

Exploring RNA Structures by Means of
Molecular Dynamics Simulations:
– The Flavin Mononucleotide Aptamer, Coaxial Stacking in a
Nicked Hairpin and Water-Mediated Base Pairs in a Duplex –

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem Rat der Biologisch-Pharmazeutischen Fakultät der
Friedrich-Schiller-Universität Jena.

von Dipl.-Chem. Christoph Schneider
geboren am 19. September 1968 in Kamp-Lintfort

Referees

- Prof. Dr. Stephan Diekmann
- Prof. Dr. Hartmut Fritzsche
- Prof. Dr. Heinz Sklenar

Day of Rigorosum: 20th of December 2000

Day of Public Defense: 26th of February 2001

Dedicated to my father

Never say, "I tried it once and it did not work."
(Lord Ernest Rutherford, 1877-1957)

Acknowledgements

I would like to thank

- ☞ Dr. Jürgen Sühnel. He gave me the opportunity to work in his group and supplied me with this thesis. He was always on hand when I needed his advice. A lot of constructive discussions have been incorporated into this thesis.
- ☞ Prof. Dr. Stephan Diekmann for support in finishing this thesis.
- ☞ the referees for their expert opinion.
- ☞ our Biocomputing Group members for the nice working atmosphere.
- ☞ Fritz Haubensack for any computational hard- and software (e.g. \LaTeX) support. His 'service' was always quick and professional.
- ☞ Dr. Gerhard Müller for his help in any computational concerns. And of course for the barbeques.
- ☞ Drs. Martin Zacharias, Felipe Pineda and Oliver Ohlenschläger for many fruitful discussions in the field of computational biophysics.
- ☞ my room mate Jan Reichert for his patience and endurance with me.
- ☞ Peter Slickers. He introduced me in the field of MD on nucleic acids and supplied me with many useful hints.
- ☞ my parents.
- ☞ my triathlon friends from Jena. I was very lucky to meet you guys. It was a wonderful time. In alphabetic order: Dominik Driesch (DD), Timo Jahn, Carsten Merz (Odlo), Clemens Rolf (Klister Clemente), Dr. Gordon Charles Ross.
- ☞ the Steinheit connection.
- ☞ my beloved friends. In chaotic order: Britta und Stefan (Straelen), Nathalie and Gonzalo (Stuttgart), Jörg (Berlin, Frankfurt, Paris, London, ...), Roger (Frankfurt), Reinhard (Stuttgart), Markus (Markdorf), Tami, Niko and Emil (Stuttgart), Karin (Stuttgart), Scheibemann (Genève), Sebastian (Boston).

Contents

1	Introduction	1
1.1	Biological role of RNA	1
1.2	Structural information on RNA	2
1.3	Molecular dynamics (MD) simulations	3
1.4	The Flavin mononucleotide (FMN) RNA aptamer	4
1.5	Coaxial stacking in a nicked RNA hairpin	5
1.6	Water-mediated uracil-cytosine (WUC) base pairs in an RNA duplex	7
1.7	Aim of the work	9
2	Basic principles of RNA structure	10
3	Computational details of MD simulations	14
3.1	General aspects	14
3.2	Technique and approaches	14
3.2.1	Molecular model and force field	14
3.2.2	Long range forces	15
3.2.3	Sampling of the conformational space	16
3.2.4	Boundary conditions	16
3.2.5	Algorithms for MD simulations	17
3.2.6	Distance constraints	17
3.2.7	Solvent and ions	17
3.2.8	Heating up and equilibration	18
3.3	Problems	18
3.3.1	System size	18
3.3.2	Accuracy	18
3.4	Systems studied in this work	19
3.4.1	The FMN-RNA aptamer	19
3.4.2	Coaxial stacking in a nicked RNA hairpin	20
3.4.3	WUC base pairs in an RNA duplex	21

4	Results and Discussion	22
4.1	The FMN-RNA aptamer	22
4.1.1	Standard MD analysis and global structural features	22
4.1.2	Backbone torsional angles	26
4.1.3	FMN and its environment	29
4.1.4	The UUCG loop	30
4.1.5	Orientation of the ribose 2'-OH group	31
4.1.6	Structural water	35
4.1.7	Backbone C-H \cdots O interactions	37
4.2	Coaxial stacking of a nicked RNA hairpin	39
4.2.1	Stability and flexibility	39
4.2.2	Helical parameters	41
4.2.3	Comparison of simulations with different counterions	46
4.3	WUC base pairs in an RNA duplex	52
4.3.1	Dynamical features of WUC	52
4.3.2	Analysis of the water dynamics	57
5	Summary and outlook	60
6	Supplement	64
	References	I

Publications

- Christoph Schneider and Jürgen Sühnel (1999).
A Molecular Dynamics Simulation of the Flavin Mononucleotide-RNA Aptamer Complex.
Biopolymers, **50**, 287–302.
- Christoph Schneider and Jürgen Sühnel (2000).
A Molecular Dynamics Simulation Study of Coaxial Stacking in RNA.
J. Biomol. Struct. Dyn., **18**, 345–352.
- Christoph Schneider, Maria Brandl and Jürgen Sühnel (2001).
Molecular Dynamics Simulation Reveals Conformational Switching of Water-Mediated Uracil-Cytosine Base.
J. Mol. Biol., **305**, 659-667.

Abbreviations

Å	Ångstroem
AMBER	Assisted Model Building with Energy Refinement
atm	atmosphere
c	continuous
C_{6/12}	suitable constants
CURVES	computer program for the study of a number of different nucleic acid structures
DNA	desoxyribonucleic acid
dt	derivative of time
ϵ_0	permittivity of free space
ϵ_r	relative permittivity (dielectric constant)
F	force
FMN	flavin mononucleotide
fs	femtosecond
H-bond	hydrogen bond
HDV	Hepatitis Delta Virus
HF	Hartree Fock
HIV	Human Immunodeficiency Virus
K	Kelvin
m	mass
M	molar
MD	molecular dynamics
n	nicked
NMR	Nuclear Magnetic Resonance
ns	nanosecond
PDB	Protein Data Bank

PME	Particle Mesh Ewald
ps	picosecond
q	partial charge
r	atom coordinates
RESP	restrained electrostatic potential
Rev	regulator of viral expression
RMSd	root-mean-square deviation
RMSf	root-mean-square fluctuation
RNA	ribonucleic acid
RRE	rev response element
SHAKE	iterative method, subject to holonomic scleronomous constraints (time independent)
t	time
T	temperature
TAR	trans-activation response
TIP3P	transferable intermolecular potential 3-point model
v	velocity
V	potential energy
WUC	water-mediated uracil-cytosine
xLeap	acronym for X-windows 'Prep Link Edit Parm'

Chapter 1

Introduction

1.1 Biological role of RNA

Nucleic acids were originally conceived purely as carriers of genetic information in the form of the genetic code. DNA was the repository of genetic information, and RNA served as a temporary copy to be decoded in the synthesis of proteins. The discovery of transfer RNA, the 'adapter' molecules that assist in the decoding of genetic messages, broadened awareness of the role of RNA. In the past few years, we have come to appreciate the functional versatility of nucleic acids and their participation in a wide range of vital cellular processes.

Today, we recognize that RNA has greater structural versatility than DNA in the variety of its species, in its diversity of structures, and in its chemical reactivity. RNA plays crucial roles in many diverse biological processes, such as the storage of information in messenger RNA (mRNA), mRNA splicing, transport via transfer RNA (tRNA), retroviral replication, processing, and expression of genetic information by means of ribosomal RNA (rRNA) or as regulatory elements (e.g. RRE, TAR).

Cech and Altman [91, 66] discovered that proteins do not have a monopoly on biological catalysis. Some RNA molecules called 'ribozymes' (group I/II introns [31], hammerhead [119], HDV [136], hairpin [85]), are able to fold up to form active sites that catalyze chemical reactions, perhaps including a key reaction of protein synthesis. Ribozymes can accelerate a multitude of chemical transformations, the most common involving phosphodiester transfer reactions [128].

In some cases, catalytic activity is stimulated by substrates as in allosteric ribozymes [131, 132]. An allosteric ribozyme, like its protein enzyme counterpart, possesses an effector-binding site that is distinct from the active site of the catalyst. The ability to endow nucleic acid catalysts with allosteric properties provides new prospects for RNA and DNA catalysts as controllable therapeutic agents or sensors of their cognate effector compounds.

These diverse spectra of RNA functions, especially the catalytic activity, led to the idea of a prebiotic RNA world [64].

A subgroup of RNA, although not detected *in vivo*, are nucleic acid 'aptamers'. These oligonucleotides have been selected *in vitro* for specific binding to a target molecule. The process through which these molecules are isolated is called *in vitro* selection or selective evolution of ligands by exponential enrichment (SELEX) [139]. Such DNAs/RNAs are a new class of potential chemotherapeutics that could block the enzymatic activity of pathologically relevant proteins. Aptamers binding to antibiotics have been selected [127] and recently an RNA pseudoknot aptamer with high specificity for HIV-1 RT has been found [87]. But the possible variety is even larger ranging from antibodies over growth factors to enzyme inhibitors [54]. So aptamers provide a wealth of information to understand principles of nucleic acid folding, three-dimensional structures and molecular interactions.

1.2 Structural information on RNA

The specific functions of RNA molecules are based on their distinct three-dimensional structures. While, for over two decades, regular helices of double-stranded RNA and the tRNA fold had been the only known examples of natural occurring RNA three-dimensional structures [124], recent years have seen an explosion of novel RNA architectures.

Dramatic technical progress in RNA synthesis and structure determination methods and increasing interest in RNA has led to the solution of many three-dimensional structures of RNA molecules and complexes between RNA and both protein and small molecules. In particular, the crystal structure of the P4-P6 domain of self-splicing group I introns has revealed a treasure of novel RNA motifs [30]. The current structure determination of the ribosome, the protein factory, is a further step in uncovering the miracles of life and is also a great success of molecular biology [109, 29].

It is now fully recognized that RNA molecules intervene at all stages of cell life, not only because of key sequence motifs but also because of intricate three-dimensional folds. This realization, and the fact that RNA lacks repair mechanisms, has promoted RNA to a potential therapeutic target. While regular A-form helices exert little potential for specific interactions with other domains, sites have been grafted into helices that comprise structurally conserved modules. Among these modules for specific interactions are: (a) variations of the Watson-Crick base-pairing scheme (i.e. mismatches), (b) triples and quadruples of interacting bases, (c) platforms with pairing between consecutive bases within one strand, (d) bulged-out residues, (e) alternate cross-stacking between bases in different strands, and (f) recurring hydrogen-bonding patterns between riboses of consecutive nucleotides in two strands.

A subunit of RNA molecules with a variety of structural modules are aptamers. New structures reveal that aptamers bind to their ligands via a combination of well-known structural motifs and unexpected interactions that provide a tantalizing glimpse into a world orchestrated by nucleic-acid chemistry. The understanding of the interplay between different structural modules and architectures will also help us to use RNA as therapeutic agents in the future. Improvements in theoretical methods for modeling and simulation of nucleic acid systems will pave the way for computer-aided rational drug design.

1.3 Molecular dynamics (MD) simulations

Molecular dynamics (MD) simulations are a theoretical tool for the microscopic description of structural and dynamic properties of molecules [103]. In addition, MD also provides a direct route from the microscopic details to macroscopic properties. The main contributions that can be gained from microscopic observations are comprehension, interpretation of experimental results and the capability to expand simulations in experimentally poorly accessible regions.

Besides being of academic interest, therefore, the information gained can be useful in technological ways. Simulations of extreme temperatures and pressures are feasible while it could be impossible to carry out experiments under these conditions. Persistent improvements of computer power make it possible to calculate and model the structures and simulate the motions of polynucleotides containing thousands of atoms, including solvent molecules and counterions for longer time-scales. Why is it so important to understand the effects of water on the shapes of biological molecules? Principally, because a molecule's structure yields clues to how it functions, helping scientists decipher the biochemical interactions that add up to life. On a more practical level, understanding the structures of biological molecules in water may one day help researchers design new drugs that act by blocking or enhancing various biochemical pathways.

MD has a long history [58, 3, 142] and has evolved into an important and widely-used theoretical tool that allows researchers in chemistry, physics, and biology to model the detailed behavior of many different types of system, including gases, liquids, solids, surfaces, clusters and, of course, biomolecules [6]. Thus MD simulations can serve as a link between structure, function and dynamics in theory and experiment.

In a MD simulation, the classical equations of motion governing the microscopic time evolution of a many-body system are solved numerically subject to boundary conditions appropriate to the geometry or symmetry of the system. MD can also be employed as a means of sampling a statistical mechanical ensemble and determining equilibrium properties. These properties include average thermodynamic quantities (pressure, temperature, volume, etc.), structure, and free energies along reaction paths. In order to provide a picture of the microscopic behavior

of a system from the laws of classical mechanics, MD requires, as an input, a description of the interparticle interactions. The quality of the results of an MD simulation depends on the accuracy of this description. For biopolymers, the interactions between bonded and nonbonded atoms are usually defined by empirical force fields.

Recently, the reliability of MD simulations of biopolymers has been substantially improved especially by the implementation of Ewald methods for the treatment of long-range electrostatics [43, 154, 52]. This has been especially important for the highly charged nucleic acids [14]. Recent MD simulations on RNA have provided useful information complementing experimental studies [10, 73].

With increasing computational power MD simulations of solvated proteins, membranes, nucleic acids, and other large biological assemblies have become feasible. Hence, the results of current MD simulations can now be directly compared to crystallographic and NMR structural data.

In recent years a lot of work has been focused on MD of proteins and of DNA. RNA, as a target for MD simulations, has lagged behind. Accordingly, the first MD simulations on DNA [94] were published before the first MD simulation on RNA which was performed by Harvey et al. [68] in 1984 on phenylalanine tRNA. A variety of further simulations on tRNA^{Asp} [7, 11, 8, 9, 12], the hammerhead ribozyme [69], an A-RNA helix r[CCAACGUUGG]₂ [33], an RNA hairpin r[GGGCGCAAGCCU] [157] and an RNA UUCG tetraloop [105] has become available.

MD simulations are not subject to crystal packing forces found in X-ray structures. They can provide additional information by the analysis of components of the system, of the dynamic and time-averaged properties and of the surrounding solvent. Moreover, MD simulations allow us to study motions which cannot be detected in the NMR experiments due to their temporally restricted nature. An introduction to the methodology of MD is given in chapter 3.

1.4 The Flavin mononucleotide (FMN) RNA aptamer

Aptamers are nucleic acid molecules selected *in vitro* to bind small molecules, peptides or proteins with high affinity and specificity [49, 57, 114, 71]. An archaic meaning of apt is 'suited' or 'adapted'. Contrary to the usual approach, where ligands are optimized for specific binding to the nucleic acid, in the aptamer case the nucleic acid structures are selected from large random libraries. Therefore, aptamers offer a unique opportunity to extend our knowledge of the basic principles of nucleic acid architecture and of nucleic acid ligand recognition [50, 139]. The great majority of DNA or RNA aptamer structures has been determined by nuclear magnetic resonance spectroscopy (NMR) [114]. Examples are complexes with co-

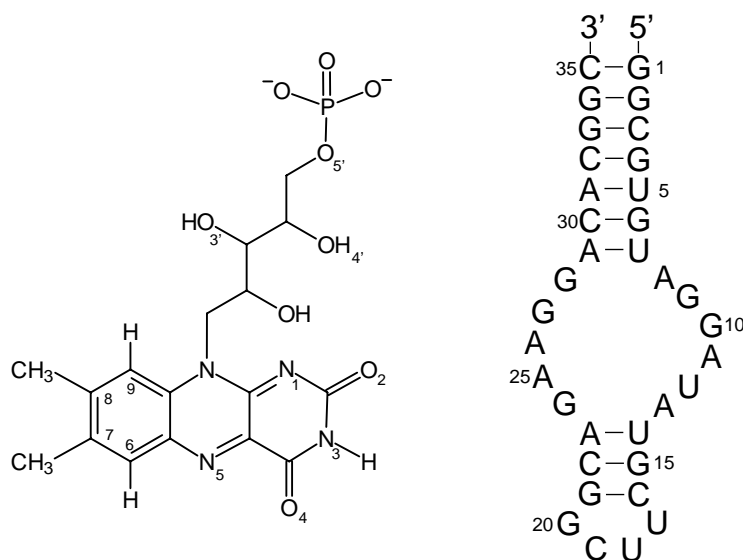


Figure 1.1: Chemical formula and numbering scheme of flavin mononucleotide (FMN) and secondary structure with numbering scheme of the 35-nucleotide RNA aptamer.

factors (FMN [55], AMP [82, 47]), amino acids (L-arginine and citrulline [152, 26]), peptides [17, 153], aminoglycoside antibiotics (Neomycin [60] and tobramycin [83]) and bronchodilators (theophylline [158]). The HIV-1 rev peptide-RRE RNA complex [17] is directly related to the corresponding aptamer [153] and, therefore, mentioned for comparison. Again, as mentioned above, aptamers for antibiotics have been found and indicate the potential as drugs.

FMN, shown in Fig. 1.1, is a cofactor that mediates redox reactions when bound as a prosthetic group in flavoproteins. The chromophore (isoalloxazine ring) of FMN serves as a transient carrier of two hydrogens at N1 and N5. Recently, it has been found that isoalloxazine derivatives, like FMN, are able to photocleave RNA specifically at G-U base pairs embedded within a helical stack [25]. The structure of the FMN-RNA aptamer was solved by NMR spectroscopy [55]. It consists of an internal loop flanked by helix regions and a UUCG hairpin loop (Fig. 1.1). The binding site is located in the internal loop region. FMN forms hydrogen bonds with an adenine and intercalates between a G-U-A base triple and a G-G mismatch.

The aim of this study is to probe the structural and dynamic features of the FMN-RNA aptamer complex including solvent-solute interactions by means of an unrestrained MD simulation.

1.5 Coaxial stacking in a nicked RNA hairpin

Coaxial stacking of helical stems is one of the prime mechanisms of RNA to build higher forms of structural organization [16]. It is a consequence of the energetically favorable interaction of the π -systems of bases, but the nature of this non-covalent interaction remains under

debate. Theoretical studies have implicated several factors as potentially important in stabilizing the face-to-face interactions as there are electrostatic (dipole-dipole and dipole-induced dipole) interactions, dispersion (momentary dipole-induced dipole) effects and solvation effects [61, 110, 111, 112]. Furthermore, *ab initio* studies of stacking interactions in DNA and RNA have been performed [133, 75, 134, 4, 59]. Also experimental studies yield contradictory results [65, 63, 34].

The coaxial stacking is found in multi-branched loops of RNA and it also occurs in structures of tRNA, the hammerhead ribozyme and of pseudoknots [79]. Coaxial stacking interactions are usually not included in free-energy minimization algorithms for RNA secondary structure prediction. Recently, however, model systems have been studied to estimate the thermodynamic contribution of this motif to the total free energy of RNA folding [145, 144, 88]. The model systems consisted of oligomers binding to a hairpin stem with a four- or five-nucleotide overhang (Fig 1.2). In these studies it has been shown that the oligomers bind approximately up

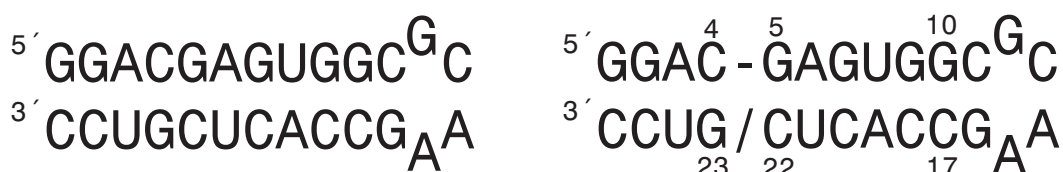


Figure 1.2: Secondary structures of the continuous (left) and nicked (right) RNA hairpins. The stacking interface in the nicked structure is indicated by /. The nicked structure differs from the continuous structure by only one missing phosphate group.

to 1000-times more tightly than predicted for a free tetramer or pentamer duplex. This means that coaxial stacking provides large, favorable free energy contributions to RNA stability. Thus far, no structural information on the RNA model systems investigated in the thermodynamic studies is available. Moreover, the coaxial stacking motif in tRNA and pseudoknots, for example, is incorporated into a larger nucleic acid environment and may thus differ from the structure of the model systems.

Some structural information for nicked DNA structures is available, however. A low-resolution X-ray structure of a nicked dodecamer DNA adopts a conformation similar to those found in the intact structure [15]. This indicates that the stacking and hydrogen bonding interactions are sufficient to overcome the effect exerted by the disruption of the backbone. It has been concluded from a gel electrophoresis study on a 139 base pair DNA fragment with a single-stranded break that the nicked duplex is kinked [93]. NMR studies and restrained MD simulations on nicked and gapped 14-mer DNA have also shown that the nicked structure is close to a canonical B-DNA but exhibits enhanced local flexibility [121]. On the other hand,

gel electrophoresis experiments on DNA [106] and crosslinking studies on RNA [37] indicate that nicked structures retain much of the rigid character of the nucleic acids with uninterrupted strands whereas gapped structures exhibit an increased flexibility.

Here, the results of four unrestrained MD simulations for a tetramer binding to a 4-nucleotide overhang at the 5'-end of a hairpin (nicked structure) and for the corresponding continuous hairpin (Fig 1.2) adopting different counterion conditions (Na^+ , Mg^{2+}) with simulation times ranging between 2 and 3 ns are presented.

1.6 Water-mediated uracil-cytosine (WUC) base pairs in an RNA duplex

Water-mediated base pairs represent a new base pair motif that has recently been identified in RNA structures [122, 77, 39, 40, 80]. In these complexes the bases are connected by direct hydrogen bonds (H-bonds) and an additional water-mediated base-base H-bond interaction. One of the first examples was found by Holbrook et al. in 1991 [77]. In the RNA duplex $(\text{r-GGACUUCGGUCC})_2$ the central GU and UC mismatches do not form an internal loop, but rather a highly regular helix. In the UC base pair H4 of cytosine is H-bonded to O4 of uracil and H3 of uracil and N3 of cytosine are linked by H-bonds to a tightly associated water molecule (temperature factor: 11.7 \AA^2). The secondary structure of the RNA duplex and a schematic drawing of the water-mediated UC (WUC) base complex is shown in Fig. 1.3.

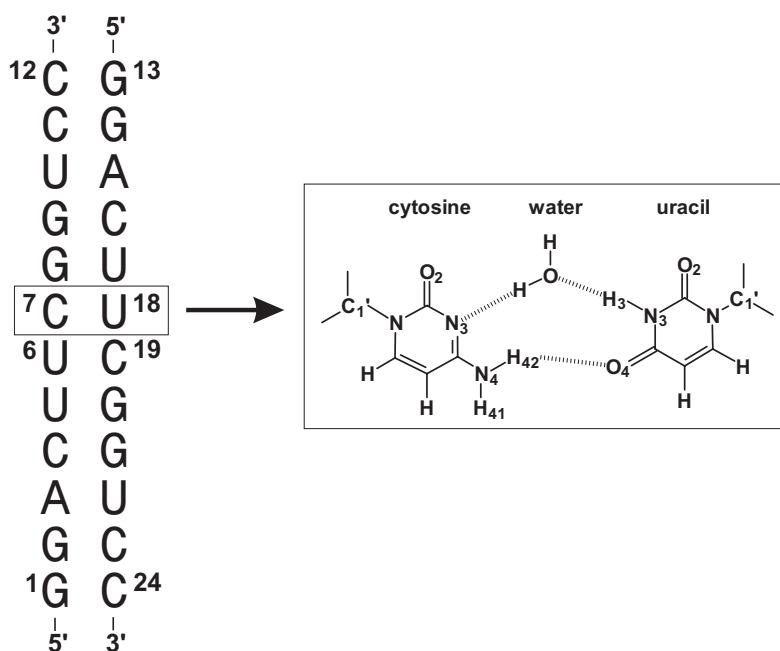


Figure 1.3: Secondary structure of the simulated RNA duplex and structural formula of the water-mediated UC base pair occurring in the duplex X-ray structure [77] (PDB code: 255d).

Due to the water molecule between the bases, the $C1'(U)\cdots C1'(C)$ distance is even slightly larger than the corresponding distances of Watson-Crick pairs in an ideal helix. Hence, even though the UC base pair is of the pyrimidine-pyrimidine type, its incorporation into a helix leads only to minor distortions of the helix geometry. Thus far, it is not clear, whether the geometry of these unusual base complexes is due to their intrinsic properties or enforced by backbone or stacking restraints exerted by the nucleic acid environment. In addition, it has to be analyzed whether or not crystal forces are important. Finally, the dynamic behavior of water-mediated pairs remains to be assessed.

The geometry of the isolated base pair calculated by quantum-chemical methods is very similar to the experimental base pair geometry in the duplex. In addition, the interaction energies of the WUC pair shows a high degree of cooperativity [23].

From these facts it has been concluded that this base pair can be regarded as a structurally autonomous building block of RNA. The quantum-chemical results indicate that neither the crystal nor the nucleic acid environment are required for the formation of this base pair [23]. They cannot prove, however, if the base pair is also stable in solution. For a different water-mediated pair there is, however, an argument in favor of the occurrence of water-mediated pairs in solution. The loop E 5S rRNA structure has been resolved both by X-ray crystallography (PDB entry: 354d, [39]) and NMR spectroscopy (PDB entry: 1a4d, [41]). The loop E structure includes three water-mediated base pairs GG, GA and UG. It turns out that in the GA pair the distance between the base donor and acceptor sites involved in the water link ($N1-H(G)\cdots N1(G)$) is with 3.88 and 3.92 Å almost the same in the NMR and X-ray structures even though, in the NMR refinement, water was not directly taken into account. This clearly shows that at least this base pair occurs in solution with a geometry similar to the X-ray structure. It should be noted, however, that this is not true for the other two water-mediated base pairs. In the GG case the $O6\cdots N7$ distance is 1.34 Å longer in the X-ray structure, whereas for the UG complex, the X-ray $N3-H\cdots H-N2$ distance is shorter by 1.0 Å. In these cases the distances between the base donor/acceptor sites differ by approximately 1 Å between the X-ray and NMR structures. MD simulations can be performed for complete medium-sized biopolymers in solution. They yield information on both the average structure and its dynamical behavior. In particular these simulations enable one to analyze short time-scales not accessible by NMR spectroscopy. Given that the simulations are performed over sufficient long simulation times, it becomes also possible to detect conformational changes. We have performed a 4 ns MD simulation of the RNA duplex $(r\text{-GGACUUCGGUCC})_2$ ([77]) in order to study the structure and dynamics of WUC base pairs.

1.7 Aim of the work

The functional and structural diversity of RNA as compared to DNA makes RNA an attractive object for studies on structure and dynamics. From a more fundamental point of view it is interesting to find out how RNA manages to generate this structural diversity having only a limited alphabet of four building blocks available. In spite of the increasing number of new three-dimensional structures their dynamics is widely unexplored. Metal ions play often a crucial role for RNA structure and stability. On the other hand, rules for the location of metal binding sites and for possible global structural effects of metal counterions have not yet been established. Water is an integral part of nucleic acid structure. Nevertheless, the analysis of hydration patterns and of the water dynamics is still in the state of infancy both from the experimental and theoretical point of view. Finally, there are relatively few studies on the conformational dynamics in RNA structure. The aim of this work is hopefully to gain more insight into the aspects mentioned for the three structure types described above.

Chapter 2

Basic principles of RNA structure

From the increasing information on three-dimensional RNA structures we have learned that RNA exhibits a greater structural versatility than DNA. Structural information on RNA can be provided in terms of primary, secondary, and tertiary structure. The primary structure is given by the one-dimensional nucleotide sequence. The secondary structure is defined by a two-dimensional representation of base-paired and unpaired regions. Finally, the tertiary structure requires data on the location of atoms in three-dimensional space. In the following a few basic facts about RNA structure are presented.

