

Molecular Dynamics Simulations of the Nascent Peptide Chain in the Ribosomal Exit Tunnel

Diploma Thesis

by

Lars Bock

Department for Theoretical and Computational Biophysics

Max Planck Institute for Biophysical Chemistry

Göttingen

2007

Contents

1	Introduction	4
2	Biological Background	8
2.1	Ribonucleic acid	8
2.2	Proteins	9
2.3	Ribosome	12
2.4	Ribosomal Exit Tunnel	14
2.5	Translational Arrest of SecM	14
2.6	<i>Haloarcula Marismortui</i>	17
3	Molecular Dynamics	18
3.1	Principles	18
3.2	Methods	21
3.2.1	Periodic Boundary Condition	21
3.2.2	Temperature and Pressure Coupling	21
3.3	Force Probe Molecular Dynamics	22
4	Methods	24
4.1	Set-up of the System	24
4.1.1	Crystal Structure	24

4.1.2	Post-Transcriptionally Modified Nucleotides	26
4.1.3	Protonation and Treatment of the Termini of the Proteins	26
4.1.4	Salt Concentration	27
4.2	Simulation System Set-up	28
4.3	Three Levels of Complexity	29
4.3.1	Three Simulation Systems	29
4.3.2	Charges Outside the Freely Simulated Area	31
4.4	Polypeptides Inside the Tunnel	34
4.5	Sequential Adding and Pushing of Amino Acids	35
4.6	Analysis	38
4.6.1	Root Mean Square Deviation	38
4.6.2	Root Mean Square Fluctuation	38
4.6.3	Autocorrelation Function	39
4.6.4	Distance Geometry	39
4.6.5	Principal Component Analysis	40
5	Results	43
5.1	Salt Concentration	44
5.2	Equilibration	45
5.3	Three levels of complexity	47
5.4	Conformations and Dynamics of Polypeptides inside the Tunnel	51
5.4.1	Modeled Polypeptides	52
5.4.2	Equilibration of Polypeptides	52
5.4.3	Conformations of the Polypeptides	54
5.4.4	Interactions of the Polypeptides with the Ribosome . .	60
5.4.5	Dynamics of the Polypeptides	64

5.4.6	Entropy Differences	67
5.5	Movement of the Polypeptide through the Tunnel	68
6	Summary and Outlook	77

Chapter 1

Introduction

The ribosome is a large macromolecular complex which synthesizes all proteins in the cell according to the genetic code, one of the most important processes in cells [1]. Proteins are long chains of amino acids folding into a unique functional three-dimensional structure. The sequence of amino acids is stored in deoxyribonucleic acid (DNA), a long polymer of nucleotides. There are twenty different standard amino acids but only four different nucleotides, so triplets of nucleotides code for the individual amino acids.

Segments of the DNA, called genes, contain the sequence information for individual proteins. These segments are transcribed to ribonucleic acid (RNA) by the RNA polymerase. The strands of RNA transport the information to the ribosome and are therefore called messenger RNA (mRNA). In a process called translation, the ribosome reads the sequence information from the mRNA and accordingly links amino acids to a growing peptide chain [2].

Besides mRNA, RNA has two other forms that are important in the process of translation. Ribosomal RNA (rRNA) is RNA which folds into a unique three-dimensional structure, forming the ribosome together with ribosomal

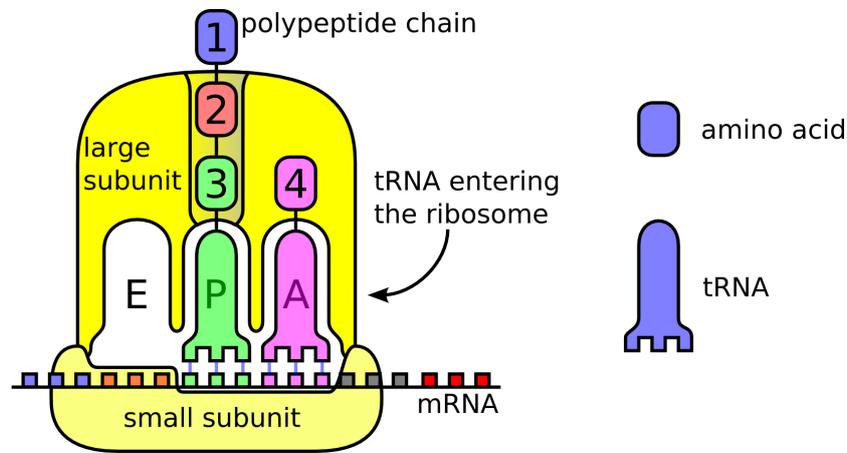


Figure 1.1: Scheme of the translating ribosome.

proteins. Transfer RNA (tRNA) is a family of folded RNA strands which covalently bind amino acids on one side and can basepair with nucleotide triplets on the opposite side. For each amino acid there is a specific tRNA to which it binds and which basepairs with the according nucleotide triplet.

To start translation, two subunits of the ribosome bind to the mRNA (see figure 1.1). Amino acids bound to tRNAs enter the ribosome and when the tRNA basepairs with the nucleotide triplet of the mRNA, mRNA and amino acid are translocated to another position inside the ribosome. The mRNA also translocates so that the next triplet is ready for basepairing. When the next matching mRNA with its amino acid enters the ribosome, the first amino acid is linked by a peptide bond to the second amino acid, breaking the bond to the first tRNA. The tRNAs and mRNA translocate and the first tRNA, now empty, exits the ribosome and a new mRNA triplet is visible [3]. This process continues elongating the polypeptide chain until a special triplet, called stop codon, causes the ribosome to free the finished polypeptide.

The growing peptide chain exits the ribosome through a tunnel which is approximately 100 Å long [4]. The tips of two ribosomal proteins form part of the tunnel wall, but the largest part is formed by rRNA.

How, structurally, the nascent chain moves through the tunnel, is largely unknown. Fluorescence resonance energy transfer (FRET) measurements suggest that the nascent polypeptide might prefold inside the tunnel, forming α -helices [5].

This question becomes particularly relevant in light of the stunning observation that there are sequences, e.g., of the protein SecM, that cause a translation arrest, thereby stalling the translation while the polypeptide is still inside the tunnel. The ribosome is disabled until the arrest is abolished [6]. Accordingly, the mechanism of this process still remains unsolved.

Further, the essential role of bacterial ribosomes in translation renders it an important target for antibiotics. A class of antibiotics, called the macrolides, specifically bind inside the tunnel and block the progression of the polypeptide through the tunnel [7]. The blocked ribosomes can not synthesize proteins anymore, which are essential for the survival of the bacterium and thus the bacterium dies. Ribosomes known to be resistant to the macrolide erythromycin, however, have a mutation in the tip of one of the proteins forming a part of the tunnel wall causing the resistance. The antibiotic still binds inside the tunnel, but yet the polypeptide can progress. Therefore, a conformational change of the tunnel, which deactivates the blockage, is proposed [8]. To shine light on this question of immediate pharmaceutical relevance also requires knowledge of the conformations of the nascent peptide chain as it moves towards the exit of the ribosome.

In this work we will therefore address the following questions by molecular dynamics simulations.

- What are the conformations and dynamics of polypeptides inside the tunnel?
- Do the conformations and dynamics depend on the sequence of the polypeptide?
- Which path through the tunnel do the growing polypeptides follow during elongation?
- What force is needed at the peptide synthesis site to push the peptide chain into and along the exit tunnel?

The work is structured as following. In chapter 2 the biological background necessary to understand this work is explained. In chapter 3 the methods of molecular dynamics simulations and force probe molecular dynamics simulations are described. The methods and steps that were necessary to perform this work are the topic of chapter 4. In chapter 5 the results are presented and discussed. The main results are summarized in chapter 6 and an outlook for possible future work is given.

Chapter 2

Biological Background

2.1 Ribonucleic acid

Ribonucleic acids (RNA) are molecules which play important roles in all living organisms. RNA is polymer of nucleotide monomers (figure 2.1) and each monomer consists of a phosphate, a ribose and one of the four bases: adenine, guanine, cytosine and uracil. The phosphate and the ribose form the backbone, where the individual monomers are linked. The oxygen at the C3 of the ribose is bound to the phosphorus of the subsequent nucleotide.

In a process called transcription, parts of the DNA, which stores the genetic information, are copied by an enzyme called RNA polymerase to complementary RNA. These pieces of RNA (mRNA) contain the information of the order of amino acids for a protein. This information is organized in codons which are groups of three consecutive bases. Each codon codes for one amino acid, except for the start and the stop codons. The start codon specifies where to begin with the translation, and the stop codon specifies the end of translation.

Most functional forms of RNA strands require a specific three-dimensional structure. This tertiary structure is stabilized by hydrogen bonds between nucleotide bases. These hydrogen bonds are called base pairs and mostly occur between adenine and uracil or between guanine and cytosine.

The role of tRNA and rRNA is directly related to the ribosome and will be described in 2.3.

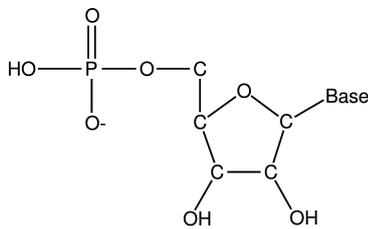


Figure 2.1: Nucleotide Monophosphate with the phosphate on the left, and the ribose in the center. The base can be one of the four bases adenine, guanine, cytosine, and uracil.

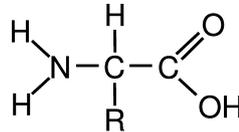


Figure 2.2: The common structure of an amino acid with the amino group on the left and the carboxyl on the right. The residue (R) determines the type of the amino acid.

2.2 Proteins

Proteins are large molecules built of amino acids (figure 2.2) which are linked by peptide bonds forming a linear chain. The peptide bond is formed by dehydration between the nitrogen of the amino group and the carbon of the carboxyl group. There are 20 different standard amino acids, which are encoded in the DNA/RNA, but they can also be posttranslationally modified. They share a common structure (backbone) and differ in their residues (side chains) which determine the physical and chemical properties of the amino

acid. Among these properties are polarity, acidity or basicity, hydrophobicity and charge.

The structure of a protein can be described using four levels. The primary structure is the sequence of amino acids in the protein, which is encoded in the DNA/RNA.

The secondary structure describes the local structure of successive amino acids which are stabilized by hydrogen bonds. The most common are α -helices and β -sheets. An α -helix is a coiled structure, where every amino acid has a hydrogen bond with the amino acid four residues earlier. This hydrogen bond is formed between the N-H and the C=O group of the backbone. A β -sheet consists of β -strands, which are sequences of amino acids, where the backbone is almost completely extended. The β -strands form a hydrogen-network with adjacent β -strands building the stabilized β -sheet structure.

The tertiary structure is the three dimensional structure of a protein determined by the sequence of amino acids and the minimum of free energy G , which the protein achieved by folding. This structure is stabilized by a lot of effects, among them are van-der-Waals interactions, electrostatic and hydrophobic effects. Often other atoms, molecules and ions play a role in stabilizing the tertiary structure.

Proteins often aggregate to larger complexes which is called the quaternary structure.

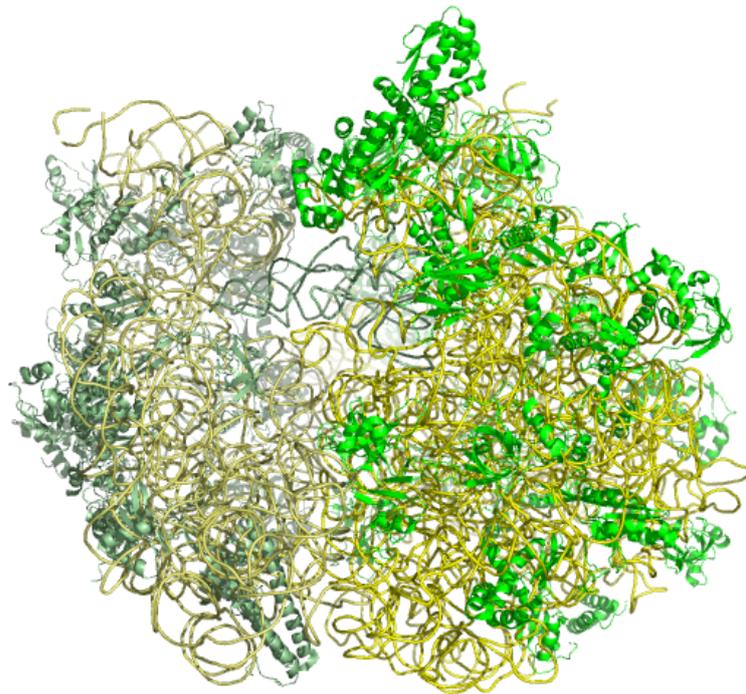


Figure 2.3: Cartoon diagram of the *Thermus thermophilus* ribosome. RNA is drawn in yellow and proteins are drawn in green. The 30S subunit is drawn in pale, the 50S subunit in saturated colors. PDB codes for the two structures are 1YL3 and 1YL4.

2.3 Ribosome

The ribosome is a large macromolecular complex which decodes the information carried by messenger ribonucleic acid (mRNA) and synthesizes proteins accordingly. It consists of several RNA molecules, the so called ribosomal RNA (rRNA), and ribosomal proteins forming two subunits. Ribosomal components are named after their sedimentation rate in an ultracentrifuge in the unit of Svedberg (S), bigger components yield a higher sedimentation rate. The procaryotic (70S) and the eucaryotic (80S) ribosomes are structurally quite similar, but differ in size and number of rRNA and protein components.

The 70S ribosome consists of a large (50S) and a small subunit (30S), where the 50S subunit is built of two rRNA strands and 34 proteins, and the 30S subunit is built of one rRNA strand and 21 proteins (figure 2.3). Most of the proteins sit on the surface of the ribosome, but have long nonglobular regions penetrating into the ribosome.

The two subunits are separate while they are inactive and form the ribosome when they attach to an mRNA to synthesize the protein.

In the translation process, besides the ribosome and mRNA, transfer RNAs (tRNAs) play a crucial role. A tRNA is a small RNA strand, which has a binding site for specific amino acids at the 3' end and a region consisting of three bases, called the anticodon which can base pair to three bases of a specific codon on the mRNA. During the translation (figure 2.4), the mRNA is read by the ribosome, and amino acids are linked according to the sequence. The peptidyl transferase reaction takes place at the peptidyl transferase center (PTC), which is a part of 50S subunit close to the intersubunit surface

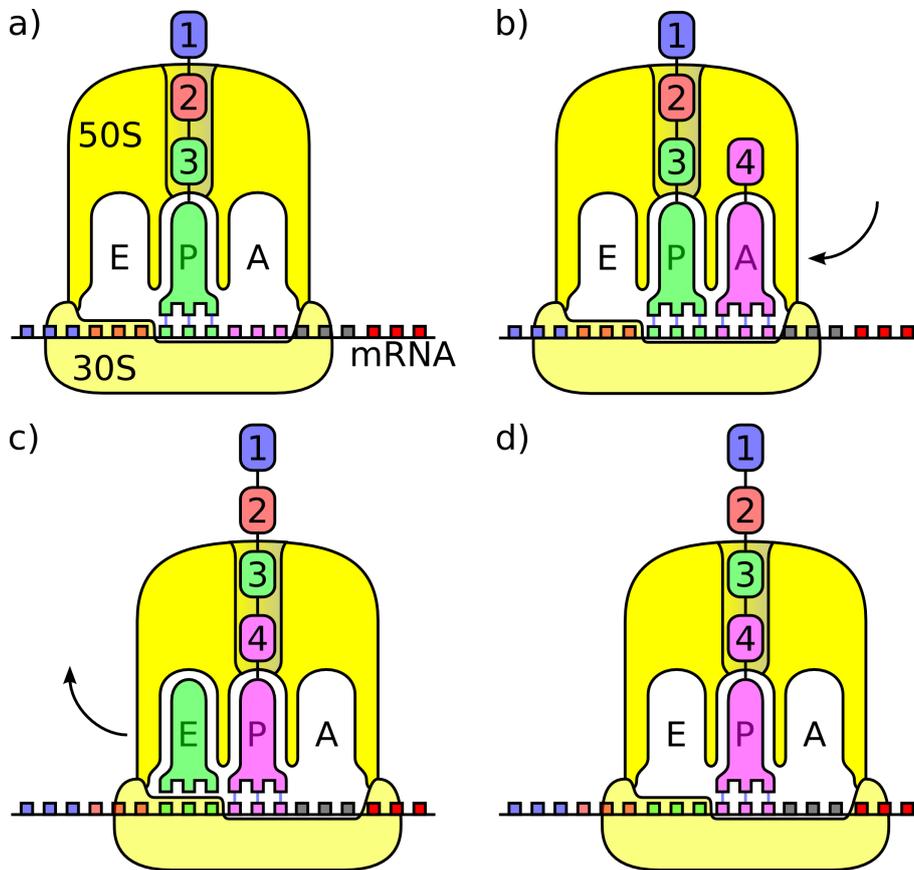


Figure 2.4: The steps of translation. The mRNA, the 50S and the 30S subunit assemble to the working ribosome. There are three tRNA binding sites: A for the aminoacyl-tRNA, P for the peptidyl-tRNA and E for exiting the ribosome. **a)** A peptidyl-tRNA is bound in the P-site, which is attached to nascent peptide chain (amino acids 1,2 and 3). **b)** A aminoacyl-tRNA, whose anti-codon region matches the exposed mRNA codon, has entered the A site. **c)** The peptide chain was linked with amino acid 4 (peptidyl-transferase reaction), the empty tRNA moved from P-site to E-site and the tRNA, which now carries the peptide chain moved to the P-site. **d)** the empty tRNA has left the E-site.

and consists only of rRNA [9]. The ribosome is thereby a ribozyme, because the catalytic region is completely built of RNA. The steps of translation are mediated by elongation factors (EF-Tu and EF-G) which hydrolyse GTP to GDP and the ribosome undergoes conformational changes during translation [10].

2.4 Ribosomal Exit Tunnel

The nascent peptide chain exits the ribosome via a tunnel which stretches through the large ribosomal subunit. The tunnel begins at the peptidyl transferase center (PTC), where the peptide bonds are formed and exits the ribosome at the opposite site of the subunit (figure 2.5 a). The length of the tunnel is approximately 100 Å and its diameter varies between 10 and 20 Å [11].

