

Molecular Dynamics Simulations of RNA Hairpins

DISSERTATION
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt dem Fachbereich
Chemische und Pharmazeutische Wissenschaften
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Jessica Koplín
aus Shillong/Indien

November 2005
(DF1)

vom Fachbereich Chemische und Pharmazeutische Wissenschaften
der Johann Wolfgang Goethe-Universität als Dissertation angenommen

Dekan: Prof. Dr. Harald Schwalbe

1. Gutachter: Prof. Dr. Gerhard Stock

2. Gutachter: Prof. Dr. Harald Schwalbe

Datum der Disputation: 11.10.2005

Für Lena, Bony, Ralf und Max

"Gäbe es nur eine Wahrheit, könnte
man von einem Thema keine Hundert Bilder machen."

Pablo Picasso

Danksagung

Mein besonderer Dank gilt meinem akademischen Lehrer Professor Dr. Gerhard Stock für die Schaffung hervorragender Arbeitsbedingungen, seine offene, freundliche Art. Er war stets ansprechbar, sei es für die Wissenschaft oder andere Themen.

Herrn Dr. Yuguang Mu und Dr. Phuong H. Nguyen danke ich für ihre große Hilfe bei den Simulationen und bei der Verarbeitung großer Datenmenge.

Professor Dr. Harald Schwalbe und Herrn Dr. Christian Richter danke ich für die zahlreichen, interessanten Diskussionen und Gespräche bei unserem gemeinsamen Projekt "UUUU Hairpin".

Dr. Alessandra Villa, Dr. Jens Wöhnert, Oliver Boden, Dr. Verena Ludwig und Alexandros Altis danke ich für das ausführliche Korrekturlesen dieser Arbeit.

Hier in der Gruppe gilt mein Dank Herrn Dr. Rainer Hegger, der bei Computerproblemen immer sofort hilfsbereit war und unserer Sekretärin Frau Claudia von Sulecki Madic für ihre Hilfe bei den verschiedensten Angelegenheiten. Nicht zuletzt danke ich allen Mitarbeitern der Arbeitskreise von Professor Dr. Gerhard Stock und Professor Dr. Joseph Wachtveitl für die stets gute Stimmung.

Last but not least, I want to thank Pawel and Danii for always motivating and encouraging me, especially in the most difficult phases of this work.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 2 | Methods | 4 |
| 2.1 | Molecular dynamics (MD) simulation | 4 |
| 2.2 | Sampling methods | 7 |
| 2.2.1 | Locally enhanced sampling (LES) | 7 |
| 2.2.2 | Replica exchange molecular dynamics (REMD) simulation . . | 8 |
| 2.3 | Principal component analysis (PCA) | 9 |
| 2.4 | Calculation of NMR parameters | 10 |
| 3 | Structure and dynamics of the UUUU loop | 14 |
| 3.1 | MD simulation of the UUUU loop | 14 |
| 3.1.1 | Computational details | 14 |
| 3.1.2 | Dynamics of the UUUU loop | 15 |
| 3.1.3 | Backbone angle analysis | 16 |
| 3.1.4 | Sugar conformations | 24 |
| 3.1.5 | NOE distances | 24 |
| 3.1.6 | Conformational dynamics | 26 |
| 3.2 | REMD simulation of the UUUU loop | 33 |
| 3.2.1 | Computational details | 33 |
| 3.2.2 | Dynamics of the UUUU loop | 33 |
| 3.2.3 | Comparison to NMR | 37 |
| 3.2.4 | Conformational analysis | 41 |
| 3.3 | Conclusions | 43 |

| | | |
|----------|---|------------|
| 4 | Molecular dynamics of the UUCG tetraloop | 45 |
| 4.1 | Computational details | 47 |
| 4.2 | The 14 mer UUCG loop | 48 |
| 4.2.1 | Dihedral angles | 49 |
| 4.2.2 | 3J coupling | 51 |
| 4.2.3 | S^2 order parameter | 51 |
| 4.2.4 | Dynamical analysis | 52 |
| 4.3 | The 10 mer UUCG loop | 59 |
| 4.3.1 | Conformational analysis | 59 |
| 4.4 | Conclusions | 62 |
| 5 | Unfolding studies of the UUUU and UUCG loop | 63 |
| 5.1 | Introduction | 63 |
| 5.2 | Observables of thermal unfolding | 64 |
| 5.3 | Conformations at high temperatures | 70 |
| 5.4 | Outlook | 80 |
| 6 | Docking Simulation | 81 |
| 6.1 | Theory of Docking | 81 |
| 6.1.1 | Energy evaluation | 81 |
| 6.1.2 | Search algorithm | 83 |
| 6.2 | Binding of aromatic ligands to TAR | 84 |
| 6.2.1 | Computational details | 84 |
| 6.2.2 | Docking results | 86 |
| 6.2.3 | Summary | 92 |
| 6.3 | Binding and MD studies of aminoglycosides | 92 |
| 6.3.1 | Outlook | 96 |
| 7 | Summary | 97 |
| 8 | Zusammenfassung | 100 |

List of Figures

| | | |
|-----|--|----|
| 2.1 | A two-dimensional periodic system. Molecules can enter and leave each box across each of the edges. In a three-dimensional example, molecules would be free to cross any of the six cube faces | 7 |
| 2.2 | Illustration of the torsional angles (upper panel), the the pseudorotational states (middle panel) and the RNA bases (lower panel). The conformation of the backbone is determined by the dihedral angles α , β , γ , δ , ϵ , and ζ . The conformation along the glycosidic bond is determined by the χ angle. The torsion angles of the sugar ring are denoted by $\nu_0 - \nu_4$ | 11 |
| 3.1 | Time evolution of the RMSD obtained for trajectories MD1 at 300 K (solid line) and for the 350 K simulation (dashed line). | 16 |
| 3.2 | Circle representation of the torsional angles of the UUUU loop as obtained from the two trajectories at 300 K, MD1 and MD2, as well as from the 350 K trajectory and the LES trajectory at 300K. | 17 |
| 3.3 | Determination of the backbone angles α , β , ϵ and ζ from experimental 3J couplings (left) and CSA cross-correlated relaxation rates (right), shown for the example of the U7 nucleotide of the UUUU loop. (a) Experimental values (as boxes) and theoretical Karplus curves (Eqs. (1) and (2), as lines) of various 3J couplings reflecting the angle β . Shown are 3J (H5'proSi,Pi) drawn as square and dashed lines, 3J (H5'proRi,Pi) drawn as filled box and dotted lines and 3J (C4'i, Pi) drawn as filled circle and lines. (b) Theoretical dependence of the CSA cross-correlated relaxation rates $\Gamma_{(C5',H5'),(Pi)}^c + \Gamma_{(C5',H55'),(Pi)}^c$ on the angles β and α . The red circles indicate the conformational regions which fulfill the experimental cross-correlated relaxation rates. Assuming $\beta = -165$ as obtained in (a), possible values for the angle α are -120, 29, and 75 (pink bars). (c) Various 3J couplings reflecting the angle ϵ . Shown are 3J (H3i,Pi+1) drawn as filled squares and dashed lines. 3J (C2i, Pi) drawn as circles and dotted lines, and 3J (C4i, Pi+1) drawn as filled circles and lines. (d) Theoretical dependence of the CSA cross-correlated relaxation rates (in green) and (in blue) on the angles ϵ and ζ . The red circles indicate the conformational regions which fulfill the experimental cross-correlated relaxation rates. | 19 |

| | | |
|------|--|----|
| 3.4 | Joint probability distribution of the α and β (upper panels) and ϵ and ζ (lower panels) as obtained for trajectory MD1 at 300 K and for the 350 K run. The red dots correspond to values of the angles that are consistent with the analysis of the experimental data in Fig 3.3. The green and blue asterisks display the values of the canonical A and B form, respectively. | 20 |
| 3.5 | Experimental and simulated ^{31}P -CSA cross correlated relaxation rates of the UUUU loop. Comparison of the NMR in black to MD at 350 K (red) and to the LES method (green) | 22 |
| 3.6 | Experimental and simulated ^3J couplings of the UUUU loop. Upper panel: Comparison of the two 300 K trajectories MD1 (gray squares) and MD2 (light triangles) to NMR data (black circles). Lower panel: Comparison of the 350 K trajectory (squares) and the LES trajectory (triangles) to NMR data (circles). | 25 |
| 3.7 | Experimental and simulated dipole-dipole cross correlated relaxation rates of the UUUU loop. Upper panel: Comparison of the two 300 K trajectories MD1 (gray squares) and MD2 (light triangles) to NMR data (black circles). Lower panel: Comparison of the 350 K trajectory (squares) and the LES trajectory (triangles) to NMR data (circles). | 26 |
| 3.8 | Relative fraction of the dynamics gained from the first PCA eigenvectors | 27 |
| 3.9 | Free energy surface (in units of kcal/mol) obtained from MD1 (upper panel), LES (middle panel) and from the 350 K (lowest panel) run, plotted as a function of the first two PCA eigenvectors. | 29 |
| 3.10 | Illustration of the secondary structures of the UUUU loop as found in NMR experiment and in the 350 K MD simulation. Dashed lines indicate that the residue points into the solvent. | 31 |
| 3.11 | Conformational dynamics of the UUUU loop as observed in the MD simulation at 350 K. (a) Projection of the trajectory on its first (black), second (red), and third (green) PCA eigenvector, respectively. (b) Distances between the loop bases U4 and U5 (black), U5 and U7 (red), U5 and U6 (green), and between U6 and U7 (blue). (c) Pseudorotation angles of residues U5 (blue) and U6 (green). | 32 |
| 3.12 | Time trace of the all atoms RMSD obtained from the REMD simulation at 300 K. These values are the deviations from the average structure. | 34 |
| 3.13 | Time trace of the RMSD of the whole UUUU system (in Å) and backbone torsion angles of the U4 residue obtained from the REMD simulation at 300 K. | 35 |
| 3.14 | Time evolution of the distances in Å obtained from the 300 K REMD simulation. The distances are calculated between the center of masses of the bases in the loop region. U4U5 is the distance between the base of residue U4 and the base of U5. The other distances are defined in the same way. | 36 |

| | | |
|------|--|----|
| 3.15 | Time course of the hydrogen bond distances (Å) in the stem of the UUUU loop. The distances are calculated between the heavy atoms N1-H1..N3 of the base pairs. | 36 |
| 3.16 | Pseudorotational angles for the loop region of the UUUU hairpin gained from REMD at 300 K (black). The probability (%) as obtained from NMR is shown for U4 in the C3'-endo, for U5, U7 in C2'-endo and U6 in C3'-endo and C2'endo state. | 38 |
| 3.17 | ³ J couplings for the UUUU loop gained from conventional MD in red (MD1) and green (MD2). ³ J couplings from REMD (blue) at 300 K in blue compared to NMR in black. | 38 |
| 3.18 | Joint probability distribution of the α and β (upper panels) and ϵ and ζ (lower panels) as obtained for trajectory REMD and for the conventional MD 350 K run. The red dots correspond to values of the angles that are consistent with the analysis of the experimental data in Fig 3.3. The green and blue asterisks display the values of the canonical A and B form, respectively. | 40 |
| 3.19 | Distribution of the first nine principal components gained from an REMD simulation at 300 K. A determination of the sub states (labeled with 1 and 2) is done using the first four principal components. | 42 |
| 3.20 | Illustration of the secondary structures of the UUUU loop as found in REMD simulation at 300 K. Conformations A, B, C and D occur with the probabilities 49%, 23%, 8% and 7%, respectively. These conformations are gained by combining all sub states gained from the first four principal components as shown in Fig. 3.20. | 42 |
| 4.1 | Stabilizing interactions derived from X-ray structure of the UUCG motif [1]. The hydrogen bond interactions contributing to the structure of the tetraloops are seen: (10) U6(O2')-G9(O6), (11) G9(N2)-U6(O2), (12) G9(N1)-U6(O2), (13) U7(O2')-G9(O6), (14) C8(O2')-C8(O2) and (15) C8(N4)-U7(PO ^{ProR}). Stacking of the bases are seen between residue U2 and C1 (U2-C1) and between U2 and C4 (U2-C4). | 46 |
| 4.2 | Secondary structure (left) and a snapshot of the 14 nt RNA investigated with conventional MD (right). | 48 |
| 4.3 | Comparison between MD (red) at 300 K and NMR [2] (black) at 298 K of the backbone torsional angles of the 14 mer. Only the dihedrals observed experimentally are reported. | 49 |
| 4.4 | Comparison between MD (red) at 300 K and NMR (black) at 298 K of the backbone torsional angles for the 14 mer UUCG loop. | 51 |
| 4.5 | Comparison between MD (red) and NMR (black) of the order parameters of the 14 mer. Left panel: at 300 K (MD) and 298 K (NMR). Right panel: at 317 K. | 52 |

| | | |
|------|--|----|
| 4.6 | The atomic fluctuations in the loop region around the initial structure of the 14 mer UUCG from a conventional MD simulation at 300 K. P: phosphor, BB: Backbone, R: Ribose | 53 |
| 4.7 | The atomic fluctuations in the loop region around the initial structure of the 14 mer UUCG from a classical MD simulation at 317 K. P: phosphor BB: Backbone, R: Ribose. | 54 |
| 4.8 | Circle representation of the backbone angles of the UUCG 14 mer at 300 K. | 55 |
| 4.9 | Circle representation of the backbone angles of the UUCG 14 mer at 317 K. | 56 |
| 4.10 | Backbone angles ϵ and ζ of the 14 mer UUCG loop at 300 K (red) and at 317 K (black). | 57 |
| 4.11 | Upper panel: Time trace of the distance between the center of mass of phosphate G9 and G10. Lower panel: Time evolution of angle ϵ of the G9 residue of the 14 mer at 300 K. | 57 |
| 4.12 | Time traces from a 400 K trajectory of the 14 mer UUCG loop. a) RMSD of the total tetraloop. b) RMSD of phosphate groups. Black: G9-G10. Red: U7-G10. Green: C5-G10 and Yellow: G2-G10. c) G9 ϵ angle. | 58 |
| 4.13 | Dynamics of the 10 mer UUCG at 300 K. a) RMSD of the total tetraloop b) Distance between the phosphate of C3 und U4 and c) Angles ϵ and ζ of the C3 residue | 59 |
| 4.14 | Distribution of the first nine principal components gained from the REMD trajectory at 300 K. The determination of the sub states (labeled with 1 and 2) is done using the first six (without component 3) principal components for the 10 mer UUCG loop. | 60 |
| 4.15 | Conformations A, B, C and D of the 10 mer occur with high population probability in the trajectory of the REMD simulation at 300 K: A=61%, B=9%, C=8% and D=5%. These conformations are gained by combining all sub states gained from the first six (except number 3) principal components as shown in Fig. 4.14. | 61 |
| 5.1 | Radius of gyration for the UUCG (black) and UUUU (red). a) For the whole system. b) For the loop region. c) For the stem area. | 64 |
| 5.2 | Upper panel illustrates the hydrogen bonds in the UUCG and UUUU loop. Middle panel depicts hydrogen bond opening in the stem of UUCG (black) and UUUU (red). Lower panel shows the hydrogen bonds of the loop region of the UUCG only. | 66 |

| | | |
|-----|---|----|
| 5.3 | Upper panel illustrates the native contacts (NC) which arise from stacking or pairing between bases. The NC exist if they are within 1.5 Å of that seen in NMR (UUCG) and of that seen in conformation 2 (Fig. 3.10) for the UUUU loop. Middle panel indicates the fraction of all native contacts (NC) for the UUCG (black) and UUUU loop (red) as a function of temperature. Lower panel shows the plots of the NC for the stem (left) and the loop region (right) of both hairpins. | 67 |
| 5.4 | Upper panel shows the fluctuation of all native contacts for the UUCG in black line and for the UUUU in red dashed line. Lower panel shows the fluctuation of the native contact for the UUCG loop when divided into the subsystems. a) Fluctuation of the NC in the loop. b) Fluctuation of the NC of the loop when coupled to the loop closing base pair (C3-G8). c) Fluctuation of the NC in the stem. d) Fluctuation of the NC of the stem when coupled to the residue U4 and G7 of the loop. e) Fluctuation of the NC of the residues of the loop closing base pair C3 and G8 and the terminal residues in the loop. | 69 |
| 5.5 | Free energy surface (in units kT) obtained from the REMD simulation at 331 K for the UUCG loop, plotted as a function of the first two PCA eigenvectors. | 71 |
| 5.6 | Free energy surface (in units kT) obtained from the REMD simulation at 331 K for the UUUU loop, plotted as a function of the first two PCA eigenvectors. | 72 |
| 5.7 | Free energy surface (in units kT) obtained from the REMD simulation at 457 K for the UUCG loop, plotted as a function of the first two PCA eigenvectors. | 74 |
| 5.8 | Free energy surface (in units kT) obtained from the REMD simulation at 413 K for the UUUU loop, plotted as a function of the first two PCA eigenvectors. | 75 |
| 6.1 | Secondary structure element of TAR-RNA | 82 |
| 6.2 | Overview of the genetic algorithm | 84 |
| 6.3 | Aromatic compounds for the inhibition of the Tat-TAR complex, the last compound 3 shows the complete ligand (Arg-X-Arg). | 85 |
| 6.4 | Preferred binding sites in the free TAR for component Nr. 16 . A) Binding site in the minor groove B) Binding site in the major groove and C) Binding in the upper stem | 87 |
| 6.5 | In the first plot the docked ligands are ranked in the order of increasing energy. The last plot shows the nine compounds plotted against the number of clusters. a) Results of the docking simulations of the free TAR with altogether 30 independent simulations. b) Results of the docking simulations of the bound TAR with altogether 50 independent simulations. | 89 |

| | | |
|------|---|----|
| 6.6 | Free energy of binding plotted as a function of the 30 ranked and clustered ligands. Note no experimental data are available on these ligands yet. . . . | 90 |
| 6.7 | Surface models of the investigated TAR ligand complex. The aromatic ligands are indicated in yellow. Three main binding sites are obtained. Binding in the major groove of the bulge is indicated as site 'A', and the binding in the minor groove region is characterized as site 'B'. Site 'C' corresponds to the binding in a region of the upper stem | 91 |
| 6.8 | The studied aminoglycosides with the 1-2 and 1-6 linkages, compared to the neamine with only two torsional angles between the sugar rings | 93 |
| 6.9 | MD simulation at 300 K. Upper panel shows the RMSD of the aminoglycosides. Middle panel exhibits the angle between two vectors which are defined by two sugar atoms. Lower panel depicts the dihedral angles of the peptide bond linkage | 94 |
| 6.10 | Snapshots of the aminoglycoside taken from a 1 ns MD trajectory. (Left shows the 1-6 linked, middle and right correspond to the 1-2 linked disaccharide and the neamine, respectively) | 95 |
| 6.11 | Left figure shows a surface of the TAR RNA (blue) with the neamine in the binding pocket of the bulge, right is an atomic representation of the binding site. Black lines indicate hydrogen bonds. | 95 |

List of Tables

| | | |
|-----|--|----|
| 3.1 | Difference of the 3J coupling constants, when the Karplus parameters of VB and LK (Van Buuren given in equation (2.21) and equation (2.20) for Lankhorst) are used. | 23 |
| 3.2 | Comparison of calculated and experimental NOEs (in Å) of the UUUU loop. 26 | |
| 3.3 | Comparison of calculated and experimental torsional angles of the UUUU tetraloop (in degrees). REMD results are obtained at 300 K and compared to conventional MD at 300 K for MD1, to the trajectory of 350 K and NMR. 39 | |
| 4.1 | Comparison of calculated at 300 K and experimental torsional angles of the UUCG tetraloop (in degrees). The REMD and the MD angles stem from the 10 mer CGCUUCGGCG and the 14 mer GGCACUUCGGUGCC, respectively. NMR_s angles are taken from the 14 mer, determined in the Schwalbe group [2], while the angles of a 12 mer GGACUUCGGUCC are obtained from Allain and Varani [3]. | 50 |
| 5.1 | Torsion angles from different states for the UUCG loop obtained from the REMD simulation at 457 K. The states are gained by applying PCA on the trajectory. | 76 |
| 5.2 | Torsion angles from different states for the UUUU loop obtained from the REMD simulation at 413 K. The states are gained by applying PCA on the trajectory. | 77 |
| 5.3 | Fraction of the native contacts (NC) (upper panel) and of the hydrogen bonds (H) (lower panel) in the UUCG loop for the states (S) at the corresponding temperature (T) | 78 |
| 5.4 | Fraction of the native contacts (NC) (upper panel) and of the hydrogen bonds (H) (lower panel) in the UUUU loop for the states (S) at the corresponding temperature (T) | 79 |
| 6.1 | $\lg(IC_{50})$ values determined experimentally and the logarithm of the inhibition constants $\lg(K_i)$ obtained theoretically are shown for the free and bound TAR | 92 |
| 6.2 | Force field parameters for torsion angles | 93 |

Chapter 1

Introduction

Molecular dynamics (MD) simulation serves as an important and widely used computational tool to study molecular systems at an atomic resolution. No experimental technique is capable of generating a complete description of the dynamical structure of the biomolecules in their native solution environment. MD simulations allow us to study the dynamics and structure of the system and, moreover, helps in the interpretation of experimental observations.

MD simulation was first introduced and applied by Alder and Wainwright in 1957 [4]. However, the first MD simulation of a macromolecule of biological interest was published 28 years ago [5]. The simulation was concerned with the bovine pancreatic trypsin inhibitor (BPTI) protein, which has served as the “hydrogen molecule” of protein dynamics because of its small size, high stability, and relatively accurate X-ray structure available in 1977 [6]. This method is now widely used to tackle larger and more complex biological systems [7, 8] and has been facilitated by the development of fast and efficient methods for treating the long-range electrostatic interactions [9], the availability of faster parallel computers, and the continuous development of empirical molecular mechanical force fields [10, 11, 12].

It took several years until the first MD simulations of nucleic acid systems were performed [13, 14, 15, 16]. These investigations, which were also performed in vacuo, clearly demonstrated the importance of proper handling of electrostatics in a highly charged nucleic acid system, and different approaches, such as reduction of the phosphate charges and addition of hydrated counterions, have been applied to remedy this shortcoming and to maintain stable DNA structures. A few years later, the first MD simulation of a DNA molecule, including explicit water molecules and counterions was published [17]. Various MD simulations on fully solvated RNA molecules with explicit inclusion of mobile ions indicated the importance of proper treatment of the environment of highly charged nucleic acids [18, 19, 20, 21].

Given the central roles of RNA in the life of cells, it is important to understand the mechanism by which RNA forms three dimensional structures endowed with properties such as catalysis, ligand binding, and recognition of proteins. Furthermore,

the increasing awareness of the essential role of RNA in controlling viral replication and in bacterial protein synthesis emphasises the potential of ribonucleic acids as targets for developing new antibacterial and new antiviral drugs. Driven by fruitful collaborations in the Sonderforschungsbereich "RNA-Ligand interactions" the model RNA systems in this study include various RNA tetraloops and HIV-1 TAR RNA. For the latter system, the binding sites of heteroaromatic compounds have been studied employing automated docking calculations [22]. The results show that it is possible to use this tool to dock small rigid ligands to an RNA molecule, while large and flexible molecules are clearly problematic.

The main part of this work is focused on MD simulations of RNA tetraloops. This hairpin motif is a ubiquitous structural motif of RNA and DNA oligonucleotides [23]. RNA hairpins emerge when the single-stranded RNA molecule folds back on itself to form a double-helical stem capped by a loop of unpaired nucleotides. They play important roles in both RNA structure and function, e.g., hairpins are thought to provide nucleation sites for RNA folding [24] and tertiary recognition sites for both proteins and nucleic acids [25, 26].

In particular, tetraloops (RNA hairpins with four nucleotides) are commonplace and have therefore been the focus of recent structural investigations, including nuclear magnetic resonance (NMR)[27, 28, 29, 30, 31, 32, 33] and molecular dynamics (MD) studies [34, 35, 36]. The focus of these studies has been the investigation of stable tetraloops, especially of UNCG, GNRA, and CUUG loops, which represent the most frequently occurring hairpins both in prokaryotic and eukaryotic RNAs. The stability of this classes of tetraloops has been attributed to the additional base pairs formed between nucleotides 1 and 4 in the loop, back stacking and several 2' OH-base hydrogen bonds [35, 37].

However, also other sequences can form hairpin structures: For example, the sequence of four repetitive uridine nucleotides, flanked by a short, three basepairs containing stem, forms a stable hairpin with a melting temperature for the sequence gUUUUc of 51.5 °C and for cUUUUg of 60.4 °C. The melting temperatures are therefore only 8.6 °C and 11.3 °C decreased compared to the most stable 12mer hairpins gUUCGc and cUUCGg, respectively [38, 39, 40]. The decrease is even less pronounced for 10mer hairpins (cUUCGg 72.6 °C compared to cUUUUg 56.8 °C) [41]. From all diribonucleotides, the stacking interaction of the UpU-dinucleotide has been reported to be the weakest [42]. Hairpins containing UUUU-tetraloop have been incorporated as flanking loops for RNA quadruplexes [43] and have been investigated in the context of the functional anticodon architecture of tRNA^{Lys}, [44, 45]. In addition, the incorporation of a UUUU tetraloop in loop II of the hammerhead ribozyme lead to a 2-3 fold increase in activity compared to the more stable GCAA tetraloop in this position [46]. The findings are furthermore supported by data from Conaty et al. [47] showing that polypyrimidine rather than very stable tetraloop structure are important for efficient cleavage activity of the hammerhead ribozyme.

The structure determination of this class of flexible tetraloops is more difficult

than for stable tetraloops, since NMR data are averaged over the ensemble of conformers. Recently, extensive NMR studies on the flexible RNA tetraloops 5'-CGCUUUUGCG-3' including NOEs, various 3J couplings, and cross-correlated relaxation rates [48, 49, 50, 51] have been performed in the Schwalbe group. It was shown that the loop nucleotide U6 of the UUUU-tetraloop is conformationally averaged between the N- and the S-type sugar pucker mode [51] thus indicating significant conformational dynamics. NMR structural parameters have also been assigned to the more stable 5'-GGCACUUCGGUGCC-3' tetraloop [33]. In order to better understand the dynamics, ^{13}C spin relaxation experiments were performed for the first time for all nucleobases and ribose moieties of this loop [52]. These NMR results of the UUCG loop indicated the highest flexibility for the second loop nucleobase and ribose. Representing a structural description of these small RNA loops in great detail, these NMR results may serve as benchmark data to study the validity and accuracy of a theoretical description of the conformational dynamics using a MD simulation.

The purpose of the extensive MD studies on the UUUU and UUCG loops in this current work is three-fold. First, by comparing experimental and theoretical data we studied the quality and performance of the theoretical model, including the force field, the conformational sampling, and the modeling of NMR data via approximate formulas such as Karplus relations. Second, combining the results of NMR and simulation, a picture for the dynamic structure of the UUUU and UUCG tetraloops is provided. Especially for the flexible UUUU loop this could not be obtained by conventional NMR derived structures. In the case of the UUCG loop for which a dynamical picture from the NMR point of view is still missing for the phosphate backbone, the MD data are thought to complete the dynamical description of this loop. Third, a temperature-dependent MD study of the 10 mers 5'-CGCUUUUGCG-3' and 5'-CGCUUCGGCG-3' is performed. We obtain from this study the temperature-dependent unfolding and we can for the first time give a structural and dynamical comparison of the denaturation process of both loops.

Chapter 2

Methods

2.1 Molecular dynamics (MD) simulation

MD simulation is a technique by which one generates the atomic trajectory of system of N particles by numerical integration of the classical Newton's equation of motion,

$$\mathbf{f}_i = m_i \frac{d^2 \mathbf{r}_i}{dt^2} \quad (2.1)$$

where m_i denotes the mass of particle i , \mathbf{r}_i its position and \mathbf{f}_i the vector sum of all forces acting on particle i at its center of mass. This is an excellent approximation for a wide range of materials. Only when we consider the translational or rotational motion of light atoms or molecules (He, H₂, D₂) or a vibrational frequency ν such that $h\nu > k_B T$, we should worry about quantum effects, where h is the Planck constant, k_B the Boltzmann constant and T the absolute temperature.

There are various algorithms [53, 54] to solve Newton's equation. The Verlet integrator [55, 56] and its leap-frog variant [57] for the translational motion are commonly used in MD. The basic idea is to write two second-order Taylor expansions for the positions $\mathbf{r}(t)$, one forward and one backward in time. Calling \mathbf{v} the velocities, \mathbf{a} the accelerations, and \mathbf{b} the third derivatives of \mathbf{r} with respect to t , one has

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + 1/2\mathbf{a}(t)\Delta t^2 + \dots \quad (2.2)$$

$$\mathbf{r}(t - \Delta t) = \mathbf{r}(t) - \mathbf{v}(t)\Delta t + 1/2\mathbf{a}(t)\Delta t^2 - \dots \quad (2.3)$$

Adding the two expressions gives

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \mathbf{a}(t)\Delta t^2 \quad (2.4)$$

This is the basic form of the Verlet algorithm. Since we are integrating Newton's equations, $\mathbf{a}(t)$ is just the force divided by the mass ($\mathbf{f} = m\mathbf{a}$), and the force \mathbf{f} is in

turn a function of the positions $\mathbf{r}(t)$. The truncation error of equation 2.4 is of the order of Δt^4 , even though third derivatives do not appear explicitly. A problem with this version of the Verlet algorithm is that velocities are not directly generated. One could compute the velocities from the positions by using

$$\mathbf{v}(t) = \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t - \Delta t)}{2\Delta t} \quad (2.5)$$

However, the error associated to this expression is of order Δt^2 rather than Δt^4 . To overcome this difficulty, some variants of the Verlet algorithm have been developed. They give rise to exactly the same trajectory, and differ in what variables are stored in memory and at what times. The leap-frog algorithm as already mentioned is one of such variants [57] where velocities are handled somewhat better. This algorithm evaluates the velocities at half-integer time steps and uses the velocities to compute the new positions.

$$\mathbf{v}(t + \Delta t/2) = \mathbf{v}(t - \Delta t/2) + \Delta t \frac{\mathbf{f}(t)}{m} \quad (2.6)$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t \mathbf{v}(t + \Delta t/2) \quad (2.7)$$

In the leap-frog formulation the velocity and the position are not known at the same point of time. Therefore, the name “leap-frog”.

The most time-consuming part of almost all MD simulations is the calculation of the force acting on every particle. The calculation of the force is pairwise additive, meaning we have to consider the contribution to the force on particle i due to all its neighbors. This implies the time needed for the evaluation scales as N^2 (number of particle). Nevertheless, there exist techniques to speed up the evaluation of forces in such a way that the computing time scales as N [53, 55].

The calculated force at each time step is defined by the so called force field, which has the complete information on the topology of the molecular system. The force field contains the parameters needed to describe the potential V of the system as a function of the position \mathbf{r} of the particles, The differentiation of V leads us to the forces, which are required to solve Newton’s equation. The potential has the following form,

$$\begin{aligned} V(\mathbf{r}) = & \frac{1}{2} \sum_i k_{b_i} (b - b_0)^2 + \frac{1}{2} \sum_i k_{\theta_i} (\theta - \theta_0)^2 \\ & + \frac{1}{2} \sum_i k_{\phi_i} (1 + \cos(n\phi - \delta)) \\ & + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right] \end{aligned} \quad (2.8)$$

where the first term determines the covalent bond, the second describes the bond angle, the third is responsible for the torsional angle, and the last term defines

the Lennard-Jones and Coulomb interactions, respectively. For example, the bond length b_0 and the force constant k_{b_i} are equilibrium values determined by the force field. b is the bond length between two atoms at time t during an MD run. The definitions of the Lennard-Jones and the Coulomb interactions do not contain harmonic potentials but are described by their van-der-Waals-Parameter A_{12} and B_{12} and by their dielectric constant ϵ in the case of the Coulomb interaction. The validity of the force field depends strongly on these parameters. Empirical force field for nucleic acids perform surprisingly well, despite the absent of dynamic polarizability and the simplicity of the models [58]. In chapter 3, for example, an MD simulation on a 15 ns timescale is shown to reproduce well the NMR NOESY data and widely applied nucleic acid force fields do an excellent job of reproducing a variety of nucleic acid structures [59, 60]. More practical questions regarding this method are whether or not a sufficient region of phase space is explored by the system trajectory to yield satisfactory time averages within a feasible amount of computer time, and whether thermodynamic consistency can be attained between simulations with identical macroscopic parameters (density, energy etc) but different initial conditions (atomic positions and velocities). Therefore, methods to speed-up the simulation also increase the sampling of available conformations, and since this is very important in many applications, several techniques have been developed to further increase conformational sampling in MD [61, 62, 63, 64]. Two of these methods will be used in this study.

MD simulations are in many aspects very similar to real experiments. In experiment and simulations the sample is connected to a measuring instrument for example to a thermometer and one can measure the property of interest during a certain time interval. In experiment we have an exchange of temperature and pressure with the surrounding, whereas in MD simulations these conditions are realized by certain correction factors. The sum of kinetic and potential energy of a simulation system would stay constant if there were no temperature adjustments. This would be equivalent to a closed system. To simulate the exchange of temperature with the surrounding, the system is coupled to a temperature bath. As the temperature can be obtained via the kinetic energy, all particle's velocities are scaled by a factor to maintain the temperature [65].

$$E_{\text{kin}}(t) = \sum_{i=1}^N \frac{1}{2} m_i v_i^2(t) = \frac{1}{2} N_f k_B T(t) \quad (2.9)$$

In practice, we would measure the total kinetic energy of the system and divide this by the number of degrees of freedom N_f ($3N-3$ for a system of N particle). The pressure coupling is also based on the same idea like the temperature coupling [65]. For the simulation the particles of interest are placed in a box, using periodic boundary conditions. Running simulation in a periodic boundary box overcomes the problem of surface effects. The box is replicated throughout space to form an infinite lattice (Fig.2.1). In the course of the simulation, as a molecule moves in the original box, its periodic image in each of the neighboring boxes moves in exactly the same

The LES approach is designed to enhance the sampling of a small subsystem of interest. Let l denote the small, and L the large subsystem, respectively, while $S = \{l \cup L\}$ comprises the entire system. Let N_L and N_l be the number of atoms in the large and small systems, respectively. Furthermore, \mathbf{q}_l are the $3N_l$ Cartesian coordinates of the small subsystem and \mathbf{q}_L are the $3N_L$ Cartesian coordinates of the large system. The potential energy of the entire system is given by

$$V(\mathbf{q}_L, \mathbf{q}_l) = V_{LL}(\mathbf{q}_L) + V_{ll}(\mathbf{q}_l) + V_{Ll}(\mathbf{q}_L, \mathbf{q}_l), \quad (2.10)$$

where V_{LL} denotes the self energy of the small subsystem, and V_{Ll} represents the interaction energy between the large subsystem and the smaller subsystem. If, in the LES method, N copies of the small system are introduced with the index i designating one of them, the LES potential is given by

$$V_{LES}(\mathbf{q}_L, \mathbf{q}_l) = V_{LL}(\mathbf{q}_L) + \frac{1}{N} \sum_{i=1}^N [V_{ll}(\mathbf{q}_l^i) + V_{Ll}(\mathbf{q}_L, \mathbf{q}_l^i)] \quad (2.11)$$

2.2.2 Replica exchange molecular dynamics (REMD) simulation

The method of parallel tempering or replica exchange simulation is a Monte Carlo scheme that has been derived to achieve good sampling of systems that have a free energy landscape with many local minima. In parallel tempering, M systems are considered. Each of these systems are in a different thermodynamic state. Usually but not necessarily, these states differ in temperature. The REMD approach has been described and successfully applied to various systems [70, 71, 72, 73, 74]. In brief, the method constructs a generalized ensemble from M noninteracting trajectories with temperatures T_m ($m = 1, \dots, M$). A state of this generalized ensemble is characterized by $X = \{\dots, \mathbf{x}_m^{(i)}, \dots\}$, where $\mathbf{x}_m^{(i)}$ represents the coordinates $\mathbf{q}_m^{(i)}$ and momenta $\mathbf{p}_m^{(i)}$ of all atoms of the i th replica at temperature T_m . The algorithm consists of two consecutive steps: (i) Independent constant-temperature simulations of each replica and (ii) exchange of the trajectories (i.e., the coordinates) between the replicas (defined by the temperature) according to a Metropolis-like criterion. Considering, for example, the exchange $X = \{\dots, \mathbf{x}_m^{(i)}, \dots, \mathbf{x}_n^{(j)}, \dots\} \rightarrow X' = \{\dots, \mathbf{x}_m^{(j)}, \dots, \mathbf{x}_n^{(i)}, \dots\}$, the criterion is given by

$$W(\mathbf{x}_m^{(i)}, \mathbf{x}_n^{(j)}) = \begin{cases} \exp(-\Delta) & \text{if } \Delta > 0 \\ 1 & \text{if } \Delta \leq 0 \end{cases} \quad (2.12)$$

$$\Delta = [E(\mathbf{q}^{(j)}) - E(\mathbf{q}^{(i)})][(k_B T_m)^{-1} - (k_B T_n)^{-1}], \quad (2.13)$$

where $E(\mathbf{q}^{(i)})$ and $E(\mathbf{q}^{(j)})$ are the potential energies of the i th and j th replica, respectively. After the exchange, the momenta $\mathbf{p}'^{(i)}$ and $\mathbf{p}'^{(j)}$ of the replicas are given

by

$$p^{(i)} = \sqrt{\frac{T_n}{T_m}} p^{(i)}, \quad p^{(j)} = \sqrt{\frac{T_m}{T_n}} p^{(j)}. \quad (2.14)$$

Iterating steps (i) and (ii), the trajectories of the generalized ensemble perform a random walk in temperature space. This facilitates an efficient conformational sampling of the energy landscape of the system, even in the presence of multiple local minima. To obtain the canonical average of any physical quantity at a specific temperature, one performs a time average over all trajectory segments of the replica corresponding to this temperature.

2.3 Principal component analysis (PCA)

To identify the most important conformational degrees of freedom of a simulation, the principal component analysis (PCA) method, also called quasiharmonic analysis or essential dynamics method, has been proven useful [75, 76, 77, 78, 79]. The approach is based on the covariance matrix

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle, \quad (2.15)$$

where x_1, \dots, x_{3N} are the mass-weighted Cartesian coordinates of the N-particle system and $\langle \dots \rangle$ denotes the average over all sampled conformations. Since the covariance matrix provides information on the correlated fluctuations of pairs of atoms, its eigenvectors and eigenvalues of σ yield the modes of collective motion and their amplitudes. To remove the overall rotation of the trajectory, its initial geometry was chosen as reference structure for the rotational fit. If the system does not undergo major large amplitude motion, the results of the PCA is quite insensitive to the choice of the reference structure.

Since we wish to focus on the conformation dynamics of the loop region, the PCA was restricted to the atoms of the four loop residues. By diagonalizing the resulting covariance matrix, we obtain the eigenvectors v_n and the eigenvalues λ_n , which are organized in descending order, i. e., λ_1 represents the largest eigenvalue. The results of the PCA on the RNA tetraloops indicate the motions along the first two eigenvectors v_1, v_2 of σ to contain around 50 % of the fluctuations. Restricting ourselves to these first two eigenvectors, the free energy surface along these vectors is given by

$$\Delta G(v_1, v_2) = -k_B T [\ln P(v_1, v_2) - \ln P_{\max}]. \quad (2.16)$$

Here $P(v_1, v_2)$ is the probability distribution obtained from a histogram of the MD data and P_{\max} denotes the maximum of the distribution, which is subtracted to ensure that $\Delta G = 0$ for the lowest free energy minimum. The representation of the free energy surface as a function of higher eigenvectors for the RNA tetraloops was found to not yield significantly more information. This can be seen in Figure 3.8, which will be discussed in chapter 3 and 5.

