Effect of posttranscriptional modifications and Mg^{2+} ions on tRNA structure and flexibility

Diplomarbeit

vorgelegt von

Christian Blau

aus

Merseburg

angefertigt am

Max Planck Institut für biophysikalische Chemie Göttingen

in der

Abteilung für theoretische und computergestützte Biophysik

31.08.2009

Contents

Part I

Introduction

A basic goal in life sciences is to understand where mechanisms in the cell, like gene replication or mutation, originate from or why certain molecules perform their function in the way they do. The tremendous complexity of life forms makes it necessary to explain mechanisms in the cell from less complex forms. A tempting Gedankenexperiment to reduce the complexity of analysis of mechanisms in the cell is to look for a system of small chemical compounds that intrinsicly shows some essentials of life. One of those molecules in question is ribonucleic acid (RNA), a macromolecule closely related to DNA. It can form a great many of both, rigid and flexible structures.

One of the properties that put RNA in the role of being a major player in early life development is its double capability of both storing information and performing function. The efficiency by which RNA performs some of its functions under different environmental conditions can hardly be explained from the few canonical building parts of that molecule alone. An extension to that set of building blocks are nucleobase modifications and ions binding at specific sites. The common bracket of this work is to examine the effect of modifications ions and temperature on RNA structure and flexibility.

The method employed in this thesis to examine those effects is molecular dynamics (MD) simulations. Configurational space is sampled by means of numerical integration of the equations of motions of molecule and solvent. From that sampling, general biophysical properties can be derived. Additionally MD offers atomistic descriptions of biomolecular processes.

RNA primary structure must be viewed as coding for a plethora of conformers in tertiary structure depending on the values regulating parameters take [1]. In this respect an understanding of the effect of temperature change on RNA structure and dynamics can help understanding RNA thermosensors and thus gene regulation and expression on the one side and the great stability some RNA structures can maintain even through a wide temperature range on the other side.

The type of RNA that is most affected by posttransscriptional modifications is transfer RNA or tRNA [2]. It was chosen as subject of MD simulations because of its pivoting role in the process of translating genetic code into proteins, its well known structure and modest size. Nucleotide modification can strongly affect the ability of tRNA to perform its biological function which may lead to death of its hosting organism.

Nucleotide modifications are conserved between cytoplasmic tRNAs of different species even though their sequences are not [3]. This speaks for a important function of these modifications even under different environmental conditions. Yet the actual function of many of these modifications and the specific way of how they work is not fully understood even though some of them are essential for life [4].

Another factor that influences tRNA stability are magnesium ions. It is known that magnesium helps tRNA folding and maintaining tertiary structure and is thereby essential for its biological function [5]. What is unknown so far is how magnesium ions affect the structure of tRNA on an atomistic level. Their effect on the overall structural stability of tRNA will be investigated along with the mutual impact of magnesium ion binding to RNA and nucleotide modifications.

By simulation of tRNA in different setups ex- or including magnesium ions and nucleotide modifications insight into atomistic effects of these as well as their impact on general dynamics of tRNA will be gained.

1 RNA in the cell

Biological implications of nucleotide modifications, ions and temperature on transfer RNA can be understood only in terms of the processes and molecules transfer RNA interacts with. The following chapter will provide some biological background of these molecules and processes.

1.1 Central Dogma of Molecular Biology

To classify RNA in the process of life, the general concept of information flow in the cell shown in Fig. 1.1 is useful. Watson and Crick published that concept known as central dogma of molecular biology in 1958 [6, 7]. It states that information flow always goes forward from nucleic acid to protein. Multiple tasks sketched in Fig. 1.2 are performed by different kinds of RNA listed in Table 1.1.

Table 1.1: Examples of RNA and their biological role, showing its involvement in multiple vital processes.

Abbreviation	Name	Process associated
mRNA	messenger RNA	information storage; translation
tRNA	transfer RNA	translation
rRNA	ribosomal RNA	translation [8]
pre-mRNA	precursor mRNA	information storage; processing
snRNA	small nuclear RNA	processing
siRNA	small interfering RNA	gene regulation
aRNA	antisense RNA	gene regulation
miRNA	micro RNA	gene regulation

Figure 1.1: Central Dogma of Molecular Biology as stated by Crick [7]. Transfers that not depicted by arrows are unlikely to be discovered. The central role of RNA in molecular biology can be immediately understood from its position in this figure. Solid arrows show general transfers, whereas dotted refer to special ones.

Figure 1.2: The central dogma of biology helps to classify the processes RNA is involved in (black and colored arrows). Gray arrows show other transfers not involving RNA.

1.2 Translation

Ribosomes are molecular machines that read genetic code transmitted to them by messenger RNA (mRNA) and produce proteins according to that code. This process is called translation. Amino acids needed to build proteins are delivered to them by transfer RNA (tRNA). For different amino acids different kinds of tRNA are used.

The ribosome comprises three different binding sites for tRNA; the accommodation site (A), the peptide site (P) and the exit site (E). See figure 1.3 for a schematic overview of the ribosome and the elongation cycle in which new amino acids are incorporated into the nascent peptide chain (NPC).

The elongation cycle starts after initiation of the ribosome. One tRNA is present then at the P-site. An additional tRNA charged with an amino acid reaches the ribosome and is checked for matching the codon on the mRNA at the A-site. If that is the case, tRNA is incorporated into the ribosome with the help of the elongation factor TU (EF-TU). After the peptide transfer reaction took place the nascent peptide chain is now bound to the A-site tRNA with an additional amino acid incorporated. Elongation factor G then helps tRNA to move from the A to the P-site and from the P to the E-site. The E-site tRNA can now leave the ribosome. From that point the cycle starts again.

New amino acids will be incorporated into the nascent peptide chain until a stop codon is read, at which point translation termination is initiated. Finally, the peptide chain leaves the ribosome and the two subunits of the ribosome dissociate.

1.3 Aminoacyl-tRNA synthease

The tRNAs that leave the ribosome need to be aminoacylated again. Aminoacyl-tRNA synthetases (aaRS) fulfill this task. These enzymes catalyze aminoacylation of tRNA [9] by the use of adenosine-triphosphate (ATP) hydrolysis to adenosine-monophosphate (AMP).

For each amino acid exists a specific aaRS which charges the cognate tRNA with high fidelity rates of one error per 10,000.

Figure 1.3: Elongation cycle of a procaryotic ribosome.

2 RNA biochemistry

To understand how RNA performs its biological tasks, a look at its structure at all levels from atomistic to tertiary structure is inevitable. The following chapter will give an introduction to RNA and define the terms used in this thesis.

2.1 Chemical components of RNA

RNA is a macromolecule consisting of a few building blocks in micro RNA up to many thousands in mRNA. Those blocks are called nucleotides.

Figure 2.1 shows that nucleotides in RNA consist of a nucleobase (blue, denoted by R), a ribose sugar (green) and a phosphate (red). Each sugar is connected via the 3' and 5' position to a phosphate oxygen, thus forming the backbone of RNA. While the phosphate and the sugar group do not change, different nucleobases determining RNA structure and function are attached at the 1' end of the sugar.

Nucleobases are purine or pyrimidine derivatives. These heterocyclic organic compounds are shown in Fig. 2.2. Their basic character is due to the lone electron pair at one of the nitrogen atoms.

There are four canonical nucleobases in RNA. These are adenine, guanine, cytosine and uracil. Three subsequent canonical nucleobases on an RNA strand can form a codon when they are involved in the information flow in the cell as depicted in Fig. 1.1.

Nucleosides arise from the linkage of the 1'-carbon atom of the ribose to the 9-nitrogen of the purine nucleobase (1-nitrogen for pyrimidines respectively) shown in figure 2.1. If a phosphate group binds a nucleoside via an ester linkage at the 3' or 5', the emerging compound is called nucleotide.

Figure 2.1: Primary structure of RNA. Phosphate group is shown in red, the sugar group in green and the blue R denotes the nucleobase.

Figure 2.2: Purine (light blue) and pyrimidine (dark blue) are the bases all nucleobases are derived from. Adenine, guanine, uracil and cytosine on the right are the canonical ones found in RNA. Zigzag lines indicate binding sites to the sugar.

Figure 2.3: Canonical Watson-Crick base pairing between cytosine and guanine.

RNA-strands are established when nucleotides join each other by an ester bond at the 5' or 3' end, respectively. Thereby they form a nucleotide chain with a 3' and a 5' end.

2.2 Secondary structure

Analogous to proteins there are multiple intramolecular interactions within an RNA chain. These interactions stabilize the secondary structure of RNA. The main contribution to intramolecular interactions are hydrogen bond interactions.

2.2.1 Base pairing

By their very shape and position of hydrogen bond donors and acceptors, nucleobases can form pairs with very high hydrogen bonding energies. An cytosine-guanine base pair is shown in Fig. 2.3 as an example; adenosine and uridine base pair in a similar manner.

Non-canonical base pairing

Alternative interactions between hydrogen bond donors and acceptors are possible [10, 11]. Those interactions are named according to the nucleoside edges involved as shown in Fig. 2.4. For these interactions symbols are introduced in secondary structure sketches according to Fig. 2.5. C-H edge and Hoogsteen edge will be used synonymously further on.

2.3 Tertiary structure

2.3.1 Base stacking

Base stacking occurs between bases that are arranged one upon another as shown in Fig. 2.6. The stacking interaction is caused by overlapping p-orbitals. Base stacking stabilizes the tertiary structure of RNA.

Figure 2.4: Edges involved in non-canonical base pairing for pyrimidines(left) and purines(right).

Figure 2.5: Symbols used to indicate base pairing. Cis conformations of the nucleobase-sugar bonds are shown with filled symbols, trans with empty. Different edges interacting with each other are shown as a combination of symbols.

Figure 2.6: Base stacking interaction indicated in red.

Part II

Theoretical concepts and methods

3 Molecular dynamics simulations

This thesis is aiming at deriving biophysical properties of RNA on an atomistic level. These thermodynamical properties can be derived from its phase space density. Therefore, a methods for sampling phase space is desirable. Methods aiming at that goal are Monte Carlo (MC) [13] or molecular dynamics (MD) [14] simulations. An outline of the approximations made that lead to the molecular dynamics simulations method this thesis relies on will follow. We will use MD simulations in this work which will be subsequently derived from basic quantum mechanics, the Schrödinger equation describing the movement of nuclei and electrons,

$$H|\psi\rangle = i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\psi\rangle\,. \tag{3.1}$$

3.1 Physical approximations

The time dependend Schrödinger equation for more than two particles can in general not be solved analytically. A numerical approach will fail due to limited computing powers. Fig. 3.1 shows where a computational approach is amenable. To be able to use numerical methods, three approximations of the Schrödinger equation are applied, resulting in an integration of Newtonian equations of motion of particles in a force field.

3.1.1 Born-Oppenheimer Approximation

Born and Oppenheimer developed an approximation for the solution of the Schrödinger equation of a system that has slow moving heavy (nuclei) and fast moving light parts (electrons). For a given set of nuclei (charges Z_i, masses M and positions R_i) and electrons (charges e,

Figure 3.1: Classification of physical systems. A computer simulation approach is amenable for the shaded fields. Adapted from van Gunsteren [12].

masses m_e and positions r_i) the Hamiltonian H reads:

$$\begin{aligned} H &= T_{\mathrm{n}} + T_{\mathrm{e}} + V_{\mathrm{n,e}} + V_{\mathrm{n,n}} + V_{\mathrm{e,e}} \\ &:= T_{\mathrm{n}} + H_{\mathrm{e}} \\ &= -\sum_{\mathrm{n}} \frac{\hbar^2}{2M_n}\nabla_n^2 - \sum_{\mathrm{e}} \frac{\hbar^2}{2m_e}\nabla_e^2 + \frac{1}{2}\sum_{\mathrm{e}}\sum_{\mathrm{e}} \frac{e^2}{|r_i - r_k|} \\ &\quad + \frac{1}{2}\sum_{\mathrm{n}}\sum_{\mathrm{n}} \frac{Z_i Z_j e^2}{|R_i - R_k|} + \sum_{\mathrm{n}}\sum_{\mathrm{e}} \frac{Z_i e^2}{|r_i - R_k|}\,. \end{aligned} \tag{3.2}$$

The Born-Oppenheimer Approximation assumes that the heavy nuclei and the light electrons states can be regarded uncoupled. The approximated Schrödinger equation for the wave function of the nuclei $|\psi_n\rangle$ is

$$\left(T_{\mathrm{n}} + V\left(\mathbf{r}\right)\right)|\psi_n\rangle = i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\psi_n\rangle\,. \tag{3.3}$$

$V\left(\mathbf{r}\right)$ is further approximated by a semi-empirical potential which is called force field.

3.1.2 A force field as approximation for the potential $V\left(\mathbf{r}\right)$

Force fields approximate the potential $V\left(\mathbf{r}\right)$ as a sum of interaction functions. These interaction terms are listed in Table 3.1.

3.1.3 Newtonian equations of motion replace the Schrödinger equation

Because it is too costly computationally to solve eq. 3.3, the Schrödinger equation of motion of the nuclei is replaced by the Newtonian equations of motion. The correspondence principle

Table 3.1: Empirical terms that approximate the electron potential. The K_i denote the force constants, r_{ij} the distance between atoms i and j, P_n the periodicity for the dihedral potential, γ its phase angle and $A(i,j)$, $B(i,j)$ the van der Waals interaction parameters.

Interaction	Force Field Term	Graphical Representation
Bonded		
bond stretching	$\sum_{\text{bonds}} \frac{1}{2} K_b \left(r_{ij} - r_0\right)^2$	
bond angle vibrations	$\sum_{\text{angles}} \frac{1}{2} K_\theta \left(\theta - \theta_0\right)^2$	
extraplanar motions	$\sum_{\text{imp}} \frac{1}{2} K_\xi \left(\xi - \xi_0\right)^2$	
dihedral torsions	$\sum_{\text{dih}} \frac{K_\phi}{2}(1 + \cos(P_n\phi - \gamma))$	
Non-bonded		
van-der-Waals	$\sum_{\text{i,j}} -\frac{A(i,j)}{r_{ij}^6} + \frac{B(i,j)}{r_{ij}^{12}}$	
Coulomb	$\sum_{\text{i,j}} \frac{q_i q_j}{4\pi\epsilon_r\epsilon_0 r_{ij}}$	

gives the classical expression

$$m\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{v} = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathbf{r} = -\nabla V(\mathbf{r}) = \mathbf{F}(\mathbf{r}) \; . \tag{3.4}$$

By that the system propagates from a set of starting coordinates to sample phase space.

3.1.4 Numerical integration of the equations of motion

Still, equation 3.4 usually cannot be solved analytically for more than three atoms. Therefore this equation is integrated numerically. The leap-frog algorithm is used in this work. It follows the following integration scheme:

$$\begin{aligned} \mathbf{v}(t + \Delta t/2) &= \mathbf{v}(t - \Delta t/2) + \frac{\Delta t}{m}\mathbf{F}(\mathbf{r}(t)) \\ \mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \Delta t\mathbf{v}(t + \Delta t/2) \, . \end{aligned}$$

For reasons of numerical stability the integration time step `Δt` is chosen an order of magnitude faster then the time-scale of the fastest motions in our simulations. The fasted motions are the bond-stretching motions of hydrogen atoms. They are constrained with the LINCS algorithm and are not taken into account. The bond-angle vibrations are timescale of 20 fs. For that reason the integration time step is `Δt` $= 20/10 = 2$ fs.

3.1.5 Finite system sizes - periodic boundary conditions

System sizes have to be reduced to feasible particle numbers, which are orders of magnitude smaller than the number of particles in a cell. By the reduction of system sizes surface effects become more and more important. The disadvantage of surface artifacts caused by walls can be ruled out by periodic boundary conditions. A unit cell (called box) spanned by the vectors $\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3$ is defined. All simulated particles reside in this box. Fig. 3.2 shows the basic concept of periodic boundary conditions, any time a particle would leave the box it will be shifted back into the box by a linear combination of $\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3$. All particles interact with their periodic images. These periodic images are copies of particles in the unit cell and occupy all possible positions with respect to shifts along all linear combinations of the box vectors.

3.1.6 Finite and infinite interaction ranges - PME electrostatics and Lennard Jones cut-offs

For long ranged interaction terms theoretically all particles and their periodic images would have to be accounted for. From that an infinite number of interaction terms arises. For potentials that fall off quickly a cut-off defines the largest distance at which particles interact in simulations. Fig. 3.3 shows that potentials that fall off slower or equal with the power of one the error for a cut-off method cannot be estimated as the harmonic sum diverges. Approximations of long range electrostatic interactions were found to greatly affect simulation results of nucleic acids and proteins [15, 16]. For that reason long range electrostatic interactions have to be dealt with good accuracy while not consuming too much computational time. For this purpose Ewald summation is used in this work. Interaction terms above a certain cut-off are approximated in reciprocal space. By that long-range interactions can be taken into account while the algorithm still runs with $\mathcal{O}$(number of particles) still [17].

Figure 3.2: Periodic boundary conditions: A particle (red) in the unit cell (black) and its periodic images (light red). When moving into one of the grey cells it is shifted back to the black cell.

Figure 3.3: For periodic conditions in one dimension the coulomb interaction between a charged particle and its periodic images is a harmonic sum and thus diverges. No matter how many periodic images are taken into account, the interactions left out are still infinite.

3.2 Set of starting coordinates

To start a simulation a set of coordinates must be chosen that describes a representative state of the molecule of interest. In this thesis they are chosen from X-ray crystallography data which can be found in the PBD (Protein Data Bank) or NDB (Nucleic Acid Database). Additional data is available from nucleic magnetic resonance (NMR) experiments [18, 19].

3.3 Constant temperature and pressure simulations

Constant pressure and temperature are the canonical conditions in which biophysics experiments take place. To come close to these conditions, our simulations employ algorithms that ensure sampling of an ensemble that is isobaric and isothermal. Another constant is the number of particles as no chemical reactions are accounted for in our model. With these three constants all simulations generate an isobaric-isothermal ensemble (N,p,T).

3.3.1 Pressure coupling

The pressure in our simulation systems is kept constant by the two different methods for equilibration and unconstraint MD simulation. The Berendsen pressure coupling [20] rescales box vectors and coordinates every step in the MD simulation. The Parinello-Rahman pressure coupling [21] rescales the box vectors and changes the equations of motion. In contrast to Berendsen the Parinello-Rahman barostat gives the right thermodynamical ensemble. For situations far from equilibrium there are large box size fluctuations which might cause the simulation to crash when a Parinello-Rahman barostat is employed. For that reason the Berendsen barostat which is more stable in this respect is used for the equilibration of the system whereas Parinello-Rahman is used for free MD simulation.