The tunnel walls mainly consist of RNA loops of the 23S rRNA, but also proteins L22, L4 and L39e contribute significantly (figure 2.5 b). Parts of L4 and L22 build the tunnel surface close to the PTC and L39e sits near the exit. L39e is the only ribosomal protein which has no globular region on the ribosomal surface, but is completely buried in the ribosome.

2.5 Translational Arrest of SecM

The Sec translocase is a protein complex mediating the transmembrane movement of pre-proteins across cellular membranes [6]. In eukaryotic cells, pre-proteins are translocated across the cytoplasmic membrane and in prokaryotic cells an equivalent process takes place at the endoplasmic reticulum

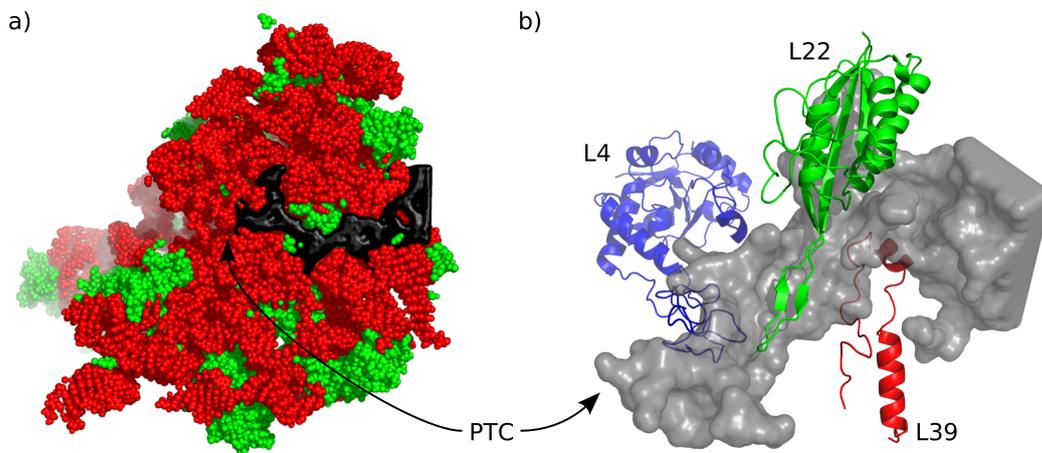


Figure 2.5: **a)** Cross-section of the 50S subunit of *Haloarcula Marismortui* (pdb code 1S72) drawn in sphere representation, RNA is colored in red, proteins in green and the tunnel in black. **b)** The proteins L4, L22, and L39 form a part of the tunnel wall.

(ER) membrane. The primary components of the Sec translocase are SecYEG and SecA. SecYEG, which consists of the integral membrane proteins SecY, SecE and SecG, builds the channel for the polypeptide and SecA is an ATPase which drives the transmembrane movement.

The Sec translocase can not only move proteins across membranes, but it also integrates membrane proteins. The signal recognition particle (SRP) is a ribonucleoprotein which is a targeting factor mainly for membrane proteins [12]. It targets the pre-protein cotranslationally to the Sec translocase.

The expression of SecA is regulated via the nascent peptide chain of SecM. The sequences of SecM and SecA reside on the same mRNA, where SecM is located upstream of SecA. The intergenic region of SecM and SecA on the mRNA can form a stem-loop secondary structure occluding the Shine Dalgarno (S-D) sequence for SecA resulting in a lower initiation rate [13]. The

S-D sequence is a sequence upstream to the start codon helping the ribosome to initiate translation.

The translation of SecM is stalled at Pro166 and the stem-loop structure of the mRNA is disturbed by the ribosome, leaving the SecA S-D sequence visible for other ribosomes, which increases the initiation rate for SecA proteins. There are two mechanisms proposed for the cancellation of the translation arrest. One is a pulling mechanism, where the translocase, to which the parts of the nascent SecM outside the ribosome are bound, generates a force on the nascent chain. Another proposed mechanism is a conformational change of the ribosome induced by the translocase. After the cancellation of the translation arrest, SecM is exported by the translocase to the periplasm where it is degraded by a protease.

The arrest point is identified to be Pro166 and the arresting sequence motif is found to be FXXXXWXXXXGIRAGP where the last amino acid is the arrest point [14]. To determine the arrest point, a stop codon was inserted at different positions of the mRNA. Insertions downstream of the arrest point can not have influence on the arrest. When the insertion is at the arrest point or upstream the translation is prevented. The necessary sequence motif was identified by codonwise mutation of the SecM mRNA. Mutants of the ribosome which allow completion of SecM have mutations in the 23S rRNA and the ribosomal protein L22. These mutated residues face the inner wall of the ribosomal exit tunnel, mostly at the constriction of the tunnel where the tips of L22 and L4 build the tunnel wall together with 23S rRNA residues.

2.6 *Haloarcula Marismortui*

Haloarcula Marismortui is a prokaryotic archaea, which originates from the Dead Sea [15]. It is extremely halophilic and has a physiological salt concentration of over 3 M to compensate for the external osmotic pressure. Therefore, all the cellular components are adapted to function at this salt concentration.

In 2000, the crystal structure of the large ribosomal subunit of *Haloarcula Marismortui* was resolved at atomic resolution [16].

Chapter 3

Molecular Dynamics

3.1 Principles

Molecular dynamics (MD) is a form of computer simulation, which describes a molecular system as a system of atoms and a potential acting upon them. Newton's equations are integrated over time to obtain information about the dynamics of the system. This approach is used in many scientific fields, especially for the description of the atomistic motion of biomolecules.

The time-dependent Schrödinger equation describes the exact motion of atoms, but analytic approaches even fail to solve the equation for the helium atom. Numerical methods are available, but computationally intense and therefore only applicable to systems comprising few atoms.

To reduce the computational effort, several approximations are required. The first is the Born-Oppenheimer approximation, which separates the electronic motion from the nucleic motion. This approximation rests on the fact that the mass of an electron is three magnitudes larger than the mass of a single nucleon. Therefore the dynamics of the electrons is much faster than the

dynamics of the nuclei, because the relaxation of the electrons within the ground state is fast with respect to the nuclear motion. Accordingly, it is sufficient to only describe the movement of the nuclei.

The second approximation is the use of a potential $V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ which describes the interatomic energies with simple functions [17]. This potential is called force field and the one used in this work has the following form:

$$\begin{aligned}
 V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = & \underbrace{\sum_{\text{bonds}} \frac{1}{2} K_b (b - b_0)^2}_{\text{bond stretching}} + \underbrace{\sum_{\text{angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2}_{\text{angle bending}} \\
 + & \underbrace{\sum_{\text{improper dihedrals}} \frac{1}{2} K_\xi (\xi - \xi_0)}_{\text{improper dihedral angle}} + \underbrace{\sum_{\text{dihedrals}} \frac{1}{2} K_\phi [1 + \cos(n\phi - \delta)]}_{\text{dihedral angle}} \\
 + & \sum_{\text{pairs}(i,j)} \left(\underbrace{\frac{C_{12}(i,j)}{r_{ij}^{12}} - \frac{C_6(i,j)}{r_{ij}^6}}_{\text{van der Waals}} + \underbrace{\frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}}_{\text{Coulomb}} \right),
 \end{aligned}$$

where N is the number of atoms, \mathbf{r}_i and q_i are the position and the charge of atom i , respectively. In the force field, there are four forces which act upon the atoms which are covalently bonded. The bond stretching term describes the force which occurs when the bond length changes with a harmonic potential. The minimum energy bond length is b_0 and the force constant is K_b . The second term describes the bond angle bending interaction with a harmonic potential which depends on the angle between two bonds involving three atoms. The equilibrium bond angle is θ_0 . There are two terms which specify interactions between four atoms. The first is the improper dihedral angle term which describes forces acting upon the atoms according to the angles between two planes, e.g., to keep aromatic rings planar. The second term is the dihedral angle term which describes the forces occurring due to

the torsion around a bond.

The last term is a sum of non-bonded interactions over all pairs of atoms. The van der Waals interaction and the Pauli repulsion are described by the Lennard-Jones potential. The charges, which are smeared across molecules, are described as partial point charges q_i which are assigned to the atoms. The electron cloud is thereby simplified to a set of point charges, and the electrostatic forces are calculated by Coulomb's law.

There are two ways to obtain the parameters for the potential $V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$. One way is to fit them to results of ab-initio quantum mechanics calculations and another way is to fit the parameters to various experimental data like free energies of solvation, NMR or x-ray data. For most potentials, and also for the one used in this work, a mixture of both approaches is used.

The third approximation is to describe the motion of all atoms $i = 1, \dots, N$ by Newton's equation of motion,

$$m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2} = \nabla_i V(\mathbf{r}_1, \dots, \mathbf{r}_N),$$

where m_i is the mass of atom i and V is the potential described above. This equation is integrated in discrete time steps with an integration step length of Δt . For all simulations described in this work we used an integration step length of $\Delta t = 2$ fs.

Velocities \mathbf{v}_i and positions \mathbf{r}_i of the atoms were updated with the Verlet algorithm [18],

$$\begin{aligned} \mathbf{v}_i(t + \frac{\Delta t}{2}) &= \mathbf{v}_i(t - \frac{\Delta t}{2}) + \frac{\mathbf{F}_i(t)}{m_i} \cdot \Delta t \\ \mathbf{r}_i(t + \Delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t + \frac{\Delta t}{2}) \cdot \Delta t, \end{aligned}$$

where $\mathbf{F}_i = -\nabla_i V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ is the force acting on atom i .

The output of the simulations was the trajectory and the interaction ener-

gies between atoms or groups of atoms. The trajectory contains the atomic positions and velocities at given intervals of the simulation time.

3.2 Methods

3.2.1 Periodic Boundary Condition

The number of atoms in a simulation is limited due to limited computational resources. To minimize artefacts due to the resulting small system size and surface effects, periodic boundary conditions were applied in all simulations. Accordingly, the atoms are put into a space-filling simulation box, which is surrounded by translated images of itself. Atoms leaving the simulation box on one side are put back into the box on the opposite side. Similarly, for the calculation of the potential also atoms which are on the other side of the boundary are taken into account.

The simulation box size has been chosen sufficiently large to avoid that molecules interact with their images. The Debye-Hückel length [19] gives a good estimate for the range of this interaction. The Debye-Hückel length for the ion concentration in our system (2 Mol/l) is 0.31 nm, such that the chosen distance of 1.5 nm between the ribosome and the boundary of the simulation box guarantees that interactions with the images are small.

3.2.2 Temperature and Pressure Coupling

Under normal conditions cells have a close to constant temperature and pressure, which is described by an NpT-ensemble. To achieve this ensemble, we need to couple temperature and pressure to given reference values, because

given a constant energy and a constant volume, the simulation would be in an microcanonical (NEV) ensemble. Consequently, we simulate with temperature and pressure coupling. To account for the time scales of energy and pressure fluctuations in the system, the coupling is not instantaneous, but a coupling time constant is introduced.

For temperature coupling we used the Berendsen temperature coupling scheme [20], where the velocity of every particle v is scaled to λv in every step with

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T} - 1 \right)},$$

where Δt is the integration step length, τ_T is the temperature coupling time constant, $T_0 = 300$ K is the reference temperature and T is the instantaneous temperature derived from the kinetic energy of all atoms.

Pressure coupling to the reference pressure $P_0 = 1$ atm was achieved by the Berendsen pressure coupling method [20], where the edges of the simulation box and the coordinates of the atoms are scaled with the factor μ .

$$\mu = 1 - \frac{\Delta t}{3\tau_P} \kappa (P_0 - P),$$

where τ_P is the pressure coupling time constant, κ is the isothermal compressibility of water and P is the instantaneous pressure derived from the velocities and forces of all atoms via the virial theorem.

3.3 Force Probe Molecular Dynamics

In this work we examined the pathway of the polypeptide chain through the ribosomal exit tunnel. We therefore pushed the polypeptide in the direction of the tunnel axis. This was achieved by the use of force probe molecular

dynamics (FPMD), a method which has been established to simulate atomic force microscopy (AFM) experiments on single molecules [21, 22], e.g., pulling a ligand out of a binding pocket [23]. The atomistic interactions and dynamics can not be observed in the AFM experiments which is the main motivation for FPMD simulations.

To model the effect of the cantilever, an additional harmonic potential V_{pull} is introduced,

$$V_{pull}(\mathbf{x}_i, t) = \frac{1}{2}k((\mathbf{x}_i - \mathbf{x}_i^0) \cdot \hat{\mathbf{n}} - vt)^2$$

where \mathbf{x}_i is the position of atom i which is in the group of atoms to be pulled, \mathbf{x}_i^0 is its position in the beginning of the simulation, k is the spring constant, the normalized vector $\hat{\mathbf{n}}$ is the direction of pulling, and v is the velocity with which the spring is moved in the pulling direction.

The additional pulling force \mathbf{F}_i acting on the atom i in the pull group is then given by

$$\mathbf{F}_i(t) = \hat{\mathbf{n}}\nabla V(\mathbf{x}_i, t) = k\hat{\mathbf{n}}((\mathbf{x}_i - \mathbf{x}_i^0) \cdot \hat{\mathbf{n}} - vt).$$

Chapter 4

Methods

4.1 Set-up of the System

4.1.1 Crystal Structure

The x-ray crystal structure of the large ribosomal subunit from *Haloarcula Marismortui*, protein data bank entry 1S72 [24], was used as a starting structure. The resolution of the structure is 2.4 Å, the R factor is 0.188, and the R_{free} factor is 0.222. R is a measure for the agreement of the modeled structure and the x-ray diffraction data. R_{free} is the same as R , but it is derived by cross-validation based on a test set consisting of a small percentage of reflections excluded from structure refinement.

The model contains two strands of RNA, the 23S rRNA and the 5S rRNA, and 29 ribosomal proteins. All nucleotides of the 5S rRNA and 2754 of the 2922 23S nucleotides are resolved. The nucleotides, which are not resolved, are not closer than 25 Å to the tunnel and were therefore not considered.

Residues of proteins L4, L22, and L39 form a part of the tunnel wall, L4 and

L22 close to the PTC and L39 close to exit of the tunnel (see figure 2.5b). L24 is located next to the opening of the tunnel. L4 and L24 are resolved in the structure, but four amino acids at the c-terminal of L22, which are more than 30 Å away from the tunnel, are not resolved.

The crystal structure misses three L39 residues at the tunnel opening. These residues have been resolved in a different crystal structure of the large ribosomal subunit of *Haloarcula Marismortui* (protein data bank entry 1YJ9 [25]). To model these residues, the atomic positions of protein L39 of the pdb entry 1YJ9 were fitted to the atomic position of L39 of the pdb entry 1S72 using the positions of the C α atoms present in both structures. The three residues of the fitted 1YJ9 structure were included into the 1S72 structure and an energy minimization was performed.

A loop of 11 amino acids of the ribosomal protein L10, which are in a region closer than 25 Å from the tunnel, are not resolved. No available *Haloarcula Marismortui* crystal structure contains this loop, suggesting that it is rather flexible. The Swiss-Model-Server (<http://swissmodel.expasy.org>) was used to build a homology model with the protein sequence and the part of the 1S72 structure describing the protein. The result was reinserted into the structure with only one minor sterical clash between an L10E and a 23S residue.

All other residues missing in the crystal structure are more than 25 Å away from the tunnel and are not expected to affect the dynamics of the studied polypeptides in the exit tunnel. We therefore did not attempt to model these remaining residues.

Besides the ribosome, the crystal structure contains 15 water molecules, 234 Mg⁺, 174 Na⁺, 46 Cl⁻, and 2 K⁺ ions, which were taken into account in all

simulations.

4.1.2 Post-Transcriptionally Modified Nucleotides

The crystal structure contains five post-transcriptionally modified nucleotides. Modified nucleotides are reported for all kinds of RNA, including tRNA, mRNA, and rRNA. Modifications like methylations and uridine isomeration can enforce or block base pairing and thereby play a role in RNA folding [27]. In the crystal structure there are four methylated nucleotides, 1-methyladenosine, 2'-O-methyluridine, 2'-O-methylguanosine, and 3-methyluridine, and one isomerized uridine, pseudouridine, are resolved. These were taken into account in all simulations.

All simulations in this work were carried out using the software package GROMACS [26] and the GROMACS port of the amber forcefield [30]. The standard AMBER-forcefield [28] contains parameters for all common amino acids and nucleotides, but lacks parameters for modified nucleotides. Parameters from the Modifieds Database Server (<http://ozone3.chem.wayne.edu/>) [29] were converted into the GROMACS format and included into the force field.

4.1.3 Protonation and Treatment of the Termini of the Proteins

As the x-ray structure does not contain hydrogen atoms, we added them to the model using `pdb2gmx` from the GROMACS simulation suite [26]. Histidines have a relatively neutral pK. So their protonation state depends on the surrounding atoms and the possibility to form hydrogen-bonds. The

molecular modeling package WHATIF [31], which takes the position of potential hydrogen-bond forming atoms into account, was used to choose the protonation state of the histidines.

Normally protein termini are charged, but some terminal amino acids of some proteins are not resolved by x-ray. If we charged the truncated termini in the model, this charge would be in a position far away from the position of the terminal in the ribosome. So we did not charge the termini, in case there were more than two amino acids missing.

4.1.4 Salt Concentration

The physiological salt concentration of *Haloarcula Marismortui*, 3 M, is about 20 times larger than in mammals or most bacteria. Because the force fields are optimized for standard conditions, it is important to examine the behavior of the water molecules and salt ions under higher salt concentrations.

We therefore simulated NaCl solutions with different concentrations (1 M, 2 M, 3 M, and 4 M). The simulation box was a cube with 4 nm edge length. The amber99 force field [30] was used and the simulation time was 3 ns each. The diffusion rate D of the ions was calculated via the Einstein-Smoluchowsky equation,

$$\langle |\mathbf{x}(t) - \mathbf{x}(0)|^2 \rangle = 2dDt,$$

where $d = 3$ is the dimension and $\mathbf{x}(t)$ is the coordinate of the ion.