2.4 Calculation of NMR parameters

The conformation in nucleotides is defined by its backbone torsion angles (α , β , γ , δ , ϵ , and ζ), ribose torsion angles ($\nu_0 - \nu_4$) and the torsion angle of the glycosidic bond χ (Fig. 2.2). The ribose sugar ring is due to the so-called ‘‘Pitzerspannung’’ not planar [80]. Changes in the sugar ring conformation (pucker or pseudorotation) is determined by two parameters, the pseudorotational phase and the maximal amplitude v_{max} [81, 82, 83]. The dependence of these parameters on the torsion angles ($\nu_0 - \nu_4$) is shown in the following

$$\tan P = \frac{(\nu_4 + \nu_1) + (\nu_3 + \nu_0)}{2\nu_2(\sin 36^\circ + \sin 72^\circ)} \quad (2.17)$$

$$v_{max} = \frac{\nu_2}{\cos P} \quad (2.18)$$

The pseudorotational cycle is divided into sectors, where each sector is 36° large. For example, if $P = 18^\circ$ then the sugar ring is in the C3'endo conformation. In the C2'-endo conformation the pseudototaional angle is around $P = 162^\circ$. In the lower panel of Fig. 2.2 the two preferred conformations of the sugar molecule in RNA are shown. In the C3'-endo conformation the C3' atom is out of plane (C1', C2', C4' and O2'), whereas in the C2'-endo conformation the C3' atom is out of plane (C1', C3', C4' and O2'). In A-RNA the C3'-endo is preferred and in B-RNA the sugar ring has the C2'-endo conformation. To verify our MD simulations, NMR structural parameters like 3J couplings cross-correlated relaxation rates, NOEs and order parameter S^2 are calculated from the trajectories and compared to experiments. We briefly go here into details only for the first two methods.

To calculate the 3J couplings shown in this study, the following Karplus relations (in units of Hz) were employed [84, 85]:

$$^3J_{H,P} = 15.3 \cos^2 \phi - 6.2 \cos \phi + 1.5, \quad (2.19)$$

$$^3J_{C,P} = 6.9 \cos^2 \phi - 3.4 \cos \phi + 0.7, \quad (2.20)$$

$$^3J_{C,P} = 8.0 \cos^2 \phi - 3.4 \cos \phi + 0.5, \quad (2.21)$$

$$^3J_{H,H} = 13.24 \cos^2 \phi - 0.91 \cos \phi + \sum_{i=1}^4 \Delta\chi_i \{0.53 - 2.41 \cos^2(\zeta_i \phi + 15.5|\Delta\chi_i|)\}. \quad (2.22)$$

Equation (2.19) was used to calculate the couplings $^3J_{H5',proS,R,P}$ ($\phi = \beta \pm 120^\circ$) and $^3J_{H3',P}$ ($\phi = \epsilon + 120^\circ$), Eq. (2.20) was used to calculate the couplings $^3J_{C4',Pi}$ ($\phi = \beta$), $^3J_{C2',P}$ ($\phi = \epsilon - 120^\circ$), and $^3J_{C4',Pi+1}$ ($\phi = \epsilon$), and Eq. (2.22) was used to calculate the couplings $^3J_{H1',H2'}$, $^3J_{H2',H3'}$, and $^3J_{H3',H4'}$, where ϕ is a function of the pseudorotation angle P of the sugar pucker and the parameters $\Delta\chi_i$ and ζ_i for the hydrogen pairs of the sugar ring are given in Ref. [84]. In all cases, the 3J couplings

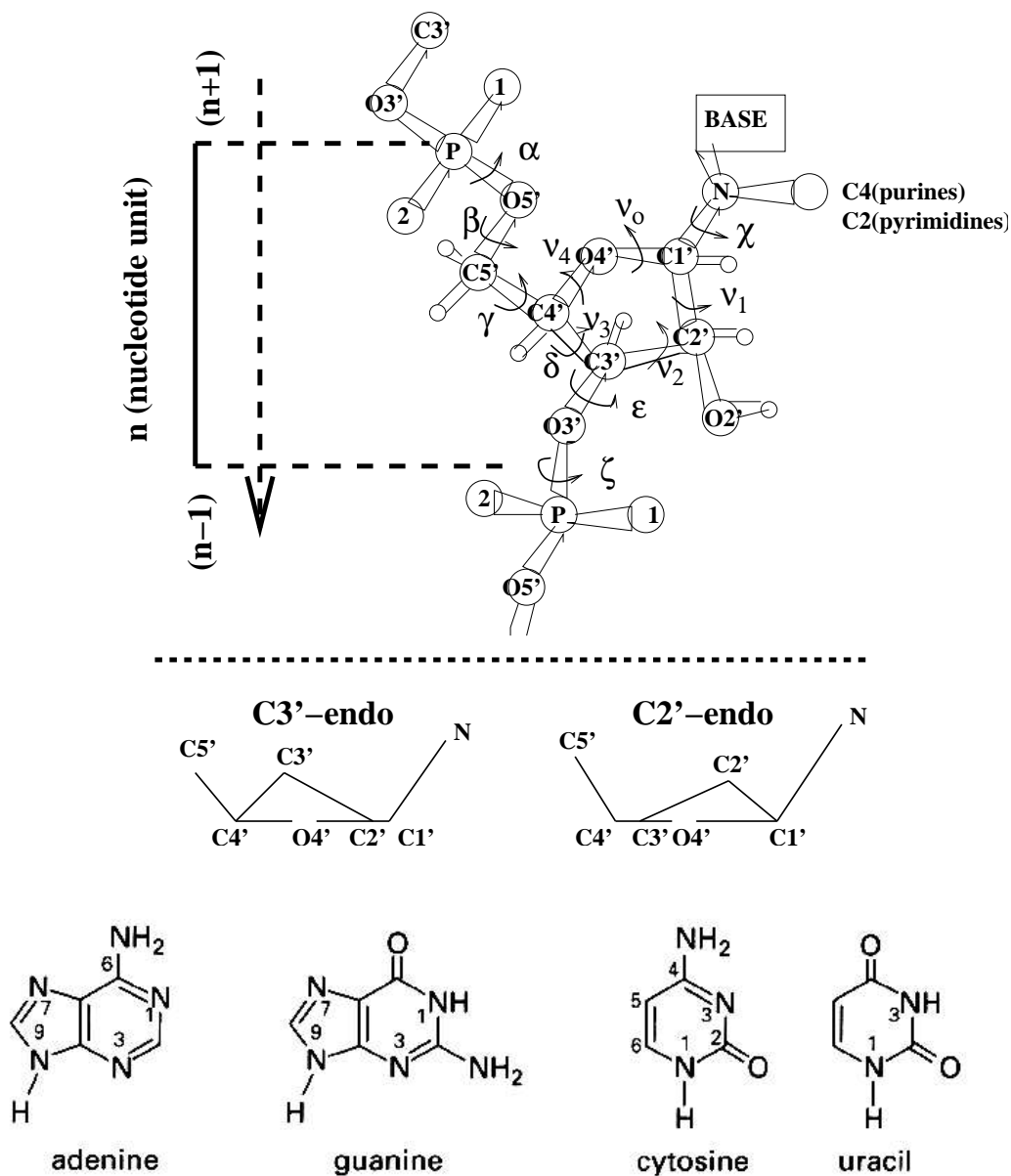


Figure 2.2: Illustration of the torsional angles (upper panel), the the pseudorotational states (middle panel) and the RNA bases (lower panel). The conformation of the backbone is determined by the dihedral angles α , β , γ , δ , ϵ , and ζ . The conformation along the glycosidic bond is determined by the χ angle. The torsion angles of the sugar ring are denoted by $\nu_0 - \nu_4$.

were obtained by averaging Eqs. (2.19)-(2.22) over all angles sampled by the MD trajectory.

To calculate the cross-correlated relaxation rate $\Gamma_{C_iH_i, C_jH_j}^c$ between the dipolar interactions of two distinct carbon-proton spin pairs (C_i, H_i) and (C_j, H_j) , we employed the expression [49, 50]

$$\Gamma_{C_iH_i, C_jH_j}^c = \frac{2}{5} \frac{\gamma_H^2 \gamma_C^2}{r_{C_iH_i}^3 r_{C_jH_j}^3} \left(\frac{\mu_0}{4\pi} \right)^2 \hbar^2 (S_{ij}^c)^2 \left(\frac{3\cos^2\theta_{ij} - 1}{2} \right) \tau_c, \quad (2.23)$$

where γ_C and γ_H are the gyromagnetic ratios, μ_0 is the susceptibility of vacuum, $r_{C_iH_i}$ and $r_{C_jH_j}$ are the carbon-proton distances, τ_c is the overall correlation time, (S_{ij}^c) is an order parameter accounting for the internal mobility of the dipole tensors of C_iH_i and C_jH_j , and θ_{ij} is the projection angle between these dipole tensors, which are oriented parallel to the respective carbon-proton bond vectors. In the calculations reported below, we have assumed rigid C-H vectors, which results in $S_{ij}^c = 1$. Furthermore, we choose $\tau_c = 1.8$ ns, as determined in experiment. From the MD simulation, the bond distances $r_{C_iH_i}$ and $r_{C_jH_j}$ as well as the projection angles θ_{ij} are obtained.

The NMR determination of the angles α and ζ in RNA have been accomplished through the structural interpretation of the cross-correlated relaxation [50] of 1H , ^{13}C -dipolar coupling and ^{31}P -chemical shift anisotropy (^{31}P -CSA). The dipole ^{31}P -CSA cross-correlated rates can be written as [50]

$$\Gamma_{CH,P}^{DD,CSA} = -\frac{2}{15} \gamma_P B_0 \tau_c \hbar \frac{\mu_0}{4\pi} \frac{\gamma_H \gamma_C}{r_{CH}^3} S_{CSA,DD}^2 \{ (\sigma_{22}^P - \sigma_{11}^P)(3\cos^2\theta_{CH,\sigma_{22}} - 1)(\sigma_{33}^P - \sigma_{11}^P)(3\cos^2\theta_{CH,\sigma_{33}} - 1) \} \quad (2.24)$$

$\Gamma_{CH,P}^{DD,CSA}$ depends on the projection angles $\theta_{CH,\sigma_{22}}$ and $\theta_{CH,\sigma_{33}}$ of the CH-dipole tensor parallel to the CH bond vector and the $(\sigma_{22} - \sigma_{11})$ and $(\sigma_{33} - \sigma_{11})$ components of the ^{31}P -CSA tensors. The components have been calibrated from model compounds with σ_{11} , σ_{22} and σ_{33} being -76 ppm, -16 ppm, and 103 ppm, respectively [50].

It is known that the internal dynamics of biological macro molecules play an important role in their biological function [86, 87]. NMR spectroscopy is one of the most powerful experimental techniques for the characterization of such dynamics [88]. In standard NMR relaxation experiments three relaxation constants, R1, R2 and NOES are measured at many sites along the backbone and in the side chains using isotropically labeled proteins, sometimes, at multiple magnetic fields. Site specific information about the amplitude and time scale of internal motions can be derived from these measurements and provide important insights into their function [89, 90]. Assuming that the global and internal motions are separable, the

widely used model-free formalism [89, 91] characterizes internal dynamics in terms of generalized order parameters S^2 and time constants, which are motional model independent measurements of the degree of spatial restriction and rate of internal motions respectively. MD is particularly powerful in simulating the ps/ns time scale motions that are relevant to NMR relaxation and provides an effective approach to evaluate various methods for analyzing NMR relaxation data. The above mentioned relaxation constants are determined by the spectral density function $J(\omega)$ [89, 92]. $J(\omega)$ is the Fourier transform of the angular auto-correlation function $C(t)$

$$J(\omega) = 2 \int_0^{\infty} C(t) \cos \omega t dt \quad (2.25)$$

The correlation function describes reorientation of the bond vector due to both the overall molecular tumbling and internal dynamics. Assuming that the overall molecular tumbling is much slower than the internal motions, one can decompose $C(t)$ as [93]

$$C(t) = C_o(t)C_I(t) \quad (2.26)$$

$$C_I(t) = \langle P_2(\tilde{\mu}(0) \cdot \tilde{\mu}(t)) \rangle \quad (2.27)$$

where $C_o(t)$ and $C_I(t)$ are the correlation functions for the overall and internal motions, respectively. Assuming furthermore, that $C_I(t)$ can be described by a single exponential, we obtain

$$C_I(t) = S^2 + (1 - S^2)e^{-t/\tau_e} \quad (2.28)$$

$$S^2 = \lim_{t \rightarrow \infty} C_I(t) \quad (2.29)$$

in which the S^2 reflects the amplitude of the internal motions and τ_e the time constant. $P_2(x) = (3x^2 - 1)/2$ is the second order Legendre polynomial and $\tilde{\mu}(t)$ describes the orientation of the bond vector in the molecular frame. In this work the S^2 values are calculated from the trajectories, using first equation (2.27) to compute the correlation function then equation (2.28) or (2.29) can be taken to calculate the order parameter S^2 .

Chapter 3

Structure and dynamics of the UUUU loop

3.1 MD simulation of the UUUU loop

3.1.1 Computational details

All simulations were performed using the AMBER6 simulation program package [94] employing the all-atom force field parm98 [11]. The UUUU loop consists of 314 atoms which were placed in a periodic truncated octahedral box of TIP3P water [95]. Including nine Na^+ ions to neutralize the system, a total of 7709 atoms is obtained. The equation of motion was integrated by using a leapfrog algorithm with a time step of 2 fs. The SHAKE algorithm [96] was used to constrain covalent bonds to hydrogen atoms with a relative geometric tolerance of 10^{-4} . A cutoff of 10 Å was chosen for the non-bonded van der Waals interactions. We used the particle-mesh Ewald method [97] to treat the long-range electrostatic interactions and updated the non-bonded interaction pair-list every 10 fs. The solute and solvent were separately weakly coupled to external temperature baths at 300 K [65], using a coupling constant of 0.2 ps. The total system was also weakly coupled to an external pressure bath at 1 atm using a coupling constant of 0.5 ps.

From about one hundred available NMR structures of the UUUU loop, four representative MD starting structures were selected by using the clustering program NMRCLUST [98]. For each structure, the equilibration protocol consisted of 200 steps of steepest-descent minimization applied to the solvent molecules with fixed solute, followed by 100 ps of MD simulation of the solvent with fixed UUUU loop, and another 500 ps simulation without position constraining of UUUU loop. During the subsequent unconstrained MD runs, two of the four NMR structures showed base-pair disruption after 0.5 and 3 ns, respectively. The simulations for the

remaining two NMR structures, referred to as MD1 and MD2, were continued for another 12 ns, whereby the coordinates were saved every 1 ps for analysis.

To improve the conformational sampling obtained in the MD simulations, two strategies were employed. First, we performed a 18 ns high-temperature MD run at 350 K, using the final structure of trajectory MD1 as starting geometry. Furthermore, the locally enhanced sampling (LES) method [69, 68] as implemented in the AMBER6 program package was employed. In order to focus the computational effort on the part of the system that is of interest (say, a flexible loop region), the LES method replaces the trajectory of this part by an ensemble of copies, which interact with the remaining system through an average over the copies. This mean-field description affects a smoothing of the energy landscape by a lowering of the barriers, thus improving the conformational sampling of the subsystem of interest. Following Refs. [68], we performed a 4 ns LES simulation using five copies of the loop region.

3.1.2 Dynamics of the UUUU loop

As a first check of quality, we have considered the root mean square distance (RMSD) of all trajectories with respect to their initial NMR starting structure. As a representative example, Fig. 3.1 shows the time evolution of the RMSD of trajectory MD1 at 300K. The RMSD is seen to fluctuate around 4.5 Å for times $\gtrsim 2$ ns, thus reflecting stable trajectory. Also shown is the RMSD of the high-temperature MD run at 350 K. Due to the increased temperature, the RMSD for this trajectory exhibits somewhat larger fluctuations. Nevertheless, the trajectory remains stable throughout the propagation. To obtain an impression of the overall behavior of the various MD runs, Fig. 3.2 shows a circle representation of the torsional angles of the four loop residues. Generally speaking, it is seen that most angles of trajectory MD1 are well described by a single mean value, thus indicating that the system predominantly stays in a single conformation. Compared to MD1, the conformational fluctuations of MD2 are significantly higher, which reflects a certain amount of conformational heterogeneity. The two trajectories at 300 K also differ in the state of the U6 sugar pucker, that is, MD1 is in C3'-endo and MD2 is in C2'-endo. Furthermore, Fig. 3.2 shows the results obtained for the two additional trajectories with enhanced sampling, i.e., the high-temperature MD run at 350 K and the LES trajectory at 300 K. Both simulations clearly sample a larger range of coordinate space than trajectories MD1 and MD2. In particular, this is true for the backbone dihedral angles α and χ as well as for the pseudorotation angles P , which reflect conformational transitions absent in the conventional runs MD1 and MD2.

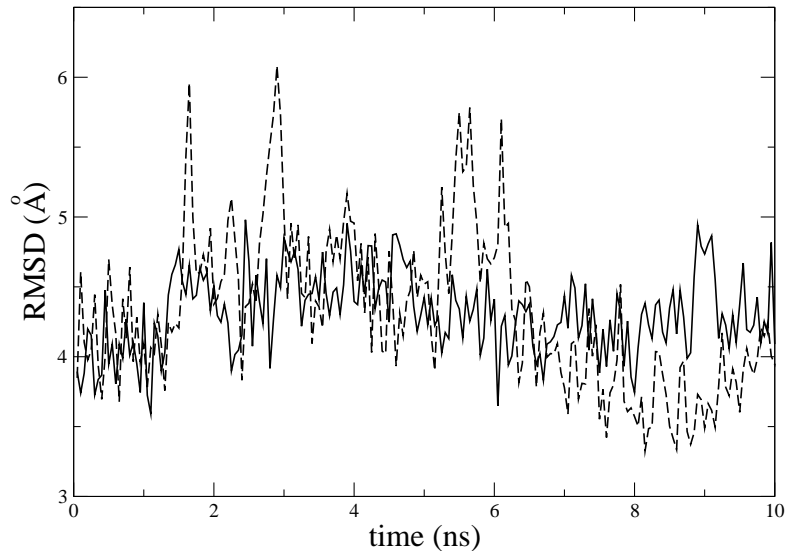


Figure 3.1: Time evolution of the RMSD obtained for trajectories MD1 at 300 K (solid line) and for the 350 K simulation (dashed line).

3.1.3 Backbone angle analysis

The comparison of the backbone angles of the UUUU loop shows an overall good agreement for the angles β , γ , and δ , but significant deviations for the angles α , ϵ and ζ . In experiment, the results were derived from a global fit of several ^3J couplings and cross-correlated relaxation rates [48, 49, 50]. From this fit, the results were chosen that either were closest to the structure of the canonical form or had the smallest RMSD fit in the absence of a complete structure calculation. MD simulations provide an additional means to analyze the experimental data.

Using nucleotide U7 of the UUUU loop as a representative example, (Fig. 3.3) shows various ^3J couplings that reflect the backbone angles β (a) and ϵ (c), as well as the ^{31}P -CSA, dipole-dipole cross-correlated relaxation rates that reflect the backbone angles α , β (b) and ϵ and ζ (d). The β angle can be obtained from the $^3\text{J}_{\text{H}5'\text{proS/R,P}}$ and $^3\text{J}_{\text{C}4',\text{P}_i}$ coupling constants. The measured couplings are consistent with a single conformation around $\beta = 165^\circ$. For the determination of α , the CSA cross-correlated relaxation rate from the C5'H5' dipole onto the phosphor CSA tensor is considered. Figure 3.3) (b) shows the α , β dependence of this rate, adopting the simple model given in Ref. [50]. Assuming that $\beta = 165^\circ$ three possible values for α are obtained, $\alpha = -120^\circ$, 29° and 75° . As $\alpha = -120^\circ$ is close to the value of the canonical conformation, this value was chosen in the interpretation of the experimental data in Ref [50]. The dihedral angle ϵ is obtained from a Karplus analysis of the coupling constants

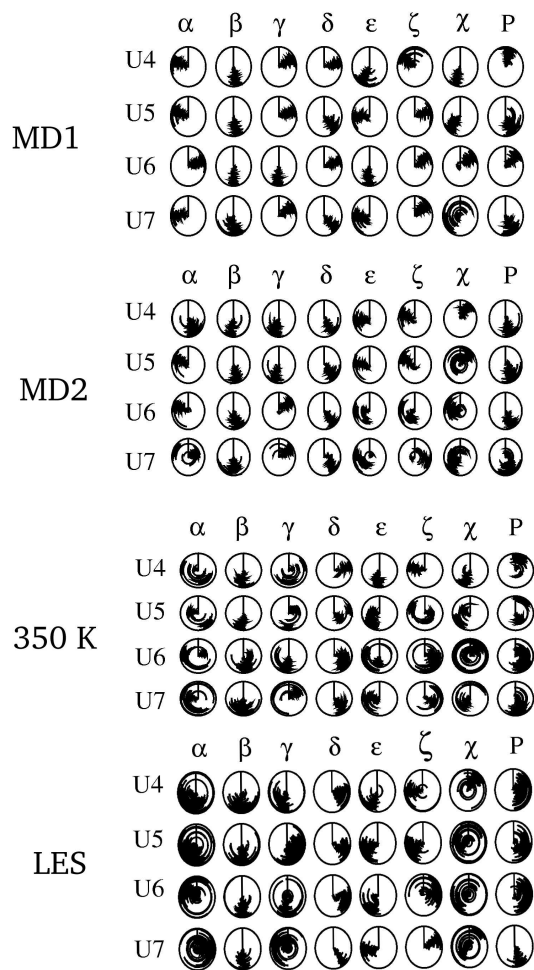


Figure 3.2: Circle representation of the torsional angles of the UUUU loop as obtained from the two trajectories at 300 K, MD1 and MD2, as well as from the 350 K trajectory and the LES trajectory at 300K.

${}^3J_{\text{H}3',\text{P}_{i+1}}$, ${}^3J_{\text{C}2',\text{P}_i}$ and ${}^3J_{\text{C}4',\text{P}_{i+1}}$ shown in Fig 3.3)(c). The fitting of the data yields two possible, solutions for the angle $\epsilon = -120^\circ$ and -237° . In the experimental analysis [50], the latter result was favored, because the RMSD obtained for $\epsilon = -237^\circ$ is lower (± 0.6) compared to (± 2.2) obtained for $\epsilon = -120^\circ$. Assuming ϵ to be either -120 or -237 , the dihedral angle ζ can be extracted from the (ϵ, ζ) -dependence of the corresponding CSA cross-correlated relaxation rate shown in Figure 3.3)(d). We obtain $\zeta = -185^\circ$, 25° or 55° for $\epsilon = -237^\circ$, and $\zeta = -225^\circ$ or -45° for $\epsilon = -120^\circ$. In the experimental analysis [50] $\zeta = -185^\circ$ was favored since it is close to the value of the canonical conformation.

Fig. 3.4 shows the joint probability distribution of the angles α and β and ϵ and ζ as obtained for trajectory MD1 and MD at 350 K. While at 300 K all four backbone angles are well described by a single mean value, the enhanced sampling calculation at 350 K reveals that several conformations may be accessible. Also shown are the values (as red dots) of the backbone angles α , β , ϵ and ζ that are consistent with the analysis of the experimental data, as well as the values (as asterisks) of the canonical A and B forms, respectively. Taking all possible experimental and calculated values shown into consideration, we find a good overall agreement of NMR and simulation. This agreement, however, is not always obtained for the values originally assigned from the experimental analysis. Considering again the U7 as a representative example, we find that β is well reproduced for all nucleotides, since it is well described by a single value in calculation and experiment. This is in contrast to the findings, for which the enhanced sampling calculation suggests that several -typically two- regions of α are thermally populated. As a consequence, two experimental values, $\alpha = -120^\circ$ and 75° are consistent with the MD data.

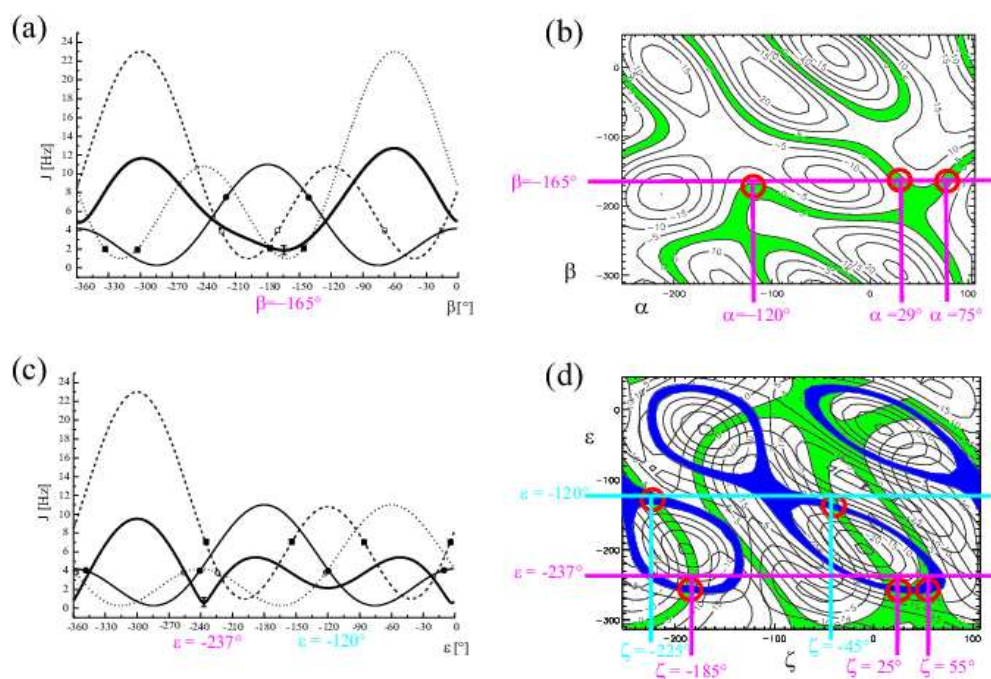


Figure 3.3: Determination of the backbone angles α , β , ϵ and ζ from experimental 3J couplings (left) and CSA cross-correlated relaxation rates (right), shown for the example of the U7 nucleotide of the UUUU loop. (a) Experimental values (as boxes) and theoretical Karplus curves (Eqs. (1) and (2)), as lines) of various 3J couplings reflecting the angle β . Shown are $^3J(\text{H5'proSi}, \text{Pi})$ drawn as square and dashed lines, $^3J(\text{H5'proRi}, \text{Pi})$ drawn as filled box and dotted lines and $^3J(\text{C4'i}, \text{Pi})$ drawn as filled circle and lines. (b) Theoretical dependence of the CSA cross-correlated relaxation rates $\Gamma_{(\text{C5'i}, \text{H5'i}), (\text{Pi})}^c + \Gamma_{(\text{C5'i}, \text{H5'i}), (\text{Pi})}^c$ on the angles β and α . The red circles indicate the conformational regions which fulfill the experimental cross-correlated relaxation rates. Assuming $\beta = -165$ as obtained in (a), possible values for the angle α are -120 , 29 , and 75 (pink bars). (c) Various 3J couplings reflecting the angle ϵ . Shown are $^3J(\text{H3i}, \text{Pi}+1)$ drawn as filled squares and dashed lines. $^3J(\text{C2i}, \text{Pi})$ drawn as circles and dotted lines, and $^3J(\text{C4i}, \text{Pi}+1)$ drawn as filled circles and lines. (d) Theoretical dependence of the CSA cross-correlated relaxation rates (in green) and (in blue) on the angles ϵ and ζ . The red circles indicate the conformational regions which fulfill the experimental cross-correlated relaxation rates.

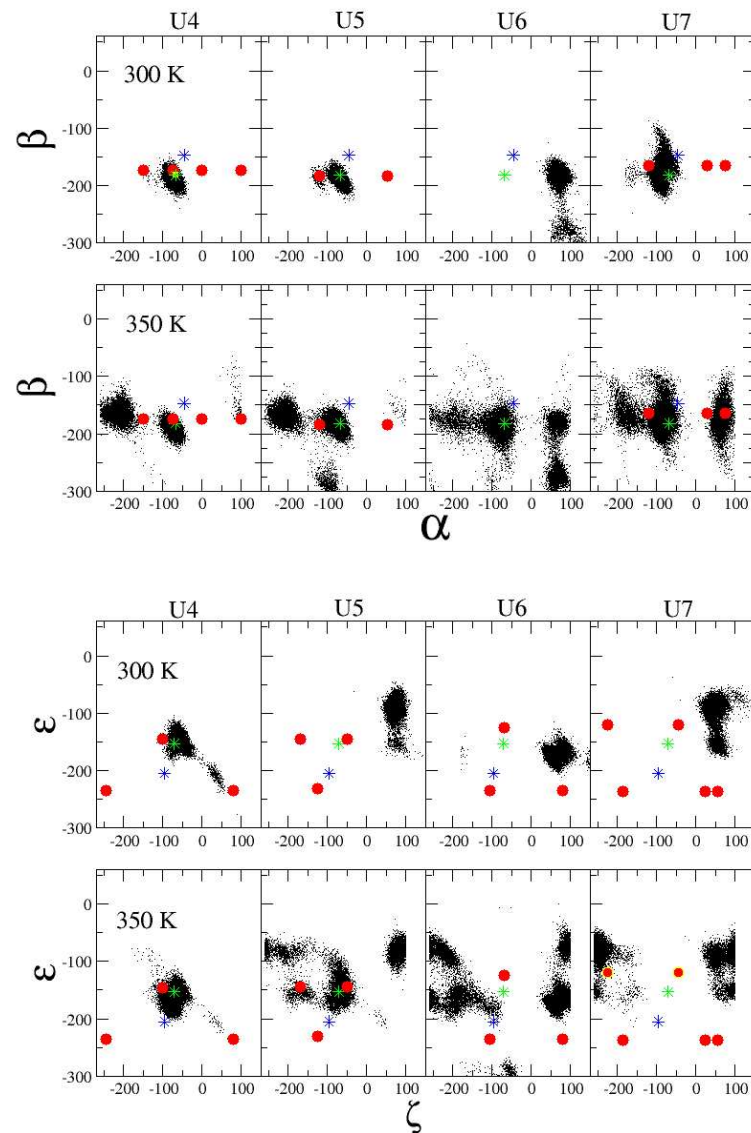


Figure 3.4: Joint probability distribution of the α and β (upper panels) and ϵ and ζ (lower panels) as obtained for trajectory MD1 at 300 K and for the 350 K run. The red dots correspond to values of the angles that are consistent with the analysis of the experimental data in Fig 3.3. The green and blue asterisks display the values of the canonical A and B form, respectively.

A clear change of the original data analysis is found for the angles ϵ and ζ , since the MD calculations seem to clearly outrule the originally adopted value $\epsilon = -237^\circ$. Assuming $\epsilon = -120^\circ$, we obtain a reasonable agreement for the ζ angle, for the nucleotides U4, U5 and U7, whereas the situation is still not clear for U6.

While the MD structures are directly obtained from the trajectories, the experimental structures are calculated from various NMR data by additionally employing theoretical models such as Karplus relations. To investigate the validity and the accuracy of the latter, it is instructive to reverse this procedure by employing the models to calculate NMR observables from the MD trajectory. As an example, Fig. 3.6 compares various 3J couplings as obtained from NMR and MD. Let us first consider the couplings $^3J_{H5'_{\text{proS/R,P}}}$ and $^3J_{C4',P_i}$, which reflect the angle β . Consistent with the perfect agreement of the experimental and calculated values for β in Fig. 3.4, the calculated couplings $^3J_{H5'_{\text{proS/R,P}}}$ match well the NMR data. Surprisingly, however, the NMR and MD results obtained for $^3J_{C4',P_i}$ differ significantly, which suggests that the Karplus relation (2.20) may not be appropriate in this case. Different Karplus relations have been tested as seen in Table 3.1, but still we were not able to significantly improve the agreement to NMR. As a second example, we consider the couplings $^3J_{H3',P}$, $^3J_{C2',P}$, and $^3J_{C4',P_{i+1}}$, which reflect the angle ϵ . While experimental and calculated 3J couplings and ϵ perfectly agree for U4, the comparison is less consistent for the remaining residues. According to the couplings $^3J_{H3',P}$ and $^3J_{C4',P_{i+1}}$, for example, the values of ϵ for U7 obtained from NMR and MD1 should match well, which, however, is not the case. The discrepancy is also not removed by improving the sampling, which suggests that either the force field or the Karplus relation (or both) cause the problem. In the case of U6, on the other hand, enhanced sampling leads to a significant improvement of the 3J couplings. This result clearly reflects the large flexibility of the U6 residue, the appropriate description of which requires advanced sampling techniques.

The backbone angles α , β , ϵ and ζ are determined from experimental 3J couplings (left) and CSA cross-correlated relaxation rates (Fig. 3.3). A comparison of the 3J couplings between MD and NMR have just been shown, therefore it is also interesting to see whether an agreement can also be found for the dipole CSA cross-correlated relaxation rates. The results of the comparison can be seen in Figure 3.5. A overall good agreement with NMR can be only be detected for $\Gamma_{(C5',H5')}^c, (P_i) + \Gamma_{(C5',H55')}^c, (P_i)$ in the case of the 350 K trajectory. The reasons for the deviations of the other rates can have many reasons. First, the only parameters that stem from MD simulations are the θ angle and the distance of the CH bond. All other parameters are taken from the experimental results that have been determined on other systems, rather than the UUUU loop [50], for example for the $^{31}\text{-P}$ -CSA tensors components. From the theoretical point of view the insufficiency of the force field may also contribute to the observed deviations.

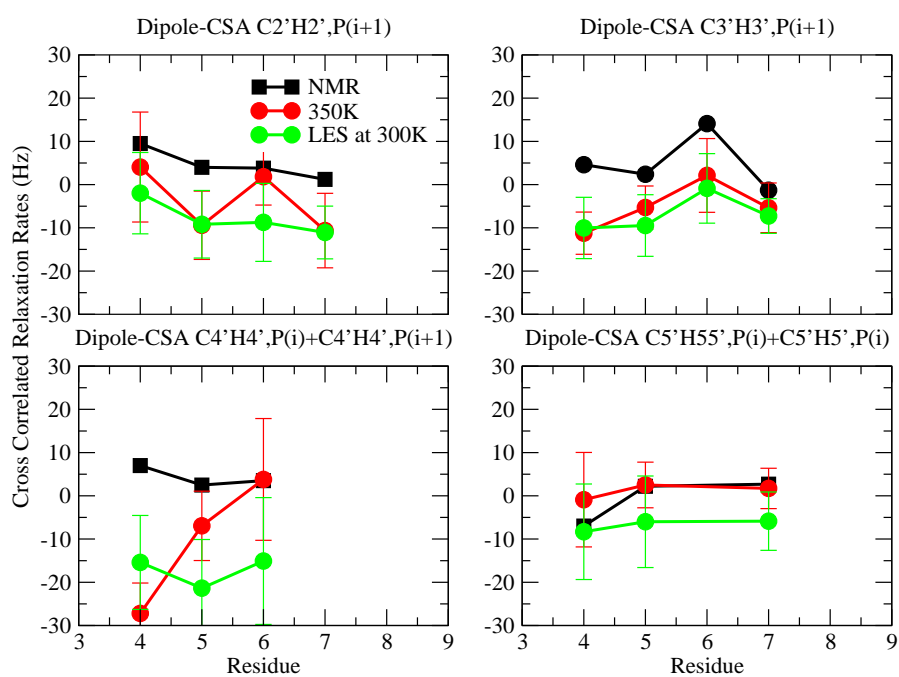


Figure 3.5: Experimental and simulated ^{31}P -CSA cross correlated relaxation rates of the UUUU loop. Comparison of the NMR in black to MD at 350 K (red) and to the LES method (green)

| Method | $C4', P_i$ | U4 | U5 | U6 | U7 |
|--------|----------------|-------|-------|-------|-------|
| MD1 | VB | 11.52 | 11.51 | 10.87 | 10.90 |
| | LK | 10.67 | 10.65 | 10.07 | 10.11 |
| MD2 | VB | 10.38 | 11.13 | 10.67 | 11.00 |
| | LK | 9.64 | 10.32 | 9.90 | 9.64 |
| 350K | VB | 11.15 | 11.37 | 8.67 | 10.30 |
| | LK | 10.33 | 10.52 | 8.07 | 9.58 |
| LES | VB | 8.59 | 9.60 | 9.97 | 10.30 |
| | LK | 8.04 | 8.94 | 9.28 | 9.57 |
| Method | $C4', P_{i+1}$ | U4 | U5 | U6 | U7 |
| MD1 | VB | 9.00 | 1.11 | 11.57 | 2.32 |
| | LK | 8.42 | 1.24 | 10.70 | 2.34 |
| MD2 | VB | 2.95 | 2.53 | 3.28 | 3.84 |
| | LK | 2.93 | 2.52 | 3.20 | 3.60 |
| 350K | VB | 10.82 | 4.57 | 7.85 | 1.67 |
| | LK | 10.04 | 4.35 | 7.32 | 1.73 |
| LES | VB | 9.00 | 4.10 | 7.71 | 2.66 |
| | LK | 8.40 | 3.96 | 7.23 | 2.66 |
| Method | $C2', P$ | U4 | U5 | U6 | U7 |
| MD1 | VB | 0.93 | 8.93 | 0.70 | 7.83 |
| | LK | 1.08 | 8.35 | 0.67 | 7.34 |
| MD2 | VB | 6.70 | 7.88 | 7.13 | 4.23 |
| | LK | 6.33 | 7.38 | 6.70 | 4.05 |
| 350K | VB | 0.67 | 5.95 | 3.07 | 8.77 |
| | LK | 0.72 | 5.62 | 2.92 | 8.19 |
| LES | VB | 1.82 | 5.82 | 2.74 | 7.05 |
| | LK | 1.81 | 5.52 | 2.68 | 6.66 |

Table 3.1: Difference of the 3J coupling constants, when the Karplus parameters of VB and LK (Van Buuren given in equation (2.21) and equation (2.20) for Lankhorst) are used.

3.1.4 Sugar conformations

Information on the state of the sugar puckers can be obtained from the couplings ${}^3J_{H1',H2'}$, ${}^3J_{H2',H3'}$, and ${}^3J_{H3',H4'}$ shown in Fig. 3.6. Except for residue U6, the couplings obtained from trajectory MD1 are seen to well reproduce the NMR data. On the other hand, trajectory MD2 again shows qualitative deviations from experiment. The fact that the NMR values for the U6 couplings lie in between the calculated values for MD1 and MD2 indicates insufficient sampling for the U6 residue. Indeed, the results for the U6 couplings of the 350 K and LES trajectories nicely match the experimental data, while the corresponding results for the U5 couplings deteriorate. The latter finding is a consequence of the fact that the 350 K trajectory also shows significant conformational heterogeneity of the U5 residue (see Fig. 3.2), which is not observed experimentally.

The sugar conformations of the UUUU loop have also been determined by dipole-dipole cross-correlated relaxation measurements [49]. Fig. 3.7 shows the relaxation rates $\Gamma_{C_1H_1,C_2H_2}^c$ and $\Gamma_{C_3H_3,C_4H_4}^c$ as obtained from NMR and various MD runs. Also shown is the ratio of the two rates, which to some extent eliminates the effects of the assumed overall correlation time τ_c and order parameters (S_{ij}^c), see Eq. (2.23). Again, trajectory MD1 is in good overall agreement with experiment, while trajectory MD2 shows qualitative deviations. Improving the sampling leads to a significantly better agreement for the 350 K trajectory, in particular for the U6 residue. Interestingly, the ratios are reproduced much better by the calculations than the individual rates. In particular, the results for trajectory MD1 are in excellent agreement experiment. As discussed in the original NMR report, the individual cross-correlated relaxation rates are influenced by dynamics, presumably sugar pucker dynamics and could not be fitted by a single overall correlation time τ_c and unity order parameters (S_{ij}^c). However, since the motions within one ribosyl moiety is correlated, formation of the ratio of $\Gamma_{C_1H_1,C_2H_2}^c/\Gamma_{C_3H_3,C_4H_4}^c$ of the cross-correlated relaxation eliminates part of this additional motion.