3.3.2 Temperature coupling

Analogously to pressure coupling there exists a Berendsen thermostat and a Nosé-Hoover thermostat [22]. The Berendsen thermostat scales the velocities of each particle. Nosé-Hoover introduces a friction term into the equations of motion with a friction parameter that evolves with time in dependence of the difference of the actual to the target temperature. Even though the actual pressure and temperature fluctuate during the simulations the Nosé-Hoover thermostat ensure a correct thermodynamical ensemble for given temperature and pressure while the Berendsen thermostat does not exactly but does not crash at larger temperature deviations. For that reason we employ Berendsen temperature coupling for equilibration while the Nosé-Hoover is used for the free MD run.

3.3.3 The isobaric-isothermal ensemble

For a system with N particles coupled to a temperature bath and held at constant pressure we find the probability of occupying a microstate to be

$$\rho\left(x_1, \ldots, x_{3N}, p_1, \ldots, p_{3N}, V\right) = \frac{\mathrm{e}^{-\beta E(x_1,\ldots,x_{3N},p_1,\ldots,p_{3N})-pV}}{\Xi} . \tag{3.5}$$

Where Ξ is the partition function for a isobaric-isothermal ensemble. All thermodynamic properties can be derived from that partition function. Of particular interest is the free energy of a system as free energy differences can be measured experimentally.

$$G = -k_B T \ln \Xi \,. \tag{3.6}$$

A molecular dynamics simulation gives a point cloud in phase space whose density approximates the phase space density of a real system. As kinetic energy terms can be assumed to cancel out and volume deviations are small analyses usually reside to conformational space. As the solvent dynamics happen on a much faster time scale than of the macromolecule, the conformational space analysis can usually be restricted to the macromolecule of interest.

4

A representation for arbitrary triclinic simulation boxes

4.1 Introduction

A constructive method for displaying the Wigner Seitz cell based on arbitrary triclinic lattice vectors was developed in course of this thesis. This visualization method helps to judge wether the choice of a simulation box is appropriate.

Computational time is proportional to the number of particles that are simulated. In our simulations the solvent surrounding the protein is explicitely taken into account. Simulation time will be reduced considerably if the number of solvent molecules is reduced. For simulations carried out under periodic boundary conditions, one way to reduce the amount of solvent is the use of non-cubic simulation boxes by the introduction of arbitrary triclinic lattice vectors.

The number of solvent molecules cannot be reduced arbitrarily. To avoid artefacts that arise from solutes that interact with its periodic images the closest distance of a molecule to its periodic image should not be smaller than a minimum threshold. The volume around a point, that is closer to that point than to any other point is defined as Voronoi cell [23] in general and as Wigner Seitz cell (WSC) for a lattice in particular. This WSC helps to visualize the distance of one molecule to its closest periodic images. Such a WSC in two dimensions and its construction are shown in Fig. 4.1.

4.1.1 A method for finding the vertices of the WSC

Without loss of generality let the point the WSC shall be constructed around be the origin of the coordinate system. A set of basis vectors $\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}$ defines the lattice and a metric tensor

$$G = (\mathbf{e_i}\mathbf{e_j})_{3\times 3}. \tag{4.1}$$

Figure 4.1: The construction of a WSC in two dimensions by perpendicular bisectors shown in different colours. On the left side two perpendicular bisectors intersect at point $\mathbf{r}$ which is not in the WSC because $\mathbf{l_1} + \mathbf{l_2}$ is closer to it than the origin. The WSC is the white polygon around the mid point on the right side.

This metric tensor defines the inner product for vectors with lattice coordinates $\mathbf{a}$ and $\mathbf{b}$: $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^T G \mathbf{b}$.

A point $\mathbf{r}$ is in the WSC if and only if for all lattice points $\mathbf{l_i}$ that point is closer to the origin than to another lattice point,

$$\begin{aligned} \langle \mathbf{r}, \mathbf{r} \rangle &< \langle \mathbf{l_i} - \mathbf{r}, \mathbf{l_i} - \mathbf{r} \rangle \\ \langle \mathbf{l_i}, \mathbf{r} \rangle &< \frac{1}{2} \langle \mathbf{l_i}, \mathbf{l_i} \rangle . \end{aligned} \tag{4.2}$$

Thus $\langle \mathbf{l_i}, \mathbf{r} \rangle = \frac{1}{2} \langle \mathbf{l_i}, \mathbf{l_i} \rangle$ defines an equation for planes that margin the WSC. The WSC is therefore a polyhedron.

The vertices of the polyhedron that defines the WSC are intersections of three planes margining the WSC. The intersection points of all combinations of three perpendicular bisectors are calculated. The intersection point is then checked for being inside the WSC.

For an intersection point $\mathbf{r}$ the following set of equations must hold

$$\begin{aligned} \mathbf{l_1}^{\mathrm{T}} G \mathbf{r} &= \frac{1}{2} \mathbf{l_1}^T G \mathbf{l_1} \\ \mathbf{l_2}^T G \mathbf{r} &= \frac{1}{2} \mathbf{l_2}^T G \mathbf{l_2} \\ \mathbf{l_3}^T G \mathbf{r} &= \frac{1}{2} \mathbf{l_3}^T G \mathbf{l_3} \ . \end{aligned} \tag{4.3}$$

Here $\mathbf{l_1}$, $\mathbf{l_2}$ and $\mathbf{l_3}$ denote the lattice points whose perpendicular bisectors to the origin are intersected.

With $\mathbf{l_i} = (l_{i1}, l_{i2}, l_{i3})^T$, $L = (l_{ij})_{3\times 3}$, linearly independent $\mathbf{l_i}$ and $\lambda = \mathrm{Diag}(LGL^T)$ this can

Figure 4.2: A reduced basis $\mathbf{g_1}', \mathbf{g_2}$ for the original basis $\mathbf{g_1}, \mathbf{g_2}$; both describing the same lattice.

be written in matrix form as

$$LG\mathbf{r} = \frac{1}{2}\lambda \tag{4.4}$$

$$\mathbf{r} = \frac{1}{2}(LG)^{-1}\lambda\ . \tag{4.5}$$

4.1.2 Lattice reduction

On the one hand condition 4.3 had to be checked for all lattice points. On the other hand only neighboring points contribute to the construction of the WSC. Lots of computational effort can be saved when only checking for these. This can be archived by lattice reduction. In a reduced lattice the neighbors of the origin are points with indices (i, j, k); $i, j, k \in \{-1, 0, 1\}$. A lattice reduction in two dimensions is shown in Fig. 4.2. An algorithm for creating such a basis efficiently in three dimensions has been found by Semaev [24].

4.1.3 Projection back to canonical coordinates

The resulting vector $\mathbf{r}$ points to the position of the vertices of the WSC in lattice coordinates. To project it back to cartesian coordinates for visualization it is necessary to reevaluate its canonical cartesian coordinates. This can be done by choosing a basis of the form

$$\mathbf{e_1} = (e_{11}, 0, 0); \mathbf{e_2} = (e_{21}, e_{22}, 0); \mathbf{e_3} = (e_{31}, e_{32}, e_{33})\ . \tag{4.6}$$

It follows

$$e_{11} = \sqrt{G_{11}} \tag{4.7}$$

$$e_{21} = G_{12}/e_{11} \tag{4.8}$$

$$e_{22} = \sqrt{G_{22} - e_{21}^2} \tag{4.9}$$

$$e_{31} = G_{13}/e_{11} \tag{4.10}$$

$$e_{32} = (G_{23} - e_{21}e_{31})/e_{22} \tag{4.11}$$

$$e_{33} = \sqrt{G_{33} - e_{31}^2 - e_{32}^2}. \tag{4.12}$$

4.2 Implementation

A c++ routine reads in the box data given in a pdb file and writes a sequence of PYMOL directives containing vertices and edges [25]. These are used in PYMOL to visualize the Wigner Seitz cell. The algorithm follows the geometrical construction. It intersects the perpendicular bisectors of three lattice points close to the origin. Then it checks wether this point is indeed in the Wigner-Seitz cell. If this is the case it is stored. After finding all vertices the edges are calculated by checking if two of the three planes that give a vertex intersection points are the same. The basic flow chart for the calculation of vertices and edges is given in Fig. 4.3.

4.3 Results

Figures 4.4-4.6 show representations of simulation systems in non-cubic boxes. For the tRNA system which can rotate freely a box which is as spherical as possible—the rhombic dodecahedron—is suitable, visualized in Fig. 4.4. For membrane pore system which is confined in orientation a hexagonal prism was chosen which is visualized in Fig. 4.5. An arbitrarily chosen cell is visualized in Fig. 4.6.

The algorithm depicted in 4.3 takes on average 300 ms on a AMD Athlon 64 X2 4400+ using one processor (2200 GHz). Excluding the time to read in box vectors from a pdb file and the time to write out the graphics directives the core routine takes 170 ms.

4.4 Discussion

The time which is needed for evaluation of the WSC is small enough for a visualization for a single frame in PYMOL. For non canonical triclinic box vectors the WSC has been correctly determined. For animation purposes of changing WSC cells due to changing box vectors a further optimization will be needed that reduces the evaluation time below the time resolution of the eye which would be less than 50ms.

There are two possible lines along improvement can take place. One might employ an algorithm that uses the Delaunay triangulation [26]. From that an evaluation of Voronoi regions is easily possible [27]. Another option is to reduce the number of neighbors used in the construction by analyzing the metric tensor properties first.

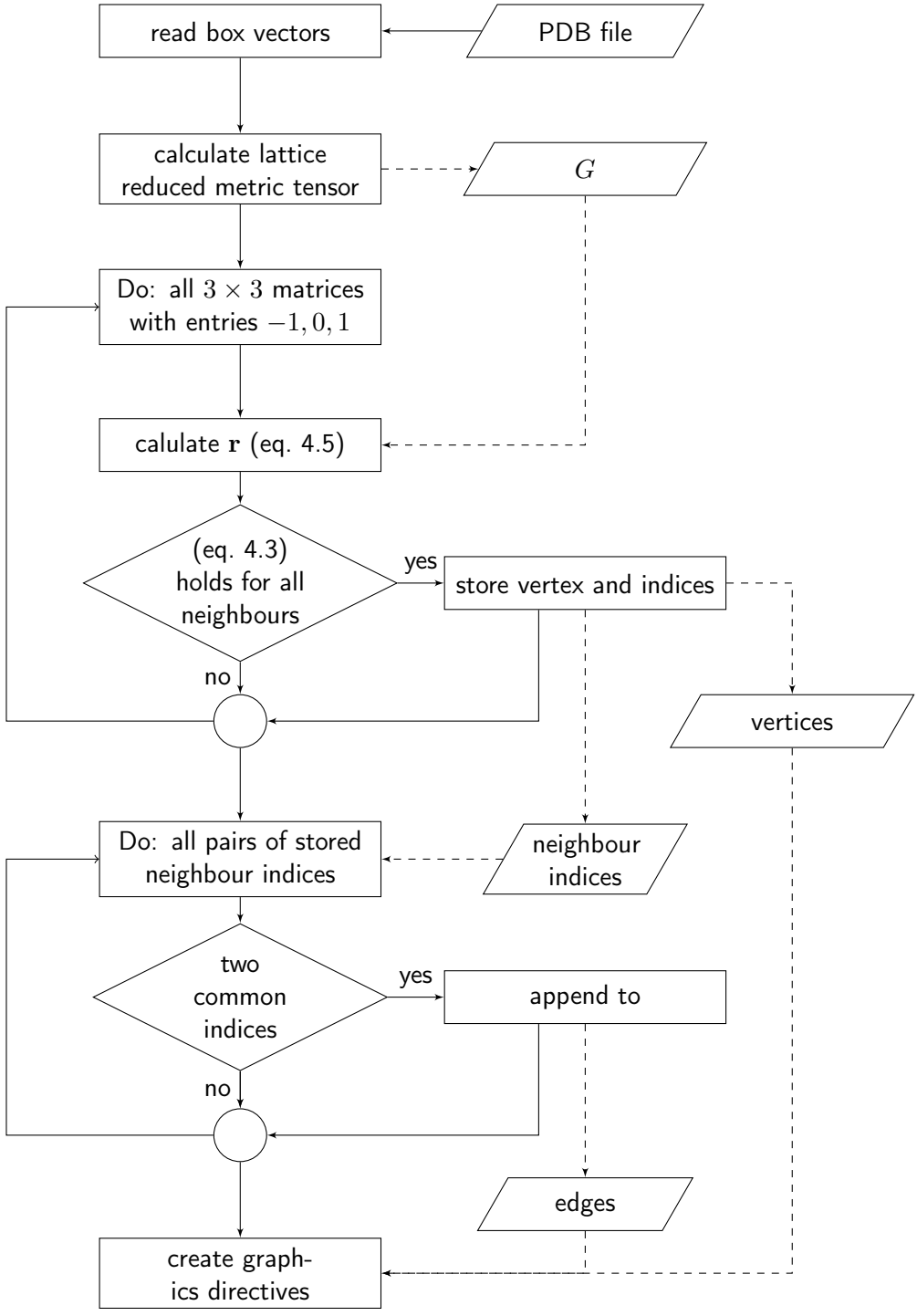

Figure 4.3: Flow chart of the algorithm used to determine vertices and edges of the polyhedron delimiting the WSC.

Figure 4.4: tRNA in solution setup in a dodecahedron box. Visualized with PYMOL using the here developed WSC algorithm on the left and the standard triclinic box already available in green and in blue the rectangular box internally used in GROMACS.

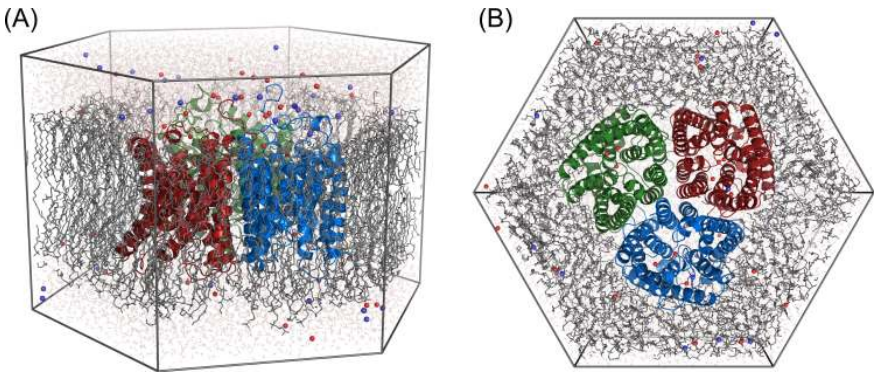

Figure 4.5: Simulation system of a membrane channel in solution in a hexagonal prism box visualized with PYMOL using the here developed WSC algorithm. See two orientations of the same system.

Figure 4.6: Different orientations of a WSC representation of an arbitrary triclinic basis (Box vector length and angles are: $a = 68\text{Å}, b = 51\text{Å}, c = 81\text{Å}, \alpha = 100°, \beta = 65°, \gamma = 80°$).

5 Analysis methods

5.1 Root mean square deviation

The root mean square deviation (RMSD) is a measure for the similarity of two structures in terms of the coordinates of the single atoms. It is defined as the quadratic mean of the deviation of the pairwisely compared coordinates, thus for two sets of N coordinates $\mathbf{x}_{i_1}$, $\mathbf{x}_{i_2}$ at times t_1 and t_2 expanding to

$$\mathtt{RMSD}_{12} = \sqrt{\frac{\sum_{i=1}^{N} (\mathbf{x}_{i_1} - \mathbf{x}_{i_2})^2}{N}}\,. \tag{5.1}$$

By its definition it is a distance measure in conformational space. It is used during this thesis to estimate equilibration times and flexibility of a structure.

As global rotational and translational movements of the system as a whole are not accounted for in RMSD analysis by fitting structures to a reference before calculating the RMSD.

5.2 Root mean square fluctuation

The root mean square fluctuation (RMSF) is a measure for the standard deviation from the mean position of the coordinates of a certain atom over a time $T - T_0$. It is defined as the quadratic mean of the deviation of the coordinates $\mathbf{x}$ of an atom from its mean position $\langle \mathbf{x} \rangle$. It is expressed as

$$\mathtt{RMSF} = \sqrt{\frac{\sum_{t=T_0}^{T} (\mathbf{x}(t) - \langle \mathbf{x} \rangle)^2}{N}}\,. \tag{5.2}$$

The RMSF is a measure for the motility of an atom over a given timespan. It can be compared to the Debye-Waller or B-factor that accounts for the thermal motion of molecules in x-ray

Figure 5.1: Principle Component Analysis in 2 dimensions on an arbitrary set of data.

crystallography by

$$\texttt{RMSF} = \sqrt{\frac{3B}{8\pi^2}}\,. \tag{5.3}$$

5.3 Principal component analysis

Principal component analysis (PCA) tries to restrict the analysis of $3N$ dimensional conformational space to subspaces that contain the most relevant information on the systems phase space behavior [28]. This is done by constructing a new orthogonal basis set. This basis set is chosen such that the variation of the projection of the analyzed points onto the first basis vector is maximal. Then the second is chosen to be orthogonal to the first one and to again maximise the variation of the data projected onto it. Figure 5.1 shows a PCA performed on a set of two-dimensional data. The new basis vectors are obtained by diagonalizing the covariance matrix of given data. The eigenvectors correspond to the basis looked for while the eigenvalues weight the information given by a system by the data projected on a certain corresponding eigenvector.

5.4 Full correlation analysis

The motions PCA modes describe do not need to be uncoupled. The reason for that lies in the orthogonality of the PCA modes. The full correlation analysis (FCA) tries to minimize the mutual information of these modes [29]. With the software employed here (g_fca) this is done by taking the PCA eigenvectors as an initial guess and then rotating them independently to minimize the mutual information of these. The coordinate transformation to the FCA basis set used in this approach is linear.

5.5 Base pair interactions

The base pair interactions have been determined by an algorithm developed by Yang et. al [30] by fitting each nucleobase to a standard reference frame `RNAView` derived from the Cambridge Structure Database [31]. The assignment of two bases to a base pairing interaction type is done by comparing the angles and distances of the assigned local coordinate systems and counting the number of hydrogen bonds.

