059-X
43. Doi, M., & Edwards, S. F. (1988). The Theory of Polymer Dynamics. Oxford: Clarendon Press. ISBN 0198520336
44. Teraoka, I. (2002). Polymer solutions: An introduction to physical properties. p45. New York: Wiley. ISBN 0-471-38929-3
45. Stirnemann, G., Giganti, D., Fernandez, J. M., & Berne, B. J. (2013). Elasticity, structure, and relaxation of extended proteins under force. Proceedings of the National Academy of Sciences, 110(10), 3847-3852. doi:10.1073/pnas.1300596110
46. Schlierf, M., Li, H., & Fernandez, J. M. (2004). The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. Proceedings of the National Academy of Sciences, 101(19), 7299-7304. doi:10.1073/pnas.0400033101
47. Walther, K. A., Grater, F., Dougan, L., Badilla, C. L., Berne, B. J., & Fernandez, J. M. (2007). Signatures of hydrophobic collapse in extended proteins captured with force spectroscopy. Proceedings of the National Academy of Sciences, 104(19), 7916-7921. doi:10.1073/pnas.0702179104
48. Guillou, J. L., & Zinn-Justin, J. (1990). Critical Exponents for the n-Vector Model in Three Dimensions from Field Theory. Current Physics–Sources and Comments Large-Order Behaviour of Perturbation Theory, 527-530. doi:10.1016/b978-0-444-88597-5.50074-0
49. Teraoka, I. (2002). Polymer solutions: An introduction to physical properties. p36. New York: Wiley. ISBN 0-471-38929-3
50. Fisher, M. E., & Hiley, B. J. (1961). Configuration and Free Energy of a Polymer Molecule with Solvent Interaction. The Journal of Chemical Physics, 34(4), 1253-1267. doi:10.1063/1.1731729
51. Eisenriegler, E., Kremer, K., & Binder, K. (1982). Adsorption of polymer chains at surfaces: Scaling and Monte Carlo analyses. The Journal of Chemical Physics, 77(12), 6296-6320. doi:10.1063/1.443835
52. Nelson, D. L., & Cox, M. M. (2008). Lehninger principles of biochemistry. New York: W.H. Freeman and Company. p.78
53. CELL BIOLOGY by the numbers. (2016). Seibutsu Butsuri, 56(6), 352-352. doi:10.2142/biophys.56.352
54. Teraoka, I. (2002). Polymer solutions: An introduction to physical properties. New York: Wiley. ISBN 0-471-38929-3

55. Knacke, R. F., & Reif, F. (1965). Solutions to problems to accompany F. Reifs Fundamentals of statistical and thermal physics. New York: McGraw-Hill.
56. Altmann, H., Büchner, F., Letterer, E., Roulet, F., Baudisch, W., Mölbert, E., . . . Bader, R. (1977). Handbuch der allgemeinen Pathologie. Mikrozirkulation = Microcirculation
57. Storm, A. J., Storm, C., Chen, J., Zandbergen, H., Joanny, J., & Dekker, C. (2005). Fast DNA Translocation through a Solid-State Nanopore. Nano Letters, 5(7), 1193-1197. doi:10.1021/nl048030d.
58. Berka K, Hanák O, Sehnal D, Banáš P, Navrátilová V, Jaiswal D, Ionescu CM, Svobodová Vařekova R, Koča J, Otyepka M: MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels. Nucleic Acids Res., 40(W1), W222-W227, 2012.
59. Wilkinson, D. J. (2012). Stochastic modelling for systems biology. Boca Raton: Taylor & Francis.
60. Travis E, Oliphant,(2006), A guide to NumPy, USA: Trelgol Publishing
61. Wes McKinney. Data Structures for Statistical Computing in Python, Proceedings of the 9th Python in Science Conference, 51-56 (2010)
62. Fredrik Johansson and others. mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 0.18), December 2013.
63. Grant, B. J., Rodrigues, A. P., Elsawy, K. M., Mccammon, J. A., & Caves, L. S. (2006). Bio3d: An R package for the comparative analysis of protein structures. Bioinformatics, 22(21), 2695-2696. doi:10.1093/bioinformatics/btl461
64. Skjærven, L., Yao, X., Scarabelli, G., & Grant, B. J. (2014). Integrating protein structural dynamics and evolutionary analysis with Bio3D. BMC Bioinformatics, 15(1). doi:10.1186/s12859-014-0399-6
65. Skjærven, L., Jariwala, S., Yao, X., & Grant, B. J. (2016). Online interactive analysis of protein structure ensembles with Bio3D-web. Bioinformatics. doi:10.1093/bioinformatics/btw482

