

Investigation of the Structure and Dynamics of RNA Systems by Molecular Dynamics Simulations

DISSERTATION
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt dem Fachbereich
Biochemie, Chemie und Pharmazie
der Johann Wolfgang GoetheUniversität
in Frankfurt am Main

von

Elisabeth Catherina Widjajakusuma
aus Ujung Pandang (Indonesien)

September 2007

vom Fachbereich Biochemie, Chemie und Pharmazie
der Johann Wolfgang GoetheUniversität als Dissertation angenommen

Dekan: Prof. Dr. Harald Schwalbe

1. Gutachter: Prof. Dr. Gerhard Stock

2. Gutachter: Prof. Dr. Harald Schwalbe

Datum der Disputation:

„Deo Gratias“

(Felix von Cantalice)

Danksagung

Den Herrn, unseren Gott Jesus Christus, möchte ich lobpreisen und ihm für die Anfertigung dieser Dissertation danken. Er hat mein Leben in Deutschland durch viele Leute, bei denen ich mich an dieser Stelle ganz herzlich bedanke, reichlich gesegnet.

Ich danke meinem Doktorvater, Prof. Dr. Gerhard Stock für die interessante Aufgabenstellung und die herzliche Aufnahme in seinem Arbeitskreis. Dank sei auch seiner Betreuung durch konstruktiven Kommentaren und aufbauenden Ermutigungen. Ferner wies er mich auf Seminare und ähnlichen Lehrveranstaltungen hin, durch welche ich meine Kenntnisse ausbauen konnte.

Dr. Alessandra Villa gilt besonderer Dank, weil sie eine sehr gute Gesprächspartnerin in allen Belangen meiner Arbeit war. Insbesondere danke ich ihr für die große Mühe, die sie sich mit der Korrektur meiner Arbeit gemacht hat, und für die kritische Durchsicht meiner Dissertation. Außerdem hat sie mir freundlicherweise ihre Daten und Bilder zum cUUCGg-Hairpin im Kapitel 4 und 5 zur Verfügung gestellt.

Prof. Dr. Harald Schwalbe und Prof. Dr. Jens Wöhnert danke ich für die gute Zusammenarbeit, besonders für die Bereitstellung von NMR-Daten, mit denen die Validerung der MD-Simulationen in dieser Arbeit ermöglicht wurde.

Weiterhin danke ich Dr. Rainer Hegger für seine Unterstützung im Umgang mit Rechnersystemen.

Mein Dank gilt außerdem der Sekretärin Frau Claudia von Sulecki für die Hilfe bei allen administrativen Tätigkeiten.

Außerdem gilt mein Dank der Prüfungskommission für die Unterstützung beim Zustandekommen der Disputation.

Dem Deutschen Akademischen Austauschdienst (DAAD) danke ich für den Sprachkurs (April 2007 bis September 2007), das Stipendium (Oktober 2004 bis September 2007), sowie die interessanten Aufträge. Frau Barbara Schwarz-Bergmann

danke ich für ihre Unterstützung als Ansprechpartnerin von DAAD bei allen Problemen und Anliegen.

Alle Berechnungen wurden mit Rechnern des Instituts für Physikalische und Theoretische Chemie der Johann Wolfgang Goethe-Universität Frankfurt am Main und des Frankfurt Center for Scientific Computing (CSC) durchgeführt. Für deren Bereitstellung bedanke ich mich sehr.

Bei Tagungsbesuchen hat mich die SFB 579 mit Reisekostenzuschüssen unterstützt. Auch dafür bedanke ich mich.

Die Katholische Universität Widya Mandala Surabaya, hat mir erlaubt in Deutschland weiterzustudieren und mir nach der Rückkehr eine Festanstellung in Indonesien versichert, wofür ich mich auch besonders bedanke.

Besonders muß ich das sehr gute Arbeitsklima hervorheben. Unter anderem hat Jessica mir die jetzige Wohnung in Frankfurt vorgestellt. Roman hat mich mit seinen Programmierkenntnissen unterstützt. Bei Sanni und Aleko bedanke ich mich für die sprachlichen Korrekturen und für deren immer währende Hilfsbereitschaft. Auch Phuong, Maja, und allen Mitglieder des AK um Prof. Josef Wachtveitl, besonders Mirka, und allen Mitglieder des AK Dr. Andreas Dreuw sei für die freundschaftlich-kollegiale Zusammenarbeit und deren immer vorhandene Hilfsbereitschaft gedankt.

Das Zusammentreffen so vieler netter Leute hat entscheidend dazu beigetragen, daß ich mich in Deutschland sehr schnell einleben konnte. Dem Pfarrer Wolf Schmidt, S.J, der mein Beichtvater ist, der Gemeinschaft Emmanuel bzw. meine Hausgemeinschaft in Frankfurt, bestehend aus Gabi und Hugo, Antonia und Ebenhard, François, Claudia und Thomas, Fransiska, Hajo, Viva, Letiana und Mark, Jean-Hugues, Emilie, und Siglinde danke ich sehr für die zahlreiche Unterstützung v.a im Gebet. Besonderen Dank gilt an Felix für die Deutschsprachenkorrektur. Auch danke ich der Gemeinschaft Emmanuel in Indonesien, die mich immer in ihren Gebeten aufnahmen und den liebevollen Anrufen von Andryani und Christina.