5.5.1 Trajectory analysis of base pair interactions

In thermodynamical equilibrium base pairs usually break and reform. For that reason it is not possible to speak of base pair interactions being present or not but rather of occupation probabilities. Every 50 ps base pair interactions were analysed and averaged subsequently. For visualization every possible interaction has a symbol assigned to it. The script package `bpi` that was developed in this thesis automises all trajectory base pair interaction analysis.

5.6 Error Estimation

For quantities x measured in the course of time the error of the mean was estimated using the standard deviation of the data σ_x. According to Janke [32] with the number of measurement points N and the autocorrelation time τ_x the error of the mean is

$$\sigma_{\bar{x}} = \sigma_x \sqrt{\frac{2\tau_x}{N}} \ .$$

The autocorrelation time τ_x was estimated using a single exponential fit to the autocorrelation function of the data.

Part III

Studies on RNA

6 Temperature dependence of RNA flexibility

6.1 Introduction

After introducing the general methods that are used in this theses a double stranded RNA helix at different temperatures will be handled as first simulation system.

A remarkable ability of RNA is to undergo large conformational changes in response to cellular signals [1]. These signals include changes in temperature. Riboswitches, e.g., regulate gene expression on temperature influence thus enabling organisms to adapt to different environments instantaneously [33]. Changes in RNA dynamics can be accessed experimentally with techniques like NMR residual dipolar couplings [34].

On the other side it is inevitable for biological systems to maintain their structure through environmental changes. The overall shape of biological complexes is constant through a wide variety of environmental conditions. One example is the ribosome of different species whose overall tertiary structure is maintained be it environments with high salt concentration *haloarcula marismortui* lives in, high temperature regions *thermus thrermophilis* inhabits or modest conditions under whose *escherichia coli* lives. The crystal structures of all those ribosomes reveal that the general shape of their RNA helix regions is similar [35–38].

MD simulation studies on RNA helices have been performed on the ribosomal 16S RNA helix 44, a standard RNA duplex of 37 base pairs length and the helix 42-44 related portion of 23S RNA [39, 40]. In these simulations at 300 K isotropic bends of all simulated helices have been observed suggesting that RNA helices are intrinsically elastic molecules.

We studied an RNA double helix by means of MD simulations as a model system to determine how a common RNA motif and its elastical properties are affected by changing environmental conditions like temperature. For that purpose we performed simulations of an A-form RNA helix in a temperature range from 275 K to 400 K.

Figure 6.1: Solvated 19 bp A form dsRNA. Simulation system setup in a rhombic dodecahedron box.

6.2 Methods

6.2.1 Simulation Setup

The structure employed for the MD simulations is an A-form RNA helix comprising 19 base pairs (pdb code `1QC0`) [41]. See Fig. 6.1 for a depiction of the simulation system in a dodecahedron box filled with water and Na^+ and Cl^- ions. For 275 K to 310 K five simulations were set up for each 5 K step. Five additional simulations at 400 K were carried out.

The amber99 force field [42] and TIP3P water model [43] were used throughout all simulations.

After an energy minimization routine to remove close contacts with water molecules and correct conformations an equilibration with the phosphor atoms restraint was performed for 10 ns. After that equilibration time a free MD was carried out for 20 ns.

6.3 Results

6.3.1 RMSD analysis

The RMSD depicted in Fig. 6.2 has been computed with respect to an MD trajectory snapshot after position restraint equilibration, 1 ns of free MD and 10 ns of free MD, respectively. This was done in order to distinguish the contributions to the RMSD from relaxation of the system to a local Gibbs free energy minimum on the one hand and phase space sampling on the other. A good agreement between all RMSDs is an indication that the contribution to the RMSD from relaxation of the system is small enough. This is a precondition for the system to be

Figure 6.2: RMSD of double stranded RNA with respect to the starting structure (black), a reference structure after 1 ns of equilibration (red) and a reference structure after 10 ns of equilibration(green). Values were averaged over 5 simulations each, the gray shaded area shows the standard deviation of the RMSD with respect to the starting structure.

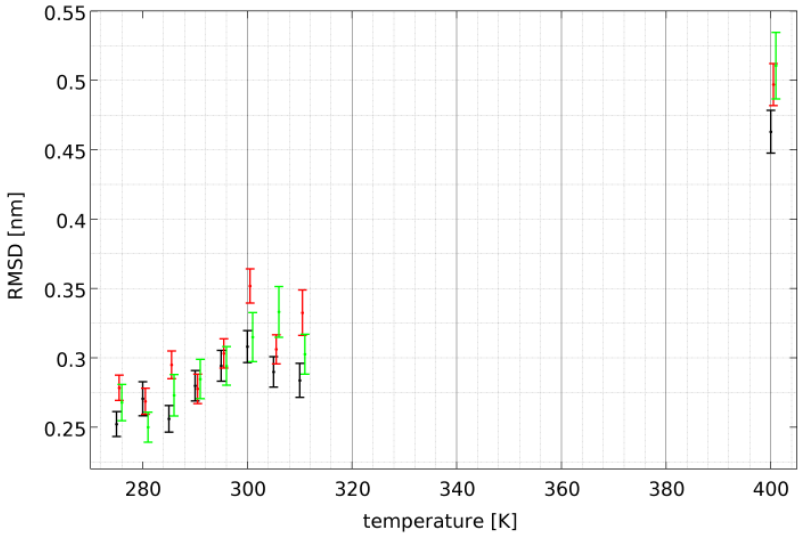

Figure 6.3: Averaged RMSD value. Averaging was done after 10 ns of equilibration. It has been calculated with respect to the starting structure (black), a reference structure after 1 ns of equilibration (red) and a reference structure after 10 ns of equilibration (green, averaging done excluding 5 ns of autocorrelation time in this case).

equilibrated [44].

For all simulations except at 400K the RMSD stays stable at 10 ns latest and RMSDs from different reference times show similar behavior from that time on. For that reason data for analysis have been acquired after 10 ns of free equilibration. The simulation at 400 K still not completely equilibrated. It is used here as a test case for an extreme parameter set to compare the relative deviations of simulations at lower temperatures to a larger distortion.

The mean RMSD gives a hint of how much of configurational space is sampled and thereby of how flexible the structure is. The mean RMSD values with respect to the starting structure and structures after 1 ns and 10 ns of free MD are shown in Fig. 6.3. The RMSD was averaged over the last 10 ns of the trajectory when taken with respect to the starting and 1 ns structure. It was averaged over the last 5 ns of the trajectory when taken with respect to the 10 ns structure. Overall the mean RMSD grows with temperature, no matter which reference structure has been chosen. The temperature dependence of the mean RMSDs is almost linear. The mean RMSD with repsect to the starting structure is usually lower than for the other reference structures. At higher temperatures the mean RMSDs deviate more then at lower temperatures. The deviations go up to 0.05 nm.

6.3.2 Hydrogen bonding

The total hydrogen bond energy between all hydrogen bond acceptors and donors in the helix was estimated using the espinosa formula. This formula is based on the distances of the hydrogen bond donor and acceptors [45]. See Fig. 6.4 for graph showing its temperature dependence. Except for 400 K where it raises to about -1.25 kJ, the total hydrogen bonding energy is independent of temperature at around -1.35 kJ at all temperatures with minimum values at 290 and 295 K. The errors in the mean hydrogen bonding energy are shown as red bars. They indicate that the drop in hydrogen bonding energy can be regarded significant.

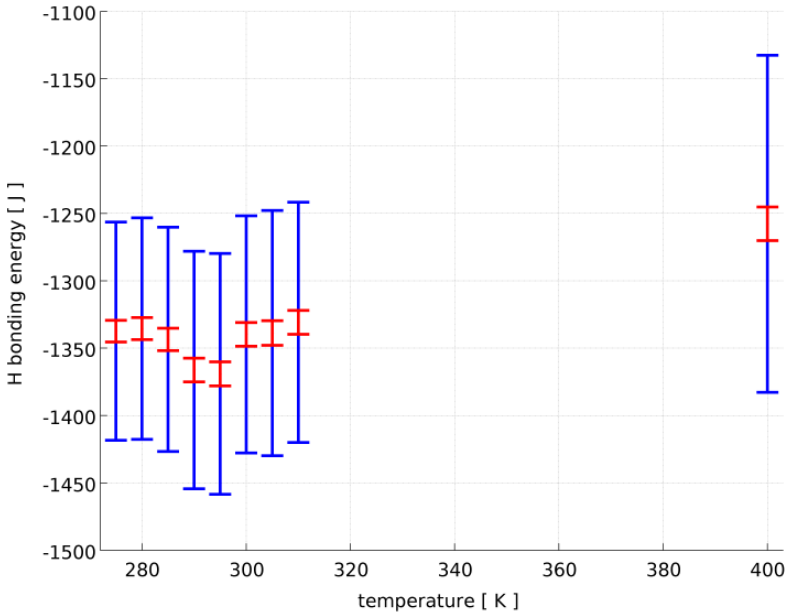

Figure 6.4: Hydrogen bonding energies as estimated by the Espinosa formula. Blue errorbars denote the standard deviation of the hydrogen bond energy in the course of time. Red error bars show the error of the mean hydrogen bonding energy.

Fluctuations in hydrogen bonding energy are indicated by the blue bars that show the standard deviation of the hydrogen bonding energy. The fluctuations of hydrogen bonding energy are in a range of 10 percent of the total hydrogen bonding energy for all temperatures except 400 K where it raises up to 20 percent.

6.3.3 Base pairing

Base pair occupation numbers were averaged over all trajectories as described in chapter 5.5.1. Results are shown in Fig. 6.5. Up to a temperature of 305 K occupancy numbers stay virtually constant – the secondary structure of the double helix is preserved. Canonical base pairs are present 96 to 99 percent of the time. It is only at 310 K that the canonical Watson-Crick base pairs become single hydrogen bonds at the base pairs A_{31} U_8 and U_{36} A_3 and a single hydrogen bond is lost in the base pair G_{35} C_4.

At 400 K the tails of the double helix begin to disintegrate. Canonical Watson-Crick base pairs become weaker Watson-Crick-edge to Watson-Crick-edge interactions. Occupation probabilities are expected to be the same at the respective mirrored base pairs e. g. $A_{22}U_{17}$ should match A_3U_{36} as a palindromic helix was simulated. This is not the case for 310 K and 400 K which indicates an insufficient sampling for higher temperatures. Nevertheless a generally symmetric behavior can be observed at the tail regions of the helix.

6.3.4 Spring model

A simple model to describe the RNA double helix is a one-dimensional harmonic spring by looking at the end-to-end distance. Two different regimes of RNA flexibility can be described. One regime describes RNA as flexible along all space directions. In the other it can be

Figure 6.5: Secondary and tertiary base pair interactions averaged over the whole trajectories. Percentages denote occupation probabilities. Symbols are used as described in 5.5.1. Marked areas denote significant changes in base pair interactions.

Figure 6.6: Free energy profile for length of a 19 bp dsRNA.

Figure 6.7: Free energy profile for length of a 19 bp dsRNA with estimated errors.

assumed to be stiff enough to be fixed to one-dimensional elongation only. A measure that tells one regime from the other is the persistence length. It is defined as the length after which correlation of the tangent vector on a polymer is lost. For double stranded RNA it is approximately 70 nm [46]. As the RNA simulated here is only approximately 5 nm in length RNA had been approximated as elastic rod in one dimension. To check if this approximation is appropriate

Measuring the end to end distance l over time gives the length distribution. This distance was defined as the distance of the centers of mass of base pairs $C_{38}G_1$ and $C_{19}G_{20}$

States corresponding to a short and a long elongation are shown in Fig. 6.9. On the one hand at significance levels $\alpha = 0.05$ a Kolmogorov-Smirnov test reveals that the hypothesis that the end-to-end distance is normally distributed must be rejected. On the other hand the commutative sum of both distributions show differences below one percent (data not shown). For that reason a gaussian distribution is a justifiable approximation. With that a spring constant K follows with the Gibbs free energy G, Boltzmann constant k, temperature T, variance of the distribution σ, occupation probability p and equilibrium length l_0 the spring constant K equals $\frac{1}{2}\frac{kT}{\sigma^2}$.

The equilibrium length and calculated spring constants are shown in 6.8. The helix gets softer the higher the temperature becomes. As long as secondary structure elements cannot break the helix becomes longer, too.

Free energy differences for states corresponding to different lengths can be estimated using $\Delta G = -kT \log \frac{p_1}{p_2}$ where p_1 and p_2 denote the probabilities for finding the system in a state one or two. The potentials gained are harmonic to a good approximation and flatten for higher temperatures. The simulation at 400 K shows a different behavior as the minimum free energy helix conformation is shorter and the free energy profile is much flatter and anharmonic than the others.

6.3.5 PCA analysis

To extract dominant modes of motion a PCA basis set has been created by analysis of all trajectory data. The eigenvalues of the correlation matrix were sorted according to their value. The first three eigenvectors on which a more detailed analysis will follow contain more than 50 percent of the overall information of the whole system.

Fig. 6.10 shows the motions the first three eigenvectors refer to by showing structures that refer to extreme values on the projection of the chosen eigenvectors and interpolations inbetween. The first eigenvector refers to a bending motion around a vector that is perpendicular to the helical axis. Eigenvector two describes a twisting motion around the helical axis. A stretching motion along the helical axis is described by eigenvector three.

The third eigenvector is highly correlated to the end-to-end distance of the helix with a correlation coefficient of 0.8 ± 0.1. This was expected by investigating the motion that is represented by the third eigenvector in PCA space indicated in Fig. 6.10.

Projections on the space spanned by the first and second eigenvector for trajectories at 275 K and 400 K are shown in Figs. 6.11 and 6.12. For 275 K and 400 K all points follow approximately a gaussian distribution independent from simulation time. At 400 K the system prefers the more bent conformations than at 400 K and is covering more of conformational space.

To gain a clear picture for all nine simulations that have been carried out, a bivariate kernel density estimator was used to transform the point clouds into local densities. From that

Figure 6.8: Spring constant and equilibrium length for dsRNA. Results are color coded, ranging from low temperatures (blue) to high temperatures (red).

Figure 6.9: Structure of the double helix at the most extended (lefthand) and shortest (right-hand) form at 275 K (left) and 400K (right)

Figure 6.10: Interpolated structures for extreme values of projections on first three eigenvectors for all trajectories.

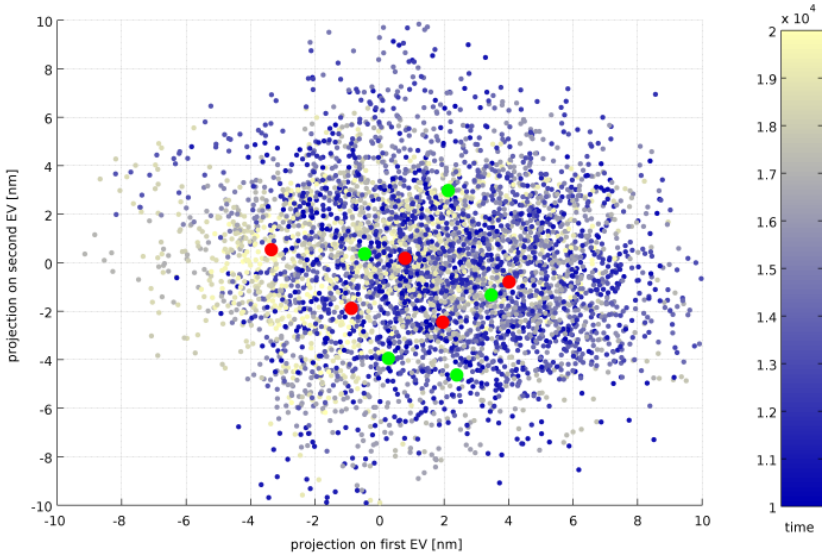

Figure 6.11: Projection of five different trajectories at 275 K on the plane spanned by eigenvectors one and two. Starting points of the simulations are shown as green dots, end points as red dots. For points corresponding to later simulation times are less saturated.

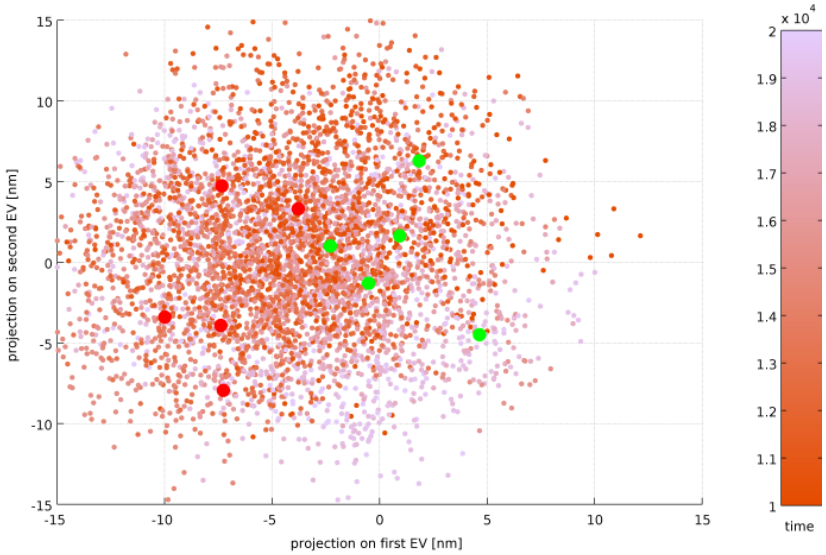

Figure 6.12: Projection of five different trajectories at 400 K on the plane spanned by eigenvectors one and two. Starting points of the simulations are shown as green dots, end points as red dots. For points corresponding to later simulation times the saturation is less.

contour lines along constant densities corresponding to constant free energies were derived. For each temperature two contours at low and high densities depicting the maximum and width of the probability distribution are shown in Figure 6.13.

The density flux in PCA space due to changing temperature was calculated by subtracting subsequent densities, omitting the 400 K simulation, and averaging over the gradient of the difference density. By that a qualitative picture on the temperature dependence of the projections in PCA space can be revealed. The vector fields in Figs. 6.14 show that flux and structures representing conformations that stay stable during temperature shift. This structures must be stabilized by enthalpic effects.

The double helix essentially explores more conformational space with raising temperature without a specially preferred direction.

Analysis of the PCA space for the first eight eigenvectors did not reveal further separation of the different trajectories in phase space but a rather harmonic behavior of the system.

6.4 Discussion

6.4.1 Equilibration

The evaluation of RMSD shows that the system is well equilibrated at 10 ns for all setups except at 400 K. The deviations of the mean RMSD indicate that a time of 1 ns of free MD is not enough for equilibration at temperatures from 300K and above.