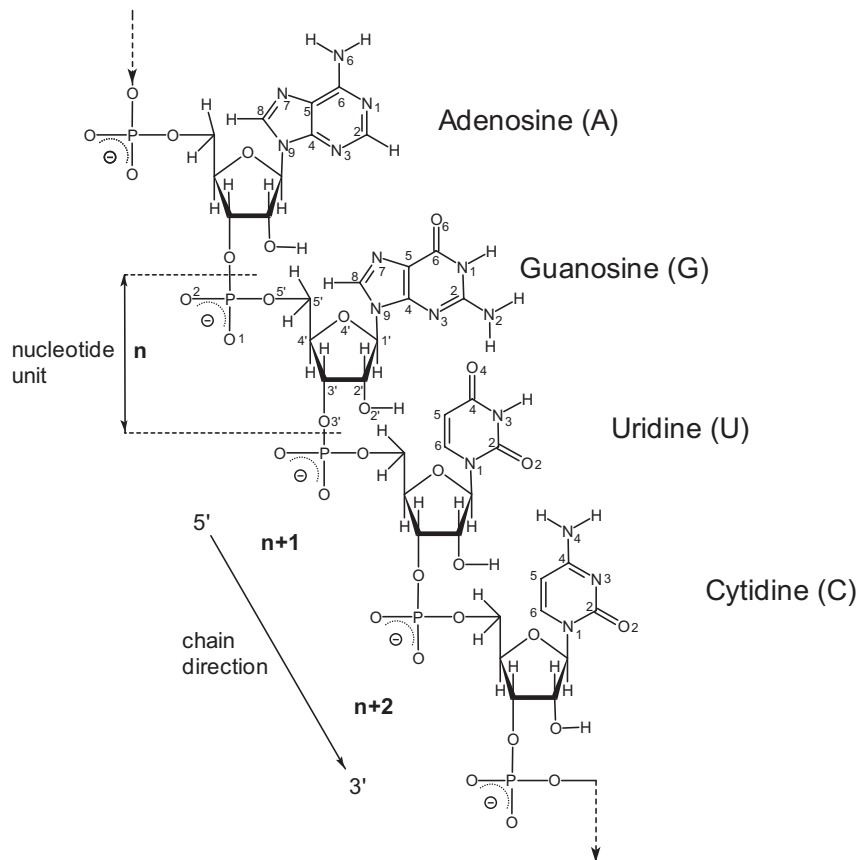


Figure 2.1: Schematic drawing and numbering of a ribonucleic acid sequence adenosine (A), guanosine (G), uridine (U) and cytidine (C).

The basic repeating unit of a nucleic acid sequence is the nucleotide. Each nucleotide is made up of three components: base, sugar, and phosphate group (Fig. 2.1 and 2.2).

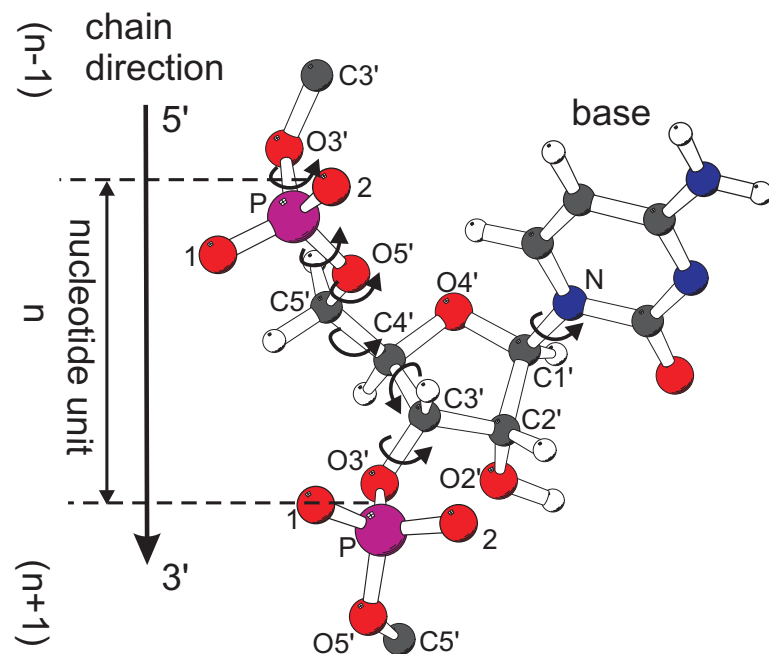


Figure 2.2: Atomic numbering and definition of torsion angles for a nucleotide unit.

Torsion angles are

$$\alpha: O3'_{n-1} - P - O5' - C5',$$

$$\beta: P - O5' - C5' - C4',$$

$$\gamma: O5' - C5' - C4' - C3',$$

$$\delta: C5' - C4' - C3' - O3',$$

$$\epsilon: C4' - C3' - O3' - P_{n+1},$$

$$\zeta: C3' - O3' - P_{n+1} - O5'_{n+1},$$

$$\chi: O4' - C1' - N1 - C2$$

(pyrimidines) and

$$\chi: O4' - C1' - N9 - C4$$

(purines).

In RNA there are mainly 4 different bases, namely adenine (A), guanine (G), uracil (U) and cytosine (C). The bases are attached to the ribose via a β -D-glycosyl linkage. The ribose sugars and the phosphate groups form together the nucleic acid backbone, where the nucleotides are connected by 5',-3'-phosphodiester bonds (Fig. 2.1). Each phosphate group has a partial charge of $'-1'$. This feature makes nucleic acids highly charged biomolecules. The main difference between DNA and RNA at the atomic level are, that RNA contains a 2'-hydroxyl group at the ribose and it uses uracil for base pairing with adenine instead of thymine.

RNA can be stabilized by association forces between bases. There are two kinds of interactions, one in the plane of the bases by means of hydrogen bonding of complementary bases (base pairing) and forces perpendicular to the base planes (base stacking) mainly driven by London dispersion forces, electrostatic interactions, solvent and hydrophobic effects.

Base pairs in nucleic acids can be formed by interactions via the Watson-Crick edge, the Hoogsteen edge and the shallow groove edge [150] (Fig. 2.3). In order to arrive at a comprehensive base-pair classification one has to take into account the χ torsion angle about the bond connecting the base with the ribose ring. The standard Watson-Crick base-pairs AU and GC are shown in Fig. 2.4. As already pointed out a great number of non-canonical base pairs like the Wobble, Hoogsteen and reverse Hoogsteen type and even water-mediated pairs can occur in RNA structures. Further mismatches like purine-purine or pyrimidine-pyrimidine interactions can emerge [27].

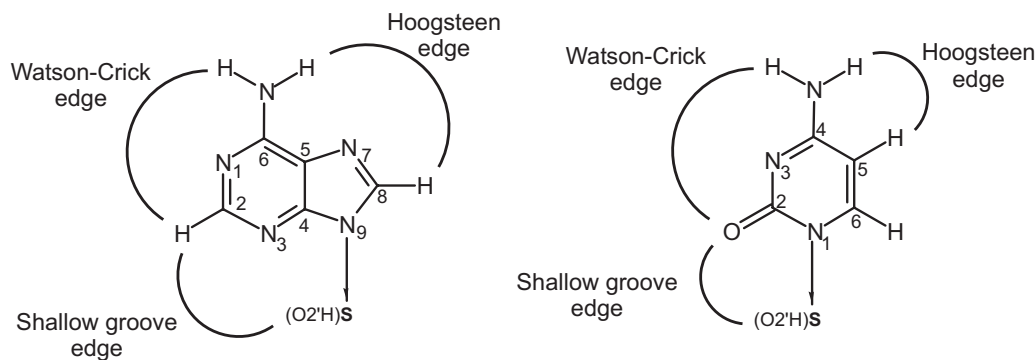


Figure 2.3: The three hydrogen-bonding edges for a purine base (left) or a pyrimidine base (right). The ribose hydroxyl O2'H frequently participates in the hydrogen-bonded pair in the shallow groove edge.

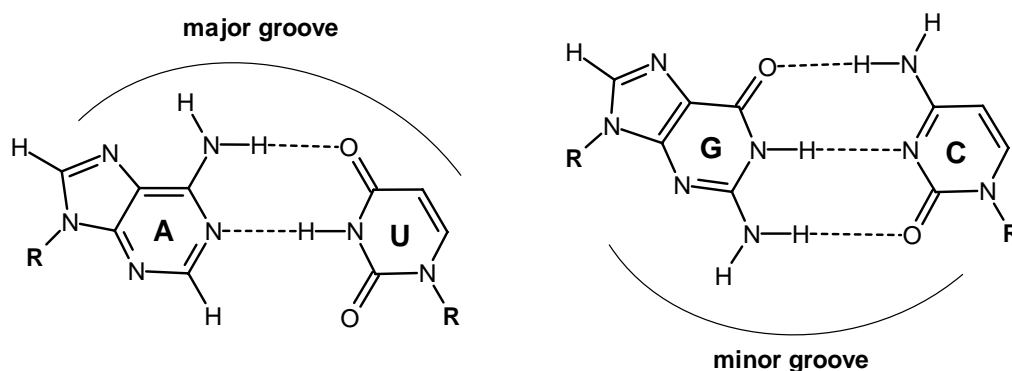


Figure 2.4: Watson-Crick base pairs of RNA. R stands for the connection to the ribose C1'.

These base pair arrangements form higher aggregates called secondary structures. RNA secondary elements are: single-strands, helices, helices with dangling ends, bulges, hairpin loops, symmetric and asymmetric internal loops, coaxial stacks and junctions (Fig. 2.5).

The coaxial stacking of helical stems, the most basic mechanism to form higher ordered organizations in RNA [16], is a consequence of the favorable stacking interactions. The contribution of coaxial stacking to RNA folding was first recognized solving the crystal structure of tRNA^{Phe} [89, 120].

Secondary structure elements can interact with each other via so-called tertiary interactions. Examples are kissing hairpins [102, 32] and hairpin-loop interactions and in a certain sense also pseudoknots [117, 42, 118]. Analyzing randomly chosen RNA from a cell shows that about 50% of its secondary elements are of double-stranded helical structure. The great majority of RNA helices adopt an antiparallel, right-handed A-conformation with two grooves. In this case the sugar pucker is in a C3'-endo orientation. That means C3' is out of the ribose plane and on the same side as C5'. A-RNA shows two grooves. The major groove is deep and narrow,

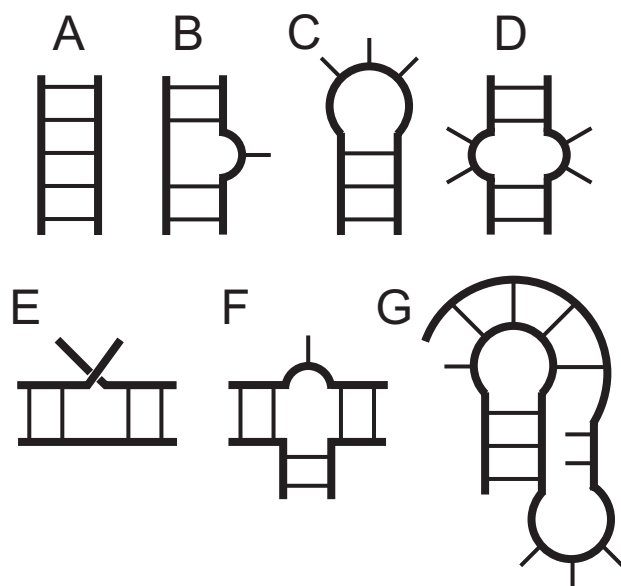


Figure 2.5: Elements of RNA secondary structure. *A* double helix, *B* single bulge, *C* hairpin loop, *D* symmetric internal loop, *E* coaxial stack, *F* three-stem junction, and *G* pseudoknot, which is a tertiary structure.

whereas the minor groove is shallow and wide. Therefore, the minor groove is often called shallow groove (Fig. 2.3). While the major groove requires distortion or widening to serve as a recognition motif [147], the minor groove has a higher impact on recognition. The helical geometry is often described in terms of rotational (tip, inclination, opening, propeller, buckle, twist, roll and tilt) and translational (x-y-displacement, stagger, stretch, shear, rise, slide and shift) parameters with regard to a helical axis [44, 45].

The major forces shaping RNA structure are H-bond interactions within base pairs, stacking interactions between different base pairs and backbone-backbone interactions.

Chapter 3

Computational details of MD simulations

3.1 General aspects

Molecular dynamics (MD) is the term used to describe the solution of the classical equations of motion for a set of molecules. Newton's second law (Equation 3.1) tells us that the velocity of the particles changes with time at a rate proportional to the force F_i and that the proportionality constant can be taken to be the reciprocal of the mass m of the particle.

$$\begin{aligned} F_i &= m_i \frac{d^2 r_i(t)}{dt^2} \\ &= - \frac{\partial V(r_1, r_2, \dots, r_N)}{\partial r_i} \end{aligned} \quad (3.1)$$

The force in turn can also be described as the gradient of the potential energy V , with respect to the atom coordinates r_i .

3.2 Technique and approaches

3.2.1 Molecular model and force field

A typical molecular force field or effective potential of a system with N atoms with the masses m_i ($i = 1, 2, \dots, N$) and the Cartesian coordinate vectors r_i has the form given in Equation 3.2. All MD simulations in this work were done with the AMBER package [115, 28] and the Cornell force field [38].

$$\begin{aligned}
V(r_1, r_2, \dots, r_N) = & \sum_{bonds} \frac{1}{2} K_b [b - b_0]^2 \\
& + \sum_{angles} \frac{1}{2} K_\theta [\theta - \theta_0]^2 \\
& + \sum_{im.torsions} \frac{1}{2} K_\xi [\xi - \xi_0]^2 \\
& + \sum_{torsions} K_\phi [1 + \cos(n\phi - \delta)] \\
& + \sum_{pairs(i,j)} \left[\frac{C_{12}(i,j)}{r_{ij}^{12}} - \frac{C_6(i,j)}{r_{ij}^6} + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \right]
\end{aligned} \tag{3.2}$$

The first term describes the covalent bond-stretching of bond b . It has an harmonic potential where the equilibrium distance b_0 and the force constant K_b depend on the bond type. The second term describes in a similar way the deformation of a bond angle θ . Two terms specify the four-body interaction of the torsion angles. An harmonic term for improper torsions ξ where transitions are forbidden (e.g. inside aromatic rings) and a cosine term for the remaining torsion angles ϕ where rotations of 360° are allowed. The last term is a sum over all atom pairs and characterizes the effective non-bonded interactions which consist of van-der-Waals and Coulomb interactions between atoms i and j with the partial charges q_i and q_j at a distance r_{ij} . C_6 and C_{12} are suitable constants, ϵ_r and ϵ_0 are the dielectric constant and the permittivity of free space respectively. The initial velocities of all atoms are chosen randomly accordingly to a selected temperature and to the Maxwell-Boltzmann distribution. They are corrected so that there is no overall momentum.

3.2.2 Long range forces

The number of terms in the summation over all atom pairs is proportional to N^2 , while the other sums are proportional to N . The efficiency of the calculation can be improved by reducing the number of summation terms. One simple method is the implementation of a cutoff criterion. Usual values for the cutoff distance vary between 6 and 10 Å. This approach is very likely justified for forces associated with the Lennard-Jones potential. However, since the Coulomb term is proportional to r^{-1} , the electrostatic forces are of long range nature and a cutoff may thus lead to significant errors especially for ion-ion interactions. Truncation of ion-ion interactions or highly charged particles beyond the cutoff radius would not be acceptable. The Ewald sum [53] was originally used for summing up the interactions between an ion and all its periodic images in crystals. It was rebuilt as a $N \log(N)$ method called particle mesh Ewald (PME) by Darden et al. [43] and finally modified [52] to be useful in simulating biomolecular systems.

This PME method is capable of efficiently calculating long-range coulombic interactions. This improvement makes it possible to perform MD simulations of nucleic acids without geometrical constraints. We have run all simulations using the PME method within AMBER 4.1 or AMBER 5.0 respectively, using a cubic B-spline interpolation order and a 10^{-5} Å tolerance for the direct space sum cutoff. The PME charge grid is chosen to be a product of powers of 2, 3, and 5, and to be slightly larger than the size of the periodic box. This speeds up the fast Fourier transform in the calculation of the reciprocal sum and leads to a grid spacing of ~ 1 Å or less. The PME method requires a neutralization of the unit cell, because the summation method is not correct when a non-neutral system is used. This is normally done by adding counterions. The nonbonded list was updated every 10 steps. A 10 Å cutoff was applied to the Lennard-Jones interactions.

3.2.3 Sampling of the conformational space

For small systems with a small number of degrees of freedom the whole conformational space can be explored in a systematic manner. If a system contains too great a degrees of freedom it is impossible to search the overall phase space. In this case, random sampling methods like the Monte-Carlo (MC) approach can be used. This algorithm has to ensure that the sequence of produced conformations must be proportional to its Boltzmann factor. The MC methods are not very efficient for macromolecular systems because a lot of the computational cost will be wasted for irrelevant conformations.

In contrast, the MD simulation produces a trajectory of the molecular system in phase space. Each step is directly linked to previous steps. Statistic equilibrium values can be obtained by averaging over the trajectory. The trajectory should be long enough to represent a statistically relevant ensemble. For each ensemble the thermodynamic variables are fixed. Common statistical ensembles are the microcanonical, or constant-NVE, the canonical, or constant-NVT, the grand canonical constant- μ VT ensemble and the isothermal-isobaric constant-NPT ensemble. The NPT ensemble is closest to nature and is used throughout this work.

3.2.4 Boundary conditions

The simulation of systems with finite size requires special treatment at the surface. Liquids or solutions require periodic boundary conditions to minimize boundary effects. The cubic box with all atoms inside is replicated throughout space to form an infinite lattice. Thus, as a molecule leaves the central box, one of its images will enter through the opposite face. Hence, the system contains no walls at the boundary of the central box and no molecules at the surface. The number of atoms in each the box is conserved. The use of periodic boundary conditions

inhibits the occurrence of long-wavelength fluctuations. The periodicity suppresses any density wave with a wavelength greater than the cube length L .

3.2.5 Algorithms for MD simulations

Computer simulations produce an ensemble of system conformations. Whether these are representative depends on how thoroughly the important parts of the phase space, which are usually of low energy, could be sampled. Hence, the sampling algorithm should be able to overcome a multiplicity of barriers on the multidimensional energy surface. The algorithm should permit a long time step δt and should duplicate the classical trajectory as closely as possible. It should further satisfy the known conservation laws for energy and momentum, be time-reversible and computationally efficient. The Verlet algorithm [142], which is widely used to integrate the equations of motion, was modified to build the half-step 'leap-frog' scheme [76]. The 'leap-frog' algorithm takes the form shown in equation (3.3) and (3.4).

$$r(t + \delta t) = r(t) + \delta t v(t + \frac{1}{2} \delta t) \quad (3.3)$$

$$v(t + \frac{1}{2} \delta t) = v(t - \frac{1}{2} \delta t) + \delta t a(t) \quad (3.4)$$

The algorithm only contains the current positions $r(t)$, the acceleration $a(t)$ and the velocity $v(t)$.

3.2.6 Distance constraints

The length of the integration time step Δt is limited by the motion with the highest frequency in the system. The highest frequencies are normally vibrations of hydrogen atoms bonded to a heavy atom. These fluctuations are of little interest when simulating biomolecular systems of appropriate size. A means to elongate the simulation time is thus the use of distance constraints. Perhaps the most widely used method of resetting the bond distance was developed by Ryckaert et al. [123] and is called SHAKE. All simulations presented in this work were carried out by using SHAKE for all bonds involving hydrogens with a tolerance of 0.0005 Å. The SHAKE method allows us to set the time step to 2 fs.

3.2.7 Solvent and ions

All MD simulations presented in this work have been carried out using water as solvent. The water model used is called TIP3P [84] and is tuned for the Cornell force field [38]. TIP3P is a 3-point water model with improved energy and density terms for liquid water. First water molecules were placed around the solute. Then counterions were placed using a coulombic

potential on a grid. Water molecules were removed when they clashed with the counterions that had just been placed. The parameters for sodium and chlorine were taken from the Cornell force field, and the magnesium parameters were adapted from Åqvist [1].

3.2.8 Heating up and equilibration

The first step was the relaxation of the water molecules and ions by minimization while the solute was kept fixed. Then the system was gradually heated up to 300 K. Constant pressure dynamics with anisotropic diagonal position scaling has been used with a reference pressure of 1 atm. Detailed input parameters are listed in the supplement (Chapter 6).

3.3 Problems

Computer simulations of biopolymer structures in solution face two major problems. First, the size of the system and second the accuracy of the molecular model described by the molecular force field.

3.3.1 System size

The simulation of molecular systems at temperatures above the zero-point level requires the generation of a statistically representative set of configurations, an ensemble. System properties are defined as ensemble means or integrals over the phase space. Since many-body systems contain various degrees of freedom it is not possible to sample the whole phase space. Thus the smaller the system the longer the time that can be simulated with the same amount of computational power. This is also important for longer equilibration times of divalent ions in comparison to monovalent ions. The choice of the approximation level should be chosen in such a way that the degrees of freedom, that are of interest for an exact calculation of an item or system property, can be calculated in a sufficient magnitude.

3.3.2 Accuracy

Assuming that the degrees of freedom is of infinite density or could be sampled over long time ranges, the accuracy is then directly dependent on the quality of presumption and approximation of the molecular model. A molecular system can be described as a system of moving mass points in an effective potential field. This potential field is normally a conservative field, that means it depends solely on the current coordinates of the mass points. The movement of mass points can be described with sufficient preciseness through the rules of classical mechanics.

Quantum effects can be omitted. Different force fields might yield different results and some implementation might lack the PME method. While the reliance of the prediction made on the basis of computer simulations obliquely several factors, a critical test should always be the comparison with experimental data where these are available.

3.4 Systems studied in this work

3.4.1 The FMN-RNA aptamer

The Cornell force field has been developed for the simulation of structures, of proteins, of nucleic acids, and related organic molecules in condensed phase. But it contains parameters only for proteins and nucleic acids. Flavin mononucleotide (FMN) is a naturally occurring cofactor that mediates redox reactions. Force field parameters for FMN have been determined by a quantum-chemical *ab initio* calculation of the electrostatic potential at the HF/6-31G* level. The partial charges were derived from a restrained electrostatic potential (RESP) fit [18]. The force field parameters were obtained according to the philosophy adopted by the AMBER 4.1 software package [116, 38]. Initial coordinates of the FMN-RNA aptamer were taken from model 1 of the experimental NMR structure (Protein Data Bank (PDB) [21] code: 1fmn [55]). The 35 nucleotide FMN aptamer was solvated in a rectangular box ($36 \times 70 \times 36 \text{ \AA}^3$) with 36 Na^+ counterions and 2607 TIP3P water molecules (9040 atoms). The counterions were placed using a Coulombic potential on a grid with the xLeap module. The aptamer was first kept fixed to relax the water molecules and ions and then heated up over 20 ps with a following 10 ps run at 300 K. After minimization and warming up an unrestrained simulation was run under periodic boundary conditions for 1700 ps at constant temperature (Berendsen coupling [19]) and constant pressure with anisotropic scaling (300 K, 1 atm) with a 2 fs time step. SHAKE was applied to all hydrogen bonds. The nonbonded pair list was updated every 10 steps. The electrostatic interactions were calculated by the PME summation with a grid spacing of approximately 1.0 Å. The box size varied during the simulation between 35-37 Å for x, 61-73 Å for y and 35-39 Å for the z direction. An average structure was generated from the last 500 ps of the simulation. It was not processed further. In order to remove overall rotation and translation of the solute, each coordinate set was first positioned for minimum mass-weighted root-mean-square deviation (RMSd) of the group on the reference coordinates. All RMSd values refer to heavy atoms only. As in many other MD simulations of RNA structures [7, 11, 8, 9, 12], the effect of Mg^{2+} was not taken into account. On the other hand, divalent cations are important for nucleic acid structures and the first attempts to take them into account have been undertaken [70, 100, 69, 72, 137, 156, 135]. We have also introduced Mg^{2+} ions in the calculations of

coaxial stacking of a hairpin in section 4.2. Other analysis tools used were Carnal from the AMBER package and CURVES 5.1 [92].

3.4.2 Coaxial stacking in a nicked RNA hairpin

An initial hairpin structure was built from 2 subunits, loop and stem. The GCAA loop was taken from an experimental tetraloop structure (Protein Data Bank entry: 1zih [21, 86]). The stem structure was built de novo with SYBYL (Tripos Ass., Inc.) in a canonical A-RNA conformation and merged with the loop. A nicked structure (in the following denoted as n) was generated by removing the phosphodiester linkage between C22 and G23 (see Fig. 1.2) from the continuous backbone hairpin (in the following denoted as c). These model structures were solvated with 2800 TIP3P water molecules. A Coulombic potential on a grid was calculated and the counterions were placed accordingly with xLEaP [125] from the AMBER package [28] to neutralize the system. Finally, either Na^+ only or Mg^{2+} and Na^+ counterions were added. The concentration of Na^+/Cl^- ions in the Na^+ only simulation were 0.51/0.07 M, and in the simulation with $\text{Na}^+/\text{Mg}^{2+}$ the concentrations of the $\text{Mg}^{2+}/\text{Na}^+/\text{Cl}^-$ ions were 0.21/0.12/0.1 M, respectively. In this manner four systems were obtained, the nicked structure with Na^+ or $\text{Na}^+/\text{Mg}^{2+}$ (nNa^+ , nMg^{2+}) and the continuous structure with Na^+ or $\text{Na}^+/\text{Mg}^{2+}$ (cNa^+ , cMg^{2+}). All simulations were run using the Sander module of AMBER 5 [28] with the Cornell force field [38]. Mg^{2+} parameters were adopted from Åqvist [1]. Counterions were placed with the 'add ions' option in xLEaP. The nonbonded pair list was updated every 10 steps and a 10 Å cutoff was applied to the Lennard-Jones interactions. All structures were minimized for 1000 steps, afterwards water molecules and ions were relaxed, while the solute was kept fixed. Then all atoms were allowed to move. The system was gradually heated up to 300 K and equilibrated for additional 110 ps. Then production runs were performed for 2 to 3 ns at constant pressure (1 atm) and constant temperature (300 K) with a 2 fs time step. The simulation times were 2.5 ns for nMg^{2+} and cMg^{2+} , 2 ns for cNa^+ , and 3 ns for nNa^+ . The PME method [52] was used for calculating the electrostatic interactions with a grid spacing of approximately 1 Å. SHAKE was applied to all hydrogen bonds. The simulated systems had a density of 1.06 g/cm³. For all four simulated systems we have generated average structures taking into account the last 500 ps of simulation time.

The CURVES 5.1 program [92] has been used for the analysis of helical parameters. Only the 11 base pairs of the stem were included in the analysis. Moreover, the base C25 in the cNa^+ average structure was ignored for the determination of the helical axis because an opening was found for the G2-C25 base pair and this would significantly affect the orientation of the global helical axis. G2 was taken into account because it has a similar orientation as the neighbouring bases. Only global rotational and translational helix parameters are presented in

this work. In view of the fact that different analysis algorithms may lead to different helical parameter values [96], these data were compared to the corresponding local CURVES parameters and to parameters determined with the program FREEHELIX [46]. The major groove width of the nicked structures cannot be measured with the CURVES algorithm because the number of successive base pairs is too small. We have, therefore, adopted a different approach. In ideal A-RNA, the distance between the P atoms of nucleotides i and $i'+6$ is 10 Å, where i and i' indicate pairing nucleotides. This leads to a groove width of 4.2 Å when subtracting twice the van-der-Waals radius of P (2.9 Å). In an irregular structure the selection of a fixed base pair difference may lead to artificial irregularities in the groove width. We have thus selected the shortest $P(i) \cdots P(i'+x)$ distance, where x can vary between 5 and 8. In order to visualize ion distributions around the RNA, the structures of all frames from the last 500 ps of simulation time were superimposed so that they best fit into the simulated RNA average structure. For this reason, the ion distributions shown include the effect of the superposition error and give only a qualitative impression. Diffusion coefficients of Na^+ and Mg^{2+} ions have been calculated from the last 500 ps of simulation time and all ions taking into account (AMBER module ptraj).

3.4.3 WUC base pairs in an RNA duplex

MD simulations of the RNA duplex $(r\text{-GGACUUCGGUCC})_2$ were performed with the AMBER program package adopting the Cornell force field [38]. The starting coordinates were taken from the experimental RNA duplex X-ray structure (PDB entry: 255d, [77]). Counterions were placed with the 'add ions' option in xLEap. We have used 22 Na^+ counterions, additional 12 Na^+/Cl^- pairs (0.1 mol/l) and 2962 TIP3P water molecules to solvate the RNA. The system density was 1.06 g/cm³ with box dimensions of $(42 \times 57 \times 42 \text{ \AA}^3)$. The total system consisted of 9696 atoms. It was heated up over 10 ps with a following equilibration run of 100 ps at 300 K. Then an unrestrained simulation was carried out over 4000 ps at constant temperature (300 K) and constant pressure (1 atm) with anisotropic scaling and a time step of 2 fs. SHAKE was applied to all H-bonds. The nonbonded pair list was updated every 10 steps. The electrostatic interactions were calculated by PME summation with a grid spacing of approximately 1 Å. Coordinates were saved at time steps of 1 ps.