Schlussendlich danke ich meinen Eltern, meiner Schwägerin, meinen Brüdern und meinen Nichten für deren Liebe in allen Lebenslagen.

Content

Danksagung

1 Introduction	1
2 Molecular Dynamics Simulations	9
2.1 Force Fields.....	10
2.1.1 Bond and angle parameters.....	13
2.1.2 Torsional parameters.....	14
2.1.3 van der Waals parameters.....	16
2.1.4 Electrostatic parameters.....	17
2.2 Classical Molecular Dynamics Simulations.....	18
2.2.1 Periodic boundary conditions.....	20
2.2.2 Temperature and pressure.....	21
2.2.3 Constraint dynamics	23
2.3 Long-Range Interactions	23
2.3.1 Reaction field	24
2.3.2 Particle mesh Ewald	25
2.4 Replica Exchange Molecular Dynamics.....	27
2.4.1 Replica exchange method	27
2.5 Trajectory Analysis.....	29
2.5.1 Root mean square deviation	29
2.5.2 Root mean square fluctuation.....	30
2.5.3 Radius of gyration.....	30
2.5.4 Solvent accessible surface.....	30
2.5.5 Radial distribution function	32
2.5.6 Principal component analysis.....	34
3 Effect of Electrostatic Treatment and Force Fields on the Conformations of Small RNA Systems	35
3.1 Simulation of two RNA duplexes.....	36
3.1.1 Computational details.....	37

3.1.2	Effect of electrostatic treatments on duplexes.....	39
3.1.3	Conclusion	41
3.2	Different AMBER force fields in the investigation of CACG hairpin	42
3.2.1	Computational details	46
3.2.2	AMBER94, AMBER98 and AMBER99 versus experimental data	47
3.2.3	Conclusion	62
3.3	Effect of electrostatic treatments on uCACGg hairpin	63
3.3.1	Computational details	63
3.3.2	Reaction field versus PME	64
3.3.3	Comparison of different ion concentrations	68
3.3.4	Conclusion	73
4	Structure, Dynamics, and Thermostability of the RNA Hairpins	
	uCACGg and cUUCGg	75
4.1	Computational details	78
4.1.1	Simulation conditions.....	78
4.1.2	Replica-exchange molecular dynamics.....	79
4.1.3	Trajectory analysis	79
4.2	Result and Discussion.....	80
4.2.1	Characterization at 300 K	80
4.2.2	Thermal unfolding	84
4.3	Conclusion	92
5	Internal motion of RNA hairpins as reflected	
	by NMR relaxation parameters	95
5.1	Theory and Computational Details.....	97
5.1.1	Molecular dynamics simulations	97
5.1.2	NMR relaxation parameters	98
5.1.3	Correlation function	99
5.1.4	Order parameter	100
5.2	Results and Discussion.....	101
5.2.1	NMR relaxation parameter of the cUUCGg hairpin	101

5.2.2	Order parameters of the cUUCGg hairpin.....	105
5.2.3	Internal dynamics if the cUUCGg hairpin.....	108
5.2.4	Comparison of the fast dynamics of the cUUCGg and uCACGg hairpins.....	113
5.3	Conclusion	118
6	Conclusion	121
	Bibliography	125
	Zusammenfassung	139
	Lebenslauf	

Chapter 1

Introduction

Nucleic acids were originally understood simply as carriers of genetic information in the form of the genetic code. DNA was the repository of genetic information, and RNA served as a temporary copy to be decoded in the synthesis of proteins. The discovery of transfer RNA, the 'adapter' molecules that assist in the decoding of genetic messages, broadened awareness of the role of RNA. In the past few years, the functional versatility of nucleic acids and their participation in a wide range of vital cellular processes have to be appreciated (Hougland, 2006; Storz, 2006).

Nowadays, we recognize that RNA has greater structural versatility than DNA in the variety of its species, in its diversity of structures, and in its chemical reactivity. RNA plays a central role in many diverse biological processes, with messenger RNA (mRNA) serving as an informational molecule and template for protein production (Atkins, 2006), ribosomal RNA (rRNA) functioning as the biological machinery for protein production (Moore, 2006; Nissen, 2000), and transfer RNA (tRNA) operating as a vehicle for amino acids during translation and in error control (Roy, 2006).