7. Supplement

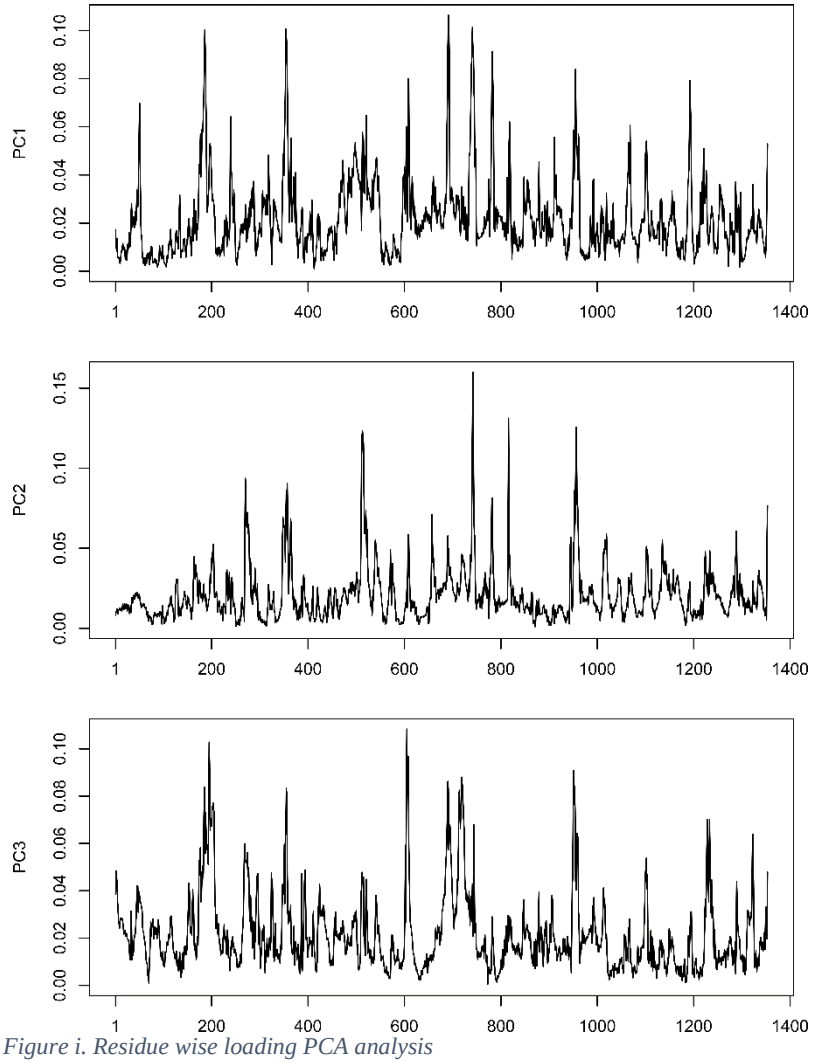

Figure i. Residue wise loading PCA analysis

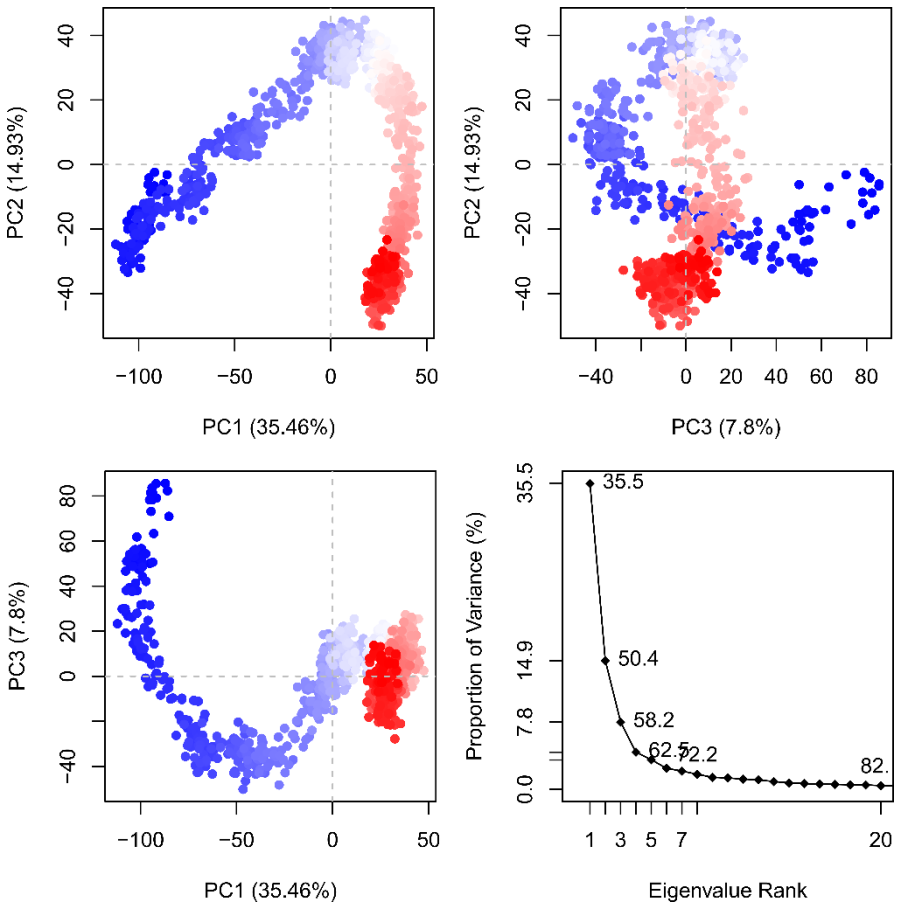