6.4.2 Influence of the force field

The AMBER99 force field that was employed in this work was parameterized to give correct free energies and secondary structure properties at 300 K. For that reason artifacts induced

Figure 6.13: Projection of all trajectories on the plane spanned by eigenvectors one and two. Two density contour lines are shown, the outer lines correspond to lower densities while inner lines correspond to lower ones.

Figure 6.14: Average change in probability density in PCA space of the first and second eigenvector when the system is simulated at higher temperatures. Arrows head from states preferably occupied at lower temperatures to states more preferably occupied at higher temperatures.

by the force field are expected to become bigger the larger the deviations from 300 K are.

6.4.3 Constant behavior over a large range of temperatures

Double stranded A-form RNA as simulated here seems to be remarkably stable over a large range of temperatures. Furthermore the overall dynamics of RNA seem to be affected much less by changes in temperature than expected. This speaks for the ability of RNA to maintain the same biological function over different environmental conditions. On the other hand it suggests that thermosensing RNA is to have structural elements that are essentially more complex than a canonical helix.

6.4.4 Hydrogen bonding and base pairing

The canonical double stranded RNA helix is stabilized to a big extend by hydrogen bonds of canonical Watson-Crick base pairs. These base pairs do virtually never break in the lower temperature range from 275 K to 310 K. Therefore the double helix stays stiff at low temperatures. Gaussian distributions of PCA coordinates or helix length indicate that the system is moving in a harmonical regime which is left at 400 K when the canonical base pairs can break.

The minimum of the hydrogen bonding energy at 295 K can be explained by the molecules having just enough kinetic energy to access the optimal hydrogen bonding distance while on the other side not yet being able to leave these minima.

6.4.5 Spring constant and persistence length

As known so far the spring constant revealed in this MD simulation has not been found experimentally. Previous experiments that have been carried out on nucleic acids to classify their elastic behavior refer to much longer polymers. The length of these polymers is much longer than the persistence length. Therefore these polymers not behave like an elastic rod in one dimension but rather isotropic. Additionally RNA used in experiment was not in the canonical A-form with standard Watson-Crick base pairs. For that reason folding into more complex shapes like bulges and formation and breaking of base pairs play an essential role on the elastic properties of RNA [47].

For further studies the persistence length of RNA could be estimated by simulating short pieces of RNA as done in this thesis. This could be done by calculating the expectation value of the cosine of the tangent vectors on polymers.

6.4.6 Entropy and enthalpy estimates

The temperature dependence of the change in Gibbs free energy ΔG for elongation of the RNA double helix was calculated in this thesis. An estimate for entropic (S) and enthalpic (H) contributions to the dynamics could be done by approximating $\mathrm{d}G = -T\,\mathrm{d}S + \mathrm{d}H$ linearly by $\Delta G = -T\Delta S + \Delta H$. However large deviations in the estimated entropies and enthalpies were found while attempting to use this method for the simulated system (data not shown). This indicates insufficient sampling for that method. For appropriate use of this method the error of the free energy estimates would have to decrease by a factor of ten at least. To decrease the error in free energy estimates by a factor of ten, simulation times would have

to be approximately a hundred (ten square) times longer. For that reason other free energy calculation methods should be employed.

6.4.7 PCA and FCA

The first two PCA modes are in good agreement with the bending motions described by previous simulations on RNA A-form helix [39, 40] even if they are much less pronounced due to the much shorter helix simulated in this thesis.

A FCA was performed on the data but did not reveal qualitatively new results and is therefore not shown here. As the system was simulated in a very harmonic regime the motions described by the PCA eigenvectors were highly uncorrelated. The mutual information matrix revealed little mutual information for the PCA modes already.

7

Effect of nucleobase modifications and ions on tRNA structure and flexibility

7.1 Introduction

The general structural differences of modified and unmodified tRNA and tRNA that is complexed by ions will be subject of the following resarch.

In the following chapter tRNA will be introduced. Then a summary of the effects of nucleotide modifications on tRNA and on magnesium ions and tRNA will follow. Previous studies and basic questions will be outlined afterwards.

7.1.1 tRNA

Transfer RNA (tRNA) is a macromolecule made of RNA containing usually 74 – 95 nucleotides. It is present in all living cells where it carries amino acids to the ribosome for protein synthesis and decodes the mRNA bound there. Apart from that, recent studies suggest that tRNA plays a role in processes like cell wall synthesis, reverse transcription, heme and cholorophyll biosyntheses, amino acid biosynthesis regulation and viral genomes [48].

The primary structure of tRNA has first been determined by Holley et. al in 1965 [49] who first proposed the now canonical cloverleaf structure which is shown with secondary and tertiary interactions in Fig. 7.1. The tertiary structure of tRNA determinded by X-ray crystallography is shown in Fig. 7.2.

The canonical tRNA structure comprises a CCA-tail at the 3' end (grey) which can bind an amino acid, an acceptor stem which includes a phosphylated 5' end (purple), a D loop (blue;

Figure 7.1: Cloverleaf structure of tRNA showing secondary and tertiary base pair interactions. CCA tail in gray, acceptor stem in purple, D loop in blue, anticodon loop in red, variable loop in lime and TΨC loop in cyan. Modified nucleosides are not capitalized.

Figure 7.2: Tertiary structure of tRNA (pdb code `1EVV`) in "cartoon" representation. The phosphate backbone is shown as ribbon while the sticks represent the orientation of the nucleosides. Color coding refers to 7.1.

Figure 7.3: Atomistic representation of tRNA. Nucleosides subject to modifications (modification sites) are shown in orange; ion binding sites are shown in green.

usually containing *D*ihydrouridine), an anticodon loop (red) containing the 3 nucleobases coding for the specific amino acid at the CCA-tail a variable loop (lime) with a differing number of bases for different tRNAs and a TΨC loop (cyan) which forms the elbow region with the D-loop for structural stability. In its tertiary fold D and TΨC loop interact via hydrogen bonds, namely interactions of G_{19} with C_{56} and G_{18} with Ψ_{55} of which the former is named Levitt base pair [50].

Transcription of tRNA is performed by RNA Polymerase III from the genome sequence in DNA [51]

Transfer RNA is aminoacylated by cognate aminoacyl-tRNA syntheases (aaRS). It then carries amino acids to the ribosome where these amino acids are subject to protein synthesis. For each species there are several kinds of RNA coding for different amino acids. Each kind of tRNA has a unique anticodon sequence to be recognized by mRNA on the ribosome, in contrast several anticodon sequences can code for the same amino acid. According to that the specific nth kind of tRNA coding for an aminoacid aa is denoted by tRNA_n^{aa}.

Transfer RNA in vivo is subject to nucleoside modifications (shown in orange in Fig. 7.3) and is furthermore complexed by Mg^{2+} (shown as green spheres in Fig 7.3).

7.1.2 Nucleoside Modifications

Derivatives of the canonical nucleosides A, U, C, and G are called modified nucleosides. They were first reported by Hotchkiss in 1948 [52]. Except for queuosine all modifications occur after transcription by proteins.

Different types of modifications are listed in table 7.1. Modified nucleotides are denoted by either a different letter from the canonical base alphabet for very common ones like dihydroudridine as D or pseudouridine as Ψ or by small letters showing the types of modification followed by its position like m^1A denoting 1-methyl-adenosine.

Modifications are reported to change physical and chemical properties such as electrostatics, base pairing and stacking, conformations of nucleotides, solvation and ion binding and the

Table 7.1: Modifications present in RNA and their suggested effects.

Modification	Effect	Reference
Uridine to pseudouridine		
	improves stacking interactions	[53, 54]
	forms additional hydrogen bond by N1	[55]
	forms bridge to adjacent phosphate residues mediated by stably coordinated water molecule	[55, 56]
	rigidifies backbone	[55]
Uridine to dihydrouridine		
	enhances conformational flexibility	[57]
Methylation		
	augments polarizability and hydrophobicity	
	enhances base stacking	[4, 53]
	sterically induces structural changes	[58]
	blocks hydrogen bonds	[59, 60]
	affects hydrogen bonding strength by introducing positive charge to nucleosides	[61]
	stabilizes C3'-endo puckering conformation	[62]
	prevents hydrolysis	[63]
Recognition related Modifications		
	regulate wobble base pairing (position 34)	[64]
	essential for reading frame maintenance (position 37)	[65]
	stabilize codon-anticodon interactions	[66]
	affect codon specifity	[67]
	increase aminoacylation specifity for cognate aaRSs	[67]

Table 7.2: Chemical and physical properties affected by nucleotide modifications. (Adapted and extended after [4])

Affected property	Description
Electrostatics	
	protonation dependent transient charges
	introduction of positive/negative charges
Base pairing and stacking	
	inhibition of non-canonical base-pairing
	enhancement of non-canonical base-pairing
	disruption of canonical base pairs
Nucleotide Conformation	
	alteration at nucleoside
	phosphodiester bond restriction
	sugar pucker
	new conformations and chemistry by interaction of modification with modified nucleoside
Coordination of solvent and ions	
	reordering of water
	coordination of metal ions
Thermodynamic de/stabilization	
	lowering entropic barrier for ligand binding
	stabilizing tertiary structure by increasing entropy

stability of the tertiary structure of tRNA. A summary of the most important changes in physical properties is listed in Table 7.2.

Modifications improve functions such as codon-anticodon-recognition and aminoacylation. A potential medical application might employ that all retroviruses utilize m^1A58-containing tRNAs to prime reverse transcription.

Three modifications are essential, and six other modifications located in the anticodon region seriously affect the phenotype [51]. However, the deletion of enzymes that provide most nucleoside modifications shows few obvious effects even though they are evolutionary high conserved [75].

New possibilities added to the highly specific tRNA recognition mechanisms by modifications could explain their impact on vitality on the one side. On the other side these modifications can help tRNA to conform the narrow frames of fit into the ribosome and processing enzymes amid diverse sequences. Modifications can be needed to provide the necessary folding pathway thereto [77]. A fact that supports the later is the rapid turnover degration of unmodified tRNA that reduces half-life of tRNA from hours to minutes [78].

A special region of interest for recent research has been the anticodon region. Most modifications are found there and this region is essential for codon-anticodon recognition. Human $tRNA_3^{Lys}$ modifications mnm^5s^2U34, t^6A37 and $\Psi 39$ order the structure of the anticodon stem loop [79]. For $tRNA^{Phe}$ a restriction of conformational space due to modifications has been found by NMR spectroscopy [80].

Modifications in yeast $tRNA^{Phe}$ which is employed in this work and their specific role known

Table 7.3: Positions of Modifications in yeast $tRNA^{Phe}$ resolved in the crystal structure with pdb code `1EVV`. Obsolete names shown in Brackets. Enzyme abbreviations are listed in Table 7.4 .

Pos.	Modification	Mutant phenotype	Modif. Enzyme	Ref.
10	m^2G	not clear	-	-
16	h^5h^6U (D)	not clear	-	-
17	h^5h^6U (D)	not essential	DUS1,2	[68, 69]
26	m_2^2G	not essential	TRM1	[70]
32	Cm	slow growth / translation	TRM7	[71]
34	Gm	slow growth / translation	TRM7	[71]
37	yW	very sick	TRM7	[72]
39	Ψ	slow growth	PUS3 (DEG1)	[73]
40	m^5C	not essential	TRM4	[74]
46	m^7h^8G	temperature dependent growth defect	TRM8/TRM82	[75]
49	m^5C	not essential	TRM4	[74]
54	m^5U	not essential	TRM2	[70]
55	Ψ	not essential	PUS4	[70]
58	m^1A	essential	TRM6/TRM61 (GCD10/GCD14)	[76]

Table 7.4: Enzyme names for tRNA modifications.

Abbreviation	Enzyme name
DUS	tRNA *d*ihydro*u*ridine *s*ynthease
PUS	tRNA *p*seudo*u*ridine *s*ynthease
TRM	*tR*NA *m*ethyltransferase

Figure 7.4: Different modes of Mg^{2+} interaction with RNA. RNA shown in black, magnesium in green and water in red(oxygen) and gray(hydrogen). Mg^{2+} can interact chelated and dehydrated (left) or via long range electrostatic interactions (right). A mixed situation where Mg^{2+} interacts via its hydration shell is shown in the middle. Adapted from Draper [86].

so far are listed in Table 7.3.

7.1.3 Mg^{2+} ions complexed to tRNA

Magnesium ions in contrast to nucleoside modifications are essential for folding of tRNA [5] and maintaining its tertiary structure [81]. As shown in hydroxyl radical cleavage experiments, even small concentrations of Mg^{2+} can have a large effect on its tertiary structure [82].

It was shown that the ability of denatured tRNA to accept amino acids can be restored by Mg^{2+} [83] and that tRNA renaturation is possible by adding uni- and divalent ions [84]. This effect is primarily due to the highly anionic nature of RNA. By compensating for electrostatic repulsion by the phosphate backbone cations are crucial for RNA folding and maintaining functional forms. Tan reports for example that the presence of Mg^{2+} loop formation is more favorable than without due to shielding of backbone charges [85].

Molecular Picture

Together with water Mg^{2+} ions can not only interact with nucleic acids in solvent from the outside but rather occupy specific binding sides at RNA [86]. These different modes of ion binding to RNA are depicted in Fig. 7.4.

For specific binding of Mg^{2+} ions to RNA there are two different modes. Outer sphere or "diffuse" binding describes binding of ions with their complete water shell sharing water of hydration with RNA. Inner sphere binding or "chelation" refers to ions binding directly to RNA after dehydration.

Using a nonlinear-Poisson-Boltzmann modell for the electrostatic potential on the crystal structure of tRNA Draper suggests that Mg^{2+} binding to tRNA is mainly governed by electrostatic interactions and that the coordination of hydrogen bonds and other short range forces play a minor role in Mg^{2+} binding to tRNA. The stabilizing effect of magnesium ions on tRNA is explained by strong coloumb binding to electronegative areas around tRNA and

Table 7.5: Modified nucleosides in simulated tRNA

Nm 2′-O-Methylnucleoside	h^6m^1A 6-Hydro-1-methyl-adenosine
m^2G N^2-Methylguanosine	m_2^2G N^2-Dimethylguanosine
yW Wybutosine	m^7h^8G N^7-Methyl-8-Hydroguanosine
m^5C 5-Methylcytidine	h^5h^6U 5,6-Dihydrouridine
m^5U 5-Methyluridine	Ψ pseudouridine

"entropic release" of Na^+ ions from tRNA. Because of their single charge more sodium ions than magnesium would bind to tRNA which leads to an increase in entropy as soon as single magnesium ions can take the place of two sodium ions [87].

Serebrov et al. support this model in their work by titration experiments combined with fluorescent labels that give hints on structural transitions [88]. Yet they propose a unique strong binding site of Mg^{2+} where Mg^{2+} might interact not only mainly via its electrostatic potential but by direct chelation to the phosphate backbone. This interaction site is formed by the phosphate groups of G_{20} and A_{21} of the D loop.

7.1.4 Previous MD studies on tRNA

The first simulations on tRNA were done by Harvey et al. who did 12 ps of molecular dynamics on $tRNA^{Phe}$. Multiple simulations tRNA have been carried out by Westhof and Auffinger et al. These include a simulations of the anticodon loop of $tRNA^{Asp}$ [89, 90] and simulations of the whole $tRNA^{Asp}$ molecule. As no structural data for magnesium ions bound to tRNA has been available in the times when these simulations were carried out all their setup lack magnesium ions. In these simulations it is shown that with the particle mesh Ewald method tertiary interactions can be maintained in 500 ps of simulation. Lahiri and Nielsson performed simulations of the anticodon domain of yeast $tRNA^{Phe}$ in presence and absence of different codons, showing that codon binding stabilizes the anticodon arm [91]. All workers point out that they could not account for magnesium bound to tRNA, neither did they investigate the effect of nucleotide modifications.

7.1.5 Questions that will be addressed

As classical molecular dynamics are employed there are certain limitations on the observable properties. While polarizations charge transfers and protonations cannot be observed [1], steric and conformational effects are amenable to analysis. Reordering of water as well as free energy differences can be accessed by MD methods but are costly for bigger systems like tRNA. Certain entropy terms can be estimated, the complete calculation of the entropic contribution of certain nucleoside modifications is still far out of scope.

Molecular dynamics can access the behavior of a molecule like tRNA in the liquid phase on an atomistic basis. General dynamical properties like the flexibility of certain domains of tRNA or dominant motions will be accessed. Furthermore local properties like tertiary interactions and base pairing patterns will be observed.

7.2 Methods

7.2.1 Simulation setups

We set up our simulations by extracting atomic coordinates from a yeast $tRNA^{Phe}$ crystal structure [92](PDB-code:1EVV). All known tRNA modifications to $tRNA^{Phe}$ have been resolved in this crystal structure. In addition to the tRNA this structure includes crystal water, magnesium ions and spermine.

Resolved crystal water was used in all simulations. The spermine that is also resolved is special to thermophilic species. It is known to stabilize tertiary structure at high temperatures

[1] Polarizable force fields and constant pH-protocols are currently under development

[93]. Spermine was removed to exclude interaction effects with modifications and magnesium ions with this ligand in order to probe the effect of nucleotide modifications and magnesium ions only.

The resulting setup including both, magnesium ions and nucleotide modifications will be denoted (I+/M+). To test for the effect of nucleotide modifications a set of atomic coordinates lacking nucleotide modifications was created by manually replacing modified nucleotides with unmodified ones. This setup including magnesium ions but lacking nucleotide modifications will be denoted (I+/M-). To test for the effect of magnesium ions these had been removed from the set of starting coordinates for the molecular dynamics simulation (I-/M+). As a test for the mutual influence of magnesium ions and nucleotide modifications a setup was created lacking both (I-/M-). Fig. 7.6 shows all different setups.

The same simulation protocol was used throughout. A rhombic dodecahedron box with a volume of 1003.98 nm^2 was filled with tRNA as described above, TIP3P water and sodium and chloride ions to a physiological concentration of 154 mM. An excess of sodium ions was used for charge neutralization.

Following a standard equilibration protocol steepest descent energy minimization was used to relax strain from the initial setup [94]. The energy minimized structure was used for a 10 ns simulation with phosphate atoms restraint. From the last nanosecond of this simulation structures were extracted at six different times. Using these structures an unrestrained MD followed for 100 ns.