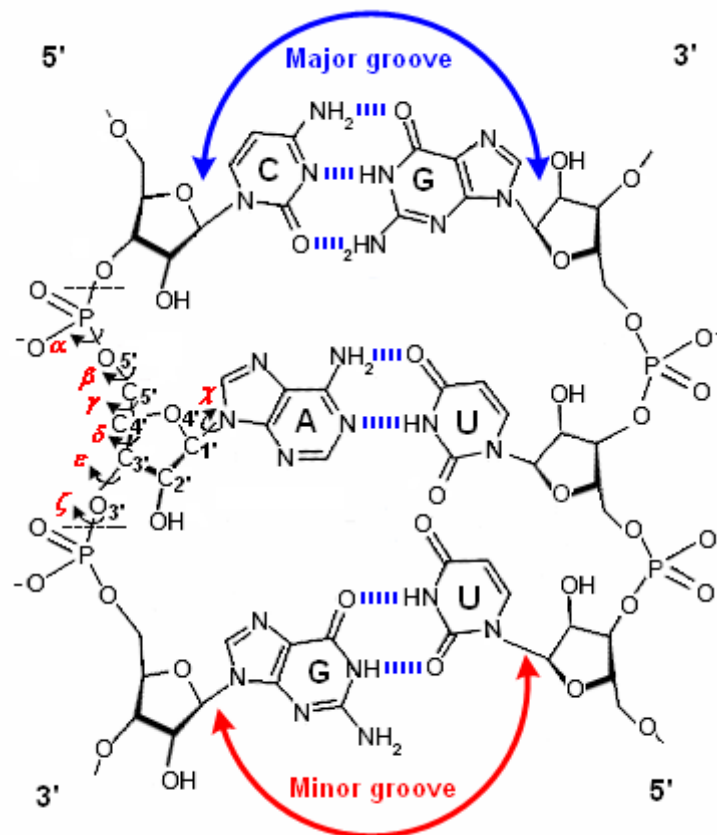


Figure 1.1 Atomic structure of RNA. Torsional angle α , β , γ , δ , ϵ , ζ of sugar-phosphate backbone atoms and glycosidic angle χ between the sugar and base moiety are shown. Base-pair Watson Crick (C–G and A–U) and a wobble pair (G–U) together with hydrogen bonds are shown. The RNA sequences are given in 5′-3′ direction of the phosphate-ribose backbone. Major and minor grooves are also illustrated.

The function of biomolecules is determined by their structures, which lead to a growing interest in the three-dimensional shapes of RNA molecules. At a two-dimensional level, RNA structure can be described and predicted by computational methods. The secondary structure is defined as the pattern of base-pairs. In RNA, there are four types of bases: adenine (A) and guanine (G) (purine), cytosine (C) and uracil (U) (pyrimidine). The hydrogen bonds which determine the geometry of pairing can be formed between the complementary Watson-Crick pairs G–C and A–U, as well as the less stable G–U wobble pair (Figure 1.1). A nucleotide sequence, simply gives the order of the nucleotides starting at the 5′-end and ending at the 3′-end. As shown in Figure 1.1 nucleic acids exhibit also two characteristic grooves which have different width and depth.

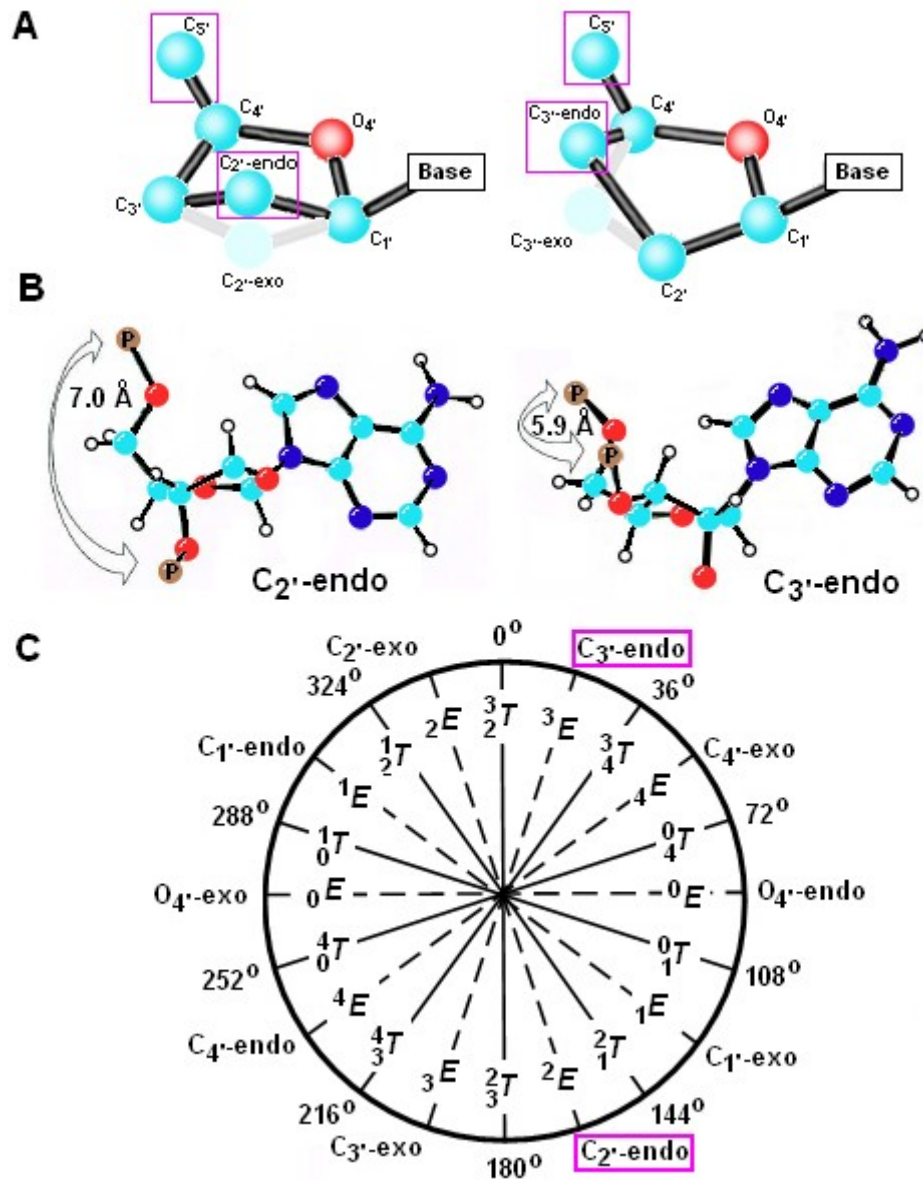
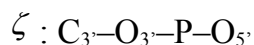
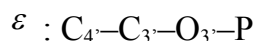
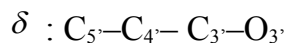
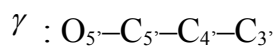
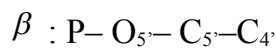