4.2 Simulation System Set-up

To prepare the system for simulation, the model of the ribosome, with all our modifications, was put into a cuboid simulation box with edge lengths 24.5 nm, 23.5 nm, and 23.1 nm, respectively, such that the distance between the ribosome and the box faces is larger than 1.5 nm. This box was then filled with water molecules and sodium and chloride ions at a concentration of 2 M using the programs `genbox` and `genion` from the GROMACS simulation suite [26].

The long range electrostatics beyond 1 nm were calculated with particle mesh Ewald (PME) [32, 33] which needs an overall neutral system. Each nucleotide carries one negative elementary charge. Amino acids arginine, lysine, and histidine, if protonated, are positively charged. Amino acids aspartic acid, and glutamic acid are negatively charged. To obtain an overall neutral system, additional ~ 2600 sodium ions were placed with `genion`. The system then contained $\sim 91\,000$ ribosomal atoms, $\sim 364\,000$ water molecules, $\sim 16\,000$ sodium, and $\sim 13\,000$ chloride atoms including those resolved in the crystal structure, summing up to a total of $\sim 1\,270\,000$ atoms.

An energy minimization using steepest descent was performed starting with a total potential energy of $2.10 \cdot 10^{11} \frac{\text{kJ}}{\text{mol}}$ converging at a total potential energy of $-2.98 \cdot 10^7 \frac{\text{kJ}}{\text{mol}}$. Then the water molecules and the ions were equilibrated for 1 ns with position restraints on the ribosomal atoms, adding a harmonic potential with a force constant $k = 1000 \frac{\text{kJ}}{\text{mol nm}^2}$.

Subsequently, the solvent and the modeled loops were equilibrated for 1 ns, keeping position restraints on the rest of the ribosome, to let the loops find a favorable position without changing the positions of other ribosomal atoms.

Then the whole system was equilibrated for 3 ns without any position restraints.

4.3 Three Levels of Complexity

4.3.1 Three Simulation Systems

Due to the large size of the system and the required length of simulation time needed to address our questions, we searched for way to reduce CPU time. To this aim, we assumed that the tunnel region is not influenced much by the outer area of the subunit except by Coulomb interaction and that it is sufficient to describe this interaction via the average positions of the atoms. To verify this assumption, we tested three different simulation systems (figure 4.1) and analyzed their influence on the dynamics of the region around the exit tunnel. Therefore the positions of parts of the system were fixed during the simulation. The temperature of the fixed atoms is 0 K and the temperature of the free atoms is 300 K. To reduce the effect of the temperature difference a layer of position restrained atoms was established at the border between fixed and free atoms.

In the first system (figure 4.1a) only atoms in the range of 20 Å around the tunnel (green) were simulated with free molecular dynamics. Position restraints were added to atoms in the range from 20 to 25 Å (pink). The remaining atoms of the ribosome were fixed (red), but the electrostatics were taken into account. The solvent was allowed to move freely inside a box with walls built of fixed argon atoms preventing the solvent atoms to leave the box. The argon atoms were chosen not to interact with the system in any

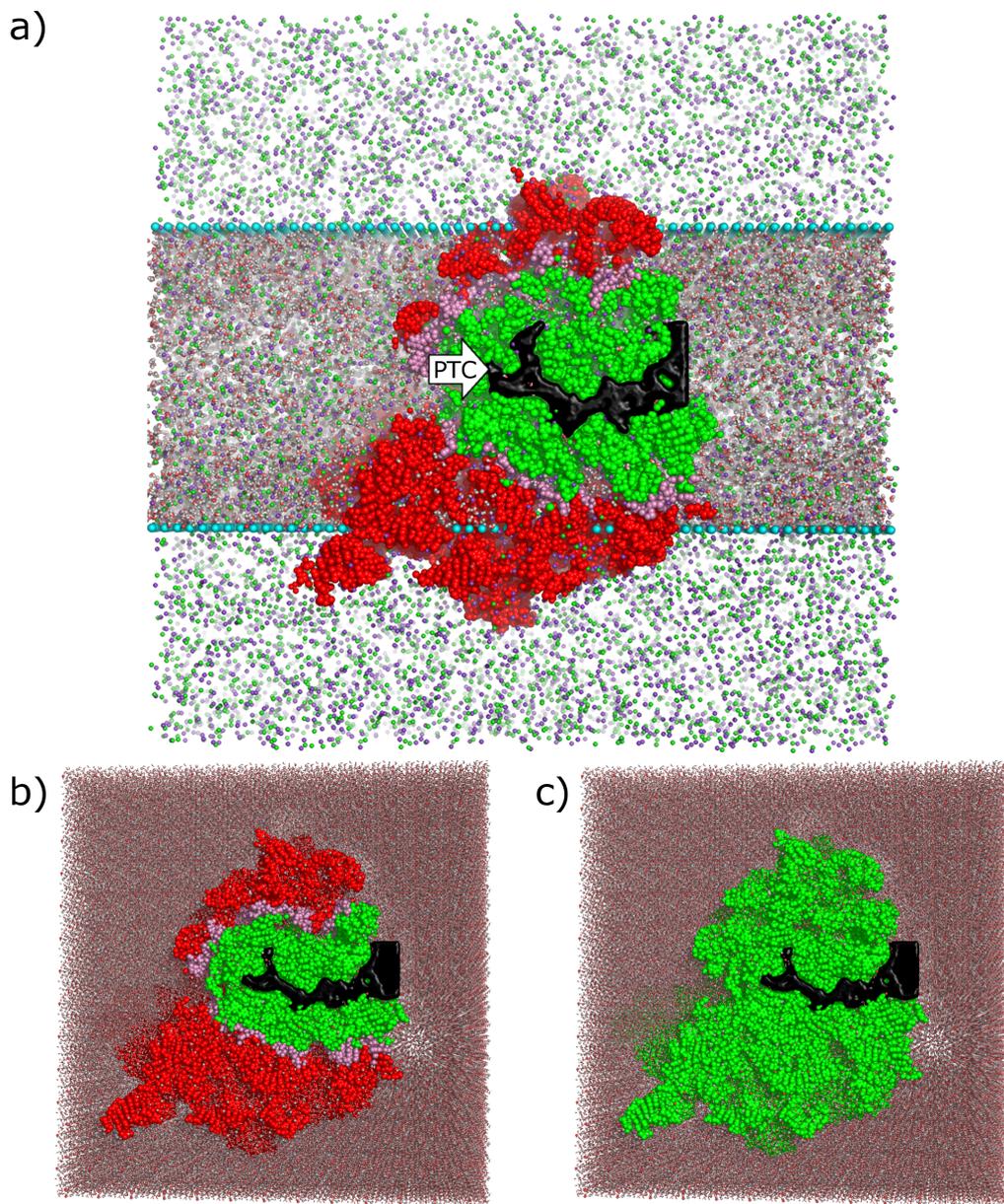


Figure 4.1: Cross section of the ribosome along the tunnel axis: The ribosome is shown in green, pink and red, the ions are in magenta and green, the tunnel in black and the argon layer in cyan. The green area of the ribosome is simulated with free MD, the pink area with position restraints, and the red area is fixed. **a)**, **b)**, and **c)** show the three different types of the simulation system.

other way. The ions outside the argon box were also fixed.

In the second system (figure 4.1b), the ribosomal atoms were treated as in the first system, but solvent molecules were simulated with free MD in the whole simulation box.

In the third system (figure 4.1c), the whole system was simulated with free MD.

4.3.2 Charges Outside the Freely Simulated Area

For the first system we restricted free MD to a box around the tunnel, but we also wanted to consider the electrostatics of the charges outside this area. Due to the lack of water and its charge screening effect the dielectric constant was lowered and we reduced the charge of the atoms in the vacuum to mimic this effect. The following treatment [34] aims at determining the factor for the charge reduction.

Point charge near a plane dielectric boundary

We assume two dielectrics separated by the xy-plane with ϵ_1 for ($z > 0$) and ϵ_2 for ($z < 0$). A point charge q is situated at $\mathbf{r}_q = (0, 0, d)$ with $d > 0$ (figure 4.2).

In an infinite dielectric ($\epsilon_1 = \epsilon_2$) the potential would be

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_1} \frac{q}{|\mathbf{r} - \mathbf{r}_q|}.$$

This would give

$$\nabla \cdot \mathbf{D} = -\epsilon_1 \Delta V = q\delta(|\mathbf{r} - \mathbf{r}_q|)$$

with $\Delta \frac{1}{r} = -4\pi\delta(r)$.

Potential for $z > 0$:

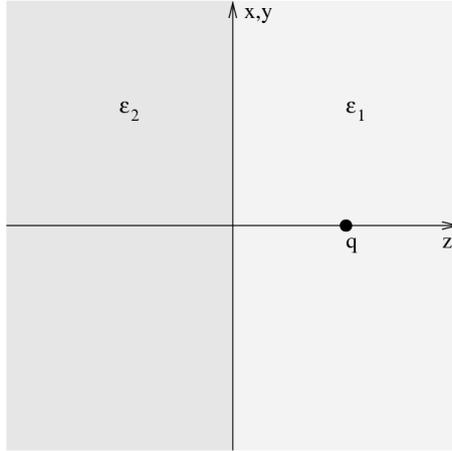


Figure 4.2: Point charge q near a plane boundary between ϵ_1 and ϵ_1 .

We now replace the interface between the two dielectrics by an image charge q_2 at $(0,0,-d)$

$$V_1(\mathbf{r}) = \frac{1}{4\pi\epsilon_1} \left[\frac{q}{(x^2 + y^2 + (z - d)^2)^{\frac{1}{2}}} + \frac{q_1}{(x^2 + y^2 + (z + d)^2)^{\frac{1}{2}}} \right].$$

Potential for $z < 0$:

For this potential we put a charge q_2 at the position $(0,0,d)$,

$$V_2(\mathbf{r}) = \frac{1}{4\pi\epsilon_2} \frac{q_2}{(x^2 + y^2 + (z - d)^2)^{\frac{1}{2}}}.$$

The potentials V_1 and V_2 fulfill the electrostatic Maxwell equations,

$$\nabla \cdot \mathbf{D} = \rho \text{ and } \nabla \times \mathbf{E} = 0.$$

$$\nabla \cdot \mathbf{D}_1 = -\epsilon_1 \Delta V_1 = q\delta(|\mathbf{r} - \mathbf{r}_q|) \text{ for } z > 0$$

$$\nabla \cdot \mathbf{D}_2 = -\epsilon_2 \Delta V_2 = 0 \text{ for } z < 0$$

$$\nabla \times \mathbf{E}_i = \nabla \times (\nabla V_i) = 0 \text{ for } i = 1, 2.$$

V_1 and V_2 must be continuous across the boundary:

$$\frac{\partial}{\partial x} V_1|_{z=0} = \frac{\partial}{\partial x} V_2|_{z=0}$$

$$\begin{aligned}
\frac{\partial}{\partial x} V_1|_{z=0} &= \frac{1}{4\pi\epsilon_1} \left[-\frac{qx}{(x^2 + y^2 + (z-d)^2)^{\frac{3}{2}}} - \frac{q_1x}{(x^2 + y^2 + (z+d)^2)^{\frac{3}{2}}} \right] \Big|_{z=0} \\
&= -\frac{1}{4\pi\epsilon_1} (q + q_1) \frac{x}{(x^2 + y^2 + d^2)^{\frac{3}{2}}} \\
\frac{\partial}{\partial x} V_2|_{z=0} &= -\frac{1}{4\pi\epsilon_2} q_2 \frac{x}{(x^2 + y^2 + d^2)^{\frac{3}{2}}} \\
&\Rightarrow \frac{q + q_1}{\epsilon_1} = \frac{q_2}{\epsilon_2}
\end{aligned}$$

Since there is no free charge at the boundary, the perpendicular component of \mathbf{D} must be continuous across the boundary:

$$\begin{aligned}
\epsilon_1 \frac{\partial}{\partial z} V_1|_{z=0} &= \epsilon_2 \frac{\partial}{\partial z} V_2|_{z=0} \\
\epsilon_1 \frac{\partial}{\partial z} V_1|_{z=0} &= \frac{1}{4\pi\epsilon_1} \left[-\frac{q(z-d)}{(x^2 + y^2 + (z-d)^2)^{\frac{3}{2}}} - \frac{q_1(z+d)}{(x^2 + y^2 + (z+d)^2)^{\frac{3}{2}}} \right] \Big|_{z=0} \\
&= \frac{1}{4\pi\epsilon_1} (q - q_1) \frac{d}{(x^2 + y^2 + d^2)^{\frac{3}{2}}} \\
\epsilon_2 \frac{\partial}{\partial z} V_2|_{z=0} &= \frac{1}{4\pi\epsilon_2} q_2 \frac{d}{(x^2 + y^2 + d^2)^{\frac{3}{2}}} \\
&\Rightarrow q - q_1 = q_2 \\
&\Rightarrow q_1 = \frac{\epsilon_1 - \epsilon_2}{\epsilon_1 + \epsilon_2} q \text{ and } q_2 = \frac{2\epsilon_2}{\epsilon_1 + \epsilon_2} q
\end{aligned}$$

Scaling factor for charges in vacuum

In our situation the charge was in vacuum outside the freely simulated area ($\epsilon_1 = 1$). For the dielectric inside the box we assumed the dielectric of water $\epsilon_2 = \epsilon_{H_2O}$. The Coulomb interaction of the charges in vacuum was too large, because the charge was not screened by water molecules. So we had to compare the potential of the simulation situation (water and vacuum) V_{sim}

with the potential of the desired situation (only water) V_{H_2O} and adjust the charge q' of the atoms in vacuum:

$$V_{sim}(\mathbf{r}) = \frac{1}{4\pi\epsilon_2} \frac{q_2}{(x^2 + y^2 + (z - d)^2)^{\frac{1}{2}}} = \frac{1}{4\pi\epsilon_{H_2O}} \frac{2\epsilon_{H_2O}}{1 + \epsilon_{H_2O}} \frac{q'}{(x^2 + y^2 + (z - d)^2)^{\frac{1}{2}}}$$

$$V_{H_2O}(\mathbf{r}) = \frac{1}{4\pi\epsilon_{H_2O}} \frac{q}{(x^2 + y^2 + (z - d)^2)^{\frac{1}{2}}}$$

$$V_{sim}(\mathbf{r}) = V_{H_2O}(\mathbf{r}) \Rightarrow q' = \frac{1 + \epsilon_{H_2O}}{2\epsilon_{H_2O}} q$$

The charge of the atoms in the vacuum had to be scaled by the factor $\frac{1 + \epsilon_{H_2O}}{2\epsilon_{H_2O}} \approx 0.506$ for $\epsilon_{H_2O} = 82$ (tip3p water). For the simulation in the system, where the water was restricted to the atoms inside the argon box, the charges of atoms outside the argon box were scaled accordingly.

4.4 Polypeptides Inside the Tunnel

To investigate the conformations and dynamics of polypeptide chains in the exit tunnel, three different peptides were modeled into the pre-equilibrated ribosome structure. The peptides were modeled by linking amino acids according to their sequence with the molecular visualization system pymol (<http://www.pymol.org>).

C-pmn-pcb, an analog of a small polypeptide attached to a tRNA (pp-tRNA) in the A-site of an *Haloarcula Marismortui* was crystallized [37]. So we modeled the peptide into the ribosome such that the position of the last amino acid agreed with the position of its analog. The rest of the polypeptide was modeled to sterically fit into the tunnel. Solvent molecules overlapping with the polypeptide atoms were deleted from the system.

For each polypeptide we ran four different simulations of 5 ns length each. The

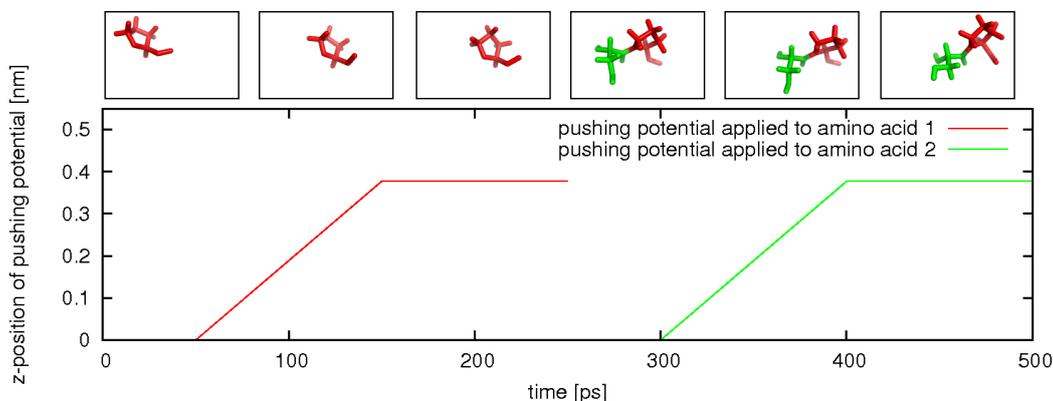


Figure 4.3: Adding and pushing the amino acids: The path of the minima of the pushing potential attached to the first two amino acids and six snapshots of two amino acids being inserted and pushed to the right.

first simulation started with the modeled polypeptide in the pre-equilibrated ribosome. From this 5 ns trajectory we took the structure at 0.5 ns, 1 ns, and 1.5 ns simulation time. With these structures we started new simulations, calculating new velocities for all atoms according to the Boltzmann distribution at 300 K. All simulations were carried out with free MD.

4.5 Sequential Adding and Pushing of Amino Acids

In the ribosome amino acids are linked sequentially forming the growing nascent peptide chain. To mimic this process, we have developed a new simulation scheme, where the amino acids were put at the location of the PTC, covalently linked to the precursor amino acid and then pushed into the direction of the tunnel.

As a starting structure the pre-equilibrated ribosome was used. The first

amino acid was placed at the position of the crystallized pp-tRNA analog [37]. Then water molecules overlapping with the amino acid were deleted from the system. If the charge of the added amino acid was not zero, a randomly chosen water molecule outside the ribosome was replaced by an ion of opposite charge.