Figure ii. Principle complement analysis of MD simulation

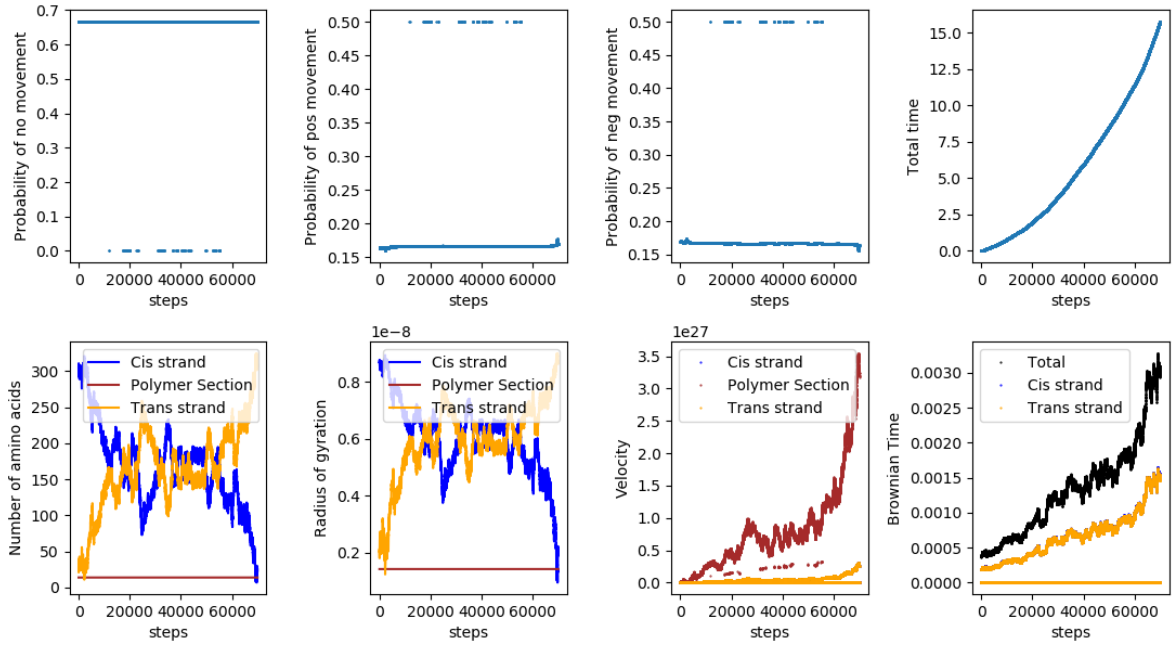

Figure iii. Example of quantities needed for calculation