Figure 1.2 Sugar pucker. Ribose conformation C_{2'}-endo and C_{3'}-endo which are determined by the position of atom C_{2'} and C_{3'} to C_{5'} are shown (A). Sugar pucker controls phosphat-phosphat distance (B). Cyan spheres indicate the C atoms, red spheres the O atoms, blue spheres the N atoms and the brown spheres the P atoms. Sugar pucker in pseudorotation are represented (C) (Saenger, 1984)

RNA residues are characterized by seven torsional angles are labeled by Greek letters (Figure 1.1). Six of them are along the backbone, and coming from the 5'-end of the molecule their definition is as follows:

$$\alpha : \text{O}_3\text{-P-O}_5\text{-C}_5'$$



The seventh one is glycosidic torsional angle χ formed by atoms of sugar and base residue ($\text{O}_{4'}-\text{C}_{1'}-\text{N}_1-\text{C}_1$ and $\text{O}_{4'}-\text{C}_{1'}-\text{N}_9-\text{C}_4$ for pyrimidines and purines respectively). Two angles are of special interest as they usually assume only very specific values. First is the torsional angle δ that lies within the sugar ring system and is therefore restricted by a ring closure criterion. If the atom is on the same side of the plain as the $\text{C}_{5'}$ the conformation is called endo, if it is on the opposite side it is called exo. Figure 1.2 shows two of the most frequent sugar conformations in RNA, C_2' -endo (left-hand-side of Figure 1.2) and C_3' -endo (right-hand-side of Figure 1.2). This conformational behavior is also called sugar-puckering represented by pseudorotation angle P (Figure 1.2 C). The sugar pucker controls the distance between phosphorous atoms in the same strand (Figure 1.2 B). Nucleotides in the standard A-RNA helix are of C_3' -endo conformation, C_2' -endo conformations occur mostly in small loops, because of the tendency to elongate the backbone. Second angle is the glycosidic torsional angle χ . Though this torsional angle is not involved in a ring system, its values are nevertheless restricted to two distinct regions, one around 0° (syn-conformation) and the other around 180° (anti-conformation) (Figure 1.3).

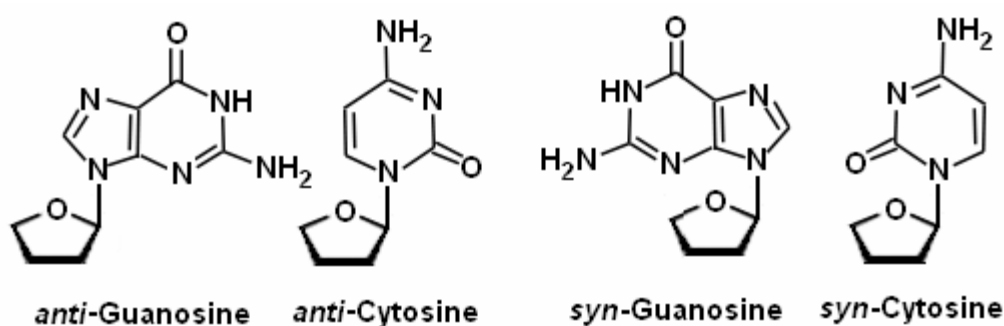


Figure 1.3 Sterically allowed orientations of purine and pyrimidine bases, that are represented by guanosine and cytosine respectively, with respect to their attached ribose units.

The glycosidic torsional angle generally adopts the anti-conformation, where the base is oriented away from the ribose. The anti-conformation generally maximizes hydrogen bonding and base stacking in nucleic acid secondary structures, while minimizing the electrostatic repulsion from phosphate groups. The syn-conformation is also a typical in RNA, and is normally found in structures where hydrogen bonding or stacking interactions counter the destabilization resulting from increased steric repulsion between the base and ribose. Also, sugar pucker is correlated to the glycosidic torsion angle (Murthy, 1999) and the syn-conformation is more readily adopted by purines than pyrimidines.