7.2.2 Implemented force field parameters

Force field parameters of modified nucleotides have been implemented into the amber99 force field using the data available from John Santa Lucias database [95, 96]. The phosphylated 5' terminus of RNA has been parameterized using standard amber parameters for dihedrals, bonds and angles. For the derivation of partial charges a self consistent field approach was employed using the program GAUSSIAN with a 6-31G basis set. To reduce the computational effort partial charges were derived for dimethylphosphate which is a good approximation to the whole phosphylated 5' terminal base.

Ion parameters for sodium and chloride were used from Dang [97].

7.3 Results

7.3.1 RMSD

In order to determine whether the simulation system was in equilibrium the RMSD (shown in Fig. 7.5) was employed. The starting structure and structures from the trajectory at five and twenty nanoseconds were chosen as reference structures for RMSD calculation.

The canonical setup that includes ions and nucleotide modifications deviates most from the starting structure at around six nanoseconds. From 20 nanoseconds on the deviations from the starting structure stay almost constant with a standard deviation from six different simulations of 0.15 nm.

A starting structure that is far away from the structures usually sampled in equilibrium would add an offset to the RMSD. To check if the configurational space sampled is in close vicinity to the starting structure RMSD was taken with respect to different reference structures.

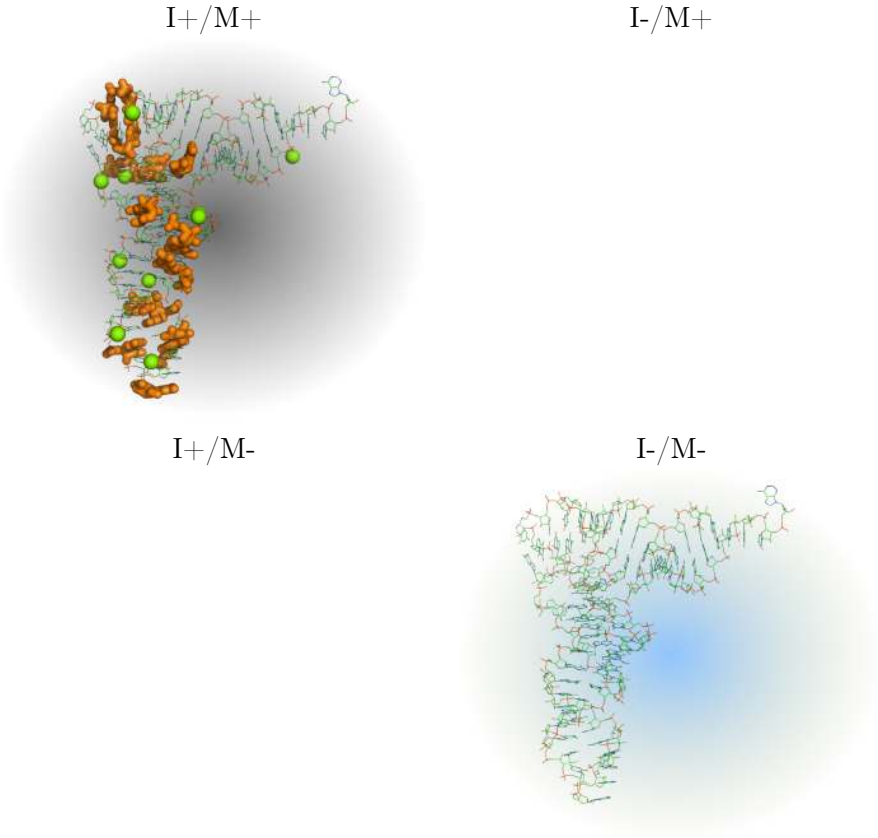

Table 7.6: Setups for tRNA simulations. Color coding consistently used throughout future illustrations: black denotes the canonical setup with magnesium ions and nucleotide modifications, red lacking ions, green lacking modifications and light blue a tRNA lacking both. Setups with/without modifications will subsequentely be denoted by M+/-, with/without Mg^{2+} ions by I+/-.

The RMSD with respect to the structure at five and at twenty nanoseconds simulation time reaches the level of the canonical RMSD after two nanoseconds for all setups except the canonical one. In the simulations that employ the canonical setup the RMSDs in a range from 100 ps to 10 ns deviate about 0.15 nm. The setup lacking nucleotide modifications deviates most from the starting structure at the last 90 ns. In a range from 3 to 30 ns of simulation the RMSD stays constant with a standard deviation of below 0.1 nm. From 30 ns on the structures extracted from the simulations deviate increasingly more.

The CCA tail that comprises the nucleotides 74-76 contributes a lot to the overall RMSD while it does not significantly influence the biologically relevant overall shape of tRNA. For that reason the RMSD was analyzed excluding the highly flexible CCA tail. See Fig. 7.6 for its time dependence. The same reference structures as above have been used in all cases.

The deviations from the starting structure are lower in all cases when the CCA tail is excluded. The general properties as described above stay conserved unless for the setup lacking nucleotide modifications but including ions. Here the increasing deviations from the starting structure in the last 70 ns of simulation are no longer present. This speaks for a different conformation of the CCA tail in this setup. The setup M-/I+ can be therfore considered sufficiently equilibrated when the CCA tail is not taken into account.

From this RMSD analysis an equilibration time of 20 ns has been derived which was used for all subsequent calculations of means of analyzed properties.

An averaging over the RMSD values after equilibration gives a measure for the extend of phase space that is sampled by the system and the deviation from the crystal structure. Mean RMSDs are shown in table 7.7.

With the CCA tail included the canonical setup shows the lowest deviations of about 0.35 ± 0.01 nm. A removal of Mg-ions increased the mean RMSD by 0.01 nm while a removal of nucleotide modifications while still including Mg-ions leads to a significant increase of the RMSD of 0.03 nm. By removing both Mg-ions and nucleotide modifications the RMSD increases by 0.04 nm.

TRNA deviates less from the starting structure in the simulations when the CCA tail is excluded. Furthermore the increase of the RMSD with respect to the starting structure is much less significant except for the setup lacking both, magnesium ions and nucleotide modifications. The setup lacking Mg-ions only but including nucleotide modifications shows even a slightly lower RMSD.

7.3.2 RMSF

The root mean square fluctuations of the tRNA as averaged over single bases give a hint on which regions of tRNA are stiff or highly flexible. Fig. 7.7 shows these RMSFs for the four simulated setups.

For all setups there are common properties of the root mean square fluctuations. The motility is very high for the CCA tail and decreases monotonically till the end of the acceptor arm. The D arm, the TΨC arm and the variable loop are relatively stiff with fluctuations between one and two Å. The bases 26 to 30 and 42 to 44 in the anticodon arm are relatively stiff while the bases 31 to 40 show increased flexibility in all setups.

Still there are significant differences in the RMSF for the different setups. The most striking differences between the different setups can be seen in the bases 34 to 44 in the anticodon arm. The bases 34 to 38 are much more motile when nucleotide modifications and ions are missing.

Figure 7.5: RMSD of tRNA with respect to the starting structure (black), a reference structure after 5 ns of equilibration (red) and a reference structure after 20 ns of equilibration(green). Values were averaged over six simulations each. The area shaded gray shows the standard deviation of the RMSD with respect to the starting structure.

Figure 7.6: RMSD of tRNA excluding the CCA tail. Computed with respect to the starting structure(black), a reference structure after 5 ns of equilibration (red) and a reference structure after 20 ns of equilibration(green). Values were averaged over six simulations each. The area shaded gray shows the standard deviation of the RMSD with respect to the starting structure.

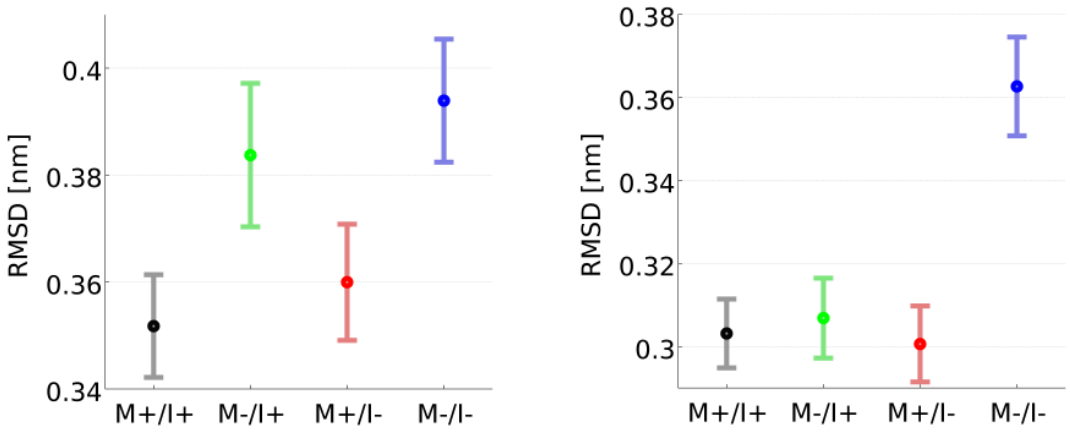

Table 7.7: Averaged RMSD of tRNA over free MD simulation time of 80 ns for six simulations each including CCA tail (left) and excluding CCA tail (right). Errors were estimated using the autocorrelation time of the RMSD.

The removal of magnesium ions only leads to an overall increased motility in the bases 37 to 44 in the anticodon stem.

The setup lacking nucleotide modifications, but including ions is the least motile one, followed by the canonical tRNA. A lack in magnesium ions leads to a slightly increased motility, if nucleotide modifications are lacking additionally motility is increased especially in the anticodon arm.

7.3.3 PCA and FCA

Treatment of the CCA tail

The highly flexible CCA tail is expected to contribute a lot to the overall motion of the system and thereby might dominate the PCA and FCA modes. For that reason PCA and FCA based on a coordinate set that included the CCA tail and one that did not include the CCA tail had been carried out. The difference in CCA tail treatment affected the results of the PCA and FCA analysis in a minor way. Therefore only the results with PCA and FCA including the CCA tail are shown here.

Set of coordinates for PCA and FCA analysis

Nucleotide modifications substitute or add atoms to nucleotides. For that reason a PCA or FCA in the $3n$ dimensional space on the set of all atomic coordinates of canonical tRNA was not possible. As the sugar and base of the nucleotides are very rigid due to their small cyclic structure the orientation and position of the sugar, the nucleobase and the phosphate atom in the backbone describe the overall tRNA structure sufficiently well. A set of atoms that is common in all setups and describes these orientations and positions can be used to define a common coordinate space for PCA or FCA. To describe the orientation of the sugar we used the coordinates C1', C3' and C4'. For the nucleotide C2, C4 and C6 were used. This summed

Figure 7.7: Per base root mean square fluctuations averaged after 20 ns of equilibration. Error bars shown here denote the statistical error.

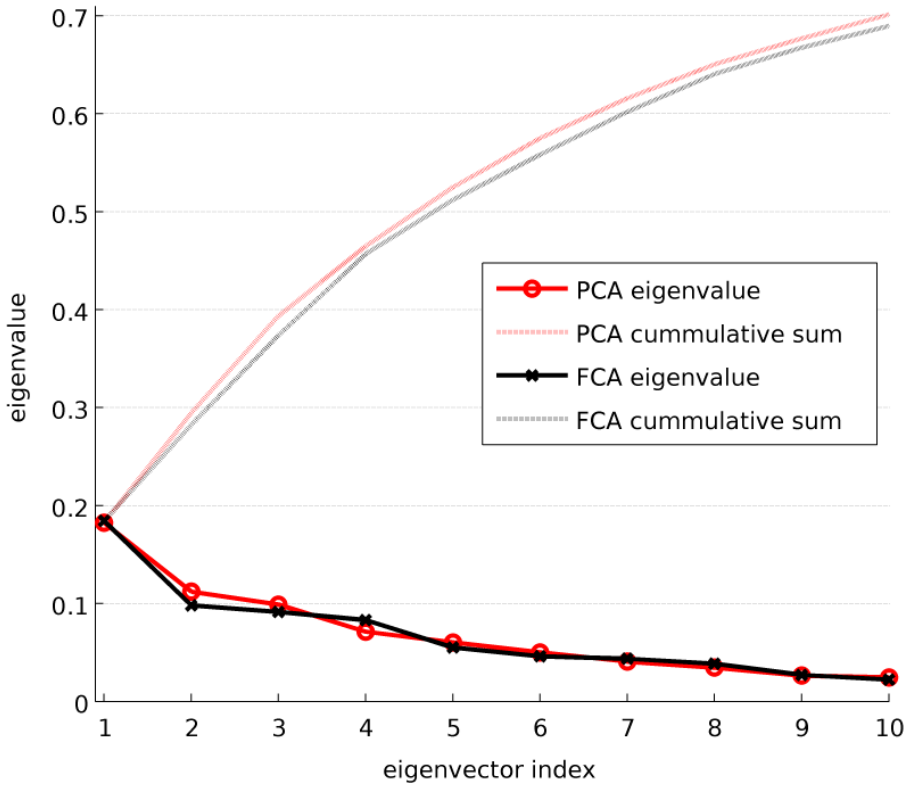

Figure 7.8: Information about the overall movement of the system captured by single modes. Modes were gained by PCA (red) and FCA (black) analysis.

up to an overall number of 532 atomic coordinates for the analysis including the CCA and 511 for analysis excluding the CCA tail, respectively.

Information content

Fig. 7.8 shows the information of the overall movement of the system the single modes include. The eigenvalues decay slowly which indicates that tRNA dynamics is governed by many small movements. Choosing a new basis set by FCA did not help to increase the fraction of the overall movement that is described by single modes, but rather lowered it slightly.

PCA

The first PCA mode describes mainly a change in the angle between the anticodon stem and the acceptor stem. It is depicted in Fig. 7.9. This is the most pronounced structural change that is present in all simulations.

A measurement of the angle between anticodon stem and acceptor stem will help to visualize this motion. As an approximation for the angle between anticodon stem and acceptor stem the angle between phosphate atoms of the backbone was chosen. The phosphate atoms that define that angle are the phosphates of G_1 at the very end of the acceptor stem, C_{56} from the TΨC loop in the ellbow region and C_{32} from the anticodon stem. The minimum angle observed was 72°, the maximum angle was 112°. In the crystal structure the angle is 80.2°.

Figure 7.9: Projection of structures to the extreme values of the first eigenvector gained by PCA analysis. The straight black arrow denotes the axis of rotation, dark blue lines indicate the geometrical properties at maximal projection values while light blue lines denote geometrical properties at minimal projection values. Angles were measured between the phosphates of G_1,C_{56} and C_{32}.

While tertiary structure changes largely when moving along this eigenvector a comparison of secondary structure of different projections on extreme values show that this is not necessarily true for secondary structure.

The histogram in Fig. 7.10 shows the distribution of the projection of different trajectories on the eigenvector in PCA space corresponding to that mode. The statistical error was estimated by creating histograms for six independent simulations. The higher the projection values the smaller the smaller angle between anticodon stem and acceptor stem. The energy landscape is close to harmonic for all setups. Distinct substates do not exist.

Lacking nucleotide modifications leads to configurations that have lower opening angles than the canonical setup. The absence of nucleotide modifications and magnesium ions leads to even lower opening angles while the absence of magnesium ions leads to a more open conformation.

The second eigenvector as shown in Fig. 7.11 describes a rotation and shortening of the anticodon stem that is accompanied by a pivoting motion of the acceptor stem. The major groove distance of the anticodon stem as defined of the distance between the phosphate atoms of the nucleotides 39 and 25 shortens from 11.0 Å to 6.4 Å. In the crystal structure this distance is 8.3 Å.

A histogram of the projections on the second eigenvector is shown in Fig. 7.12. Small eigenvalues refer to a prolonged anticodon stem and the acceptor stem pivoted towards the flexible loop. While the removal of ions does not change the distribution of projections on the

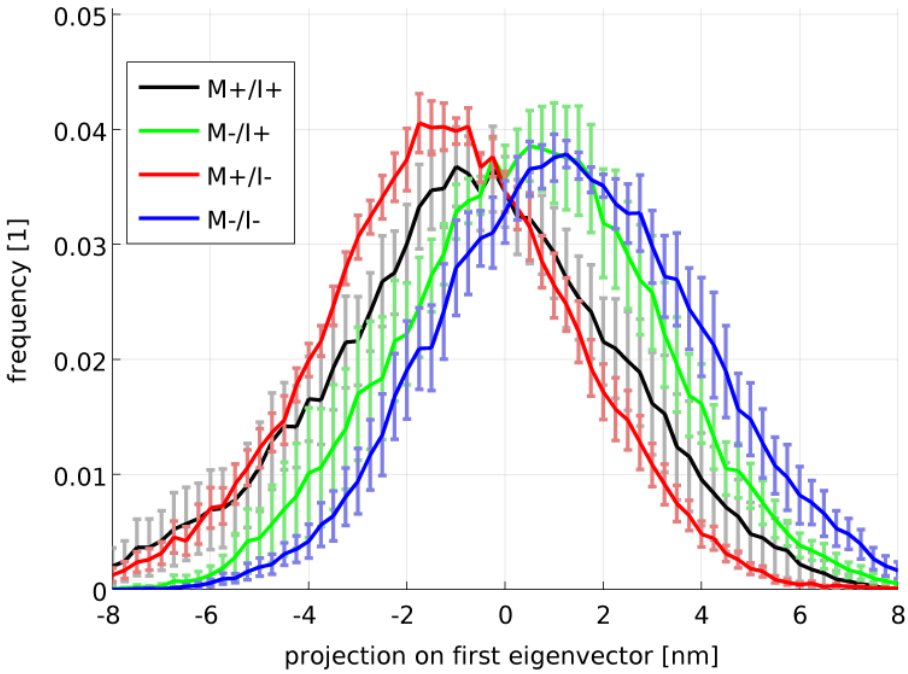

Figure 7.10: Histogram of projection of trajectory on the first eigenvector gained by PCA analysis.

second eigenvector the removal of nucleotide modifications while including magnesium ions shifts the whole distribution to higher values. Surprisingly the probability distribution for setup that includes ions and nucleotide modifications is very close to the canonical one.

A two dimensional representation of the projections of the first and second eigenvector is shown in Fig. 7.13. Densities were estimated using a kernel density estimator.

While the canonical setup and the setup that lacks magnesium ions explore essentially the same configurational space, the setups that lack nucleotide modifications or nucleotide modifications and magnesium ions explore more configurations with higher projection values on the first eigenvector.

The lack of ions keeps the setup without nucleotide modifications and ions restraint to structures that have lower projection values on the second eigenvector while the setup that includes magnesium ions but lacks nucleotide modifications favors structures with higher projections on the second eigenvector.