3.1.5 NOE distances

For the UUUU loop, a total of 79 NOEs have been measured. For all intra-nucleotide NOEs as well as for NOEs associated with base-paired nucleotides, we have found excellent agreement between the distances obtained from experiment and all MD calculations (data not shown). The remaining nine structural (i.e., inter-nucleotide but not base-paired) NOEs are listed in Tab. 6.1. The results for trajectories MD1 and for the 350 K run are again in good overall agreement with experiment and show only one prominent deviation, U5-H3';U6-H6. The agreement of NMR and trajectory MD2 is only slightly worse, with two nonmatching NOEs, C3-H3';U4-H6 and U4-H2';U5-H6. The LES simulation differs in all three mentioned NOEs from the NMR data. It is noted that in virtually all cases the calculated NOE distances are larger

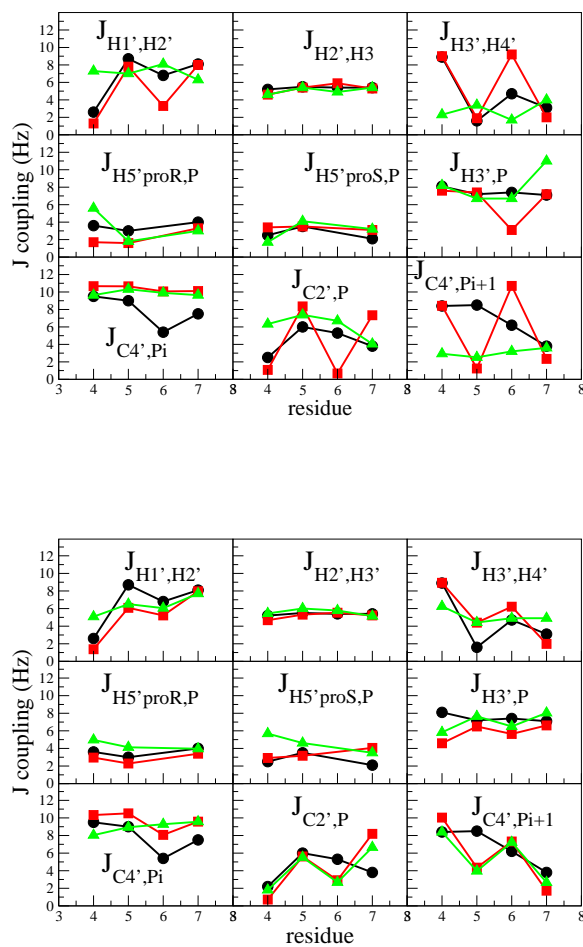


Figure 3.6: Experimental and simulated 3J couplings of the UUUU loop. Upper panel: Comparison of the two 300 K trajectories MD1 (gray squares) and MD2 (light triangles) to NMR data (black circles). Lower panel: Comparison of the 350 K trajectory (squares) and the LES trajectory (triangles) to NMR data (circles).

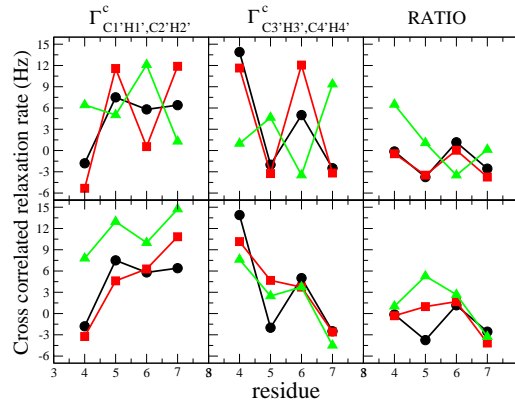


Figure 3.7: Experimental and simulated dipole-dipole cross correlated relaxation rates of the UUUU loop. Upper panel: Comparison of the two 300 K trajectories MD1 (gray squares) and MD2 (light triangles) to NMR data (black circles). Lower panel: Comparison of the 350 K trajectory (squares) and the LES trajectory (triangles) to NMR data (circles).

than the experimentally measured ones. This finding indicates some conformational heterogeneity, where conformations with larger distances are suppressed in the $\langle r^{-6} \rangle$ averaging process of the NOE measurement [99].

| NOE | MD1 | MD2 | LES | 350 K | NMR |
|--------------|---------------|----------------|---------------|---------------|---------------|
| C3-H6;U2-H2' | 4.1 ± 0.4 | 4.1 ± 0.4 | 3.8 ± 0.3 | 4.2 ± 0.5 | 2.0 ± 0.5 |
| C3-H6;U2-H3' | 4.0 ± 0.4 | 3.4 ± 0.4 | 3.5 ± 0.3 | 3.6 ± 0.5 | 4.0 ± 1.0 |
| G2-H1;G8-H1 | 3.9 ± 0.4 | 3.4 ± 0.3 | 3.7 ± 0.4 | 3.6 ± 0.4 | 4.5 ± 1.0 |
| G10-H1;G2-H1 | 4.5 ± 0.6 | 4.2 ± 0.5 | 4.4 ± 0.5 | 4.9 ± 0.9 | 4.0 ± 1.0 |
| U4-H6;C3-H3' | 3.0 ± 0.3 | 7.9 ± 0.4 | 6.7 ± 1.2 | 3.6 ± 0.7 | 3.0 ± 0.5 |
| U5-H6;U4-H2' | 3.6 ± 0.4 | 11.1 ± 1.0 | 7.0 ± 1.4 | 5.8 ± 1.0 | 3.0 ± 0.5 |
| U6-H6;U5-H3' | 7.8 ± 0.8 | 4.7 ± 2.3 | 8.4 ± 1.4 | 8.7 ± 1.3 | 4.0 ± 1.0 |
| G8-H8;U7-H1' | 4.0 ± 0.8 | 4.8 ± 1.1 | 3.9 ± 0.8 | 4.0 ± 1.5 | 3.0 ± 0.5 |
| C9-H6;G8-H3' | 3.9 ± 0.5 | 4.1 ± 0.4 | 3.3 ± 0.3 | 3.6 ± 0.5 | 2.0 ± 0.5 |

Table 3.2: Comparison of calculated and experimental NOEs (in Å) of the UUUU loop.

3.1.6 Conformational dynamics

Although trajectory MD1 is in good overall agreement with the NMR data, the above results have shown that improved sampling strategies such as LES or high-temperature simulations are necessary to catch the experimentally observed transitions of the U6 sugar ring. Moreover, Fig. 3.2 revealed that these simulations exhibit conformational heterogeneity with respect to the backbone dihedral angles α and χ . This finding indicates the possibility of further conformational transitions of the

UUUU loop. In the following, we adopt the 350 K trajectory to investigate the nature of this conformational dynamics in some detail. The reason that we prefer the high-temperature run over the LES simulation is twofold. The analysis as well as the interpretation of a standard MD trajectory are more straightforward, and the overall results are in better agreement the performed replica exchange MD simulation (REMD), which will be discussed in the following section.

To identify and characterize the various conformations adopted by the 350 K simu-

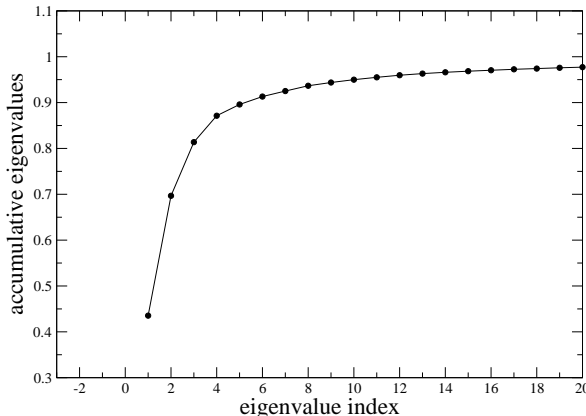


Figure 3.8: Relative fraction of the dynamics gained from the first PCA eigenvectors

lation, we have performed a PCA of all atoms of the UUUU loop (see Methods). As can be seen in Figure 3.8. The relative fluctuations are calculated from the relative cumulative sum of the eigenvalues. Already with the first two eigenvectors we can get 70% of fluctuation of the trajectory. Figure 3.9 shows the resulting free energy surface as a function of the first two PCA eigenvectors. From MD1 there is only one minimum, meaning the UUUU loop occurs only in one conformation. The LES trajectory has several minima, where most of the conformations have an unusual geometry in the loop area, this can be due to the nature of the method. However, the structure of the large minimum resembles slightly the structure of MD1 and structure 1 (see last panel) of the 350K run. The free energy landscape of the 350 K run exhibits several minima, labeled by the numbers 1-4. To clearly separate minima 3 and 4, it is necessary to invoke a third PCA eigenvector (data not shown). The minima of free energy correspond to various conformational structures displayed in Fig. 3.10. The population probabilities of the four states are 14, 20, 40 and 17 %, which correspond to free energy differences of less than 1 kcal/mol. The minima of free energy correspond to various conformational structures displayed in 3.10. The four structures found in the simulation are seen to differ mainly in the stacking interactions of their bases: Structure 1 exhibits a stacking of the bases of residues U4 and U5; it essentially reflects the average structure found in trajectory MD1.

Structures 2 and 3 represent the most populated conformations of the loop. Both structures show a stacking of the bases U5 and U7, but differ in the position of the U6 residue: In structure 2, U6 points into the solvent (as in structure 1); in structure 3, U6 is close (but not stacking) to U7. Finally, structure 4 exhibits a stacking of the bases of residues U6 and U7. All structures show a stable stacking interaction between the bases of the loop residue U4 and the stem residue C3, respectively. The initial NMR structure that was used as starting geometry for the simulation is qualitatively similar to structure 1. Interestingly, however, no NOEs revealing intensive base stacking could be observed, not even between the U4 and C3 bases. However, some experimental evidence on stacking interactions is provided by the H6 chemical shifts (U7=7.81 ppm; U5=7.71 ppm; U4=7.68 ppm; U6=7.52 ppm) [100], which show that U7 provides the most downfield proton reflecting some stacking, while U6 provides the most shielded proton, presumably due to the lack of any stacking interactions [101]. It should be noted that there is also very little stacking observed for trajectory MD2, which has two residues (U5 and U6) pointing into the solvent. However, the structure of trajectory MD2 was hardly sampled in the 350 K and the LES simulations, which may also explain the unsatisfactory agreement of this trajectory with experiment.

To reveal the underlying dynamics of this conformational heterogeneity, Fig. 3.11 shows the time evolution of (a) the projection of the MD trajectory on the first two PCA eigenvectors and (b) the stacking interactions of base pairs U4-U5, U5-U6, U5-U7, and U6-U7 as monitored by the distances between their centers of mass. Considering a distance $\lesssim 5$ Å as indicative for base stacking, it is found that the bases U4 and U5 are stacked throughout the trajectory, the bases U5 and U7 are stacked during $2 \text{ ns} \lesssim t \lesssim 15 \text{ ns}$, the bases U6 and U7 are stacked for $t \gtrsim 15 \text{ ns}$, and the bases U5 and U6 are not stacked at all. As anticipated in the discussion of the PCA energy surface and the corresponding conformational structures of the UUUU loop, there is a clear correlation between the dynamics of the principal components shown in (a) and the dynamics of the base distances in shown (b). The time evolution of these quantities can be roughly divided up in four periods, (1) 0-2.5 ns, (2) 2.5-6 ns, (3) 8-15 ns, and (4) 15-18 ns, which are found to directly correspond to the free energy minima and structures shown in Figs. 3.9 and 3.10, respectively. Moreover, the conformational rearrangement from structure 2 to structure 3 during times 6-8 ns is reflected in a shallow low-energy region between the free energy minima 2 and 3 in Fig. 3.9. Hence the time evolution of the base distances account almost completely for the main conformational dynamics described by the first few PCA eigenvectors.

Figure 3.11(c) also shows the time evolution of the pseudorotation angle P of the sugar rings U5 and U6. The data for the U6 angle nicely reproduce the experimental finding that the U6 sugar pucker is to 60 % in the C2'-endo state (corresponding to $P \approx 150^\circ$) and to 40 % in the C3'-endo state (corresponding to $P \approx 20^\circ$). As discussed above, the 350 K simulation also predicts significant conformational heterogeneity of the U5 sugar ring, which does not match the experimental result

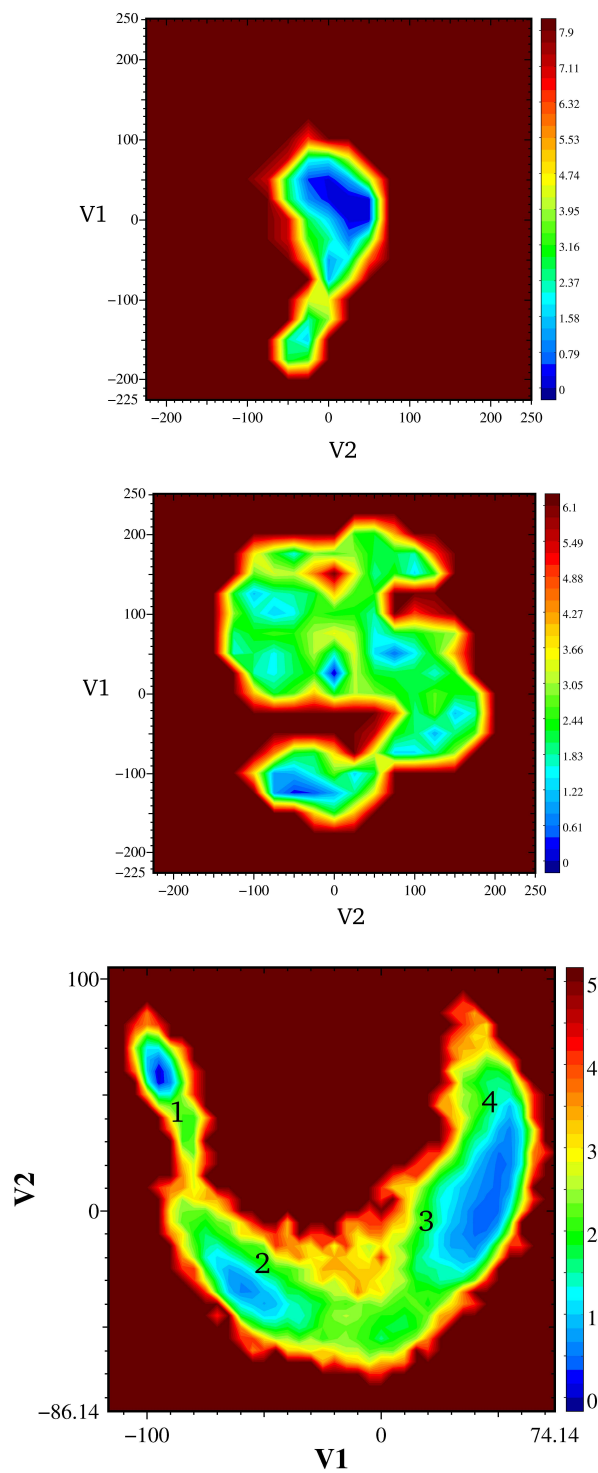


Figure 3.9: Free energy surface (in units of kcal/mol) obtained from MD1 (upper panel), LES (middle panel) and from the 350 K (lowest panel) run, plotted as a function of the first two PCA eigenvectors.

that U5 sugar pucker is to 90 % in the C2'-endo state. Moreover, the MD simulation predicts the C3'-endo state for the U4 sugar and the C2'-endo state for the U7 sugar, which again is in good agreement with experiment.

Most interestingly, however, Fig. 3.11 clearly shows that there is no correlation between the dynamics of the sugar rings and the stacking dynamics of the loop bases. We have also considered the time evolution of the torsional angles χ , which describe the relative orientation of the sugar ring and the base. Throughout the simulation, we could only find very little correlation of the χ angles with the dynamics of the sugar rings or the stacking of the bases (data not shown). This finding is somewhat surprising because of the well-known fact that the A- and B-forms of DNA and RNA are clearly correlated with the conformation of the sugar rings, i.e., C3'-endo state for A-forms and C'-endo state for B-forms.

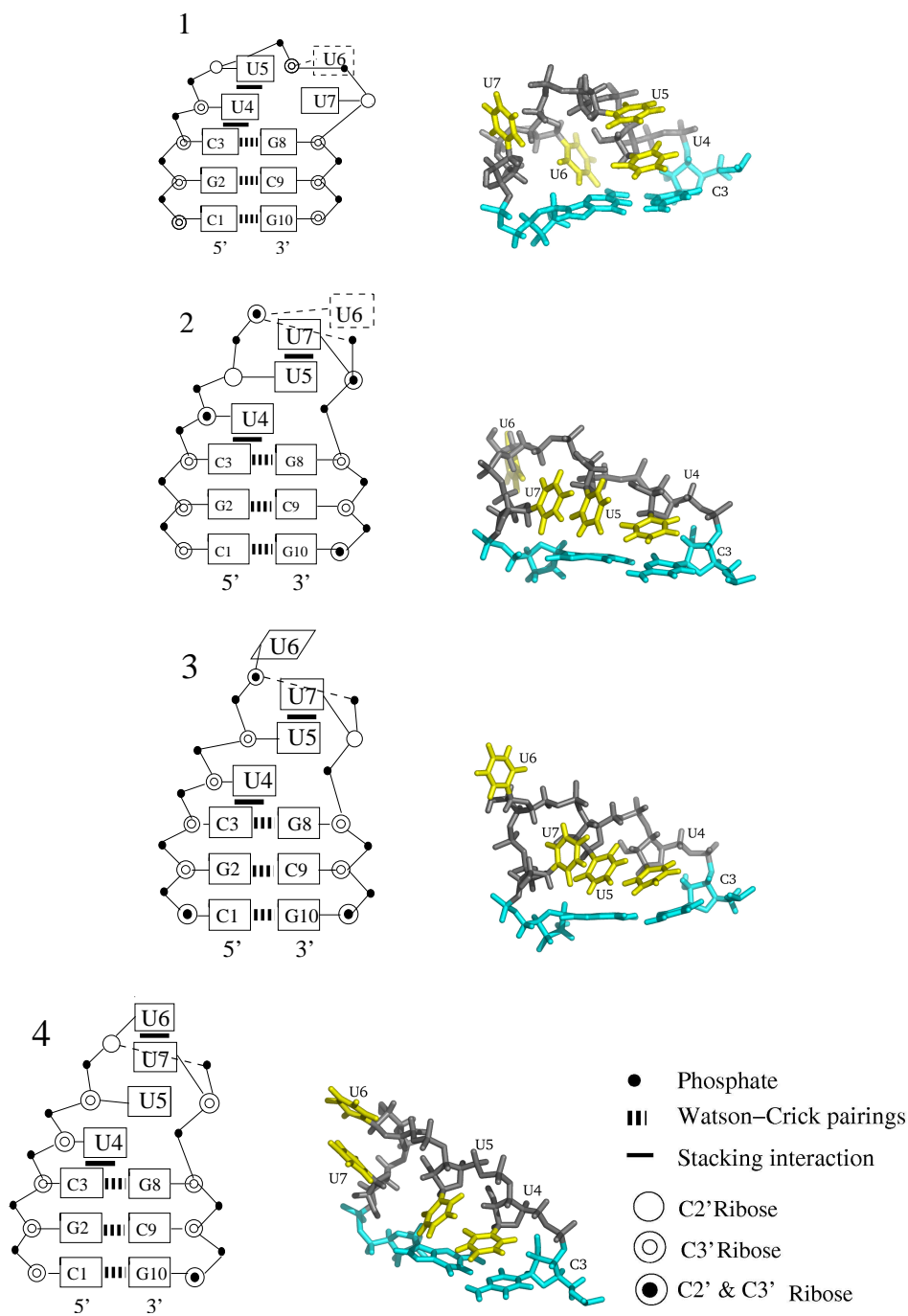


Figure 3.10: Illustration of the secondary structures of the UUUU loop as found in NMR experiment and in the 350 K MD simulation. Dashed lines indicate that the residue points into the solvent.

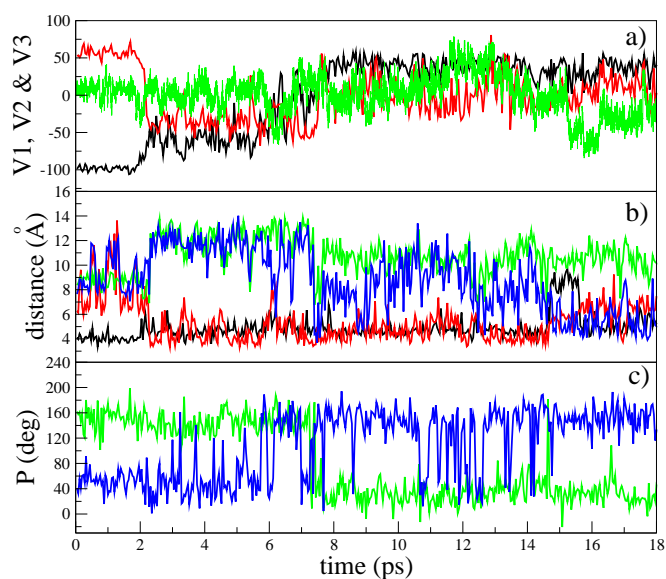


Figure 3.11: Conformational dynamics of the UUUU loop as observed in the MD simulation at 350 K. (a) Projection of the trajectory on its first (black), second (red), and third (green) PCA eigenvector, respectively. (b) Distances between the loop bases U4 and U5 (black), U5 and U7 (red), U5 and U6 (green), and between U6 and U7 (blue). (c) Pseudorotation angles of residues U5 (blue) and U6 (green).

3.2 REMD simulation of the UUUU loop

3.2.1 Computational details

REMD approach is applied using the algorithm described in chapter 2. To achieve the optimal performance of the algorithm for a given system, the temperatures of the replicas are chosen such (i) that the lowest temperature is small enough to sufficiently sample the states of low energy, (ii) that the highest temperature is large enough to overcome energy barriers of the system and (iii) that the acceptance ratio of criterion (2.12) is sufficiently high. In the case of the tetraloops we employed the procedure suggested in Ref. [70] and run several short simulations at temperatures $T = 280, 320, 370, 400,$ and 460 K. From these simulations, we obtained the total potential energy E as a function of temperature T . Requesting an acceptance ratio of 20-30 %, the temperature distribution for the replicas was then obtained by solving Eq. (2.13) iteratively. This way, we obtained a total of 40 temperatures from 295 K to 453 K, with a gap of 4-6 K between adjacent replicas.

As starting structure for the UUUU loop we obtain from the 350 K conventional MD run at 10 ns and constant volume for all 40 replicas, each replica was run independently at its own temperature for 200 ps. Then, the exchange procedure between the replicas was turned on, using a time step of 1.5 ps between two attempts of exchange. This time interval should be large enough compared to the coupling time of the heat bath (0.1 ps), such that the trajectory is roughly equilibrated before the next exchange is attempted. Each replica was run about 8 ns, yielding a total sampling time of 320 ns for the UUUU and the UUCG loop. The data were collected every 0.1 ps.

3.2.2 Dynamics of the UUUU loop

We have seen in the previous section that conventional MD was able to reproduce experimental data for certain observables. The pseudorotation of the U6 ribose however was not in accordance with NMR, this could be seen in the discrepancy of certain ^3J couplings. The application of the REMD to the hairpin should overcome one limitation of conventional MD, which is the insufficient sampling of the important configurational space. As we have seen previously, the hairpin from the MD1 trajectory could come up with only one conformation. Enhancing the sampling by increasing the temperature and applying LES lead to more conformations but certain observables were still not in good agreement with the NMR derived ones. The results that are presented here are from the 300 K trajectory of the REMD simulation. One of the most widely used measures to judge the performance of a simulation is the RMSD, as we have already seen in the previous chapter. A plot of all atoms RMSD versus time is shown in Fig. 3.12. The simulation shows very little movement from the initial structure for the first 2.5 ns with the RMSD values

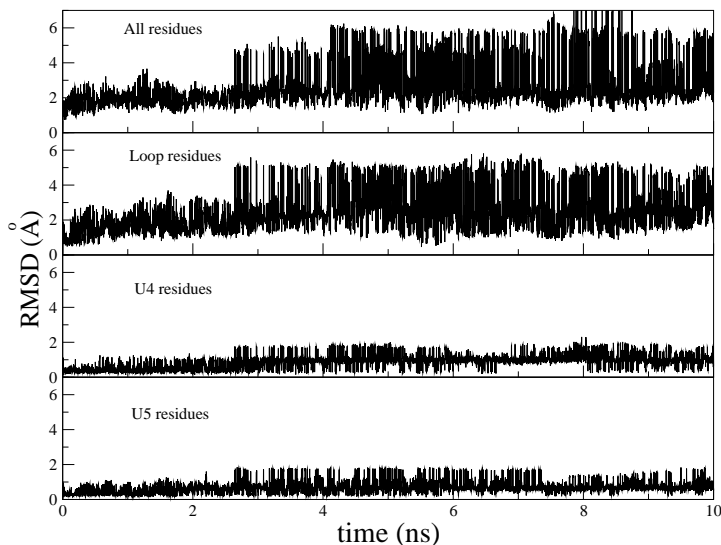


Figure 3.12: Time trace of the all atoms RMSD obtained from the REMD simulation at 300 K. These values are the deviations from the average structure.

ranging between 1 and 2.5 Å. For longer time we can see that the RMSD values increase to almost 8 Å. The big change of the RMSD at around 2.6 ns is caused by a transition in the loop region. (Second panel: Fig. 3.12).

Further decomposition of the RMSD into the individual residues shows a correlation between the RMSD of the whole system with the RMSD of the U4 and U5 residues. Therefore, the next step is to identify for these residues the observables, that lead to such a significant change in the RMSD. The first attempt is to check the dihedral angles for these residues. Indeed, a correlation of the backbone angles could be observed. In the case of U4 (Fig. 3.13) four dihedral angles χ , δ , ϵ and ζ cause the observed transition of the RMSD. The U5 residue on the other hand show a similar time trace for the χ angle only (data not shown). Hence, residue U4 is predominantly responsible for the changes observed in the total RMSD. This finding is somewhat surprising as one would expect that the dihedral of residue U6 to be responsible for the large changes in the RMSD, as it has been shown previously to be the most flexible residue in this loop. Still, a correlation between the total RMSD and the U6 residue can be observed.

In conventional MD it has been shown that the base to base distances can be used to describe the conformations of this system. The base to base distance between U5 and U6 (Fig. 3.14) also correlates with the time trace of the total RMSD. The time trace of the distance U4-U5 does not have a transition at around 2.5 ns, which means that the transition of the distance U5-U6 can only be caused by U6, which slowly drifts towards U5. In this case a change of the distance from around 12 Å to almost 5 Å is visible. Stacking between U4 and U5 was a predominant interaction in MD1 and the 350K simulations, and can be observed in the REMD

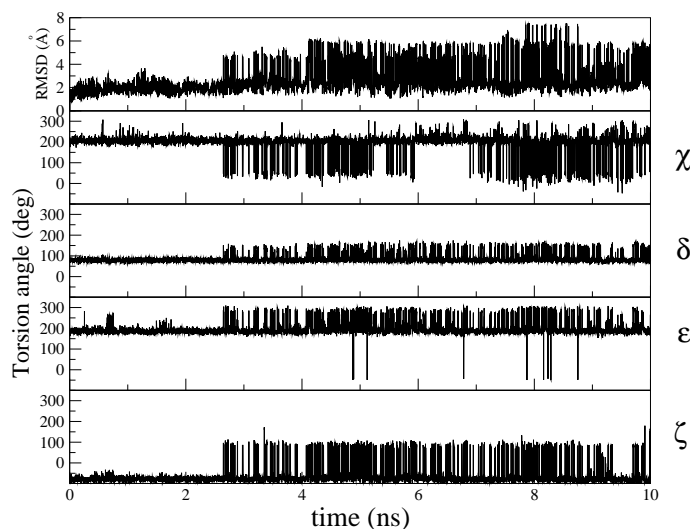


Figure 3.13: Time trace of the RMSD of the whole UUUU system (in Å) and backbone torsion angles of the U4 residue obtained from the REMD simulation at 300 K.

simulation, too (Fig. 3.14). However, reduced stacking interactions between these residues can be observed around 4.5-8 ns, enabling the U5 to move freely, whereas the U4 residue remains still very rigid, due its prevailing stacking interaction with the base of residue C3 (data not shown). The distance between U6 and U7 also states the high flexibility of the U6 residue, whereas the distance between U5 and U7 does not show a significant change in their values throughout the trajectory.

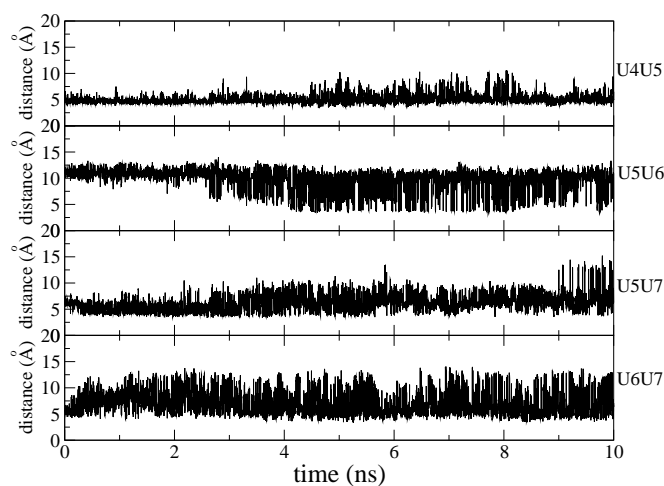


Figure 3.14: Time evolution of the distances in \AA obtained from the 300 K REMD simulation. The distances are calculated between the center of masses of the bases in the loop region. U4U5 is the distance between the base of residue U4 and the base of U5. The other distances are defined in the same way.

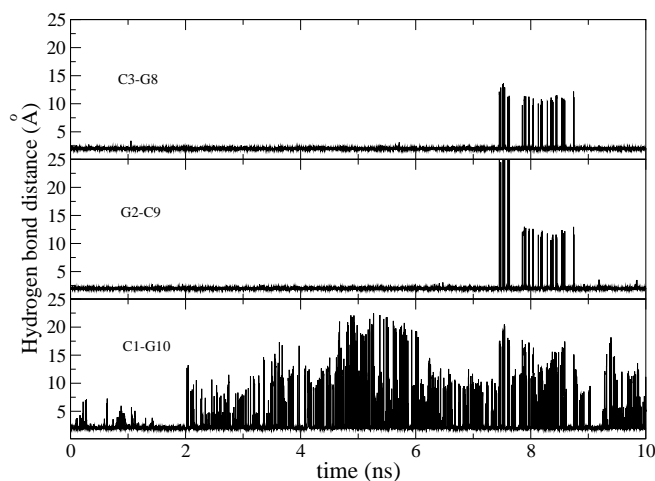


Figure 3.15: Time course of the hydrogen bond distances (\AA) in the stem of the UUUU loop. The distances are calculated between the heavy atoms N1-H1..N3 of the base pairs.

Even though the stem area does not contribute to the changes in the RMSD, the stem area exhibit especially for the terminal base pair some significant dynamics. The closing and the second base pair maintain most of the time a planar geometry in the simulation (Fig. 3.15), A disruption of the hydrogen bonds can be observed with a higher frequency throughout the entire simulation for the terminal base-pair.

3.2.3 Comparison to NMR

Conventional MD simulations at 300 K show that residue U4 and the U6 are in the C3' endo conformations of the ribose, however only U4 is in agreement with NMR. The ribose of residue U6 exists in NMR in C2' and C3' endo. Residue U5 and U7 showed ribose C2' endo conformations in accordance with NMR. Enhancing the temperature to 350 K resulted in two ribose conformations in agreement to NMR. This means that the MD1 trajectory was obviously under-sampled. However, for residues U5 and U7 a transition from C2' endo to C3' endo was observed at this temperature. Therefore to check whether these conformations do also exist using enhanced sampling method, REMD is implemented. And indeed we monitor a transition between the two ribose conformations for all residues, whereas U4 and U5 riboses seem to prefer the C3' endo conformation. U6 and U7 are observed to prefer both ribose conformations equally (Fig. 3.16).

Base on these observations the averaged J coupling constants ${}^3J_{H1',H2'}$ and ${}^3J_{H3',H4'}$ in the following plot (Fig. 3.17) are in better agreement for residues U4, U6 and U7. U5 on the other hand has a large deviation from NMR as we have seen due to its strong preference in our simulation for the C3' endo conformation (Fig. 3.16). The couplings ${}^3J_{H5',proS/R,P}$ and ${}^3J_{C4',Pi}$ reflect the angle β . The agreement of the experimental and calculated values for β can be seen in Table 3.3, the calculated couplings ${}^3J_{H5',proS/R,P}$ match well the NMR data. As we have already seen from conventional MD the results obtained for ${}^3J_{C4',Pi}$ and ${}^3J_{C2',P}$ which correspond to β and ϵ differ significantly despite enhanced sampling and overall good agreement with NMR for these angles (Table 3.3). This can be as already mentioned due to the Karplus relation (2.20) which may not be appropriate in this case.

Fig. 3.18 shows the joint probability distribution of the angles α and β and ϵ and ζ as obtained for the REMD simulation at 300 K and MD at 350 K. We have seen from conventional MD at 300 K (Fig. 3.4) that all four backbone angles are well described by a single mean value, the enhanced sampling calculation at 350 K revealed that several conformations may be accessible.

Also shown are the values (as red dots) of the backbone angles α , β , ϵ and ζ that are consistent with the analysis of the experimental data, as well as the values (as asterisks) of the canonical A and B forms, respectively. Taking all possible experimental and calculated values shown into consideration, we found for the 350 K a good overall agreement of NMR and simulation.

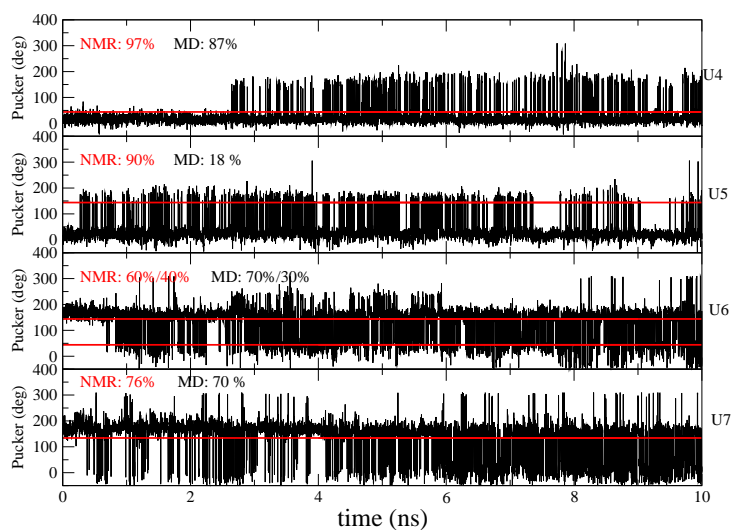


Figure 3.16: Pseudorotational angles for the loop region of the UUUU hairpin gained from REMD at 300 K (black). The probability (%) as obtained from NMR is shown for U4 in the C3'-endo, for U5, U7 in C2'-endo and U6 in C3'-endo and C2'endo state.

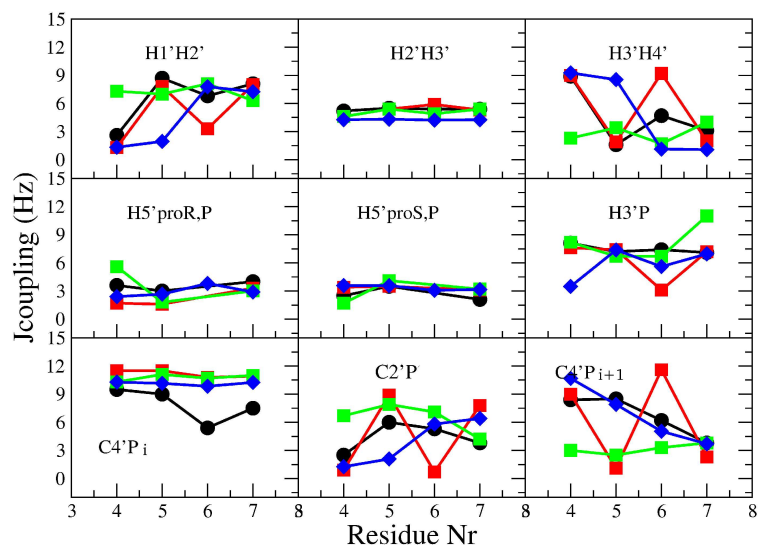


Figure 3.17: 3J couplings for the UUUU loop gained from conventional MD in red (MD1) and green (MD2). 3J couplings from REMD (blue) at 300 K in blue compared to NMR in black.

The 10 ns REMD simulation (Fig. 3.18) depicts in most cases a similarity to the 350 K trajectory. Therefore, the agreement of the REMD and 350 K trajectory with NMR data is in most cases the same. For ϵ and ζ angles of the U6 residue, the REMD seems to be slightly better than the 350 K trajectory. This can be seen for $\epsilon = -125^\circ$ and $\zeta = -70^\circ$ angle (Table 3.3). Considering U7 as a representative example, we find that β is well reproduced for all nucleotides, since it is well described by a single value in calculation and experiment. This is in contrast to the findings, for which the enhanced sampling simulations suggest that several -typically two- regions of α are thermally populated. As a consequence, two experimental values, $\alpha = -120^\circ$ and 75° are consistent with the enhanced sampling data.

| Torsion angle | Method | U4 | U5 | U6 | U7 |
|---------------|--------|------------------|-------------------|-----------------|-----------------------------|
| α | REMD | -143 | -148 | -116 | -153 |
| | MD1 | -73 | -70 | -71 | -87 |
| | 350 K | -133 | -132 | -154 | -109 |
| | NMR | -75 -150, 0, 100 | -120, 52 | | -120, 29, 75 |
| β | REMD | -179 | -179 | 171 | -181 |
| | MD1 | -187 | -188 | 173 | -175 |
| | 350 K | -180 | -177 | 154 | -173 |
| | NMR | -174 | -184 | 180 | -165 |
| γ | REMD | 130 | 120 | 167 | 116 |
| | MD1 | 63 | 72 | 183 | 58 |
| | 350 K | 117 | 122 | 120 | 91 |
| | NMR | 57 | 68 | | 53 |
| δ | REMD | 85 | 90 | 125 | 128 |
| | MD1 | 80 | 143 | 75 | 138 |
| | 350 K | 87 | 117 | 102 | 136 |
| | NMR | 67 | 109 | 112 | 118 |
| ϵ | REMD | -230 | -240 | -193 | -245 |
| | MD1 | -166 | -268 | -176 | -260 |
| | 350 K | -150 | -115 | -153 | -193 |
| | NMR | -235,-146 | -231,-145 | -235,-125 | -237,-120 |
| ζ | REMD | -77 | -99 | -85 | -221 |
| | MD1 | -54 | -268, | 63 | 49 |
| | 350 K | -79 | -184 | 125 | 92 |
| | NMR | -245,-100,80 | -125, -170 -50 | -105, -70 80 | -185, -255 25, -45 55 |

Table 3.3: Comparison of calculated and experimental torsional angles of the UUUU tetraloop (in degrees). REMD results are obtained at 300 K and compared to conventional MD at 300 K for MD1, to the trajectory of 350 K and NMR.