RNA secondary structure can be classified in a few types of structural motif (Figure 1.4) (Batey, 1999). One of the most common structural motifs in RNA is the hairpin, comprising of double stranded stem and a single stranded loop. Stability of hairpins depends on the nature of the closing base-pair (Serra, 1993), on the length of the loop

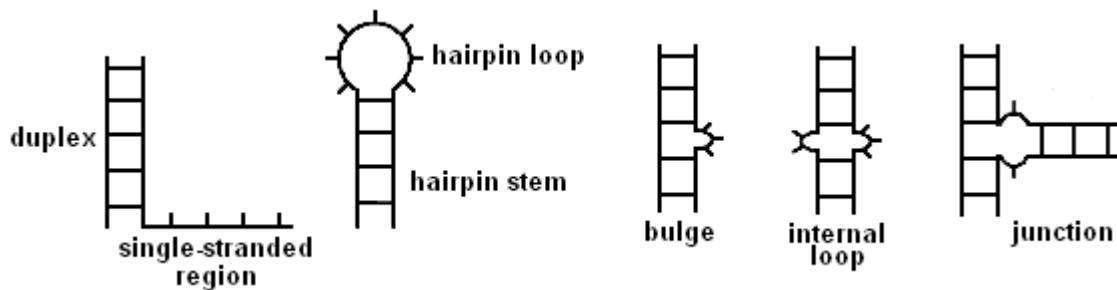


Figure 1.4 Secondary motif structures in RNA

and on its sequence. Hairpin loops exist in various sizes, ranging from three and four-membered types in ribosomal RNA up to loop sizes of 7, 8 or 9 nucleotides in tRNAs. In ribosomal RNA the most frequent hairpins consist of four nucleotides (Uhlenbeck, 1990) and are, therefore, often referred to as tetraloops.

Throughout this thesis the secondary structure encountered is mainly a tetraloop hairpin. The stability and structure of RNA tetraloops have been extensively studied. Notable for their abundance are three types of tetraloops: GNRA, UNCG, and CUUG (N stands for any base and R for a purine, either G or A) (Proctor, 2002; Wolters 1992). Known stable tetraloops were found, including sequences belonging to the UNCG motif closed by a CG base-pair. Among unusually stable hairpin loops, the UUCG tetraloop closed by a CG base-pair is the most stable sequence known (Proctor, 2002; Antao,

1991). The cUUCGg hairpin has first been identified in biochemical experiments (Tuerk, 1988) and later identified as common in biology using phylogenetic comparison (Woese, 1990). The CG closing base-pair makes a significant contribution to stability; for instance, a CG to GC change results in a relative stability at 37°C of +2.3 kcal mol⁻¹ for tetraloop formation (Antao, 1991).

Recently, the uCACGg hairpin was also observed to adopt UNCG-like structures (Ohlenschläger, 2004; Du, 2003; Proctor 2002). Thus, it suggests that tetraloops having the motif YNMG (Y stands a pyrimidine, M for C or A) are interchangeable because they are able to adopt highly similar three-dimensional structures in the context of various closing base-pairs or loop sizes (Proctor, 2002). Interestingly, the U–G closing base-pair in the stem and the A residue in the loop change significantly the stability and the biological function of the hairpin (Ohlenschläger, 2004). A more detailed the NMR structure of a uCACGg hairpin that occurs naturally within the cloverleaf RNA structure of the 5'-UTR of Coxsackievirus B3 (Ohlenschläger, 2004; Du, 2003). This hairpin is the major determinant for interaction between the cloverleaf RNA and viral 3C protease, which is an essential part of a ribonucleoprotein complex that plays a critical role in the regulation of viral translation and replication.

Molecular dynamics (MD) simulations are a theoretical tool for the microscopic description of structural and dynamic properties of molecules. In addition, MD also provides a complete microscopic description of the atomic structure and motions of macromolecules and solvent, and a direct route from the microscopic details to macroscopic properties (Frenkel, 2002). The main contributions that can be gained from microscopic observations are comprehension, interpretation of experimental results and the capability to expand simulations in experimentally poorly accessible regions.

Different elements are required to come together in order to perform dynamics simulations. An initial set of atomic coordinates and velocities and a description of the interaction between atoms in the system are required. These interatomic interactions can be well approximated by empirical functions, these come together to form a description of the forces acting on the atoms called the force field. The outcome of the simulation consists in the trajectory of all the atoms during the time covered by the simulation. This trajectory can then be analyzed to reach a new understanding of the system based on the

atomistic description of the system offered by MD. The trajectory also enables the calculation of observables which can then be compared to experimental value, such as NMR spectra (Beveridge *et al.*, 2000). Therefore, MD simulations can serve as a link between structure, function and dynamics in theory and experiment.

Recently, many accomplishments in MD simulations of biopolymers have been achieved to improve its performance and increase its reliability (Auffinger, 2006; van Gunsteren 2006; Norberg, 2003; Orozco, 2003; Frenkel, 2002; Gunsteren, 1999). MD has been very successful in modeling processes involving RNA systems such protein-RNA binding (Beveridge, 2006; Hall, 1992), nucleic acid hydration and ionic effects (Lyubartsev, 2004; Rueda, 2004; Várnai, 2004; Auffinger, 2000), and in investigation of structures and dynamics of small RNA systems (Špačková, 2006; Villa, 2006; Koplin 2005).