Fig. 7.14 shows the motion the third eigenvector describes. The anticodon stem loop is moving outwards with the bases (WB)G_{37}, A_{36} and A_{35}. This is accompanied by flipping of U_{16}. The angle chosen was that one between the phosphate atoms of A_{35}, U_{47} and U_{54}. It is 158° for the smallest value of projection on the third eigenvector and 113.4° for the biggest one. The angle defined is 149.1° in the crystal structure.

A histogram of the projections on the third eigenvector is depicted in Fig. 7.15. A removal of nucleotide modifications leads to a broader distribution of projections on the third eigenvector. The removal of magnesium ions broadens the distribution and shifts it to higher values. A removal of both broadens the distribution even more and shifts it further to higher values.

Figure 7.11: Projection of structures to the extreme values of the second eigenvector gained by PCA analysis. The straight black and gray arrows denote axis of rotation, dark blue lines indicate the geometrical properties at maximal projection values while light blue lines denote geometrical properties at minimal projection values.

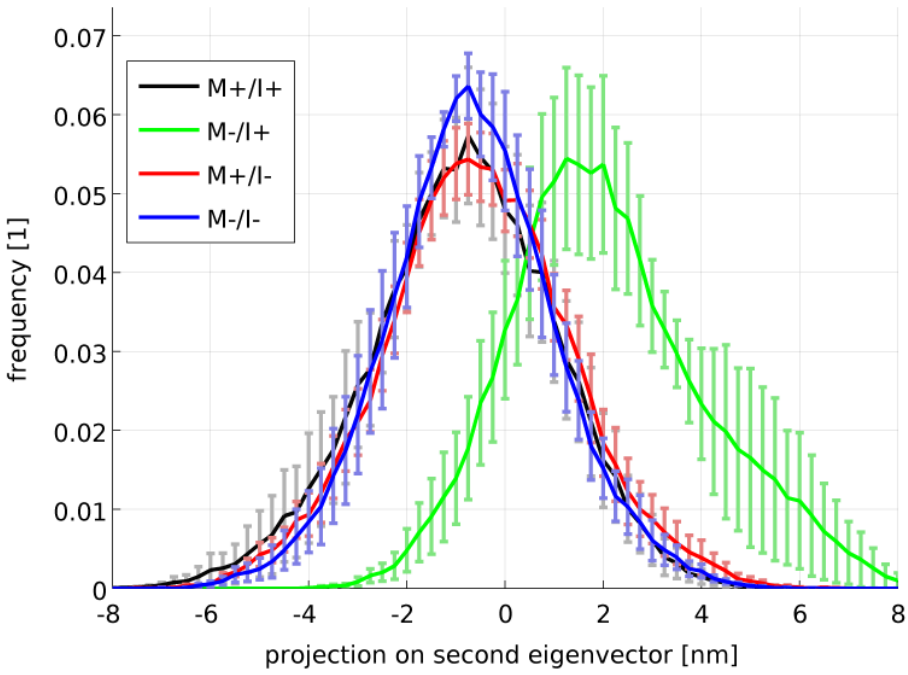

Figure 7.12: Histogram of projection of trajectory on the second eigenvector gained by PCA analysis.

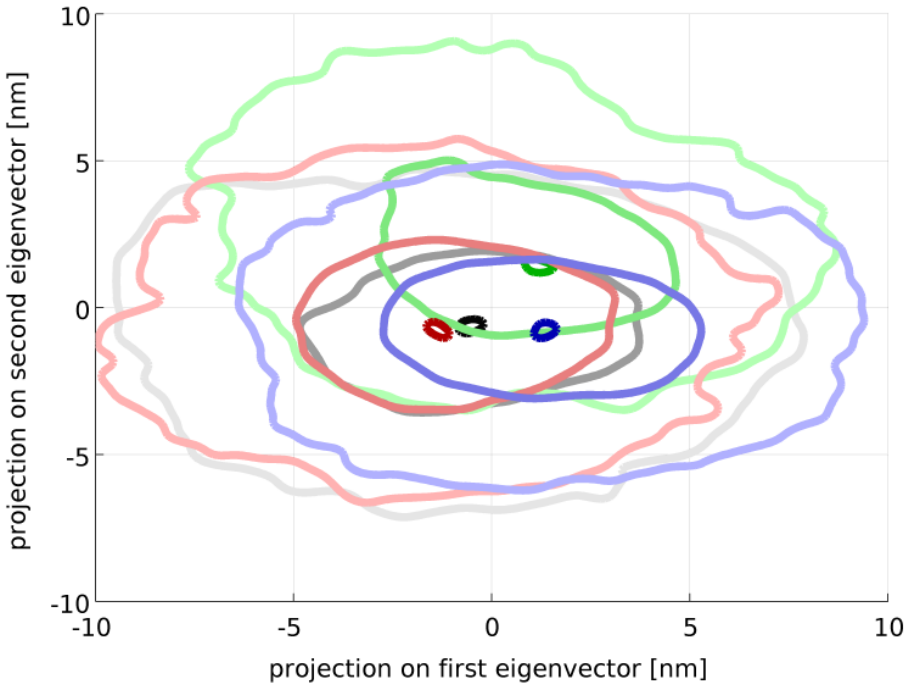

Figure 7.13: Estimated density from projection of trajectory on the first and second eigenvector gained by PCA analysis. Lines denote contours of constant density. Saturation is directly proportional to the density.

Entropy Estimate

The configurational entropy of the tRNA was estimated using the Schlitter formula [98]. For this a PCA of every single simulation was carried out. The entropy was then estimated for all of these. Fig. 7.16 shows the averaged results and their statistical errors. The canonical setup has the lowest entropy, followed by the setups lacking either nucleotide modifications or magnesium ions with a configurational entropy that is approximately four percent higher. With a seven percent increase the setup that lacks both, magnesium ions and nucleotide modifications has the highest entropy.

FCA

The FCA modes are very similar to the PCA modes in this case and describe essentially the same motions. The movement that is described by the third PCA eigenvector is described by the second FCA eigenvector and vice versa. For that reason FCA modes are not discussed here explicitely.

7.3.4 Base pairing

Occupancy numbers for base pairs were determined using the algorithm described above for every single trajectory after equilibration. Then they were averaged over all simulations with the same setup. By that average occupancies for base pairs were derived.

To check if secondary structure that was present in the crystal structure was preserved

Figure 7.14: Projection of structures to the extreme values of the third eigenvector gained by PCA analysis. The straight black arrow denotes the axis of rotation, dark blue lines indicate the geometrical properties at maximal projection values while light blue lines denote geometrical properties at minimal projection values.

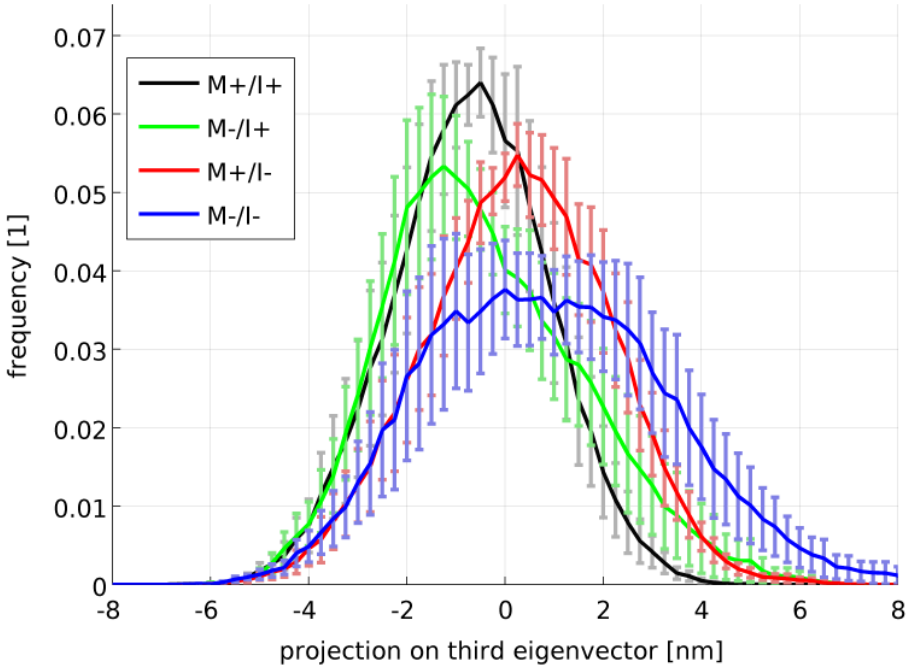

Figure 7.15: Histogram of projection of trajectory on the third eigenvector gained by PCA analysis.

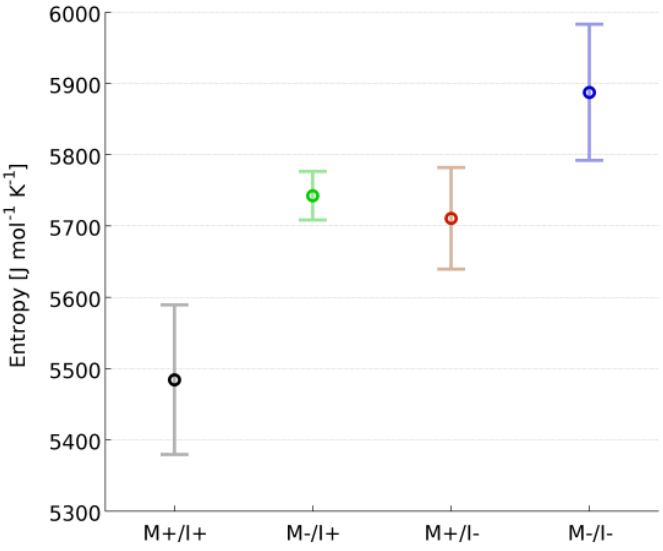

Figure 7.16: Configurational entropy as estimated by the Schlitter formula.

during the simulations the average occupancies were subtracted from the crystal structures occupancy numbers (which are all 100%). Differences bigger than 20 % in occupancy are shown in Fig. 7.17.

Different occupation numbers in the simulation than in the crystal structure are to be expected as occupation numbers for the crystal structure can only be 100% or 0 %. The transition from crystal to liquid phase and to higher temperatures is unfavorable for canonical base pairs which are therefore expected to be less present in simulation than in crystal structures. Most differences affect single hydrogen bonds, especially in the anticodon loop and in interactions of the variable loop with the D stem.

A significant change in base pairing occurs in the G_{18}-U_{55} the G_{19}-C_{56} and A_{58}-U_{54} base pairs in the ellbow region of tRNA where interactions resolved in the crystal structure become weaker by loosing single hydrogen bonds. At the begin of the anticodon stem the canonical Watson-Crick interaction between U_{33} and A_{36} is weakened. In the anticodon loop the interaction between U_{33} and A_{36} is stronger in the simulations than in the crystal structure, an extra hydrogen bond is formed.

Most secondary structure elements are maintained during the simulations.

A lack in magnesium ions leads to only few changes in base pairing. The G_{19}-C_{56} interaction in the ellbow region shows more canonical behavior. In the anticodon stem loop the hydrogen bonding network between C_{32}, G_{37}, U_{33} and A_{36} is broken.

For the setup that lacks nucleotide differences in base pair occupancy probabilities are shown in Fig. 7.19. Similar to the setup that lacks magnesium ions the hydrogen bonding interactions in the anticodon stem loop are weaker here between C_{32}, G_{37} and U_{33}. The hydrogen bonds between A_{38} and C_{32} are lost and the Watson-Crick-Watson-Crick edge interaction of A_31 with U_{39} in the anticodon stem is lost which might be seen as a first step for disintegration of the RNA helix in this place. In contrary to the setup lacking modifications the Watson-Crick-Hoogsteen-edge interaction between U_{32} and A_{32} is strengthend. The modification of U_{39} to P_{39} might be needed to stabilize the anticodon stem.

Figure 7.17: Differences in base pairing between simulation and crystal structure. Only differences bigger than 20 % in occupancy are shown. Blue symbols denote higher occupancy numbers in the crystal structure while red denotes higher occupancy numbers in the simulation average with canonical tRNA.

Figure 7.18: Differences in base pairing between simulations of canonical tRNA and tRNA lacking magnesium ions. Only differences bigger than 20 % in occupancy are shown. Blue symbols denote higher occupancy numbers in the canonical tRNA while red denotes higher occupancy numbers in the simulation without magnesium ions.

In the ellbow region the base pair interaction between G_{18} and U_{55} is almost lost. More canonical base pairing occurs between G_{19} and C_{56}.

G_{45} can establish more single hydrogen bonds with C_{11}, G_{24} and C_{25}. The Watson-Crick-Watson-Crick-edge interaction between G_{26} and A_{26} is strengthend.

In the TΨC arm an Watson-Crick-Hoogsteen-edge interaction between A_{58} and U_{54} is strengthend.

The difference in base pairing in the setup lacking nucleotide modification and magnesium to the canonical setup is shown in Fig. 7.20. In the ellbow region the interaction between G_{18} and U_{55} is weakened in favor of a single hydrogen bond. Additionally the Watson-Crick-Hoogsteen-edge interaction between U_{54} and A_{58} is stronger than in the canonical setup, similar to the setup lacking nucleotide modifications only. In the anticodon stem to the base pair C_{32}-A_{38} all interaction changes are similar to the setup lacking nucleotide modifications only. In the anticodon loop U_{33} base pairs less with G_{37} and A_{36}.

7.3.5 Orientations of bases U_{16} and A_{58}

U_{16} which is modified to dihydrouridine in canonical tRNA resides in 5Å distance to a magnesium ion (See Fig. 7.22 for a representation of that area). In PCA mode three it was seen flipping out and in. To see the impact of its modification and the magnesium ion in its close vicinity its local conformational change was monitored. This was done using the dihedral angle between the center of mass of the nucleotide and the center of mass of the sugar of subsequent base. The definition of this angle shown in Fig. 7.21 follows Banavali and MacKerell in their definition of the COM pseudo-dihedral angle but omits the base of the opposing strand of the helices used in their definition as in the region of U_{16} RNA is not in a helical conformation [99]. This angle will be called orientation angle. The orientation angle between the bases A_{16} and U_{16} is shown in Fig. 7.23. Fig. 7.24 shows the orientation angle between U_{16} and U_{17} respectively.

For the canonical setup the orientation angle between bases G_{15} and U_{16} is restricted to a range between -20° and 150° with a pronounced maximum at 10° and a local maximum at 120°. These to states are well separated. There is a favorable flipped in conformation and a much less favorable flipped out conformation.

With the nucleotide modification removed there are slightly more orientations that can be explored. The global maximum is shifted towards50° and the barrier between the flipped in and flipped out states vanishes.

When the magnesium ion is removed all orientations can be explored with low probabilities for angles between 50° and 100°.

A removal of nucleotide modifications and magnesium ions allows all orientational angles to be explored with a minimum between -180° and -100°.

The orientation angle between bases U_{16} and U_{17} are shown in Fig. 7.24. In the canonical setup these bases prefer a anti-parallel conformation with a very broad distribution of favorable states from 150° to -80° and second interval from 50°to 130°. When nucleotide modifications are removed the distribution of orientation angles becomes broader with unfavorable orientation angles only in the range from -80° to -140°. The removal of the magnesium ion makes the parallel conformation much more favorable than the antiparallel one. angles smaller than -100° and bigger than 100° become very unfavorable. The setup that lacks nucleotide modifications and magnesium ions allows both parallel and antiparallel conformations but makes orientational angles from 10° to 150° very unfavorable.

Figure 7.19: Differences in base pairing between simulations of canonical tRNA and tRNA lacking nucleotide modifications. Only differences bigger than 20 % in occupancy are shown. Blue symbols denote higher occupancy numbers in the canonical tRNA while red denotes higher occupancy numbers in the simulation without nucleotide modifications.

Figure 7.20: Differences in base pairing between simulations of canonical tRNA and tRNA lacking nucleotide modifications and magnesium ions. Only differences bigger than 20 % in occupancy are shown. Blue symbols denote higher occupancy numbers in the canonical tRNA while red denotes higher occupancy numbers in the simulation without magnesium ions and nucleotide modifications.

Figure 7.21: Definition of base orientation angle, generalized from Banavali and MacKerell [99].

Figure 7.22: Close up of area highly affected by nucleotide modifications and magnesium ions.

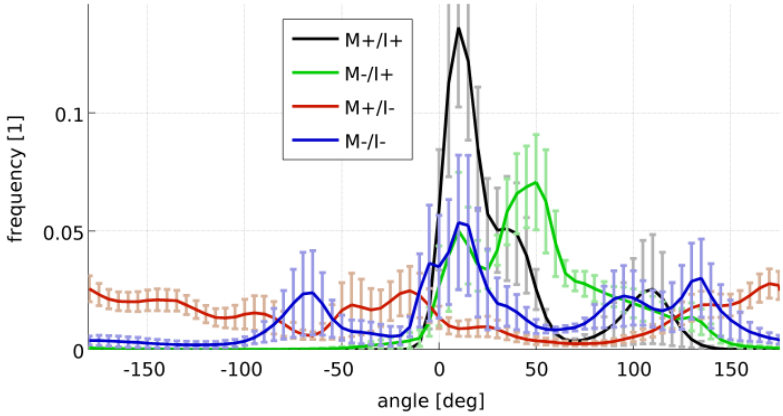

Figure 7.23: Orientation angle between bases 15 and 16 in tRNA.

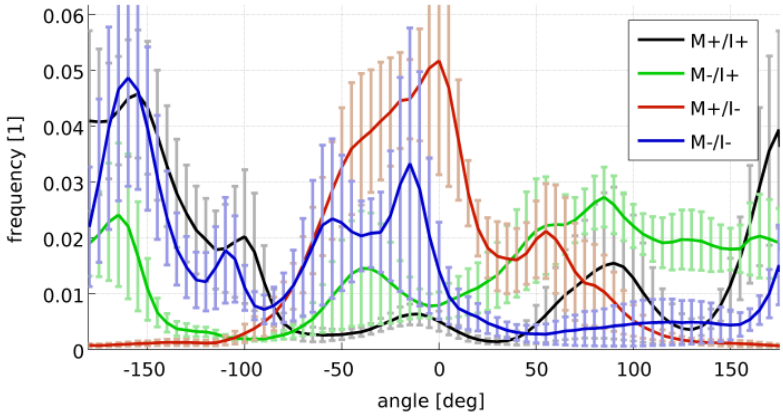

Figure 7.24: Orientation angle between bases 16 and 17 in tRNA.