of Monte Carlo simulation

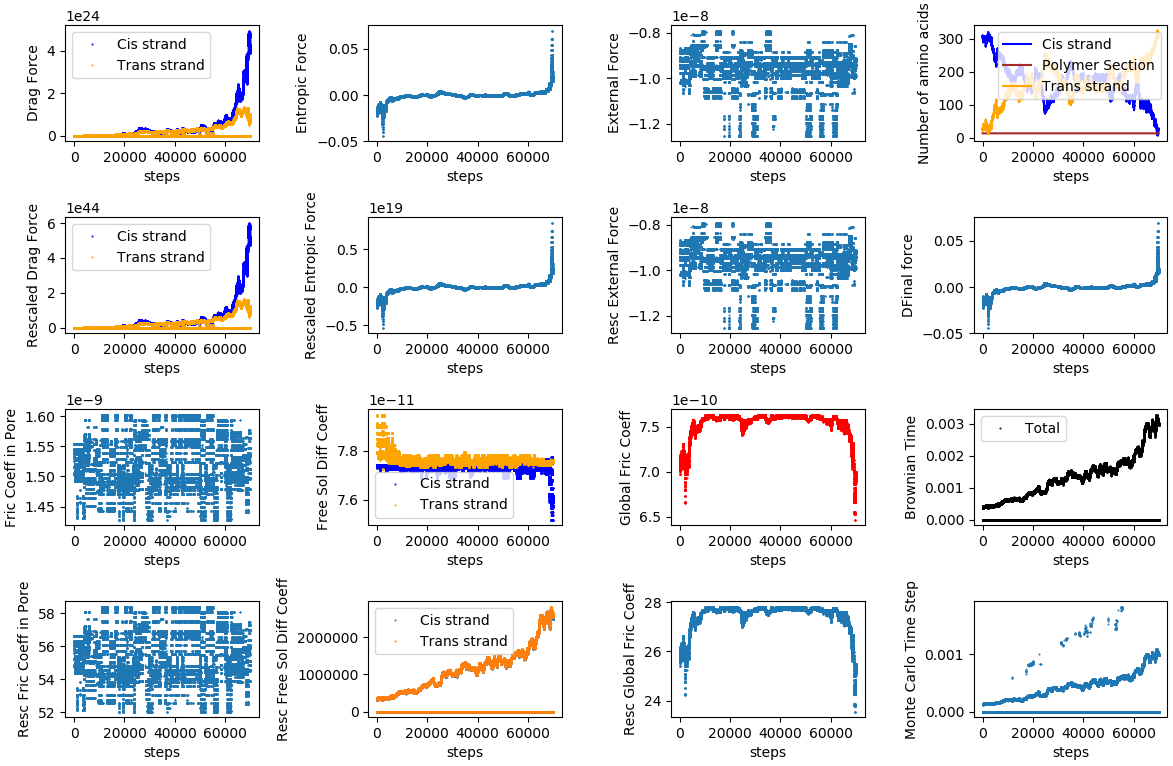

Figure iv. Example of quantities needed for calculation of Monte Carlo simulation