In this thesis, the structures and dynamics of small RNA systems are investigated by means MD simulations. These studies were motivated by the NMR experiments of Schwalbe and co worker, who explore the structures and dynamics of RNA systems by using NMR techniques. To achieve a microscopic picture of the structure and dynamics underlying these experiments, we performed MD simulations. The thesis is organized as follows. First, an introduction to the methodology of MD simulation is given in Chapter 2. In Chapter 3, the simulation protocol and validation of the model system are studied. The three version of AMBER force field and the two different methods to treat the electrostatic interaction are used to perform the simulations. The results are evaluated by comparison with the available experimental data. Then the effects of ion concentrations and ion types on the stability of RNA hairpins are investigated. Once the reliability of the molecular models and computational procedures has been established, MD simulations are used to describe the structural and dynamical properties of small RNA hairpins and to understand their folding/unfolding. In the Chapter 4, the replica exchange method is used to study the structure and the thermostability of two RNA hairpins: uCACGg and cUUCGg 14-mer RNA hairpins. In Chapter 5, the internal fast motion of the hairpins is examined.

The data and pictures of cUUCGg hairpin in Chapter 4 and 5 are given kindly by Dr. Alessandra Villa.

Chapter 2

Molecular Dynamics Simulations

A short overview to the methods which are used in the present thesis is presented. Section 2.1 gives an introduction to the force field. The accuracy of the force field used is important for molecular dynamics (MD) simulations, so over the years the empirical force field parameters have been continuously refined and optimized. Here the derivation of force field model from AMBER force field is focused. Section 2.2 deals with the methodology of MD simulations. Some basic ideas involved in MD simulations of molecular system are discussed. In Section 2.3 the treatment of the long-range electrostatic interactions is described. In particular, reaction field (RF) and particle mesh Ewald (PME) are discussed in details. A proper treatment of long-range forces in a simulation is essential for polyelectrolyte systems such as nucleic acids systems.

To understand how nucleic acids fold, that is, the sequences of structural changes they undergo before reaching their final native structure, one can use the replica exchange molecular dynamics (REMD) method. The REMD method allows the simulation to overcome the energy barriers which are separating local minima as it

moves into its folded state. Section 2.4 provides a description of the replica exchange molecular dynamics (REMD) approach. This method is considered to be one of the most promising and efficient methods to sample conformational states.

Finally, the analysis of MD trajectories used in this thesis is outlined in Section 2.5. This section provides an overview of the tools used to evaluate structural and dynamical properties of a system of interest.

2.1 Force Fields

The most accurate description of conformational and energetic properties of molecules can be achieved by the use of quantum mechanical (QM) methods. QM methods deal with the electrons in a system, so that even if some of the electrons are ignored (as in the semi-empirical schemes) a large number of particles must still be considered and the calculations are time-consuming. QM can also provide a detail or to a certain degree of accuracy that is unnecessary for the application. In molecular mechanical force fields, however, the electrons in an atom or molecule are not treated explicitly. Instead, their effect is included into an empirical potential which depends only on the atomic (nuclear) positions.

A force field consists of classical potential energy functions and the associated adjustable parameters. It is used to describe the relationship of the structure to the energy of the system of interest. For macromolecular systems, typically a potential energy has a general form (Jorgensen, 2006; MacKerrel, 2004; Leach, 1999; Cornell, 1995; Weiner, 1984):

$$V = V_{bond} + V_{angle} + V_{tor} + V_{nb} + V_{other} \quad (2.1)$$

where V_{bond} and V_{angle} stand for the stretching and bending energies; V_{tor} is used to represent the energy profile of rotations around chemical bonds; V_{nb} stands for the non-bonded interaction energies such as van der Waals potential and electrostatic potential; and V_{other} accounts for any other types of interaction such as symmetry restraints, external potential *etc.* Each term in equation 2.1 is described more detail in this section.