Figure 7.25: Surrounding of base A58 who is subject to nucleotide modifications.

While in canonical tRNA the orientations of base U_{16} and U_{17} are well defined with U_{16} flipped in and U_{17} flipped out, the removal of magnesium ions leads to many more possible configurations for U_{16} and changes their behavior from anti-cooperative to cooperative.

The nucleotide modifications bases U_{16} and U_{17} lead to a confinement of their respective orientational angles. As compared with the setup lacking those modifications the favorable orientational angles are much clearer defined.

The setup lacking both, nucleotide modifications and magnesium ions leads to a much broader spectrum of favorable orientational angles, too.

Because of the vital role that has been reported on the nucleotide modification to A_{58} [76] the orientational angles of bases G_{57} and A_{58} was examined. Figs. 7.25 and 7.25 show the surrounding of A_{58}.

While the orientational angle distribution of bases A_{58} and U_{59} show little deviations for different setups (data not shown), the distribution of orientational angles of bases G_{57} and A_{58} shows differences for setups with the nucleotide modification removed.

Canonical tRNA or tRNA with magnesium ions removed all orientations between 0° and 40° occur with a broad maximum at 25°. A removal of the nucleotide modification leads to a narrower distribution of base orientations. Orientations from 0 to 10° are no longer sampled.

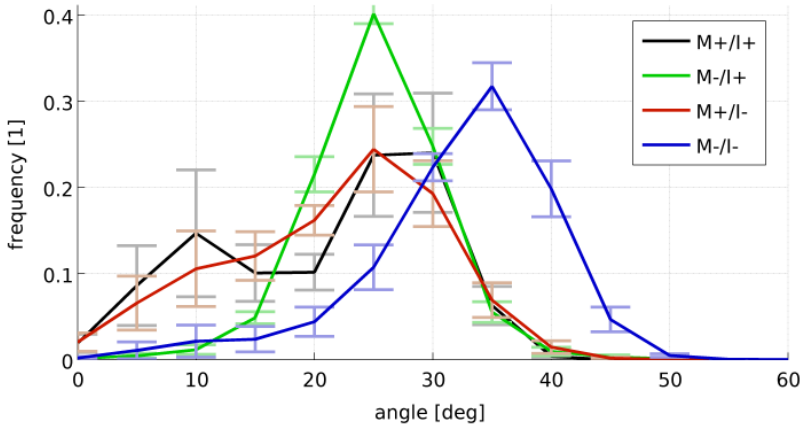

Figure 7.26: Dihedral angle between bases G57 and A58 in tRNA.

When ions are removed additionally the most favorable orientation is shifted to 35°.

The methylation of adenosine 58 leads to more possible configurations for A_{58}.

7.4 Discussion

7.4.1 Dynamics in liquid phase compared to the crystal phase

As the RMSD shows, it is possible to perform stable simulations of tRNA dynamics on the decananosecond timescale. Differences in the mean structures of the simulation in liquid to the crystal structure occur for various reasons. First the force field used is an approximative description of the system. Even though great progress in RNA simulation has been made in the last ten years by correct treatment of electrostatics and correction for ion parameters and dihedral angle interaction terms there are still differences to be expected from these approximations. Second the crystals are grown at temperatures of 100K in contrast to 300K at which the simulations were performed. This will largely reduce the flexibility of the whole molecule. The molecule might be trapped local energy minima that can be easily overcome at 300K. Third crystallization is performed under very high ion concentrations to stabilize the conformation of the molecule which might lead to a very different electrostatic surrounding than in the liquid phase. Fourth crystal structure close contacts to periodic images of the molecule induces steric barriers to certain conformations. These close contacts that might have hindered conformations in the crystal phase that are present in the simulation will now be discussed.

The CCA tail is in close contact with the variable loop and the anticodon stem of the tRNA, namely residues 42 to 45. On the other side it is in close contact with the anticodon of the tRNA YG37 and A36. The CCA tail is therefore trapped in the crystal structure.

Furthermore residues 67 to 71 are in vicinity to residues 51-53 and 64 of the TΨC loop.

G19 and C56 are close to the D arm of the tRNA. The guanosine at position 34 is in close contact with the phosphate backbone of A at position 62 of its symmetry related image. This induces base stacking interactions of the anticodon bases. Bigger opening angles between anticodon stem and acceptor stem are unfavorable in the crystal structure. The crystallographic symmetry favors angles that are close to 90° which might explain the deviation of the angles

found to the crystal structure. A comparison to crystal structures of similar molecules that have been derived under different crystallizing conditions might be useful.

7.4.2 Modes of motion

The dynamic behavior of tRNA relevant for its binding to perform its biological function. TRNA binds to aminoacyl-tRNA-syntheases, elongation factor TU, and the ribosome where it moves through the intersubunit space.

High resolution structural data for binding of tRNA binding to aaRSs is not available yet, though attempts have been made to determine binding sites by tRNA docking to aaRSs [100]. The picture of highly flexible aaRSs can be accompanied by the findings of this thesis by that of a very flexible tRNA.

TRNA in the A/T position that is still bound to EF-Tu reveals that the opening movement which is described by the first eigenvector of the PCA eigenvector as well as the kinking movement of the anticodon loop that is described by the third eigenvector are essential for its binding [101]. The same kinking movement is essential for P/E binding of tRNA to the ribosome [102].

The A-site accommodation of the tRNA that comprises the movement from A/T state of tRNA to the A/A state is a crucial step in protein synthesis. From the data acquired by Sanbonmatsu et al. it has become apparent that the kinking movement of tRNA is accompanied by a rotation of the anticodon arm as was seen exactly in the movement described by the second and third PCA mode [103].

The modelled A/A, P/P and E/E positions in the ribosome are very close to the crystal structure with respect to the modes of motions of tRNA [104]. As initial structures for crystallographic refinement the crystal structures of tRNA in solution are used. Taking into account the modes of movement determined in this thesis can help improve the quality of fits to crystallographic or cryo-EM data.

Magnesium ions constrains the second mode of motion of tRNA as revealed by PCA analysis while allowing for the first. As the second mode describes the kinking movement of the anticodon loop that is essential to tRNA accommodation and movement through the ribosome a lack of magnesium ions could thus lower the efficiency of protein synthesis.

The lack of nucleotide modifications has its biggest effect on the second PCA mode. A change in the movement described by this mode might again hinder movement of tRNA through the intersubunit space as the there preferred conformation with a shortened anticodon stem.

A biological reasoning that would explain the great flexibility that was observed in the opening angle between anticodon stem and acceptor stem was not found in this thesis. This flexibility could give rise to new binding modes to aaRSs or EF-TU.

7.4.3 Base pairing

Oliva et al. proposed that nucleotide modifications and Mg_{2+} binding play interchangeable roles in stabilizing the G_{15} C_{48} Levitt base pair in archeosines [105]. Yet for yeast $tRNA^{Phe}$ that was used in our simulations we could not observe significant alterations in this base pair but found its tertiary interactions rather unaltered throughout all setups. An analysis of this interaction regarding single hydrogen bonds rather than general tertiary interactions might yield further results.

Figure 7.27: Close contacts to other tRNA molecules in the crystal. Atoms in 4 Å range are shown as grey spheres, the residues belonging to them as grey lines. Where interactions might affect nucleobase orientation those are shown in green lines and cartoon representation, where the phosphate backbone is affected smudge green is used.

Nobles proposes that Ψ_{55} directly contributes to tertiary interactions between the D and the TΨC loop while not contributing to secondary interactions in the TΨC loop alone [106]. Our results have similar implications. The favorable tertiary interactions of P_{55} with G_{18} between TPC and D domain gets completely lost when modifying Ψ_{55} to U. Additionally he proposes a stabilizing role of the modification at U_{54} for the TPC arm. The effects in our simulations were too small to support or contradict this hypothesis.

The anticodon stem loop seems to have a sensitive hydrogen bonding pattern that can be easily disturbed by either nucleotide modifications or magnesium missing. The changes in base paring at the anticodon stem loop where the canonical Watson-Crick interaction is lost can lead to a destabilization of the helical fragments of the anticodon stem. This would explain the decrease in thermal stability that was observed in the anticodon stem loop when Ψ_{39} is replaced by U_{39} [53].

In both regions pseudouridine seems to play a major role in stabilizing tRNA structure. This is in good agreement with previous research on the role of pseudouridine [56]. A stable anticodon region may help proper tRNA binding on the ribosome and by that increase the translational accuracy [55].

7.4.4 Enthalpic or entropic stabilization of tRNA by magnesium ions or nucleotide modifications

The configurational entropy estimate by the Schlitter formula did yield the lowest configurational entropy for the canonical setup and increases of configurational entropy in all other cases.

To relate our simulations to melting experiments the free energy difference between a ordered and a disordered state is looked at. The free energy difference to a disordered state is

$$\Delta G = \Delta H - T\Delta S$$

where $\Delta H = H_{\text{native}} - H_{\text{disordered}}$ is the enthalpy difference between native and disordered state and $\Delta S = S_{\text{native}} - S_{\text{disordered}}$ is the entropy difference.

As it can be assumed that the enthalpy and entropies are almost equal for all setups in the disordered state, the difference of the free energy difference between the canonical tRNA and tRNA lacking nucleotide modifications, magnesium ions or both is

$$\Delta\Delta G = (H_{\text{canonical}} - H_{\text{other}}) - T(S_{\text{canonical}} - S_{\text{other}}).$$

As known from melting experiments $\Delta G_{\text{canonical}} < \Delta G_{\text{other}}$. With $S^{\text{Schlitter}}_{\text{canonical}} < S^{\text{Schlitter}}_{\text{other}}$ the estimate would yield $H_{\text{canonical}} < H_{\text{other}}$. This speaks for an enthalpic stabilization of tRNA by magnesium ions and nucleotide modifications.

However solvent entropy and interaction of tRNA with solvent might play a major part in stabilizing tRNAs structure. Methods to estimate these contributions are currently under development but could yet not be employed in this thesis.

7.4.5 The flipping of single bases

In all crystal structures of tRNA the bases U_{16} and U_{17} are flipped in opposite directions with U_{16} a orientation angle between bases 15 and 16 of 50°. This conformation is not favored

in the simulations but is close to the ensemble average. The simulation results would rather favor a two state model for the flipping of these bases.

The anticooperative behavior of the bases in the could be needed for accurate binding to aaRS or EF-TU or for defining the correct distance between the ellbow regions of A/A-site and P/P-site tRNAs in the ribosome as these bases from tRNA in the A/A-site are in close contact(3 Å distance) to U_{47} of the P/P-site tRNA. The lack of magnesium ions could effect the movement of tRNA through the ribosome by changing the orientation of bases 16 and 17 of $tRNA^{Phe}$. The effect of nucleotide modifications seems to be not so dominant in this case. Still their impact on the orientation of bases 16 and 17 might lower rates for the incorporation of new peptides into the peptide chain.

The effect of nucleotide modifications on the orientation of A_{58} could be the cause for conformational rearrangements in the ellbow region of tRNA. A different conformation in the ellbow region will lead to a different conformation of the whole tRNA as the ellbow region defines how anticodon stem and acceptor stem are positioned with respect to each other. To be able to further estimate the effect of this very modification simulations with only that modification removed will be useful.

7.4.6 Further studies

To study the effect of outer sphere binding of magnesium ions simulations with different concentrations of magnesium ions might be useful. Free energy calculations could estimate binding energies of magnesium ions that have been predicted experimentally and identify binding sites and modes.

To distinguish the effect of the single nucleotide modifications simulations with only single modifications being present or absent can give better insight into their specific role. Free energy calculations could give hints on their stabilizing or destabilizing effects on tRNA.

Solvent-RNA interactions are another open field for studies. The analysis of the water shell around magnesium ions and nucleotide modifications might give an atomistic explanation of the way they function. Hydrogen bonding patterns and energies of solvent-RNA interactions as well as estimates of solvent entropy terms will give better insight into the thermodynamic stabilization of tRNA by ions and nucleotide modifications.

As certain nucleobases become protonated or deprotonated simulations with a constant pH-protocoll could give better results [54]. Furthermore they could reveal the effect of a change in pH on tRNA structure.

Part IV

Concluding remarks

8 Conclusions and outlook

In this thesis three factors that influence RNA structure and flexibility were analyzed. This factors were the change of temperature, analyzed on an RNA double strand, and a change in the presence of magnesium ions and nucleotide modifications on tRNA.

Significant changes on an atomistic level as well on their general dynamics were observed for all three factors.

The double stranded RNA simulations provided a reference frame for further simulations on RNA. The knowledge already gained from the canonical A-from helix could help bioengineering techniques, for example the building of pre-stressed tensegrity structures [107].

The simulations of tRNA in solution could be related to its biological function in the course of this thesis. By that possible implications of nucleotide modifications and magnesium ions in tRNA on its biological function could be found. Steps could be made towards a more complete understanding of tRNA biology which in turn might prove useful for the development of new antibiotics that might act on tRNA modifying enzymes or in the understanding of the evolution of tRNA and its modifications [108, 109].

An even better connection between atomistic pictures revealed by in silico and in vitro experiments will be made possible by the current development of new experimental techniques that provide high resolution data in time and space as in NMR, cryo electron microscopy or FRET experiments on the one hand and increasing computer powers and enhanced analysis and simulation tools for in silico experiments on the other hand. A combination of both kinds of experiments will be a scope of future work.

9 Acknowledgements

I would like to thank my supervisor Prof. Helmut Grubmüller for ever encouraging discussions and Gerrit Groenhof for co-supervision. Furthermore I'd like to thank (in alphabetical order) Lars Bock for accompanying me through my whole time as diploma student and proofreading, Stephanus Fengler for thorough discussions on theory related topics, Dirk Matthes for proofreading and hints on intersting papers, Lars Schaefer for helping me to get started, Martin Stumpe for his superb qualities as office mate and Martin Vesper for proofreading, valuable discussions and entertainment.

Bibliography

[1] H. M Al-Hashimi. Dynamics-based amplification of RNA function and its characterization by using Nmr spectroscopy, *ChemBioChem*, **6**(9), 1506–1519, 2005.

[2] R. J Maraia, N. H Blewett, and M. A Bayfield. It's a mod mod tRNA world, *Nat. Chem. Biol.*, **4**(3), 162–164, 2008.

[3] M Sprinzl, C Horn, M Brown, A Ioudovitch, and S Steinberg. Compilation of tRNA sequences and sequences of tRNA genes, *Nucleic Acids Research*, **26**(1), 148–153, 1998.

[4] P. F Agris. The importance of being modified: Roles of modified nucleosides and Mg^{2+} in RNA structure and function, *Progress in neuro-psychopharmacology & biological psychiatry*, **53**, 79–129, 1996.

[5] E. J Maglott, S. S Deo, A Przykorska, and G. D Glick. Conformational transitions of an unmodified tRNA: Implications for RNA folding, *Biochemistry*, **37**(46), 16349–16359, 1998.

[6] F Crick. *The biological replication of macromolecules.* In *Symp. Soc. Exp. Biol*, volume 12, pages 138–163, 1958.

[7] F Crick. Central Dogma Of Molecular Biology, *Nature*, **227**(5258), 561–, 1970.

[8] T. R Cech. Structural biology - The ribosome is a ribozyme, *Science*, **289**(5481), 878–879, 2000.

[9] J Perona. Sructural Basis For Transfer-RNA Aminoacylation By Escherichia-Coli Glutaminyl-transfer RNA-Synthease, *Biochemistry*, **32**(34), 8758–8771, 1993.

[10] K Hoogsteen. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine, *Acta Crystallographica*, **16**(9), 907–916, Sep 1963.

[11] W Traub and D Elson. RNA Composition And Base Pairing, *Science*, **153**(3732), 178–, 1966.

[12] W. E van Gunsteren and H. J. C Berendsen. Computer Simulation Of Molecular Dynamics Methodology Applications and Perspectives In Chemistry, *Angewandte Chemie. International edition in English*, **29**(9), 992–1023, 1990.

[13] N Metropolis. The Monte Carlo Method, *J. Am. Stat. Assoc.*, **44**(247), 335–341, 1949.

[14] B. J Alder and T. E Wainwright. Studies in Molecular Dynamics. I. General Method, *The Journal of Chemical Physics*, **31**(2), 459–466, 1959.

[15] P Auffinger and E Westhof. RNA solvation: A molecular dynamics simulation perspective, *Biopolymers*, **56**(4), 266–274, 2000.

[16] J Norberg and L Nilsson. Molecular dynamics applied to nucleic acids, *Acc. Chem. Res.*, **35**(6), 465–472, 2002.

[17] T Darden, D York, and L Pedersen. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems, *J. Chem. Phys.*, **98**, 10089–10092, 1993.

[18] F. C Bernstein, T. F Koetzle, G. J. B Williams, E. F Meyer, M. D Brice, J. R Rodgers, O Kennard, T Shimanouchi, and M Tasumi. Protein Data Bank - Computer-Based Archival File For Macromolecular Structures, *J. Mol. Biol.*, **112**(3), 535–542, 1977.

[19] H. M Berman, W. K Olson, D. L Beveridge, J Westbrook, A Gelbin, T Demeny, S.-H Hsieh, A. R Srinivasan, and B Schneider. The Nucleic-Acid Database - A Comprehensive Relational Database of 3-dimensional Structures of Nucleic-Acids, *Biophys. J.*, **63**(3), 751–759, 1992.

[20] H. J. C Berendsen, J. P. M Postma, W. F van Gunsteren, A DiNola, and J. R Haak. Molecular dynamics with coupling to an external bath, *J. Chem. Phys.*, **81**(8), 3684–3690, October 1984.

[21] S Nose and M. L Klein. Constant Pressure Molecular-dynamics for Molecular-systems, *Mol. Phys.*, **50**(5), 1055–1076, 1983.

[22] W Hoover. Canonical Dynamics - Equilibrium Phase-Space Distributions, *Physical review. A, Atomic, molecular, and optical physics*, **31**(3), 1695–1697, 1985.

[23] G Voronoi. Nouvelles applications des paramètres continus à la théorie des formes quadratiques, *Journal für die Reine und Angewandte Mathematik*, **133**, 97–178, 1907.

[24] I Semaev. A 3-dimensional lattice reduction algorithm, *Lecture notes in computer science*, **2146**, 181–193, 2001.

[25] W DeLano. *The PyMOL Molecular Graphics System*, 2002.