Chapter 4

Results and Discussion

4.1 The FMN-RNA aptamer

4.1.1 Standard MD analysis and global structural features

The unrestrained MD simulation leads to a stable structure (Fig. 4.1).

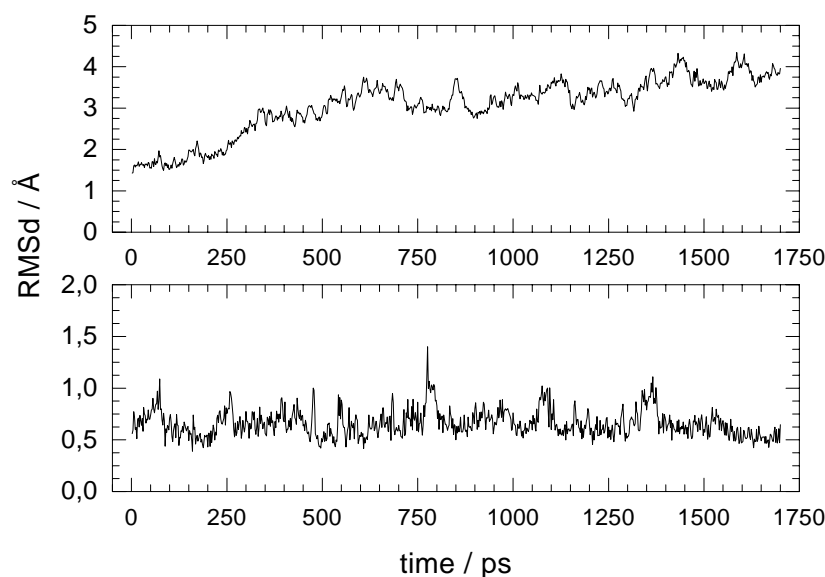


Figure 4.1: Time-course of the RMSd (in Å) of the whole RNA aptamer (top) and for FMN only (bottom) against the NMR starting structure (heavy atoms only).

The RMSd value for the comparison of the simulated average structure and the NMR starting structure is 3.7 Å. If one analyzes different parts of the complex separately, the following RMSd values are obtained: hairpin loop (residues 16–21): RMSd = (2.2 ± 0.2) Å, first and last three nucleotides (residues 1–3,33–35): RMSd = (3.3 ± 0.2) Å, FMN binding site (residues 9,

10, 12, 25, 26, 27): $\text{RMSd} = (1.2 \pm 0.1) \text{ \AA}$, FMN (residue 36): $\text{RMSd} = (0.6 \pm 0.1) \text{ \AA}$, total (residues 1-36): $\text{RMSd} = (3.7 \pm 0.3) \text{ \AA}$. These values indicate that the binding site including FMN stays close to the starting structure and that the hairpin loop and the first and last three nucleotides of the aptamer exhibit the largest differences. For FMN this can also be seen from the time-course of the RMSd values shown in Fig. 4.1.

The backbone superposition of the start and average structures in Fig. 4.2 shows a good overall agreement between the NMR and simulated structures. However, the relative orienta-

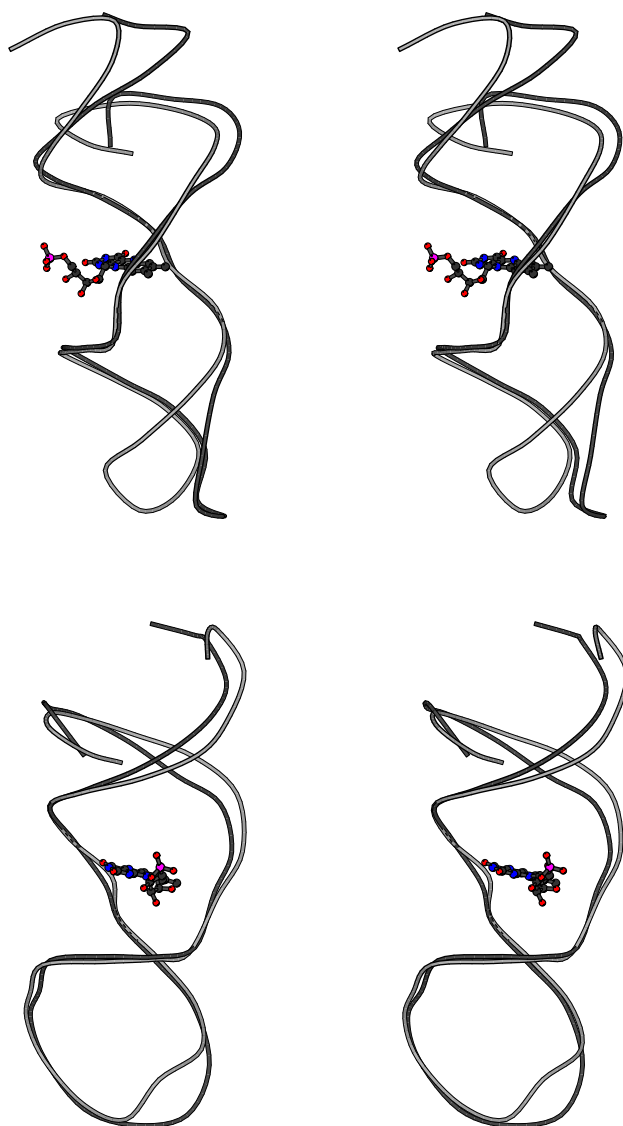


Figure 4.2: Stereoplots of the superposition of the model 1 NMR structure (light grey) and the simulated average structure (dark grey). FMN is shown in a balls-and-sticks representation. The structure in the bottom plot is rotated by 90° about the z direction as compared to the structure shown on the top.

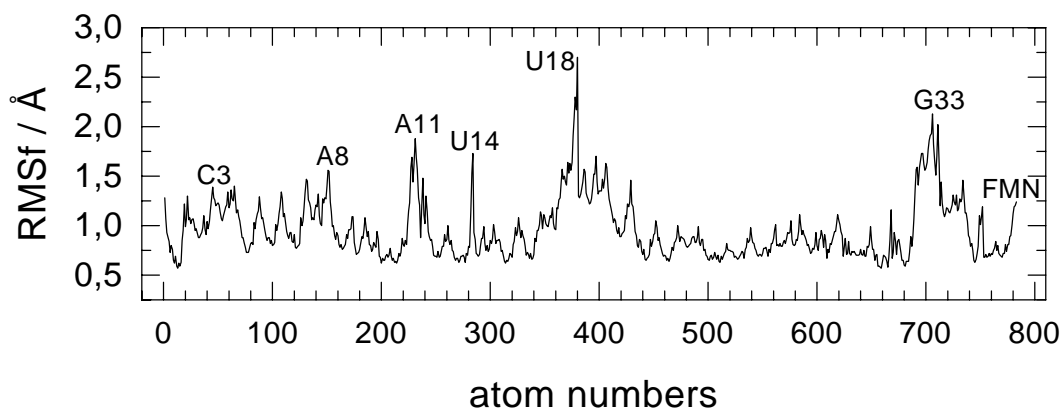
tion of the hairpin loop towards the stem differs and the first and last three nucleotides 1-3 and 33-35 exhibit a structural reorganization not seen in the backbone superposition. In the NMR structure the base pairs G1-C35 and U5-A31 are almost planar. G4-C32 is slightly distorted and the two base pairs C3-G33 and G2-C34 show a more serious distortion. By contrast, in the

Table 4.1: Atomic RMS fluctuations (RMSf) relative to the atom average positions for nucleotide atom types.

Type	RMSf (Å)	Type	RMSf (Å)	type	RMSf (Å)
O1P	1.60	O5'	1.30	N1	1.12
O2P	1.49	C4'	1.26	C2	1.12
P	1.39	C3'	1.26	N2	1.12
O2'	1.38	C2'	1.24	N3	1.11
O3'	1.37	O4'	1.22	C4	1.09
C5'	1.32	C1'	1.19	N6	1.01

simulated structure the nucleotides 1-3 and 33-35 show no base pairing at all. The NMR data clearly show that the three terminal base pairs of the stem are intact. Therefore, it may be that the simulation cannot properly describe this end effect.

The individual atomic motions within the RNA aptamer were also examined. The fluctuations

**Figure 4.3:** Atomic fluctuations (RMSf) for all heavy atoms relative to their average structure during the last 500 ps.

of an atom relative to its average position (RMSf) reveal the extent of motion experienced by this atom. In Table 4.1 the RMSf values for most atom types averaged over all nucleotides are given and in Fig. 4.3 the corresponding values for all individual heavy atoms are shown.

Table 4.2: Mean distances and standard deviations (in Å) of the experimental structure family (NMR) and of the last 500 ps of the simulation (MD) for both FMN–RNA and cross-strand RNA–RNA NOEs identified in the NMR experiments.

FMN–RNA							
FMN	RNA	NMR	MD	FMN	RNA	NMR	MD
H3	H3(U12)	3.4 ± 0.2	3.4 ± 0.3	H8M3	H8(G9)	3.9 ± 0.8	4.2 ± 0.8
H3	H8(A25)	3.8 ± 0.1	4.0 ± 0.3	H8M1	H1'(G9)	5.4 ± 0.6	5.7 ± 0.7
H3	H2'(A25)	5.4 ± 0.1	5.2 ± 0.4	H8M2	H1'(G9)	5.8 ± 0.8	5.8 ± 0.7
H3	H8(A26)	3.3 ± 0.2	3.0 ± 0.2	H8M3	H1'(G9)	5.3 ± 0.6	5.8 ± 0.7
H3	H61(A26)	2.6 ± 0.2	3.0 ± 0.2	H8M1	H2'(G9)	3.1 ± 1.0	3.5 ± 0.8
H3	H62(A26)	4.3 ± 0.2	4.7 ± 0.2	H8M2	H2'(G9)	3.5 ± 0.7	3.5 ± 0.8
H3	H8(G27)	5.3 ± 0.4	3.8 ± 0.4	H8M3	H2'(G9)	3.2 ± 0.5	3.6 ± 0.8
H6	H1(G10)	3.8 ± 0.1	3.4 ± 0.3	H8M1	H3'(G9)	3.6 ± 0.5	3.7 ± 0.7
H9	H1'(G9)	4.9 ± 0.2	4.4 ± 0.3	H8M2	H3'(G9)	3.9 ± 0.9	3.8 ± 0.8
H9	H2'(G9)	2.6 ± 0.2	2.4 ± 0.2	H8M3	H3'(G9)	3.6 ± 0.7	3.8 ± 0.8
H9	H1'(G10)	3.1 ± 0.2	3.3 ± 0.4	H8M1	H8(G10)	3.9 ± 0.5	4.2 ± 0.8
H7M1	H1(G10)	4.6 ± 0.5	4.2 ± 0.8	H8M2	H8(G10)	4.0 ± 0.7	4.1 ± 0.8
H7M2	H1(G10)	3.8 ± 0.5	4.1 ± 0.8	H8M3	H8(G10)	4.2 ± 0.9	4.2 ± 0.7
H7M3	H1(G10)	3.9 ± 0.9	4.2 ± 0.8	H8M1	H1'(G10)	4.6 ± 0.6	4.9 ± 0.6
H8M1	H8(G9)	4.4 ± 0.6	4.0 ± 0.7	H8M2	H1'(G10)	4.9 ± 0.6	4.8 ± 0.7
H8M2	H8(G9)	4.5 ± 0.6	4.2 ± 0.8	H8M3	H1'(G10)	5.1 ± 0.7	4.9 ± 0.7
RNA–RNA							
RNA	RNA	NMR	MD	RNA	RNA	NMR	MD
H8(A25)	H3(U12)	3.0 ± 0.2	3.0 ± 0.2	H2'(A25)	H2(A13)	5.4 ± 0.1	5.2 ± 0.4
H61(A25)	H3(U12)	4.4 ± 0.2	4.5 ± 0.2	H1'(A25)	H2(A13)	4.1 ± 0.5	3.2 ± 0.4
H62(A25)	H3(U12)	2.7 ± 0.2	2.8 ± 0.2	H2(A25)	H2(A13)	3.8 ± 0.5	4.0 ± 0.4
H3'(G24)	H1(G10)	2.6 ± 0.3	2.5 ± 0.2	H3(U12)	H1(G24)	5.5 ± 0.1	6.5 ± 0.4
H1(G24)	H1(G10)	6.6 ± 0.3	6.7 ± 0.3	H6(U12)	H1(G24)	3.7 ± 0.4	4.3 ± 0.4
H8(G24)	H1(G10)	3.5 ± 0.2	4.0 ± 0.3	H5(U12)	H1(G24)	2.9 ± 0.2	3.3 ± 0.4
H1(G28)	H2(A8)	3.4 ± 0.7	3.0 ± 0.3				

Each nucleotide has a peak which is due to the especially flexible phosphate group, whereas the ribose and base parts are more rigid. The calculated values are similar in magnitude to those found in other simulations [157, 105]. Upon closer inspection it turns out, however, that this order may be reversed for bases exposed to the solvent (A11, G33, U18) [78]. The data shown in Fig. 4.3 indicate that the looped out base A11 (atoms 218-239), two nucleotides of the loop region (atoms 365-404, nucleotides 18-19) and the nucleotides 33 and 34 (atoms 690-732) are the most flexible parts of the complex. In Table 4.2 both cross-strand NOEs within the RNA aptamer and intermolecular NOEs between FMN and RNA are compared to the atomic distances of the simulated average structure and are found to be in good agreement. Finally, it has to be pointed out that the FMN-A26 interaction as well as the base triple G10-U12-A25 and the mismatches A8-G28 and A13-G24 are maintained during the simulation. From these data we can conclude that the simulation preserves the overall structural features of the NMR aptamer complex.

4.1.2 Backbone torsional angles

The sugar-phosphate backbone geometry of nucleic acids is characterized by bond lengths, bond angles and torsion angles. The torsion angles include the angles ν_0 to ν_4 in the ribose ring, the torsions α (O3'-P-O5'-C5'), β (P-O5'-C5'-C4'), γ (O5'-C5'-C4'-C3'), δ (C5'-C4'-C3'-O3'), ϵ (C4'-C3'-O3'-P), ζ (C3'-O3'-P-O5') and finally the glycosidic torsion angle χ , which describes the torsion around the bond linking the backbone to the base (see Fig. 2.2 on page 11). For an ideal A-RNA the conformational parameters α to ζ adopt the values $\alpha = 292^\circ$, $\beta = 178^\circ$, $\gamma = 54^\circ$, $\delta = 82^\circ$, $\epsilon = 207^\circ$, $\zeta = 289^\circ$ [124]. Fig. 4.4 shows the time-course of these angles and of χ over the last 500 ps for nucleotides 2 to 35.

These data provide a comprehensive overview on the dynamical behavior of the backbone conformation of each individual nucleotide. A further often used geometrical quantity describing the backbone torsions is the pucker phase angle P, which can be calculated from the ν angles. For standard A-RNA it adopts values between 0° and 40° which corresponds to the C3'-*endo* conformation ($\delta \approx 82^\circ$). The alternative C2'-*endo* conformation ($\delta \approx 139^\circ$) preferably found in B-DNA corresponds to P values between 145° and 190° . It is occasionally found in RNA structures as well. During the simulation, dynamically stable P values were obtained. From Fig. 4.4 one can see that the purine nucleotides A and G exhibit no C2'-*endo* conformation, whereas the five pyrimidine nucleotides U7, U18, C19, C32 and C34 were found to adopt primarily this geometry. In principle, NMR measurements can discriminate between C2'-*endo* and C3'-*endo* conformations, because C2'-*endo* shows a large coupling constant between H1' and H2' protons. In the NMR experiments the corresponding cross peaks have been identified for A11, U18 and C19. However, in 4 out of the 5 finally generated NMR structures, the

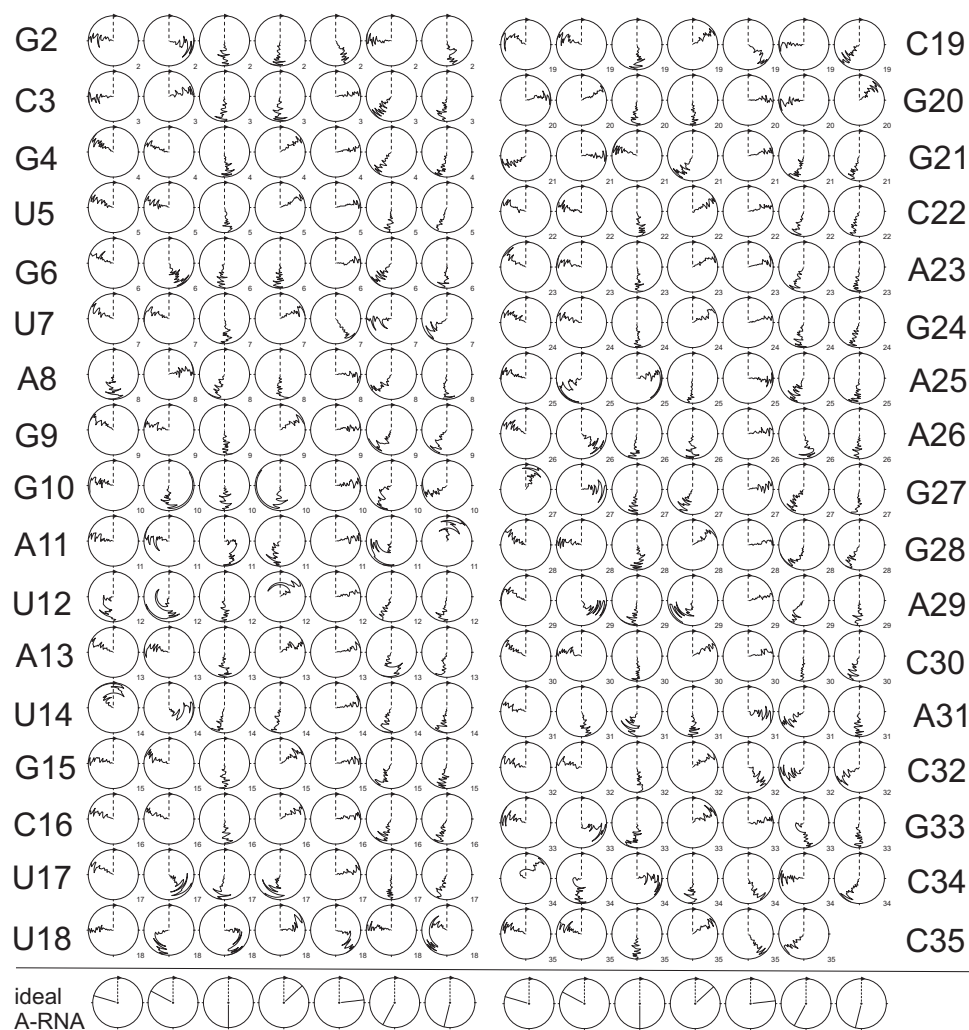


Figure 4.4: Time-course of the torsion angles $\zeta, \alpha, \beta, \gamma, \delta, \epsilon, \chi$ for residues G2 to C35 over the last 500 ps. The trajectory starts at the center of the circle (1200 ps) and ends at the circumference (1700 ps) and the angles are displayed in a clockwise direction between 0° and 360° . See Fig. 2.2 for torsion angles reference.

pucker phase angle of A11 is C3'-endo. In the MD simulation the preferred orientation of A11 is also C3'-endo. For the loop bases U18 and C19 the P values in the simulation agree with the NMR observations. The simulation predicts, however, that also U7, C32 and C34 assume the C2'-endo conformation. Here, it has to be pointed out that the simulation leads to a complete structural reorganization of the first and last 3 nucleotides of the sequence. This may be an artifact of the simulation. In addition to those cases where C3'-endo is the preferred conformation over the whole trajectory, there are events where the C2'-endo conformation is adopted for a short time. One example is the flexible nucleotide A31 showing puckering events with 55 ps

and 140 ps duration to the C2'-*endo* state. Moreover, C34 and C35 exhibit four correlated short repuckers from C2'-*endo* to C3'-*endo* within a 50 ps interval. The χ angle describes the torsion about the glycosidic bond between the ribose and the base (O4'-C1'-(N1/N9)-(C2/C4)). The standard nucleotide conformation is *anti* with χ angles between 110° and 180° . Surprisingly, A11 adopts a *syn* conformation which is known to be sterically hindered in a C3'-*endo* conformation [113, 62]. On the other hand, recently even *syn* pyrimidines have been found [98]. So, we predict that *syn* conformations can occur in RNA structures. However, the detailed origin of their occurrence remains to be clarified. As mentioned above, standard values of the torsion

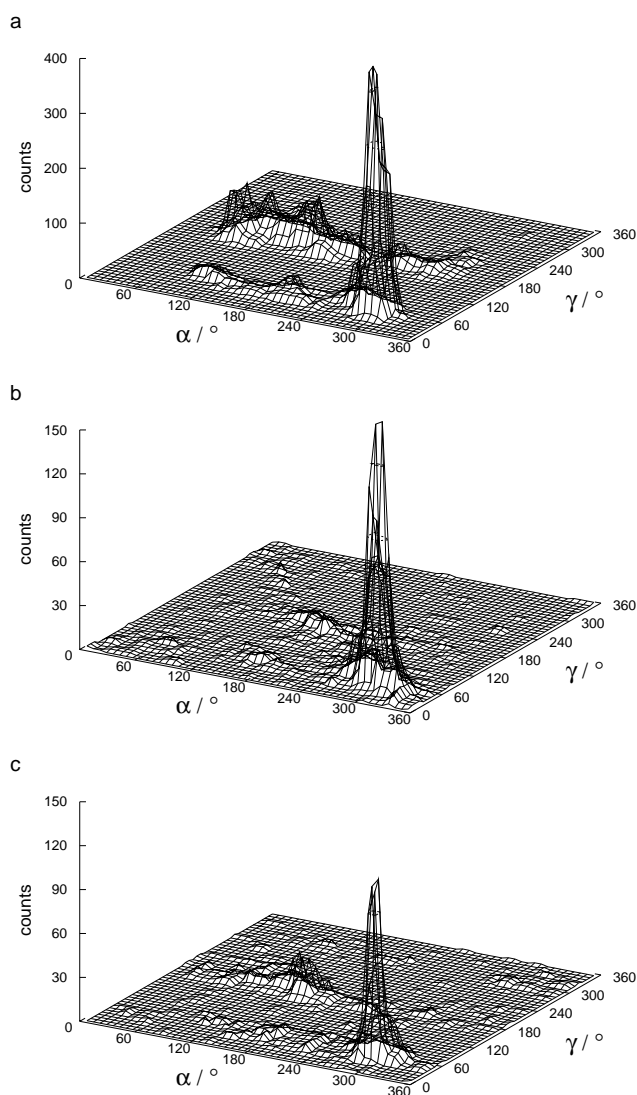


Figure 4.5: Two-dimensional distribution of the nucleic acid torsional angles α and γ . a) MD simulation of the FMN-RNA aptamer over the last 500 ps, b) 59 X-ray RNA structures from the NDB, for which torsion angle report files were available, c) NMR RNA aptamer structures from the PDB (PDB codes: 1am0, 1eht, 1fnn, 1koc, 1kod, 1raw, 1tob, 1ull; all models of individual structures were taken into account).

angles α and γ for an A-RNA in C3'-*endo* conformation are 292° and 54° , respectively, which corresponds to the '*-gauche*' and '*+gauche*' conformations. However, in the MD simulation a substantial number of cases is found where γ is in the trans conformation ($\approx 180^\circ$) and α varies over a large angle range (Fig. 4.5).

In these cases the distance $O2'(n) \cdots O5'(n+1)$ between neighbour nucleotides can be rather short. A survey of torsion angles in mono-, di- and trinucleotides has shown that indeed the great majority of γ values is around 60° (*+gauche*) [62]. In addition, however, a small number of cases with γ angles around 180° (*trans*) and 300° (*-gauche*) was found as well. The advantage of these small structures is their high resolution. On the other hand, there may be differences to large structures. We have, therefore, scanned 59 RNA structures for which the Nucleic Acid Database (NDB) has provided torsion angle reports [20] and the RNA aptamer entries of the Protein Data Bank (PDB codes: 1am0, 1eht, 1fmn, 1koc, 1kod, 1raw, 1tob, 1ull; together 50 models) [2]. The NDB only includes X-ray entries and the RNA aptamer structures available in the PDB at the time of analysis were exclusively NMR structures. The results are shown in Fig. 4.5. What we see is that for the X-ray RNA structures there are γ torsional angles around 180° even though its fraction is small. However, in the RNA aptamer structures there is a larger number of backbone orientations with γ around 180° . Examples for structures with γ in a *trans* conformation are a high-resolution RNA duplex X-ray structure (PDB code: 259d, NDB code: arh074; G13) [48], an NMR RNA hairpin (PDB code: 1uuu; A6, G8, G13, A15 of model 1) [130] and an NMR RNA aptamer structure (PDB code: 1raw; A10, A12-14, C15, G17, C18, ... of model 1) [47]. So far we do not know, whether the frequent occurrence of this unusual backbone conformation is a typical feature of RNA aptamer structures or nothing more than an artifact of structure determination by NMR or molecular dynamics. Finally this observation may also be a consequence of general differences between crystallized and solution structures.

4.1.3 FMN and its environment

The recognition specificity for FMN is associated with hydrogen bonding of the uracil like edge of the isoalloxazine ring ($O2/N3-H3$) to the Hoogsteen edge of an adenine at the intercalation site. There is extensive stacking between the isoalloxazine-A26 recognition alignment and the flanking G9-G27 mismatch and the G10-U12-A25 base triple in the complex. The time-course of hydrogen bond (H-bond) distances shown in Fig. 4.6ab indicates stable H-bonds in the FMN binding site between FMN and A26.

In addition, however, there is a movement of the FMN phosphate towards G27 which leads to two new H-bonds between $O5'/OxP(FMN)$ ($x=1,2$) and $N1/N2(G27)$ (Fig. 4.6cd and Fig. 4.7a). Note, however, that the side chain of FMN is not critical for binding, since the aromatic part of the molecule alone exhibits a similar affinity to the 35-mer RNA aptamer [24]. Contrary to the NMR refinement, the MD simulation takes water into account. Water molecules with long residence times are good candidates of 'structural water' with a particular structural or even functional importance. We have found two long-lived water molecules mediating either atoms within FMN or between FMN and RNA. One water bridges $O2(FMN)$ and $O4'(FMN)$

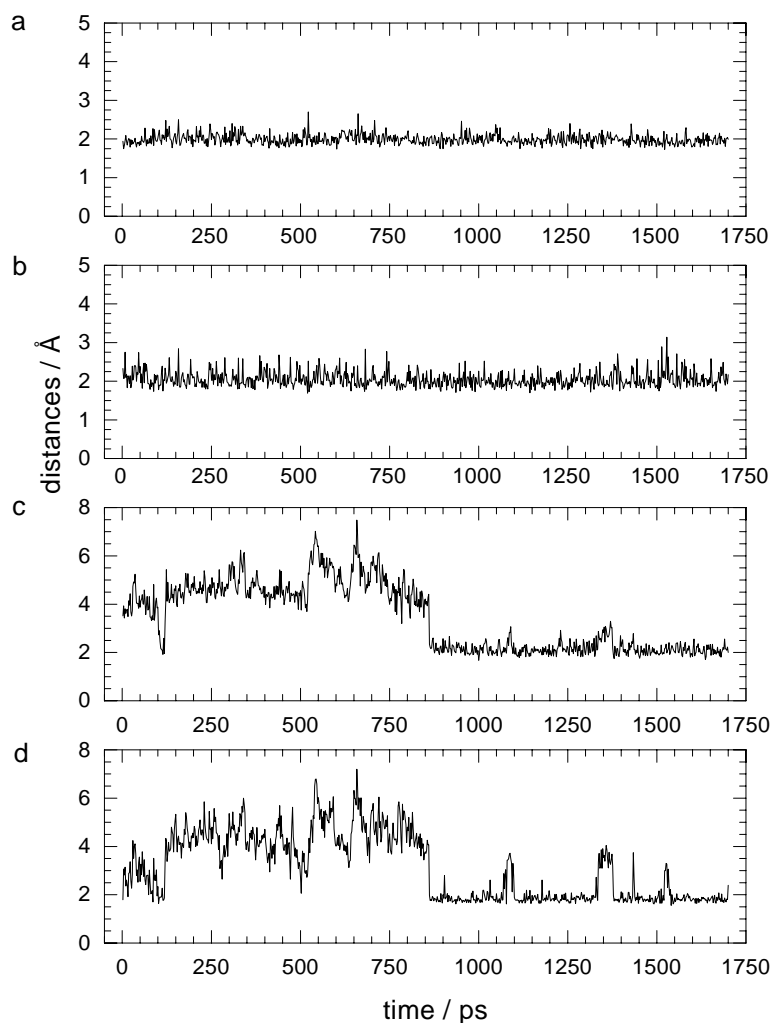


Figure 4.6: Time-course of hydrogen bond distances between FMN and A26, G27. *a) H3(FMN)···N7(A26), b) O2(FMN)···H61(A26), c) O5'(FMN)···H1(G27), d) O2P(FMN)···H21(G27).*

over 600 ps (Fig. 4.7ab-2) and the other one links O4(FMN) with O2P(A26) for about 460 ps (Fig. 4.7ab-3). From these results we conclude that water-mediated hydrogen bond interactions provide an additional means for stabilizing the FMN binding site.