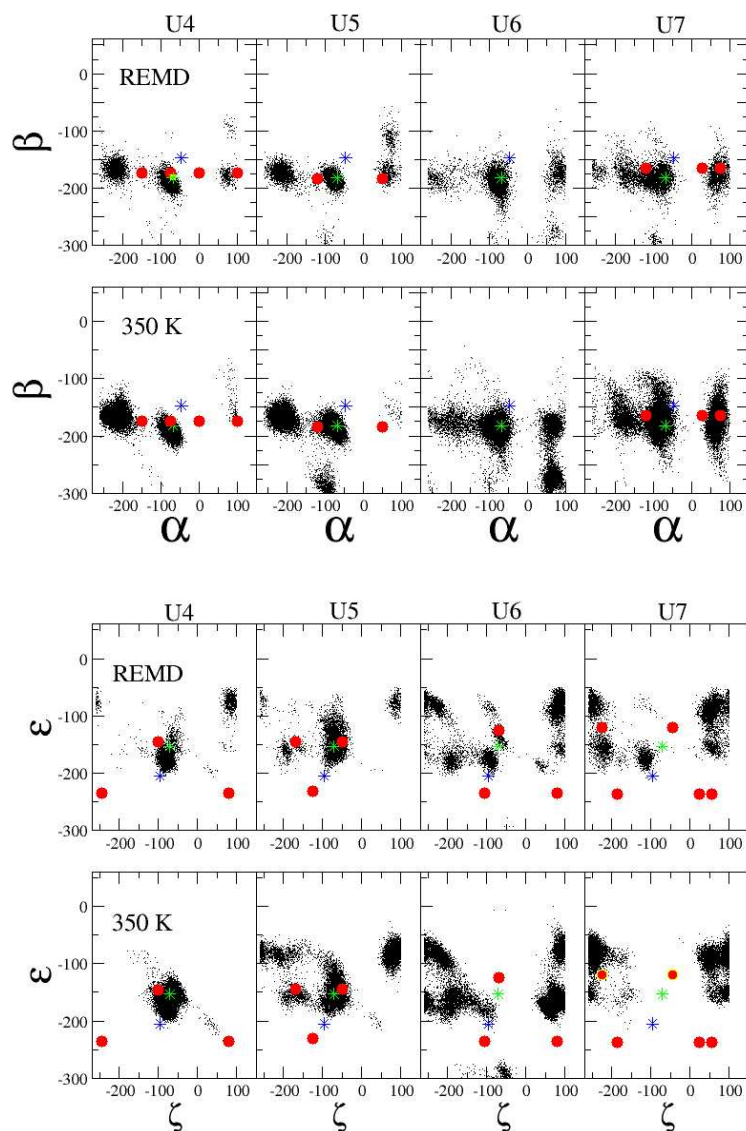


Figure 3.18: Joint probability distribution of the α and β (upper panels) and ϵ and ζ (lower panels) as obtained for trajectory REMD and for the conventional MD 350 K run. The red dots correspond to values of the angles that are consistent with the analysis of the experimental data in Fig 3.3. The green and blue asterisks display the values of the canonical A and B form, respectively.

3.2.4 Conformational analysis

In the previous section, we used the conventional MD at 350 K to characterize the conformations of the loop region of UUUU tetraloop. Four conformations were obtained. These structures were stabilized only by stacking interactions. The strongest stacking is between the U4 and C3 residue, followed by U5-U7. Another one which was observed between residue U4 and U5 occurred with a population probability of around 14 %. In this section, PCA (Fig. 3.19) of the REMD trajectory at 300 K resulted in four main conformations (Fig. 3.20). The conformations are achieved in the following. First, the higher principal components are plotted by their distributions. Then all Gaussian shaped distributions, in this case components higher than number 4, are omitted in the analysis of the conformations. The first four components are therefore used for the characterization of the conformations. Defining for each of the four components ambiguously sub-states 1 and 2 and a combination of all possible sub-states result in 16 states. However, the population probabilities of many of these states are low and we restrict ourselves to the states with higher populations, as shown (Fig 3.20).

A comparison of the angles from REMD with conventional MD, indicated that the REMD samples more angles, meaning the conventional MD at 350 K is still under-sampled. The four conformations we detect here (Fig 3.20) are similar for A, B and D to the conformations from 350 K. They distinguish from each other primarily in their stacking behavior of residue U5 and U7 and the position of the U6 base. Additional conformations that are gained from REMD that 350 K MD lacked, correspond to conformation C, where residue U4 points into the solvent. Characteristics of conformation A which has the highest probability is the strong stacking interaction between U3 and U4 and a weaker interaction between nucleobase U5 and U7. Here the U6 base is still pointing into the solvent. This conformation is also observed at the 350 K trajectory to occur with the highest probability. Conformation B is stabilized by U6-U7, U4-U5 stacking and both of them could also be found in conventional MD. Conformation D occurs with only 7% and a weak stacking between U5 and U7 can also be observed. Interestingly, the conformation that hosts U4-U5 stacking, as was seen in conventional MD does not appear at least in the first 4 four states obtained from REMD. From this conformational analysis, we can conclude that the UUUU loop is stabilized in the loop area mainly by stacking between C3 and U4 and by U5 and U7. At the same time the flexible U6 base prefers to be in close vicinity of U7.

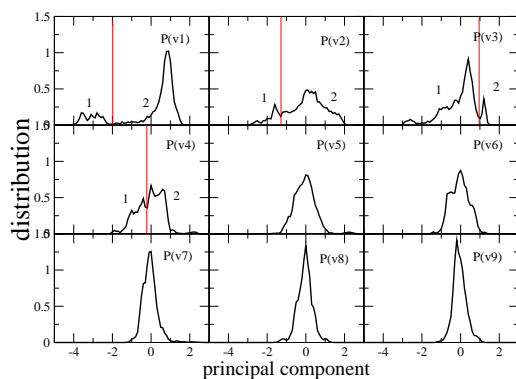


Figure 3.19: Distribution of the first nine principal components gained from an REMD simulation at 300 K. A determination of the sub states (labeled with 1 and 2) is done using the first four principal components.

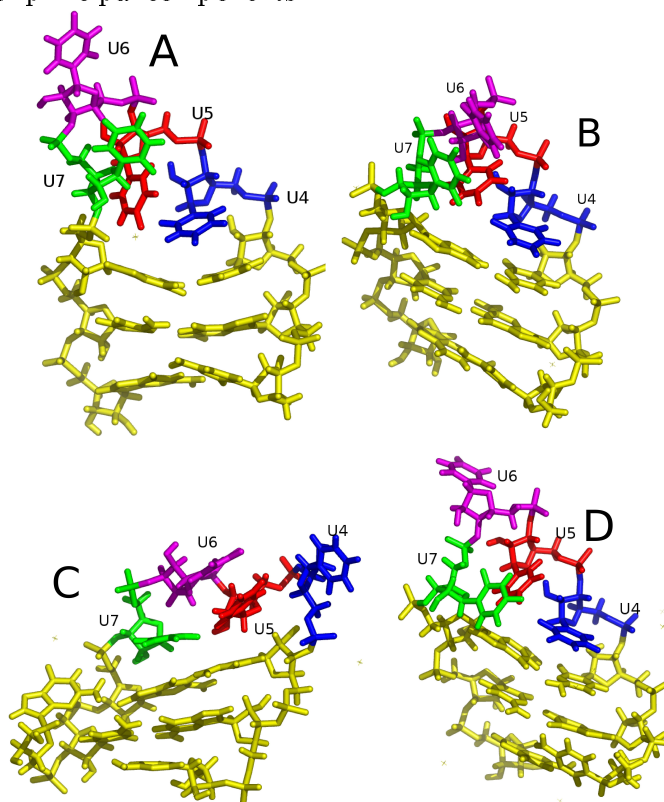


Figure 3.20: Illustration of the secondary structures of the UUUU loop as found in REMD simulation at 300 K. Conformations A, B, C and D occur with the probabilities 49%, 23%, 8% and 7%, respectively. These conformations are gained by combining all sub states gained from the first four principal components as shown in Fig. 3.20.

3.3 Conclusions

We have presented a detailed and comprehensive MD study of the structure and the dynamics of the RNA UUUU loop, by performing MD runs at 300K and at 350K, as well as enhanced sampling using LES and REMD simulation. As a first goal of this work, we have studied the performance of the theoretical model, that is, the quality of the force field and the conformational sampling achieved. As a second goal, we have employed both experimental and theoretical data to draw a dynamical picture of the UUUU loop.

Concerning the sampling achieved in the MD simulations, it has been shown that a conventional room-temperature MD run on a 10 ns time scale cannot account for the conformational heterogeneity of a flexible RNA tetraloop. This is clearly demonstrated by the calculations of the ^3J couplings of the U6 residues in Fig. 3.6, where the results of 300 K runs MD1 and MD2 are either too high or too low compared to experiment, while all three enhanced-sampling simulations yield average values close to experiment. Moreover, the time traces of the 350 K and the LES simulation indicate that even these enhanced-sampling methods do not yield converged data for the UUUU loop on a 10 ns time scale. Therefore the REMD simulation was inevitable. Nonetheless, the overall results of the 350 K simulation are in good agreement with the 300 K data of the REMD simulation, and therefore appear to draw a realistic picture of the conformational heterogeneity of the RNA tetraloop. Concerning the quality of the force field, we have found that the 300 K trajectory MD1 is in surprisingly good overall agreement with experiment, despite of its insufficient sampling. The simulation yields only a single nonmatching NOE and with the exception of the under sampled U6 residue ^3J couplings and dipole-dipole cross correlated relaxation rates that well reproduce the NMR data in most cases. The overall good agreement is also reflected in the fact that trajectory MD1 is able to perfectly recover the first four backbone dihedral angles α , β , γ , and δ , while there are some deviations for the angles ϵ and ζ . Enhancing the sampling in the 350 K, LES and REMD simulations, the results for the U6 residue improve as expected, while the corresponding results for the U5 residue deteriorate, since the U5 sugar ring is too flexible compared to experiment and the converged REMD trajectory seems to favor the C3' endo than the observed C2' endo in NMR. While the REMD could improve the agreement to NMR for ϵ and ζ , the γ angle however exists in the simulations in two conformations, leading to a high deviation of its angle value to NMR. Nonetheless, considering the unusual amount of experimental results that are available for the UUUU tetraloop, it seems fair to say that the AMBER98 force field does an surprisingly good job in reproducing the NMR data.

To characterize the structure and the conformational heterogeneity of the UUUU tetraloop, we have performed a principal component analysis of the MD1, LES, 350 K and REMD trajectory. The MD1 trajectory has a free energy surface with one main minimum, which corresponds to stacking between the base U3 and U4 and another between U4 and U5. The 350 K trajectory also exhibits this conformation

which is similar to the NMR proposed conformation. From this trajectory due to its enhanced sampling three more conformations were obtained. These conformations also comprise the U3 and U4 stacking, whereas the flexibility of U5 and U6 cause certain stacking interactions with U7. PCA of the REMD simulation also supports the finding of the conventional MD at 350 K, where U5 and U7 as well as U6 and U7 base stacking seem favorable. REMD simulation additionally exhibits conformation C in which U4 points into the solvent and is therefore no longer stabilized by base stacking with C3. On the other hand stacking between U4 and U5 could be rarely detected in the REMD trajectory in comparison to its high occurrence in conventional MD. This can presumably be due to the NMR initial structure that was used in conventional MD, that caused the structure in MD1 to be trapped in the energy minimum. Enhancing simulations are therefore in no doubt very important for the proper handling of sampling.

Chapter 4

Molecular dynamics of the UUCG tetraloop

The UUCG loop is known to have an exceptionally high stability ($T_m \sim 70^\circ$) [40]. This tetraloop constitutes a highly abundant structural motif and has been structurally characterized both by NMR spectroscopy [39, 3] and X-ray crystallography [1, 102]. Theoretical studies have also been performed on this tetraloop [36, 103, 34, 35, 104]. However, a detailed description of the dynamics of this motif has not been completely exposed. In the Schwalbe group NMR work on the 14 mer 5'-GGCACUUCGGUGCC-3' loop [33, 52] has been performed. Various NMR structural parameters [33] have been obtained and NMR studies on the dynamics of the sugar and base moieties have been done [52]. However, data on the backbone dynamics are still missing. In this work various conventional MD simulations on the 14 mer and on the 10 mer 5'-CGCUUCGGCG-3' have been performed in order to support the NMR dynamical assignment and at the same time these NMR data were used as benchmark to validate the theoretical model. Additionally we performed REMD simulation on the 10 mer UUCG loop to enhance sampling.

The structural features of the loop region and the loop closing base pair are depicted schematically in Figure 4.1. The residues U2 and G5 form a base pair, this base pair is stabilized by two interbase hydrogen bonds (**11** and **12** in Figure 4.1) and one hydrogen bond (**10** in Figure 4.1), which is between the base of G5 and the ribose of U2. The loop is formed by U3 and C4, which both exhibit the C2'-endo ribose repuckering, extending the backbone by about 2 Å [105], thus bridging the two sides of the stem. U3-C4 interaction is stabilized by a hydrogen bond (**15** in Figure 4.1). The conformation of residue C5 is also stabilized by an intraresidual hydrogen bond (**14** in Figure 4.1) and a stacking interaction with U2. The stem region of both oligonucleotides have proper base pairing.

MD simulations of the 14 mer at 300 K and 317 K are performed comparable to NMR experiment at 298 K and 317 K [52]. The MD results are compared with the NMR 3J coupling constants and the torsional angles of the hairpin.

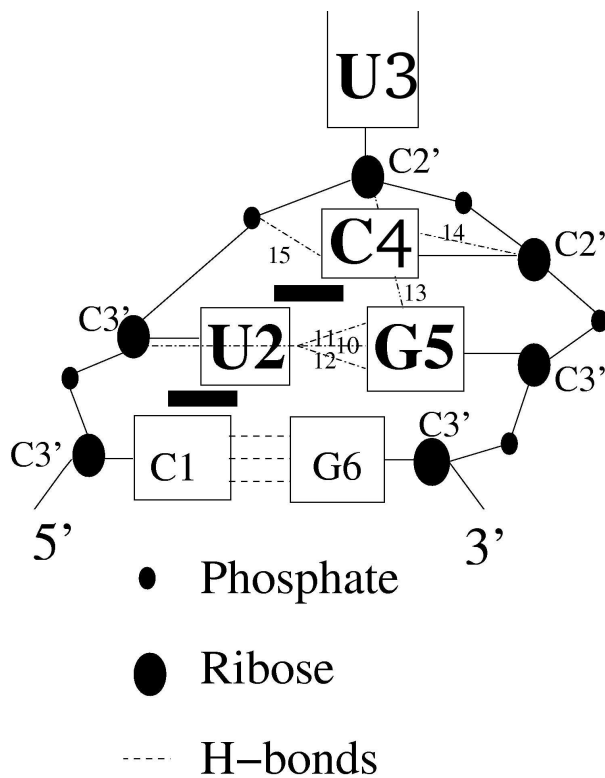


Figure 4.1: Stabilizing interactions derived from X-ray structure of the UUCG motif [1]. The hydrogen bond interactions contributing to the structure of the tetraloops are seen: (10) U6(O2')-G9(O6), (11) G9(N2)-U6(O2), (12) G9(N1)-U6(O2), (13) U7(O2')-G9(O6), (14) C8(O2')-C8(O2) and (15) C8(N4)-U7(PO^{ProR}). Stacking of the bases are seen between residue U2 and C1 (U2-C1) and between U2 and C4 (U2-C4).

Based on these agreements, MD simulation is used to clarify the dynamics in the backbone area. A high temperature trajectory at 400 K is used to confirm the convergence of the trajectories at 300 K and 317 K and to therefore support the findings of the trajectories at lower temperatures. Moreover, the 10 mer UUCG loop which differs from the 14 mer in its length and stem sequence is studied to check whether a different stem exhibits a different dynamical behavior of the loop moieties.

4.1 Computational details

All simulations were performed using the AMBER6 simulation program package [94] employing the all-atom force field parm98 [11]. The 14 mer and 10 mer UUCG loops consist of 447 and 319 atoms, respectively. These atoms were placed in a periodic truncated octahedral box of TIP3P water [95]. Including 13 Na^+ ions for the 14 mer and 9 Na^+ ions for the 10 mer to neutralize the system, a total of 6820 atoms for the 14 mer (6416 atoms for the 10 mer) is obtained. The equation of motion was integrated by using a leapfrog algorithm with a time step of 2 fs. The SHAKE algorithm [96] was used to constrain the length of the covalent bonds to hydrogen atoms with a relative geometric tolerance of 10^{-4} . A cutoff of 10 Å was chosen for the non-bonded van der Waals interactions. We used the particle-mesh Ewald method [97] to treat the long-range electrostatic interactions. The non-bonded interaction pair-list was up-dated every 10 fs. The Berendsen thermostat and barostat were used to keep the system at constant temperature and pressure. The solute and solvent were coupled separately to external temperature baths at 300 K [65], using a coupling constant of 0.2 ps. The system was also coupled to an external pressure bath at 1 atm using a coupling constant of 0.5 ps. For the initial structure first a canonical A-form RNA of the 14 mer was created using the *nucgen* command of the Amber software, then the loop region gained from x-ray [1] is superimposed onto the A-RNA stem. The equilibration protocol consisted of 500 steps of steepest-descent minimization applied to the solvent molecules with fixed solute, followed by 200 ps of MD simulation of the solvent with fixed UUCG loop, and another 500 ps simulation without position constraining of UUCG loop. After equilibration had been performed, 5 ns MD simulation followed. The structure at 5 ns was then taken as the initial structure for the MD simulations. To achieve the optimal performance of the REMD algorithm for a given system, the temperatures of the replicas are chosen such (i) that the lowest temperature is small enough to sufficiently sample the states of low energy, (ii) that the highest temperature is large enough to overcome energy barriers of the system and (iii) that the acceptance ratio of criterion (2.12) is sufficiently high. In the case of the tetraloops we employed the procedure suggested in Ref. [70] and run several short simulations at temperatures $T = 280, 320, 370, 400, \text{ and } 460$ K. From these simulations, we obtained the total potential energy E as a function of temperature

T . Requesting an acceptance ratio of 20-30 %, the temperature distribution for the replicas was then obtained by solving Eq. (2.13) iteratively. This way, we obtained a total of 40 temperatures from 295 K to 453 K, with a gap of 4-6 K between adjacent replicas.

In total we performed 20 ns for the 14 mer at 300 K, 317 K and 400 K. We have achieved around 20 ns from conventional MD simulations and 10 ns from REMD simulations for the 10 mer loop.

4.2 The 14 mer UUCG loop

The MD simulations of the 14 mer (5'-GGCACUUCGGUGCC-3') indicate proper base pairing for all stem residues and several hydrogen bonds and stacking interactions could be observed in accordance to experiments regarding the UUCG family (see detailed structural description above). As an example a snapshot after 5 ns simulation at 300 K of the 14 mer is reported in Figure 4.2 together with the secondary structure. As the main aim of this chapter is to predict the dynamics of the backbone of the UUCG loop, a comparison of certain structural parameters between NMR and MD should occur first, in order to secure that the applied theoretical model is valid. Torsional angles (Tab. 4.1), 3J coupling constants and generalized order parameters S^2 are observables that are compared with NMR for the 14 mer UUCG. The gained values from MD are averages of the whole trajectory.

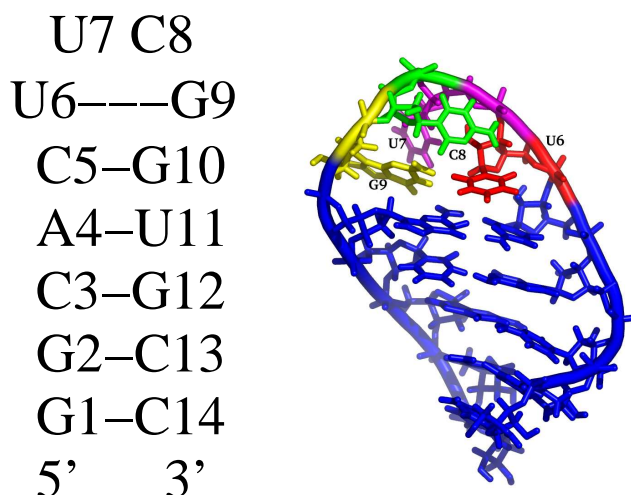


Figure 4.2: Secondary structure (left) and a snapshot of the 14 nt RNA investigated with conventional MD (right).

4.2.1 Dihedral angles

The dihedral angles (see description in chapter 2) of the backbone, the ribose and the base are in overall good agreement with NMR values for the 14 mer [33], except for β and the ζ angles exhibit for two residues a deviation of the MD values from the corresponding NMR value (Fig. 4.3). While the β angles show a large deviation for G10 (slight deviation for U7) from the NMR derived angles for the 14 mer, we notice (Table 4.1) that the same β angles are in agreement with NMR data of the 12 mer of Varani [3] for these residues, as can be seen for $\beta = 112^\circ$ and 92° for U7 and G10, respectively. On the other hand the ζ angles from the MD simulations of residues C8 and G9 do not agree either with the work of Varani or with the data from the Schwalbe group [2]. Work by Miller and Kollman [36] on a 12 mer UUCG using AMBER 4.1 force field loop indicated an angle of 57° for G9 which is similar to our result and vacuum simulation by Williams and Hall [103] gained an angle of -83° for C8 (-68° from our MD simulation).

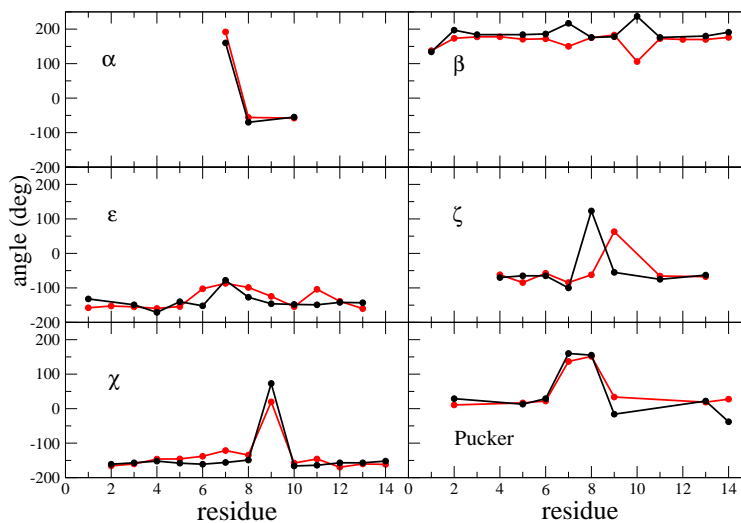


Figure 4.3: Comparison between MD (red) at 300 K and NMR [2] (black) at 298 K of the backbone torsional angles of the 14 mer. Only the dihedrals observed experimentally are reported.

| Torsion angle | Method | C3(C5) | U4(U6) | U5(U7) | C6(C8) | G7(G9) | G8(G10) |
|---------------|---------|--------|--------|--------|--------|------------|------------|
| α | REMD | -86. | -83. | -156. | -83. | 76. | -153. |
| | MD | -73. | -73. | -120. | -63. | 73. | -78. |
| | NMR_s | | | -200. | -70. | | -55. |
| | NMR_v | -71. | -80. | -142. | -55. | -58./34. | -118./112. |
| β | REMD | 174. | 171. | 167. | 181. | 182. | 165. |
| | MD | 171. | 172. | 150. | 175. | 185. | 106. |
| | NMR_s | 184. | 186. | 217. | 176. | 178. | 238. |
| | NMR_v | 163. | 181. | 112. | 174. | 136./-145. | 236./92. |
| γ | REMD | 75. | 79. | 71. | 65. | 185. | 143. |
| | MD | 61. | 63. | 56. | 59. | -66. | 26. |
| | NMR_s | | | | | | |
| | NMR_v | 63. | 49. | 63. | 45. | 189./-24 | 192./99. |
| δ | REMD | 83 | 88. | 143. | 141. | 88. | 83. |
| | MD | 76. | 76. | 135. | 138. | 81. | 78. |
| | NMR_s | | | | | | |
| | NMR_v | 92. | 92. | 133. | 136. | 91. | 96. |
| ϵ | REMD | -152. | -157.5 | -98. | -101. | -147. | -157. |
| | MD | -151. | -50. | -88. | -91. | -74 | -153. |
| | NMR_s | -140. | -152. | -78. | -127. | -147. | -148. |
| | NMR_v | -148. | -165. | -101. | -114. | -147. | -151. |
| χ | REMD | -154. | -166. | -155. | -136. | -152. | -152. |
| | MD | -156. | -148. | -58. | -134. | 35. | -143. |
| | NMR_s | -158. | -161. | -156. | -149. | 73. | -166. |
| | NMR_v | -140. | -140. | -149. | -126. | 61. | -165. |
| ζ | REMD | -70. | -76. | -97. | -103. | 74. | 146. |
| | MD | -66. | -61. | -91. | -68. | 70. | 23. |
| | NMR_s | -65. | -65. | -100. | 135. | -55. | |
| | NMR_v | -67. | -92. | -58. | 105. | -41. | -68. |
| Pucker | REMD | 12. | 33. | 165. | 157. | 32. | 6. |
| | MD | 17. | 24. | 148. | 152. | 44. | 18. |
| | NMR_s | 13. | 29. | 160. | 155. | -16. | -55. |
| | NMR_v | 20. | 19. | 143. | 149. | 50. | -5 |

Table 4.1: Comparison of calculated at 300 K and experimental torsional angles of the UUCG tetraloop (in degrees). The REMD and the MD angles stem from the 10 mer CGCUUCGGCG and the 14 mer GGCACUUCGGUGCC, respectively. NMR_s angles are taken from the 14 mer, determined in the Schwalbe group [2], while the angles of a 12 mer GGACUUCGGUCC are obtained from Allain and Varani [3].

4.2.2 3J coupling

The 3J coupling constants of the 14 mer were calculated from the MD simulation using the Karplus relations (Eqs. 2.19,2.21,2.22,2.20) [84, 85]. The averaged calculated 3J values together with the experimental values are reported in Figure 4.4. From the coupling constants, $^3J_{H1',H2'}$ and $^3J_{H3',H4'}$, the pseudorotations of the ribose can be determined. These couplings are in overall good agreement with NMR data as also was the case for the pucker angles (Fig. 4.3). The

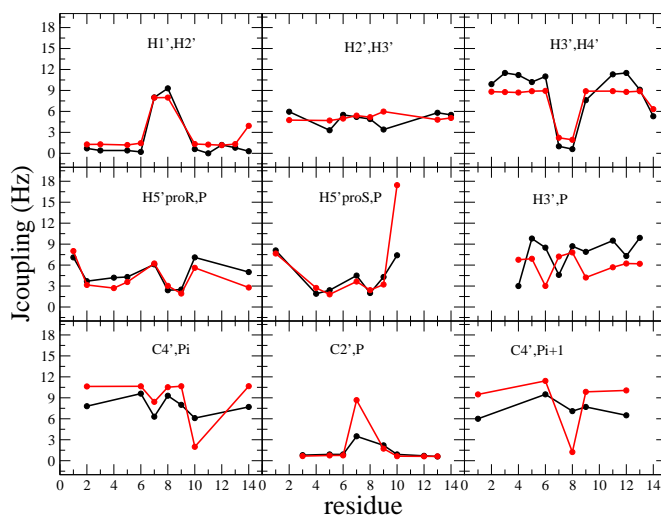


Figure 4.4: Comparison between MD (red) at 300 K and NMR (black) at 298 K of the backbone torsional angles for the 14 mer UUCG loop.

couplings $^3J_{H5'proS/R,P}$ and $^3J_{C4',Pi}$ reflect the angle β . The large deviation observed for $^3J_{H5'proS,P}$ of G10 correlates with the disagreement in the β angle value for this residue, as shown in Figure 4.3. For the UUCG loop as well in the case of the UUUU loop (Chapter 3) $^3J_{C4',Pi}$ and $^3J_{C2',P}$, which correspond to β and ϵ , differ significantly from NMR despite overall good agreement with NMR derived angles (Table 3.3). Note that in the case of the UUUU loop enhanced sampling technique was also used. The deviation observed can therefore be due to the Karplus relation (Eq. 2.20) which may not be appropriate in the case of $^3J_{C,P}$.

4.2.3 S^2 order parameter

A dynamical picture of the base and ribose moiety have been determined via ^{13}C NMR relaxation for all residues [52]. The squared generalized order parameter S^2 reflects the amplitude of the internal motions. In this context the S^2 parameter is calculated and compared to NMR data (Fig. 4.5). To be consistent with NMR the

C8-H, C6-H vectors of the bases and the C1'-H vectors of the ribose are taken for the calculation of the S^2 values. An agreement can be observed between NMR and MD results. Both MD and NMR indicate for the base and the ribose of U7 the highest flexibility, which relates to small S^2 values. This also agrees with the structural

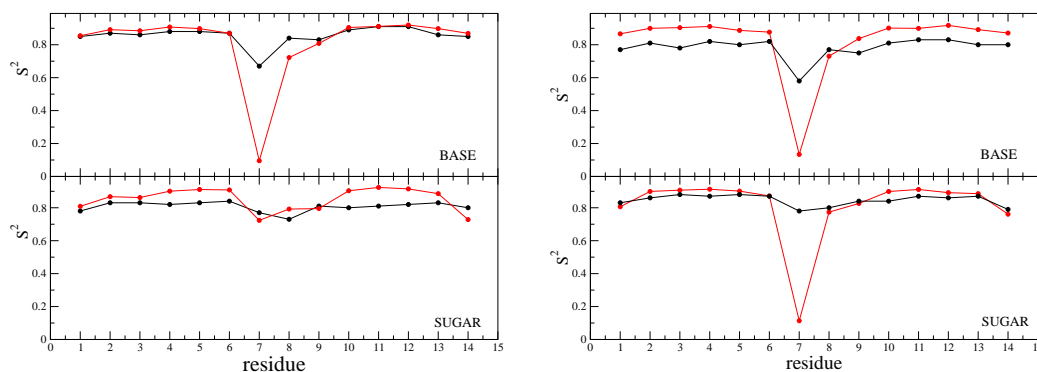


Figure 4.5: Comparison between MD (red) and NMR (black) of the order parameters of the 14 mer. Left panel: at 300 K (MD) and 298 K (NMR). Right panel: at 317 K.

features of the UUCG motif [1, 106] where all residues are rigid ($0.86 < S^2 < 0.97$) in both the base and the ribose moiety except for the base U7, which is protruding into the solution ($S^2 = 0.66$ in NMR and $S^2 = 0.08$ in MD). The S^2 values indicate uniform flexibility in the stem region, whereas the terminal nucleotides (G1 and C14) are of higher flexibility than the internal ones. The riboses exhibit in general slightly more rigidity than the bases. U7 shows in MD high flexibility as can be observed in the extremely low S^2 values (in comparison to NMR) for both temperatures (300 K and 317 K). One explanation for the very low S^2 values in MD is that the motional behaviour of U7 may have not converged during simulation, due to its high flexibility, leading to overestimations of the order parameters.

4.2.4 Dynamical analysis

The next step is to identify the molecular motion that resulted in the shown S^2 values. Figure 4.6 shows the root mean square fluctuations of the atoms in the loop region around the initial structure. From this picture, the atoms (205-215) of nucleobase U7 undergo large changes, as can be seen in the high atomic fluctuation value of around 3\AA . As we have seen previously a small S^2 value reflects a high motional dynamics of the C-H bond vector of interest. Therefore, a large atomic fluctuation corresponds to a small S^2 value. This means that the previously seen lowest S^2 of the U7 nucleobase is in agreement to the largest atomic fluctuation. Also in agreement is the finding that the atoms of the U6 residue (even though not significantly) have the lowest atomic fluctuation and hence a relatively high S^2 of around 0.9 than all

other loop residues. This is due to base stacking interaction between U6 and C5 observed in our simulations and in X-ray experiment [1].

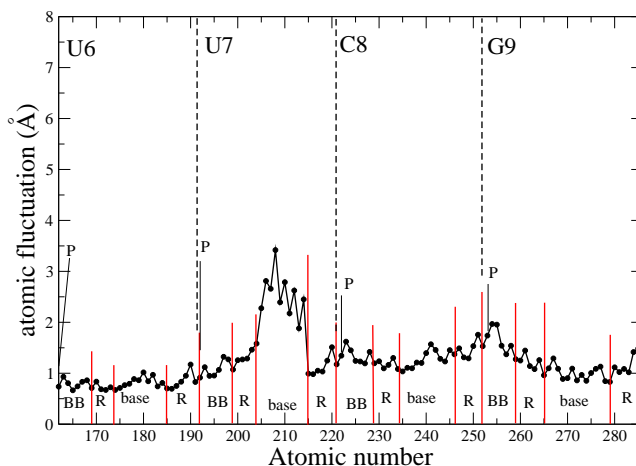


Figure 4.6: The atomic fluctuations in the loop region around the initial structure of the 14 mer UUCG from a conventional MD simulation at 300 K. P: phosphor, BB: Backbone, R: Ribose

In general, there is no significant difference in the atomic fluctuations between the atoms of the C8 residue and the G9 residue, but a slightly higher fluctuation for the C8 base can be observed. This agrees with the experimentally observed slightly higher S^2 value for C8 than for G9 base atoms. However, the highest atomic fluctuations of the C8 and G9 do not arise from the nucleobases but instead from the backbone atoms. Especially the phosphate group seems to be responsible for the flexibility of these residues (Fig. 4.6), as can be observed in the high atomic fluctuation of the phosphate group. In particular the atomic fluctuation of G9 phosphate atoms is higher than of all other atoms in this nucleotide. From the MD simulation at 317 K this observation is even more pronounced, especially for the C8 phosphate group (Fig. 4.7). It is interesting to see that the high flexibility of the ribose and nucleobase of U7 does not induce a flexibility in the phosphate backbone, as it remains relatively stable with an atomic root mean square value of around 1 Å, similarly to the ones at 300 K.

A circle representation of the dihedral and pseudorotational angle distributions for the 14 mer UUCG loop at 300 K is shown in Figure 4.8. The pseudorotational angles of the stem region are in the C3'-endo conformations except for the terminal residue C14 which exists in the two RNA preferred ribose conformations (C2'-endo/C3'-endo). In the loop region U7 and C8 residues can be found in the C2'-endo conformations, whereas U6 and G9 prefer the C3'-endo. This is also the case for the 317 K simulation, whereas U7 seems to have a stronger repuckering in this case.

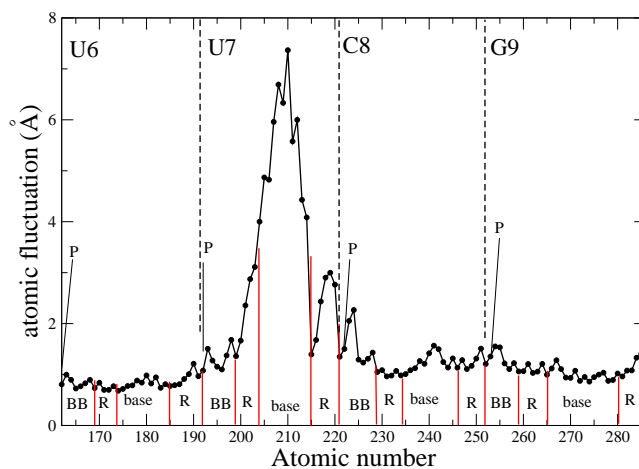


Figure 4.7: The atomic fluctuations in the loop region around the initial structure of the 14 mer UUCG from a classical MD simulation at 317 K. P: phosphor BB: Backbone, R: Ribose.

The χ angle exhibits the highest flexibility in the loop residue U7 (Fig. 4.8), which correlates well with the low S^2 value of the base (Fig. 4.5) observed both in MD and NMR. The α and γ angles exhibit in more than three residues transitions between two conformations, while the β , δ and ζ angles fluctuate around their averaged values. The ϵ angles at 300 K appear to be stable except for G9 residue where twice a transition from around 200° to 280° is observed (Fig. 4.9), this can be seen clearly in the time trace presentation of the ϵ angle in Figure 4.10. Therefore, this finding supports the observation of the high atomic rms fluctuation of the phosphate group in residue G9. It should be reminded that the ϵ and ζ angles represent the phosphate backbone. At 317 K the ϵ of G9 stays stable, instead the ζ of the C8 exhibits stronger motions.

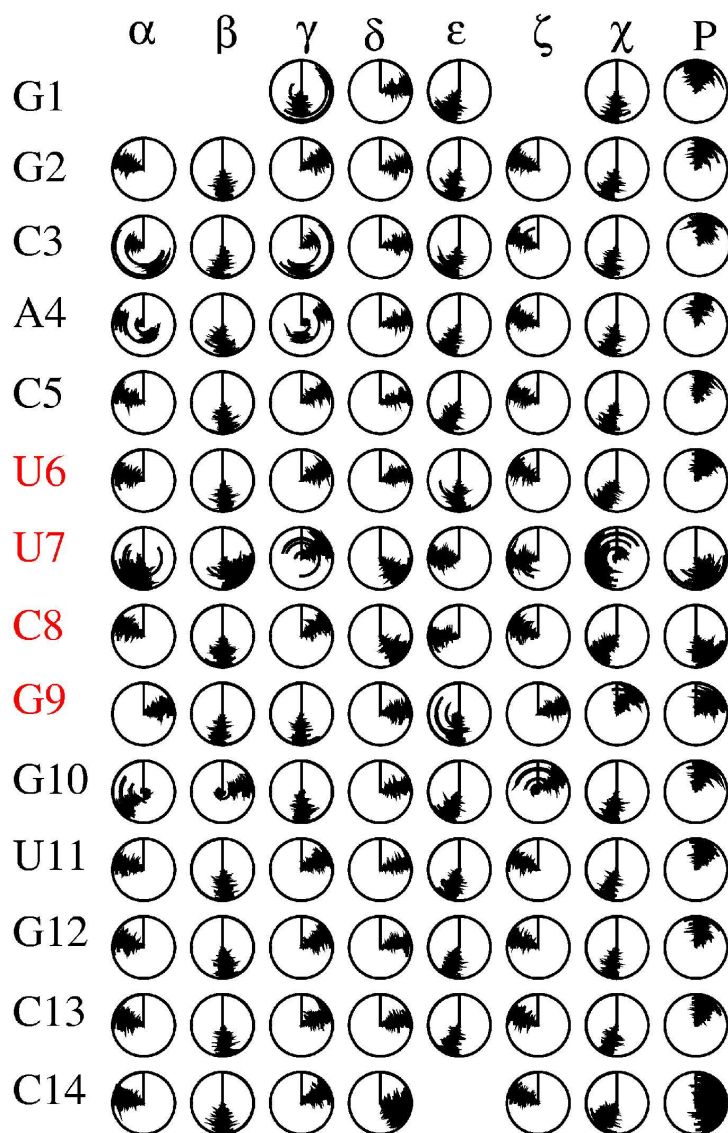


Figure 4.8: Circle representation of the backbone angles of the UUCG 14 mer at 300 K.