After energy minimization the system was equilibrated for 50 ps with position restraints on the backbone atoms of the new amino acid to allow equilibration of the amino acid side chain, the solvent around it, and the nearby ribosomal atoms. Subsequently, an FPMD (Force Probe Molecular Dynamics) pushing potential V_{push} was applied to the backbone atoms of the amino acid. The pushing potential moved with constant velocity v by 3.77 Å into the direction of the tunnel axis, which is the distance between C α atoms of consecutive amino acids in an extended polypeptide. The simulation time for the pushing was 100 ps or 200 ps for different simulations. While the amino acid was pushed, a harmonic potential V_{res} was applied to the backbone atoms, depending on the coordinates perpendicular to the pulling direction to keep the amino acid from moving sideways.

$$V_{push}(z_i, t) = \frac{1}{2}k(z_i - z_i^0 - vt)^2,$$

$$V_{res}(x_i, y_i) = \frac{1}{2}k_{res} [(x_i - x_i^0)^2 + (y_i - y_i^0)^2],$$

where $\mathbf{x}_i = (x_i, y_i, z_i)$ is the position of the amino acids backbone atoms, k is the force constant of the pulling potential, $\mathbf{x}_i^0 = (x_i^0, y_i^0, z_i^0)$ is the starting position of atom i , and $k_{res} = 10000 \frac{kJ}{\text{mol nm}^2}$ is the force constant of V_{res} .

Subsequently, the pushing potential was kept at the final position for another 100 ps.

Then the second amino acid was placed at the same position as the first. A covalent peptide bond was established to the first amino acid. Then the process of preparation, energy minimization, and pushing was repeated (figure 4.3), where position restraints and pulling potential were only applied to the most recently added amino acid, such that the rest of the peptide chain could freely explore a path through the tunnel.

To avoid drifting and rotation of the ribosome, center of mass translation and rotation around the center of mass were removed.

By this procedure we added amino acids according to the sequence of the SecM, amino acids 132–166, and Bpp, amino acids 1–36, peptides using different spring constants, $k = 2000 \frac{kJ}{\text{mol nm}^2}$ and $k = 8000 \frac{kJ}{\text{mol nm}^2}$ of the pulling potential. The two spring constants were used, because it was unclear whether the force resulting from the soft spring would suffice to push the growing polypeptide chain. To estimate the influence of the spring velocity v , simulations with a pulling time of 100 ps and 200 ps were calculated. Table 4.1 shows the parameters of the simulations.

sequence	$k \left[\frac{kJ}{\text{mol nm}^2} \right]$	$v \left[\frac{m}{s} \right]$	total simulation time [ns]
SecM	2000	3.77	8.75
SecM	8000	3.77	8.75
SecM	8000	1.885	15.75
Bpp	2000	3.77	9
Bpp	8000	3.77	9
Bpp	8000	1.885	16.2

Table 4.1: Parameters for the growing peptide chain simulations.

4.6 Analysis

4.6.1 Root Mean Square Deviation

The root mean square deviation (rmsd) is a measure for the similarity of two structures. The rmsd between two structures at times t_1 and t_2 was calculated by

$$rmsd(t_1, t_2) = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\mathbf{r}_i(t_1) - \mathbf{r}_i(t_2)\|^2},$$

where N is the number of atoms and $\mathbf{r}_i(t)$ is the position of atom i at time t . For most analyses in this work we were not interested in the overall rotation or translation of the molecules, so we first fitted the atom positions of one structure (t_2) on the atom positions of a reference structure (t_1) and then calculated the rmsd. This way we only measured the internal deviation of the atoms.

The rmsd to the starting structure ($t_1 = 0$) as a function of time $t_2 = t$ ($t_1 = 0$) typically increases until the system reaches equilibrium.

4.6.2 Root Mean Square Fluctuation

As a measure for the deviation of an atom from its mean position, root mean square fluctuation (rmsf) was used,

$$rmsf(x_i) = \sqrt{\frac{1}{T} \sum_{t_j}^T (x_i(t_j) - \langle x_i(t) \rangle)^2},$$

where x_i is a coordinate of the atom, T is the total time and t_j are the individual time frames.

4.6.3 Autocorrelation Function

To estimate the time difference Δt_s which separates independent measurements, in a series of measurements we used the autocorrelation function,

$$a(\Delta t) = \frac{\sum_t^N (x(t) - \langle x \rangle)(x(t + \Delta t) - \langle x \rangle)}{\sum_t^N (x(t) - \langle x \rangle)^2},$$

where t is the index of the measurement, N is the total number of measurements, and $x(t)$ is the value for measurement t . We fitted an exponential function $f(\Delta t) = e^{-\frac{\Delta t}{\tau}}$ to the autocorrelation function and used the decay constant τ as an estimate for Δt_s .

4.6.4 Distance Geometry

Distance Geometry [35] is a method for generating coordinates from the pairwise distances of n points. The problem of finding coordinates \mathbf{x}_i of the points compatible with these distances is connected to the diagonalization of the gramian matrix \mathbf{G} which is defined by

$$\mathbf{G} := \mathbf{X}^T \mathbf{X},$$

where $\mathbf{X} = (\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_n)$, where \mathbf{x}_i is the coordinate of point i , and $i, j \in [1, n]$. Therefore

$$G_{ij} = \mathbf{x}_i \cdot \mathbf{x}_j.$$

By the law of cosines one obtains:

$$\begin{aligned} (\mathbf{x}_i - \mathbf{x}_j)^2 &= \mathbf{x}_i^2 + \mathbf{x}_j^2 - 2|\mathbf{x}_i||\mathbf{x}_j| \cos(\angle(\mathbf{x}_i, \mathbf{x}_j)) = \mathbf{x}_i^2 + \mathbf{x}_j^2 - 2\mathbf{x}_i \cdot \mathbf{x}_j \\ \Rightarrow \mathbf{x}_i \cdot \mathbf{x}_j &= \frac{1}{2} (\mathbf{x}_i^2 + \mathbf{x}_j^2 - (\mathbf{x}_i - \mathbf{x}_j)^2) = \frac{1}{2} (d_{1i}^2 + d_{1j}^2 - d_{ij}^2), \end{aligned}$$

where d_{ij} is the distance between points i and j , and the origin of the coordinate system is chosen to be in an arbitrary point, here point number 1. So it is possible to calculate the gramian matrix by using the distances between the points.

Diagonalization of \mathbf{G} gives

$$\mathbf{L} = \mathbf{Y}\mathbf{G}\mathbf{Y}^T,$$

where \mathbf{L} denotes the matrix of eigenvalues and \mathbf{Y} is the corresponding transformation matrix. We choose

$$\tilde{\mathbf{X}} = \sqrt{\mathbf{L}}\mathbf{Y},$$

which gives

$$\tilde{\mathbf{X}}^T\tilde{\mathbf{X}} = \left(\sqrt{\mathbf{L}}\mathbf{Y}\right)^T \sqrt{\mathbf{L}}\mathbf{Y} = \mathbf{Y}^T\mathbf{L}\mathbf{Y} = \mathbf{G}.$$

Comparing this equation to the definition of the gramian matrix gives $\tilde{\mathbf{X}} = \mathbf{X}$.

Therefore

$$\mathbf{X} = \sqrt{\mathbf{L}}\mathbf{Y}$$

are the sought-after coordinates.

4.6.5 Principal Component Analysis

MD simulations produce a huge amount of data, e.g., the positions of all atoms for each time frame. So it is very important to separate the information which is relevant for answering the questions at hand from the irrelevant information. To study conformational changes of a molecule, most of the times one will not be interested in small fluctuations of side chains, but in large correlated motions of the whole molecule. A technique to separate degrees of freedom with large fluctuation from degrees of freedom with low

fluctuation is the principal component analysis (PCA) which is also called covariance analysis or essential dynamics [36].

This method uses the covariance matrix C , which expresses the correlation between atomic positions,

$$C = \langle (\mathbf{x} - \langle \mathbf{x} \rangle)(\mathbf{x} - \langle \mathbf{x} \rangle)^T \rangle$$

$$\Rightarrow C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle,$$

where x_i are the atomic coordinates with $i, j \in [1, 3N]$ and N is the number of atoms. $\langle \rangle$ denotes an average over time. As C is a symmetric matrix, it can be diagonalized by an orthogonal coordinate transformation T :

$$\mathbf{x} - \langle \mathbf{x} \rangle = T\mathbf{q} \Rightarrow \mathbf{q} = T^T(\mathbf{x} - \langle \mathbf{x} \rangle)$$

$$\Rightarrow (T\mathbf{q})^T = \mathbf{q}^T T^T = (\mathbf{x} - \langle \mathbf{x} \rangle)^T$$

T transforms C into the diagonal matrix $\Lambda = \langle \mathbf{q}\mathbf{q}^T \rangle$ of eigenvalues λ_i :

$$C = TT^T C TT^T = TT^T \langle (\mathbf{x} - \langle \mathbf{x} \rangle)(\mathbf{x} - \langle \mathbf{x} \rangle)^T \rangle TT^T$$

$$= T \langle T^T(\mathbf{x} - \langle \mathbf{x} \rangle)(\mathbf{x} - \langle \mathbf{x} \rangle)^T T \rangle T^T$$

$$= T \langle \mathbf{q}\mathbf{q}^T \rangle T^T$$

The columns of T are the eigenvectors of the covariance matrix C , also called principal or essential modes. The total positional fluctuation can be expressed by the eigenvalues λ_i :

$$\sum_i \langle (x_i - \langle x_i \rangle)^2 \rangle = \langle (\mathbf{x} - \langle \mathbf{x} \rangle)^T (\mathbf{x} - \langle \mathbf{x} \rangle) \rangle$$

$$= \langle \mathbf{q}^T T^T T \mathbf{q} \rangle = \langle \mathbf{q}^T \mathbf{q} \rangle = \sum_i \langle q_i^2 \rangle = \sum_i \lambda_i$$

The eigenvalues describe the variance of the position along the eigenvectors. We sort the eigenvalues and thus the eigenvectors, so that $\lambda_1 \geq \lambda_2, \dots, \lambda_{3N}$. Accordingly, the first eigenvectors represent the motion with the largest positional fluctuation.

By projecting the trajectory on the principal modes the principal components $q_i(t)$ are obtained:

$$\mathbf{q}(t) = T^T(\mathbf{x}(t) - \langle \mathbf{x}(t) \rangle)$$

Chapter 5

Results

In this chapter the results of our work are presented. First we tested the effects of large salt concentrations in our simulations on the diffusion of the ions. Then we compared the three different simulation systems of the ribosome to choose one system for carrying out the simulations.

To address the question whether the conformations and the dynamics of the polypeptides depend on their sequence, several simulations with different polypeptides inside the tunnel were carried out. To examine the differences of the polypeptides, we compared the conformations which the polypeptides adopted, calculated the interaction energies with the ribosome, and estimated entropy differences between different polypeptides. To investigate the dynamics of the polypeptides Principal Component Analyses were performed. Several simulations of the nascent peptide chain growth and movement through the tunnel were performed to study the mechanical work necessary to push the chain through the tunnel and to examine the pathway of the growing polypeptides.

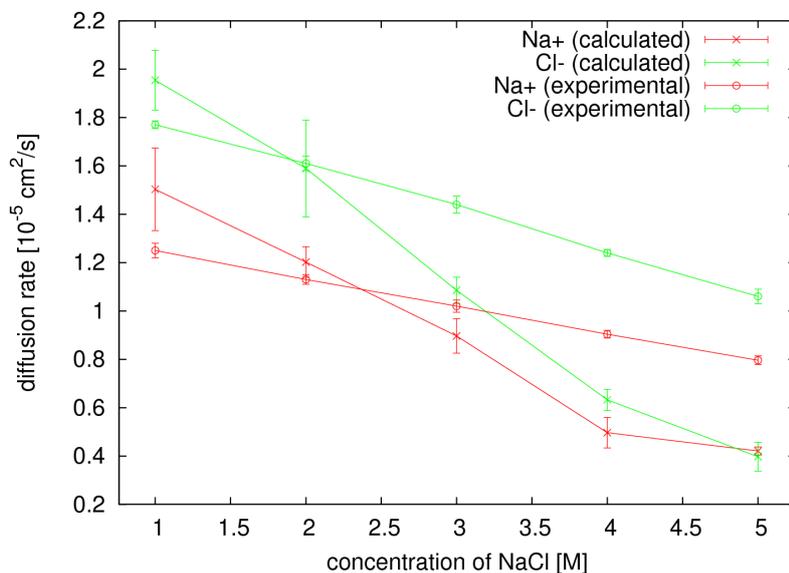


Figure 5.1: Comparison of experimental and calculated diffusion coefficients for different concentrations of sodium chloride.

5.1 Salt Concentration

Since the physiological salt concentration of *Haloarcula Marismortui* is extremely large, we simulated NaCl solutions of different concentrations for 3 ns each and compared the diffusion coefficient of the ions in the simulations with experimental results.

For the calculation of the diffusion coefficient D of sodium chloride ions, their mean square deviation from the starting position over time was calculated. Then we performed linear regressions over three 1 ns intervals and calculated the diffusion coefficients via the Einstein-Smoluchowsky equation and their mean and variance were calculated.

The experimental values for the diffusion coefficients were measured at a temperature of 25 °C via the measurement of the concentration decay of radioactive ions from a capillary which was put in bath with non-radioactive

ions [38, 39]. We expected the calculated diffusion coefficient for low concentrations to be larger than the measured ones, because the self diffusion coefficient for the tip3p water model, which is reported to be $5.19(8) \cdot 10^5 \text{cm}^2/\text{s}$ [40], is larger than the measured diffusion coefficient $2.272 \cdot 10^5 \text{cm}^2/\text{s}$ [41]. If the diffusion coefficient of water molecules is overestimated in the simulations, the diffusion coefficient of ions should also be too large.

Figure 5.1 compares the calculated diffusion rates with the measured ones. As can be seen, the calculated values were larger than the experimental ones for low concentrations, but decreased faster with increasing salt concentration. At a concentration of 2 M they agreed best, and for larger concentrations the calculated values were lower than the experimental ones. Consequently, we chose to perform our simulations at a concentration of 2 M.

5.2 Equilibration

After energy minimization, equilibration of the solvent, and equilibration of the modeled loops the system was equilibrated without any position restrains for 3 ns. To check if the system converged to an equilibrium during this time and to compare regions with different distances to the tunnel, we calculated the rmsd for atoms in these regions. We chose the distance ranges, smaller than 2 nm, between 2 and 2.5 nm, and more than 2.5 nm, because these are the regions where the atoms were chosen to be free, position restrained and fixed in the argon-box-simulation system, respectively. The rmsd was calculated for every atom of the ribosome, but before each calculation of the rmsd, the positions of the protein and RNA backbone atoms were fitted to their positions in a reference structure. The rmsd is shown in figure 5.2.

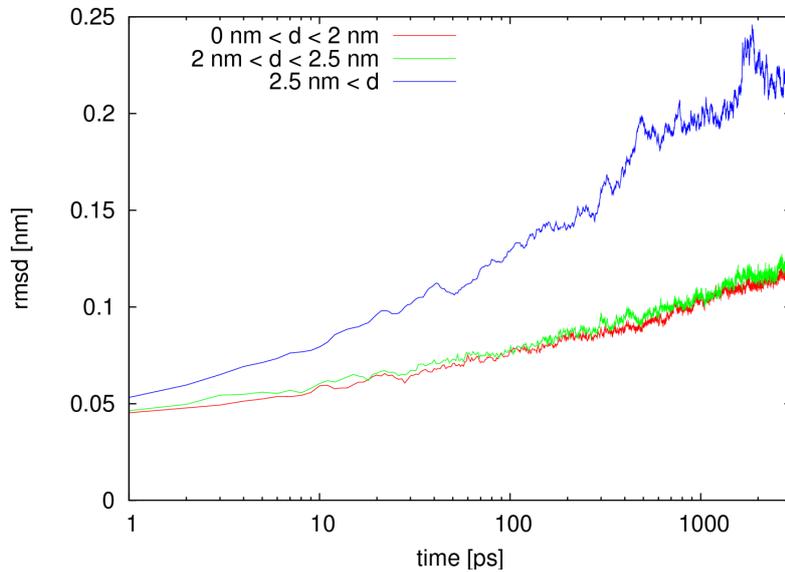


Figure 5.2: Structural Deviation of the ribosome during a 3 ns equilibration phase: Shown is the rmsd for atoms in different distances d to the tunnel against the simulated time.

Results

The rmsd reached at the end of the simulation is larger for the outer regions, which means that the equilibrated model deviates more from the crystal structure [16] at the outer region than close to the tunnel. The rmsd for atoms in the region between 2 and 2.5 nm is similar to the region closer than 2 nm. As can be seen in the semi logarithmic plot, the rmsd rises faster than $\ln(t)$, which shows that the simulation did not converge during the simulation time of 3 ns.

Discussion

The fact that the rmsd for atoms more than 2.5 nm away from the tunnel was larger than for the atoms closer than 2.5 nm, could have several reasons. The ribosome was crystallized at a temperature of 100 K and not in physiological solution, which should affect the conformation. It is also likely that the limitations of the force field led to a different conformation. As the rmsd was, with values in the range of 2.5 Å, rather small and the computational effort is large for simulating a system of this size, we did not equilibrate the system any longer.

The fluctuations of the rmsd for the outer region were larger than for the region close to the tunnel. This shows that the outer regions are more flexible than the region containing the tunnel.

5.3 Three levels of complexity

To reduce the computational effort and to investigate if it is necessary to explicitly simulate the whole system, three different simulation systems were set up. They differed in the size of the system subjected to molecular dynamics. Free system denotes the system where all atoms were treated with free MD. Free/position-restrained/fixed system denotes the system where all the water molecules, all the ions, and only ribosomal atoms in a region of 20 Å around the tunnel were freely simulated. Argon box system denotes the system where water molecules and ions were only allowed to move inside the argon box.

To investigate the efficiency of these systems, a simulation of 100 integration

steps was carried out for each system. Each simulation was calculated in parallel on 20 CPUs. Then we compared two of the systems to examine the influence of the approximations of the system on the dynamics of the region containing the tunnel.

Results

system	time per step [s]	days per ns
free	1.74	10.06
free / position-restrained / fixed	1.91	11.05
argon box	0.65	3.74

Table 5.1: Comparing the three simulation systems: benchmarks based on test runs with 100 steps.