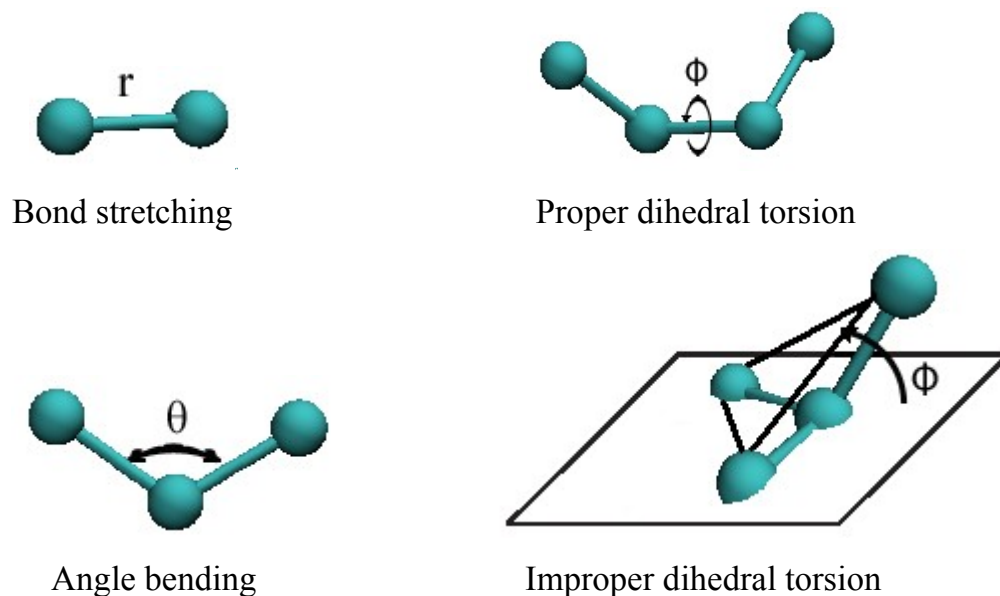


Figure 2.1 Schematic representation of the bonded interaction terms contributing to the force field.

The basis of a force field is the choice of atom types which are simplification of the real physico-chemical properties of atoms (Leach, 1999). A group of atoms which are enough similar, both chemically and physically, is treated the same in the molecular mechanic refinement. The decisions on atom types are based on an agreement between possessing the most accurate representation of some molecules and having manageable number of types. Several force fields classify the atoms into atom types on basis atomic number, hybridization, and neighboring environment. To represent its atom types, every classification uses specific symbolic codes. Different force fields can design the atom types with a different atom classification. AMBER force field uses the atom types which are described in Cornell, 1995 and Weiner, 1984.

The force field description of the interatomic forces can be distinguished into two categories: the bonded terms and the non-bonded terms. Figure 2.1 shows schematically the contribution of bonded interactions. The bonded terms refer to atoms that are linked by covalent bonds. The non-bonded terms consider interaction between non-bonded atoms or atoms separated by three or more covalent bonds. Force fields usually divide non-bonded interactions into electrostatic interactions and the van der Waals (vdW) interactions. The electrostatic interaction is due to the unequal distributions of charge in a molecule, and the vdW interaction consists of all attractive and repulsive forces

between atoms or molecules which are not covered by electrostatic interactions (Stone, 1997).

A number of force fields are developed for different aspect of bioorganic chemistry and with specific sets of data (Wang, 2001). For example, the MM family of force field is widely used for calculation on small molecules (Allinger, 1996; Allinger, 1989; Lii, 1989); the MMFF force field is used in pharmaceutical applications (Halgren, 1996a,b,c). Force fields used for nucleic acids are CHARMM (Foloppe, 2000; MacKerell, 2000; 1995; Brooks, 1983), AMBER (Wang, 2000; Cheatham, 1999; Cornell, 1995; Weiner, 1984), BMS (Langley, 1998), GROMOS (Soares, 2004; Daura, 1998) and OPLS (Jorgensen, 1988). While in CHARMM, AMBER, GROMOS and OPLS the parameters are optimized on the simplest molecule possible and then applied to larger or more complex molecules, in BMS the parameters are optimized to the best reproduce the conformational energy of a large molecule. Different force fields also use different level of detail and can be classified as all-atom, united-atom or coarse-grained. All-atom force field treat every atom (including hydrogen) explicitly, united-atom force fields combine each aliphatic carbon and associated hydrogens into a single particle, and coarse-grained force fields describe larger molecule units (such as amino acid side chains and whole water molecules) as a single particles. The recent versions of CHARMM, AMBER, BMS, and OPLS have an all-atom description, and only GROMOS is united-atom.

Partial atomic charges are designed to reproduce the electrostatic properties of a molecule and a model for the charge distribution. The AMBER and BMS force field derive charges from restrained fitting to the electrostatic potential, the CHARMM and OPLS using supramolecular approach (MacKerell, 2004), where the charges are optimized to reproduce interaction energies and geometries of the model compound with individual water molecules, and in GROMOS the partial charges are adjusted to reproduce the hydration free enthalpies in water (Oostenbrink, 2004).

The vdW parameters are the most difficult to derive in a force field due to both a lack of relevant experimental data to sufficiently constraint the parameters and the need for electron correlation and large basis sets for QM methods to accurately calculate the attractive dispersion contribution at small atomic separations (Bordner, 2003). In

AMBER, GROMOS, and OPLS the vdW parameters are obtained from liquid simulation, in CHARMM from experimental base-pairing geometries, *ab initio* interaction energies and geometries between water and model compound, and from heat of sublimation of same base analogs (Daura, 1998; Hobza, 1997).

In order to verify the generality of the force fields, in this case for nucleic acids, many MD simulations have been carried out on different force fields; CHARMM, AMBER and BMS force fields (Reddy, 2003; Cheatham, 2001). The GROMOS united atom force field is also still widely used in MD simulations that include explicit solvent representation (Dolenc, 2005; Soares, 2004). The analysis of simulation using OPLS force field shows overall good agreement with experiment (Zichi, 1995).