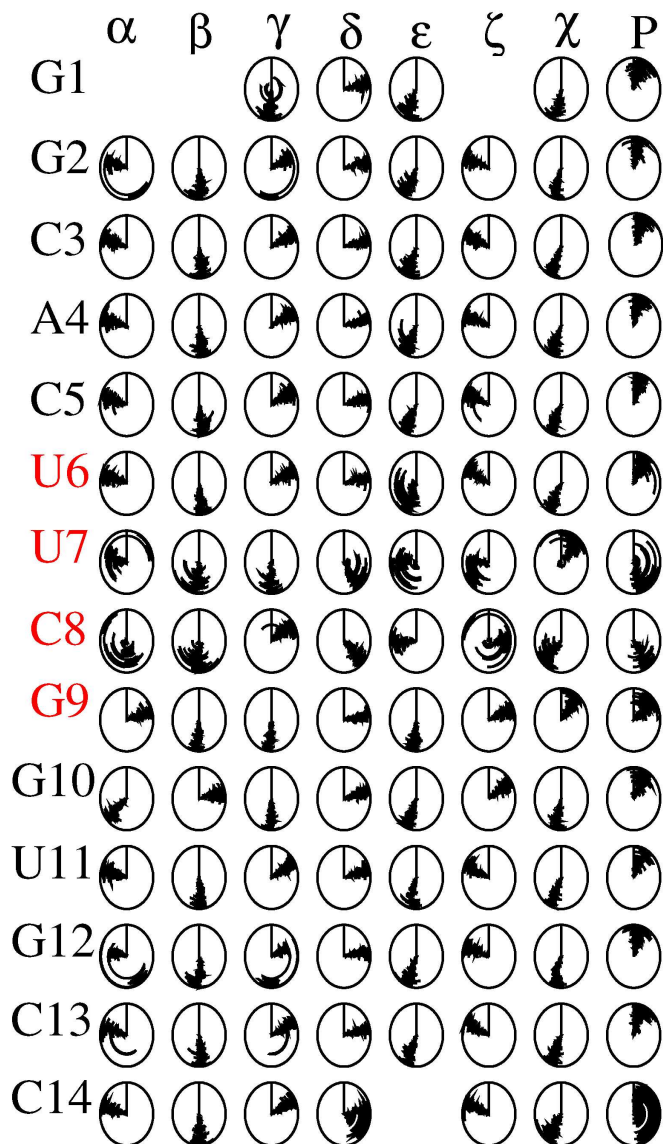


Figure 4.9: Circle representation of the backbone angles of the UUCG 14 mer at 317 K.

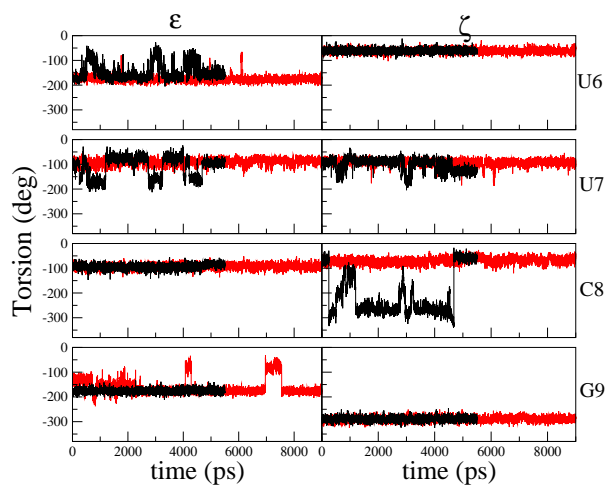


Figure 4.10: Backbone angles ϵ and ζ of the 14 mer UUCG loop at 300 K (red) and at 317 K (black).

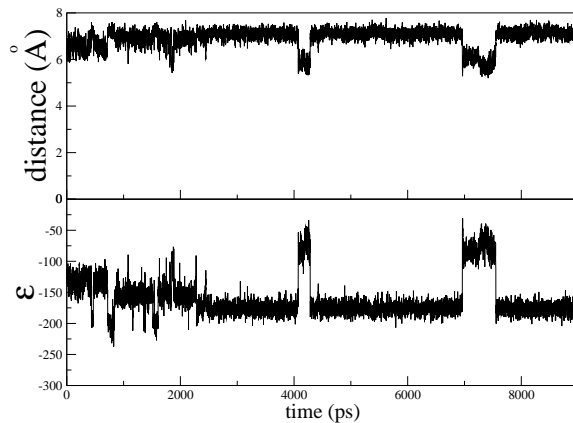


Figure 4.11: Upper panel: Time trace of the distance between the center of mass of phosphate G9 and G10. Lower panel: Time evolution of angle ϵ of the G9 residue of the 14 mer at 300 K.

The distance of the phosphate backbone of G9 relative to its neighbouring phosphate, to the corresponding nucleobase, and to the ribose has also been checked. Indeed, there were no remarkable variations in the distance values except for the distance between the phosphate of G9 and G10 (Fig. 4.11). The time trace of this distance (lower panel of Figure 4.11) is correlated to the time trace of the ϵ angle.

At higher temperature namely at 317 K as we have seen in Figure 4.7 the rms atomic fluctuation of the phosphate group C8 was remarkably larger than the base and the ribose of the same residue and the G9 residue. This high fluctuation can be related to the high flexibility of the ζ angle as can be monitored in Figure 4.9 and 4.10 at 317 K.

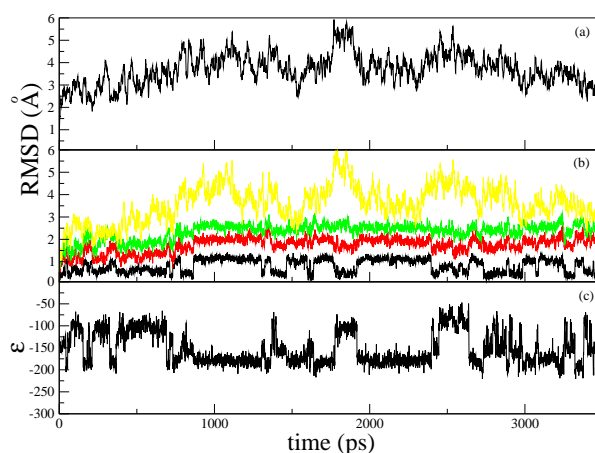


Figure 4.12: Time traces from a 400 K trajectory of the 14 mer UUCG loop. a) RMSD of the total tetraloop. b) RMSD of phosphate groups. Black: G9-G10. Red: U7-G10. Green: C5-G10 and Yellow: G2-G10. c) G9 ϵ angle.

It is not clear whether the trajectories at 300 K and 317 K (Fig. 4.10) are converged. To clarify this fact, a high temperature MD simulation at 400 K was performed. Proper base pairing in the stem could be observed, despite of the high temperature. The time traces of the root mean square deviation (RMSD) in Figure 4.12 could be an indication that the trajectory is converged. This trajectory also shows a correlation between the ϵ angle (Fig. 4.12c) of G9 and the G9-G10 phosphate RMSD (Figure. 4.12b in black). Another obvious correlation can be visualized when the RMSD (Fig. 4.12a) of the total system is compared to the ϵ angle of G9 (Fig. 4.12c), in fact the main transition of the RMSD at around 1800 ps is caused by the transition of the G9 ϵ angle.

When decomposing the total RMSD into smaller parts by calculating the RMSD (Fig. 4.12b) of the phosphate residues downstream of G10 together with G10, a clear description of the total RMSD from these individual subsystems can be seen. Therefore, the total dynamics of the 14 mer is merely caused by the phosphate backbone between G9 and G10. This is especially for the 300 K and 400 K trajectories well

pronounced. The 317 K trajectory on the other has not converged, and therefore the correlation the total RMSD and the backbone of G9 could not be observed here.

4.3 The 10 mer UUCG loop

In the previous section, we observed in the case of the 14 mer UUCG loop a backbone dynamics of the last loop residue (G9 in Figure 4.2) to be responsible for the flexibility of the RNA molecule at 300 K and 400 K. It is therefore interesting to see whether a UUCG loop with a slightly different stem will show the similar dynamical behavior. In order to achieve that we will take a look at the correlation between the backbone ϵ angle of last loop nucleotide (G7 in Figure 4.13) and the total RMSD. For this reason the 10 mer tetraloop 5'-CGCUUCGGCG-3' (Fig. 4.13) will be investigated using results from conventional MD and REMD simulations at 300 K.

4.3.1 Conformational analysis

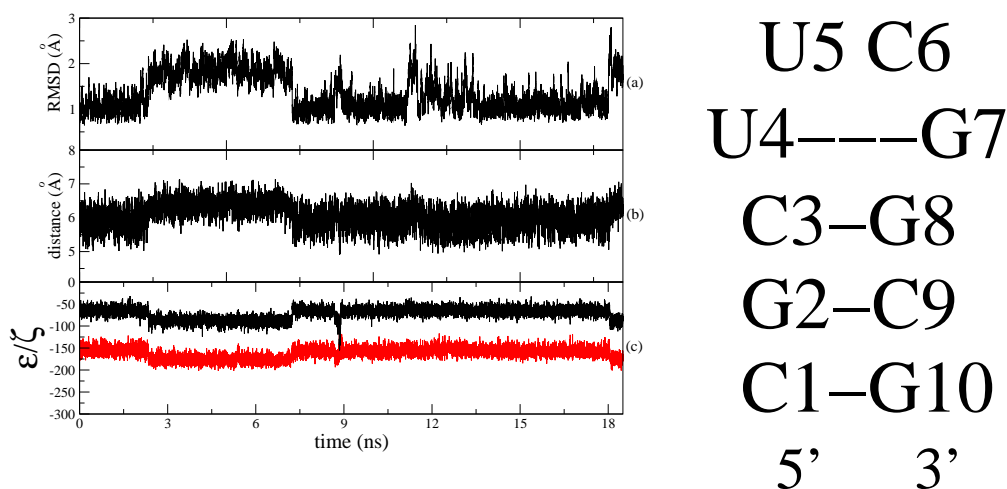


Figure 4.13: Dynamics of the 10 mer UUCG at 300 K. a) RMSD of the total tetraloop b) Distance between the phosphate of C3 and U4 and c) Angles ϵ and ζ of the C3 residue

A 20 ns MD simulation of the 10 mer has been performed. Indeed, we could observe for this system a correlation of the total RMSD with the ϵ angle. In the case of the 14 mer we saw that the backbone of the last loop nucleotide is responsible for the dynamics of the system. However, in the case of the 10 mer it seems that the backbone of the first loop residue with its adjacent loop closing nucleotide C3 is responsible for the total dynamics (Fig. 4.13).

An REMD simulation of 10 ns is used to characterize the conformations of the 10 mer UUCG loop. PCA was performed (Fig 4.14) on the REMD trajectory at 300 K. The analysis gives four main conformations (Fig 4.15). The conformations are achieved from the distribution of the higher principal components (Fig 4.14). Then components with Gaussian shaped distribution, in this case component number 3 and components after the sixth principal one, are omitted in the analysis of the conformations (Fig 4.14). The first four components are therefore used for the characterization of the conformations as shown in Figure 4.15. Defining for each of the four components ambiguously sub-states 1 and 2 and a combination of all possible sub-states resulted in 16 states. However, the population probabilities of many of these states are low which is why only the states with higher occurrence are shown in Figure 4.15.

From the four states that are obtained from PCA, conformation A occurs with the highest population probability. Both A and C conformations exhibit U4-G7 base pairing and C3-U4 stacking, while only conformation A exhibits stacking between C6 and U4. In conformation B and D, a slight disruption of the loop base pair (U4-G7) can be seen. U4 is therefore more exposed to solvent than in A and C. In the following chapter the relation of U4 pointing into the solvent with the disruption of the C3-U4 stacking and the U4-G7 base pairing will be deeply discussed.

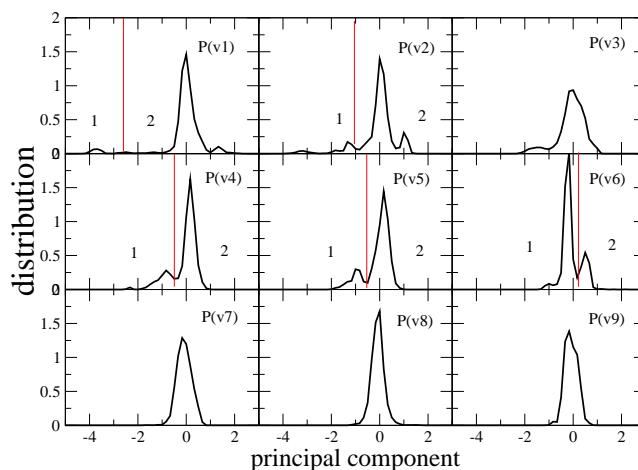


Figure 4.14: Distribution of the first nine principal components gained from the REMD trajectory at 300 K. The determination of the sub states (labeled with 1 and 2) is done using the first six (without component 3) principal components for the 10 mer UUCG loop.

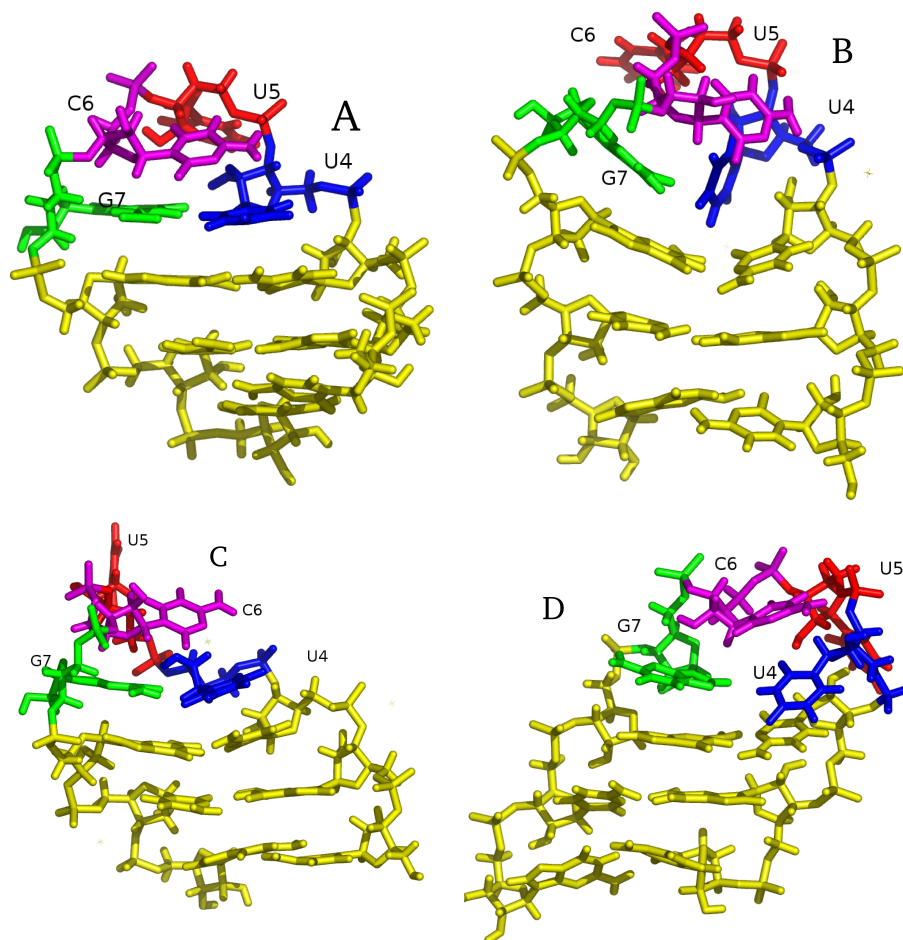


Figure 4.15: Conformations A, B, C and D of the 10 mer occur with high population probability in the trajectory of the REMD simulation at 300 K: A=61%, B=9%, C=8% and D=5%. These conformations are gained by combining all sub states gained from the first six (except number 3) principal components as shown in Fig. 4.14.

4.4 Conclusions

We have presented a detailed and comprehensive MD study of the structure and the dynamics of the RNA UUCG loop, by performing conventional MD as well as enhanced sampling using REMD simulation. Results from conventional MD simulation on a 14 mer are in overall good agreement with experimental data. A dynamical picture of the loop could be drawn from the results gained from the atomic fluctuations and the backbone angles. We monitored high flexibility in the backbone of the first loop nucleotide U in the case of the 10 mer and in the last loop residue G for the 14 mer. The difference in these two systems is in the second base pair of the stem (adjacent to the loop closing base pair); here the A-U base pair of the 14 mer is substituted by a G-C in the 10 mer.

The uracil base of the second loop residue is always pointing for both systems into the solvent, therefore this residue shows the highest flexibility. Even though the highest atomic fluctuation is observed for the uracil base, the total RMSD of the tetraloop even at 400 K, depicts stronger correlation with the phosphate backbone than with this residue. It seems that the ϵ angles of the backbone is responsible for the flexibility of this stable tetraloop. It has still to be clarified whether the backbone of the first residue of the loop or the last loop nucleotide or both are responsible for the flexibility of the system. REMD simulations of the 10 mer indicate a slightly higher preference for the first loop nucleotide. To be able to understand better the MD results on the 14 mer, simulations of the 10 mer with an A-U base pair at the second position will be performed.

Chapter 5

Unfolding studies of the UUUU and UUCG loop

5.1 Introduction

In this chapter a comparison of the two tetraloops in terms of thermal unfolding is given. Recently, it has been shown experimentally that the 12 nucleotide UUCG unfolds slower than the counterpart UUUU loop [40]. However, a detailed description of the structural dynamics of the unfolding process is still not clear. Therefore, we use the advantage of the REMD simulation method to study the unfolding mechanism of the 10 nucleotide UUCG and UUUU hairpin.

Experimental results have shown that the UUUU loop is less stable than the UUCG with the melting temperature T_m of 329.8 K for the UUUU and 345.6 K for the UUCG tetraloop [41]. As we have seen from the simulations at 300 K for the two tetraloops, UUUU is stabilized in the loop region by merely three stacking interactions. In the case of the UUCG various hydrogen bondings and additional stacking interactions are responsible for its thermodynamic stability.

Here we would like to present a few temperature-dependent observables, for instance the hydrogen bonds, the existence of native contacts (base pairing and stacking interactions relative to a reference structure) (Fig. 5.3) and the radius of gyration. These are good “reaction coordinates” which can be used to describe the unfolding of the tetraloops. The experimentally observed melting temperature difference ΔT_m of about 15 K between the hairpins is in agreement with our finding.

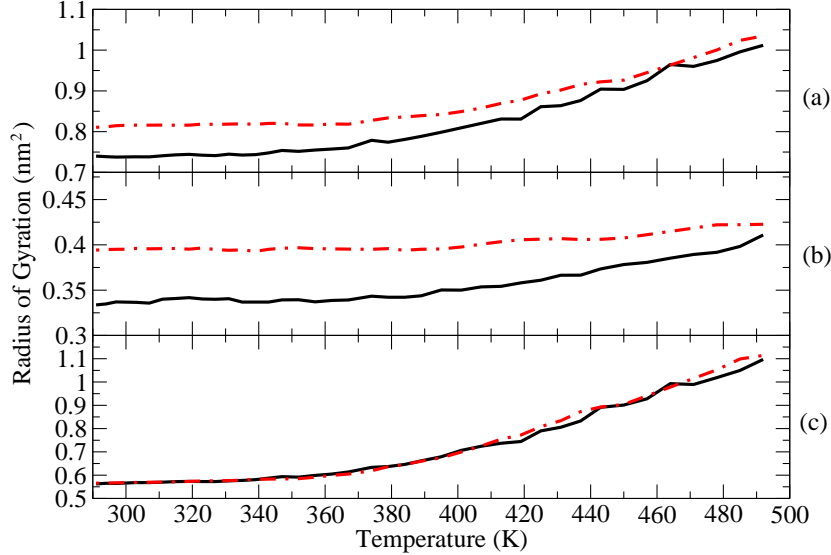


Figure 5.1: Radius of gyration for the UUCG (black) and UUUU (red). a) For the whole system. b) For the loop region. c) For the stem area.

5.2 Observables of thermal unfolding

In order to monitor the overall shape of the hairpins the radius of gyration (R_g) was calculated. R_g is calculated as the following

$$R_g = \sqrt{\frac{\sum_{i=1}^N m_i (R_i - R_c)^2}{\sum_{i=1}^N m_i}}, \quad (5.1)$$

where R_i is the atomic position of atom i (with mass m_i) from the center of mass R_c . The difference between the two hairpins in their R_g values can be seen in Figure 5.1b, which corresponds to the loop region. The UUCG has significantly lower R_g values compared to the UUUU loop. This is obvious as the UUCG loop has a more compact structure resulting from the loop base pair and the stacking interactions between the loop nucleotides. The UUUU loop exhibits large R_g values even at lower temperature which is a consequence of lack of interactions between the loop residues, causing the bases to easily protrude into the solvent and therefore enlarging the overall geometry. The main interactions, which were observed for the UUUU hairpin are merely U4-U5, U3-U4 stacking and a weak stacking of U5-U7 (Figure 5.3, see also chapter 3). Even though a similar overall shape of the stems (Figure 5.3c) can be witnessed, the analysis of another observable, shown below, clearly exhibits a difference between the two hairpins.

Experimental observations indicate that the UUCG is stabilized in the loop region by various hydrogen bonds and stacking interactions [1]. On the other hand the UUUU loop contains only stacking interactions between the loop residues and lacks any stabilizing hydrogen bonds in the loop area (See also chapter 3). In the upper panel of Figure 5.2 the observed hydrogen bonds in the UUCG and UUUU hairpin are depicted schematically. In the middle panel, base pair disruption is characterized by the fraction of the hydrogen bonds between atom N3 and H1 (C1-G10,G2-C9,C3-G8). The terminal base pair is highly flexible and therefore this base pair disrupts first. The second and the third base pair (loop closing base pair) indicate a simultaneously breaking of the hydrogen bonds. The temperature dependence of the fraction of hydrogen bonds in the loop of UUCG is depicted in the lower panel of Figure 5.2. It can be seen that hydrogen bonds 10, 11 and 12 break later than 13, 14 and 15. The first ones are stronger as residue U4 stacks with C3 and G7 stacks with G8, resulting in a stabilization of the mentioned hydrogen bonds.

A similar observable to the hydrogen bonds but also takes base stacking interactions into account are the native contacts (NC). A NC (Fig. 5.3) exists between two stacked or paired bases if their center of mass separation is within 1.5 Å relative to an NMR structure for the UUCG and relative to conformation 2 (Fig. 3.10) for the UUUU loop. In the middle panel of Figure 5.3 the averaged NC for the UUUU loop is lower than UUCG for temperature ≥ 360 K. From the melting curves of the hydrogen bonds (Fig. 5.2) and the native contacts (Fig. 5.3) we can approximately extract the melting temperature by choosing the temperature at which 50% of the structures still form native contacts or hydrogen bondings. We obtained T_m of around 400 K for both hairpins. These melting temperatures are significantly higher than experimental ones. Nevertheless difference in the melting temperature ΔT_m of 15 K is in good agreement with experiment. Furthermore, the unfolding behavior of the tetraloops which is monitored experimentally [40] is similar to our results .

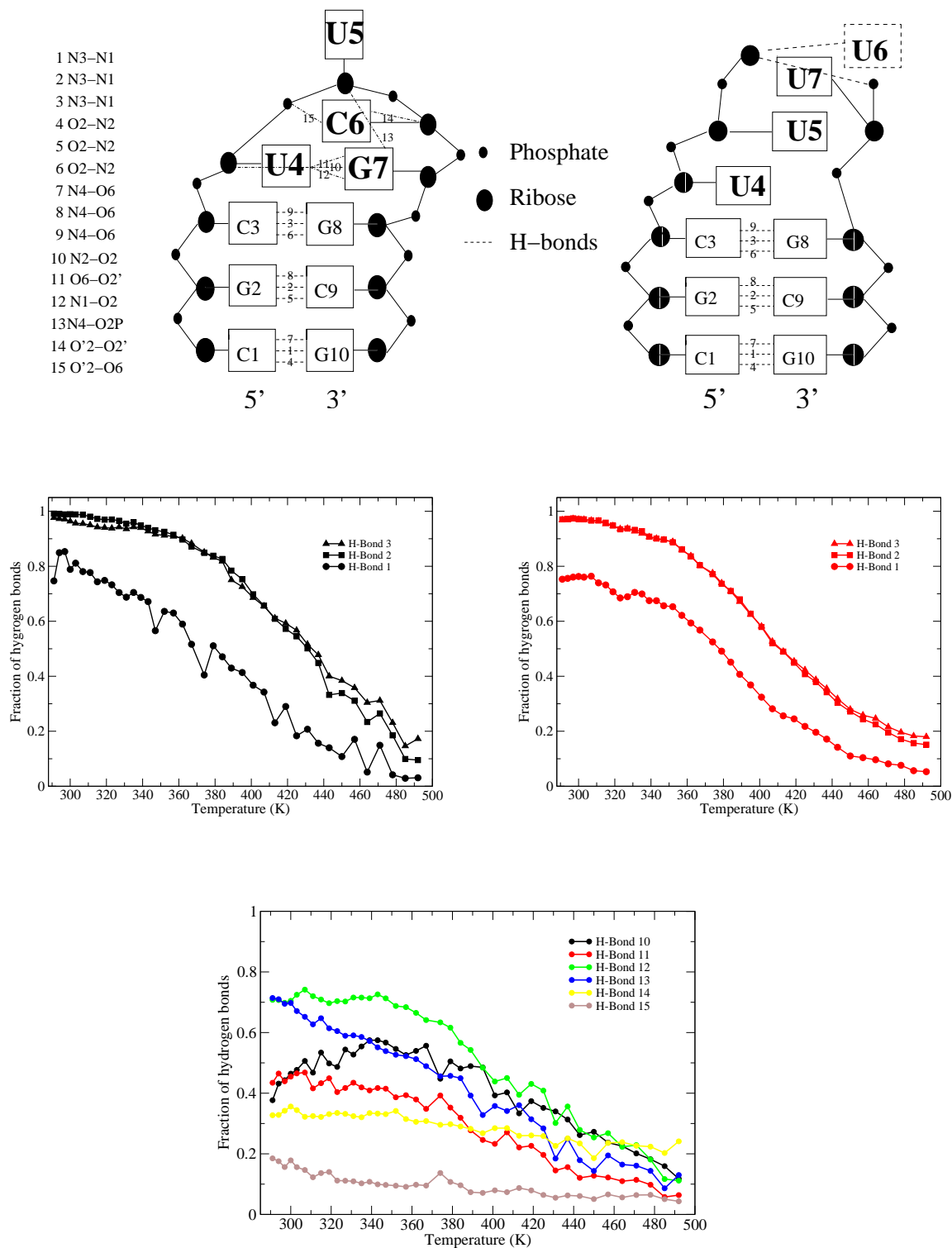


Figure 5.2: Upper panel illustrates the hydrogen bonds in the UUCG and UUUU loop. Middle panel depicts hydrogen bond opening in the stem of UUCG (black) and UUUU (red). Lower panel shows the hydrogen bonds of the loop region of the UUCG only.

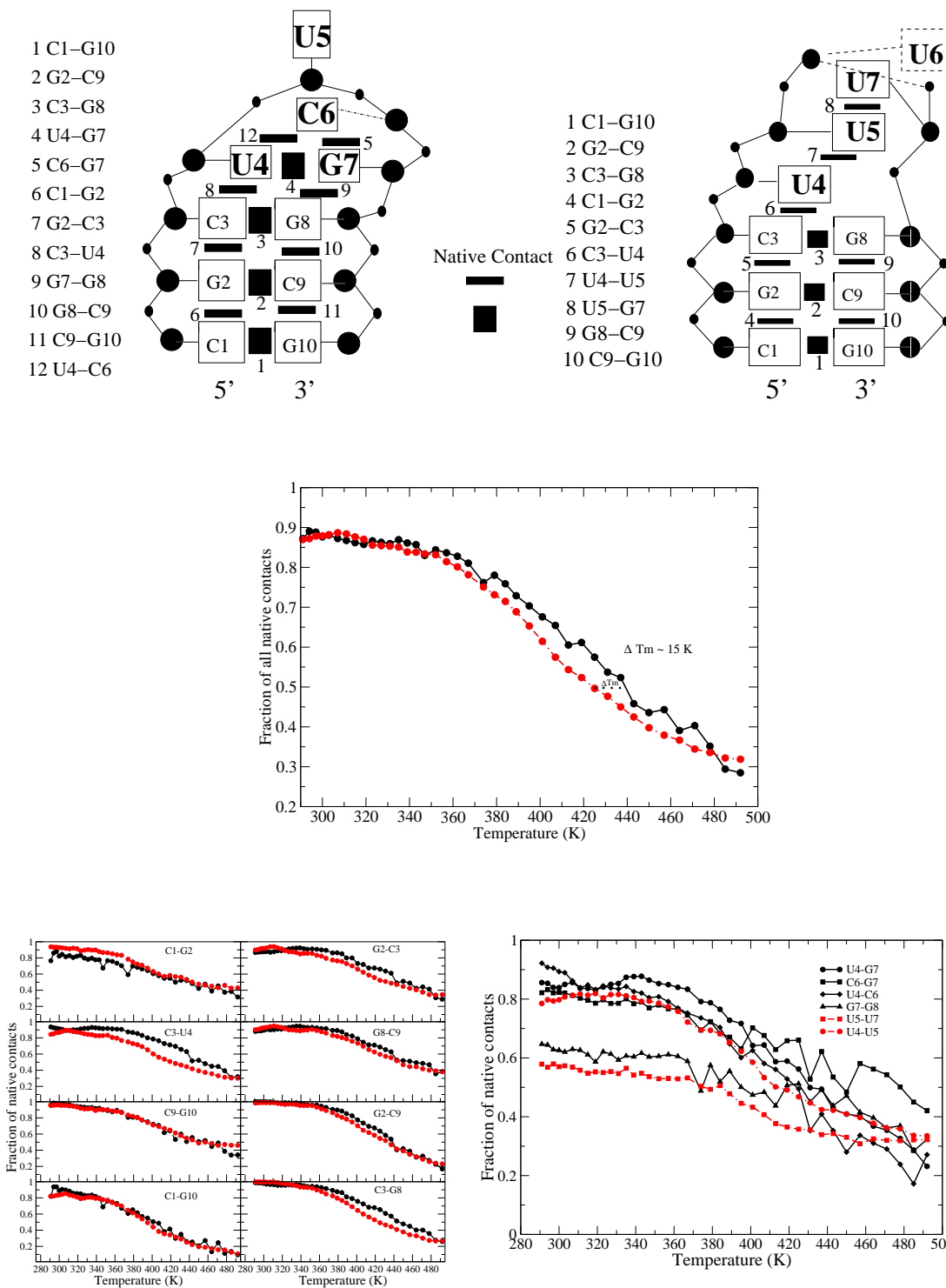


Figure 5.3: Upper panel illustrates the native contacts (NC) which arise from stacking or pairing between bases. The NC exist if they are within 1.5 \AA of that seen in NMR (UUCG) and of that seen in conformation 2 (Fig. 3.10) for the UUUU loop. Middle panel indicates the fraction of all native contacts (NC) for the UUCG (black) and UUUU loop (red) as a function of temperature. Lower panel shows the plots of the NC for the stem (left) and the loop region (right) of both hairpins.

It appears that simulations generally tend to overestimate the melting temperatures. Pitera and Swope obtained a much higher T_m value for their miniprotein system, using AMBER and a GB/SA solvation model [107]. Garcia and Sanbonmatsu obtained relatively high melting temperature for their peptide system in explicit water with the standard AMBER force field but after a slight modification of the force field the theoretically determined T_m value was comparable to the experimental ones [108]. The discrepancy in our case can also be due to the force field, signifying an over stabilization of the folded state in our computational model.

A closer analysis of the results shown in Figure 5.3 reveals a small minimum around 320 K for the NC C3-U4. Even though this minimum seems to be of small relevance, the fluctuation of the native contacts, however, results in a peak that stems from the dynamics of the C3 and U4 residues (Fig. 5.4). This peak comes from the mobility of the backbone residue U4, causing the U4 base to point into the solvent. This finding is in accordance to the results obtained from conventional MD at 400 K (see previous chapter), where the backbone dynamics of residue C3 and U4 was assumed to be responsible for the flexibility of the UUCG. The fluctuation of the NC of the UUCG does not exhibit a smooth curve for temperatures ≥ 400 K, as can be seen in the various spikes (upper panel of Fig. 5.4). These spikes result from the NC 5, 9 and 12, which corresponds to weak stacking interactions. It is noticeable that the curves of the NC formed by terminal base pairs (C1-G10), (C1-G2), (C9-G10) are almost the same for both hairpins. This means that the difference in the loop residues does not result in the difference for the terminal base pair.

From the melting curves of the loop region in shown in Figure 5.3 (lower right panel) we can clearly see that the UUUU loop has less native contacts than the UUCG loop for temperatures above 360 K. At lower temperatures we have for both systems about 90% of the structures containing all their native contacts. The other 10% mainly stem from the disruption of the terminal base pair and the breaking of the native contacts between U4 and C3 and between U4 and U7 for the UUCG loop.

Principal component analysis also shows one of the conformations of the UUCG loop containing the base of residue U4 pointing into the solvent. This conformation stems from the flexibility of the backbone between C3 and U4, as we have discussed previously. The unfolding of both loops occurs in two recognizable steps. The terminal base pair opens first, causing the second and the loop closing base pair to unfold almost simultaneously (Fig. 5.2). At the same time the loop residues are losing their interactions and become more flexible. Here we observe the main difference between the two loops. From Fig. 5.3 we can clearly see that the most stable native contact in the loop region of UUUU loop is U4-U5. But in comparison to the most stable native contact of the UUCG loop, that is U4-G7, the NC U4-U5 of UUUU loop melts already at lower temperature. From this picture it is also obvious that the additional native contacts NC 6 and NC 12 in the UUCG loop are also responsible for the higher stability of this loop.

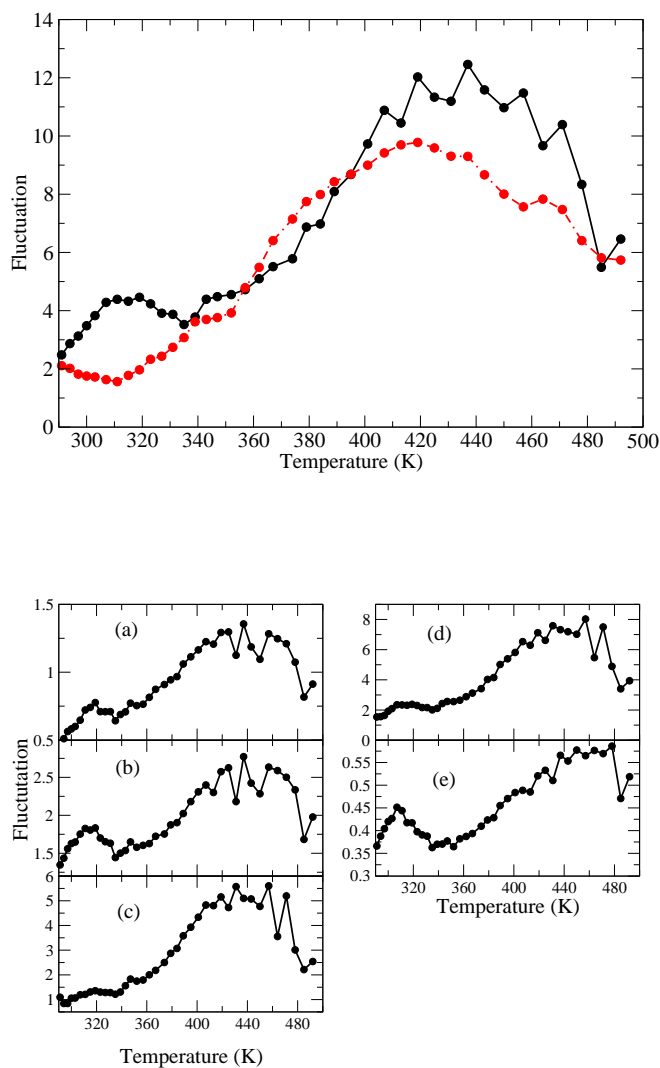


Figure 5.4: Upper panel shows the fluctuation of all native contacts for the UUCG in black line and for the UUUU in red dashed line. Lower panel shows the fluctuation of the native contact for the UUCG loop when divided into the subsystems. a) Fluctuation of the NC in the loop. b) Fluctuation of the NC of the loop when coupled to the loop closing base pair (C3-G8). c) Fluctuation of the NC in the stem. d) Fluctuation of the NC of the stem when coupled to the residue U4 and G7 of the loop. e) Fluctuation of the NC of the residues of the loop closing base pair C3 and G8 and the terminal residues in the loop.

5.3 Conformations at high temperatures

The free energy surfaces spanned by the first two principal components have been studied for two temperatures, 331 K and 413 K (457 K for the UUCG loop). These temperatures have been chosen due to the following reason. At these temperatures we expect the system to exhibit significant change in the conformations. Approximately at around 331 K the systems begin to lose their first native contacts and hydrogen bonds (Fig. 5.2 and 5.3). At around 413 K and 457 K, respectively, the UUUU and UUCG loop are around their melting temperatures. In this section the results from the trajectories at 331 K and 413 K (457 K) are shown.

When comparing the two tetraloops at 331 K, the free energy landscape of the UUCG loop seems to be more rugged than that of the UUUU loop (Fig. 5.5,5.6). The UUCG loop exist in about 80% of the conformations in the folded state, in which all essential native contacts and hydrogen bonds are still persisting. On the other hand, the UUUU loop exhibits at this temperature more structures that have lost certain native contacts. This is another indication for the lower stability of this loop. The free energy landscape of the two hairpins (Fig. 5.7 and 5.8) exhibits a single prominent minimum, the rest of the landscape is relatively smooth. This minimum corresponds in both tetraloops to the native folded state. The rest of the free energy landscape is mostly due to partially or totally unfolded structures. The unfolded structures occur with very low probabilities.

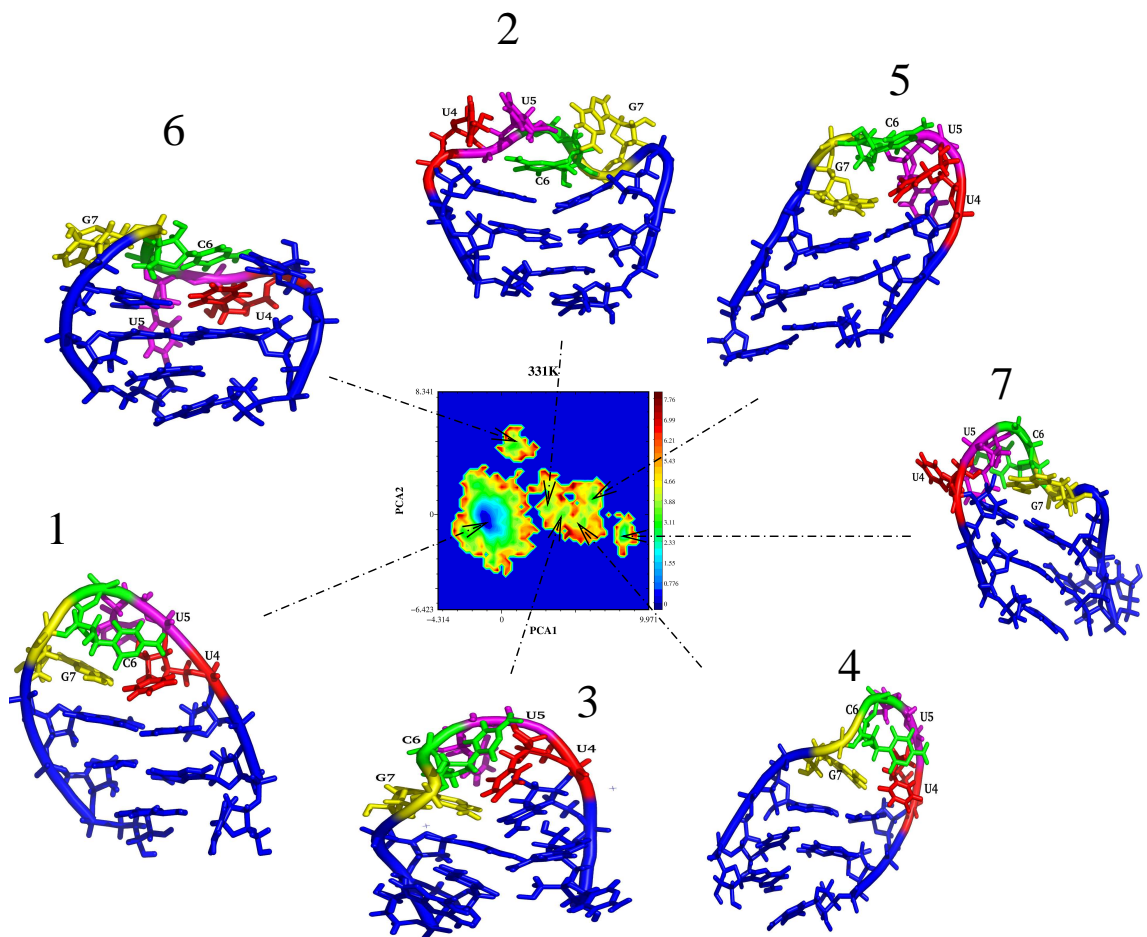


Figure 5.5: Free energy surface (in units kT) obtained from the REMD simulation at 331 K for the UUCG loop, plotted as a function of the first two PCA eigenvectors.