4.1.4 The UUCG loop

Hairpins are a frequent element of RNA secondary structure [151]. They are thought to provide nucleation sites for RNA folding [138] and tertiary recognition sites for both proteins and nucleic acids [140]. The refinement protocol used by Fan et al. [55] for the determination of the loop structure in the FMN aptamer was based on a loop structure published in 1991 [141]. This structure was then allowed to move freely during the refinement process. Recently, the earlier loop structure has been corrected [5] and it has been shown that the hydrogen bonding pattern is different from the original structure. Miller et al. [105] have performed a MD study starting from the earlier structure to check whether the simulation leads to the correct structure,

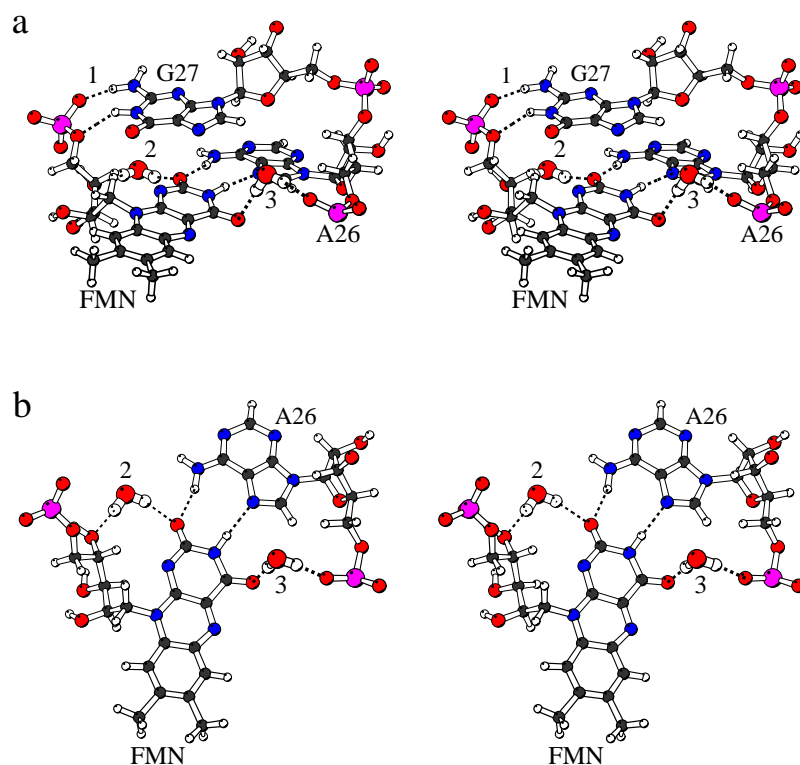


Figure 4.7: Stereo plots of the FMN binding site. Residue G9 was omitted for clarity. Dashed lines indicate hydrogen bonds. a) 1 shows the hydrogen bonds between FMN phosphate and H21/H1(G27); 2 shows a water bridge between O2 and O4' of FMN; 3 shows a water mediating O4(FMN) and O2P(A26). b) Top view of the FMN-A26 pair with the two long-lived water molecules).

but they did not succeed. However, they obtained a conversion from the earlier to the new loop conformation when the ribose sugars were replaced by deoxyribose. In Table 4.3 atom distances within the hairpin loop are compared for the simulated average structure of the aptamer, the starting structure (PDB code: 1fnn; model 1) and the corrected NMR loop structure (PDB code: 1hlx).

Our simulated structure shows both the bifurcated hydrogen bond between O2(U17) and H1/H21(G20) and the additional hydrogen bond between HO2'(U17) and O5'(U18) as found in the new loop structure. Nevertheless, there are also distances in the simulated structures more close to the NMR structure (O2'(U17), O6(G20)). In general, however, the loop region of the simulated average structure is similar to the corrected loop structure.

4.1.5 Orientation and donor and acceptor properties of the ribose 2'-OH group

The occurrence of the 2'-OH group in RNA is the only chemical difference between RNA and DNA within the backbone. The conventional functions ascribed to the 2'-OH group include stabilization of the sugar-phosphate backbone by intrastrand HO2'(n)···O4'(n+1) hydrogen bonds and locking of the sugar pucker [81]. For helices, it has been found that 2'-OH groups can be

Table 4.3: Comparison of distances and standard deviations (in Å) within the hairpin UUCG loop for the NMR aptamer structure family (PDB code: 1fnn), the last 500 ps of the simulation (MD) and the NMR loop structure family (NMR, PDB code: 1hlx).

atom 1	atom 2	NMR (1fnn)	MD	NMR (1hlx)
O2(U17)	N1(G20)	3.2 ± 0.3	2.9 ± 0.1	2.9 ± 0.1
O2(U17)	N2(G20)	3.2 ± 0.6	3.3 ± 0.4	3.0 ± 0.2
O4(U17)	N2(G20)	6.9 ± 0.2	6.7 ± 0.3	7.0 ± 0.2
O2'(U17)	O6(G20)	3.1 ± 0.1	3.4 ± 0.5	2.6 ± 0.1
O2'(U17)	O5'(U18)	4.3 ± 0.5	2.9 ± 0.3	2.9 ± 0.1
O1P(U18)	N4(C19)	3.4 ± 0.1	4.2 ± 0.7	4.5 ± 1.1

both stabilizing or destabilizing [146]. The destabilization is, however, only found for DNA-RNA hybrids and for a few extreme sequences and may thus be not typical for most of the other RNA structures. Recently, it has been shown in an analysis of a high-resolution X-ray structure of an RNA duplex that it is much more extensively hydrated than the corresponding DNA duplex. This hydration is primarily aided by 2'-OH-groups and, hence, indicates another role of the 2'-OH group [48]. We have performed a comprehensive analysis of possible interactions of the 2'-OH group within one residue, between neighbour and non-neighbour residues and with water. Our analysis has shown that the acceptor potential of O2' is less important for intramolecular interactions. Moreover, this simulation supports the view that inter-ribose stabilization via $O2'-H(n) \cdots O4'(n+1)$ hydrogen bonds is rather the exception than the rule. This is in line with MD results obtained by Auffinger and Westhof [13]. These authors have shown that for a C3'-endo conformation the preferred orientation of the 2'-OH hydrogen atom is towards the O3' atom of the same nucleotide. Two other possible but less frequent orientations are towards O4' and the base part. We have performed the same type of analysis for the aptamer structure for nucleotides adopting both C3'-endo and C2'-endo conformations (Fig. 4.8).

The results depend critically on the value of the backbone torsion angle γ of the next nucleotide. For the C3'-endo standard conformation with the γ angle in the +gauche range, the orientation of the 2'-OH group is similar to the results obtained by Auffinger and Westhof [13]. However, the preferred orientation is not directly towards O3' but slightly shifted towards O5'. This may be due to a repulsion between the hydrogen atom and C3' or a slight attraction with a phosphate oxygen O_xP. Moreover, the different sets of partial charges, used by Auffinger and Westhof [13] and in our simulation, may also lead to slightly different results. Nevertheless, the results are roughly similar. On the other hand, a completely different orientation pattern is ob-

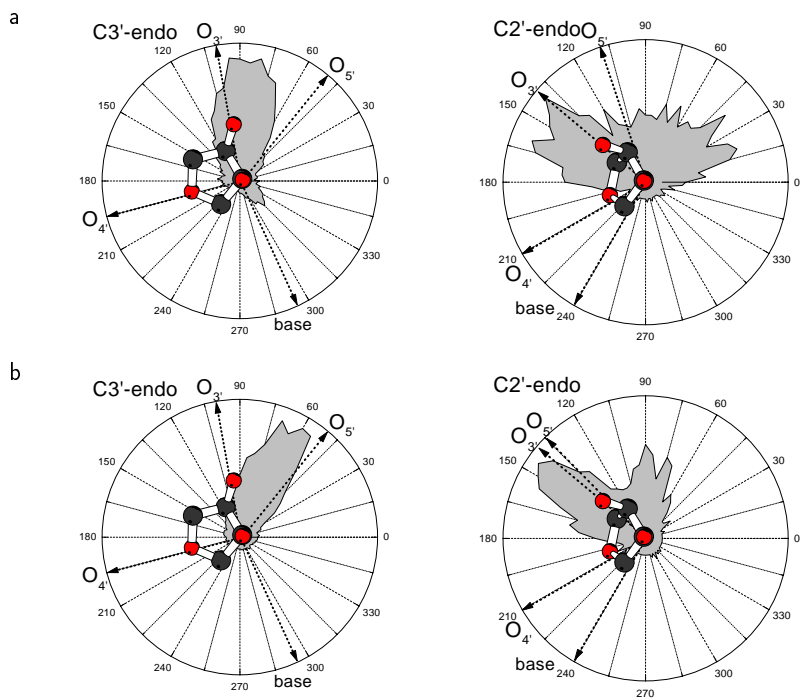


Figure 4.8: Orientation of the hydrogen atom of the 2'-OH group in C3'-endo and C2'-endo nucleotides of the FMN-RNA aptamer for the last 500 ps. a) γ angle of the next nucleotide in 5'-3' direction: +gauche (standard),

b) γ angle of the next nucleotide in 5'-3' direction: trans. The torsion angles have been measured with the atoms H2'-C2'-O2'-HO2'. The five residues 7, 18, 19, 32, 34 are in the C2'-endo conformation and all other nucleotides adopt the C3'-endo conformation. For the C3'-endo nucleotides the number of next-neighbour nucleotides with γ in +gauche or trans orientation is approximately equal. The C2'-O2' bond is oriented perpendicular to the paper plane.

tained for a backbone conformation with the γ angle in the trans range. In this case the preferred orientation of the 2'-OH hydrogen is not towards O3' of the same nucleotide but towards O5' of the next nucleotide. Again, the distribution maximum is not exactly located in the O5' direction but slightly moved away. For the standard backbone conformation (nucleotide n: C3'-endo, nucleotide n+1: γ in +gauche orientation) the next hydrogen bond acceptor is O3' of the same nucleotide. On the other hand for the unusual backbone conformation (nucleotide n: C3'-endo, nucleotide: n+1: γ in trans orientation) the O5' atom of the neighbour nucleotide has almost the same distance as the O3' atom towards the rotating hydrogen atom of O2'. For the rare cases of C2'-endo puckered riboses, the 2'-OH hydrogen atom orientation displays a broadened area with a peak in the O3' direction for γ in the +gauche range of the next nucleotide. Contrary to the C3'-endo case, for C2'-endo and γ of the next nucleotide in trans, the preferred orientation of the 2'-OH hydrogen atom is approximately towards O3'. In this case O5' lies behind O3' (Fig. 4.8). However, the results for the C2'-endo conformation were obtained from very few nucleotides. Therefore, the statistical significance is limited. As already noted, the unusual backbone conformation described above leads to a relatively short O2'(n)···O5'(n+1) distance,

comparable in length with the $O2'(n)\cdots O3'(n)$ (standard conformation). Therefore, $O5'$ is an additional candidate for accepting the donor potential of $2'$ -OH. Indeed, a considerable part of $HO2'(n)\cdots O5'(n+1)$ interactions fulfills the H-bond criteria used here. Due to unfavorable angle geometries this does not hold for the $HO2'(n)\cdots O3'(n)$ interaction.

Recently, NMR studies on orientation, dynamics and hydration of the $2'$ -OH group in RNA and DNA/RNA hybrids have been reported [67]. In that study it was concluded that the hydrogen atom does not point towards $O3'$, but to the hydrogen atom of $C1'$ and $C2'$. Hence, the

Table 4.4: *H-bond interactions between the RNA 2'-OH group and water for all nucleotides (HO2'-donor; O2'-acceptor). The values given are H-bond percentages calculated by dividing the total number of H-bonds in all frames through the number of frames. Values higher than 100% indicate more than one H-bond per nucleotide (distance $D\cdots A < 3.5 \text{ \AA}$, angle $D-H\cdots A > 135^\circ$).*

Residue	HO2'	O2'	Total	Residue	HO2'	O2'	total
G1	23.3	87.7	111.0	U18	41.7	101.7	143.4
G2	17.1	124.2	141.3	C19	43.2	83.7	126.9
C3	78.2	116.8	195.0	G20	54.6	120.8	175.4
G4	73.1	117.4	190.5	G21	42.7	117.2	159.9
U5	29.7	118.1	147.8	C22	76.0	119.9	195.9
G6	60.6	97.7	158.3	A23	79.3	126.6	205.9
U7	61.1	101.7	162.8	G24	12.1	58.6	70.7
A8	74.6	118.0	192.6	A25	72.7	136.6	209.3
G9	36.0	102.4	138.4	A26	25.7	126.5	152.2
G10	32.1	14.8	46.9	G27	68.4	121.8	190.2
A11	16.4	21.1	37.5	G28	42.8	128.5	171.3
U12	56.5	102.1	158.6	A29	74.8	120.6	195.4
A13	49.3	130.6	179.9	C30	40.1	130.5	170.6
U14	65.3	106.5	171.8	A31	23.5	80.8	104.3
G15	73.4	117.5	190.9	C32	16.8	49.4	66.2
C16	45.3	115.1	160.4	G33	20.8	80.1	100.9
U17	15.9	54.0	69.9	C34	60.1	71.8	131.9
				C35	97.4	109.5	206.9
Average					48.6	100.9	149.5

NMR and MD data yield contradicting results. Preliminary results of quantum-chemical ab initio calculations show that the orientations towards $O3'/O5'$ are energetically favoured by about

9 kcal/mol over an orientation towards ribose C1' (Meyer et al., unpublished results). The great majority of aptamer 2'-OH groups is exposed and, therefore, accessible to the solvent. In Table 4.4 on page 34 hydrogen bond percentages for the interaction with water are given both for the donor and acceptor function of the 2'-OH group. They were calculated by counting the total number of H-bonds in all frames, according to a particular geometrical criterion, divided by the number of frames. Values larger than 100% indicate the formation of more than one H-bond. The H-bond percentages averaged over all nucleotides are 100% for the acceptor function and 48% for the donor function of O2'. From a closer inspection of the data it turns out that the acceptor percentages of nucleotides 10, 11, 17, 24 and 32 are below 60% and thus significantly smaller than for the remaining nucleotides. There is an obvious explanation for this fact. The O2' atoms of these nucleotides are buried inside the RNA and form intramolecular H-bonds. Hence, their water accessibility is reduced. The conclusion from this analysis is that the great majority of 2'-OH groups is exposed and that all exposed 2'-OH groups are hydrated. In these cases the acceptor function is preferred. This latter result is especially interesting because X-ray studies do not provide information on hydrogen atom positions. Therefore, they give no information whether the 2'-OH group or water is the preferred donor in the 2'-OH-water interaction. However, from the statement made by Egli et al. [48], that water molecules are better acceptors than ribose O4' atoms, we conclude that these authors have assumed that the 2'-OH group primarily acts as donor towards water. As already noted, the MD simulations predict that in the 2'-OH-water interaction the water molecules are the preferred donors. We do not observe long-lived hydrogen bonds between water and 2'-hydroxyl groups except for one bridging O2'(G1) and O2P(A29). The residence times for water molecules at a 2'-OH group are usually below 200 ps. Recently water residence time of solvated azurin have been calculated by means of MD simulation [97]. They found a correlation between residence time τ and solvent accessible surface, where the mean residence times were all below 40 ps. But they also observed long residence times for crystallographic water with $\tau = 250$ ps.

4.1.6 Structural water

As already noted above water molecules with long residence times are good candidates for 'structural water' with a particular structural or even functional importance. We have, therefore, identified all water molecules with a residence time greater than 450 ps, Table 4.5.

In addition to the already mentioned cases within FMN and between FMN and RNA, there are a variety of further long-lived water molecules. The structural motifs indicate that in most cases the water is surrounded by more than two potential donor or acceptor atoms. Water can both act as donor and acceptor. The most long-lived waters interact with acceptor atoms in the RNA backbone. Nevertheless, there are also examples of water-mediated bases, either within

Table 4.5: Residues connected with long-lived water molecules. Hydrogen-bonding criteria (distance $D \cdots A < 3.5 \text{ \AA}$, angle $D - H \cdots A > 135^\circ$). Values in bold represent long-lived interactions for which the sum of effective residence time of all water atoms is longer than 450 ps.

Residue	atom	(H1/H2 or O)	Residue	atom	(H1/H2 or O)
WAT1766			WAT 2076		
G1	O6	450/610	A29	N1	606/478
C35	O4'	138/174	U7	O2	472/590
C34	O2'	188/44	C30	O2	164/176
G2	O6	94/106	A8	N7	28/24
WAT86			WAT2592		
A29	O2P	800/354	G1	N7	304/268
G1	N3	234/416	G4	O6	370/378
G2	N7	110/416	C3	N4H41	384 (O)
WAT474			WAT684		
A29	O2P	480/770	A13	O2P	542/48
G1	O2'	684/398	U12	O2P	8/510
G28	O3'	56/74	U12	O5'	-/80
G28	O5'	72/48	U12	O3'	54/2
WAT696			WAT737		
U5	O4	472/260	G2	O1P	208/426
G6	O6	294/502	G28	O2P	456/104
G6	N7	70/102	G1	O3'	120/296
G1	O5'	122/94	G28	N7	224/154
WAT2303					
A13	O2P	332/418			
A11	O2'	106/90			

a base pair (A29, U7) or between neighbour nucleotides within one strand (U5, G6; G1, G4). Whereas water molecules bridging backbone atoms are not very surprising, water-mediated base pairs are rather unusual. Recently, however, the first examples have been identified in experimental structures [77, 39]. See section 4.3 in this work.

In our simulation, the hydrogen bond between H3(U7) \cdots N1(A29) opens and a water molecule is inserted bridging O2(U7) and N1(A29). This is surprising, because a Watson-Crick base

pair is opened. It seems that the neighbouring backbone geometry enforces the A-U base pair opening. This loss of stability is compensated by the bridging water molecule and new hydrogen bonds between H62(A29)···O5'(G1), H6(U7)···O5'/O2P(U7) and N7(A29)···HO5'(G1). A general feature of the environment of long-lived 'structural' water are at least two and in most cases three or four potential acceptor atoms. In this way the water has various possibilities of forming hydrogen bonds and this contributes to the stability of the water positions and to the overall structure [148].

4.1.7 Backbone C-H···O interactions

Recently, it has been claimed that short C-H···O contacts may play a structural or functional role in biopolymer structures [143]. From an MD simulation of a tRNA loop and a subsequent analysis of X-ray structures, Auffinger and Westhof have concluded that there may be a stabilizing C2'(H)(n)···O4'(n+1) interaction in the RNA backbone [13]. A systematic search for short C-H···O contacts in RNA structure has identified the interaction C5'(H)(n+1)···O2'(n) as another frequent next-neighbour backbone-backbone contact with comparable geometrical properties [22]. We have therefore analyzed both interaction types in the aptamer structure (Fig. 4.9).

For the C2'(H)···O4' interaction we find results that are in line with the data obtained by Auffinger and Westhof [13]. Contrary to the C2' donor group C5' has two hydrogen atoms. The distance/angle distribution for C5'(H1)···O2' shows two maxima at 2.4-2.8 Å and 120-150° and at 4.5-6.0 Å and 20-50°. The great majority of these interactions is maintained during the simulation and flipping events are rare. The short distances are obtained for torsion angles $\alpha = 270-300^\circ$ (*-gauche*) and γ around 60° (*+gauche*) like in typical A-RNA. For the long distance maximum (4.5 - 6.0 Å) α may continuously vary between 60° and 300°. However, then γ can be around 180°, which leads to an O2'(n)···O5'(n+1) interaction. From the corresponding distribution for the C5'(H2)···O2' interaction it turns out that the overwhelming majority of these contacts is rather long with the hydrogen pointing away from O2'.

Finally, the C2'(H)···O4' distances are plotted versus the C5'(H1)···O2' distances. This plot clearly shows that in a large number of cases short C2'(H)···O4' distances are accompanied by short C5'(H)···O2' distances. In addition, short C2'(H)···O4' contacts may be accompanied by long C5'(H)···O2' distances, whereas the reverse is not true. In addition to the already known short C2'(H)···O4' contacts both an analysis of experimental structures and the MD data reported here have identified an additional C5'(H)···O2' interaction within the RNA backbone as promising candidates for an attractive hydrogen bond interaction. This is very interesting for an understanding of RNA structures in general. The C2'(H)···O4'/C5'(H)···O2' motif connects two segments of the backbone which form a seven-membered ring specific for

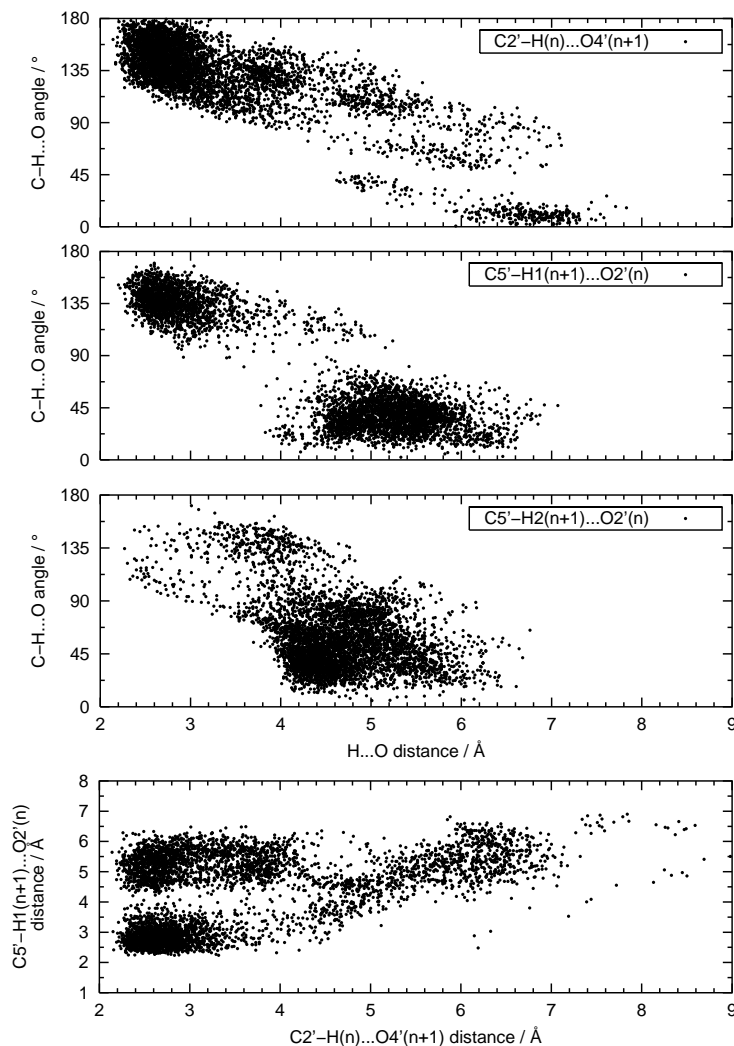


Figure 4.9: Distance/angle distributions of the interactions $C2'-H2'(n)\cdots O4'(n+1)$, $C5'-H1'(n+1)\cdots O2'(n)$ and $C5'-H2'(n+1)\cdots O2'(n)$ and the correlated distance/distance distributions for the interactions $C2'-H2'(n)\cdots O4'(n+1)$ versus $C5'-H1'(n+1)\cdots O2'(n)$.

RNA. It is, therefore, tempting to speculate that this motif contributes to the differences in the properties of RNA and DNA. The MD simulations show that both interactions are dynamically stable and can occur simultaneously at geometries for which a weak attractive hydrogen bond interaction is not impossible. Therefore, these interaction might affect both backbone rigidity and thermodynamic stability of RNA structures.

4.2 Coaxial stacking of a nicked RNA hairpin

4.2.1 Stability and flexibility

The stability of simulated structures can be evaluated by comparing their geometries with the geometry of the starting structure by means of the root mean square deviation (RMSd) of heavy atoms. The time course of the RMSd values for the four simulated systems ($n\text{Na}^+$, $n\text{Mg}^{2+}$, $c\text{Na}^+$, $c\text{Mg}^{2+}$) is shown in Fig 4.10.

In the final part of the simulation the RMSd values fluctuate about an almost constant average

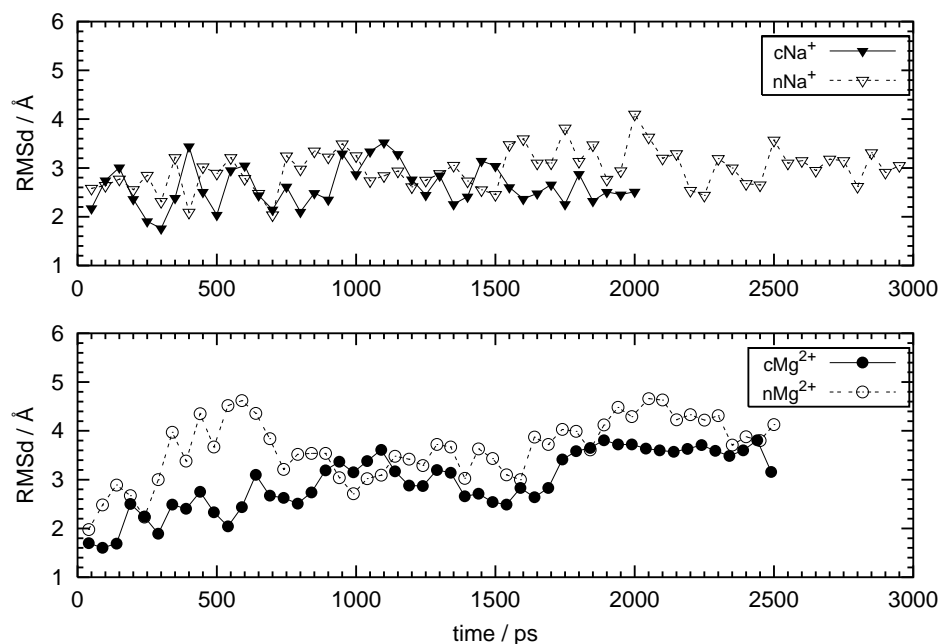


Figure 4.10: Time course of the root-mean-square deviation (RMSd) for the continuous ($c\text{Na}^+$, $c\text{Mg}^{2+}$) and nicked ($n\text{Na}^+$, $n\text{Mg}^{2+}$) hairpin structures with Na^+ and Mg^{2+} counterions.

value. This indicates that in all four cases the simulation leads to stable structures. In our analysis we focus first on differences between the simulations of the continuous and of the interrupted hairpin structure for the systems with Na^+ counterions. This system is directly related to the experimental conditions of the thermodynamic studies.