[26] B. N Delaunay. Sur la sphère vide, *Bulletin of Academy of Sciences of the Ussr 7*, **6**, 793–800, 1934.

[27] D Watson. Computing the n-dimensional Delaunay tessellation with application to Voronoi polytopes, *The computer journal*, **24**(2), 167–172, 1981.

[28] K Pearson. On lines and planes of closest fit to systems of points in space., *Philos. Mag.*, **2**(7-12), 559–572, 1901.

[29] O. F Lange and H Grubmueller. Full correlation analysis of conformational protein dynamics, *Proteins*, **70**(4), 1294–1312, 2008.

[30] H Yang, F Jossinet, N Leontis, L Chen, J Westbrook, H Berman, and E Westhof. Tools for the automatic identification and classification of RNA base pairs, *Nucleic Acids Res.*, **31**(13), 3450–3460, 2003.

[31] O Kennard, F Allen, S Bellard, M Brice, B Cartwright, A Doubleday, H Higgs, T Hummelink, B Hummelink-Peters, W Motherwell, J Rodgers, and D Watson. The Cambridge crystallographic data centre: computer-based search, retrieval, analysis and display of information, *Acta Crystallographica, Section B (Structural Crystallography and Crystal Chemistry)*, **B35**, 2331–9, 10 1979.

[32] (John von Neumann Institute of Computing.). volume 10. pages 423–445. W Janke. 2002.

[33] F. J Grundy and T. M Henkin. Regulation of gene expression by effectors that bind to Rna, *Current opinion in microbiology*, **7**(2), 126–131, 2004.

[34] M Getz, X Sun, A Casiano-Negron, Q Zhang, and H. M Al-Hashimi. NMR studies of RNA dynamics and structural plasticity using Nmr residual dipolar couplings, *Biopolymers*, **86**(5-6), 384–402, 2007.

[35] M. M Yusupov, G. Z Yusupova, A Baucom, K Lieberman, T. N Earnest, J. H. D Cate, and H. F Noller. Crystal structure of the ribosome at 5.5 angstrom resolution, *Science*, **292**(5518), 883–896, 2001.

[36] M Selmer, C. M Dunham, F. V Murphy, A Weixlbaumer, S Petry, A. C Kelley, J. R Weir, and V Ramakrishnan. Structure of the 70S ribosome complexed with mRNA and tRNA, *Science*, **313**(5795), 1935–1942, 2006.

[37] A Korostelev, S Trakhanov, M Laurberg, and H. F Noller. Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements, *Cell*, **126**(6), 1065–1077, 2006.

[38] A Korostelev and H. F Noller. The ribosome in focus: new structures bring new insights, *Trends Biochem. Sci.*, **32**(9), 434–441, 2007.

[39] K Reblova, F Lankas, F Razga, M. V Krasovska, J Koca, and J Sponer. Structure, dynamics, and elasticity of free 16S rRNA helix 44 studied by molecular dynamics simulations, *Biopolymers*, **82**(5), 504–520, 2006.

[40] F Razga, J Koca, A Mokdad, and J Sponer. Elastic properties of ribosomal RNA building blocks: molecular dynamics of the Gtpase-associated center rRNA, *Nucleic Acids Res.*, **35**(12), 4007–4017, 2007.

[41] P. S Klosterman, S. A Shah, and T. A Steitz. Crystal structures of two plasmid copy control related RNA duplexes: An 18 base pair duplex at 1.20 A resolution and a 19 base pair duplex at 1.55 A resolution, *Biochemistry*, **38**(45), 14784–14792, 1999.

[42] W. D Cornell, P Cieplak, C. I Bayly, I. R Gould, J Kenneth M. Merz, D. M Ferguson, D. C Spellmeyer, T Fox, J. W Caldwell, and P. A Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995), *J. Am. Chem. Soc.*, **118**(9), 2309–2309, 1996.

[43] W. L Jorgensen, J Chandrasekhar, J. D Madura, R. W Impey, and M. L Klein. Comparison of Simple Potential Functions for Simulating Liquid Water., *J. Chem. Phys.*, **79**(2), 926–935, 1983.

[44] L Stella. Equilibration and sampling in molecular dynamics simulations of biomolecules, *The Journal of chemical physics*, **109**(23), 10115–10117, 1998.

[45] E Espinosa, E Mollins, and C Lecomte. Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities, *Chem. Phys. Lett.*, **285**(3-4), 170–173, 1998.

[46] P Kebbekus. Persistence Length Of Rna, *Biochemistry*, **34**(13), 4354–4357, 1995.

[47] J Liphardt. Reversible unfolding of single RNA molecules by mechanical force, *Science*, **292**(5517), 733–737, 2001.

[48] R Giegé. Toward a more complete view of tRNA biology, *Nature structural & molecular biology*, **15**(10), 1007–1014, 2008.

[49] R. W Holley, J Apgar, G. A Everett, J. T Madison, M Marquisee, S. H Merrill, J. R Penswick, and A Zamir. Structure of a Ribonucleic Acid, *Science*, **147**(3664), 1462–, 1965.

[50] S. H Kim, F. L Suddath, G. J Quigley, A McPherson, J. L Sussman, A. H. J Wang, N. C Seeman, and A Rich. 3-Dimensional Tertiary Structure of Yeast Phenylalanine Transfer-RNA, *Science*, **185**(4149), 435–440, 1974.

[51] A Hopper. tRNA transfers to the limelight, *Genes & development*, **17**(2), 162–180, 2003.

[52] R. D Hotchkiss. The Quantitative Separation Of Purines, Pyrimidines, And Nucleosides BY Paper Chromatography, *The Journal of biological chemistry*, **175**(1), 315–332, 1948.

[53] C Yarian, M Basti, R Cain, G Ansari, R Guenther, E Sochacka, G Czerwinska, A Malkiewicz, and P Agris. Structural and functional roles of the N1-and N3-protons of Psi at tRNA's position 39, *Nucleic Acids Res.*, **27**(17), 3543–3549, 1999.

[54] P. C Durant and D. R Davis. Stabilization of the anticodon stem-loop of tRNA(LyS,3) by an A(+)-C base-pair and by pseudouridine, *J. Mol. Biol.*, **285**(1), 115–131, 1999.

[55] M Charette and M. W Gray. Pseudouridine in Rna: What, where, how, and why, *IUBMB Life*, **49**(5), 341–351, 2000.

[56] (ASM Press.). chapter 6, pages 103–112. P Auffinger and E Westhof. 1998.

[57] J. J Dalluge, T Hashizume, A. E Sopchik, J. A McCloskey, and D. R Davis. Conformational flexibility in Rna: The role of dihydrouridine, *Nucleic Acids Res.*, **24**(6), 1073–1079, 1996.

[58] E Kierzek and R Kierzek. The thermodynamic stability of RNA duplexes and hairpins containing N-6-alkyladenosines and 2-methylthio-N-6-alkyladenosines, *Nucleic Acids Res.*, **31**(15), 4472–4480, 2003.

[59] S Steinberg and R Cedergren. A correlation between N-2-dimethylguanosine presence and alternate tRNA conformers, *RNA*, **1**(9), 886–891, 1995.

[60] R Micura, W Pils, C Höbartner, K Grubmayr, M.-O Ebert, and B Jaun. Methylation of the nucleobases in RNA oligonucleotides mediates duplex-hairpin conversion, *Nucleic Acids Res.*, **29**(19), 3997–4005, 2001.

[61] M Helm, R Gieg, and C Florentz. A Watson-Crick base-pair-disrupting methyl group (m(1)A9) is sufficient for cloverleaf folding of human mitochondrial tRNA(Lys), *Biochemistry*, **38**(40), 13338–13346, 1999.

[62] G Kawai, Y Yamamoto, T Kamimura, T Masegi, M Sekine, T Hata, T Iimori, T Watanabe, T Miyazawa, and S Yokoyama. Conformational Rigidity Of Specific Pyrimidine Residues In Transfer-RNA ARISES From Posttranscriptional Modifications That Enhance Steric Interaction Between The Base And The 2'-Hydroxyl Group, *Biochemistry*, **31**(4), 1040–1046, 1992.

[63] J. A Kowalak, J. J Dalluge, J. A McCloskey, and K. O Stetter. The Role of Posttranscriptional Modification in Stabilization of transfer-RNA from Hyperthermophiles, *Biochemistry*, **33**(25), 7869–7876, 1994.

[64] P. F Agris. Decoding the genome: a modified view, *Nucleic Acids Res.*, **32**(1), 223–238, 2004.

[65] S Goto-Ito, T Ito, R Ishii, Y Muto, Y Bessho, and S Yokoyama. Crystal Structure of archaeal tRNA(m1G37)methyltransferase aTrm5, *Proteins*, **72**(4), 1274–1289, 2008.

[66] A Konevega, N Soboleva, V Makhno, Y Semenkov, W Wintermeyer, M Rodina, and V Katunin. Purine bases at position 37 of tRNA stabilize codon-anticodon interaction in the ribosomal A site by stacking and Mg2+-dependent interactions, *RNA-A Publication Of The RNA Society*, **10**(1), 90–101, JAN 2004.

[67] R Ishitani, S Yokoyama, and O Nureki. Structure, dynamics, and function of RNA modification enzymes, *Curr. Opin. Struct. Biol.*, **18**(3), 330–339, 2008.

[68] A. C Bishop, J Xu, R. C Johnson, P Schimmel, and V de Crecy-Lagard. Identification of the tRNA-dihydrouridine synthase family, *The Journal of biological chemistry*, **277**(28), 25090–25095, 2002.

[69] F Xing, M. R Martzen, and E. M Phizicky. A conserved family of Saccharomyces cerevisiae synthases effects dihydrouridine modification of tRNA, *RNA*, **8**(3), 370–381, 2002.

[70] M. J Johansson and A. S Byström. Dual function of the tRNA(m(5)U(54))methyltransferase in tRNA maturation, *RNA*, **8**(3), 324–335, 2002.

[71] L Pintard, F Lecointe, J M.Bujnicki, C Bonnerot, H Grosjean, and B Lapeyre. Trm7p catalyses the formation of two 2 '-O-methylriboses in yeast tRNA anticodon loop, *The Embo journal*, **21**(7), 1811–1820, 2002.

[72] G R.Bjoerk, K Jacobsson, K Nilsson, M J.O.Johansson, A S.Bystroem, and O P.Persson. A primordial tRNA modification required for the evolution of life?, *The Embo journal*, **20**(1-2), 231–239, 2001.

[73] F Lecointe, G Simos, A Sauer, E. C Hurt, Y Motorin, and H Grosjean. Characterization of yeast protein Deg1 as pseudouridine synthase (Pus3) catalyzing the formation of Psi(38) and Psi(39) in tRNA anticodon loop, *The Journal of biological chemistry*, **273**(3), 1316–1323, 1998.

[74] Y Motorin and H Grosjean. Multisite-specific tRNA : m(5)C-methyltransferase (Trm4) in yeast Saccharomyces cerevisiae: Identification of the gene and substrate specificity of the enzyme, *RNA*, **5**(8), 1105–1118, 1999.

[75] A Alexandrov, E. J Grayhack, and E. M Phizicky. tRNA m7G methyltransferase Trm8p/Trm82p: Evidence linking activity to a growth phenotype and implicating Trm82p in maintaining levels of active Trm8p, *RNA*, **11**(5), 821–830, 2005.

[76] J Anderson, L Phan, and A. G Hinnebusch. The Gcd10p/Gcd14p complex is the essential two-subunit tRNA(1-methyladenosine) methyltransferase of Saccharomyces cerevisiae, *Proc. Natl. Acad. Sci. U. S. A.*, **97**(10), 5173–5178, 2000.

[77] M Helm. Post-transcriptional nucleotide modification and alternative folding of Rna, *Nucleic Acids Res.*, **34**(2), 721–733, 2006.

[78] D. R Engelke and A. K Hopper. Modified view of tRNA: Stability amid sequence diversity, *Mol. Cell*, **21**(2), 144–145, 2006.

[79] M Sundaram, P. C Durant, and D. R Davis. Hypermodified nucleosides in the anticodon of tRNA(Lys) stabilize a canonical U-turn structure, *Biochemistry*, **39**(41), 12575–12584, OCT 17 2000.

[80] J. W Stuart, K. M Koshlap, R Guenther, and P. F Agris. Naturally-occurring modification restricts the anticodon domain conformational space of tRNA(Phe), *J. Mol. Biol.*, **334**(5), 901–918, 2003.

[81] E. V Puglisi and J. D Puglisi. Probing the conformation of human tRNA(3)(Lys) in solution by Nmr, *FEBS Lett.*, **581**(27), 5307–5314, NOV 13 2007.

[82] M Barciszewska, G Rapp, C Betzel, V Erdmann, and J Barciszewski. Structural changes of tRNA and 5S rRNA induced with magnesium and visualized with synchrotron mediated hydroxyl radical cleavage, *Mol. Biol. Rep.*, **28**(2), 103–110, 2001.

[83] T Lindahl, A Adams, and J. R Fresco. Rrenaturation of transfer ribonucleic acids through site binding of magnesium, *Proceedings of the National Academy of Sciences of the United States of America*, **55**(4), 941, 1966.

[84] J Fresco. Tertiary structure in transfer ribonucleic acis, *Cold Spring Harbor symposia on quantitative biology*, **31**, 527, 1966.

[85] Z Tan. Salt dependence of nucleic acid hairpin stability, *Biophys. J.*, **95**(2), 738–752, 2008.

[86] D. E Draper. A guide to ions and RNA structure, *RNA*, **10**(3), 335–343, 2004.

[87] V. K Misra and D. E Draper. Mg2+ binding to tRNA revisited: The nonlinear Poisson-Boltzmann model, *Journal of molecular biology*, **299**(3), 813–825, 2000.

[88] V Serebrov, R Clarke, H Gross, and L Kisselev. Mg2+-induced tRNA folding, *Biochemistry*, **40**(22), 6688–6698, JUN 5 2001.

[89] P Auffinger, S Louise-May, and E Westhof. Multiple molecular-dynamics simulations of the anticodon loop of tRNA(ASP) in aqueous-solution with counterions, *Journal of the American Chemical Society*, **117**(25), 6720–6726, 1995.

[90] P Auffinger and E Westhof. H-bond stability in the tRNA(Asp) anticodon hairpin: 3 ns of multiple molecular dynamics simulations, *Biophys. J.*, **71**(2), 940–954, 1996.

[91] A Lahiri and L Nilsson. Molecular dynamics of the anticodon domain of yeast tRNA(Phe): Codon-anticodon interaction, *Biophysical Journal*, **79**(5), 2276–2289, 2000.

[92] L Jovine, S Djordjevic, and D Rhodes. The crystal structure of yeast phenylalanine tRNA at 2.0 angstrom resolution: Cleavage by Mg2+ in 15-year old crystals, *J. Mol. Biol.*, **301**(2), 401–414, 2000.

[93] Y Terui, M Ohnuma, K Hiraga, E Kawashima, and T Oshima. Stabilization of nucleic acids by unusual polyamines produced by an extreme thermophile, Thermus thermophilus, *The Biochemical journal*, **388**, 427–433, 2005.

[94] Y Hashem and P Auffinger. A short guide for molecular dynamics simulations of RNA systems, *Methods*, **47**(3), 187–197, 2009.

[95] R Aduri, B. T Psciuk, P Saro, H Taniga, B Schlegel, and J SantaLucia. AMBER force field parameters for the naturally occurring modified nucleosides in Rna, *J. Chem. Theory Comput.*, **3**(4), 1464–1475, 2007.

[96] J SantaLucia. *modifieds database server http://ozone3.chem.wayne.edu/*, 2007.

[97] L. X Dang. MECHANISM AND THERMODYNAMICS OF ION SELECTIVITY IN AQUEOUS-SOLUTIONS OF 18-CROWN-6 ETHER - A MOLECULAR-DYNAMICS STUDY, *Journal of the American Chemical Society*, **117**(26), 6954–6960, 1995.

[98] J Schlitter. Estimation Of Absolute And Relative Entropies Of Macromolecules Using THE Covariance-MATRIX, *Chem. Phys. Lett.*, **215**(6), 617–621, 1993.

[99] N Banavali. Free energy and structural pathways of base flipping in a DNA GCGC containing sequence, *Journal of molecular biology*, **319**(1), 141–160, 2002.

[100] L Klipcan, I Levin, N Kessler, N Moor, I Finarov, and M Safro. The tRNA-induced conformational activation of human mitochondrial phenylalanyl-tRNA synthetase, *Structure*, **16**(7), 1095–1104, 2008.

[101] W Li, X Agirrezabala, J Lei, L Bouakaz, J. L Brunelle, R. F Ortiz-Meoz, R Green, S Sanyal, M Ehrenberg, and J Frank. Recognition of aminoacyl-tRNA: a common molecular mechanism revealed by cryo-EM, *EMBO Journal*, **27**(24), 3322–3331, 2008.

[102] W Li and J Frank. Transfer RNA in the hybrid P/E state: Correlating molecular dynamics simulations with cryo-EM data, *Proc. Natl. Acad. Sci. U. S. A.*, **104**(42), 16540–16545, 2007.

[103] K Sanbonmatsu, S Joseph, and C Tung. Simulating movement of tRNA into the ribosome during decoding, *Proceedings of The National Academy of Sciences of the United States*, **102**(44), 15854–15859, Nov 1 2005.

[104] G Yusupova, L Jenner, B Rees, D Moras, and M Yusupov. Structural basis for messenger RNA movement on the ribosome, *Nature*, **444**(7117), 391–394, 2006.

[105] R Oliva, A Tramontano, and L Cavallo. Mg2+ binding and archaeosine modification stabilize the G15-C48 Levitt base pair in tRNAs, *RNA*, **13**(9), 1427–1436, 2007.

[106] K Nobles, C Yarian, G Liu, R Guenther, and P Agris. Highly conserved modified nucleosides influence Mg2+-dependent tRNA folding, *Nucleic Acids Res.*, **30**(21), 4751–4760, NOV 1 2002.

[107] T Liedl, D. E Ingber, and W. M Shih. Pre-stressed Tensegrity Structures built from DNA, *Biophysical Journal*, **96**(3, Supplement 1), 20a – 20a, 2009.

[108] F Wei, S Li, and H. R Ma. Computer simulation of tRNA evolution, *Journal of Physics A: Mathematical and Theoretical*, **42**(34), 345101, 2009.

[109] M. E Saks, J. R Sampson, and J Abelson. Evolution of a transfer RNA gene through a point mutation in the anticodon, *Science*, **279**(5357), 1665–1670, 1998.