Benchmarks of the test runs of the three different systems are shown in table 5.1, where time per step is the computation time for one MD step (2 fs) and days per ns are the days needed for the computation of 1 ns simulation time. Using the second system, we would gain no speed in comparison to the free system, the calculations would be even slower. In GROMACS 3.3.1 [26] all the forces were calculated first and the resulting velocity was then only applied to atoms which are not fixed. The calculation of the forces is the most time consuming part of the simulation, therefore, there is no increase in efficiency by using the second system. Because GROMACS is open source software it would have been possible to change GROMACS such that these forces are not calculated, but the effort would have been too large for us. The fact that test simulation using this system was even slower could be due to

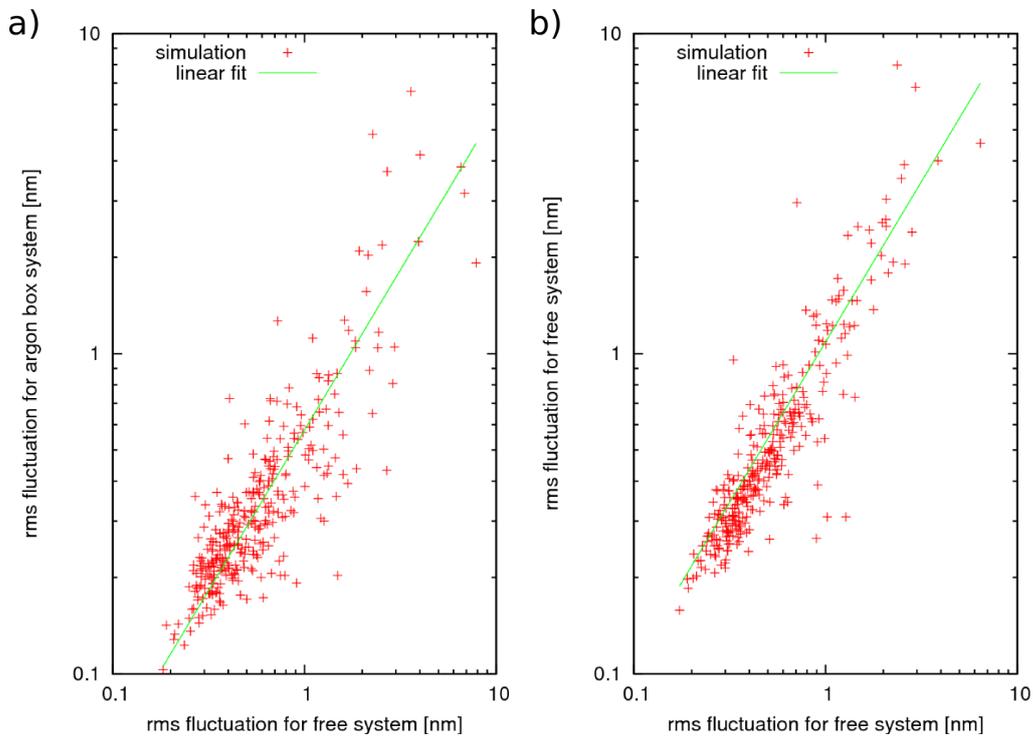


Figure 5.3: Comparison of the free and the argon box system: **a)** Rmsf of the atoms in the range of 10 \AA calculated from a 1 ns trajectory of the free and of the argon box system. **b)** Rmsf for the atoms calculated from the first 0.5 ns and the last 0.5 ns of the trajectory of the free system.

the additional calculation of the potential for the position restrained atoms. The test simulation using the third system, the argon box system, was 2.7 times faster than the one using the free system. The number of atoms in the system was reduced to $\sim 30 \%$, because the water molecules outside the argon box were not considered.

To test to which extent the approximations done in the argon box system affect the dynamics of the atoms in the region of the tunnel, two simulations were carried out, one in the argon box system and one in the free system, with a simulation length of 1 ns each and starting with the equilibrated system.

Then the rmsf of the phosphorus of the nucleotides and the $C\alpha$ atoms of the amino acids, which are closer than 10 Å to the tunnel, were calculated for both simulations. In figure 5.3a the rmsf of the argon box system atoms is plotted against their rmsf in the free system. For comparison, figure 5.3b shows the plot of the rmsf of the free system atoms calculated from the first 0.5 ns against their rmsf in the last 0.5 ns of the trajectory.

If the considered atoms in the free system were as flexible as the atoms in the argon box system, one would expect a line through origin with a slope of 1, given a trajectory that is long enough. We fitted a line through origin to the points and the calculated slope was 0.58 for the comparison of free and the argon box system and 1.09.

Discussion

The limited simulation time results in a spread of the data points in the rmsf plots around the line. The spread is larger in the plot comparing the two different systems (see figure 5.3b). This additional spread can be explained by the effects of the different systems.

The shift of the lines in figure 5.3a and 5.3b shows that the motion of the atoms is dramatically damped in the argon box system. Therefore we carried out all simulations in the free system, despite the fact that simulations using the argon box system would be calculated 2.7 times faster than simulation calculated using the free system.

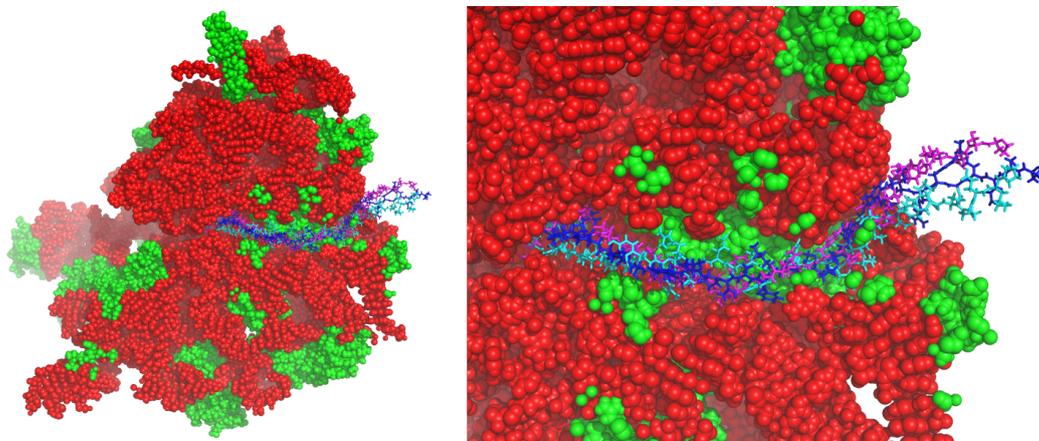


Figure 5.4: Polypeptides modeled into the tunnel. A cross section of the ribosome is shown. Atoms of the ribosomal proteins are represented by green and rRNA atoms by red spheres. The polypeptides SecM166, SecM164, and Bpp are drawn in blue, magenta and cyan sticks, respectively.

5.4 Conformations and Dynamics of Polypeptides inside the Tunnel

Three different polypeptides, SecM166, SecM164, and Bpp, were modeled into the tunnel. The polypeptide SecM induces a translation arrest, when the amino acid Pro166 is added to the peptide. SecM166 denotes the polypeptide at the arrest point. SecM164 denotes the polypeptide two amino acids before the arrest point. Bpp denotes a small non-arresting polypeptide. To examine whether differences in conformations and dynamics of the polypeptides inside the tunnel can be seen, they were simulated with four different starting structures each and with a simulation time of 5 ns per simulation.

5.4.1 Modeled Polypeptides

The first polypeptide built into the tunnel was a segment of the SecM sequence, amino acids 132–166, from now on referred to as SecM166. Pro166 is the amino acid where the translation arrest occurs and the length of the polypeptide was chosen such that the end of the peptide is outside the tunnel. The second polypeptide consists of the amino acids 132–164 of the SecM protein, from now on referred to as SecM164. This polypeptide describes the translation state two amino acids before the translation arrest occurs. As the third polypeptide we chose amino acids 2–26 of bovine pancreatic polypeptide (Bpp), a short peptide, which forms an α -helix in solution. Bpp was chosen, because its length is similar to the other polypeptides used, because we can observe whether it forms an α -helix inside the tunnel or not, and because it is a non-arresting sequence. The modeled polypeptides inside the ribosomal tunnel are shown in figure 5.4.

5.4.2 Equilibration of Polypeptides

The modeled starting structures were equilibrated for 5 ns. From the trajectory snapshots were taken at 0.5 ns, 1 ns, and 1.5 ns (figure 5.5a), respectively. With these snapshots as starting structures, new simulations were started with new velocities. To check how well the polypeptides were equilibrated we calculated the rmsd for all the trajectories.

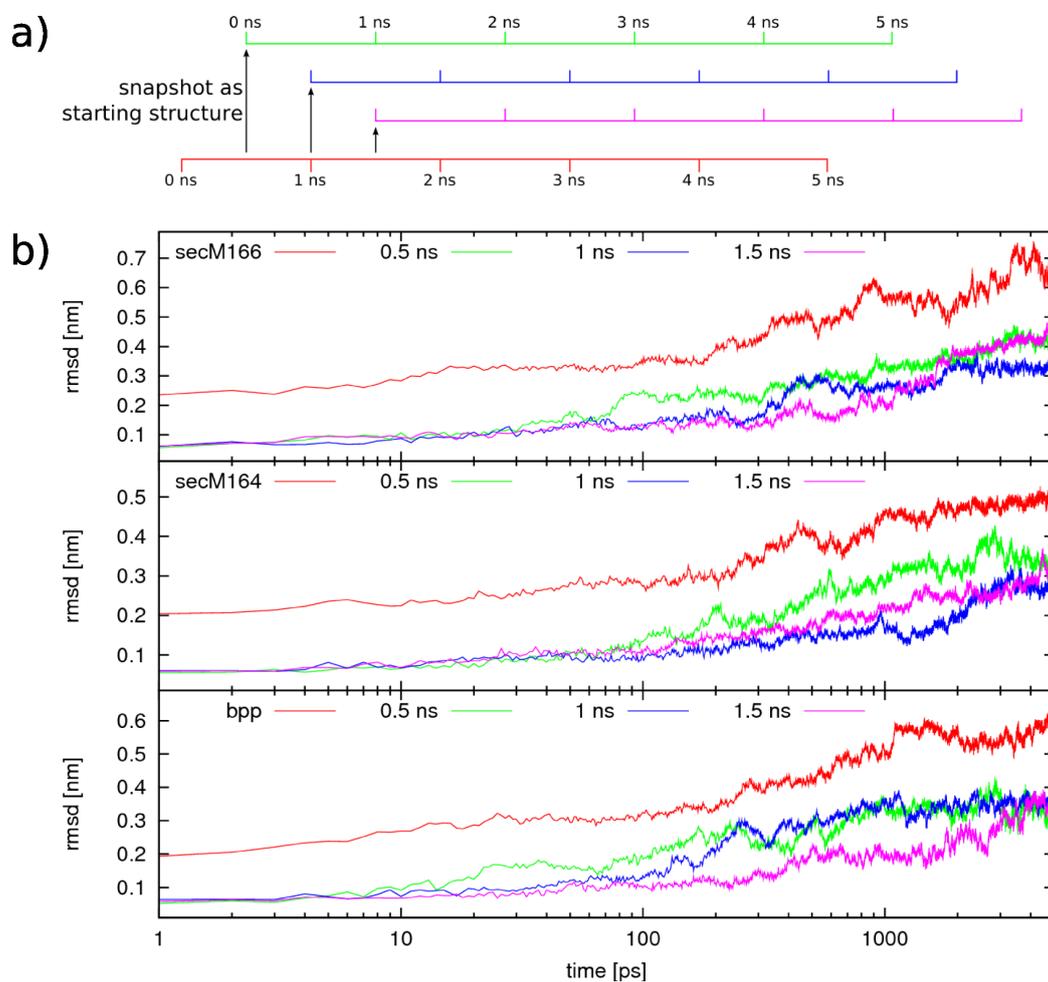


Figure 5.5: Equilibration of polypeptides inside the tunnel: **a)** Scheme of the simulations with different starting structures. The first simulation (red line) started with the modeled structure. For the second simulation (green) a snapshot of the first trajectory at 0.5 ns was taken as the starting structure. For the third (blue) and the fourth (pink) simulation, snapshots at 1 ns and 1.5 ns were taken as starting structures. **b)** Rmsd of the polypeptides, SecM166, SecM164, and Bpp, inside the tunnel for four different starting structures each.

Results

After fitting the polypeptide backbone atom positions to their positions in the starting structure, the rmsd was calculated for all polypeptide atoms with respect to the starting structure of each simulation. Figure 5.5 shows the rmsd for all simulations of the three polypeptides. The rmsd of the first trajectory (red) of each polypeptide is larger than the rmsd of the other trajectories. The rmsd of several trajectories flattens at the end.

Discussion

The rmsd is larger for the first trajectory, because it started with the modeled structure and was partially equilibrated before the snapshots for starting structure of the other simulations were taken. Thus the following simulations started with a pre-equilibrated structure and the trajectories did not move in configurational space as far as the first trajectory.

For further analyses, only the last 2 ns of each of the simulations were used to minimize the dependence on the starting structure, if not stated otherwise.

5.4.3 Conformations of the Polypeptides

To estimate the influence of the starting structure and to analyze the sequence dependence of the conformations of the polypeptides, we compared several structures from the trajectories of the twelve simulations and measured their similarity by rmsd calculation.

If the fully equilibrated conformations of the different polypeptides were similar, we would expect that the rmsd of pairs of structures of different polypeptides would decrease with increasing simulation time. This would show that

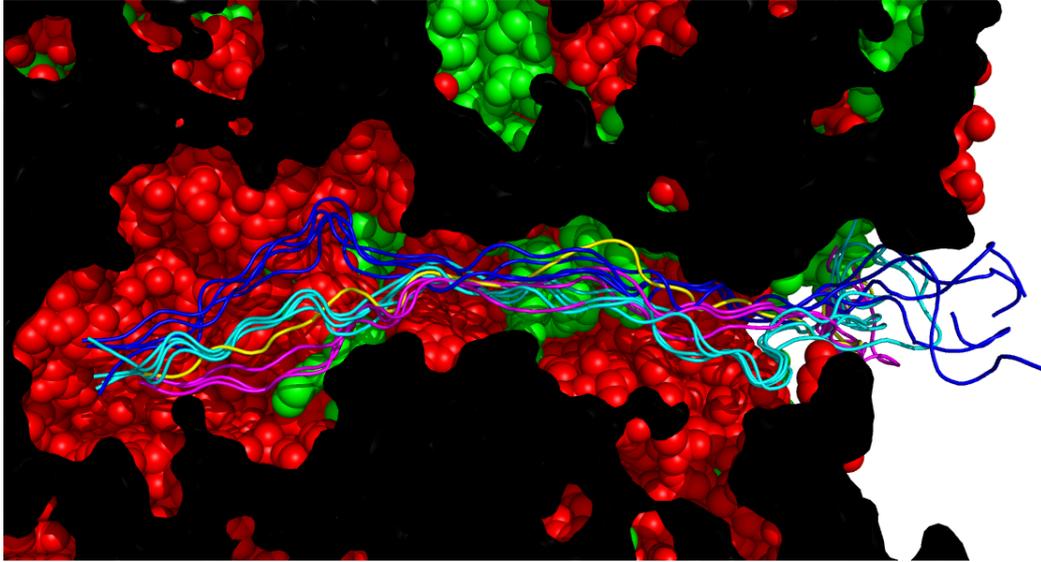


Figure 5.6: Cross section of the ribosome with the end structures of the polypeptides from all trajectories inside the tunnel. Atoms of the ribosomal proteins are represented by green and rRNA atoms by red spheres. The end structures of the polypeptides SecM166, SecM164, and Bpp are drawn in blue, magenta and cyan cartoon representation, respectively. The end structure of the polypeptide SecM164 from the trajectory starting with the snapshot taken at 0.5 ns is colored in yellow. The cross section plane is drawn in black.

the similarity of the structures increases with increasing time.

If the equilibrated conformations of the different polypeptides were different, we would expect that the different trajectories of one polypeptide would explore a compact region in the conformational space, independent of the starting structure of the trajectory. In this case the rmsd of different trajectories of one polypeptide would be small in comparison to the rmsd of different trajectories of different polypeptides.

Results

Figure 5.6 shows a superposition of the end structures of the polypeptides inside the exit tunnel from all twelve trajectories. As can be seen, the different end structures of one polypeptide show similar conformations inside the tunnel. On the contrary, the end structures of different polypeptides show different conformations, especially in the part of the tunnel which is close the PTC.

To measure the difference of the conformations of the polypeptides with respect to the tunnel, and not only their internal conformations, before each rmsd calculation the positions of the backbone atoms of ribosomal proteins and rRNA, excluding the polypeptide, were fitted to their positions in a reference structure. For all rmsd calculations only the backbone of the first 26 amino acids of the polypeptides, counted from the PTC, was taken into account, because these amino acids are located inside the tunnel and showed a relatively stable conformation.

The first row of table 5.2 shows the two polypeptides taken into account.

	rmsd of the modeled polypeptides	mean of rmsds of pairs of end structures from two polypeptides
SecM166-SecM164	0.982 nm	1.04 nm
SecM166-Bpp:	0.459 nm	0.65 nm
SecM164-Bpp:	0.908 nm	0.82 nm

Table 5.2: Comparison of structures for different polypeptides.

For the second row the rmsd of the two modeled structures was calculated.

The third row contains the mean of the rmsd of each end structure of the first polypeptide with each end structure of the second polypeptide. Table

	rmsd of end structures with respect to the modeled structure	rmsd of pairs of end structures
SecM166:	0.40 nm	0.26 nm
SecM164:	0.30 nm	0.25 nm
Bpp:	0.37 nm	0.20 nm

Table 5.3: Comparison of structures of one polypeptide.

5.3 shows the means of the rmsd for different structures of one polypeptide, listed in the first row. The second row shows the mean of the rmsd of the end structures in reference to the modeled structure. The third row contains the mean of the rmsd of all pairs of end structures of the polypeptide.

The results show that the end structures of different polypeptides differ in the same range as the modeled starting structures. Accordingly, the polypeptides did not find a common conformation independent of their sequence. The mean of the rmsds comparing the end structures of one polypeptide is smaller than the mean of the rmsds of the end structures and the modeled structures. So the distance the polypeptides travelled in conformational space is larger than the distance of the end points of their ways, which means that we ended in a similar conformation although we started with different structures and different initial velocities.