In Fig 4.11 the values of the root mean square fluctuations (RMSf) of the structure for the last 500 ps of the simulation time are shown. The RMSf values reveal the extent of motion of an atom around its equilibrium position. Hence, these data provide information on flexibility or rigidity of molecules or parts of them. For almost all nucleotides the peaks arise from the sugar-phosphate backbone. The increased flexibility of the backbone as compared to the bases is well-known from other MD simulations on nucleic acids [100, 126] and from temperature

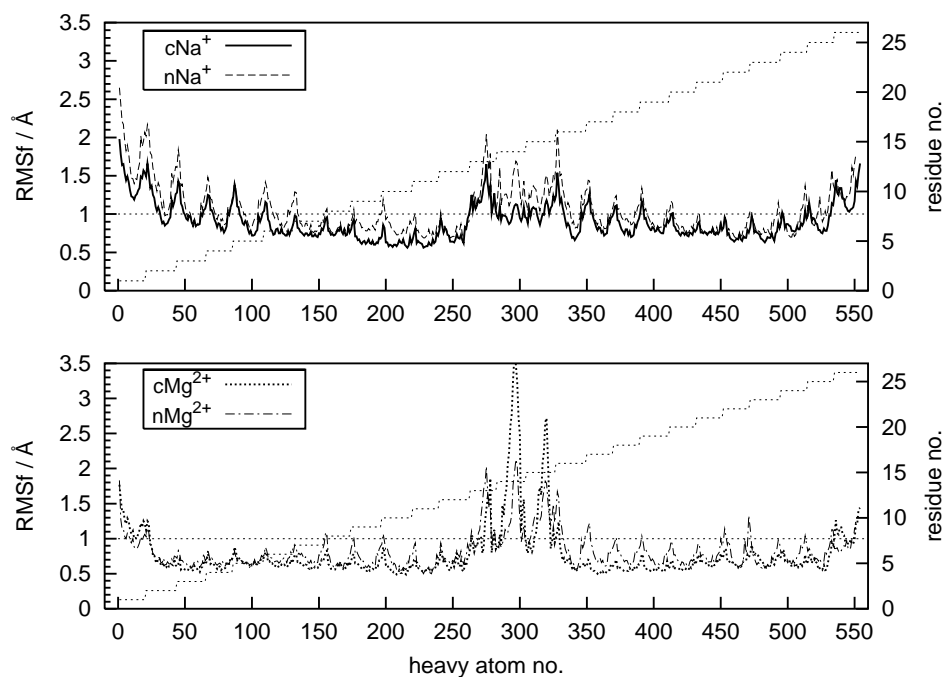


Figure 4.11: Root-mean-square fluctuations (RMSf) of heavy atoms relative to their average position over the last 500 ps for the continuous (cNa^+ , cMg^{2+}) and nicked (nNa^+ , nMg^{2+}) hairpin structures with Na^+ and Mg^{2+} counterions. Dotted steps indicate a residue (right axis). For each nucleotide first the atoms of the sugar-phosphate backbone and then the base atoms are given.

factors of X-ray structures. As expected the greatest flexibility is found for the end and loop regions of the hairpin structure.

On the other hand, the stem is relatively rigid. In the context of coaxial stacking, it is important to identify possible differences in flexibility between the continuous and nicked structures. From the data shown in Fig 4.11 for nNa^+ and cNa^+ it is evident these differences are small. Interestingly, the RMSf values of the nucleotides forming the stacking interface (nucleotides 4,5,22,23) are especially small. In the remaining part of the molecule the flexibility of the nicked structure is generally slightly larger as compared to the continuous hairpin. However, this difference is smaller than 0.3 \AA in the stem region and only slightly increased in the loop and end regions. This means that according to the MD simulations the lack of the phosphate group in the nicked structure does not lead to a substantially increased flexibility. These results are in line with the conclusion drawn from the gel electrophoresis and crosslinking studies [106, 37] but contrast with the NMR and restrained MD investigations by Roll [121].

4.2.2 Helical parameters

In order to detect structural differences between the four simulated systems we have performed a comprehensive analysis of geometrical base pair, base step and backbone parameters. The backbone parameters did not show any marked differences between the continuous and nicked hairpin structures. Hence, only data for rotational and translational helical parameters are displayed in Figs 4.12 and 4.13.

The opening found in the G2-C25 base pair of the cNa⁺ average structure is reflected in many helix parameters. It is, however, ignored in the following discussion. The great majority of helical parameters is similar for the continuous (cNa⁺) and nicked (nNa⁺) structures. A marked local effect at the stacking interface is seen for the base step parameters slide, shift and twist. All three are reduced in the nicked structure. In addition, the propeller twist for the two base pairs at the interface exhibits a slightly increased angle in the nicked structure. Recently, it has been pointed out that different algorithms for analyzing nucleic acid conformations may yield contradictory results [96]. We have thus compared the global CURVES results to local CURVES and FREEHELIX parameters [46]. Except for the slide the local CURVES and the FREEHELIX results are similar to the differences in the helical parameters calculated with the global CURVES approach. From Figs 4.12 and 4.13 it can be further seen that for the nicked structure the Y displacement, inclination and tip angles vary between the overhang and the remaining part of the hairpin. For example, in case of the nicked structure the Y displacement decreases at the stacking interface and remains then almost constant for the remaining part of the structure. The FREEHELIX Y displacement shows the same behaviour in a qualitative sense. However, in this case the Y displacement value is larger for the overhang part of the nicked structure analysis and almost identical to the continuous structure in the remaining part of the hairpin. On the contrary, the FREEHELIX analysis yields no significant differences between the continuous and nicked structures for the tip angle. Finally, the inclination value is slightly larger for all base pairs of the the nicked structure. A large roll angle is often related to a bend of the helical axis [46]. We see large roll parameters for the first three base pair steps (Fig 4.12). On the other hand, there are no significant differences in the local CURVES and FREEHELIX roll angles of the nicked and continuous structures. According to the local CURVES and FREEHELIX there is no significant difference between the rise of the nicked and continuous structures. The global CURVES data shown in Fig 4.13 seem to indicate that the rise is slightly increased in the overhang part of the nicked structure. We conclude from the analysis that the reduced twist and shift as well as the slightly increased propeller twist are the primary helical parameters describing the structural difference between the nicked and continuous structures.

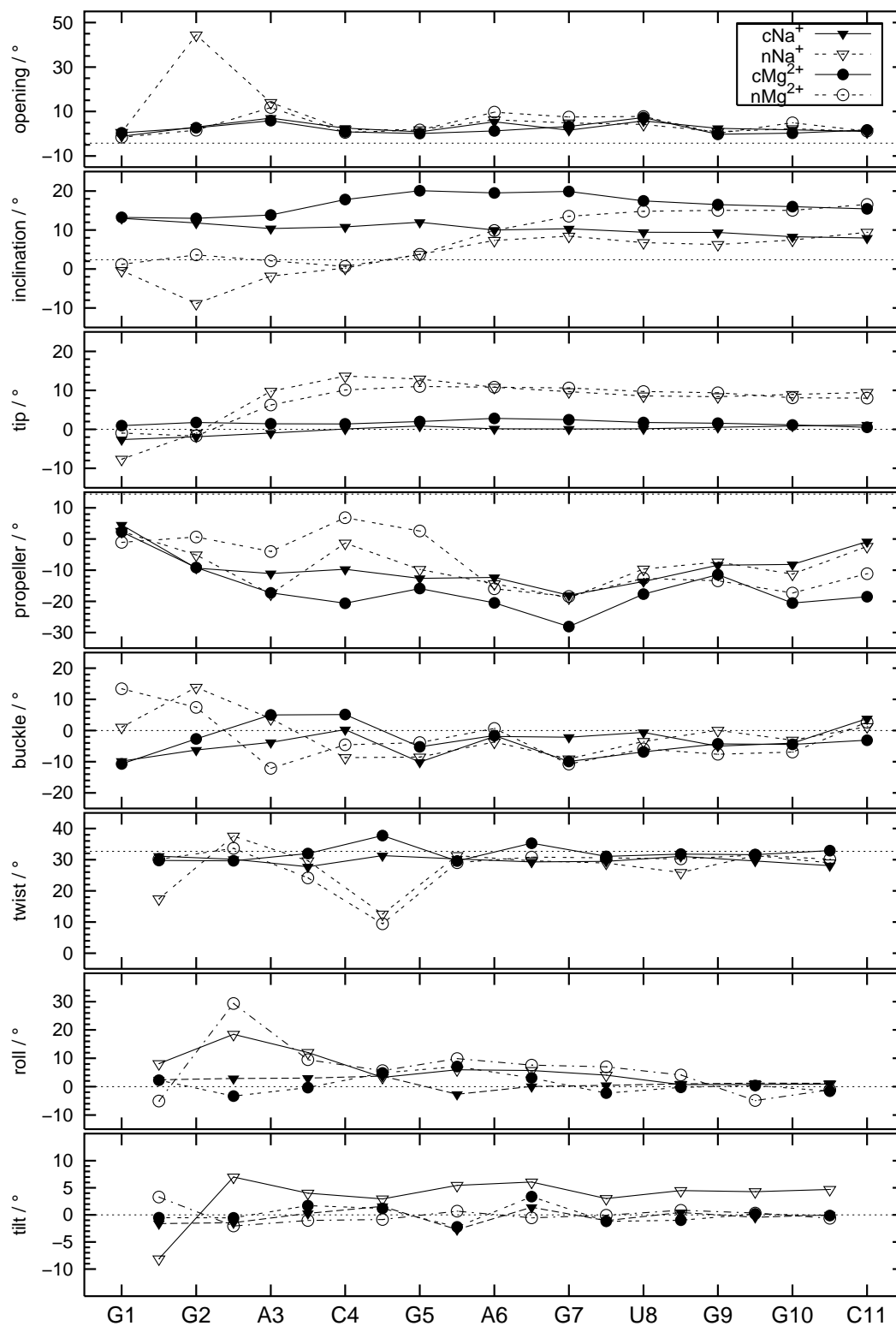


Figure 4.12: Global rotational helix parameters for the simulated average structures (last 500 ps) determined with CURVES: continuous (cNa^+ , cMg^{2+}) and nicked (nNa^+ , nMg^{2+}) hairpin structures with Na^+ and Mg^{2+} counterions. The dotted lines indicate values of the helical parameters for a canonical A-RNA.

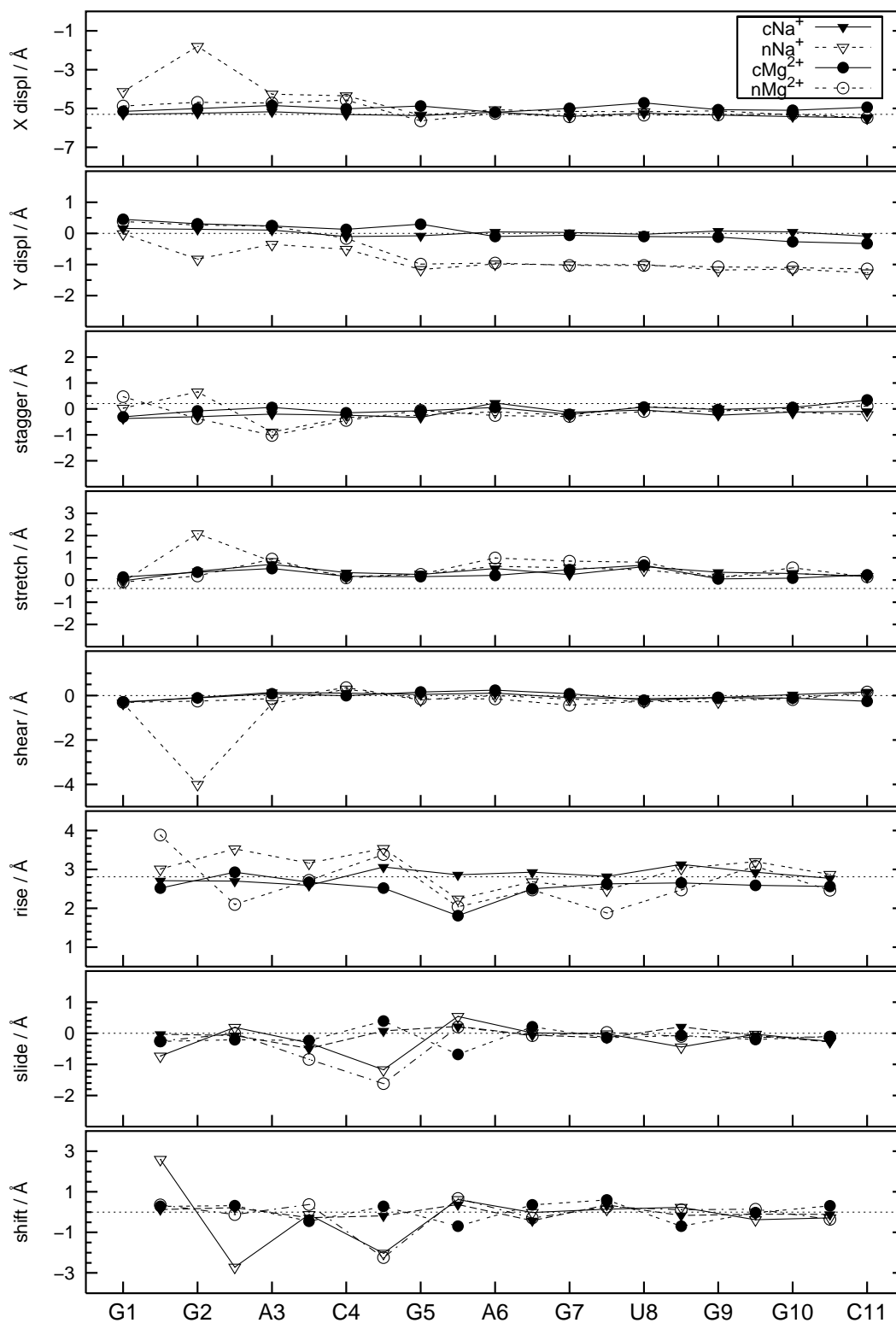


Figure 4.13: Global translational helix parameters for the simulated average structures (last 500 ps) determined with CURVES: continuous (cNa^+ , cMg^{2+}) and nicked (nNa^+ , nMg^{2+}) hairpin structures with Na^+ and Mg^{2+} counterions. The dotted lines indicate values of the helical parameters for a canonical A-RNA.

The analysis of helical parameters with CURVES and FREEHELIX has been supplemented by a further direct geometrical analysis. In Fig 4.14 a superposition of the base pairs G5-C22/C4-G23, which form the stacking interface in the nicked structure, is shown for the continuous (Fig 4.14a) and nicked (Fig 4.14b) structures.

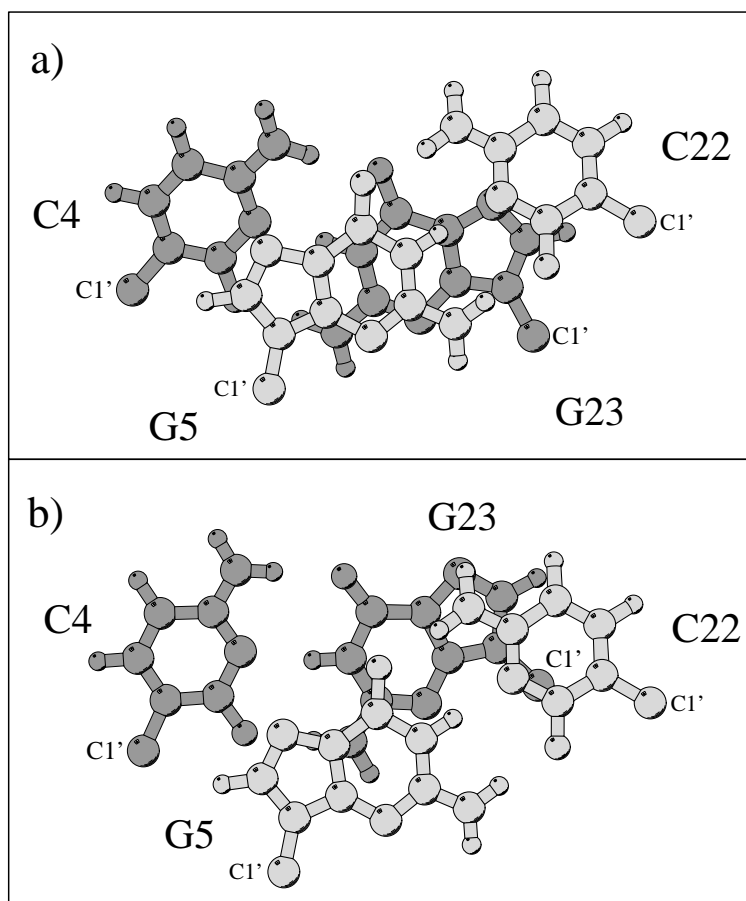


Figure 4.14: *Top view of the base pairs G5-C22 (light grey) and C4-G23 (dark grey) for the continuous (a) and nicked (b) hairpin structures (Na^+ counterions).*

The local geometry of the G5-C22 and C4-G23 base pair is practically identical for the continuous and interrupted hairpins. However, their relative orientation is substantially different. Whereas, the C1' positions of C4 differ only slightly, the C1' position of G23 in the nicked structure is shifted as compared to the uninterrupted structure. It moves towards the projection of the C22:C1' atom onto the C4-G23 base plane. In addition, it can be seen that the cytosine amino group is located approximately above the center of the G23 five-membered ring. Taken all facts together the stacking interface in the coaxially stacked nicked hairpin structure is characterized by a reduced twist and shift, and a slightly increased propeller twist. The unusual small twist and shift parameters lead to an improved overlap between C22 and G23 in the nicked structure.

The local geometric changes around the stacking interface result in a global structural effect (Fig 4.15). The continuous hairpin has a more or less straight helical axis (kink angle: 5°). On



Figure 4.15: Superposition of ribbon drawings and helical axes for the continuous and nicked hairpin structure from the Na^+ simulation ($c\text{Na}^+$: light grey; $n\text{Na}^+$: dark grey).

the other hand, the corresponding axis for the nicked structure exhibits a marked kink of 39° . A kinked helical axis with a kink angle of $(15 \pm 15)^\circ$ was also found in the gene 32 messenger RNA pseudoknot of bacteriophage T2 [79]. These authors have analyzed the possible bending of the quasi-continuous helix in other structures with a coaxial stacking motif, including the hammerhead ribozyme [119] (PDB code: 1hnh) and tRNA^{Phe,Asp} [149].

The kink angles were found to vary between 4° and 26° . This clearly shows that all experimental structures with the coaxial stacking motif are characterized by a bending of the coaxially stacked helical stems. The extent of bending and the structural details of the stacking interface may, however, be affected by local variations like Watson-Crick or non-Watson-Crick base pairs or even additional unpaired nucleotides at the stacking interface or by effects from the global nucleic acid environment. For example, in the T2 pseudoknot structure the stacking interface is formed by the base pairs U28-G16 and C7-A15 [79]. In this case the C7-U28 P-P distance and the C1'-C1' distance as well are increased as compared to an ideal A-RNA structure. Otherwise, there would be a clash between the phosphate groups of A8 and U28. By contrast, in the nicked hairpin the C1'-C1' distance between C22 and G23 is decreased as

compared to the continuous structures. The comparison of a nicked structure with the related continuous hairpin allows to derive information on the structural role of phosphate groups in RNA. As outlined above, the absence of the phosphate group in the nick hairpin has structural consequences but does not affect the RNA structure in a dramatic way. This means that obviously the hydrogen bond and stacking interactions are strong enough to keep the RNA structure in shape. It is well-known that in many cases the geometries of isolated base pairs calculated with quantum-chemical methods are in good agreement with experimental base pair geometries found in nucleic acid structures [104, 74, 23]. From this fact it can be concluded that, at least in regular helical regions, the sugar-phosphate backbone does almost not affect the base pair geometry. By contrast, theoretical studies on stacked base dimers yield geometries that do not agree with the structures found in nucleic acids [74]. Thus, either the geometries of stacked base pairs are different to the corresponding geometries of bases or the sugar-phosphate backbone affects the stacking geometry substantially. We find that the rise in the stacking interface does not substantially differ from other base steps. This means that the distance between the planes of the stacked base pairs is not significantly affected by the removal of a phosphate group in one strand. On the other hand, a major change in the relative orientation of the base pairs in the stacking interface is found. Hence, it can be concluded that an intact sugar-phosphate backbone enforces the base pairs to adopt a geometry that differs from the optimal geometry of stacked base pairs without backbone restraints.

4.2.3 Comparison of simulations with different counterions

The thermodynamic studies were performed in a sodium cacodylate/EDTA buffer [145]. Therefore, this system did not contain Mg^{2+} ions. On the other hand, the special role of Mg^{2+} ions for RNA structure and dynamics is well-known [108, 56]. To this end we have performed simulations both with Na^+ and Mg^{2+} as counterions. For a statistically meaningful representation of ion distributions around nucleic acids long simulation times are required. This is a particular problem for a divalent cation like Mg^{2+} because of its low diffusion coefficient as compared to Na^+ or Cl^- . For that reason only very few MD simulations with Mg^{2+} as counterions have been performed so far [100, 137, 156, 135, 69]. We have performed 2.5 ns MD simulations of the continuous and nicked hairpin structures with Mg^{2+} as counterions. To our knowledge this is the longest MD simulation time for an RNA/ Mg^{2+} system used thus far. Nevertheless, we are aware of the possible limitations of this approach and consider the ion distribution around RNA obtained from the simulations as a snapshot that is not necessarily identical to the distribution arrived at much longer times. An advantage of the simulated systems is that one can directly study the effect of counterions keeping all other properties constant. By contrast structural consequences of different counterions in experimental structures are often derived from

a comparison to ideal structures. The following discussion is primarily focussed on the continuous hairpin structures simulated with Na^+ and Mg^{2+} as counterions (cNa^+ , cMg^{2+}). Fig. 4.10 indicates that both the Mg^{2+} and the Na^+ simulations lead to stable structures. The Mg^{2+} simulations show a larger RMSd value as the Na^+ systems relative to the start structure. It should be noted, however, that the starting geometry does not correspond to an experimental structure. For the majority of nucleotides the Mg^{2+} simulations lead to smaller RMSf values than the Na^+ ones indicating a greater rigidity of the structures (Fig. 4.11). A similar result was obtained for DNA [100]. This phenomenon is, however, reversed for the loop nucleotides C13, A14 and A15, because the adenine stack between A14 and A15 that stabilizes the loop is here disrupted in the Mg^{2+} simulations. The helical parameters of the Mg^{2+} structures are roughly similar to the Na^+ data (Figs. 4.12 and 4.13). An exception is the opening of the base pair G2-C25 found in the nNa^+ simulation but not in the Mg^{2+} case. The structures from the Mg^{2+} simulation are more kinked than their Na^+ counterparts. This effect is pronounced in the continuous hairpin where the kink angles between the two stem parts are 5° in the Na^+ simulation and 17° with Mg^{2+} as counterions. The nicked structure is already substantially kinked with Na^+ ions (39°). In this case Mg^{2+} has only a minor effect (43°). It is interesting to note that a Mg^{2+} induced kink was also found in an experimental B-DNA structure [107]. Recently, it has been pointed out that Mg^{2+} ions may narrow the major groove width in RNA structures [39]. We have, therefore, measured the major groove widths of the nicked and continuous hairpin structures obtained from the Mg^{2+} and the Na^+ simulations as described in the Methods section (Fig. 4.16).

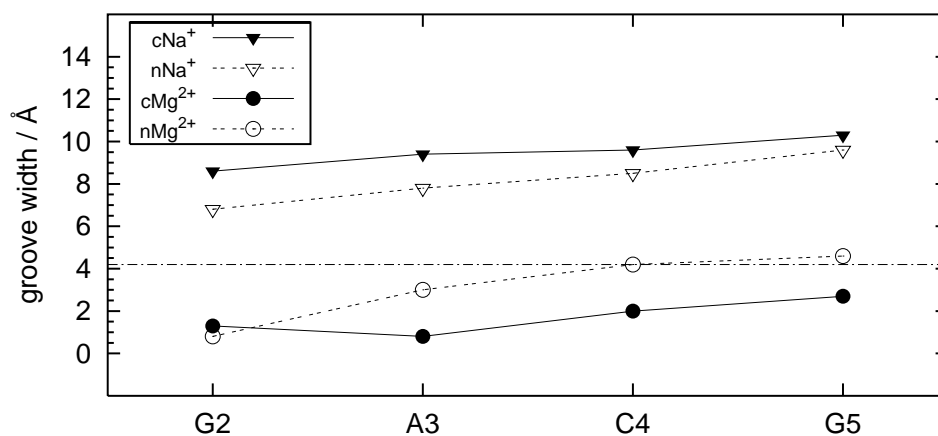


Figure 4.16: Widths of the major groove for the cNa^+ , nNa^+ , cMg^{2+} , nMg^{2+} simulations. The groove widths correspond to the shortest distances between phosphor atoms P_i and $P_{i'+x}$, where i and i' indicate nucleotides belonging to one base pair and x may vary between 5 and 8, less twice the van der Waals radius of P ($2 \times 2.9 \text{ \AA}$).

It turns out that the major groove width obtained from the Na^+ simulation is larger than for an ideal A-RNA. Interestingly, however, Mg^{2+} leads to a substantial narrowing of the major groove both for the continuous and nicked hairpins. This is in line with findings from experimental structures [39]. In summary, we find in the simulations that the replacement of Na^+ by Mg^{2+} counterions has structural consequences for the continuous hairpin including an increased kink of the helical axis and a narrowing of the major groove.

Experiments on the ion distribution in the minor groove of DNA resulted in different conclusions. Chiu et al. found no evidence for partial ion substitution for solvent within the minor groove spine of hydration [36], while Shiu et al. detected two splines, where the primary spline is partially occupied by sodium ions [129]. In MD simulations it has been shown, that sodium ions can intrude into the spline of hydration [155].

We have analyzed the distribution of Na^+ and Mg^{2+} ions in a volume defined by a 5 Å distance from any of the RNA atoms of the average structures (Figs. 4.17 and 4.18). These data were taken from the last 500 ps of the trajectories recorded.

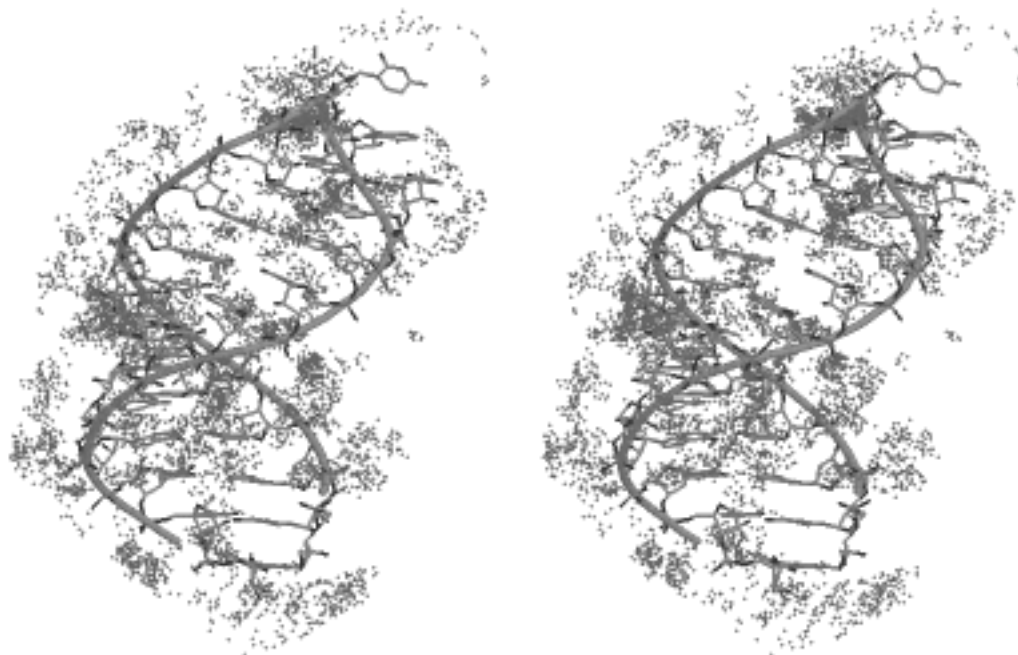


Figure 4.17: Stereoplot of the distribution of Na^+ ions around the continuous hairpin structure ($n\text{Na}^+$). Ions with a distance larger than 5 Å to RNA atoms are not shown.