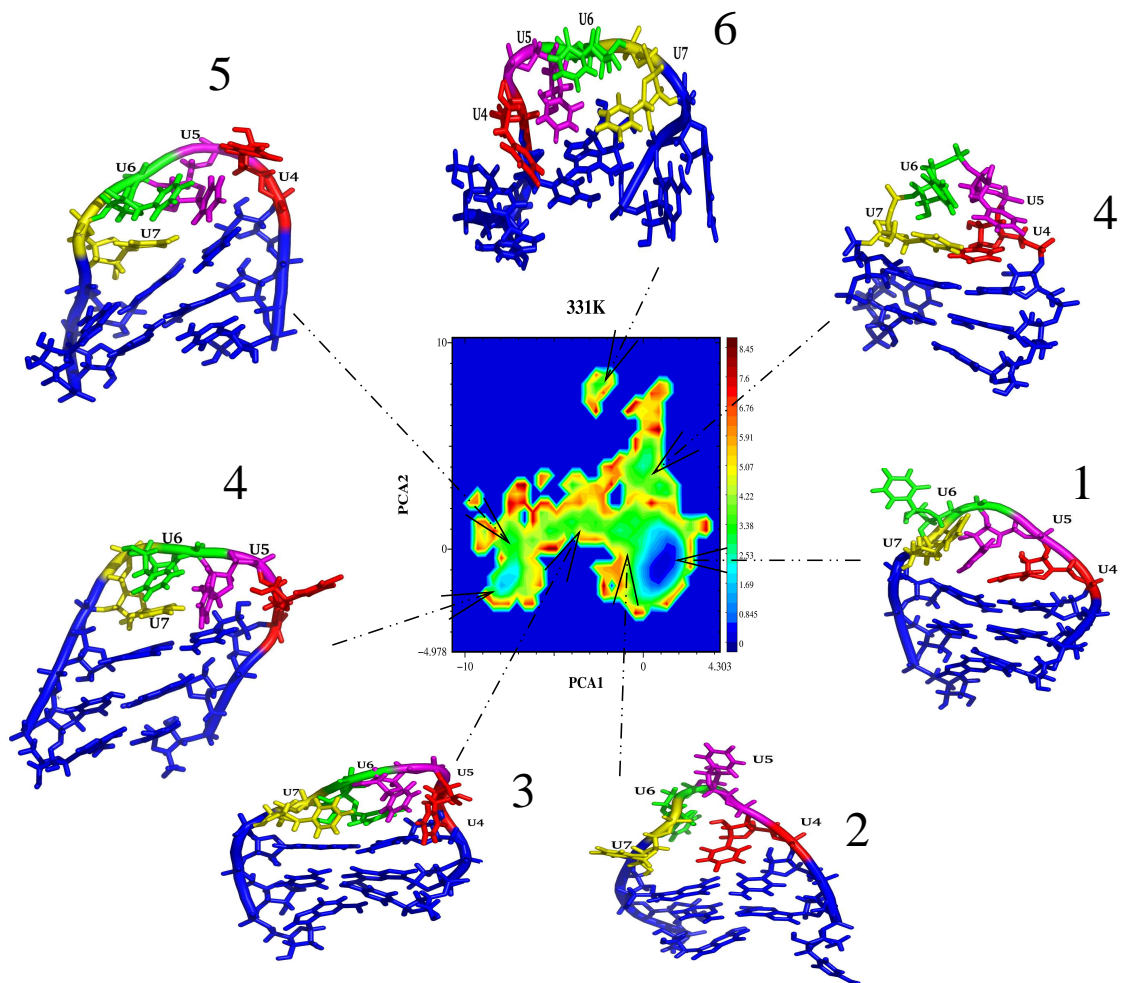


Figure 5.6: Free energy surface (in units kT) obtained from the REMD simulation at 331 K for the UUUU loop, plotted as a function of the first two PCA eigenvectors.

For the states obtained from the free energy surface at the studied temperatures, the backbone and the pseudorotational angles are shown for the loop residues and the residues of the loop closing base pair (Tab. 5.2,5.1). The result of the torsional angles shows how the angles change when going, for example, from the folded state 1 to the unfolded state 5. A big difference of the backbone angles can be observed for all angles between state 1 and 5. Also, a significant change can be observed in the pucker angles. Nevertheless, the overall picture of the conformation of state 2 is still similar to those of the folded one (state 1). In state 2 (413K) of the UUUU loop (Fig. 5.6) all base pairs are intact, except the loop region exhibits a strong deviation structurally from state 1. Therefore, most angles in the loop region show a big difference between the two states. In the case of state 2 (Fig. 5.7) of the UUCG loop where the terminal base pair is disrupted and the positions of the loop nucleobases are not significantly perturbed, many torsional angles of the loop region exhibit different values than in state 1. For example, the change of the pucker angle of G7 from 5° to 149° and β angle in the same residue from 176° to -151° . On the other hand, certain angles do not show any prominent deviation to the ones from state 1, like the pucker angles in residues C3, U4, C6 and G8, the δ angles in C3, U4, U5, C6 and G8 and the ζ angles in C3, U4, U5, and C6 (Tab. 5.2). Therefore, although the main features of the UUCG loop region are still maintained in state 2, certain dynamics of the loop can still be observed for some backbone angles.

The fraction of the native contacts and hydrogen bonds of the hairpins are depicted in the Tables 5.3 and 5.4 for the obtained states at the three temperatures. For example, native contact 9 of the UUCG loop could be observed with 65% of the structures in state 1, whereas all other native contacts are existing with probabilities of more than 80%. At 331 K more than half of the states still indicate proper base pairing between the terminal residues, the rest have disrupted base pair, as can be seen in the probabilities of NC 1 and H1. Therefore, the tables of the backbone angles, the native contacts and the hydrogen bond provide a detailed description of the conformations obtained from the first two principal components.

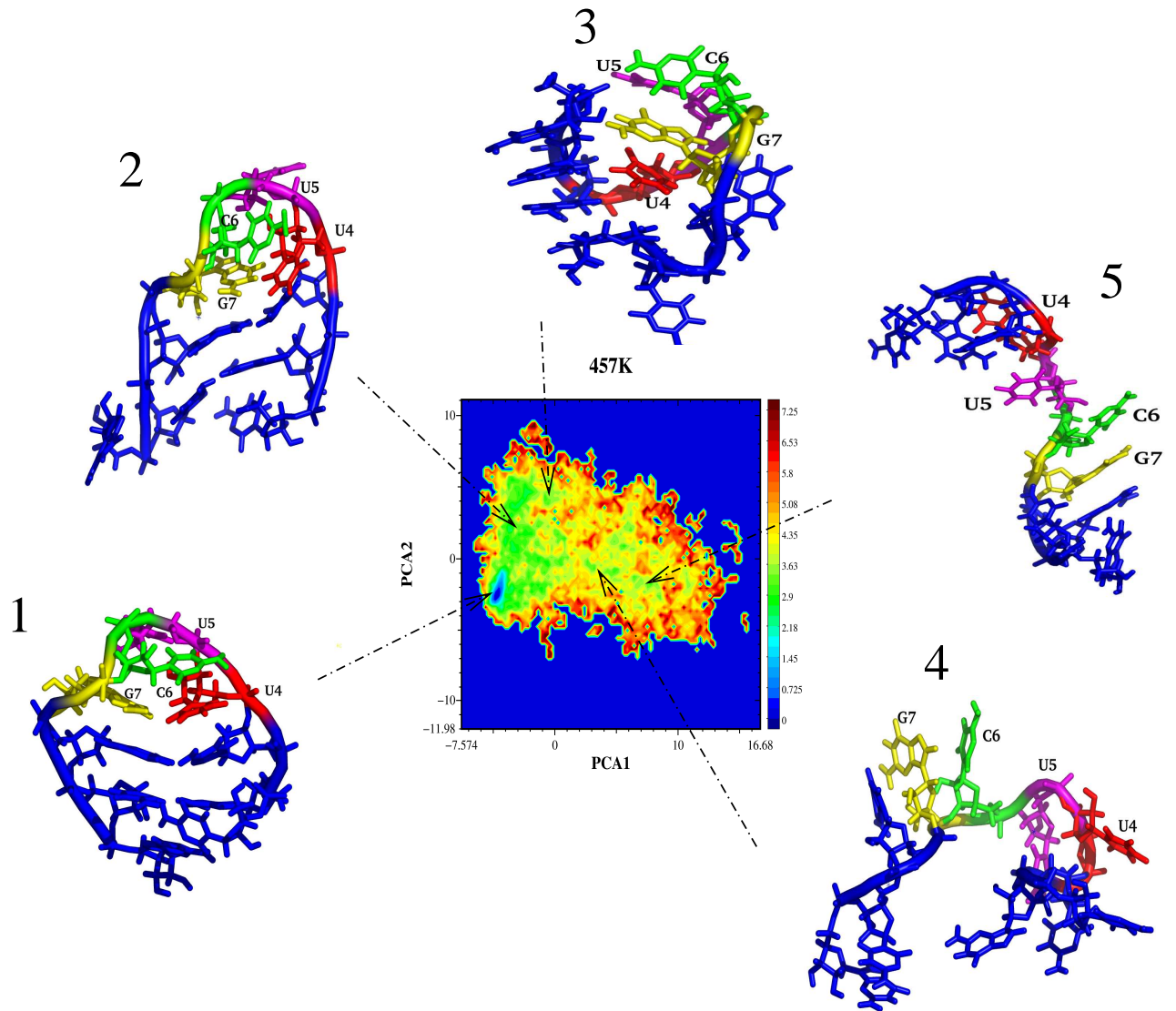


Figure 5.7: Free energy surface (in units kT) obtained from the REMD simulation at 457 K for the UUCG loop, plotted as a function of the first two PCA eigenvectors.

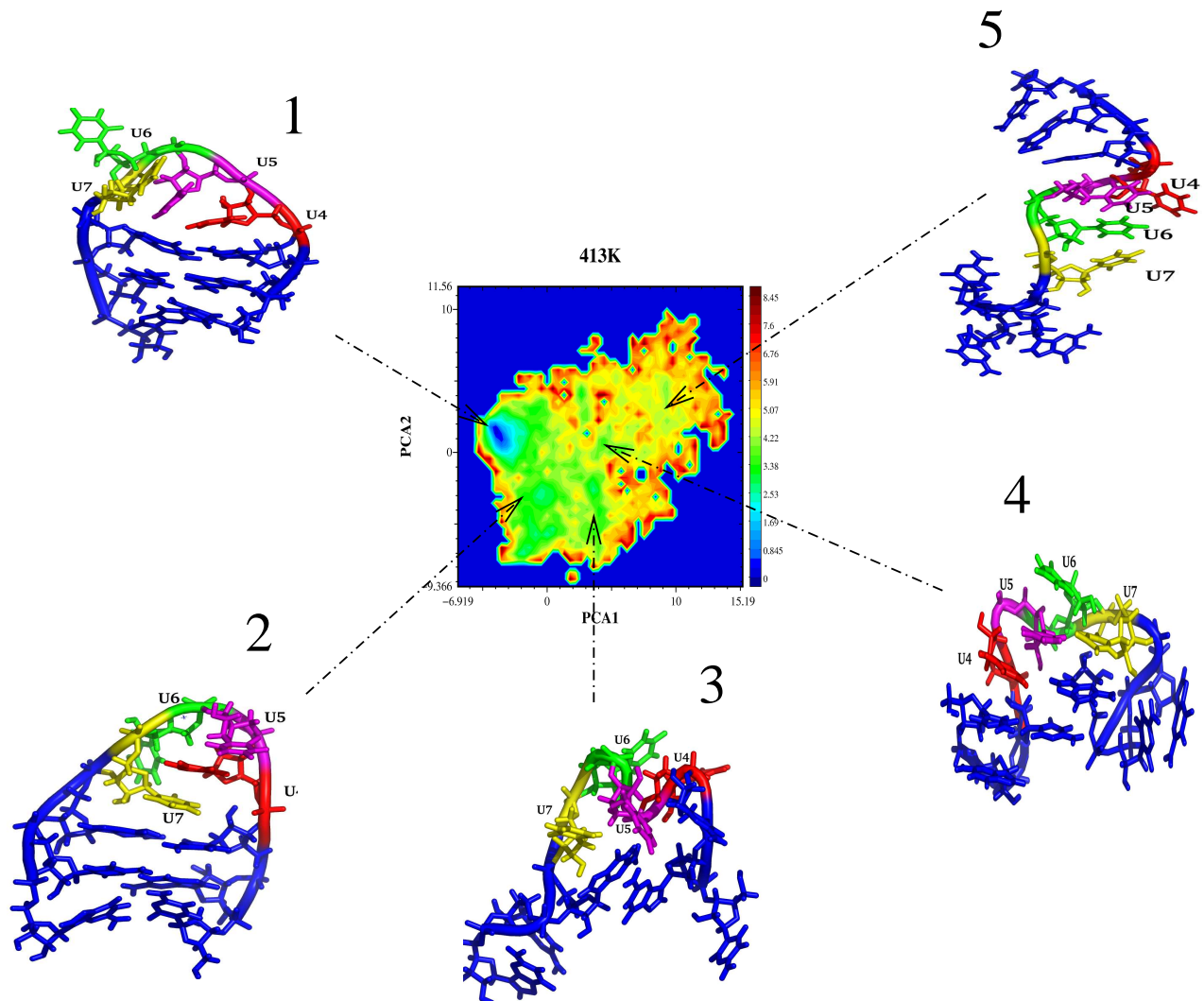


Figure 5.8: Free energy surface (in units kT) obtained from the REMD simulation at 413 K for the UUUU loop, plotted as a function of the first two PCA eigenvectors.

| torsion | State | C3 | U4 | U5 | C6 | G7 | G8 |
|------------|-------|------|------|------|------|------|------|
| α | 1 | -70 | -73 | -155 | -63 | 98 | 74 |
| | 2 | -72 | 148 | 170 | -59 | 70 | 108 |
| | 3 | -68 | 89 | 63 | 98 | 140 | -69 |
| | 4 | 63 | 76 | -112 | 132 | -177 | -52 |
| | 5 | 57 | -78 | 81 | -157 | -136 | 106 |
| β | 1 | 164 | 154 | 141 | -169 | 176 | -154 |
| | 2 | -173 | -165 | 180 | -164 | -151 | -150 |
| | 3 | 171 | -174 | 172 | 176 | 173 | -177 |
| | 4 | 149 | 155 | 66 | -164 | 178 | 147 |
| | 5 | -175 | 172 | -179 | 166 | 163 | -169 |
| γ | 1 | 64 | 71 | 50 | 46 | -172 | -153 |
| | 2 | 53 | -154 | 37 | 52 | -170 | 178 |
| | 3 | 69 | -165 | -157 | -165 | -174 | 65 |
| | 4 | 54 | -135 | -175 | 177 | 173 | 47 |
| | 5 | 164 | 66 | -167 | 47 | 81 | -172 |
| δ | 1 | 79 | 82 | 142 | 147 | 74 | 84 |
| | 2 | 100 | 86 | 162 | 131 | 156 | 85 |
| | 3 | 145 | 142 | 123 | 102 | 151 | 144 |
| | 4 | 132 | 100 | 72 | 84 | 128 | 129 |
| | 5 | 148 | 121 | 99 | 132 | 104 | 75 |
| ϵ | 1 | -133 | -164 | -94 | -90 | 156 | -158 |
| | 2 | -168 | 177 | -91 | -176 | -97 | -172 |
| | 3 | -92 | -163 | -138 | -58 | -138 | -146 |
| | 4 | -159 | -168 | -172 | -105 | -98 | -115 |
| | 5 | -96 | -121 | -146 | -74 | -140 | 150 |
| ζ | 1 | -63 | -50 | -100 | -53 | 46 | 17 |
| | 2 | -90 | -44 | -107 | -59 | 131 | -94 |
| | 3 | -91 | 77 | -143 | -87 | -67 | -63 |
| | 4 | 113 | 27 | -68 | 58 | 90 | -74 |
| | 5 | 85 | -58 | 67 | -79 | 66 | -60 |
| χ | 1 | -124 | -76 | -56 | -47 | 167 | -154 |
| | 2 | -8 | -29 | -180 | -101 | -144 | -88 |
| | 3 | -58 | 25 | 12 | -48 | 87 | -36 |
| | 4 | 176 | 117 | -85 | -127 | 133 | -89 |
| | 5 | -138 | 161 | -103 | -147 | 133 | -81 |
| Pucker | 1 | -8 | 40 | 168 | 164 | 5 | -9 |
| | 2 | -42 | 40 | -148 | 180 | 149 | 9 |
| | 3 | 165 | 150 | 125 | 43 | -172 | -170 |
| | 4 | 166 | -12 | 23 | 20 | 124 | 124 |
| | 5 | 156 | 135 | 3 | -156 | -26 | 16 |

Table 5.1: Torsion angles from different states for the UUCG loop obtained from the REMD simulation at 457 K. The states are gained by applying PCA on the trajectory.

| Torsion | State | C3 | U4 | U5 | U6 | U7 | G8 |
|------------|-------|------|------|------|------|------|------|
| α | 1 | -74 | 154 | 127 | -91 | -147 | -94 |
| | 2 | 88 | 170 | -131 | 70 | -116 | 96 |
| | 3 | -77 | 81 | 77 | 80 | -179 | -107 |
| | 4 | 76 | -86 | 136 | 45 | -72 | 120 |
| | 5 | -81 | 73 | 94 | 58 | -99 | 59 |
| β | 1 | 163 | -177 | -152 | 169 | -171 | 180 |
| | 2 | -118 | -147 | 80 | 87 | 78 | 173 |
| | 3 | 175 | 162 | -129 | -159 | 154 | 176 |
| | 4 | -81 | 166 | -168 | -156 | 177 | -178 |
| | 5 | 176 | 132 | 173 | -168 | -175 | 128 |
| γ | 1 | 58 | 169 | -176 | -164 | 5 | 45 |
| | 2 | -153 | -165 | -175 | -164 | -172 | -170 |
| | 3 | 64 | 45 | 56 | 47 | -152 | 50 |
| | 4 | -164 | 67 | 53 | -179 | 45 | 37 |
| | 5 | 64 | 65 | -170 | -170 | 69 | -179 |
| δ | 1 | 84 | 87 | 78 | 149 | 147 | 86 |
| | 2 | 72 | 82 | 144 | 67 | 128 | 103 |
| | 3 | 86 | 149 | 147 | 164 | 73 | 78 |
| | 4 | 72 | 72 | 138 | 118 | 146 | 97 |
| | 5 | 66 | 139 | 144 | 85 | 112 | 145 |
| ϵ | 1 | -157 | 179 | -152 | -63 | -98 | -153 |
| | 2 | 168 | -170 | -60 | 171 | -122 | -162 |
| | 3 | -77 | -155 | -145 | -86 | -171 | -140 |
| | 4 | -169 | -169 | -53 | -103 | -88 | -124 |
| | 5 | 166 | -156 | -56 | -161 | -68 | -83 |
| ζ | 1 | -77 | -77 | -66 | -61 | 112 | 61 |
| | 2 | -103 | -109 | 39 | 80 | 64 | -71 |
| | 3 | 136 | 92 | 82 | 61 | -51 | -84 |
| | 4 | 78 | -113 | 39 | 84 | -102 | -79 |
| | 5 | -71 | 29 | -46 | 71 | -83 | -152 |
| χ | 1 | -131 | -132 | -126 | -112 | 14 | -167 |
| | 2 | -115 | -139 | 56 | -138 | 52 | -92 |
| | 3 | -140 | -82 | -85 | -131 | 85 | -59 |
| | 4 | -131 | -158 | -121 | -68 | 48 | 164 |
| | 5 | -133 | -101 | 61 | -121 | -98 | 66 |
| Pucker | 1 | 13 | -8 | 17 | 168 | -151 | 16 |
| | 2 | 30 | 37 | -178 | 32 | 141 | -36 |
| | 3 | 41 | -170 | -162 | 171 | 47 | 15 |
| | 4 | 54 | 18 | 151 | 116 | 170 | -45 |
| | 5 | 4 | 156 | 160 | 43 | 106 | 162 |

Table 5.2: Torsion angles from different states for the UUUU loop obtained from the REMD simulation at 413 K. The states are gained by applying PCA on the trajectory.

| T | S | NC1 | NC2 | NC3 | NC4 | NC5 | NC6 | NC7 | NC8 | NC9 | NC10 | NC11 | NC12 |
|-------|---|------|------|------|------|------|------|------|------|------|------|------|------|
| 300 K | 1 | 0.93 | 1.00 | 1.00 | 0.94 | 0.87 | 0.88 | 1.00 | 0.99 | 0.65 | 1.00 | 0.99 | 0.94 |
| | 2 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 1.00 | 1.00 | 1.00 |
| | 3 | 0.89 | 1.00 | 1.00 | 0.05 | 0.31 | 0.82 | 0.14 | 0.00 | 0.04 | 0.03 | 0.55 | 0.00 |
| | 4 | 1.00 | 1.00 | 1.00 | 0.92 | 0.93 | 1.00 | 0.01 | 0.97 | 1.00 | 0.01 | 1.00 | 0.98 |
| | 5 | 0.92 | 1.00 | 1.00 | 0.00 | 0.72 | 0.52 | 0.01 | 0.00 | 0.01 | 0.00 | 0.91 | 0.00 |
| | 6 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.99 | 1.00 | 0.98 | 0.00 | 1.00 | 1.00 | 1.00 |
| 331 K | 1 | 0.90 | 1.00 | 1.00 | 0.95 | 0.84 | 0.83 | 1.00 | 0.99 | 0.65 | 1.00 | 0.95 | 0.89 |
| | 2 | 0.00 | 0.96 | 0.62 | 0.32 | 0.88 | 0.96 | 0.86 | 0.37 | 0.19 | 0.96 | 0.12 | 0.36 |
| | 3 | 0.00 | 1.00 | 0.92 | 0.89 | 0.58 | 0.73 | 0.80 | 0.91 | 0.24 | 0.87 | 0.93 | 0.91 |
| | 4 | 0.76 | 0.85 | 0.88 | 0.26 | 0.58 | 0.79 | 0.34 | 0.57 | 0.18 | 0.42 | 0.97 | 0.38 |
| | 5 | 0.97 | 0.73 | 0.91 | 0.85 | 0.94 | 0.80 | 0.28 | 0.93 | 0.77 | 0.20 | 0.97 | 0.97 |
| | 6 | 0.11 | 1.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.05 | 0.02 | 0.15 | 1.00 | 1.00 | 1.00 |
| | 7 | 0.82 | 1.00 | 1.00 | 0.00 | 0.62 | 0.64 | 0.07 | 0.00 | 0.02 | 0.01 | 0.86 | 0.00 |
| 457 K | 1 | 0.61 | 0.88 | 0.87 | 0.84 | 0.77 | 0.64 | 0.91 | 0.88 | 0.54 | 0.91 | 0.80 | 0.76 |
| | 2 | 0.24 | 0.37 | 0.50 | 0.35 | 0.58 | 0.45 | 0.34 | 0.31 | 0.41 | 0.25 | 0.34 | 0.20 |
| | 3 | 0.04 | 0.00 | 0.08 | 0.07 | 0.42 | 0.31 | 0.33 | 0.25 | 0.29 | 0.31 | 0.33 | 0.13 |
| | 4 | 0.00 | 0.00 | 0.02 | 0.05 | 0.50 | 0.53 | 0.15 | 0.25 | 0.48 | 0.26 | 0.39 | 0.17 |
| | 5 | 0.04 | 0.01 | 0.32 | 0.12 | 0.34 | 0.30 | 0.31 | 0.28 | 0.19 | 0.27 | 0.25 | 0.03 |

Table 5.3: Fraction of the native contacts (NC) (upper panel) and of the hydrogen bonds (H) (lower panel) in the UUCG loop for the states (S) at the corresponding temperature (T)

| T | S | NC1 | NC2 | NC3 | NC4 | NC5 | NC6 | NC7 | NC8 | NC9 | NC10 | |
|-------|---|------|------|------|------|------|------|------|------|------|------|------|
| 300 K | 1 | 0.93 | 1.00 | 1.00 | 0.95 | 1.00 | 1.00 | 0.89 | 0.57 | 1.00 | 0.98 | |
| | 2 | 0.29 | 1.00 | 1.00 | 0.99 | 0.64 | 0.27 | 0.42 | 0.71 | 0.58 | 0.92 | |
| | 3 | 0.08 | 0.99 | 0.99 | 0.51 | 0.99 | 1.00 | 1.00 | 0.11 | 1.00 | 0.45 | |
| | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 0.68 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 |
| | 5 | 0.81 | 1.00 | 1.00 | 0.69 | 0.98 | 0.95 | 0.55 | 0.65 | 1.00 | 0.96 | |
| | 6 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.62 | 0.50 | 0.49 | 1.00 | 1.00 | |
| | 7 | 0.93 | 1.00 | 1.00 | 1.00 | 0.50 | 0.44 | 0.80 | 0.61 | 0.51 | 0.93 | |
| | 8 | 1.00 | 1.00 | 1.00 | 1.00 | 0.04 | 0.04 | 0.81 | 0.01 | 0.12 | 1.00 | |
| 331 K | 1 | 0.91 | 1.00 | 1.00 | 0.93 | 1.00 | 0.98 | 0.90 | 0.55 | 1.00 | 0.97 | |
| | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 0.01 | 0.00 | 0.54 | 0.83 | 0.03 | 0.79 | |
| | 3 | 0.05 | 0.77 | 1.00 | 0.98 | 0.36 | 0.00 | 0.40 | 0.36 | 0.32 | 0.90 | |
| | 4 | 0.82 | 0.96 | 0.82 | 0.84 | 0.82 | 0.35 | 0.14 | 0.78 | 0.87 | 0.82 | |
| | 5 | 0.00 | 0.91 | 0.91 | 0.92 | 1.00 | 0.91 | 0.89 | 0.15 | 1.00 | 0.51 | |
| | 6 | 0.00 | 0.00 | 0.00 | 0.97 | 0.02 | 0.01 | 1.00 | 1.00 | 1.00 | 0.00 | |
| | 7 | 0.36 | 0.97 | 1.00 | 0.36 | 1.00 | 0.83 | 0.87 | 0.81 | 1.00 | 0.92 | |
| 413 K | 1 | 0.80 | 0.97 | 0.99 | 0.79 | 0.98 | 0.95 | 0.84 | 0.46 | 0.98 | 0.92 | |
| | 2 | 0.43 | 0.75 | 0.37 | 0.69 | 0.22 | 0.14 | 0.24 | 0.26 | 0.33 | 0.59 | |
| | 3 | 0.01 | 0.09 | 0.01 | 0.34 | 0.06 | 0.07 | 0.29 | 0.47 | 0.16 | 0.55 | |
| | 4 | 0.02 | 0.00 | 0.00 | 0.49 | 0.07 | 0.01 | 0.62 | 0.46 | 0.51 | 0.41 | |
| | 5 | 0.00 | 0.00 | 0.00 | 0.52 | 0.21 | 0.24 | 0.24 | 0.12 | 0.32 | 0.40 | |

Table 5.4: Fraction of the native contacts (NC) (upper panel) and of the hydrogen bonds (H) (lower panel) in the UUUU loop for the states (S) at the corresponding temperature (T)

5.4 Outlook

The conformational dynamics of two small RNA hairpins have been studied using a novel computational tool, the replica exchange molecular dynamics simulation method. We have attempted to provide insights into the dynamics of nucleic acid hairpin unfolding. The method provides atomistic framework which is confirmed relatively well by experimental data. The melting temperature in experiment could not be correctly reproduced, but ΔT_m between the UUUU and the UUCG loop is in good agreement with experiment. The high stability of the UUCG loop is caused mainly by the hydrogen bonds H10, H11 and H12 between loop residues U4-G7, the stacking interactions between C3-U4, and between G7-G8. On the other hand, the flexible UUUU loop is stabilized merely by two main stacking interactions, namely between C3-U4 and between U4-U5. In both hairpins unfolding is initiated by the terminal base pair, which is then followed by the opening of the second and the loop closing base pair. The stability of the UUCG can also be seen in the stem area, where its native contacts (except for the terminal base pair) exist with higher probability than in the case of the UUUU loop at the same temperature. Therefore, the stability in the loop region also induces a stability in the nearest neighbor residues of the stem. Future tasks will be concerned with the folding of the tetraloops. Starting from unfolded structures gained from the shown REMD simulations, we should be able to clarify whether the folding of the UUUU loop occurs later than the UUCG loop as has been observed experimentally [40].

Chapter 6

Docking Simulation

The Tat-TAR protein interaction is essential for HIV replication because the binding of Tat to TAR (Fig. 6.1) is required for activating transcription of the HIV genome [109]. Finding therapeutic agents that disrupt the Tat-TAR interaction therefore would provide a strategy for inhibiting HIV replication. Based on an NMR-derived structural model of HIV-1 TAR[110], we have used AutoDock [22] to explore the binding of small aromatic ligands as potential drugs to inhibit the Tat-TAR complex in collaboration with the Organic Chemistry group of Prof. Göbel at the University of Frankfurt am Main. In the synthesis performed [111, 112], all ligands contain a heteroaromatic element X flanked by two arginines (Arg-X-Arg), as can be seen for the last compound in Fig. 6.3. The tripeptide is negatively charged. This characteristic should enable the ligand to bind to the phosphate backbone, whereas the aromatic group should bind to the bulge cavity in an intercalating manner. The docking results shown in this work contain the ligands without the two arginines.

6.1 Theory of Docking

6.1.1 Energy evaluation

The Autodock program [22] combines a positional, orientational and conformational search engine with a grid-based method for energy evaluation. The used Autodock 3.0 [113] supports the simulated annealing search technique and also incorporates a hybrid genetic algorithm and a local search engine. Rapid energy evaluation is achieved by pre calculating atomic affinity potentials for each atom type in the substrate molecule in the manner described by Goodford [114]. In Autodock [22, 115, 113] the receptor is rigidly embedded in a three-dimensional grid and a probe atom is placed at each grid point. The energy of interaction of this single atom with the protein is assigned to the grid point. An affinity grid is calculated for

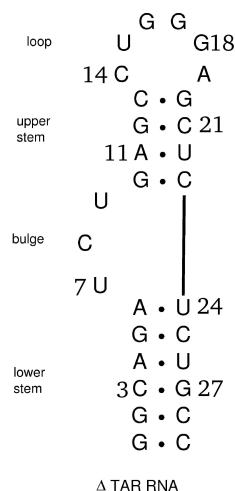


Figure 6.1: Secondary structure element of TAR-RNA

each type of atom in the ligand, typically carbon, oxygen, nitrogen and hydrogen, as well as a grid of electrostatic potential. The energetics of a particular ligand configuration is then found by tri-linear interpolation of affinity values of the eight grid points surrounding each of the atoms in the substrate. The electrostatic interaction is evaluated similarly, by interpolating the values of the electrostatic potential and multiplying by the charge on the atom.

For each grid point the pairwise interaction energy of the probe is summed over all protein atoms within a non-bonded cutoff radius and stored. For a given probe atom i and all protein atoms j , within a nonbonded cutoff distance R_{cut} , van der Waals energies are calculated using a traditional Lennard-Jones 12-6 potential. Hydrogen bonds are treated with a traditional 12-10 potential. AutoDock calculates electrostatic-interaction energy grid maps using a Coulomb potential. A measure of the unfavorable entropy of ligand binding due to the restriction of conformational degrees of freedom is added to the potential energy function. This term is proportional to the number of sp^3 bonds in the ligand, N_{tor} [116]. A last term takes the desolvation effect into account, this is based on a sigmoidal distance-dependent dielectric function [117].

The empirical scoring function implemented in Autodock 3.0 that should estimate the binding affinity of a given receptor-ligand is defined as followed.

$$\begin{aligned} \Delta G = & \Delta G_{vdw} \sum E_{vdw} + \Delta G_{hbond} \sum E_{hbond} \\ & + \Delta G_{elec} \sum E_{elec} + \Delta G_{tor} \sum E_{tor} + \Delta G_{sol} \sum E_{sol} \end{aligned} \quad (6.1)$$

where the five ΔG terms on the right-hand side are coefficients empirically determined using linear regression analysis from a set of receptor-ligand complexes with known binding constants. The summations are performed over all pairs of ligand atoms, i , and receptor atoms, j , in addition to all pairs of atoms in the ligand that

are separated by three or more bonds.

6.1.2 Search algorithm

The search method that was used in the docking simulations is termed Lamarckian genetic algorithm. The basic idea behind genetic algorithms (GA) [118] comes from natural genetics and biological evolution [119]. In molecular docking [113], the particular arrangement of a ligand and a protein can be defined by a set of values describing the translation, orientation, and conformation of the ligand with respect to the protein. These are the ligands state variables and, in the GA, each state variable corresponds to a gene. The ligand's state variable corresponds to the 'genotype', whereas its atomic coordinates correspond to the 'phenotype'. The total interaction energy of the ligand with the protein is called 'fitness'. The 'fitness' is calculated using the energy function. In the case of molecular docking the chromosome is composed of a string of real-valued genes. There are three Cartesian coordinates for the ligand translation, four variables defining a quaternion specifying the ligand orientation, and one real-value for each ligand torsion. Quaternions [120] are used in AutoDock instead of Euler angles [121] to define the orientation. The genetic algorithm begins by creating a random population of individuals, where the user defines the number of individuals in the population. For each random individual in the initial population, each of the three translation genes for x, y and z is given a uniformly distributed random value between the minimum and maximum x, y and z extends of the grid maps, respectively. The four genes defining the orientation are given a random quaternion, consisting of a random unit vector and a random rotation angle between -180° and $+180^\circ$. The second stage consists of a loop over generations, repeating until the maximum number of generations or the maximum number of energy evaluation is reached. A generation consists of five stages. The first is mapping and fitness evaluation followed by selection, crossover, mutation, and elitist selection. In the Lamarckian GA, each generation is followed by local search, being performed on a user-defined proportion of the population.

Mapping translates from each individual's genotype to its corresponding phenotype, and occurs over the entire population. This allows each individual's fitness to be evaluated. This is the sum of the interaction energy between the ligand and the protein, and the intramolecular interaction energy of the ligand. Subsequently, a selection is made to decide which individuals will reproduce. This means that individuals that have better-than-average fitness receive proportionally more offspring. Crossover and mutation are performed on random members of the population. An optimal user-defined integer parameter elitism determines how many of the top individuals automatically survive into the next generation. The vast majority of genetic algorithms mimic the major characteristics of Darwinian evolution and apply Mendelian genetics. This is illustrated on the right-hand side of Fig. 6.2. An inverse mapping exists, i.e., one that yields a genotype from a given phenotype. The local

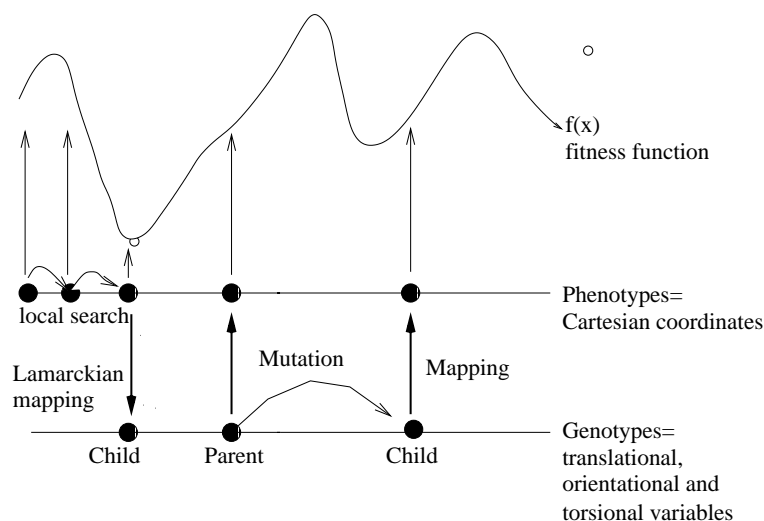


Figure 6.2: Overview of the genetic algorithm

search can be finished by replacing the individual with the result of the local search. This called the Lamarckian genetic algorithm (LGA), and is an allusion to Jean Batiste Lamarck's assertion that phenotypic characteristics acquired during an individual's lifetime can become heritable traits [122]. In the hybrid genetic algorithm and local search optimizations, the result of the local search is always used to update the fitness associated with an individual in the GA selection algorithm. Local search is shown on the left-hand side of Fig. 6.2. It is normally performed in phenotypic space and employs information about the fitness landscape. Sufficient iterations of the local search arrive at a local minimum, and an inverse mapping function is used to convert from its phenotype to its corresponding genotype. In the case of molecular docking [113], however, local search is performed by continuously converting from the genotype to the phenotype, so inverse mapping is not required. Nonetheless, this molecular variation of the genetic algorithm still qualifies as Lamarckian, because any adaptations of the ligand acquired during the local search will be inherited by the offspring.