To examine in which way the structures varied over the simulation time, we averaged the polypeptide backbone coordinates over 250 ps intervals of the whole 5 ns trajectories and calculated pairwise rmsds for all resulting

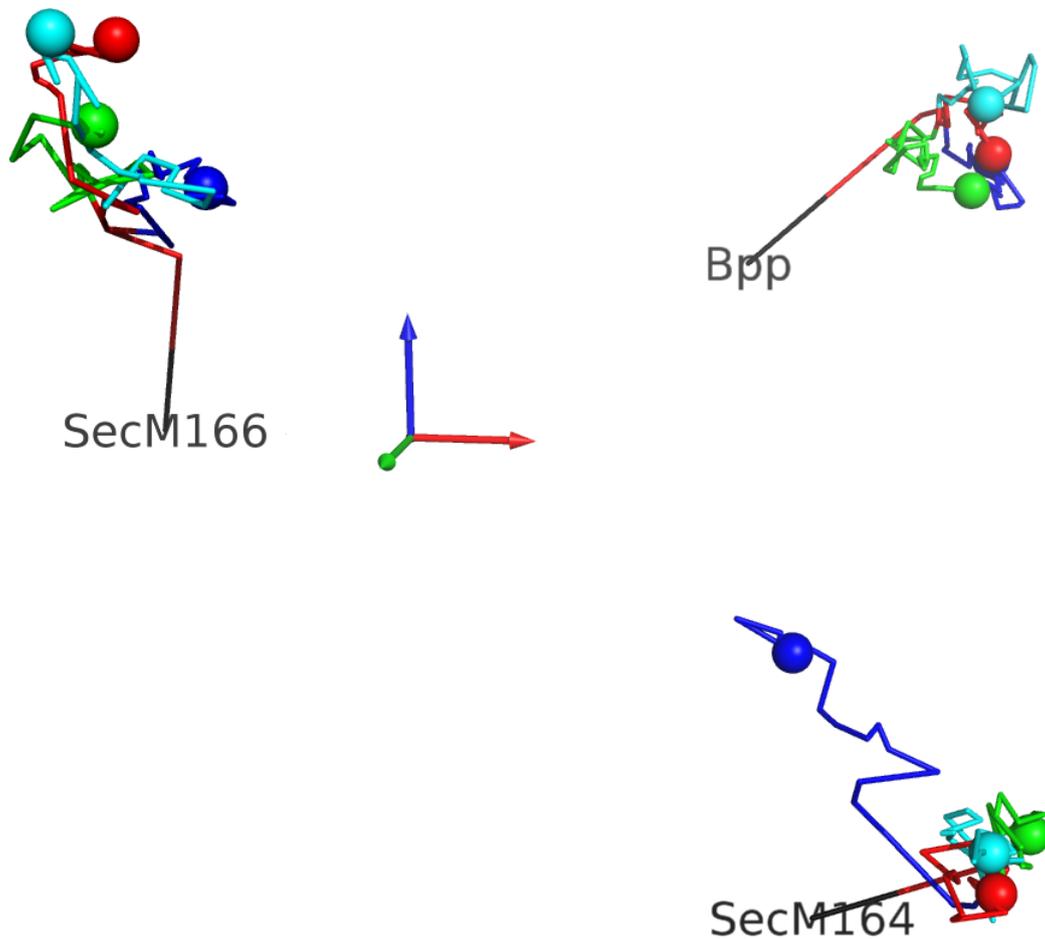


Figure 5.7: Comparison of the trajectories: The first three eigenvectors are drawn as red, green and blue arrows, respectively. The coordinates of the averaged polypeptide backbone structures are projected on these eigenvectors and linked chronologically with lines. The modeled starting structure for each polypeptide is colored in black. Red lines represent the trajectories starting from this structure. Blue, green, and cyan lines represent the trajectories starting at the snapshots from the first trajectory at times 500 ps, 1000 ps and 1500 ps, respectively. The positions of the end structures are drawn as spheres.

structures including the modeled starting structures. As there were three different polypeptides, four 5 ns simulations each, we obtained 240 averaged structures and 3 modeled structures. Thus we obtained a distance matrix with a dimension of 243.

Using Distance Geometry [35], we obtained a set of eigenvectors describing the structural variation and the coordinates of the structures in this coordinate systems. The distance between two points in this coordinate system is the rmsd of the two structures. The first three eigenvectors described more than 93 % of the structural variation. We projected the coordinates of the structures on these eigenvectors, obtaining a three dimensional coordinate set.

In these coordinates (figure 5.7) the trajectories quickly move away from the modeled polypeptides, showing that the modeled structures were not in a free energy minimum. Most of the trajectories of one sequence sampled a compact region of space, which suggests that these polypeptides were trapped in at least a local minimum. The trajectory of polypeptide SecM164, which started with the 0.5 ns snapshot of the first trajectory (represented by blue lines), explored a region far away from the other trajectories and might have explored a different minimum. In figure 5.7 the end structure of this trajectory is colored in yellow and it can be seen that its conformation varies significantly from the other conformations (magenta) of the SecM164 polypeptide. As can be seen, the trajectories of different polypeptides were not approaching each other.

Discussion

The fact that the different trajectories of different polypeptides were not approaching each other during the simulation time and that trajectories of the same polypeptide sampled a rather compact region of space suggests that the fully equilibrated polypeptides have different conformations.

But the limited simulation time and the fact that the starting structures of the trajectories of one polypeptide were all snapshots from one trajectory, separated by only 500 ps, and thus were not independent, show the need for longer simulations with independent starting structures to strengthen our conclusions.

5.4.4 Interactions of the Polypeptides with the Ribosome

A reason for the different conformations of different polypeptides in the tunnel might be the interactions of the amino acids with the ribosomal atoms. To compare the interactions of the different polypeptides with the ribosome, the Lennard-Jones and the Coulomb energies of the atoms of the polypeptides with the ribosomal atoms were calculated. For the atoms of each amino acid of the polypeptide, these energies were summed up for each time frame. Thus we obtained the interaction energy of every amino acid with ribosomal atoms as a function of time.

The mean and variance of the interaction energy were calculated for each amino acid. To estimate the number of statistically independent measurements, we calculated the autocorrelation function $a(\Delta t)$ for the energy as a

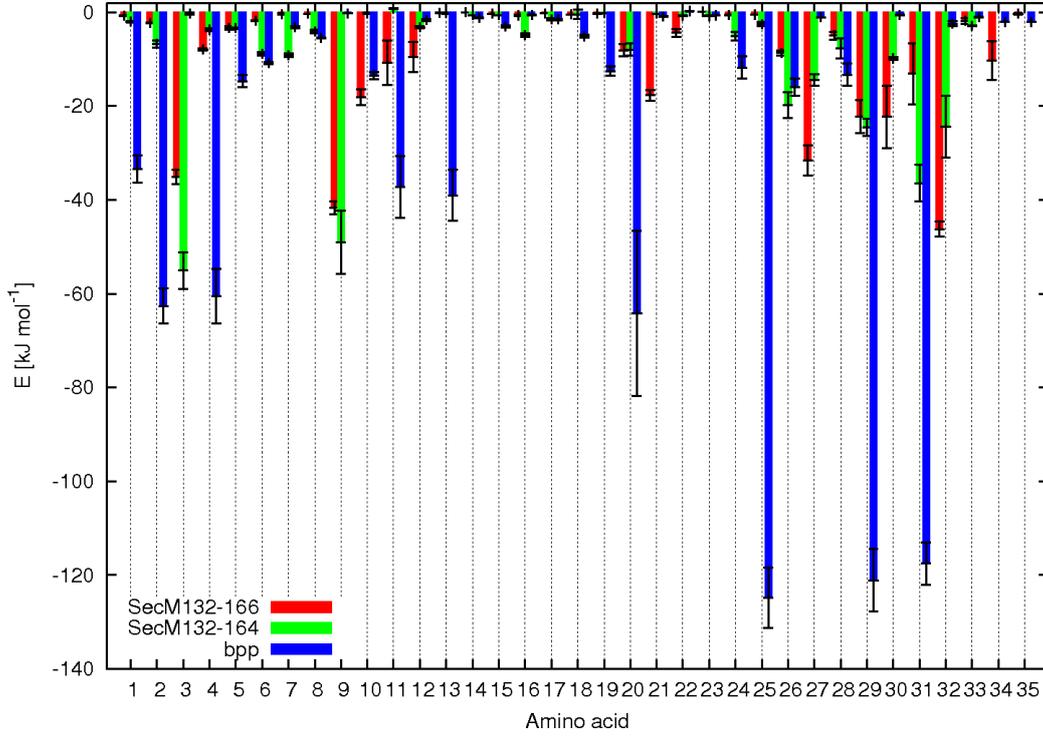


Figure 5.8: Interaction of the polypeptides with the ribosome. The averaged sum of the Lennard-Jones and Coulomb energies is plotted against the amino acid index for the polypeptides SecM166, SecM164, and Bpp. The amino acid index is counted beginning with the amino acid at the peptidyl transferase center (PTC).

function of time $E(t)$,

$$a(\Delta t) = \frac{\sum_t^N (E(t) - \langle x \rangle)(E(t + \Delta t) - \langle x \rangle)}{\sum_t^N (E(t) - \langle E \rangle)^2}.$$

Then we fitted an exponential decay and used the decay constant as an estimation of the time interval of independent measurements. The variance was divided by the number of independent measurements yielding the error estimate for the interaction energies.

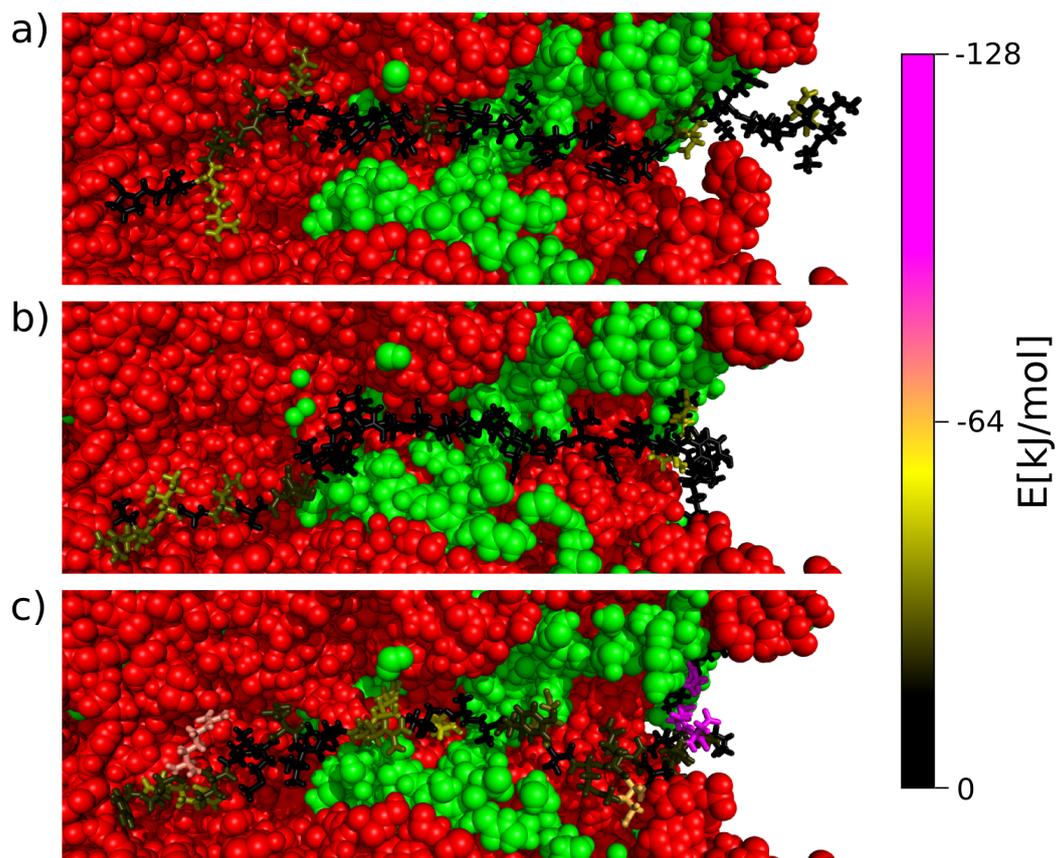


Figure 5.9: Interaction energies of the polypeptides with the ribosome: Cross section of the ribosome with an end structure of each polypeptide inside the tunnel. Ribosomal protein atoms are represented by green and rRNA atoms by red spheres. The amino acids of SecM166 (**a**), SecM164 (**b**), and Bpp (**c**) are drawn in sticks colored according to their interaction energies.

Results

The averaged interaction energies are plotted in figure 5.8. Figures 5.9a, 5.9b, and 5.9c show the polypeptides SecM166, SecM164, and Bpp, respectively, inside the tunnel. The amino acids of the polypeptides are colored according to their average interaction energies. As can be seen, the averaged interaction energies varied in a range of $0 \frac{\text{kJ}}{\text{mol}}$ and less than $-120 \frac{\text{kJ}}{\text{mol}}$. The energies of the amino acids of different polypeptides differed significantly during our simulations.

The overall interaction energy of the amino acids of the non-arresting polypeptide Bpp was lower than the interaction energies of the SecM polypeptides, so it was bound more tightly to the atoms forming the tunnel wall. Bpp amino acids close to the exit of the tunnel showed the lowest interaction energies. But also inside the tunnel several Bpp amino acids had lower energies than SecM amino acids. The amino acids of the SecM polypeptides had the lowest energies in the amino acids close to the PTC and close to the exit of the tunnel.

Discussion

During our simulations there were significant differences of interaction energies between the different polypeptides which suggests that their different conformations could be stabilized by the interactions with the ribosomal atoms. As the interaction energies of the SecM amino acids in the first third of the tunnel were larger than the energies inside the rest of the tunnel, these amino acids could be important for the mechanism of the translation arrest.

5.4.5 Dynamics of the Polypeptides

To analyze whether the dynamics of the polypeptides inside the tunnel depend on their sequence, we performed a PCA analysis on the trajectories of the twelve simulations. The backbone atoms of amino acids 1-26, counted from the PTC, were taken into account. The polypeptide trajectories were fitted to the positions of the polypeptide backbone atoms of a reference structure for the analysis of their internal dynamics. To analyze the dynamics of the polypeptides with respect to the tunnel, the trajectories were fitted to positions of C α and P atoms of amino acids or nucleotides closer than 3 Å to the polypeptides. The covariance matrices of the backbone atoms were calculated using the combined trajectory of all simulations to obtain a set of eigenvectors where the eigenvectors with the largest eigenvalues describe the largest fluctuations of the backbone atoms of the three polypeptides.

Results

In figure 5.10 the eigenvalues of the first one hundred eigenvectors are plotted. The fast decay of the eigenvalues shows that a huge amount of the peptide dynamics can be described by the first eigenvectors. Further analysis was restricted to the first two eigenvectors which describe 73 % and 83 % of the internal fluctuations and the fluctuations with respect to the tunnel, respectively.

Figure 5.11 shows the projections of the trajectories on the eigenvectors from the PCA of the internal motion. Figure 5.12 shows the projections of the trajectories on the eigenvectors from the PCA of the motion with respect to the tunnel. The complete trajectories (0-5 ns) and the last two ns (3-5 ns)

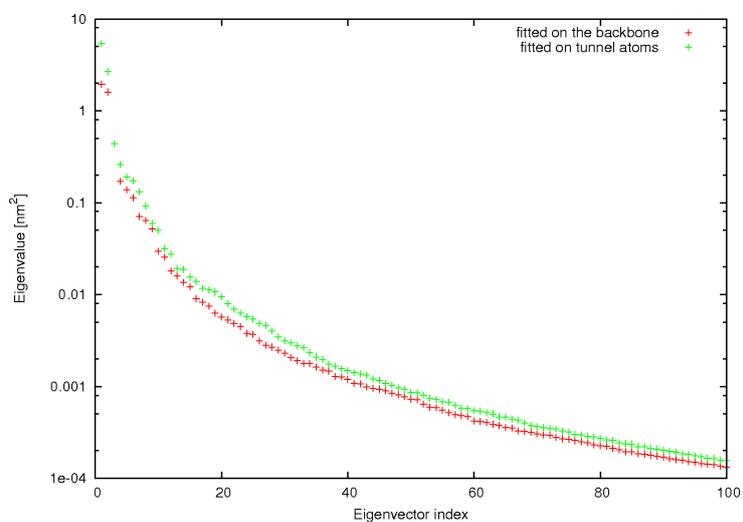


Figure 5.10: Eigenvalues for the first 100 eigenvectors of the covariance matrixes calculated from the polypeptide trajectories.

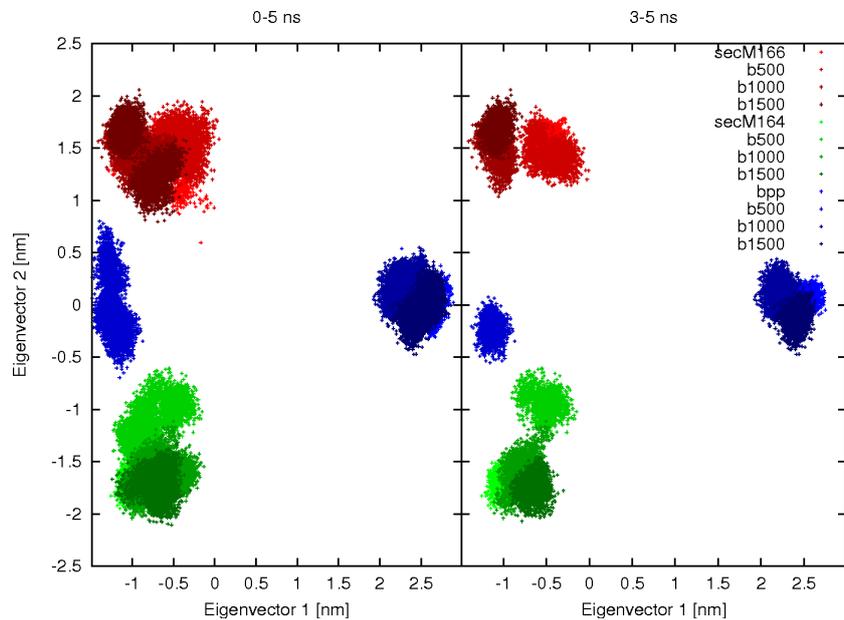


Figure 5.11: Internal dynamics of the polypeptides: Trajectories projected on the first two eigenvectors.