It is obvious that the Na^+ ions show a relatively diffuse distribution, whereas the Mg^{2+} ions are more localized. This is in line with the higher diffusion coefficient of Na^+ as compared to Mg^{2+} . Calculated values are 10^{-5} cm²/s for Na^+ ($c\text{Na}^+$) and 0.5×10^{-5} cm²/s for Mg^{2+} ($c\text{Mg}^{2+}$). These values are smaller than the experimental bulk data (Na^+ : 1.3×10^{-5} cm²/s; Mg^{2+} : 0.7×10^{-5} cm²/s) [95] but show the same ordering. Moreover, it is well-known that the

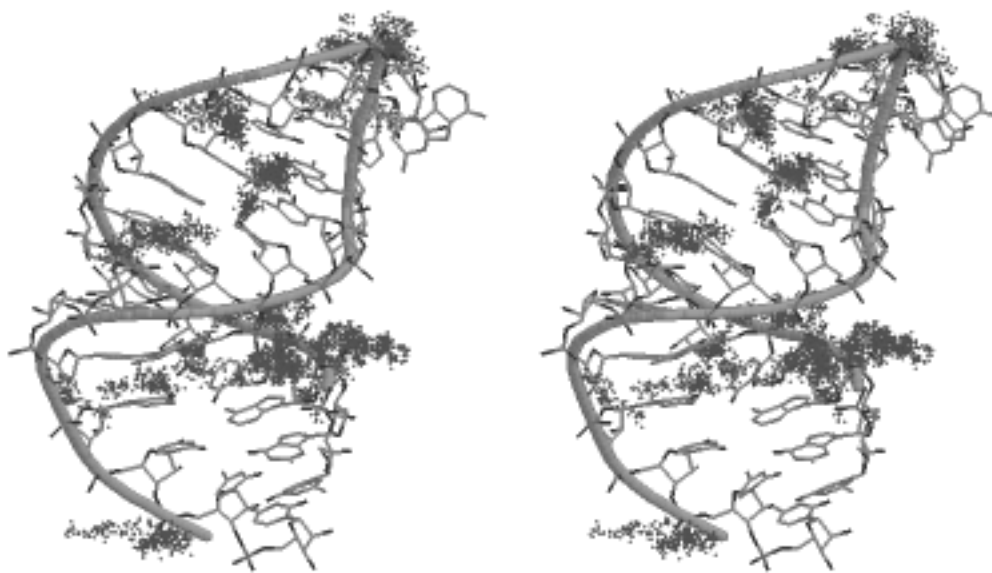


Figure 4.18: Stereoplots of the distribution of Mg^{2+} ions around the continuous hairpin structure (nMg^+). Ions with a distance larger than 5 Å to RNA atoms are not shown.

diffusion in the direct vicinity of biopolymers is lower as in the bulk [101]. Despite these facts it is not unlikely that the simulation times are too short for a statistically significant distribution of Mg^{2+} ions. Hence the distribution shown in Fig. 4.18 should be viewed with caution.

In order to get an impression on the movement of individual metal ions during the simulation in Fig. 4.19 the time course of the ion/RNA distance over the last 500 ps of the simulation time is shown for selected Mg^{2+} /RNA and Na^+ /RNA interactions. As noted above the Na^+ /RNA distance shows rather large fluctuations. On the other hand, the Mg^{2+} ions are tightly connected to a specific binding site. The Mg^{2+} /RNA distances are in the range between 1.9 Å and 2.4 Å for a direct contact and between 4 Å and 5 Å for a water-mediated interaction. The extent of fluctuations is comparable for the direct and water-mediated interaction of Mg^{2+} with phosphate oxygen atoms. On the other hand, the average distance of the interaction between Mg^{2+} and the base atoms O6 of G16 is slightly larger and also the fluctuations are more pronounced in this case. We have also analyzed the ion/RNA distances for the total simulation time and get the same results as shown for the last 500 ps. In other words, the Mg^{2+} ions are placed according to the electrostatic potential of RNA in the setup procedure and do not change their average position during the total simulation time of 2.5 ns. Calculations of the electrostatic surface potential have shown that in A-form nucleic acids the major groove is more negative than the minor groove [35]. One would thus expect that Mg^{2+} ions prefer major groove binding sites. It has been found by experimental studies and Monte Carlo simulations that the affinity of DNA for cation binding decreases in the order $Ca^{2+} > Mg^{2+} \gg Na^+$ [90]. In addition,

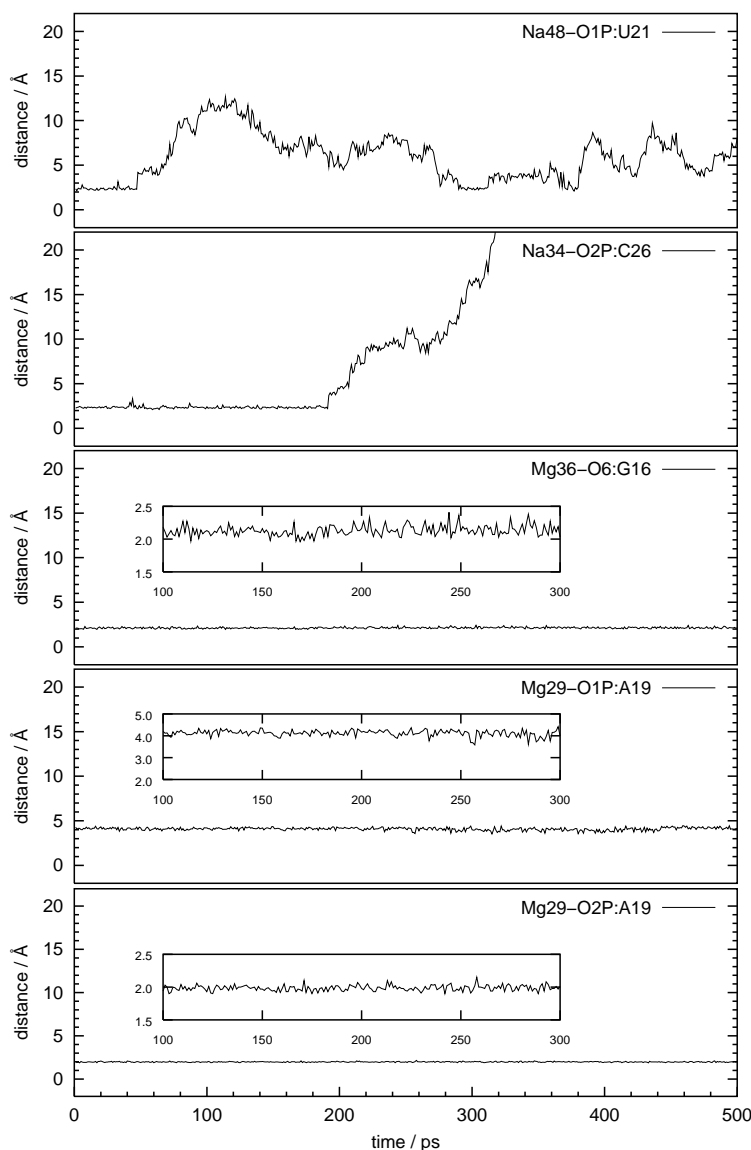


Figure 4.19: Time course of selected Na^+ /RNA and Mg^{2+} /RNA distances over the last 500 ps of simulation time for the continuous hairpin structures. The two Na^+ examples include interactions with phosphate oxygen atoms. Among the three Mg^{2+} /RNA interactions two represent direct contacts ($\text{Mg}36/\text{G}16:\text{O}6$; $\text{Mg}29/\text{A}19:\text{O}2\text{P}$) and one corresponds to a water-mediated link ($\text{Mg}29/\text{A}19:\text{O}1\text{P}$).

MD simulations of DNA with Li^+ , Na^+ , and Cs^+ counterions have revealed that the preferred binding sites are dependent on the ion type [99]. Cs^+ ions bind directly to the bases in the minor groove. Na^+ ions also bind preferentially to bases in the minor groove through one water molecule, however. On the contrary, the preferred binding sites of the Li^+ ions are the backbone phosphate groups. In several crystal structures metal binding sites have been identified. A discussion of relevant structures is provided in the reviews by Misra and Draper [108] and by Feig and Uhlenbeck [56]. However, the crystallographic studies do not yield exhaustive information on all metal binding sites. Moreover, the number of RNA structures determined with nuclear magnetic resonance (NMR) spectroscopy is rapidly increasing. This technique does usually not yield direct information on metal ion binding. Therefore, theoretical approaches can complement the experimental data in a useful manner [72]. We have found only one Mg^{2+} ion (not

shown in Fig. 4.18) that is located more than 5 Å away from the RNA and could thus be classified as an example of diffuse binding according to the terminology of Misra and Draper [108]. The other 11 Mg^{2+} ions show either inner or outer sphere binding primarily in the major groove of the RNA hairpin. There is only one base/ Mg^{2+} ion binding site. In this case the metal ion is bound to O6 of C16 via a direct and to N7 via a water-mediated contact. The preferred Mg^{2+} ion binding sites are, however, the phosphate oxygens pointing towards the major groove. The typical pattern is one direct contact accompanied by an additional water-mediated interaction. There are two different binding motifs: a) interaction with the two phosphate oxygen atoms of one and the same nucleotide (G5, G7, G10, C13 (loop), A14, C20, C22, C25); b) interaction with two phosphate oxygens of one nucleotide and two additional water-mediated contacts with a nucleotide on the same side of the major groove and across the groove (A3/C18/A19, A3/C17/C18). We do not see the clamp motif with two direct Mg^{2+} /RNA interactions across the major groove [51]. This is not surprising, however, because this motif is found at a site with two bulged out bases. We cannot exclude that the Mg^{2+} ions will change their average positions after longer simulation times. However, at least for 2.5 ns, they remain in their starting positions. This means that the ion placement procedure during the system setup is of utmost importance for the outcome of the simulations. It remains to be clarified whether the standard ion placement procedure in MD simulations is sufficient or other approaches for the prediction of magnesium binding sites in RNA are required [72].

4.3 WUC base pairs in an RNA duplex

4.3.1 Dynamical features of WUC

We have first analyzed the time course of the RMSd of the simulated structure as compared to the starting geometry (Fig. 4.20). The RMSd value oscillates around a stable mean value of

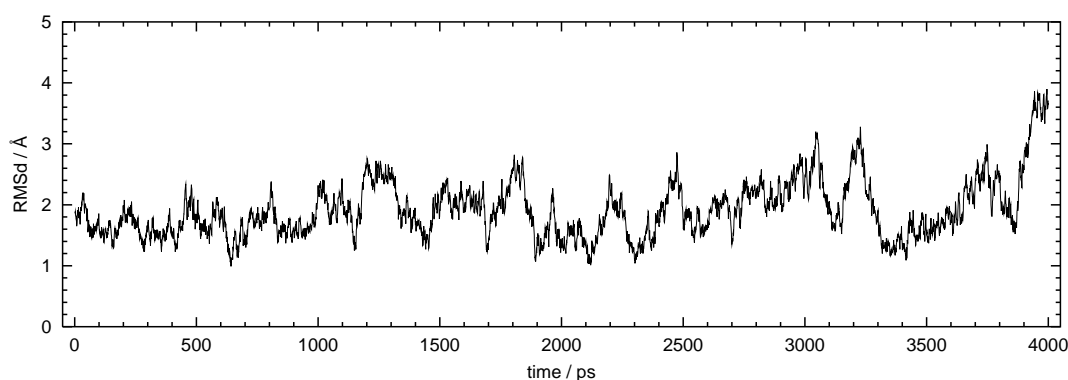


Figure 4.20: Time course of the root-mean-square deviation (RMSd) of the trajectory against the starting structure.

2 Å. This value indicates that the overall geometry of the complex remains more or less stable during the simulation. It is assumed that the increase of the RMSd value found at the end of the simulation is a temporary event.

In order to study the structure and dynamics of base pairs we have analyzed the time course of various inter-base distances. In Fig. 4.21 on page 53 results for the two WUC pairs (U6-C19, C7-U18) and for the flanking GU base pairs (U5-G20, G8-U17) are shown. For the GU pairs the H1(G)···O2(U) and the H3(U)···O6(G) distances of the direct inter-base H-bonds are displayed. For the WUC pairs the inter-base distances H42(C)···O4(U) and H3(U)···C(N3) are monitored. The GU pairs exhibit the stable behavior typical for base pairs with two or three direct H-bonds. The H-bond distances fluctuate around a value of about 2 Å which is in agreement with values found in X-ray and NMR structures. Occasionally, very short base pair opening events are found. The only longer opening event occurs for the G8-U17 pair in the time range between 1000 and 1200 ps.

By contrast, the WUC pairs shows a completely different dynamics. They adopt different conformations that are stable over simulation times between 500 and 1500 ps and switch between these conformations from time to time. In the time range between 800 and 4000 ps altogether four conformations can be identified. Conformation I is characterized by a direct H-bond between one hydrogen atom of the cytosine amino group and the O4 atom of uracil with an average distance of about 2 Å. The distance between the water donor and acceptor sites

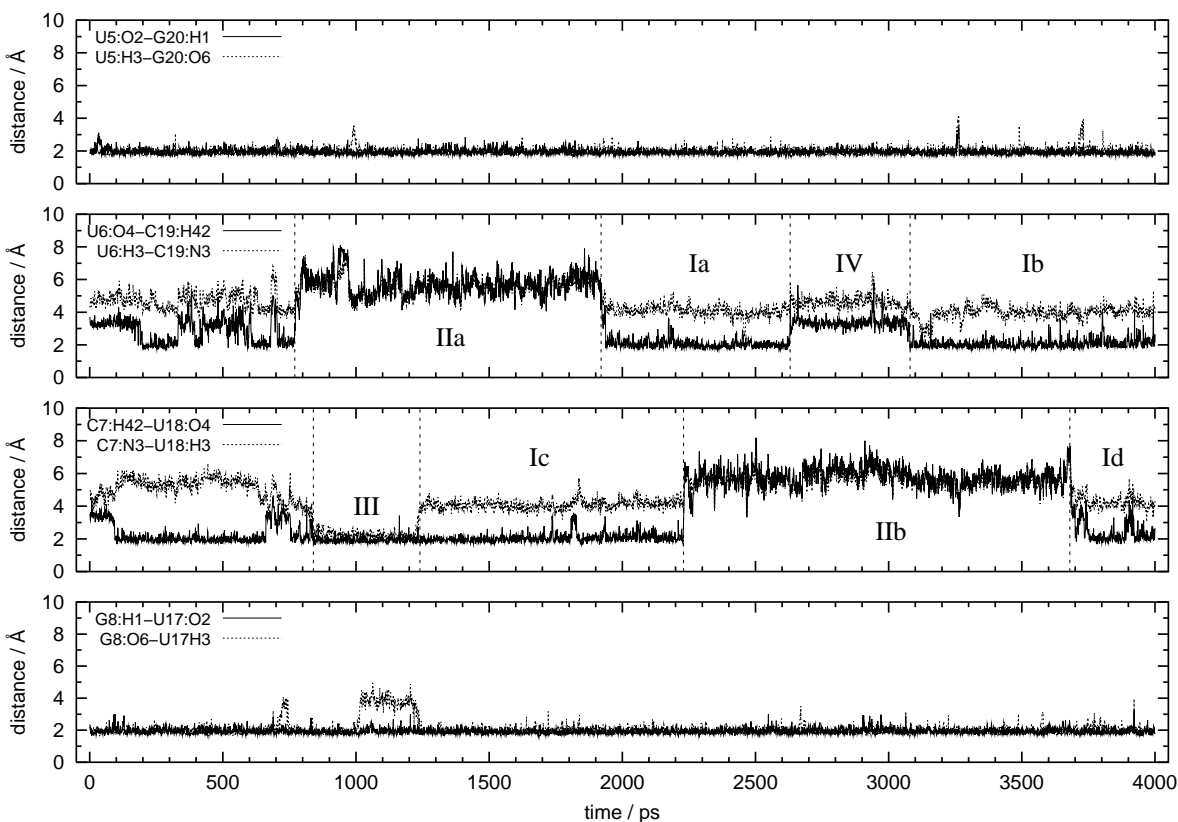


Figure 4.21: Time course of selected inter-base distances for the central UC and the flanking GU pairs. Roman numerals indicate different base pair conformations.

H3(U) and N3(C) is around 4 Å. In conformation II both of these distances are substantially increased. In Fig. 4.22a simulated average structures of conformations I and II are compared to the experimental RNA duplex structure. The superposition was done for the P, O1P, O2P, O3', and O5' backbone atoms of the flanking GU base pairs. It can be seen that conformation I corresponds to the WUC pair found in the X-ray structure ([77]).

On the other hand, in conformation II the relative orientation of U and C is completely different to I. The direct H-bond between the exocyclic amino group of cytosine and O4 of uracil is broken and the shortest inter-base distances occur between the cytosine amino group and O2 of uracil. It is also obvious that uracil shows a much larger movement than cytosine in passing from I to II. We have performed a systematic analysis of backbone torsional angles in both conformations. It turned out that indeed for cytosine there are no major differences between I and II. On the other hand, for uracil significant changes occur for ϵ and ζ (U18: $\epsilon I = -153^\circ$, $\epsilon II = -170^\circ$, $\zeta I = -66^\circ$, $\zeta II = -82^\circ$). Interestingly, conformation II bears resemblance to the geometry of a WUC pair found in the X-ray structure of the sarcin/ricin domain from *E. coli* 23S rRNA (PDB code: 483d, [40]). This can be seen from the superposition in Fig. 4.22b.

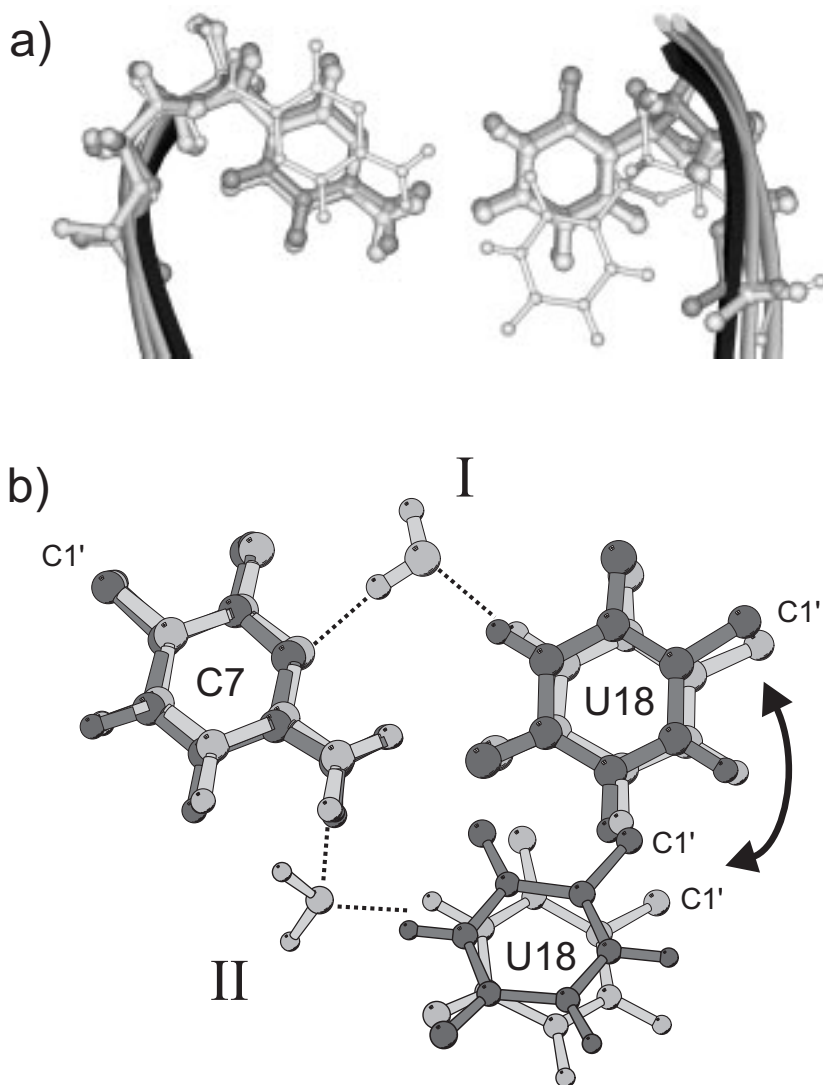


Figure 4.22: a) Superposition of the experimental RNA duplex structure (PDB code: 255d, dark grey) with two simulated average structures (light grey) where either both UC pairs adopt conformations I (simulation time 1920 - 2230 ps; thick sticks) or one UC pair switches to conformation II (simulation time 2230 - 2630 ps; thin sticks). The superposition was done for the P,O1P, O2P, O3', and O5' backbone atoms of the flanking GU base pairs. Only the base pair C7-U18 is shown.

b) Superpositions of the simulated average structures of conformations I (simulation time 1920-2230 ps; light grey, thick sticks) and II (simulation time 2230-2630 ps; light grey, thin sticks) with the experimental geometries of the UC pairs in the A-RNA duplex (PDB code: 255d, [77] (dark grey; thick sticks) and the sarcin/ricin domain 23 S rRNA structure (PDB code: 483d, [39]; dark grey, thin sticks). In order to show the relative orientation of bases cytosine was superimposed. Note, that in the simulation both cytosine and uracil are moving in passing from conformation I to II. The water molecules were manually placed, hydrogen bonds are shown in dotted lines. Conformation I is similar to the UC base pair geometry in the A-RNA duplex and conformation II has a similar geometry as the water-mediated UC pair in the sarcin/ricin 23 S rRNA structure.

In this case the UC pairs of the two experimental structures and of the average structures of conformations I and II are shown by superimposing cytosine (C7). Note, however, that the UC pair in this structure is surrounded by a different nucleic environment as compared to the

simulated structure. Not only the flanking base pairs are different. The UC pair also occurs in a flexible region. Nevertheless, the relative orientation of U and C is similar in the experimental and simulated structure. Whereas, however, U18 of the RNA duplex and of conformation I is approximately located in one plane, the planes of U18 are different for the experimental sarcin/ricin structure and for conformation II.

In addition to the dominating conformations I and II, we find a UC pair with two direct H-

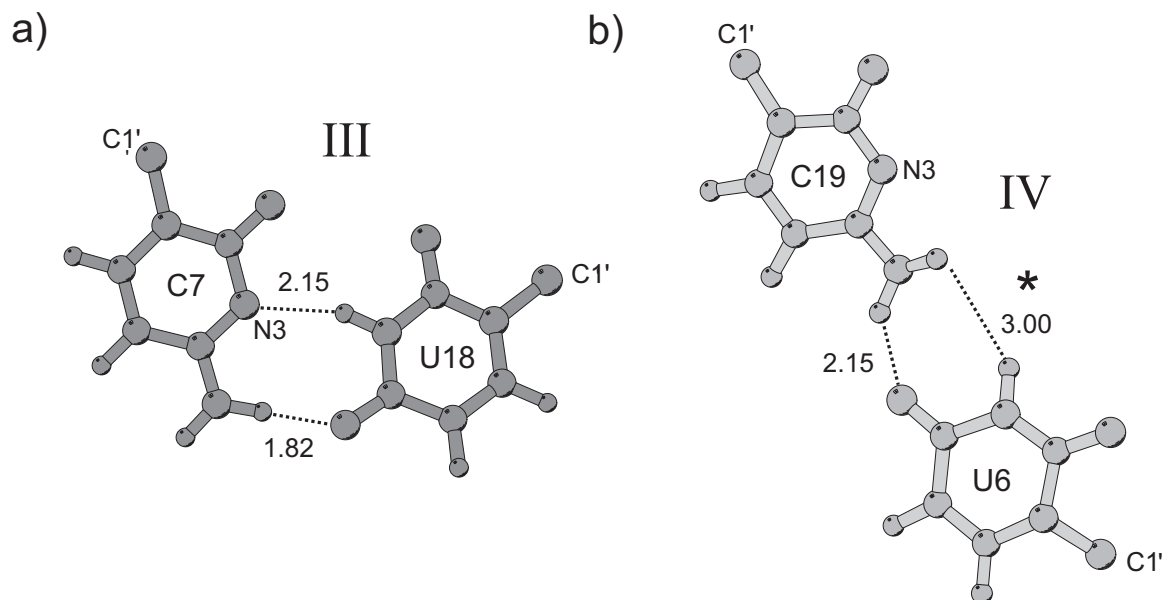


Figure 4.23: a) Direct base pair of conformation III of C7-U18. Both hydrogen bonds are direct and not mediated by single water molecule. This has also consequences for the C1'-C1' distance (see Fig. 4.24) of C7/U18 and for the neighbour pairs U6-C19 and G8/U17 which are obviously reduced. b) Conformation IV of U6-C19. One hydrogen bond with a length of 2.15 Å and a hydrogen-hydrogen distance are displayed. An asterisk (*) stands for a possible water-oxygen.

bonds (III) (Fig. 4.23a) for a short time of 400 ps and a further conformation that is similar to I but exhibits an increased H3(U)···N3(C) distance (IV) (Fig. 4.23b).

In order to get an impression on the flexibility of the four central base pairs of the duplex structure we have monitored their C1'···C1' distances (Fig. 4.24). The fluctuations in this distance are much more pronounced for the water-mediated pairs than for the flanking GU pairs. They are not only due to conformational switches, but also to an increased flexibility within the conformers. As expected, conformation III (see Fig. 4.23) with two direct H-bonds leads to a drastic reduction of the C1'···C1' distance and this geometrical constraint also affects the neighboring base pairs. In this way the long base pair opening of G8-U17 observed between

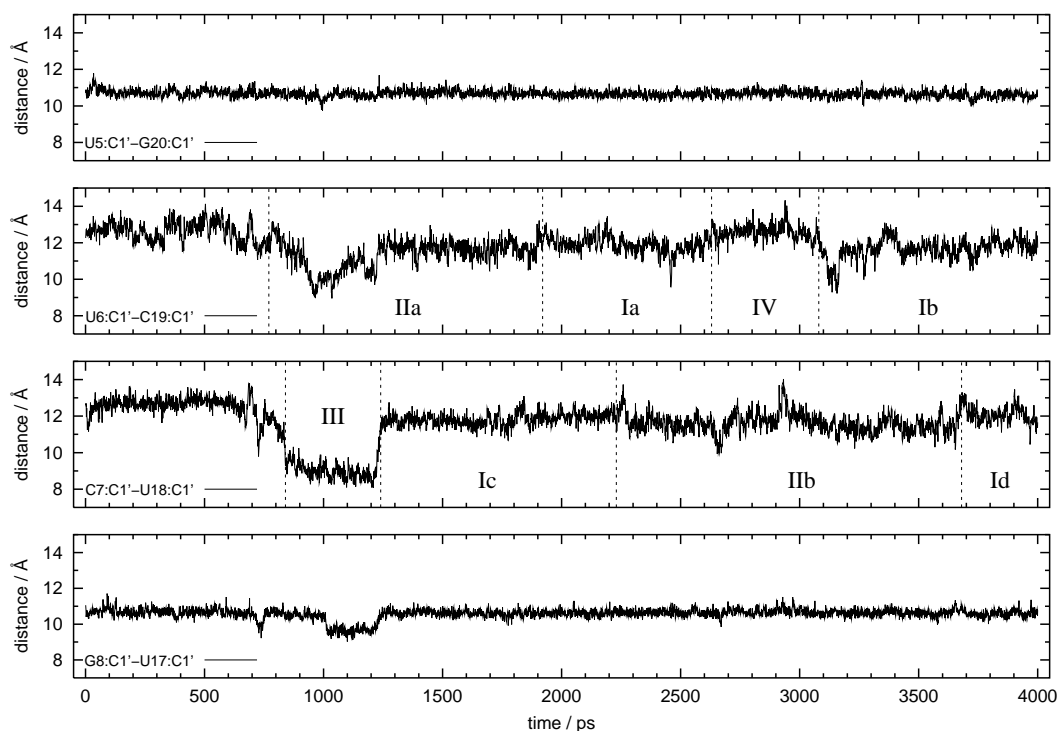


Figure 4.24: Time course of $C1' \cdots C1'$ distances for the central UC and the flanking GU pairs of the simulated duplex structure.

1000 and 1200 ps is induced. In this time range a water molecule is inserted into the GU pair. In the X-ray structure the $C1' \cdots C1'$ distance for the GU pairs is 10.31 Å. This value is by about 0.4 Å smaller than in the simulated average distance (10.67 Å). For the water-mediated UC pair the experimental distance is 11.7 Å which is in good agreement with the simulation results for conformation I. For example, the average $C1' \cdots C1'$ distance for conformations Ic,d is 11.89 Å. The $C1' \cdots C1'$ distance of the UC pair in the sarcin/ricin domain structure is 10.7 Å. This distance has to be compared to the mean $C1' \cdots C1'$ distance of conformations IIa and IIb. It is approximately 1 Å smaller than found in the simulated average structure of conformation II. It should be noted again, however, that the nucleic acid environment in this structure is different to the RNA duplex for which the simulations were performed.

4.3.2 Analysis of the water dynamics

Thus far we have analyzed the base pair geometries without taking water directly into account. In a next step, we have selected all water molecules that fulfill H-bond criteria (H-bond donor and acceptor heavy atom distance $D \cdots A < 3.5 \text{ \AA}$, angle $D - H \cdots A > 120^\circ$) to both of the base donor or acceptor sites for more than 10% of the lifetime of conformations I or II. The results are displayed in Table 4.6.