6.2 Binding of aromatic ligands to TAR

6.2.1 Computational details

All compounds first underwent geometry optimization using quantum mechanics with HF/6-31G* basic set, then the electrostatic potential (ESP) of the molecule was obtained using Gaussian-98 [123]. The RESP (restrained electrostatic potential) fit [124] was employed to gain the atomic charges. The basic idea with electrostatic

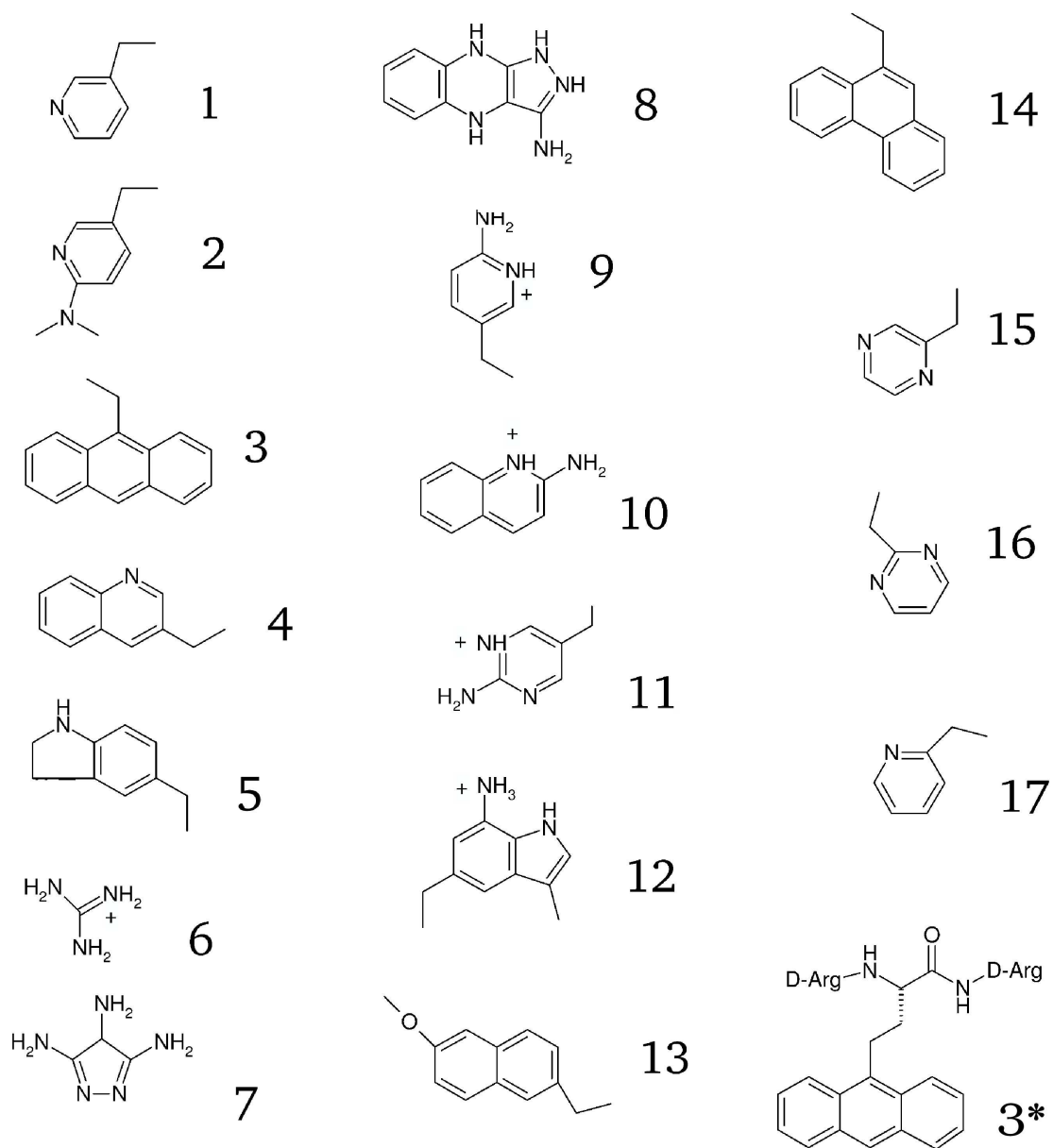


Figure 6.3: Aromatic compounds for the inhibition of the Tat-TAR complex, the last compound **3** shows the complete ligand (Arg-X-Arg).

potential fit charges is that a least squares fitting algorithm is used to derive a set of atom-centered point charges which best reproduce the electrostatic potential of the molecule. The charges for the macromolecule was taken from the Amber98 [11] force field for nucleic acids.

The grid for energy evaluation in Autodock was set in the center of the ligand with dimensions of 100 points \times 100 points \times 100 points and spacing of 0.374 Å. The ligand initial translation, quaternion, and torsion steps of 2.0 Å, 50° and 50°, respectively, were chosen with a reduction factor of 1 per cycle. The Lamarckian genetic algorithm parameters were used. Each docking simulation consisted of 30 independent runs in the case of the structure of free TAR [125] and 50 independent runs for the structure of the bound TAR [110]. The parameters used were 30 runs of a population size of 50, a termination criterion of a maximum of 27000 generations or a maximum of 1.3×10^6 energy evaluations, whichever came first. ΔG is the predicted free energy of binding plotted as a function of the cluster number (upper plot).

Analysis of the docking results is accomplished by 'structure binning'. The docked conformations are first sorted in order of increasing energy (ranking). The coordinates of the lowest-energy conformation are compared by RMSD with those of all other conformations, and grouped into the same cluster with an rmsd tolerance of 1 Å. This is repeated until all conformations have been clustered.

6.2.2 Docking results

In the calculations, we had to omit the two arginines for all 17 ligands because the docking results of the ligands containing the peptide group didn't show the expected stacking interaction of the aromatic component with any nucleotide bases in the TAR-RNA. Nevertheless, it was interesting to investigate whether these pure aromatic compounds can bind to the bulge of the TAR RNA and exhibit intercalating effects. So far, experimentally determined IC_{50} values [111, 112] are available for 9 ligands (Fig. 6.3). The IC_{50} value measures the concentration of the inhibitor required to reduce the binding of a ligand by half. The IC_{50} value is, however, not very suitable for theoretical studies as it depends on the amount of ligand available to the receptor, and this makes comparisons between data obtained under different conditions difficult. Nonetheless, we can indirectly use our docking results to interpret the IC_{50} values, especially to answer the question, what the reason could be for compound **16** (Fig. 6.3) to have the best IC_{50} , of only 2 μM and the highest *in vivo* activity, respectively [111]. Considering the fact that compound **16** comprises a structure similar to compound **1** and **15**, the last two compounds, however have much higher IC_{50} values than compound **16**, meaning less binding affinity to the TAR-RNA.

The results of the set of docking simulations with the first nine compounds are shown in Fig. 6.4, 6.5 and 6.7. All the compounds bind to the free as well as to the bound TAR mainly in bulge area. These binding conformations can be found for all

compounds either in the first cluster, which indicates the highest binding affinity or in the second cluster with the second lowest binding energy. The results we show

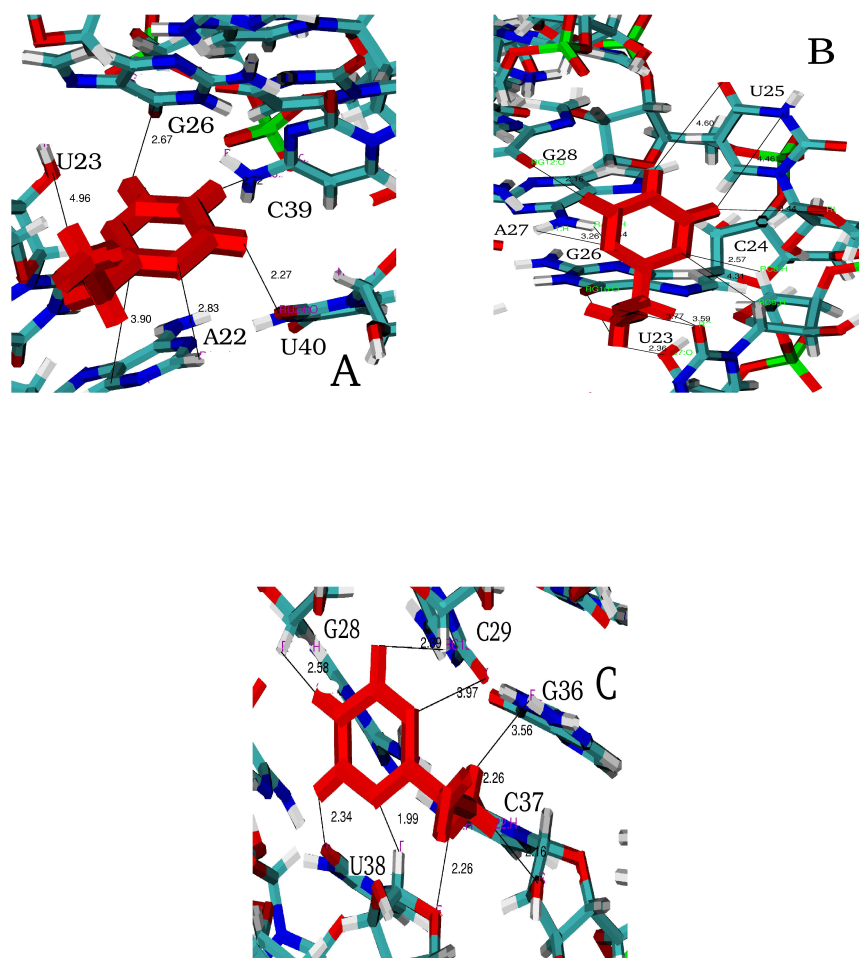


Figure 6.4: Preferred binding sites in the free TAR for component Nr. **16** . A) Binding site in the minor groove B) Binding site in the major groove and C) Binding in the upper stem

here are mainly for the binding to the the free TAR. In Fig. 6.4 the two different binding sites are shown for compound **16** , for which experimental results indicate the highest binding affinity. The binding site C corresponds to the binding in the upper stem where the aromatic ligands stabilize their binding through electrostatic interactions and hydrogen bonds. The main binding region can be found in the bulge area (Fig. 6.4A,B), where the ligands bind through stacking, hydrogen bonds and electrostatic interactions to the target. The binding to the bulge can be either from the minor groove or from the major groove as can be seen in Fig. 6.7 for compound Nr. **15** . The binding site in the minor groove of the bulge can be seen in Fig. 6.4A. Some of the ligands stack either to G26 or to the A22 which closes the bulge to the lower stem. Electrostatic interactions can be observed between the ligand and

the residues G26, C39, U40, A22 and U23. Hydrogen bonds can be observed in the case of compound **16** (Fig. 6.4A) with residue C39. The binding site in the major groove of the bulge (Fig. 6.4B) can be characterized by the interactions of the ligands with residue U25 whose base is looped out, meaning pointing into the solvent. Most of the ligands stack with their aromatic rings to the base of U25 from the major groove. Another preferred binding site of these aromatic ligands that could be monitored in our automated docking studies is accounted to the upper stem area that involves strong electrostatic interactions between the ligand and the residues G28, C29, G36, C37 and U38 which can be again seen for compound **16** in Fig. 6.4C. Evaluation of the structural quality of the results is not possible in an exact manner, since no NMR structural reference can be given. Nevertheless an NMR solution structure of acetylpromazine to the free TAR resulted in an insertion of the three membered ring between base pairs G26-C39 and A22-U40 [126]. Recently a series of compounds containing a bi-aryl heterocycle was found to induce a novel, inactive TAR conformation by stacking between the base pairs at the site of the bulge within TAR [127]. Even though these molecules don't belong to the derivatives of our ligands, our aromatic rings seem to reproduce the stacking interactions. From the upper plot in Fig. 6.5 it is clear that the ligands having the lowest free energy of binding belong to the two and three ring system, whereas the smaller ligands demonstrate ΔG values that are smaller than -7 kcal/mol. Having in mind the fact that we docked only the aromatic rings without the peptide group, and therefore treating the ligand as a rigid body without torsional degrees of freedom, we cannot directly compare our ΔG values with the experimentally determined IC_{50} values. Nonetheless if we compare the ligands with similar scaffolds, like **1**, **15**, **16** and **17**, in terms of their docked conformational populations, these smaller rings exhibit different binding specificity to the target. **16** has the lowest number of clusters in the bound as well as in the free TAR (lower plot of Fig. 6.5). Docking of **16** to the bound TAR, where 50 docked conformations are existing, the first cluster which has the lowest docked energy contains 23 similar structures, within a cluster tolerance of 1Å. In comparison to compound **15** that could only come with 17 similar structures in the first cluster. The first two clusters belong in the case of **16** to the binding in the bulge area as can be seen in Fig. 6.4A,B. The experimentally determined IC_{50} values and a recently performed NMR experiment where a titration of compound **16** to an argininamide bound TAR RNA indicated a strong inhibition by compound **16** [111]. The plot in Fig. 6.6 is shown for the ligands for which we do not have IC_{50} values. From our docking simulations we can predict for compounds **9**, **10**, **11** and **12** as the best bulge binders. They show strong binding affinities with docked energy of less than -9 kcal/mol and a small number of clusters.

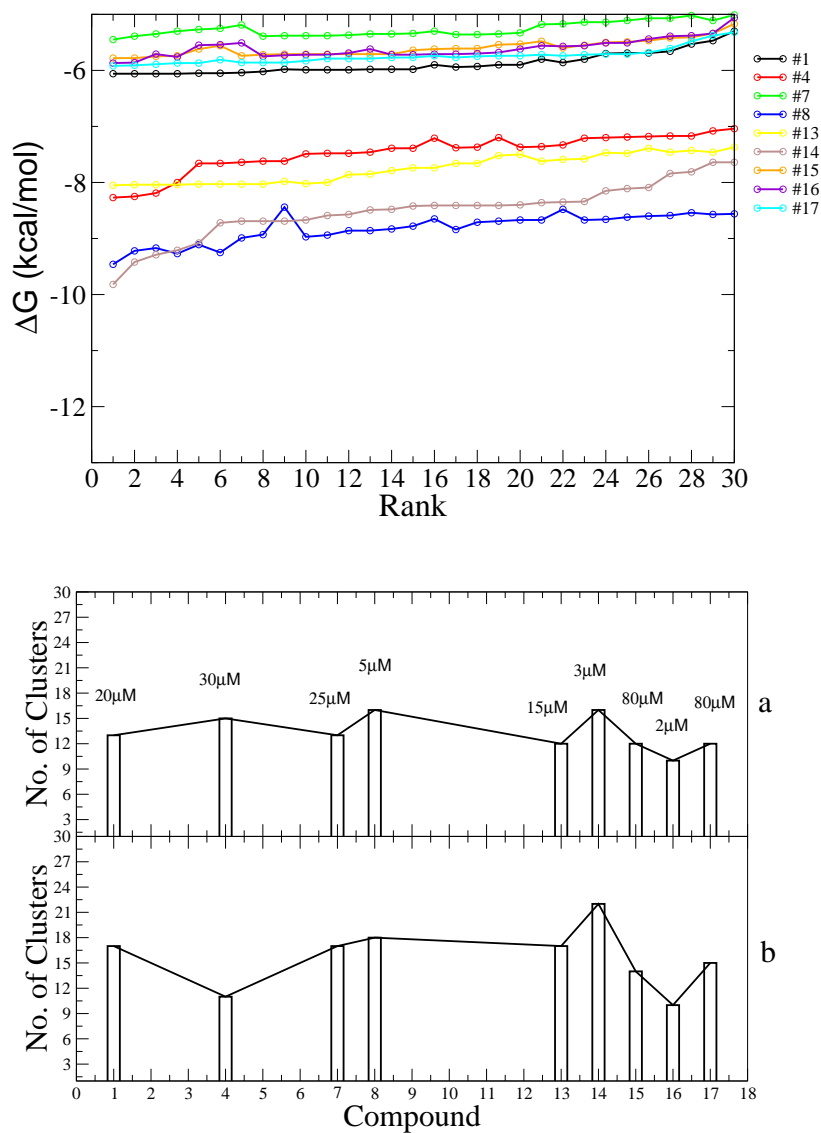


Figure 6.5: In the first plot the docked ligands are ranked in the order of increasing energy. The last plot shows the nine compounds plotted against the number of clusters. a) Results of the docking simulations of the free TAR with altogether 30 independent simulations. b) Results of the docking simulations of the bound TAR with altogether 50 independent simulations.

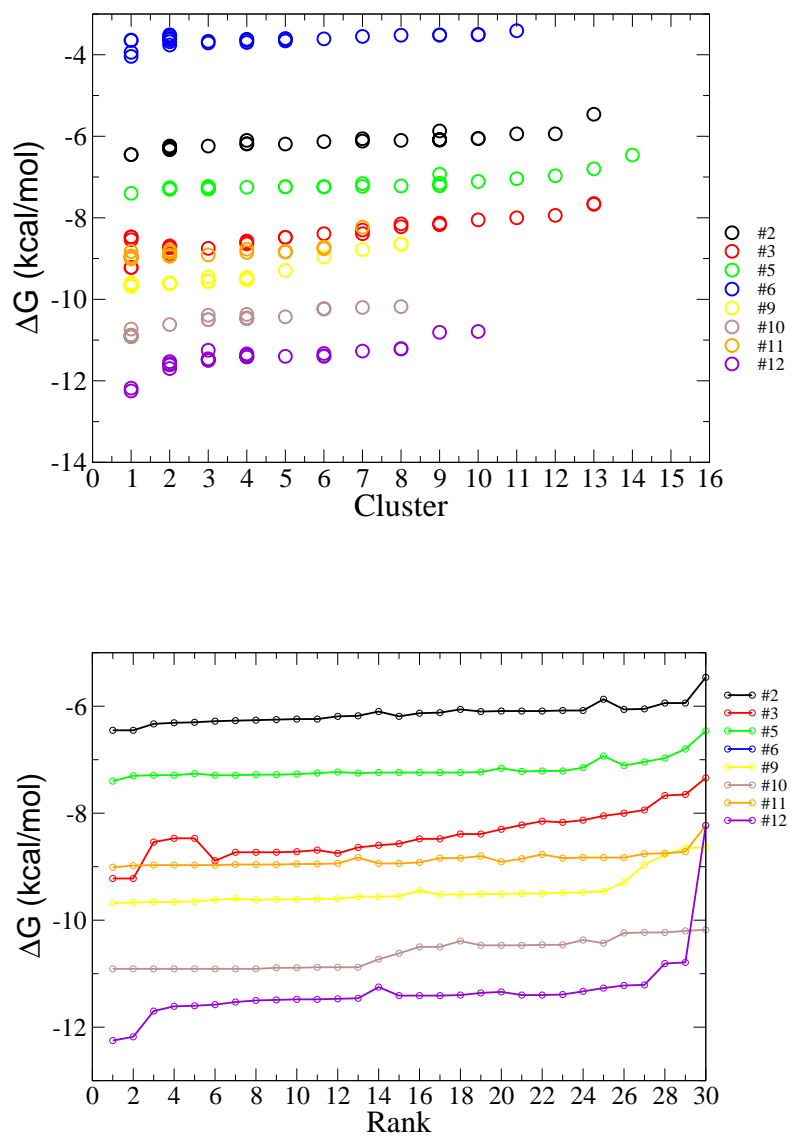


Figure 6.6: Free energy of binding plotted as a function of the 30 ranked and clustered ligands. Note no experimental data are available on these ligands yet.

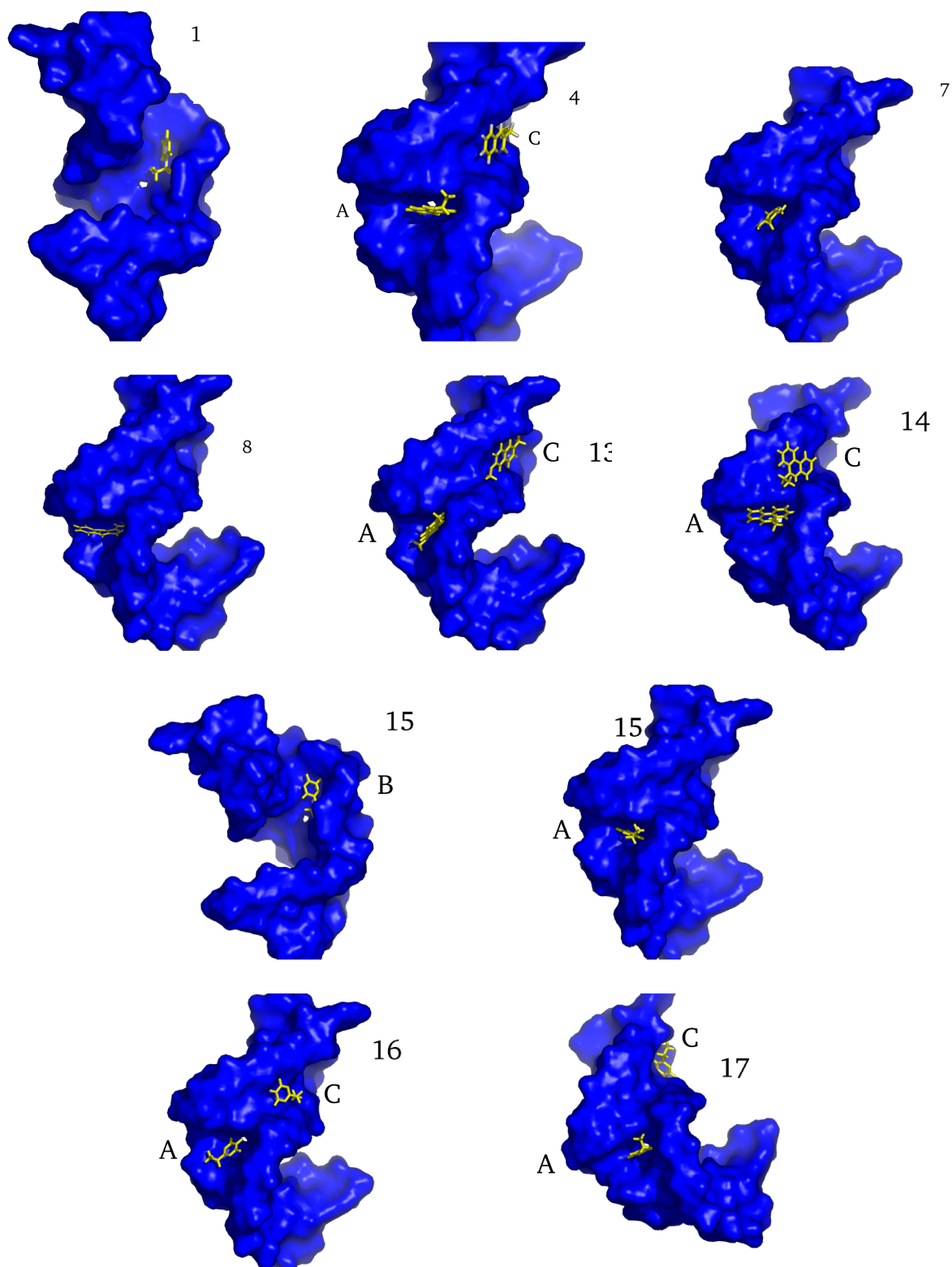


Figure 6.7: Surface models of the investigated TAR ligand complex. The aromatic ligands are indicated in yellow. Three main binding sites are obtained. Binding in the major groove of the bulge is indicated as site 'A', and the binding in the minor groove region is characterized as site 'B'. Site 'C' corresponds to the binding in a region of the upper stem

| Compound | 1 | 4 | 7 | 8 | 13 | 14 | 15 | 16 | 17 |
|----------------------|------|------|------|------|------|------|------|------|------|
| $\lg(IC_{50})$ | 4.69 | 4.52 | 4.60 | 5.30 | 4.82 | 5.52 | 4.10 | 5.69 | 4.10 |
| $\lg(K_i)$ free TAR | 3.05 | 5.49 | 3.58 | 7.27 | 5.90 | 6.34 | 5.21 | 4.22 | 5.24 |
| $\lg(K_i)$ bound TAR | 4.51 | 6.37 | 6.20 | 8.60 | 6.92 | 8.89 | 4.53 | 4.55 | 5.56 |

Table 6.1: $\lg(IC_{50})$ values determined experimentally and the logarithm of the inhibition constants $\lg(K_i)$ obtained theoretically are shown for the free and bound TAR

6.2.3 Summary

We have studied the binding sites of heteroaromatic compounds in the TAR RNA using automated docking. The results show that it is possible to use this tool to dock small rigid ligands to an RNA molecule, while large and flexible molecules are clearly problematic. The rigid planar ligands were able to stack in the bulge of TAR and should therefore inhibit the Tat-TAR interaction. Preliminary NMR results indicate that one of these aromatic molecules could successfully compete with argininamide in the binding to TAR. In nice agreement, the docking results of this ligand showed the highest specificity in the binding to TAR.

6.3 Binding and MD studies of aminoglycosides

As we have seen previously targeting RNA sequences using small molecule drugs is a topic of significant interest. Aminoglycoside antibiotics, as a class, have long been known to bind RNA. They exert their antibacterial effects at least in part by binding to specific target sites in the bacterial ribosome. Aminoglycosides also interact with a large number of other RNAs including the two essential elements of the HIV genome, Rev responsive element (RRE) and TAR [128, 129]. For example, the binding of neomycin B to HIV TAR in the minor groove leads to conformational changes in TAR, thus restricts HIV-1 transactivator protein, Tat, binding at the major TAR-RNA groove [129].

In collaboration with the Wittmann group who synthesize aminoglycoside analogs, the so-called “pseudo disaccharides”, where the sugar rings are connected either via a ‘pseudo’1-2 or a 1-6 peptide bond linkage, a prediction of the flexibility of the linkages should be made. For this task we used MD simulations and simulated three aminoglycosides in explicit solvent using AMBER6 using the same input parameters as in the case of the RNA tetraloops (see also chapter 3-5). One of the aminosugars is neamine, which is part of the antibiotic neomycin, on which extensive docking have already been done [130]. Therefore, neamine can be used as a comparison to the pseudo disaccharides of interest.

Before the MD simulations had been performed, certain force field parameters had

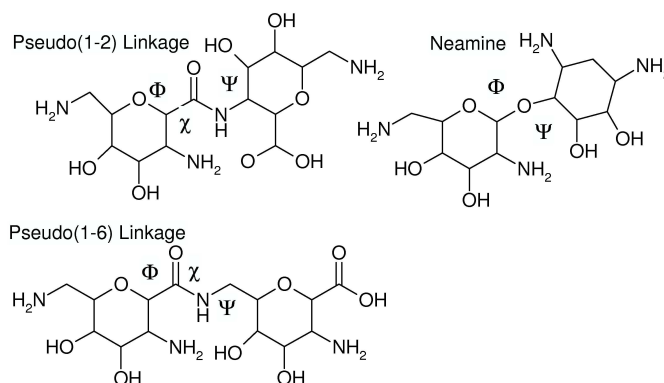


Figure 6.8: The studied aminoglycosides with the 1-2 and 1-6 linkages, compared to the neamine with only two torsional angles between the sugar rings

| Torsion | IDIVF | PK | PHASE | PN |
|-----------|-------|-------|---------|----|
| N-C-EC-H2 | 1.0 | 3.96 | 43.63 | -1 |
| N-C-EC-H2 | 1.0 | 1.63 | -65.25 | 2 |
| O-C-EC-OS | 1.0 | -0.14 | 123.67 | -1 |
| O-C-EC-OS | 1.0 | 0.72 | -194.19 | -2 |
| O-C-EC-OS | 1.0 | -0.54 | 0.0 | -3 |
| O-C-EC-OS | 1.0 | -0.1 | -10.54 | 4 |

Table 6.2: Force field parameters for torsion angles

to be calculated. For these systems we used a force field that is specialized on sugar molecules, namely the GLYCAM [131] force field. In order to treat the peptide bond properly between the sugar rings (Fig. 6.8), torsion parameters for these had to be calculated. For this purpose a rotational potential for the sugar molecules (1-6 and 1-2) with quantum mechanical calculations using GAUSSIAN [123] was derived. These parameters can be found in Table 6.2. The AMBER parameters IDIVF, PF, PN and PHASE are used to define the torsional potential energy function. Each bonded series of atoms I-J-K-L must have at least one set of dihedral parameters in the force field, The torsional energy function formula is

$$E_{\text{tors}} = (\text{PK}/\text{IDVIF})(1 + \cos(\text{PN} * \text{phi} + \text{PHASE})) \quad (6.2)$$

First we monitored the time trace of the RMSD . The highest flexibility can be seen for the 1-6 linked sugar (Figure 6.9), which is due to the twist of the two sugar rings relatively to each other. But this is somehow obvious as the 1-6 linked disaccharide has one extra bond between the rings, in comparison to the 1-2 linked molecule. The neamine is similar to the 1-2 linked sugar which comprises a relatively rigid linkage. As we can also see from the plot in the lowest panel of Figure 6.9, the dihedral angle is responsible ψ for the flexibility in the 1-6 linked disaccharide, whereas the other two angles ϕ and χ fluctuate around one single value.

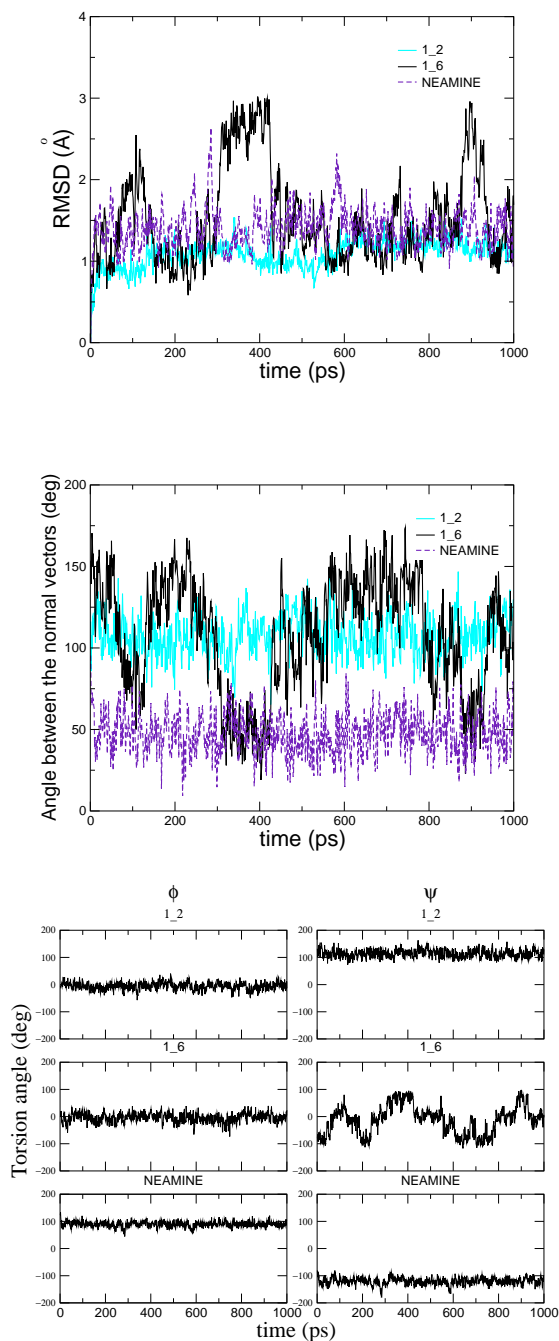


Figure 6.9: MD simulation at 300 K. Upper panel shows the RMSD of the aminoglycosides. Middle panel exhibits the angle between two vectors which are defined by two sugar atoms. Lower panel depicts the dihedral angles of the peptide bond linkage

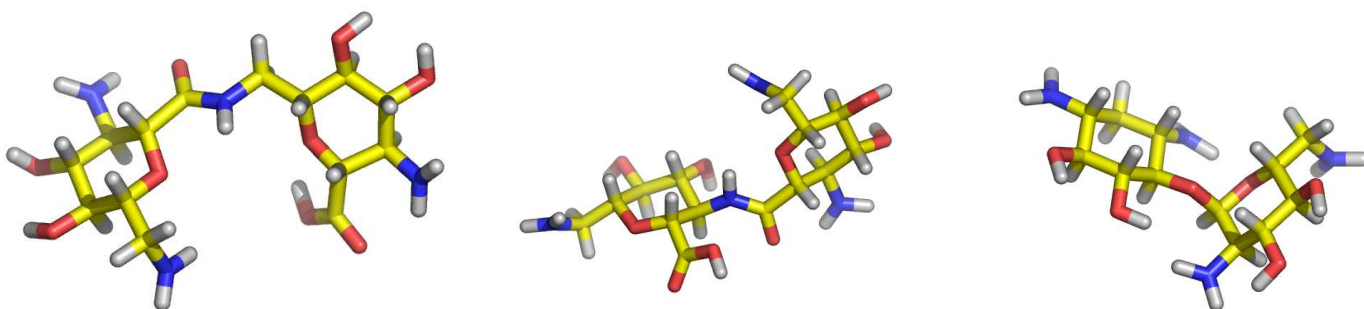


Figure 6.10: Snapshots of the aminoglycoside taken from a 1 ns MD trajectory. (Left shows the 1-6 linked, middle and right correspond to the 1-2 linked disaccharide and the neamine, respectively)

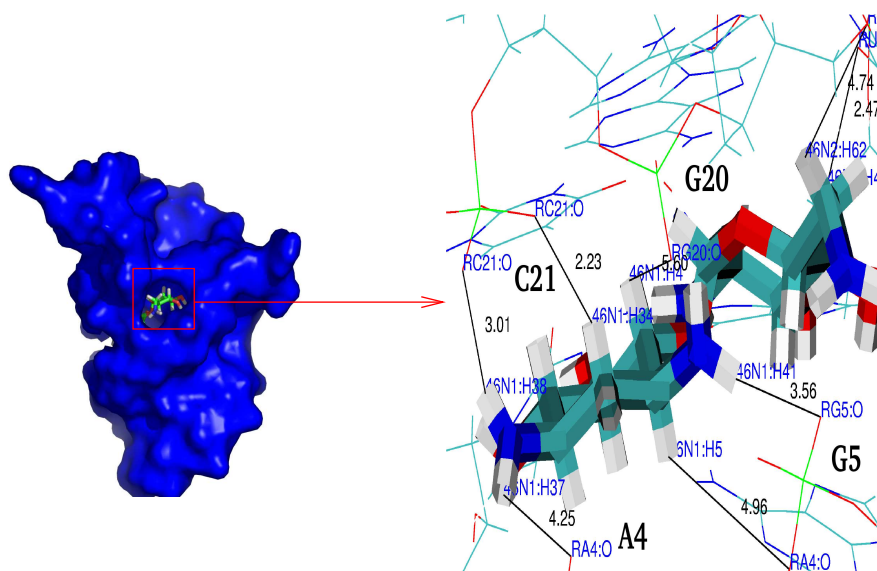


Figure 6.11: Left figure shows a surface of the TAR RNA (blue) with the neamine in the binding pocket of the bulge, right is an atomic representation of the binding site. Black lines indicate hydrogen bonds.

Docking simulations were performed for these disaccharides as in the case of the aromatic rings (see previous section). Even though we found for all three molecules a binding to the bulge, certain parts of the upper and lower stem regions were seen to be bound, too. Here we show how neamine binds to the TAR bulge area in close contact to C21, G20, A4 and G5. A free energy of binding is observed in the bulge area to be around -11.0 kcal/mol.

6.3.1 Outlook

MD simulations of these aminosugars have shown that the 1-6 linked molecule proves to be the most flexible of the three. This characteristics may be of importance for the ligand to target its proper binding site, but as studies have shown for neomycin which contains the less flexible neamine, binding to TAR RNA caused the inhibition of Tat [129]. Therefore, we can postulate that perhaps the 1-2 linked sugar as it is comparable to the neamine in its flexibility, may be a better candidate in the inhibition process.

Unfortunately, a proper binding prediction could not be made from the docking experiments, this may be due to the weakness of this method. MD simulations should therefore be more appropriate, however these can only be fulfilled when proper initial structures of the TAR-ligand complex are existing, either using better docking methods which are specific for RNA and sugar molecules or using NMR or X-ray structure.

Chapter 7

Summary

The main part of this work has been concerned with the theoretical description of the conformational dynamics and structural investigations of RNA tetraloops and RNA ligands. Three goals could be achieved in the study of the RNA tetraloops. First, the theoretical model including the force field and conformational sampling could be validated by comparing with experimental NMR data. Second, a dynamical picture of the UUCG and UUUU loops could be drawn from MD simulations. Third, a temperature dependent unfolding study of the two tetraloops exhibited higher stability for the UUCG loop in comparison to the UUUU loop in accordance with experimental observation.

Conventional MD simulations of the UUUU at various temperatures, as well as enhanced sampling simulations exhibited overall good agreement with NMR data. This is a good indication that a proper performance of the theoretical model could be obtained, including the quality of the force field and the conformational sampling. Due to the high flexibility of the UUUU loop, it was difficult to extract from NMR data alone the structural conformations for this loop. However, a combination of the experimental and theoretical data could draw a dynamical picture for the UUUU loop. Principal component analysis of a 300 K trajectory has a free energy surface with one main minimum (conformation 1), which corresponds to a conformation in which stacking interactions could be observed between U3 and U4 and between U4 and U5. Enhanced sampling by a high-temperature simulation at 350 K resulted in four conformations, where one conformation corresponds to conformation 1 and the other three conformations show U3-U4 stacking with additional stacking for residues U5 and U6 with U7.

Enhanced sampling using replica exchange simulation resulted in similar conformations as from the 350 K trajectory, however conformation 1 could not be observed. On the other hand in replica exchange simulations a conformation occurred, which was not seen in the conventional MD simulations. This conformation does not exhibit stacking between U3 and U4, instead U6 exhibits strong stacking interactions with U5 and U7.

Concerning the quality of the force field, we have found that the 300 K trajectory is in surprisingly good overall agreement with experiment, despite of its insufficient sampling. The simulation yields only a single nonmatching NOE and with the exception of the under sampled U6 residue 3J couplings and dipole-dipole cross correlated relaxation rates that well reproduce the NMR data in most cases. The overall good agreement is also reflected in the fact that the enhanced sampling trajectories, 350 K and replica exchange simulation are able to perfectly recover the backbone dihedral angles. Enhancing the sampling improve the results of residue U6 significantly, while the corresponding results for the U5 residue deteriorate, since the U5 sugar ring is too flexible compared to experiment and the converged replica exchange simulation trajectory seems to favor the C3' endo than the observed C2' endo in NMR. Nonetheless, considering the unusual amount of experimental results that are available for the UUUU tetraloop, it seems fair to say that the AMBER98 force field does an surprisingly good job in reproducing the NMR data.

It was also interesting to study stable tetraloops and compare with the flexible UUUU loop. For this reason we have presented a detailed and comprehensive MD study of the UUCG tetraloop. As in the case of the UUUU loop we first applied conventional MD simulations followed by a replica exchange simulation simulation. Simulations with two different stem residues have been performed. One is a 14 mer RNA on which NMR data from the Schwalbe group are compared to our simulations at various temperatures. Another loop is a 10 mer that is identical in the stem residues to the UUUU loop. An overall good agreement with NMR was also observed for this hairpin. The difference in these two systems lies in the second base pair of the stem, (adjacent to the loop closing base pair) where an A-U is substituted by a G-C base pair in the case of the 14 mer.

A dynamical picture of the loop region of the UUCG could be drawn by the results gained from the atomic fluctuations and the backbone angles. We monitored a higher flexibility for the phosphate backbone of the first loop nucleotide U and the last loop residue G, for the 10 mer and 14 mer, respectively. The uracil base of the second loop residue, which protrudes into the solvent exhibits high atomic fluctuations, in accordance to low S^2 values in NMR. However, it is surprising to see that the total RMSD of the trajectories depicts a stronger correlation with the phosphate backbone rather than with this flexible uracil base. The dynamics of the phosphate backbone stems from the ϵ angles of the loop base pair (U-G) which seems to maintain the flexibility of this rather stable tetraloop. It is still not clear whether the first or the last loop nucleotide is responsible for the flexibility of the system because replica exchange simulation of the 10 mer indicate a slightly higher preference for the first loop nucleotide U.

An unfolding study have been performed for the UUUU and UUCG loop using replica exchange simulation, by monitoring temperature dependent observables. Hence, insights into the dynamics of nucleic acid hairpin unfolding could be provided, that cannot currently be monitored experimentally. In this study we could not reproduce the absolute melting temperature, but the melting temper-

ature difference between the UUUU and the UUCG loop is in accordance with experiment. The higher stability of the UUCG loop is primarily caused by various hydrogen bonds between the loop base pair residues U and G and various stacking interactions. In contrast, the UUUU loop is stabilized merely by two main stacking interactions. The unfolding of the loops are both initiated by the terminal base pair, which is then followed by the simultaneous opening of the second and the loop closing base pair.

A prediction of the binding sites of heteroaromatic and aminoglycosidic compounds in the TAR RNA using automated docking have been made. Results indicate the possibility of using this tool to dock small rigid ligands to RNA targets, while large and flexible molecules seem to be clearly problematic. The planar aromatic ligands interact with the bulge of TAR in an intercalating manner. Preliminary NMR results indicate that one of these aromatic molecules could successfully compete with argininamide in the binding to TAR. In nice agreement, the docking results of this ligand resulted in the highest specificity in the binding to TAR.