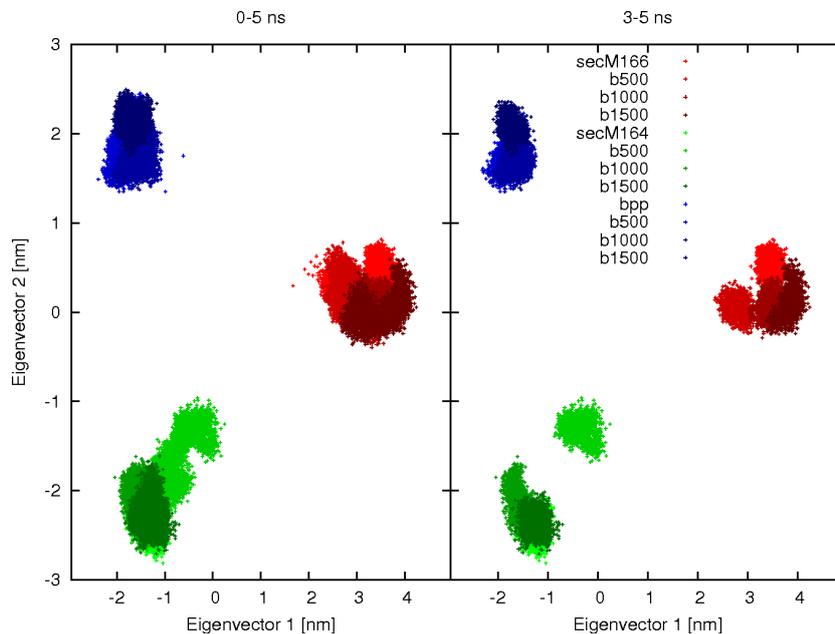


Figure 5.12: Dynamics of the polypeptides in reference to the tunnel: Trajectories projected on the first two eigenvectors.

were projected separately to display their development in time.

Discussion

In both plots the projections of trajectories of different polypeptides are clearly separated. This could be due to the fact that the simulations were not converged or that the dynamics of the polypeptides are different. The sampled regions in the first two eigenvectors of the individual trajectories of one polypeptide, represented by different shades, are close to each other and overlap, except for one trajectory of the Bpp polypeptide and one trajectory of the SecM164 polypeptide which also showed a different conformation (see 5.4.3).

5.4.6 Entropy Differences

To estimate the relative entropies of the polypeptides in the tunnel, the method of Schlitter [42] was used. The phase space density of a system is fitted by a multivariate Gaussian density to the trajectory using the covariance matrix. For an approximate calculation of entropy difference ΔS between two conformations A and B , with covariance matrices $\sigma_{\mathbf{A}}$ and $\sigma_{\mathbf{B}}$, we used:

$$\Delta S = S_B - S_A \sim \Delta S' = S'_B - S'_A = 0.5k \ln \left(\frac{\det(\sigma_{\mathbf{B}} + \mathbf{M}^{-1} \frac{\hbar^2}{kTe^2})}{\det(\sigma_{\mathbf{A}} + \mathbf{M}^{-1} \frac{\hbar^2}{kTe^2})} \right),$$

where \mathbf{M} is the mass matrix which contains the masses on the diagonal and is zero elsewhere, k is the Boltzmann constant, $T = 300$ K is the temperature, and e is Euler's number. S' is a measure for the size of the sampled phase space.

The covariance matrix for each polypeptide was calculated for different intervals of the last two nanoseconds of the four simulations to see how much the results depend on the sampling.

Results

polypeptide A	polypeptide B	$T\Delta S' = T(S'_B - S'_A)$		
		3-4 ns	4-5 ns	3-5 ns
Bpp	SecM166	7.10 $\frac{\text{kJ}}{\text{mol}}$	15.05 $\frac{\text{kJ}}{\text{mol}}$	9.69 $\frac{\text{kJ}}{\text{mol}}$
SecM166	SecM164	5.17 $\frac{\text{kJ}}{\text{mol}}$	-6.96 $\frac{\text{kJ}}{\text{mol}}$	2.36 $\frac{\text{kJ}}{\text{mol}}$
Bpp	SecM164	12.27 $\frac{\text{kJ}}{\text{mol}}$	8.10 $\frac{\text{kJ}}{\text{mol}}$	12.05 $\frac{\text{kJ}}{\text{mol}}$

Table 5.4: Entropy difference estimation: $T\Delta S'$ for different intervals of the trajectories.

The result for the calculation of $T\Delta S'$ for all pairs of different polypeptides and for three time intervals are shown in table 5.4.

Discussion

The values for $T\Delta S'$ varied largely for different time intervals and thus largely depended on the sampling. The variation was in the same range as the obtained values, which shows that the trajectories are too short to estimate the entropy differences by this method.

5.5 Movement of the Polypeptide through the Tunnel

The pathway of growing nascent peptide chains through the ribosomal exit tunnel is not known. To examine this pathway, six simulations with elongating polypeptides inside the tunnel were performed. To mimic the effect of polypeptide synthesis, the first amino acid was placed at the peptidyl transferase center (PTC) and was then pushed by the distance which the $C\alpha$ atoms of two adjacent amino acids have in an extended polypeptide. The amino acids were sequentially added to the growing polypeptide and then pushed into the tunnel.

Results

In figure 5.13 the z -positions of the $C\alpha$ atoms, where z is the direction of the tunnel axis, and the forces applied to the most recently added amino acids by the pushing potential are plotted for the simulation of the SecM polypeptide

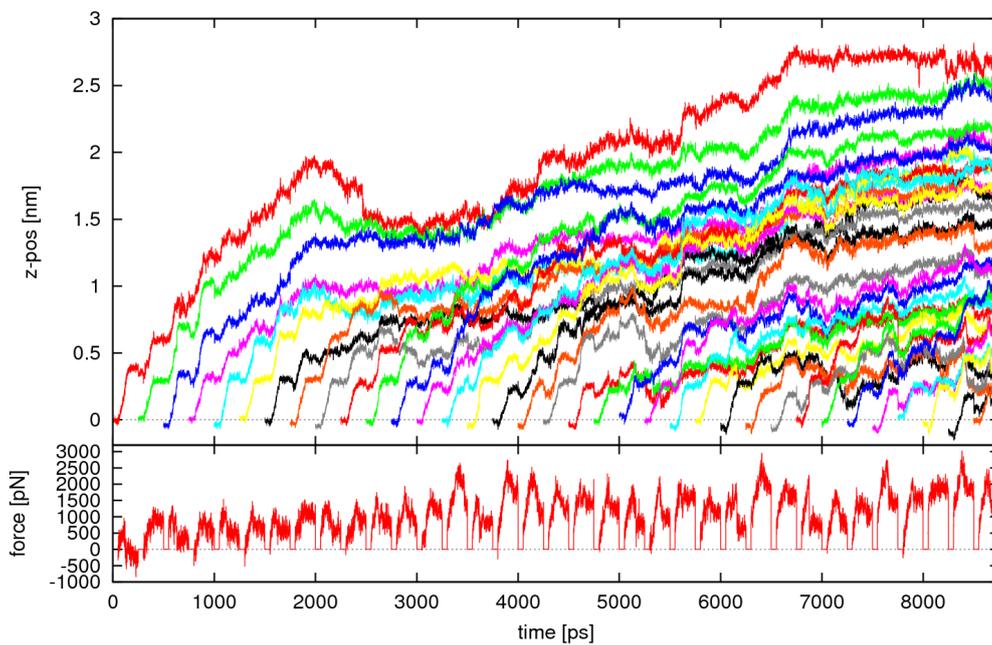


Figure 5.13: Movement of the growing polypeptide through the exit tunnel: Sequential adding and pushing of SecM amino acids with spring constant $k = 8000 \frac{kJ}{\text{mol nm}^2}$ and pushing velocity $v = 3.77 \frac{m}{s}$. Z-position of the polypeptide C α atoms and the forces acting on the amino acids.

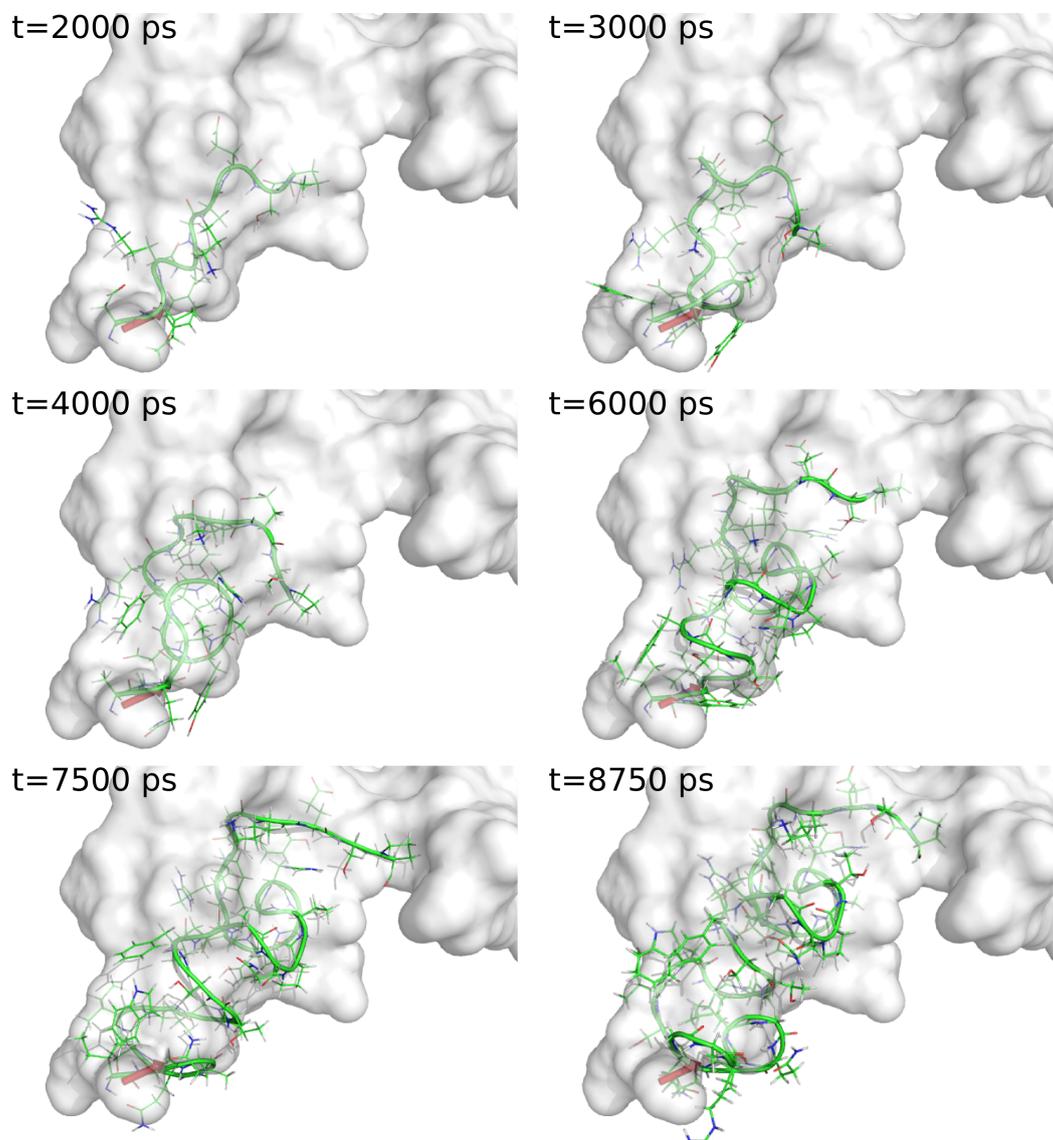


Figure 5.14: Simulation of nascent chain growth and movement through the exit tunnel: Sequential adding and pushing of SecM amino acids with spring constant $k = 8000 \frac{kJ}{\text{mol nm}^2}$ and pushing velocity $v = 3.77 \frac{m}{s}$. Snapshots from the trajectory are shown, where the backbone of the polypeptide is drawn in green cartoon representation and the amino acids are drawn as lines. The tunnel surface which was calculated from the equilibrated ribosome (see 5.4.2) is drawn in grey. The red arrow starts at the position where the amino acids were added and where the pulling potential started. The tip of the arrow indicates the end position of the pulling potential.

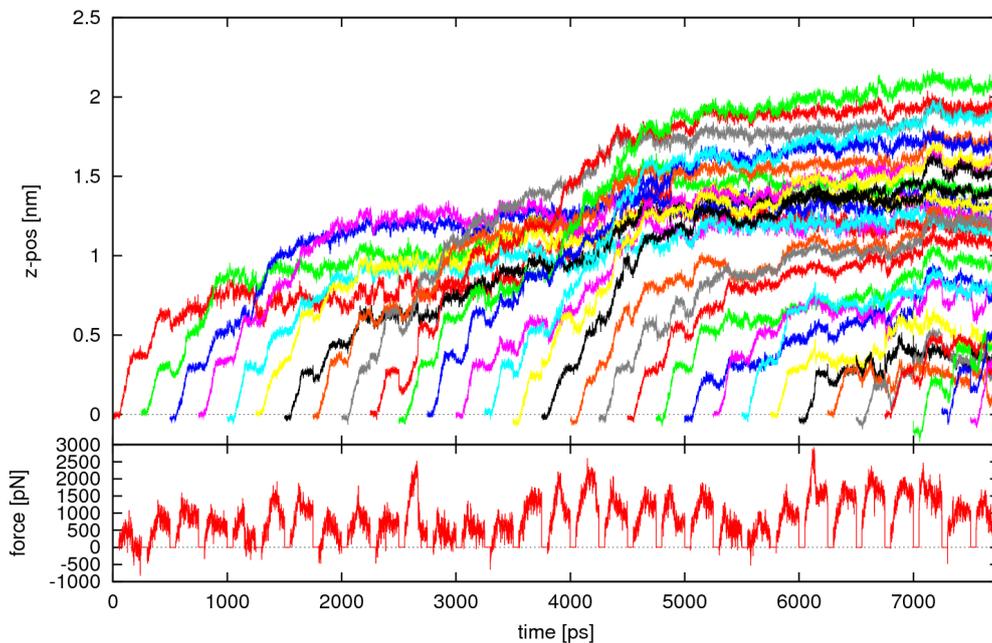


Figure 5.15: Movement of the growing polypeptide through the exit tunnel: Sequential adding and pushing of Bpp amino acids with spring constant $k = 8000 \frac{kJ}{mol \text{ nm}^2}$ and pushing velocity $v = 3.77 \frac{m}{s}$. Z-position of the polypeptide $C\alpha$ atoms and the forces acting on the amino acids.

with a spring constant of $k = 8000 \frac{kJ}{mol \text{ nm}^2}$ for the pushing potential. Figure 5.14 shows snapshots of the trajectory at different times.

In the first 2000 ps the growing polypeptide chain moved rather straight through the tunnel. Then the tip started to bend towards the tunnel wall and stayed there until $t = 6000$ ps. Meanwhile the polypeptide grew and formed a closely packed curved structure. The curvature and the close packing increased the forces on the tip and at ~ 4000 ps it moved forward again, thus finding its way in the direction of the tunnel.

Figure 5.15 shows the z-positions of the $C\alpha$ atoms of the growing Bpp

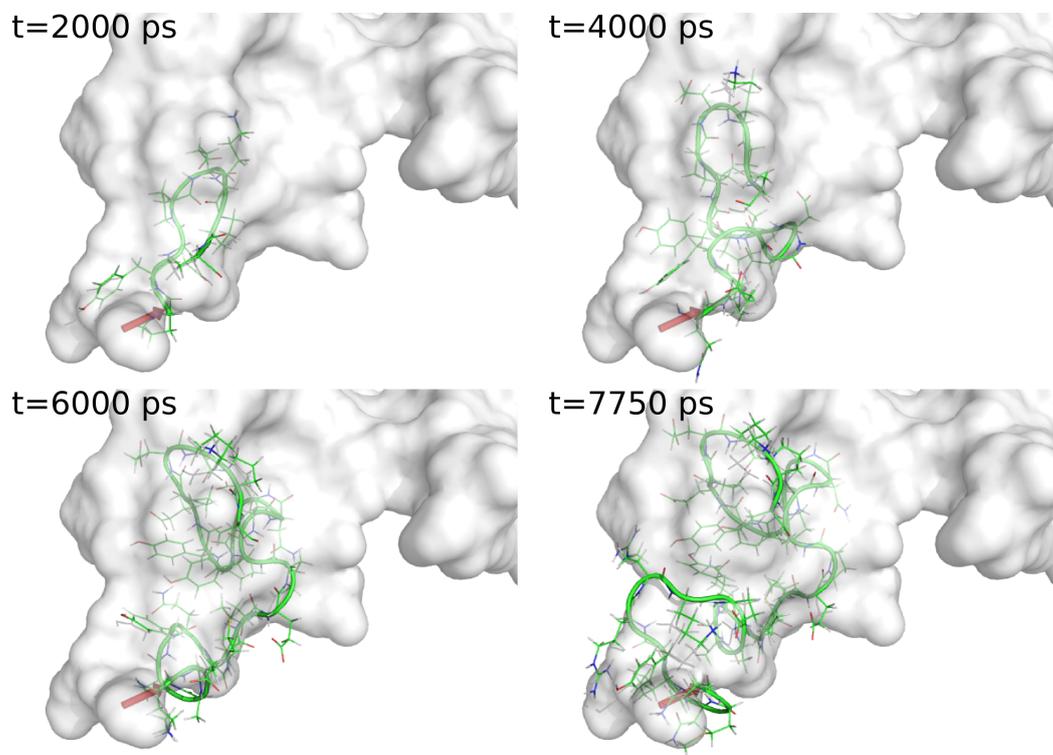


Figure 5.16: Simulation of nascent chain growth and movement through the exit tunnel: Sequential adding and pushing of Bpp amino acids with spring constant $k = 8000 \frac{kJ}{\text{mol nm}^2}$ and pushing velocity $v = 3.77 \frac{m}{s}$. Snapshots from the trajectory are shown, where the backbone of the polypeptide is drawn in green cartoon representation and the amino acids are drawn as lines. The tunnel surface which was calculated from the equilibrated ribosome (see 5.4.2) is drawn in grey. The red arrow starts at the position where the amino acids were added and where the pulling potential started. The tip of the arrow indicates the end position of the pulling potential.

polypeptide and the forces which were applied to the most recently added amino acid as a function of simulation time t . The spring constant for the harmonic pushing potential was $k = 8000 \frac{\text{kJ}}{\text{mol nm}^2}$ and the potential was moved with constant velocity $v = 3.77 \frac{\text{m}}{\text{s}}$. Figure 5.16 shows four snapshots of the trajectory. The simulation was not yet completed so the results are only shown for the first 31 amino acids, whereas the complete polypeptide consists of 36 amino acids.