Table 4.6: Lifetimes of conformations I and II and residence times of long-lived water molecules involved in these conformations (in ps).

Conformation	Base pair	Water Number	Conformation lifetime (from-to)	Water residence time	Total residence time
Ia	U6-C19	88	710 (1920-2630)	211 (1925-2136)	632
		643		421 (2208-2629)	
Ib	U6-C19	2939	920 (3080-4000)	102 (3167-3269)	821
		2673		322 (3281-3602)	
		1964		397 (3604-4000)	
Ic	C7-U18	1255	1006 (1225-2231)	1006 (1225-2231)	
Id		1911	320 (3680-4000)	270 (3684-4000)	
IIa	U6-C19	273	1150 (770-1920)	123 (774-931)	749
		1980		347 (1047-1474)	
		761		279 (1474-1837)	
IIb	C7-U18	426	1450 (2230-3680)	184 (2233-2463)	1227
		2146		409 (2471-2954)	
		1461		260 (2937-3231)	
		352		374 (3237-3652)	

Only water molecules are taken into account that fulfil simultaneously H-bond criteria (H-bond donor and acceptor heavy atom distance $D \cdots A < 3.5 \text{ \AA}$, angle $D - H \cdots A > 120^\circ$) to both bases and have a residence time exceeding 10% of the lifetime of a particular conformation. See Fig. 4.21 for more information on the occurrence of conformations I and II. The from-to difference indicates the first and last frame for which the H-bond criteria with a water molecule are fulfilled. The residence times of long-lived water molecules amounts to 77-94% of the 'from-to' difference. This is due to the fact that the H-bonds criteria may be violated for a short time.

It should be clarified that in conformation I water is both acceptor and donor in the minor groove (shallow edge, see Fig. 2.3) and in conformation II it is a double acceptor in the major groove (Watson-Crick edge). Within conformations Ic and Id only one water is included. By

contrast, conformations Ia, Ib, IIa, and IIb involve two to four long-lived water molecules. In all cases, however, over the major part of the lifetime of a particular conformation water mediated base-base H-bond contacts are observed. The most extreme case is found for conformation Ic. In this case the residence time of water 1255 is practically identical to the lifetime of the conformation (≈ 1 ns). This is the longest water residence time we have ever seen in a MD simulation. Note, that the total residence times of water molecules in a particular conformation are smaller than the 'from-to' residence times. The 'from-to' times indicate the first and last occurrence of a water molecule fulfilling the H-bond criteria. Within this time range the H-bond criteria may be violated for a short time. A detailed view of two representative examples of the water dynamics is shown in Fig. 4.25.

For conformation Ic the water remains in place over the whole lifetime of the conforma-

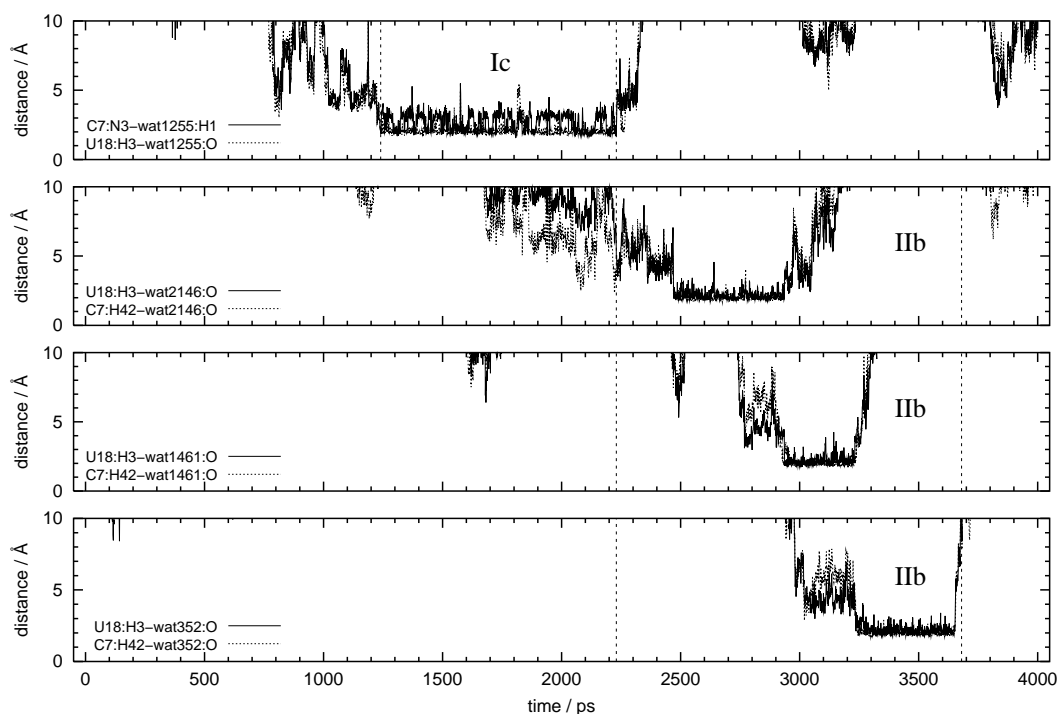


Figure 4.25: Time course of water-base H-bond distances for selected long-lived water molecules incorporated in conformations I and II of the water-mediated UC pair. In conformation Ic the water residence time is identical to the conformation lifetime. For conformation IIb the exchange of three out of the four water molecules involved is shown.

tion. On the other hand, in conformation IIb four water molecules are involved (Table 4.6). In Fig. 4.25 the dynamics of the water exchange for three of them is shown. First, water 2146 is linked to the UC pair. It is then replaced by water 1461 which is in turn replaced by water 352. This water exchange occurs without major effects on the base pair geometry. Fig. 4.25 points to a further interesting feature of the water-base interaction which is explained in Fig. 4.26. In

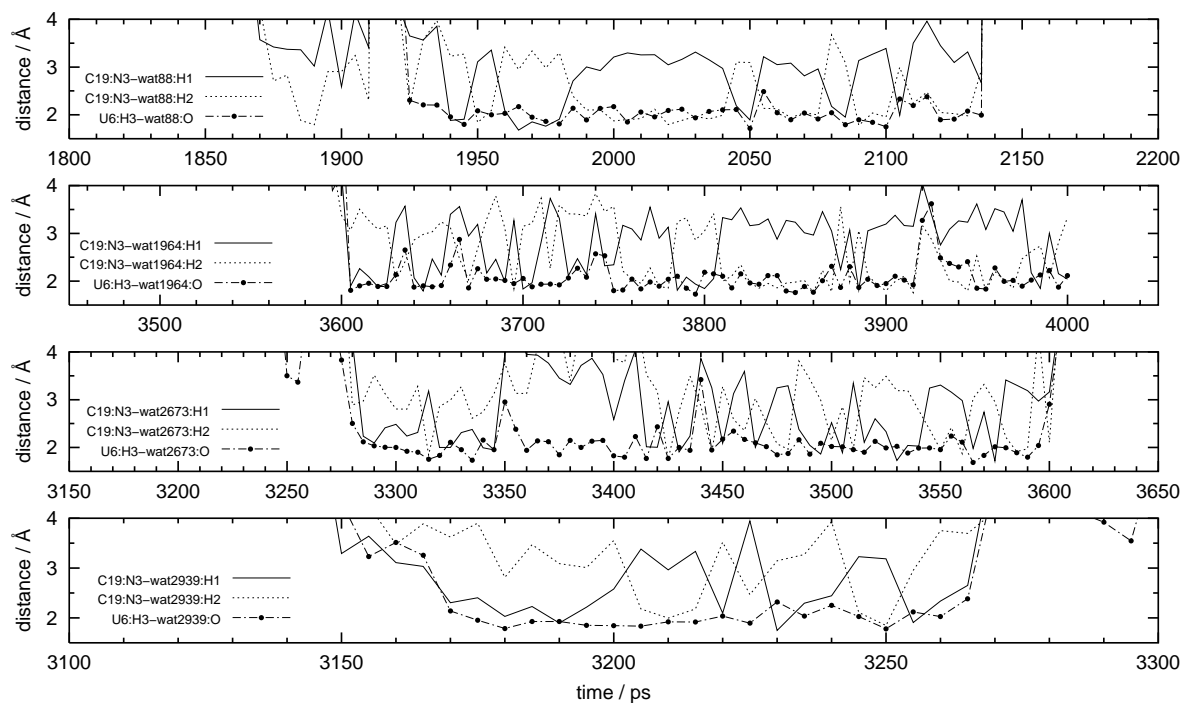


Figure 4.26: Four examples for water molecules mediating U6 and C19 in different time frames are shown. Plotted are the distances of the water hydrogens to C19:N3 (solid and dotted line) and the water oxygen to U6:H3 (line with black points). While the distance between water oxygen and U6:H3 is constant, the water hydrogen distances to C19:N3 alternate between H1 short ($\approx 2 \text{ \AA}$) – H2 long ($\approx 3.5 \text{ \AA}$) and vice versa, which can be explained by a rotation of the water molecule around the virtual U6:H3–wat:O axis.

conformation IIb both U and C donate H-bonds to the water oxygen and both $\text{H} \cdots \text{O}(\text{water})$ H-bond distances exhibit only minor fluctuations around a value of 2 \AA . A different situation is observed for conformation I where water both accepts and donates an H-bond. As for conformation II the $\text{H} \cdots \text{O}(\text{water})$ distance is on the average 2 \AA . Yet, the distance between the water hydrogen H1 and the N3 acceptor site of cytosine adopts two different average distances around 2 and 4 \AA . A closer look shows that a short distance of water hydrogen 1 is usually accompanied by a large distance for hydrogen 2 and vice versa. This means that the water molecule is rotating from time to time thereby disrupting an H-bond and forming a new one with the other water hydrogen. During this process the H-bond distance between the base donor site and the water oxygen is almost unaffected.

Chapter 5

Summary and outlook

The FMN-RNA aptamer

The unrestrained molecular dynamics simulation of the FMN-RNA aptamer leads to a stable structure. Both NOE-derived cross-strand RNA and RNA-FMN distances are maintained during the simulation and the simulated structure has all the qualitative features of the experimental structure including the G10-U12-A25 base triple and the A13-G24, A8-G28 and G9-G27 mismatches. The main differences between the simulated average and NMR structure are the base pair opening and reorganization of the last three base pairs. Moreover, the orientation of the hairpin loop towards the remaining part of the structures is different. The simulation predicts that the flexible phosphoglycerol part of FMN moves towards G27 and forms hydrogen bonds. The simulated average structure of the UUCG hairpin loop within the aptamer shows a few differences to the NMR aptamer structure. However, it is very similar to an independently solved NMR structure of the hairpin alone.

An analysis of the backbone conformation shows that the majority of nucleotides adopts the standard C3'-*endo* and anti conformation. However, for five pyrimidine nucleotides the preferred conformation is C2'-*endo* and for the looped out nucleotide A11 the preferred orientation is syn. Within the RNA backbone, a relatively large number of conformations with the backbone torsional angle γ in trans orientation is found. A survey of all experimental RNA X-ray structures shows that this backbone conformation occurs but is less frequent than that found in the simulation. Experimental NMR RNA aptamer structures have a higher fraction of this conformation compared to the X-ray structures, which is, however, still smaller than in the simulation. The backbone conformation of nucleotides $n+1$ with the torsional angle γ in trans orientation leads to a relatively short distance between 2'-OH(n) and O5'($n+1$), enabling hydrogen bond formation.

There are structural long-lived water molecules in the FMN binding pocket forming H-

bonds with FMN and between FMN and RNA. Hence, water-mediated hydrogen bonds contribute to the free energy of binding in addition to stacking and direct H-bonds between FMN and RNA. In addition, long-lived water is found bridging primarily backbone atoms. A general feature of the environment of long-lived 'structural' water are at least two, and in most cases, three or four potential acceptor atoms.

The 2'-OH group is mainly involved in interactions with the solvent and acts mainly as an acceptor, both for the C3'-*endo* and the C2'-*endo* pucker conformation. There are almost no intra-strand O2'H(n)··O4'(n+1) hydrogen bonds within the RNA backbone. A systematic analysis of the orientation of the 2'-OH hydrogen atom shows that for the standard backbone conformation the preferred orientation is approximately towards the O3' atom of the same nucleotide. If, however, the γ angle of nucleotide n+1 is in the trans range, then the preferred orientation of the 2'-OH(n) hydrogen is towards O5'(n+1) and thus allows for inter-residue hydrogen bond formation.

For the standard backbone conformation we find dynamically stable interactions between C2'(H)(n)··O4'(n+1) and C5'(H)(n+1)··O2'(n). This motif is a possible candidate for contributing to both an increased rigidity and greater thermodynamic stability of RNA compared with DNA.

Coaxial stacking in an RNA hairpin

MD simulations of an RNA tetramer binding to a 4-nucleotide overhang at the 5'-end of a hairpin (nicked structure) and for the corresponding continuous hairpin with either Na⁺ or Mg²⁺ counterions have led to a structural model for the coaxially stacked hairpin structure.

The stacking interface is characterized by a reduced twist and shift, and a slightly increased propeller twist as compared to the continuous hairpin. The modified slide and twist leads to an improved overlap between C22 and G23 in the nicked structure and may thus explain the favorable free energy contribution of the helix-helix interface found in thermodynamic studies on this system [145, 144, 88]. The local geometry changes around the stacking interface result in a global structural effect. The continuous hairpin has an almost straight helical axis. On the other hand, the nicked structure exhibits a marked kink of 39°. This finding is in qualitative agreement with the geometry of coaxial stacking motifs found in tRNAs and pseudoknots [79]. The structural details of the stacking interface may be affected, however, by other features such as the base pair type. Interestingly, the flexibility of the nicked structures is similar to the one of the continuous hairpin.

Compared with Na⁺ counterions, Mg²⁺ ions have a direct structural effect on the continuous RNA hairpin. They induce a narrowing of the major groove and an increased kink angle

of the helical axis. But the nicked structure of the Mg^{2+} simulation is kinked at almost the same quantity as in the Na^+ simulation. Accordingly, cutting the backbone and excising the phosphate has a minor effect in Mg^{2+} simulations.

The narrowing of the deep major groove might simplify an access to the minor groove, which seems more important as a recognition site in RNA. The Na^+ ions show a relatively diffuse distribution, whereas the Mg^{2+} ions are localized at specific binding sites. The majority of these sites are located in the major groove. The preferred RNA atoms interacting via direct or water-mediated contacts with Mg^{2+} are the phosphate oxygens.

Water-mediated base pairs in a RNA helix

A 4 ns MD simulation of the RNA duplex (r-GGACUUCGGUCC)₂ reveals a switching of the two water-mediated UC base pairs between different conformations. One of the two dominating conformations found in the simulation corresponds to the geometry of water-mediated UC pairs of the duplex X-ray structure with water acting both as H-bond donor and acceptor in the minor groove (shallow edge). The H-bond between the water oxygen and the N3-H3 donor site of uracil is retained during the lifetime of this conformation. However, the water molecule rotates from time to time thereby disrupting the H-bond between one water hydrogen and the N3 acceptor site of cytosine and forming a new H-bond involving the other water hydrogen.

Another conformation bears resemblance to a water-mediated UC base pair found in the X-ray structure of the 23S rRNA sarcin/ricin domain. In this case two base donor sites form H-bonds with water oxygen in the major groove (Watson-Crick edge). To switch between both conformations, the bases have to shift over several Ångstrom relative to another. For a very short time a direct UC base pair is also found with two direct hydrogen bonds and a further conformation that is similar to the starting structure but exhibits an elongated H3(U)···N3(C) distance.

These results show that the water-mediated UC base pairs are characterized by a different dynamics as compared to the Watson-Crick and non-canonical guanine-uracil pairs with two or three direct H-bonds. To our knowledge this is the first MD demonstration of a conformational switching of base pairs. The conformational variability and greater flexibility of the water-mediated UC base pairs as compared to conventional pairs is very likely a general property of these unusual pairs. This should be taken into account when considering these structural elements as building blocks of RNA and as recognition motifs.

The MD simulations performed in this work have explored the conformational space around the experimental FMN-RNA aptamer, generated a structural model for a nicked hairpin, for which an experimental structure is not yet available, and finally identified a thus far unknown conformational back-and-forth switching in water-mediated base pairs. In addition, hydration patterns and counterion effects have been analysed. The simulation times varied between 1 and 4 ns. This type of analysis is typical for current MD simulations. However, with increasing computer power the feasible simulation times can hopefully be increased thereby opening up the way for a simulation of RNA folding.

Chapter 6

Supplement

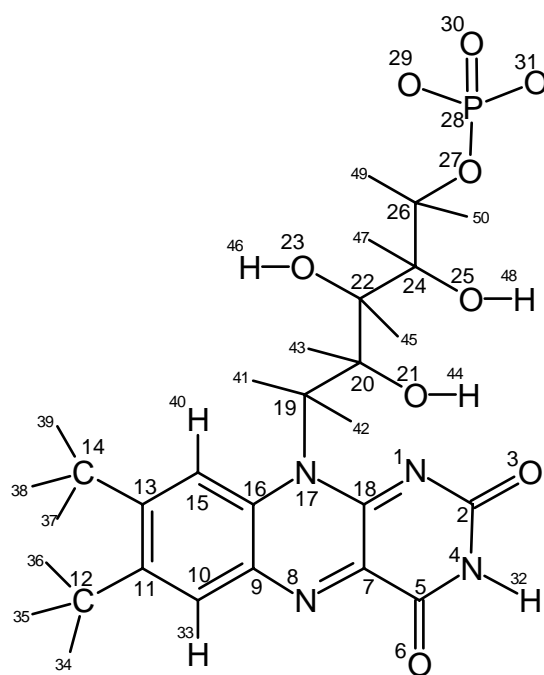


Figure 6.1: Chemical formula and numbering scheme of flavin mononucleotide (FMN) for potential allocation.

Table 6.1: Names, potentials, atom numbers and charges for FMN in consistence with the Cornell force field [38].

Names	Potential	Number	Charge	Names	Potential	Number	Charge
N1	NC	1	-0.3256	C5'	CT	26	0.3213
C2	C	2	0.5487	O5'	OS	27	-0.6748
O2	O	3	-0.4946	P	P1	28	1.5646
N3	NA	4	-0.3841	O1P	O2	29	-1.0097
C4	C	5	0.3733	O2P	O2	30	-1.0097
O4	O	6	-0.5356	O3P	O2	31	-1.0097
C4A	CP	7	0.5182	H3	H	32	0.3046
N5	NC	8	-0.5504	H6	HA	33	0.1821
C5A	CB	9	0.3630	H7M1	HC	34	0.0781
C6	CA	10	-0.2662	H7M2	HC	35	0.0781
C7	CA	11	0.0484	H7M3	HC	36	0.0781
C7M	CT	12	-0.2638	H8M1	HC	37	0.0863
C8	CA	13	0.1857	H8M2	HC	38	0.0863
C8M	CT	14	-0.2716	H8M3	HC	39	0.0863
C9	CA	15	-0.3043	H9	HA	40	0.1773
C9A	CB	16	-0.0075	H1'1	HC	41	0.1040
N10	N*	17	-0.0989	H1'2	HC	42	0.1040
C10	CP	18	0.0537	H2'	H1	43	0.0401
C1'	CT	19	0.0059	HO2	HO	44	0.4447
C2'	CT	20	0.2136	H3'	H1	45	0.0349
O2'	OH	21	-0.6791	HO3	HO	46	0.5027
C3'	CT	22	0.1914	H4'	H1	47	0.1002
O3'	OH	23	-0.7893	HO4	HO	48	0.5179
C4'	CT	24	0.0095	H5'1	H1	49	-0.0007
O4'	OH	25	-0.7267	H5'2	H1	50	-0.0007

Table 6.2: Bond parameters for FMN.

bond	K_r^a	r_{eq}^b	bond	K_r^a	r_{eq}^b
O2-P1	525.0	1.480	OS-P1	230.0	1.610
NC-CP	415.0	1.390	CP-NC	415.0	1.390
CP-N3	388.0	1.410	CP-CP	360.0	1.480
C -CP	410.0	1.444	CP-N*	388.0	1.410

^a kcal/(mol Å²).^b Å.**Table 6.3:** Angle parameters for FMN.

angle	K_θ^a	θ_{eq}^b	angle	K_θ^a	θ_{eq}^b
CT-OS-P1	100.0	120.50	O2-P1-O2	140.0	109.50
O2-P1-OS	100.0	109.50	NA-C -NC	70.0	120.00
CP-CP-NC	60.0	120.00	NC-CP-C	63.0	120.00
CB-NC-CP	70.0	120.00	CA-CB-NC	70.0	120.00
N3-CB-CB	63.0	120.00	N3-CB-CA	63.0	120.00
N3-CP-NC	60.0	120.00	N3-CP-CP	70.0	120.00
CP-NC-C	70.0	120.00	CP-N3-CB	70.0	120.00
CP-N3-CT	70.0	120.00	CT-N3-CB	70.0	120.00
HC-CT-N3	50.0	109.50	CP-C -NA	70.0	114.10
CP-C -O	80.0	125.30	C -CP-CP	63.0	120.70
CP-N*-CT	70.0	120.00	HC-CT-N*	50.0	109.50
CP-N*-CB	70.0	120.00	N*-CB-CA	63.0	120.00
N*-CP-CP	70.0	120.00	N*-CP-NC	60.0	120.00
C -NC-NB	0.0	120.00			

^a kcal/(mol radian²).^b deg.

Table 6.4: Torsional parameters for FMN.

Normal torsions									
torsion	paths ^a	$V_n/2^b$	δ^c	n^d	torsion	paths ^a	$V_n/2^b$	δ^c	n^d
X -OS-P1-X	3	0.75	0.0	3.0	CP-CP-NC-C	2	5.20	180.0	2.0
CB-NC-CP-C	2	5.20	180.0	2.0	CB-N3-CP-NC	4	3.40	180.0	2.0
CB-N3-CP-CP	4	3.40	180.0	2.0	N3-CP-NC-C	2	5.20	180.0	2.0
CP-CP-NC-CB	2	5.20	180.0	2.0	CP-N3-CB-CB	4	3.40	180.0	2.0
CP-N3-CB-CA	4	3.40	180.0	2.0	CT-N3-CB-CB	4	3.40	180.0	2.0
CT-N3-CB-CA	4	3.40	180.0	2.0	CT-N3-CP-NC	4	3.40	180.0	2.0
CT-N3-CP-CP	4	3.40	180.0	2.0	X -CP-CP-X	4	26.60	180.0	2.0
X -C -CP-X	4	8.70	180.0	2.0	CB-N*-CP-CP	4	3.40	180.0	2.0
CT-N*-CP-CP	4	3.40	180.0	2.0	N*-CP-NC-C	2	5.20	180.0	2.0
CT-N*-CP-NC	4	3.40	180.0	2.0	CB-N*-CP-NC	4	3.40	180.0	2.0
Improper torsions									
torsion		$V_n/2$	δ	n	torsion		$V_n/2$	δ	n
CA-CB-CA-HA		1.1	180.0	2.0	CA-CT-CA-CA		1.1	180.0	2.0
CB-CA-CA-HA		1.1	180.0	2.0	NA-CB-C -O		1.1	180.0	2.0
C -C -NA-H		1.1	180.0	2.0	NC-NA-C -O		1.1	180.0	2.0
NA-CP-C -O		1.1	180.0	2.0					

^a Number of bond paths that the total $V_n/2$ is divided into.

^b Magnitude of torsion in kcal/mol.

^c Phase offset in deg.

^d The periodicity of the torsion.

Table 6.5: *List of parameters and settings used in the MD simulations of the FMN aptamer.*

2607 TIP3P water molecules, 36 Na⁺ ions
minimization over 500 steps, conjugate gradient method
gradually heating over 20 ps from 1 K to 300 K, time step 1 fs
further 10 ps of equilibration at 300 K
time step for production: 2 fs
SHAKE bond length constraints applied to bonds involving hydrogens
SHAKE tolerance: 0.0005 Å
temperature: 300 K
Berendsen T coupling constant: 0.25 ps for solvent and solute
pressure: 1 atm
Berendsen pressure coupling constant: 0.5 ps, molecule-based
Ewald grid size: 36 × 72 × 36
Ewald grid interpolation with cubic B-splines
Ewald direct sum tolerance: 0.00001
van der Waals cut-off radius: 10.0 Å
update of non-bonded pair list every 10 steps
dielectric constant: 1, distance independent
coordinates were saved every 2 ps
total main simulation time: 1700 ps

Table 6.6: *List of parameters and settings used in the Coaxial stacking simulations.*

cNa ⁺ : 2776 TIP3P water molecules, 29 Na ⁺ , and 4 Cl ⁻ ions
nNa ⁺ : 2754 TIP3P water molecules, 28 Na ⁺ , and 4 Cl ⁻ ions
cMg ²⁺ : 2808 TIP3P water molecules, 12 Mg ²⁺ , 7 Na ⁺ , and 6 Cl ⁻ ions
nMg ²⁺ : 2811 TIP3P water molecules, 12 Mg ²⁺ , 6 Na ⁺ , and 6 Cl ⁻ ions
average density 1.06 g/cm ³
minimization over 1000 steps, conjugate gradient method
gradually heating over 10 ps from 1 K to 300 K, time step 1 fs
further 100 ps of equilibration at 300 K
time step for production: 2 fs
SHAKE bond length constraints applied to bonds involving hydrogens
SHAKE tolerance: 0.0005 Å
temperature: 300 K
Berendsen T coupling constant: 0.5 ps for solute and 0.2 ps for solvent
pressure: 1 atm
Berendsen pressure coupling constant: 1.0 ps, molecule-based
Ewald grid size: 72 × 36 × 36
Ewald grid interpolation with cubic B-splines
Ewald direct sum tolerance: 0.00001
van der Waals cut-off radius: 10.0 Å
update of non-bonded pair list every 10 steps
dielectric constant: 1, distance independent
coordinates were saved every ps
total main simulation time for cNa ⁺ : 2000 ps
total main simulation time for nNa ⁺ : 3000 ps
total main simulation time for cMg ²⁺ : 2500 ps
total main simulation time for nMg ²⁺ : 2500 ps

Table 6.7: *List of parameters and settings used in the WUC simulations.*

2962 TIP3P water molecules, 34 Na⁺ ions, 12 Cl⁻ ions
minimization over 1000 steps, conjugate gradient method
average density 1.06 g/cm³
gradually heating over 10 ps from 1 K to 300 K, time step 1 fs
further 100 ps of equilibration at 300 K
time step for production: 2 fs
SHAKE bond length constraints applied to bonds involving hydrogens
SHAKE tolerance: 0.0005 Å
temperature: 300 K
Berendsen T coupling constant: 0.5 ps for solute and 0.2 ps for solvent
pressure: 1 atm
Berendsen pressure coupling constant: 0.2 ps, molecule-based
Ewald grid size: 42 × 57 × 42
Ewald grid interpolation with cubic B-splines
Ewald direct sum tolerance: 0.00001
van der Waals cut-off radius: 10.0 Å
update of non-bonded pair list every 10 steps
dielectric constant: 1, distance independent
coordinates were saved every ps
total main simulation time: 4000 ps

References

- [1] Åqvist, J. (1990) Ion-Water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.*, **94**, 8021–8024.
- [2] Abola, E. E., Sussman, J. L., Prilusky, J., and Manning, N. O. (1997) Protein Data Bank archives of three-dimensional macromolecular structures. *Methods Enzymol.*, Academic Press, San Diego.
- [3] Alder, B. J. and Wainwright, T. E. (1957) . *J. Chem. Phys.*, **26**, 1208.
- [4] Alhambra, C., Luque, F. J., Gago, F., and Orozco, M. (1997) Ab initio study of stacking interactions in A- and B-DNA. *J. Phys. Chem. B*, **101**, 3846–3853.
- [5] Allain, F. H. and Varani, G. (1995) Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol.*, **250**, 333–353.
- [6] Allen, M. P. and Tildesley, D. J. (1987) Computer simulations of liquids. , Oxford University Press, New York.
- [7] Auffinger, P., Louise-May, S., and Westhof, E. (1995) Multiple molecular dynamics simulations of the anticodon loop of tRNA^{Asp} in aqueous solution with counterions. *J. Am. Chem. Soc.*, **117**, 6720–6726.
- [8] Auffinger, P., Louise-May, S., and Westhof, E. (1996) Hydration of C-H groups in tRNA. *Faraday Discuss.*, **103**, 151–173.
- [9] Auffinger, P., Louise-May, S., and Westhof, E. (1996) Molecular dynamics simulations of the anticodon hairpin of tRNA^{Asp}: Structuring effects of C-H···O hydrogen bonds and of long-range hydration forces. *J. Am. Chem. Soc.*, **118**, 1181–1189.
- [10] Auffinger, P., Louise-May, S., and Westhof, E. (1999) Molecular dynamics simulations of solvated yeast tRNA^{Asp}. *Biophys. J.*, **76**, 50–64.
- [11] Auffinger, P. and Westhof, E. (1996) H-bond stability in the tRNA^{Asp} anticodon hairpin: 3 ns of multiple molecular dynamics simulations. *Biophys. J.*, **71**, 940–954.