Chapter 8

Zusammenfassung

Molekulardynamik (MD) Simulation ist eine wichtige und gern verwendete Computermethode zur Erforschung von molekularen Systemen in atomarer Auflösung. Es gibt keine experimentelle Methode, die in der Lage ist, eine komplette Beschreibung der dynamischen Struktur der Biomoleküle in ihren natürlichen, wässrigen Lösungen zu geben. MD-Simulation verhilft uns nicht nur zum Verständnis von Dynamik und Struktur, sondern unterstützt uns außerdem bei der Interpretation von experimentellen Beobachtungen.

MD-Simulation wurde zuerst von Alder und Wainwright im Jahre 1957 eingeführt und angewandt [4]. Dennoch wurde diese Methode erst vor 28 Jahren für biologische Makromoleküle eingesetzt [5]. Die Grundlage dafür war die Lösung der Röntgenstruktur [6] des Trypsininhibitorproteins der Rinderbauchspeicheldrüse (BPTI), das aufgrund seiner kleinen Größe und hohen Stabilität, als das “Wasserstoffmolekül” bezeichnet wurde.

MD-Simulationen werden heutzutage zur Untersuchung von großen und komplexen biologischen Systemen eingesetzt [7, 8]. Sie wurden verbessert durch die Entwicklung von schnellen und effizienten Methoden in der Behandlung von langreichweitigen elektrostatischen Wechselwirkungen [9], durch den Einsatz von schnellen Parallelrechenmaschinen und die ständige Weiterentwicklung von empirischen molekular-mechanischen Kraftfeldern [10, 11, 12].

Es dauerte mehrere Jahre bis die ersten MD-Simulationen von Nukleinsäuren durchgeführt werden konnten [13, 14, 15, 16]. Diese Untersuchungen, die im Vakuum durchgeführt wurden, zeigten deutlich die Wichtigkeit der korrekten Behandlung der Elektrostatik vor allem in stark geladenen Systemen, wie z.B. den Nukleinsäuren. Verschiedene Ansätze wurden untersucht, um das System zu verbessern und stabile DNA-Strukturen zu erhalten, z.B. wie die Reduktion der Phosphatladungen und die Zugabe von Gegenionen. Ein paar Jahre später wurde die erste MD-Simulation mit Explizitwassermolekülen und Gegenionen von einem DNA-Molekül durchgeführt und publiziert [17]. Mehrere MD-Simulationen von solvatisierten RNA Molekülen mit expliziten Gegenionen wiesen auf die Wichtigkeit

der korrekten Beschreibung der Umgebung von stark geladenen Nukleinsäuren hin [18, 19, 20, 21].

RNA spielt eine zentrale Rolle im Leben der Zelle, daher ist es unter anderem wichtig den Mechanismus der Entstehung der drei dimensional RNA Struktur und deren Funktionen in der Katalyse, Ligandenbindung und Proteinerkennung zu verstehen. Auch getrieben durch fruchtbare Kollaborationen im Sonderforschungsbereich “RNA-Liganden Wechselwirkungen” wurden RNA-Modellsysteme wie RNA Tetraloops und HIV-1 TAR untersucht.

Der Hauptteil dieser Arbeit beschäftigt sich mit der MD-Simulation von RNA Tetraloops (RNA Hairpins mit vier Nukleotiden). Dieses Hairpinmotiv ist ein ubiquitäres Strukturmotiv von RNA und DNA Oligonukleotiden [23]. RNA Tetraloops entstehen, wenn sich ein einsträngiges RNA-Molekül in einen doppelt-helikalen Stamm und eine Schleife (Loop) aus vier ungepaarten Nukleotiden faltet. Diese Schleife spielt eine Rolle in der Struktur und Funktion von RNA. Hairpins stellen Nukleationsstellen für die RNA Faltung [24] und tertiäre Erkennungsstellen für Proteine und Nukleinsäuren zur Verfügung [25, 26]. Aufgrund ihrer Bedeutung sind sie in der letzten Zeit das Ziel von zahlreichen strukturellen Untersuchungen, angefangen durch kernmagnetische Resonanz (NMR) [27, 28, 29, 30, 31, 32, 33] bis hin zu MD-Simulationen [34, 35, 36]. Diese Untersuchungen konzentrierten sich auf stabile Tetraloops, insbesondere auf UNCG, GNRA, und CUUG Loops, die mit erhöhter Häufigkeit in prokaryontischen und eukaryontischen RNAs auftreten. Die Stabilität dieser Klassen von Tetraloops ist an den hohen Schmelztemperaturen erkennbar und wird durch das zusätzliche Basenpaar zwischen dem Nukleotid 1 und 4 im Loop, verschiedene Stapelwechselwirkungen und mehrere 2'-OH-Base-Wasserstoffbrückenbindungen hervorgerufen [35, 37].

Andere Nukleotidsequenzen können ebenfalls Hairpinstrukturen bilden. So bildet zum Beispiel die Sequenz aus vier repetitiven Uridinnukleotiden, geschlossen durch einen kurzen Stamm aus drei Basenpaaren, einen stabilen Hairpin mit einer Schmelztemperatur von 51.5 °C für gUUUUc und von 60.4 °C für cUUUUg. Die Schmelztemperaturen sind daher nur 8.6 °C und 11.3 °C tiefer im Vergleich zu den stabilsten 12 mer Hairpins gUUCGc and cUUCGg [38, 39, 40]. Die Abnahme ist sogar weniger ausgeprägt für 10 mer Tetraloops (cUUCGg 72.6 °C im Vergleich zum cUUUUg 56.8 °C) [41]). Von allen Diribonukleotiden wurde die Basenstapelwechselwirkung des UpU-Dinukleotids als die schwächste erkannt [42]. Hairpins, die ein UUUU Tetraloop enthalten, wurden als flankierende Loops in RNA Quadruplexen eingeführt [43] und wurden gefunden in der funktionalen Antikodonarchitektur von tRNALys [44, 45]. Zusätzlich führte die Inkorporation des UUUU Tetraloops in Loop II des Hammerheadribozyms zu einer 2-3 fachen Erhöhung der Aktivität im Vergleich zum stabileren GCAA Tetraloop [46]. Dieses Ergebnis wird zusätzlich durch die Daten von Conaty et al. [47] unterstützt, welche zeigen, dass Polyuridine anstelle der sehr stabilen Tetraloopstruktur für eine effiziente Spaltung durch das Hammerheadribozym zuständig sind.

Die Strukturbestimmung dieser Art von flexiblen Tetraloops ist weitaus schwieriger als für stabile Tetraloops, weil NMR-Daten über ein Ensemble von Konformeren gemittelt werden. Kürzlich wurden detaillierte NMR Untersuchungen am 10 mer Tetraloop 5'-CGCUUUUGCG-3' durchgeführt. Die gemessenen Strukturparameter beinhalten NOEs, mehrere 3J Kopplungen und kreuz-korrelierte Relaxationsraten [48, 49, 50, 51] und wurden in der Gruppe von Prof. Schwalbe durchgeführt. Es wurde gezeigt, dass das Loopnukleotid U6 des UUUU Tetraloops eine konformationelle Mittelung zwischen der N- und S-Konformation der Zuckerpseudorotation aufweist [51], was auf eine signifikante konformationelle Dynamik hindeutet. NMR strukturelle Bestimmung des stabileren 5'-GGCACUUCGGUGCC-3' Tetraloop [33] und dynamisches Verständnis der Basen und Ribosen durch NMR ^{13}C Spinrelaxationsexperimenten wurde auch von der Schwalbe Gruppe geliefert [52]. Diese NMR Resultate für den UUCG-Loop zeigen die größte Dynamik für den zweiten Nucleobasen- und Riboserest an. Diese detaillierte strukturelle NMR Beschreibung des kleinen RNA Loops kann somit als Anhaltspunkt für die Untersuchung der Gültigkeit und Genauigkeit der theoretischen Beschreibung der konformationellen Dynamik durch MD-Simulationen dienen.

Die vorliegende Arbeit verfolgt daher drei Ziele. Erstens können wir durch den Vergleich von experimentellen und theoretischen Daten die Qualität und Güte des theoretischen Modells, bezüglich des Kraftfeldes, des konformationellen Samplings und der theoretischen Berechnung von NMR-Daten durch Einsatz von Näherungsformeln wie die Karplusrelation, prüfen. Zweitens, kann durch die Verknüpfung von NMR und MD Resultaten eine dynamische Struktur des UUUU und UUCG Tetraloops gewonnen werden, die durch die konventionellen NMR bestimmten Strukturen insbesondere für den flexiblen UUUU-Loop nicht abgeleitet werden konnte. Im Falle des UUCG-Loops, für den eine NMR Beschreibung der Rückgratdynamik noch fehlt, sollen die MD-Daten diese unterstützen. Drittens soll eine temperaturabhängige Untersuchung des 10 mers 5'-CGCUUUUGCG-3' und 5'-CGCUUCGGCG-3 durch die Verwendung der Replika austauschsimulation durchgeführt werden. Somit kann eine temperaturabhängige Entfaltung beider Loops verfolgt und ein dynamischer und struktureller Vergleich gemacht werden. Verschiedene Trajektorien von konventionellen MD-Simulationen zeigten jedoch unzureichendes Sampling auf, weswegen wir den Einsatz von verstärktem Sampling wie die Replika austauschsimulation benötigten.

Eine detaillierte und umfassende MD Untersuchung der Struktur und Dynamik des UUUU- und UUCG-Loops wurden durch den Einsatz von mehreren MD-Simulationen, sowie von verstärkten Samplingmethoden wie LES und REMD bewerkstelligt. Bezüglich der Qualität des Kraftfeldes wurde eine überraschend gute Übereinstimmung mit den experimentellen Daten für den UUUU-Loop gefunden, trotz des unzureichenden Samplings. Die Simulation zeigte einen einzigen NOE Wert, der nicht in Übereinstimmung mit den NMR-Daten gebracht werden konnte und mit der Ausnahme des "under-sampled" U6 Restes. 3J Kopplungen und

Dipol-Dipol kreuz-korrelierte Relaxationsraten konnten im Einklang mit NMR-Daten reproduziert werden. Die gute Übereinstimmung zeigte sich weiterhin in der Tatsache, dass die Trajektorien in der Lage waren, die Rückgrattorsionswinkel zu reproduzieren. Die Erhöhung des Samplings durch Einsatz von 350 K, LES and REMD Simulationen resultierte in einer besseren Übereinstimmung des flexiblen U6 Restes mit NMR-Daten, wohingegen die Resultate des U5 Restes abweichen, da der Zuckerring zu flexibel ist im Vergleich zum Experiment, und die REMD Trajektorie scheint eher die C3'-endo Konformation im Gegensatz zu der im NMR beobachteten C2'-endo Konformation zu bevorzugen. Der Einsatz von experimentellen und theoretischen Daten erlaubte eine dynamische Beschreibung des Hairpins. Dies wurde deutlicher durch die Resultate der Hauptkomponentenanalyse (principal component analysis) für den flexiblen UUUU-Loop. Mehrere Konformationen wurden gefunden, die durch reine Basenstapelwechselwirkungen ihre Stabilität erhalten im Gegensatz zum stabilen UUCG-Loop, der zusätzlich Wasserstoffbrückenbindungen aufweist. Im UUUU-Loop bestehen Stapelwechselwirkungen zwischen C3 und U4, U4 und U5 und in einigen Konformationen treten Stackinginteraktionen zwischen U5 und U7, und U6 und U7 auf.

Umfassende MD-Simulationen wurden auch für den stabilen UUCG-Loop durchgeführt. Der Zweck dieser Untersuchungen besteht darin, die Dynamik des UUCG-Loops besser zu verstehen, strukturelle und dynamische Unterschiede zum UUUU-Loop zu charakterisieren und wie im Falle des UUUU-Loops die Validität und Genauigkeit des theoretischen Modells durch den Vergleich mit NMR Daten zu prüfen. In diesem Zusammenhang wurden zwei UUCG Tetraloops unterschiedlicher Sequenzen untersucht. Ein 10 mer mit der folgenden Sequenz 5'-CGCUUCGGCG-3' ist im Stamm identisch mit dem des UUUU-Loops, um die Vergleichbarkeit zu garantieren. Ein UUCG-Loop mit der Sequenz 5'-GGCACUUCGGUGCC-3', wurde von der Schwalbe-Gruppe mittels NMR untersucht [33, 52]. Diese NMR-Daten konnten bisher dynamische Informationen über Basen und Ribosen liefern. Mit Hilfe der MD-Simulation soll nun auch eine Aussage über die Dynamik im Rückgrat gemacht werden.

Zu diesem Zweck wurden MD Trajektorien unterschiedlicher Temperaturen untersucht. Atomare Fluktuationen zeigten die höchste Flexibilität für die Basen und den Ribosering des zweiten flexiblen U-Nukleotids im Loop in guter Übereinstimmung mit kleinen S^2 Werten im Experiment. Der Zeitverlauf der RMSD zeigte eine Dynamik im Backbone im Bereich des ersten (U) und letzten (G) Loopnukleotids. Diese Dynamik ist auf den ϵ Winkel zurückzuführen. Es ist überraschend, dass das flexible U-Nukleotid nicht wesentlich zum gesamten Verlauf des RMSD beiträgt. Nichtsdestotrotz muss noch genauer geklärt werden, zu welchem Anteil das erste und letzte Loopnukleotid an der Dynamik beteiligt sind. Bisher zeigten die Trajektorien des 10 mers eine Dynamik verstärkt im Backbone des ersten Loopnukleotids, im Gegensatz zu den Trajektorien des 14 mers, die lediglich im Bereich des letzten Loopnukleotids starke Backbonedynamik aufweisen. Konvergierte REMD-Simulationen des 14 mers zeigten eine starke Präferenz für das erste Loopnukleotid,

obwohl das letzte Loopukleotid nicht unwesentlich an der Dynamik beteiligt ist. Die REMD-Simulationen des 10 mer UUUU- bzw. UUCG-Loops wurden verwendet, um eine temperaturabhängige Entfaltung dieser Hairpins näher zu studieren. Hieraus konnten strukturelle und dynamische Einblicke in die Entfaltung der Hairpins gewonnen werden, die experimentellen Methoden bisher verborgen blieben. Interessant ist die Tatsache, dass die Schmelztemperaturdifferenz ΔT_m in guter Übereinstimmung mit experimentellen Beobachtungen reproduziert werden konnte. Die Entfaltung wurde in beiden Hairpins durch das terminale Basenpaar eingeleitet, wobei sich das zweite und das loop-schliessende Basenpaar gleichzeitig öffnen, gefolgt vom ersten bzw. letzten Nukleotid im Loopbereich.

Im letzten Teil dieser Arbeit wurden Untersuchungen an RNA-bindende Liganden durchgeführt. RNA bildet wohl definierte drei dimensionale Strukturen mit Bindungstaschen, ähnlich wie in Proteinen. Eine Vorhersage der Bindungsstellen von heteroaromatischen und aminoglycosidischen Substanzen an TAR RNA wurde durch automatisiertes Docking gemacht. Die Ergebnisse der Dockingsimulationen zeigen die Anwendbarkeit dieser Methode für rigide Liganden, während sie für flexible Moleküle nicht unumstritten ist. Die planaren aromatischen Liganden wechselwirken im Bulgebereich der TAR RNA über Interkalationen. Erste NMR Resultate [111] weisen auf eine erfolgreiche Bindung eines dieser Liganden an TAR hin. In guter Übereinstimmung zeigten die Dockingresultate für diesen Ligand die höchste Bindungsspezifität in der Bindung an TAR.

Bibliography

- [1] Ennifar, E., Nikulin, A., Tishchenko, S., Serganov, A., Nevskaya, N., Garber, M., Ehresmann, B., Ehresmann, C., Nikonov, S., Dumas, P. The crystal structure of UUCG tetraloop. *J. Mol. Biol.* 304:35, 2000.
- [2] Fürtig, B. Diploma Thesis, Johann Wolfgang Goethe-Universität, Frankfurt, 2003.
- [3] Allain, F., Varani, G. Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol.* 250:333, 1995.
- [4] Alder, B., Wainwright, T. Phase transition of a hard sphere system. *J. Chem. Phys.* 27:1208, 1957.
- [5] McCammon, J., Gellin, B., Karplus, M. Dynamics of folded protein. *Nature (London)* 267:585, 1977.
- [6] Deisenhofer, J., Steigermann, W. Crystallographic refinement and the structure of the bovine pancreatic trypsin inhibitor at 1.5 Å resolution. *Acta Cryst.* 238, 1975.
- [7] de Groot, B., H. Grubmüller. Water permeation across biological membranes. *Science* 294:2353, 2001.
- [8] Roux, B. Computational studies of the gramicidin channel. *Acc. Chem. Res.* 35:366, 2002.
- [9] Essmann, U., Perera, L., Berkowitz, M., Darden, T., Lee, H., Pedersen, L. A smooth particle mesh EWALD method. *J. Chem. Phys.* 103:8577, 1995.
- [10] Langley, D. Molecular dynamics simulations of environment and sequence dependent DNA conformation: the development of the BMS nucleic acid force field and the comparison with empirical force field. *J. Biomol. Struct. Dyn.* 16:487, 1998.
- [11] Cheatham, T., Cieplak, P., Kollman, P. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* 16:845, 1999.

- [12] Foloppe, N., MacKerell, A. All-atom empirical force field for nucleic acids. 1) Parameter optimization based on small molecule macromolecular target data. *J. Comput. Chem.* 21:86, 2000.
- [13] Levitt, M. Computer simulation of DNA double helix dynamics. *Cold Spring Harbor Symp. Quantum Biol.* 47:251, 1983.
- [14] Tidor, B., Irikura, K., Brooks, B., Karplus, M. Dynamics of DNA oligomers. *J. Biomol. Struct. Dyn.* 1:231, 1983.
- [15] Prabhakaran, M., Harvey, S., Mao, B., McCammon, J. Molecular dynamics of phenylalanine transfer RNA. *J. Biomol. Struct. Dyn.* 1:537, 1983.
- [16] Nilsson, L., Karplus, M. Molecular dynamics simulation of anticodon arm of phenylalanine transfer RNA. *NATO Asi Ser. Ser. C.)* 110:151, 1986.
- [17] Seibel, G., Singh, U., Kollman, P. A molecular dynamics simulation of double helical B-DNA including counterions and water. *Proc. Natl. Acad. Sci. USA* 82:6537, 1985.
- [18] Lee, H., Darden, T., Pedersen, L. Accurate crystal molecular dynamics simulations using particle-mesh-ewald:RNA dinucleotides - ApU and GpC. *Chem. Phys. Lett.* 243:229, 1995.
- [19] Zichi, D. Molecular dynamics of RNA with the OPLS force field. Aqueous simulation of a hairpin containing a tetranucleotide loop. *J. Am. Chem. Soc.* 117:2957, 1995.
- [20] Auffinger, P., Westhof, E. RNA hydration: Three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA^{Asp} anticodon hairpin. *J. Mol. Biol.* 269:326, 1997.
- [21] Auffinger, P., Louis-May, S., Westhof, E. Molecular dynamics simulations of solvated yeast tRNA^{Asp}. *Biophys. J.* 76:50, 1999.
- [22] Goodsell, D., Olson, A. Automated docking of substrates to proteins by simulated annealing. *Prot. Struct. Func. Genet.* 8:195, 1990.
- [23] Woese, C., Winker, S., Gutell, R. Architecture of ribosomal RNA: Constrains on the RNA tetraloops. *Proc. Natl. Acad. Sci. USA* 87:8467, 1990.
- [24] Uhlenbeck, O. Tetraloops and RNA folding. *Nature (London)* 346:613, 1990.
- [25] Legault, P., Li, J., Mogridge, J., Kay, L., Greenblatt, J. NMR Structure of the Bacteriophage γ N Peptide/*boxb* RNA Complex: Recognition of a GNRA Fold by an Arginine-Rich Motif. *CELL* 93:289, 1998.

- [26] Jagath, J., Matassova, N., Leeuw, E. D., Warnecke, J., Lentzen, G., Rodnina, M., Luirink, J., Wintermeyer, W. Important role of the tetraloop region of the 4.5S RNA in SRP binding to its receptor FtsY. *RNA* 7 2:293, 2001.
- [27] Heus, H., Pardi, A. Structural features that give rise to the unusual stability of RNA hairpins. *Science* 253:191, 1991.
- [28] Cheong, C., G. Varani, G., Jr., I. T. Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Science* 346:680, 1990.
- [29] Jucker, F., Pardi, A. Solution structure of the CUUG hairpin loop: A novel RNA tetraloop motif. *Biochemistry* 34:14416, 1995.
- [30] Jucker, F., Heus, H. A., Yip, P., E, H. M., Pardi, A. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.* 264:968, 1996.
- [31] Butcher, S., Dieckmann, T., Feigon, J. Solution structure of a GAAA tetraloop receptor RNA. *EMBO J.* 16:7490, 1997.
- [32] Du, Z., Yu, J., Andino, R., James, T. Extending the family of UNCG-like tetraloop motifs: NMR structure of a CACG tetraloop from coxsackievirus B3. *Biochemistry* 42:4373, 2003.
- [33] B. Fürtig, Richter, C., Bermel, W., Schwalbe, H. New NMR experiments for RNA nucleobase resonance assignment and chemical shift analysis of an RNA UUCG tetraloop. *J. Biomol. NMR* 28:69, 2004.
- [34] Williams, J., Hall, K. Experimental and theoretical studies of the effects of deoxyribose substitutions on the stability of the UUCG tetraloop. *J. Mol. Biol.* 297:251, 2000.
- [35] Williams, J., Hall, K. Experimental and computational studies of the G[UUCG]C RNA tetraloop. *J. Mol. Biol.* 297:1045, 2000.
- [36] Miller, J., Kollman, P. 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, 1997.
- [37] Zacharias, M. Simulation of the structure and dynamics of nonhelical RNA motifs. *Curr. Opin. Struct. Biol* 10:311, 2000.
- [38] Antao, V. P., Lai, S., Tinoco, I. J. Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucl. Acids Res.* 19:5901, 1991.
- [39] Procter, D., Schaak, J., Bevilacqua, J., Falzone, C., Bevilacqua, P. Isolation and characterization of a family of stable RNA tetraloops with the motif YNMG that participate in tertiary interactions. *Biochemistry* 41:12062, 2002.

- [40] Proctor, D., Ma, H., Kierzek, E., Kierzek, R., Gruebele, M., Bevilacqua, P. Folding thermodynamics and kinetics of YNMG RNA hairpins: specific incorporation of 8-bromoguanosine leads to stabilization by enhancement of the folding rate. *Biochemistry* 43:14004, 2004.
- [41] Wörner, K. PhD Thesis, Johann Wolfgang Goethe-Universität, Frankfurt, 1997.
- [42] Norberg, J., Nilsson, L. Stacking free energy profiles for all 16 natural ribonucleoside monophosphates in aqueous solution. *J. Am. Chem. Soc.* 117:10832, 1995.
- [43] Liu, H., Matsugami, A., Katahira, M., Uesugi, S. A dimeric RNA quadruplex architecture comprised of two G:G(:A):G:G(:A) hexads, G:G:G:G tetrads and UUUU loops. *J. Mol. Biol.* 322:955, 2002.
- [44] Stuart, J., Gdaniec, Z., Guenther, R., Marszalek, M., Sochacka, E., Malkiewicz, A., Agris, P. Functional anticodon architecture of human tRNA^{Lys,3} includes disruption of intraloop hydrogen bonding by the naturally occurring amino acid modification, t6a. *Biochemistry* 39:13396, 2000.
- [45] Durant, P., Davis, D. Stabilization of the anticodon stem-loop of tRNA^{Lys,3} by an A-C base-pair and by pseudouridine. *J. Mol. Biol.* 285:115, 1999.
- [46] Persson, T., Hartmann, R., Eckstein, F. Selection of hammerhead ribozyme variants with low Mg²⁺ requirement: Importance of stem-loop ii. *ChemBioChem* 3:1066, 2002.
- [47] Conaty, J., Hendry, P., Lockett, T. Selected classes of minimised hammerhead ribozyme have very high cleavage rates at low Mg²⁺ concentration. *Nucl. Acids Res.* 27:24, 1999.
- [48] Richter, C., Reif, B., Wörner, K., Quant, S., Marino, P., Engels, J., Griesinger, C., Schwalbe, H. A new experiment for the measurement of ⁿJ(C,P) coupling constants including ³J(C4'_i,P_i) and ³J(C4'_i,P_i+1) in oligonucleotides. *J. Biomol. NMR* 12:223, 1998.
- [49] Felli, I., Richter, C., Griesinger, C., Schwalbe, H. Determination of RNA sugar pucker mode from cross-correlated relaxation in solution NMR spectroscopy. *J. Am. Chem. Soc.* 121:1956, 1999.
- [50] Richter, C., Reif, B., Griesinger, C., Schwalbe, H. NMR spectroscopic determination of angles α and ζ in RNA from CH-dipolar coupling, P-CSA cross-correlated relaxation. *J. Am. Chem. Soc.* 122:12728, 2000.

- [51] Duchardt, E., Richter, C., Reif, B., Glaser, S., Engels, J., Griesinger, C., Schwalbe, H. Measurement of $^2J(\text{H,C})$ - and $^3J(\text{H,C})$ -coupling constants by α/β selective HC(C)H-TOCSY. *J. Biomol. NMR* 21:117, 2001.
- [52] Duchardt, E. PhD Thesis, MIT, USA, 2004.
- [53] Frenkel, D., Smit, B. Understanding molecular simulations. San Diego, California, USA: Academic press. 2002.
- [54] Berendsen, H., van Gunsteren, W., Zwinderman, H., Guertsen, R. Simulations of proteins in water. *Ann. N. Y. Acad. Sci.* 482:269, 1986.
- [55] Verlet, L. Computer "experiments" on classical fluids. I. Thermodynamical properties of lennard-jones molecules. *Phys. Rev.* 159:98, 1967.
- [56] Verlet, L. Computer "experiments" on classical fluids. ii. equilibrium correlation functions. *Phys. Rev.* 165:201, 1968.
- [57] van Gunsteren, W., Berendsen, H. A leap frog algorithm for stochastic dynamics. *Mol. Simulation* 1:173, 1988.
- [58] Koplín, J., Mu, Y., Richter, C., Schwalbe, H., Stock, G. Structure and dynamics of an RNA tetraloop: A joint molecular-dynamics and NMR study. *subm. to Structure*.
- [59] Cheatham, T., Young, M. Molecular dynamics simulations of nucleic acids: successes, limitations and promise. *Biopolymers* 56:232, 2001.
- [60] Reddy, S., Leclerc, F., Karplus, M. DNA polymorphism: a comparison of force fields for nucleic acids. *Biophys. J.* 84:1421, 2003.
- [61] Miranker, A., Karplus, M. Functionality maps of binding sites: a multi copy simultaneous search method. *Prot. Struct. Func. Genet.* 11:29, 1991.
- [62] Stultz, C., Karplus, M. MCSS functionality maps for flexible protein. *Prot. Struct. Func. Genet.* 37:512, 1999.
- [63] H. Grubmüller. Predicting slow structural transitions in macromolecular systems-conformational flooding. *Phys. Rev. E* 52:2893, 1995.
- [64] Zhou, R., Berne, B., Germain, R. The free energy landscape of β hairpin folding in explicit water. *Proc. Natl. Acad. Sci. USA* 98:14931, 2001.
- [65] Berendsen, H., van Gunsteren, W. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684, 1984.
- [66] Allen, M., Tildesly, D. Computer simulations of liquids. New York, USA: Oxford university press. 1987.

- [67] Roitberg, A., Elber, R. Modeling side chains in peptides and proteins: Application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *J. Chem. Phys.* 95:9277, 1991.
- [68] Simmerling, C., Miller, J., Kollman, P. Combined locally enhanced sampling and Particle Mesh Ewald as a strategy to locate the experimental structure of a nonhelical nucleic acid. *J. Am. Chem. Soc.* 120:7149, 1998.
- [69] Elber, R., Karplus, M. Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc.* 112:9161, 1990.
- [70] Sanbonmatsu, K. Y., Garcia, A. E. Structure of met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins* 46:225, 2002.
- [71] Sugita, Y., Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314:141, 1999.
- [72] Mitsutake, A., Sugita, Y., Okamoto, Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* 60:96, 2001.
- [73] Garcia, A. E., Sanbonmatsu, K. Y. Exploring the energy landscape of a β hairpin in explicit solvent. *Prot. Struct. Func. Genet.* 42:345, 2001.
- [74] Garcia, A. E., Onuchic, J. N. Folding a protein in a computer: An atomic description of the folding/unfolding of protein A. *Proc. Natl. Acad. Sci. USA* 100:13898, 2003.
- [75] Ichiye, T., Karplus, M. Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* 11:205, 1991.
- [76] Garcia, A. E. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* 68:2696, 1992.
- [77] Amadei, A., Linssen, A. B. M., Berendsen, H. J. C. Essential dynamics of proteins. *Proteins* 17:412, 1993.
- [78] Hayward, S., Kitao, A., Hirata, F., Go, N. Effect of solvent on collective motions in globular proteins. *J. Mol. Biol.* 234:1207, 1993.
- [79] Becker, O. M. Geometric versus topological clustering: An insight into conformation mapping. *Proteins* 27:213, 1997.
- [80] Quinkert, G., Egert, E., Griesinger, C. *Aspekte der Organischen Chemie*. Basel: Verlag Helvetica Chimica Acta. 1995.

- [81] Haasnoot, C., de Leeuw, F., de Leeuw, H., Altona, C. The relationship between proton-proton NMR coupling constants and substituent electronegativities. *Org. Magn. Reson.* 15:43, 1981.
- [82] Altona, C., Sundaralingam, M. Conformational analysis of the sugar ring in nucleosides and nucleotides. new description using the concept of pseudorotation. *J. Am. Chem. Soc.* 94:8205, 1972.
- [83] Kilpatrick, J., Pitzer, K., Spitzer, R. The thermodynamics and molecular structure of cyclopentane. *J. Am. Chem. Soc.* 69:2483, 1947.
- [84] Wijmenga, S., van Buuren, B. The use of NMR methods for conformational studies of nucleic acids. *Prog. Nuc. Magn. Reson. Spec.* 32:287, 1998.
- [85] Lankhorst, P., Haasnoot, C., Erkelens, C., Altona, C. C-13 NMR in conformational-analysis of nucleic-acid fragments 2. a reparametrization of the karplus equation for vicinal NMR coupling-constants in CCOP and HCOP fragments. *J. Biomol. Struct. Dyn.* 1:1387, 1984.
- [86] Brooks, C., Karplus, M., Pettitt, B. *Proteins: A theoretical perspective of dynamics, structure, and thermodynamics.* New York: John Wiley and Sons. 1988.
- [87] Ishima, R., Torchia, D. Protein dynamics from NMR. *Nat. Struct. Biol.* 7:740, 2000.
- [88] Osborne, M., Schnell, J., Benkovic, S., Dyson, H., Wright, P. Backbone dynamics in dihydrofolate reductase complexes: Role of loop flexibility in the catalytic mechanism. *Biochemistry* 40:9846, 2001.
- [89] Lipari, G., Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* 104:4546, 1982.
- [90] Palmer, A. I. NMR probes of molecular dynamics: Overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.* 30:129, 2001.
- [91] Clore, G., Szabo, A., Bax, A., Kay, L., Driscoll, P., Gronenborn, A. M. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* 112:4989, 1990.
- [92] Lipari, G., Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* 104:4559, 1982.

- [93] Wallach, D. Effect of internal rotation on angular correlation functions. *J. Chem. Phys.* 47:5258, 1967.
- [94] Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, 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., Tsui, V., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., Kollman, P. A. *AMBER 6*. San Francisco: University of California. 1999.
- [95] Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R., Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926, 1983.
- [96] Ryckaert, J., Cicotti, G., Berendsen, H. Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes. *J. Comput. Phys.* 23:327, 1977.
- [97] Darden, T., York, D., Petersen, L. Particle mesh ewald: An $n \log(n)$ method for ewald sums in large systems. *J. Chem. Phys.* 98:10089, 1993.
- [98] Kelly, L., Gardner, S., Sutcliffe, M. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally-related subfamilies. *Protein Engineering* 9 9:1063, 1996.
- [99] Neuhaus, D., Williams, M. *The Nuclear Overhauser effect in structural and conformational analysis*. New York: Wiley-VCH. 2000.
- [100] Richter, C. PhD Thesis, Johann Wolfgang Goethe-Universität, Frankfurt, 1999.
- [101] Cromsigt, J., Hilbers, C., Wijmenga, S. Prediction of proton chemical shifts. *J. Biomol. NMR* 21:11, 2001.
- [102] Nissen, P., Hansen, J., Ban, N., Moore, P., Steitz, T. The structural basis of ribosome activity in peptide bond synthesis. *Science* 289:920, 2000.
- [103] Williams, J., Hall, K. Unrestrained stochastic dynamics simulations of the UUCG tetraloop using an implicit solvation model. *Biopolymers* 76:3192, 1999.
- [104] Cheng, X., Hornak, V., Simmerling, C. Improved conformational sampling through an efficient combination of mean-field simulation approaches. *J. Phys. Chem. B* 108:426, 2004.
- [105] Sanger Principles of nucleic acid structure. New York, USA: Springer Verlag. 1984.

- [106] Varani, G., Cheong, C., Tinoco, I. J. Structure of an unusually stable RNA hairpin. *Biochemistry* 30:3280, 1991.
- [107] Pitera, J., Swope, W. Understanding folding and design: Replica exchange simulations of trp-cage miniproteins. *Proc. Natl. Acad. Sci. USA* 100:7587, 2003.
- [108] Garcia, A., Sanbonmatsu, K. α -helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. USA* 99:2782, 2002.
- [109] Carroll, R., Peterlin, B., Derse, D. Inhibition of human immunodeficiency virus type 1 Tat activity by coexpression of heterologous trans activation. *J. Virol.* 66:2000, 1992.
- [110] Puglisi, J., Tan, R., Calnan, B., Frankel, A. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* 257:76, 1992.
- [111] Ludwig, V. PhD Thesis, Johann Wolfgang Goethe-Universität, Frankfurt, 2005.
- [112] Krebs, A. PhD Thesis, Johann Wolfgang Goethe-Universität, Frankfurt, 2004.
- [113] Morris, G., Goodsell, D., Haliday, R., Huey, R., Hart, W., Belew, R., Olson, A. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19:1639, 1998.
- [114] Goodford, P. A computational procedure for determining energetically favourable binding sites on biologically important macromolecules. *J. Med. Chem.* 28:849, 1985.
- [115] Morris, G., Goodsell, D., Haliday, R., Huey, R., Olson, A. Distributed automated docking of flexible ligands to proteins: parallel applications of autodock 2.4. *J. Comput.-Aided Mol. Des.* 10:293, 1996.
- [116] Böhm, H.-J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* 8:243, 1994.
- [117] Mehler, E., Solmajer, T. Electrostatic effects in proteins: comparison of dielectric and charge model. *Protein Engineering* 4:903, 1991.
- [118] Holland, J. *Adaptation in natural and artificial systems.* University of Michigan Press, Ann Arbor, MI, 1985.
- [119] Michalewicz, Z. Genetic algorithms + data structures = evolution programs. In Glassner, A.S. (Ed.) *Graphics Gems*, Springer Verlag, New York, 1996.
- [120] Mailliot, P.-G. Academic Press, New York, 1990, 498.

- [121] Watt, A., Watt, M. Advanced animation and rendering techniques-theory and practice. ACM Press, New York, NY, and Addison-Wesley Publishing Company, Wokingham, UK.,1992.
- [122] Lamarck, J. Zoological philosophy. Macmillan, London, 1914.
- [123] Gaussian 98, Revision A.9, Pittsburgh PA, 1998. Gaussian, Inc.
- [124] Bayly, C., Cieplak, P., Cornell, W., Kollman, P. A well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: The RESP model. *J. Phys. Chem.* 97:10269, 1993.
- [125] Aboul-ela, F., Karn, J., Varani, G. The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.* 253:313, 1995.
- [126] Du, Z., Lind, K., James, T. Structure of TAR RNA complexed with a Tat-TAR interaction nanomolar inhibitor that was identified by computational screening. *Chem. Biol.* 9:707, 2002.
- [127] Murchie, A., Davis, B., C. Isel, M. A., Drysdale, M., Bower, J., Potter, A., Starkey, I., Swarbrick, T., Mirza, S., Prescott, C., Vaglio, P., Aboul-ela, F., Karn, J. Structure-based drug design targeting an inactive RNA conformation: Exploring the flexibility of HIV-1 TAR RNA. *J. Mol. Biol.* 336:2005, 2004.
- [128] Puglisi, D., Chen, L., Frankel, A., Williamson, J. Role of RNA structure in arginine recognition of TAR RNA. *Proc. Natl. Acad. Sci. USA* 90:3680, 1993.
- [129] Lapidot, A., Vijayabaskara, V., Litovchicka, A., Yub, J., James, T. Structure activity relationships of aminoglycoside-arginine conjugates that bind HIV-1RNAs as determined by fluorescence and NMR spectroscopy. *FEBS letters* 577:415, 2004.
- [130] Hermann, T., Westhof, E. Docking of cationic antibiotics to negatively charged pockets in RNA folds. *J. Med. Chem.* 42:1250, 1999.
- [131] Woods, R., Dwek, R., Edge, C., Fraserreid, B. Molecular mechanical and molecular dynamical simulations of glycoproteins and oligosaccharides. 1. GLYCAM-93 parameter development. *J. Phys. Chem.* 99:3832, 1995.

Lebenslauf

I. Persönliche Daten

Name: Koplín
Vorname: Jessica
Geburtsdatum: 24.11.1974
Geburtsort: Shillong, Indien
Familienstand: ledig, 1 Kind (9 Jahre)

II. Schulbildung und Studium

1. Grundschule, Indien: 1980-1986
2. Katoomba Public School, Australien: 1986-1987
3. Drais Realschule, Karlsruhe: 1987-1992
4. Claus v. Stauffenberg, Gymnasiale Oberstufe, Rodgau: 1992-1995
5. Technische Hochschule Darmstadt, Bauingenieurwesen: 1995-1996
6. Johann Wolfgang Goethe Universität, Frankfurt a. M.
Biochemiestudium: 1996-2001
7. Diplomarbeit am Institut für Biophysikalische Chemie
(Prof. Dr. H. Rüterjans): 2001
8. Beginn der Doktorarbeit am Institut für Physikalische und
Theoretische Chemie unter der Leitung von Prof. Dr. G. Stock: 8/2001

III. Vorträge

1. Eingeladene Rednerin in Polanica Zdroj, 2004. Vortrag zum Thema:
"Structure and Dynamics of an RNA tetraloop"
- Conference on New Materials, Polen
2. Vortrag in Kleinwalsertal, Institutsseminar, 2004, Österreich
3. Kurzvortrag in Weilburg, Seminar des SFBs 579, 2003, Deutschland

IV. Posterpräsentationen

1. Hünfeld, Deutschland: "MD simulation vs. replica exchange MD "
-Computer Simulations and Theory of Macromolecules, 2004
2. Frankfurt a. M., Deutschland: "MD simulation of RNA and ligands"
-RNA Structure, Function and Ligand Interaction, 2003

3. Hünfeld, Deutschland: “MD simulation vs. NMR”
-Computer Simulations and Theory of Macromolecules, 2003
4. London, England, King’s College: “RNA ligand interactions”
-Summer School on Molecular Simulations, 2002

V. Publikationen

1. Jessica Koplín, Yuguang Mu, Christian Richter, Harald Schwalbe and Gerhard Stock ”Structure and Dynamics of an RNA tetraloop: A joint molecular dynamics and NMR study. *STRUCTURE* 13 (9): 1255-1267 SEP 2005
2. Jessica Koplín, Phuong H. Nguyen and Gerhard Stock
”Structure and dynamics of an RNA tetraloop: a replica exchange molecular dynamics study”, in Vorbereitung