In the first 1000 ps the tip of the Bpp polypeptide started to bend and formed a loop which was rather stable during the whole simulation. This loop then moved into the direction of the tunnel, while the polypeptide grew. The Bpp polypeptide in this simulation did not move as far into the tunnel as the SecM polypeptide did.

The forces applied to the amino acids by the pushing potential were integrated over the way z the amino acids moved during this time. For the integration of the force, we assumed a function formed by straight lines between successive data points and calculated the area between this function and the z -axis. Thus we derived an estimate of the mechanical work used to push the amino acids. In figure 5.17 this work is plotted for different simulations. The mechanical work increased with the length of the polypeptide, but the ratio of mechanical work and number of amino acids decreased. Besides the work needed during the two simulations we described, the work for two simulations with a slower pushing velocity $v = 1.885 \frac{\text{m}}{\text{s}}$ is shown. These simulations were still in the beginning, but as can be seen the mechanical work did not differ significantly from the other simulations.

Before we carried out the simulations with spring constant $k = 8000 \frac{\text{kJ}}{\text{mol nm}^2}$,

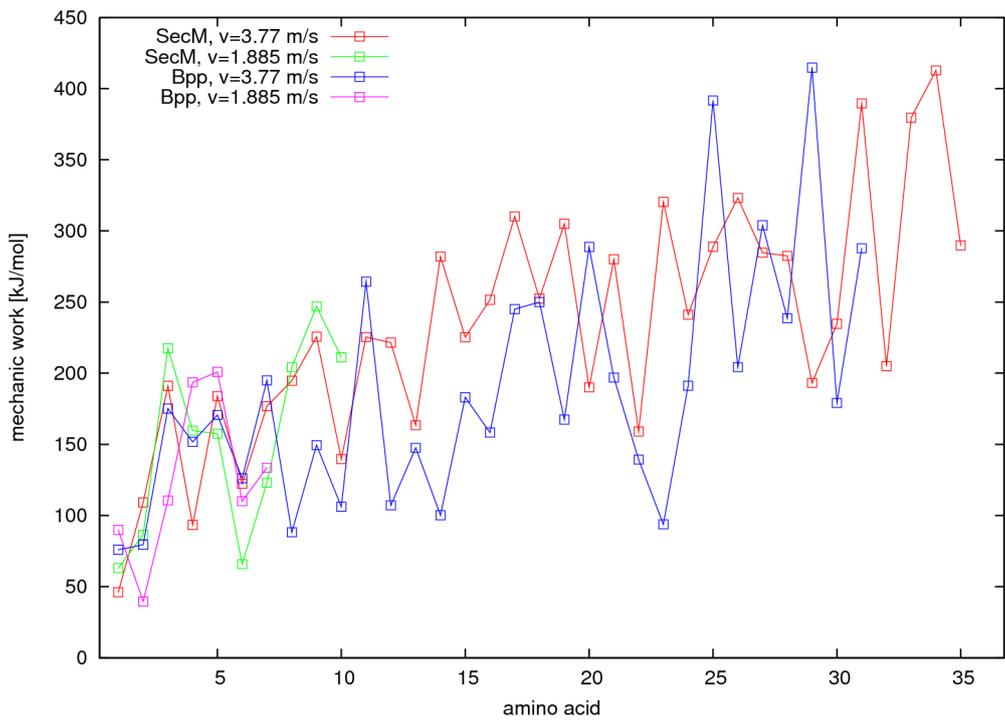


Figure 5.17: Comparison of the growing peptide chain simulation: Mechanical work applied to the amino acids in simulations with spring constant $k = 8000 \frac{\text{kJ}}{\text{mol nm}^2}$ for the pushing potential, different pushing velocities, and different polypeptides.

simulations with a spring constant of $k = 2000 \frac{kJ}{\text{mol nm}^2}$ were carried out. A lower spring constant results in a lower force, given the same distance of the pushed atoms to the minimum of the pushing potential. In our simulations the pushing potential moved only a small distance and was then kept at this position. Therefore, the maximum force which could be applied to the amino acid depended on the spring constant. In the simulations with the lower spring constant, the resulting force did not suffice to push the polypeptide chain after the polypeptide chain reached a certain length. Therefore, we used a spring constant of $k = 8000 \frac{kJ}{\text{mol nm}^2}$ which resulted in forces large enough to push the polypeptide into the tunnel.

Discussion

As we showed, it is possible to simulate the movement of a growing peptide chain in the ribosomal exit tunnel using the simulation scheme we developed. In our simulations the two growing polypeptides moved on a different pathway through the tunnel. This difference could be due to the different amino acid sequences or due to statistical influences, like positions of water molecules and ions, and small differences in the conformation of the ribosome. The analysis of more trajectories, comparing different trajectories of the same polypeptide and comparing different trajectories of different polypeptides, could allow to answer the question of sequence dependency.

As the pushing of the polypeptide was much too fast and the polypeptide could not fully equilibrate between the addition of new amino acids, simulations with a lower pushing velocity were started to estimate the influence of the velocity on the pathway and the mechanical work necessary to push

the polypeptide. For our simulations and for the first few amino acids there was no significant difference in the mechanical work for different velocities, so their influence might be small.

Chapter 6

Summary and Outlook

Proteins are synthesized by the ribosome at the peptidyl transferase center where the amino acids are covalently bound and form the growing polypeptide chain which leaves the ribosome via the exit tunnel. In this work molecular dynamics simulations of nascent peptide chains inside the ribosomal tunnel are presented. The polypeptide SecM induces a translation arrest while being synthesized by the ribosome after the addition of a certain amino acid. To study differences in conformations, energetics, and dynamics between arresting and non-arresting peptide chains, three different peptides were considered: polypeptide SecM166 as the SecM peptide at the arrest point, SecM164 as the SecM peptide synthesized up to two amino acids before the arrest point, and Bpp as a non-arresting sequence.

The modeled polypeptides inside the ribosomal tunnel were equilibrated. Our analysis of the trajectories, comparison of the conformations of the polypeptides with rmsd calculation and measurement of the interaction energies of the polypeptide amino acids with the ribosomal atoms, suggests that the converged conformations of the polypeptides differ significantly. Principal com-

ponent analysis of the trajectories was performed and we showed that the dynamics of the polypeptides in our simulations were different for different polypeptides but similar for different simulations with the same polypeptide, suggesting that the dynamics of fully equilibrated polypeptides depend on their amino acid sequence.

For each polypeptide four simulations with different starting structures were carried out to obtain a better sampling of the conformational space. Due to the large computational effort, the simulation times were restricted and the simulations did not completely converge.

To examine the pathway of the polypeptide through the tunnel and the mechanical work necessary to push the peptide through the tunnel, we developed a new simulation scheme, where the amino acids were sequentially linked to the growing polypeptide and then pushed into the direction of the tunnel using Force Probe Molecular Dynamics.

As the pushing potential only affected the most recently added amino acid, the rest of the growing peptide chain was free to "explore" the tunnel. The simulation, where the the amino acids were added according to the SecM166 sequence was finished. It was found that the polypeptide moved through the first part of the tunnel during the simulation. The mechanical work needed to push the polypeptide chain by the length of one amino acid was found to increase with the length of the nascent peptide chain, whereas the rate, mechanical work by length of polypeptide, was seen to decrease with increasing polypeptide length. The simulations of the growing peptide chain led to closely packed polypeptide conformations which differed markedly from the extended equilibrated conformations. The differences likely resulted from

the short simulation times which did not allow complete equilibration during each cycle of amino acid addition in the growing peptide simulations, or from the dependence on the extended modeled polypeptides in the equilibration simulations.

With the established simulation systems and the development of the simulation scheme for growing peptides, this work laid the basis for further studies of nascent polypeptides. For the equilibration of the polypeptides inside the tunnel, longer simulations are necessary to further support our conclusions. Also, analyses of the interactions of polypeptide amino acids and ribosomal residues and comparison with mutation data could increase the insight into the translation arrest mechanism.

Growing peptide chain simulations with lower pushing velocities would help to estimate the influence of the limited simulation time on the pathway of the polypeptide. Simulations with longer polypeptides would allow to investigate the passage through the whole tunnel.

Further into the future, our work will enable to study the mechanism of antibiotics and escape mutations. Erythromycin, e.g., is an antibiotic which binds within the tunnel and thereby hinders the growth of the polypeptide. A three residue deletion at the tip of the ribosomal protein L22, which builds a part of the tunnel wall, renders the ribosome resistant to the antibiotic. As crystal structures of the large ribosomal subunit with bound erythromycin and of the large ribosomal subunit of the mutated ribosome are available, it is possible to set up growing peptide chain simulations with erythromycin and with or without the mutation to investigate the mechanism of erythromycin and the mechanism of the resistance.

Thanks

I wish to thank Helmut Grubmüller for the idea of the exciting topic and for the great support during my work in the department, all my colleagues, especially Frank Wiederschein, for constant help, advice and a lot of fun. I would like to thank Martin Vesper for a lot of help and discussion on Distance Geometry, and Christoph Schmidt for being the main corrector of this thesis.

Bibliography

- [1] Berg, J. M., Tymoczko, J. L., Stryer, L., *Biochemistry* , (2006).
- [2] Ramakrishnan, V., “Ribosome Structure and the Mechanism of Translation” *Cell* **108**(4), (2002): 557–572.
- [3] Noller, H. F., Yusupov, M. M., Yusupova, G. Z., Baucom, A., Cate, J. H., “Translocation of tRNA during protein synthesis” *FEBS Lett.* **514**(1), (2002): 11–16.
- [4] Voss, N. R., Gerstein, M., Steitz, T. A., Moore, P. B., “The Geometry of the Ribosomal Polypeptide Exit Tunnel” *Journal of Molecular Biology* **360**, (2006): 893–906.
- [5] Woolhead, C., McCormick, P., Johnson, A., “Nascent Membrane and Secretory Proteins Differ in FRET-Detected Folding Far inside the Ribosome and in Their Exposure to Ribosomal Proteins” *Cell* **116**(5), (2004): 725–736.
- [6] Mori, H., Koreaki I., “The Sec protein-translocation pathway” *TRENDS in Microbiology* **9**(10), (2001): 494–500.
- [7] Hansen, J. L., Ippolito, J. A., Ban, N., Nissen, P., Moore, P. B., Steitz, T. A., “The Structures of Four Macrolide Antibiotics Bound to the Large Ribosomal Subunit” *Molecular Cell* **10**, (2002): 117–128.

- [8] Davydova, N., Streltsov, V., Wilce, M., Liljas, A., Garber, M., “L22 Ribosomal Protein and Effect of Its Mutation on Ribosome Resistance to Erythromycin” *Journal of Molecular Biology* **322**, (2002): 635–644.
- [9] Steitz, T. A., Moore, P. B. “RNA, the first macromolecular catalyst: the ribosome is a ribozyme” *TRENDS in Biochemical Sciences* **28**(8), (2003): 411–418.
- [10] Mitra, K., Frank, J. “Ribosome Dynamics: Insights from Atomic Structure Modeling into Cryo-Electron Microscopy Maps” *Annu. Rev. Biophys. Biomol. Struct.* **35**, (2006): 299–317.
- [11] Nissen, Poul, Hansen, Jeffrey, Ban, Nenad, Moore, Peter B., Steitz, Thomas A., “The Structural Basis of Ribosome Activity in Peptide Bond Synthesis” *Science* **289**, (2000): 920–930.
- [12] Driessen, A. J. M., Manting, E. H., van der Does, C., “The structural basis of protein targeting and translocation in bacteria” *Nature Structural Biology* **8**(6), (2001): 492–498.
- [13] Nakatogawa, H., Murakami, A., Koreaki, I., “Control of SecA and SecM translation by protein secretion” *Current Opinion in Microbiology* **7**, (2004): 145–150.
- [14] Nakatogawa, H., Koreaki, I., “The Ribosomal Exit Tunnel Functions as a Discriminating Gate” *Cell* **108**, (2002): 629–636.
- [15] Eisenberg, H., Mevarech, M., Zaccari, G., “Biochemical, Structural, and Molecular Genetic Aspects of Halophilism” *Advances in Protein Chemistry* **43**, (1992): 1–62.

- [16] Ban, N., Nissen, P., Hansen, J., Moore, P. B., Steitz, T. A., “The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution” *Science* **289**, (2000): 905–920.
- [17] van Gunsteren, Wilfred F., Berendsen, Herman J. C., “Computer Simulation of Molecular Dynamics: Methodology, Applications and Perspectives in Chemistry” *Angewandte Chemie International Edition* **29**, (1990): 992–1023.
- [18] Verlet, L., “Computer Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules” *Phys. Rev.* **159**, (1967): 98–103.
- [19] Debye, P., Hückel, E., “The theory of electrolytes. I. Lowering of freezing point and related phenomena” *Physikalische Zeitschrift* **24**, (1923): 185–206.
- [20] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., Haak, J. R., “Molecular dynamics with coupling to an external bath” *J. Chem. Phys.*, **81**(8), (1984): 3684–3690.
- [21] Finer, J. T., Simmons, R. M., Spudich, J. A., “Single myosin molecule mechanics: piconewton forces and nanometer steps” *Nature* **368**, (1994): 113–119.
- [22] Yin, H., Wang, M. D., Svoboda, K., Landick, R., Block, S. M., Gelles, J., “Transcription against an applied force” *Science* **270**, (1995): 1653–1657.
- [23] Grubmüller, H., Heymann, B., Tavan, P., “Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force” *Science* **271**, (1996): 997–999.
- [24] Klein, D. J., Moore, P. B., Steitz, T. A., “The Roles of the Large Ribosomal Proteins in the Structure, Assembly, and Evolution of the Large Ribosomal Subunit” *Journal of Molecular Biology* **340**, (2004): 141–177.

- [25] Tu, D., Blaha, G., Moore, P. B., Steitz, T. A. "Structures of MLSbK Antibiotics Bound to the Mutated Large Ribosomal Subunit Provide a Structural Explanation for Resistance" *Cell* **121**, (2005): 257–270.
- [26] Lindahl, E., Hess, B., van der Spoel, D., "GROMACS 3.0: a package for molecular simulation and trajectory analysis" *Journal of Molecular Modeling* **7**, (2001): 306–317.
- [27] Helm, Mark, "Post-transcriptional nucleotide modification and alternative folding of RNA" *Nucleic Acid Research* **34**(2), (2006): 721–733.
- [28] Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B. Woods, R. J., "The Amber biomolecular simulation programs" *J. Comput. Chem.* **26**, (2005): 1668–1688.
- [29] Aduri, R., Psciuk, B. T., Saro, P., Taniga, H., Schlegel, H. B., SantaLucia, J., Jr., "AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA" *J. Chem. Theory and Comput.* **3**, (2007): 1464–1475.
- [30] Sorin, E. J., Pande, V. S., "Exploring the Helix-Coil Transition via All-Atom Equilibrium Ensemble Simulations" *Biophysical Journal* **88**, (2005): 2472–2493.
- [31] Vriend, G., "WHAT IF: A molecular modeling and drug design program" *J. Mol. Graph* **8**(1), (1990): 52–56.
- [32] Darden, T., York, D., Pedersen, L., "Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems." *J. Chem. Phys.* **98**, (1993): 10089–10092.

- [33] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., Pedersen, L. G., “A smooth particle mesh ewald potential.” *J. Chem. Phys.* **103**, (1995): 8577–8592.
- [34] Fließbach, T., *Elektrodynamik, Lehrbuch zur Theoretischen Physik*, (2000): 277–278.
- [35] Havel, T. F., Kuntz, I. D., Crippen, G. M., “The theory and practise of distance geometry” *Bulletin of Mathematical Biology* **45**, (1983): 665–720.
- [36] Amadei, A., Linssen, A. B. M., Berendsen, H. J. C., “Essential Dynamics of Proteins” *PROTEINS: Structure, Function and Genetics* **17**, (1993): 412–425.
- [37] Schmeing, T. M., Seila A. C., Hansen, J. L., Freeborn, B., Soukup, J. K., Scaringe, S. A., Strobel, S. A., Moore, P. B., Steitz, T. A., “A pre-translocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits” *Nature Structural Biology* **9**(3), (2002): 225–230.
- [38] Wang, J. H., Miller, S., “Tracer Diffusion in Liquids. II. The Self-diffusion as Sodium Ion in Aqueous Sodium Chloride Solutions” *J. Chem. Soc.* **74**, (1952): 1611–1612.
- [39] Wang, J. H., “Tracer Diffusion in Liquids. III. The Self-diffusion of Chloride Ion in Aqueous Sodium Chloride Solutions” *J. Chem. Soc.* **74**, (1952): 1612–1615.
- [40] Mahoney, M. W., Jorgensen, W. L., “Diffusion constant of the TIP5P model of liquid water” *J. Chem. Phys.* **114**, (2001): 363–366.
- [41] Eisenberg, D., Kauzmann, W., “The structure and properties of water” *Oxford University Press, London* (1969)

- [42] Schlitter, J., "Estimation of absolute and relative entropies of macromolecules using the covariance matrix" *Chemical Physics Letters* **215**(6), (1993): 617–621.