- [12] Auffinger, P. and Westhof, E. (1997) RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA^{Asp} anticodon hairpin. *J. Mol. Biol.*, **269**, 326–341.
- [13] Auffinger, P. and Westhof, E. (1997) Rules governing the orientation of the 2'-hydroxyl group in RNA. *J. Mol. Biol.*, **274**, 54–63.
- [14] Auffinger, P. and Westhof, E. (1998) Simulations of the molecular dynamics of nucleic acids. *Curr. Opin. Struct. Biol.*, **8**, 227–236.
- [15] Aymami, Coll, van der Marel, G., van Boom, J., Wang, A., and Rich (1990) Molecular structure of nicked DNA: a substrate for DNA repair enzymes. *Proc. Natl. Acad. Sci. U. S. A.*, **87**, 2526–2523.
- [16] Batey, R. T., Rambo, R. P., and Doudna, J. A. (1999) Tertiäre Motive bei Struktur und Faltung von RNA. *Angew. Chem.*, **111**, 2472–2491.
- [17] Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D., and Williamson, J. R. (1996) Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science*, **273**, 1547–1551.
- [18] Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the RESP model. *J. Phys. Chem.*, **97**, 10269–10280.
- [19] Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. v., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
- [20] Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. H., Srinivasan, A. R., and Schneider, B. (1992) The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.*, **63**, 751–759.
- [21] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–42.
- [22] Brandl, M., Lindauer, K., Meyer, M., and Sühnel, J. (1999) C-H···O and C-H···N Interactions in RNA Structures. *Theor. Chem. Acc.*, **101**, 103–113.

- [23] Brandl, M., Meyer, M., and Sühnel, J. (1999) Quantum-Chemical Study of a Water-Mediated Uracil-Cytosine Base Pair. *J. Am. Chem. Soc.*, **121**, 2605–2606.
- [24] Burgstaller, P. and Famulok, M. (1994) Isolation of RNA aptamers for biological cofactors by in vitro selection. *Angew. Chem. Int. Ed. Engl.*, **33**, 1084–1087.
- [25] Burgstaller, P., Hermann, T., Huber, C., Westhof, E., and Famulok, M. (1997) Isoalloxazine derivatives promote photocleavage of natural RNAs at G·U base pairs embedded within helices. *Nucleic Acids Res.*, **25**, 4018–4027.
- [26] Burgstaller, P., Kochoyan, M., and Famulok, M. (1995) Structural probing and damage selection of citrulline- and arginine- specific RNA aptamers identify base positions required for binding. *Nucleic Acids Res.*, **23**, 4769–4776.
- [27] Burkhard, M. E., Turner, D. H., and Tinoco Jr., I. (1999) Structures of Base Pairs Involving at Least Two Hydrogen Bonds. *The RNA World - Second Edition*, Cold Spring Harbor Laboratory Press, New York.
- [28] Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham III, T. E., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Ferguson, D. M., Radmer, R. J., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. (1997) AMBER 5. , University of California, San Fransisco, San Francisco, California.
- [29] Cate, J., Yusupov, M., Yusupova, G., Earnest, T., and Noller, H. (1999) X-ray crystal structures of 70S ribosome functional complexes . *Science*, **285**, 2095–2010.
- [30] Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678–1685.
- [31] Cech, T. R. (1993) Structure and mechanism of the large catalytic RNAs: group I and group II introns and ribonuclease P. *The RNA World*, , Cold Spring Harbor Laboratory Press, New York.
- [32] Chang, K. Y. and Tinoco Jr., I. (1997) The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mol. Biol.*, **269**, 52–66.
- [33] Cheatham III., T. E. and Kollman, P. A. (1997) Molecular dynamics simulations highlight the structural differences among DNA:DNA, RNA:RNA, and DNA:RNA hybrid duplexes. *J. Am. Chem. Soc.*, **119**, 4805–4825.

- [34] Chen, E. C., Chen, E. S., and Wentworth, W. E. (1990) The role of electron donors and acceptors in base stacking in DNA and RNA. *Biochem. Biophys. Res. Commun.*, **171**, 97–101.
- [35] Chin, K., Sharp, K. A., Honig, B., and Pyle, A. M. (1999) Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. *Nat. Struct. Biol.*, **6**, 1055–1061.
- [36] Chiu, T. K., Kaczor-Grzeskowiak, M., and Dickerson, R. E. (1999) Absence of minor groove monovalent cations in the crosslinked dodecamer C-G-C-G-A-A-T-T-C-G-C-G. *J. Mol. Biol.*, **292**, 589–608.
- [37] Cohen, S. B. and Cech, T. R. (1998) A quantitative study of the flexibility contributed to RNA structures by nicks and single-stranded gaps. *RNA*, **4**, 1179–1185.
- [38] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz Jr., K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
- [39] Correll, C. C., Freeborn, B., Moore, P. B., and Steitz, T. A. (1997) Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell*, **91**, 705–712.
- [40] Correll, C. C., Wool, I. G., and Munishkin, A. (1999) The Two Faces of the Escherichia coli 23 S rRNA Sarcin/Ricin Domain: The Structure at 1.11 Å Resolution. *J. Mol. Biol.*, **292**, 275–287.
- [41] Dallas, A. and Moore, P. B. (1997) The loop E-loop D region of Escherichia coli 5S rRNA: the solution structure reveals an unusual loop that may be important for binding ribosomal proteins. *Structure*, **5**, 1639–1635.
- [42] Dam, E., Pleij, K., and Draper, D. (1992) Structural and functional aspects of RNA pseudoknots. *Biochemistry*, **31**, 11665–11676.
- [43] Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald: an N log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
- [44] Dickerson, R. E. (1989) Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res.*, **17**, 1797–1803.
- [45] Dickerson, R. E. (1989) Definitions and nomenclature of nucleic acid structure parameters. *J. Biomol. Struct. Dyn.*, **6**, 627–634.

- [46] Dickerson, R. E. and Chiu, T. K. (1997) Helix bending as a factor in protein/DNA recognition. *Biopolymers*, **44**, 361–403.
- [47] Dieckmann, T., Suzuki, E., Nakamura, G. K., and Feigon, J. (1996) Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA*, **2**, 628–640.
- [48] Egli, M., Portmann, S., and Usman, N. (1996) RNA hydration: a detailed look. *Biochemistry*, **35**, 8489–8494.
- [49] Ellington, A. D. (1994) RNA selection. Aptamers achieve the desired recognition. *Curr. Biol.*, **4**, 427–429.
- [50] Ellington, A. D. and Szostak, J. W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
- [51] Ennifar, E., Yusupov, M., Walter, P., Marquet, R., Ehresmann, B., Ehresmann, C., and Dumas, P. (1999) The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure Fold. Des.*, **7**, 1439–1449.
- [52] Essmann, U., Perera, M., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.*, **103**, 8577–8593.
- [53] Ewald, P. (1921) Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.*, **64**, 253–287.
- [54] Famulok, M. and Mayer, G. (1999) Aptamers as Tools in Molecular Biology and Immunology . Current Topics in Microbiology and Immunology, , Springer-Verlag, Heidelberg.
- [55] Fan, P., Suri, A. K., Fiala, R., Live, D., and Patel, D. J. (1996) Molecular recognition in the FMN-RNA aptamer complex. *J. Mol. Biol.*, **258**, 480–500.
- [56] Feig, M. and Pettitt, B. M. (1999) Sodium and chlorine ions as part of the DNA solvation shell [In Process Citation]. *Biophys. J.*, **77**, 1769–1781.
- [57] Feigon, J., Dieckmann, T., and Smith, F. W. (1996) Aptamer structures from A to ζ . *Chem. Biol.*, **3**, 611–617.
- [58] Fermi, E., Pasta, J., and Ulam, S. (1955) . *Los Alamos Preprints*, pages LA–1940.

- [59] Florian, J., Sponer, J., and Warshel, A. (1998) Thermodynamic Parameters for Stacking and Hydrogen Bonding of Nucleic Acid Bases in Aqueous Solution: Ab Initio/Langevin Dipoles Study. *J. Phys. Chem. B*, **103**, 884–892.
- [60] Fourmy, D., Recht, M. I., Blanchard, S. C., and Puglisi, J. D. (1996) Structure of the A site of Escherichia coli 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science*, **274**, 1367–1371.
- [61] Friedman, R. A. and Honig, B. (1995) A free energy analysis of nucleic acid base stacking in aqueous solution. *Biophys. J.*, **69**, 1528–1535.
- [62] Gelbin, A., Schneider, B., Clowney, L., Hsieh, S.-H., Olson, W. K., and Berman, H. M. (1996) Geometric parameters in nucleic acids: Sugar and phosphate constituents. *J. Am. Chem. Soc.*, **118**, 519–529.
- [63] Gellman, S. H., Haque, T. S., and Newcomb, L. F. (1996) New evidence that the hydrophobic effect and dispersion are not major driving forces for nucleotide base stacking. *Biophys. J.*, **71**, 3523–3526.
- [64] Gesteland, R. F. and Atkins, J. F. (1993) The RNA world. Cold Spring Harbor Laboratory Press, New York.
- [65] Guckian, K. M., Schweitzer, B. A., Ren, R. X.-F., Sheils, C. J., Tahmassebi, D. C., and Kool, E. T. (2000) Factors Contributing to Aromatic Stacking in Water: Evaluation in the Context of DNA. *J. Am. Chem. Soc.*, **122**, 2213–2222.
- [66] Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849–857.
- [67] Gyi, J. I., Lane, A. N., Conn, G. L., and Brown, T. (1998) The orientation and dynamics of the C2'-OH and hydration of RNA and DNA:RNA hybrids. *Nucleic Acids Res.*, **26**, 3104–3110.
- [68] Harvey, S. C., Prabhakaran, M., Mao, B., and McCammon, J. A. (1984) Phenylalanine transfer RNA: molecular dynamics simulation. *Science*, **223**, 1189–1191.
- [69] Hermann, T., Auffinger, P., and Westhof, E. (1998) Molecular dynamics investigations of hammerhead ribozyme RNA. *Eur. Biophys. J.*, **27**, 153–165.
- [70] Hermann, T. and Heumann, H. (1995) Determination of nucleotide distances in RNA by means of copper phenanthroline-generated hydroxyl radical cleavage pattern. *RNA*, **1**, 1009–1017.

- [71] Hermann, T. and Patel, D. J. (2000) Adaptive recognition by nucleic acid aptamers. *Science*, **287**, 820–825.
- [72] Hermann, T. and Westhof, E. (1998) Exploration of metal ion binding sites in RNA folds by Brownian-dynamics simulations. *Structure*, **6**, 1303–1314.
- [73] Hermann, T. and Westhof, E. (1999) Simulations of the dynamics at an RNA-protein interface. *Nat. Struct. Biol.*, **6**, 540–544.
- [74] Hobza, P. and Sponer, J. (1999) Structure, Energetics, and Dynamics of the Nucleic Acids Base Pairs: Nonempirical Ab Initio Calculations. *Chem Rev*, **99**, 3247–3276.
- [75] Hobza, P., Sponer, J., and Polasek, M. (1995) H-bonded and stacked DNA base pairs: cytosine dimer. An ab initio second-order Mller-Plesset study. *J. Am. Chem. Soc.*, **117**, 792–798.
- [76] Hockney, R. W. (1970) The potential calculation and some applications. *Methods comput. Phys.*, **9**, 136–211.
- [77] Holbrook, S. R., Cheong, C., Tinoco Jr., I., and Kim, S. H. (1991) Crystal structure of an RNA double helix incorporating a track of non- Watson-Crick base pairs. *Nature*, **353**, 579–581.
- [78] Holbrook, S. R. and Kim, S. H. (1997) RNA crystallography. *Biopolymers*, **44**, 3–21.
- [79] Holland, J. A., Hansen, M. R., Du, Z., and Hoffman, D. W. (1999) An examination of coaxial stacking of helical stems in a pseudoknot motif: the gene 32 messenger RNA pseudoknot of bacteriophage T2. *RNA*, **5**, 257–271.
- [80] Hung, L., Holbrook, E. L., and Holbrook, S. R. (2000) The crystal structure of the rev binding element of HIV-1 reveals novel base pairing and conformational variability. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 5107–5112.
- [81] Jeffrey, G. A. and Saenger, W. (1991) Hydrogen Bonding in Biological structures. , Springer Verlag, Berlin.
- [82] Jiang, F., Fiala, R., Live, D., Kumar, R. A., and Patel, D. J. (1996) RNA folding topology and intermolecular contacts in the AMP-RNA aptamer complex. *Biochemistry*, **35**, 13250–13266.
- [83] Jiang, L., Suri, A. K., Fiala, R., and Patel, D. J. (1997) Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex. *Chem. Biol.*, **4**, 35–50.

- [84] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
- [85] Joseph, S., Berzal-Herranz, A., Chowrira, B. M., Butcher, S. E., and Burke, J. M. (1993) Substrate selections rules for the hairpin ribozyme determined by *in vitro* selection, mutation, and analysis of mismatched substrates. *Genes Dev.*, **7**, 130–138.
- [86] Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H., and Pardi, A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.*, **264**, 968–980.
- [87] Kensch, O., Connolly, B. A., Steinhoff, H. J., McGregor, A., Goody, R. S., and Restle, T. (2000) HIV-1 Reverse Transcriptase-Pseudoknot RNA Aptamer Interaction Has a Binding Affinity in the Low Picomolar Range Coupled with High Specificity. *J. Biol. Chem.*, **275**, 18271–18278.
- [88] Kim, J., Walter, A. E., and Turner, D. H. (1996) Thermodynamics of coaxially stacked helices with GA and CC mismatches. *Biochemistry*, **35**, 13753–13761.
- [89] Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H., Seeman, N. C., and Rich, A. (1974) Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science*, **185**, 435–440.
- [90] Korolev, N., Lyubartsev, A. P., Rupprecht, A., and Nordenskiöld, L. (1999) Competitive binding of Mg(2+), Ca(2+), Na(+), and K(+) ions to DNA in oriented DNA fibers: experimental and monte carlo simulation results. *Biophys. J.*, **77**, 2736–2749.
- [91] Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, **31**, 147–157.
- [92] Lavery, R. and Sklenar, H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.*, **6**, 63–91.
- [93] Le Cam, E., Fack, F., Menissier-de Murcia, J., Cognet, J. A., Barbin, A., Sarantoglou, V., Revet, B., Delain, E., and de Murcia, G. (1994) Conformational analysis of a 139 base-pair DNA fragment containing a single-stranded break and its interaction with human poly(ADP-ribose) polymerase. *J. Mol. Biol.*, **235**, 1062–1071.
- [94] Levitt, M. (1983) Computer simulations of DNA double helix dynamics. *Cold Spring Harb. Symp. Quant. Biol.*, **47**, 251–262.

- [95] Lide, D. R. (1997) CRC Handbook of Chemistry and Physics. , CRC Press, New York.
- [96] Lu, X. J., Babcock, M. S., and Olson, W. K. (1999) Overview of nucleic acid analysis programs. *J. Biomol. Struct. Dyn.*, **16**, 833–843.
- [97] Luise, A., Falconi, M., and Desideri, A. (2000) Molecular dynamics simulation of solvated azurin: Correlation between surface solvent accessibility and water residence times. *Proteins*, **39**, 56–67.
- [98] Lynch, S. R. and Tinoco Jr., I. (1998) The structure of the L3 loop from the hepatitis delta virus ribozyme: a syn cytidine. *Nucleic Acids Res.*, **26**, 980–987.
- [99] Lyubartsev, A. P. and Laaksonen, A. (1998) Molecular dynamics simulations of DNA in solutions with different counterions. *J. Biomol. Struct. Dyn.*, **16**, 579–592.
- [100] MacKerell Jr., A. D. (1997) Influence of magnesium ions on duplex DNA structural, dynamic, and solvation properties. *J. Phys. Chem. B.*, **101**, 646–650.
- [101] Makarov, V. A., Feig, M., Andrews, B. K., and Pettitt, B. M. (1998) Diffusion of solvent around biomolecular solutes: A molecular dynamics simulation study. *Biophys. J.*, **75**, 150–158.
- [102] Marino, J. P., Gregorian, J., R. S., Csankovszki, G., and Crothers, D. M. (1995) Bent helix formation between RNA hairpins with complementary loops. *Science*, **268**, 1448–1454.
- [103] McCammon, J. A. and Harvey, S. C. (1987) Dynamics of proteins and nucleic acids. , Cambridge University Press, Cambridge.
- [104] Meyer, M. and Sühnel, J. (1997) Quantum-chemical *ab initio* study on the adenine-difluorotoluene complex – a mimic for the adenine-thymine base pair. *J. Biomol. Struct. Dyn.*, **15**, 619–624.
- [105] Miller, J. L. and Kollman, P. A. (1997) Theoretical studies of an exceptionally stable RNA tetraloop: observation of convergence from an incorrect NMR structure to the correct one using unrestrained molecular dynamics. *J. Mol. Biol.*, **270**, 436–450.
- [106] Mills, J. B., Cooper, J. P., and Hagerman, P. J. (1994) Electrophoretic evidence that single-stranded regions of one or more nucleotides dramatically increase the flexibility of DNA. *Biochemistry*, **33**, 1797–1803.

- [107] Minasov, G., Tereshko, V., and Egli, M. (1999) Atomic-resolution Crystal Structures of B-DNA Reveal Specific Influences of Divalent Metal Ions on Conformation and Packing. *J. Mol. Biol.*, **291**, 83–99.
- [108] Misra, V. K. and Draper, D. E. (1998) On the role of magnesium ions in RNA stability. *Biopolymers*, **48**, 113–135.
- [109] Moore, P. B. (1998) The three-dimensional structure of the ribosome and its components. *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 35–58.
- [110] Norberg, J. and Nilsson, L. (1995) Stacking free energy profiles for all 16 natural ribodinucleoside monophosphates in aqueous solution. *J. Amer. Chem. Soc.*, **117**, 10832–10840.
- [111] Norberg, J. and Nilsson, L. (1996) Influence of adjacent bases on the stacking-unstacking process of single-stranded oligonucleotides. *Biopolymers*, **39**, 765–768.
- [112] Norberg, J. and Nilsson, L. (1998) Solvent influence on base stacking. *Biophys. J.*, **74**, 394–402.
- [113] Olson, W. K. (1973) Syn-anti effects on the spatial configuration of polynucleotide chains. *Biopolymers*, **12**, 1787–1814.
- [114] Patel, D. J., Suri, A. K., Jiang, F., Jiang, L., Fan, P., Kumar, R. A., and Nonin, S. (1997) Structure, recognition and adaptive binding in RNA aptamer complexes. *J. Mol. Biol.*, **272**, 645–664.
- [115] Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Debolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Comm.*, **91**, 1–41.
- [116] Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Ferguson, D., Seibel, G., Singh, U. C., and Kollman, P. (1994) AMBER 4.1. , University of California, San Fransisco, San Francisco, California.
- [117] Pleij, C. W., Rietveld, K., and Bosch, L. (1985) A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.*, **13**, 1717–1731.
- [118] Pleij, C. W. A. (1994) RNA pseudoknots. *Curr. Opin. Struct. Biol.*, **4**, 337–344.

- [119] Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature*, **372**, 68–74.
- [120] Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F., and Klug, A. (1974) Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature*, **250**, 546–551.
- [121] Roll, C., Ketterle, C., Faibis, V., Fazakerley, G. V., and Boulard, Y. (1998) Conformations of nicked and gapped DNA structures by NMR and molecular dynamic simulations in water. *Biochemistry*, **37**, 4059–4070.
- [122] Rould, M. A., Perona, J. J., and Steitz, T. A. (1991) Structural basis of anticodon loop recognition by glutamyl-tRNA synthetase. *Nature*, **352**, 213–218.
- [123] Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*, **23**, 327–341.
- [124] Saenger, W. (1984) Principles of nucleic acid structure. , Springer-Verlag, New York.
- [125] Schafmeister, C. E. A. F., Ross, W. S., and Romanowski, V. (1995) Leap. , University of California, San Fransisco, San Francisco, California.
- [126] Schneider, C. and Sühnel, J. (1999) A molecular dynamics simulation of the flavin mononucleotide-RNA aptamer complex. *Biopolymers*, **50**, 287–302.
- [127] Schroeder, R., Waldsich, C., and Wank, H. (2000) Modulation of RNA function by aminoglycoside antibiotics. *EMBO*, **19**, 1–9.
- [128] Scott, W. G. (1998) RNA catalysis. *Curr. Opin. Struct. Biol.*, **8**, 720–726.
- [129] Shui, X., McFail-Isom, L., Hu, G. G., and Williams, L. D. (1998) The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry*, **37**, 8341–8355.
- [130] Sich, C., Ohlenschläger, O., Ramachandran, R., Görlach, M., and Brown, L. R. (1997) Structure of an RNA hairpin loop with a 5'-CGUUUCG-3' loop motif by heteronuclear NMR spectroscopy and distance geometry. *Biochemistry*, **36**, 13989–13400.
- [131] Soukup, G. A. and Breaker, R. R. (1999) Design of allosteric hammerhead ribozymes activated by ligand-induced structure stabilization. *Structure Fold. Des.*, **7**, 783–791.
- [132] Soukup, G. A. and Breaker, R. R. (1999) Engineering precision RNA molecular switches. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 3584–3589.

- [133] Sponer, J. and Kypr, J. (1991) Different intrastrand and interstrand contributions to stacking account for roll variations at the alternating purine-pyrimidine sequences in A-DNA and A-RNA. *J. Mol. Biol.*, **221**, 761–764.
- [134] Sponer, J., Leszczynski, J., and Hobza, P. (1996) Nature of nucleic acid-base stacking: Nonempirical ab initio and empirical potential characterization of 10 stacked base dimers. Comparison of stacked and H-bonded base pairs. *J. Phys. Chem.*, **100**, 5590–5596.
- [135] Sprou, D., Young, M. A., and Beveridge, D. L. (1999) Molecular dynamics studies of axis bending in d(G5-(GA4T4C)2-C5) and d(G5-(GT4A4C)2-C5): effects of sequence polarity on DNA curvature. *J. Mol. Biol.*, **285**, 1623–1632.
- [136] Thill, G., Vasseur, M., and Tanner, N. K. (1993) Structural and Sequence Elements Required for the Self-Cleaving Activity of the Hepatitis Delta Virus Ribozyme. *Biochemistry*, **32**, 4254–4262.
- [137] Torres, R. and Bruice, T. (1998) Molecular dynamics study displays near in-line attack conformations in the hammerhead ribozyme self-cleavage reaction. *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 11077–11078.
- [138] Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., d'Aubenton Carafa, Y., Uhlenbeck, O. C., Tinoco Jr., I., and et, a. I. (1988) CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci. U. S. A.*, **85**, 1364–1368.
- [139] Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
- [140] Varani, G. (1995) Exceptionally stable nucleic acid hairpins. *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 379–404.
- [141] Varani, G., Cheong, C., and Tinoco Jr., I. (1991) Structure of an unusually stable RNA hairpin. *Biochemistry*, **30**, 3280–3289.
- [142] Verlet, L. (1967) Computer 'experiments' on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, **159**, 98–103.
- [143] Wahl, M. C. and Sundaralingam, M. (1997) C-H···O hydrogen bonding in biology. *Trends Biochem. Sci.*, **22**, 97–102.

- [144] Walter, A. E. and Turner, D. H. (1994) Sequence dependence of stability for coaxial stacking of RNA helices with Watson-Crick base paired interfaces. *Biochemistry*, **33**, 12715–12719.
- [145] Walter, A. E., Turner, D. H., Kim, J., Lyttle, M. H., Muller, P., Mathews, D. H., and Zuker, M. (1994) Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl. Acad. Sci. U. S. A.*, **91**, 9218–9222.
- [146] Wang, S. and Kool, E. T. (1995) Origins of the large differences in stability of DNA and RNA helices: C-5 methyl and 2'-hydroxyl effects. *Biochemistry*, **34**, 4125–4132.
- [147] Weeks, K. M. and Crothers, D. M. (1993) Major groove accessibility of RNA. *Science*, **261**, 1574–1577.
- [148] Westhof, E. (1988) Water: an integral part of nucleic acid structure. *Annu. Rev. Biophys. Biophys. Chem.*, **17**, 125–144.
- [149] Westhof, E., Dumas, P., and Moras, D. (1988) Restrained refinement of two crystalline forms of yeast aspartic acid and phenylalanine transfer RNA crystals. *Acta Crystallogr A*, **44**, 112–123.
- [150] Westhof, E. and Fritsch, V. (2000) RNA folding: beyond Watson-Crick pairs. *Structure*, **8**, R55–R65.
- [151] Woese, C. R., Winker, S., and Gutell, R. R. (1990) Architecture of ribosomal RNA: constraints on the sequence of tetra-loops. *Proc. Natl. Acad. Sci. U. S. A.*, **87**, 8467–8471.
- [152] Yang, L. and Pettitt, B. M. (1996) B to A transition of DNA on the nanosecond time scale. *J. Phys. Chem.*, **100**, 2564–2566.
- [153] Ye, X., Gorin, A., Ellington, A. D., and Patel, D. J. (1996) Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 rev peptide-RNA aptamer complex. *Nat. Struct. Biol.*, **3**, 1026–1033.
- [154] York, D., Darden, T., and Pedersen, L. (1993) The effect of long-range electrostatic interactions in simulations of macromolecular crystals: a comparison of the Ewald and truncated list methods. *J. Chem. Phys.*, **99**, 8345–8348.
- [155] Young, M., Jayaram, B., and Beveridge, D. (1997) Intrusion of counterions into the spine of hydration in the minor groove of B-DNA: fractional occupancy of electronegative pockets. *J. Am. Chem. Soc.*, **119**, 59–69.

- [156] Young, M. A. and Beveridge, D. L. (1998) Molecular dynamics simulations of an oligonucleotide duplex with adenine tracts phased by a full helix turn. *J. Mol. Biol.*, **281**, 675–687.
- [157] Zichi, D. A. (1995) Molecular dynamics of RNA with the OPLS force field. Aqueous simulation of a hairpin containing a tetranucleotide loop. *J. Am. Chem. Soc.*, **117**, 2957–2969.
- [158] Zimmermann, G. R., Jenison, R. D., Wick, C. L., Simorre, J. P., and Pardi, A. (1997) Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. *Nat. Struct. Biol.*, **4**, 644–649.

Lebenslauf

Christoph Schneider

- 19.09.1968** geboren in Kamp-Lintfort, Kreis Wesel
ledig
- 1975-1979** Grundschule in Friedrichshafen
- 1979-1988** Gymnasium in Markdorf
- 1988** Allgemeine Hochschulreife
- 1988-1989** Grundwehrdienst
- 1989-1997** Studium der Chemie
Universität Stuttgart
- 1997** Hochschulabschluss zum Diplom-Chemiker
Universität Stuttgart
- 1997-2000** Wissenschaftlicher Mitarbeiter am
Institut für Molekulare Biotechnologie, Jena
- 2000** Promotion, Titel der Dissertation:
*Exploring RNA Structures by Means of Molecular Dynamics Simulations:
– The Flavin Mononucleotide Aptamer, Coaxial Stacking in a Nicked Hairpin
and Water-Mediated Base Pairs in a Duplex –*
Friedrich-Schiller-Universität Jena

Ich erkläre, daß ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Hilfsmittel und Literatur angefertigt habe.

Jena, 10. Juli 2000

Christoph Schneider

Ich erkläre, daß ich mich mit der vorliegenden Arbeit an keiner anderen Hochschule um den akademischen Grad *doctor rerum naturalium* (Dr. rer. nat.) beworben habe und daß ich weder früher noch gegenwärtig die Eröffnung eines Verfahrens zum Erwerb des o.g. akademischen Grades an einer anderen Hochschule beantragt habe.

Jena, 10. Juli 2000

Christoph Schneider