In this thesis, the AMBER force field is used to perform MD simulations. In the next discussion, the contributions of the AMBER force field are reported in detail. Moreover, the parameterization procedure is described. To understand the applicability of a given force field it is important to be aware of the approaches used in its development.

2.1.1 Bond and angle parameters

The analytical function of potential function for a covalent bond is:

$$V_{bond} = K_r (r - r_{eq})^2 \quad (2.2)$$

where V_{bond} is potential energy for bond deformation; r is the bond length; r_{eq} is the equilibrium distance; K_r is bond stretching force constant. This function describes the bond stretching energy, which in most cases at ordinary temperatures and in the absence of chemical reactions is sufficiently small for the harmonic approximation to apply (Figure 2.2).

Values of r_{eq} are taken from microwave and X-ray structural data on small molecular fragments that make up proteins and nucleic acids. Many of the K_r come from normal-mode calculation, in which the K_r values is varied to give the best fit to experimental frequencies. The remaining K_r are calculated via a direct linear interpolation. For example, linear interpolation between the “pure” C-C single bond from acetone ($r_{eq} = 1.507 \text{ \AA}$, $K_r = 337 \text{ kcal / mol \AA}^2$) and the “pure” C=C double bond from propene

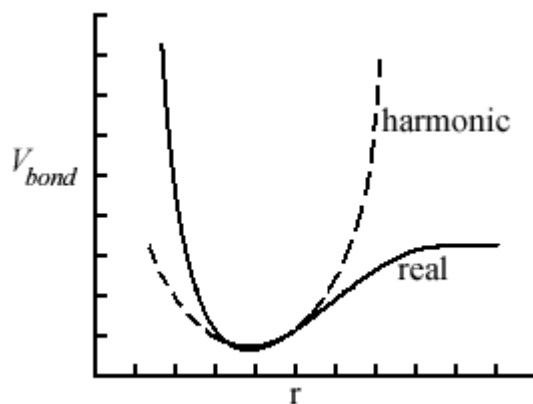


Figure 2.2 The potential energy V_{bond} as a function of the bond length r . The solid line is for simple harmonic potential and the dashed line for real potential.

($r_{eq} = 1.336 \text{ \AA}$, $K_r = 570 \text{ kcal / mol \AA}^2$). These predicted scaled values are compared to the calculated frequencies for stretching mode from the normal mode analysis. For example, the interpolation predicted stretching force constant for benzene ($r_{eq} = 1.40 \text{ \AA}$) is $475 \text{ kcal / mol \AA}^2$, and the calculated value, which gives the best fit to the experimental frequencies, is $469 \text{ kcal / mol \AA}^2$.

The potential energy function for deviation of angles from their reference value is:

$$V_{angle} = K_{\theta} (\theta - \theta_{eq})^2 \quad (2.3)$$

where V_{angle} is potential energy for angle deformation; θ is the angle between three atoms; θ_{eq} is the equilibrium angle; K_{θ} is bond bending force constant. This potential function is also described by a harmonic potential. θ_{eq} values are chosen from X-ray structural data on appropriate reference compounds. The normal mode calculations also played a large role in the choice of θ_{eq} and K_{θ} values. For example, all K_{θ} (CT-CT-X), where C2 is any sp^3 carbon and X can be any atom type, is assumed the same as K_{θ} (C2-C2-OS) derived for tetrahydrofuran, where OS is sp^3 oxygen in ethers.

2.1.2 Torsional parameters

The torsional potential V_{tor} of the equation 2.1 is commonly expressed as a cosine series expansion:

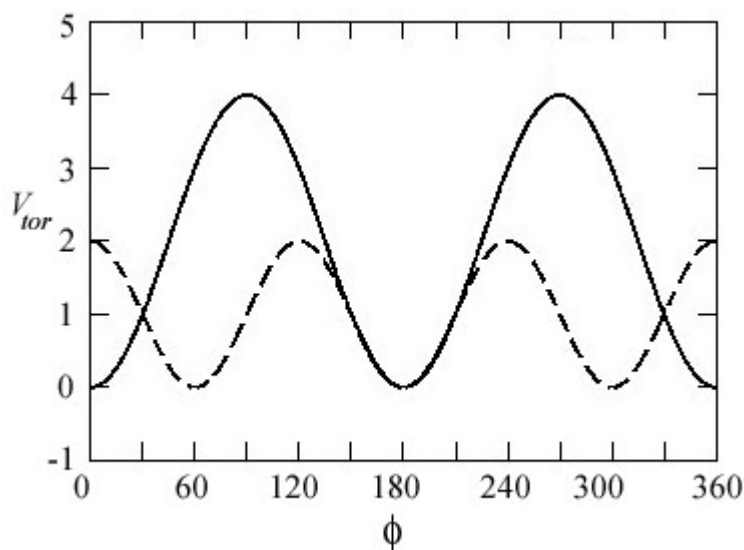


Figure 2.3 Potential energy V_{tor} as a function of the torsion angle ϕ . The solid line is for potential with $K_\phi = 4$, $n = 2$, and $\gamma = 180$, and the dashed line is for potential with $K_\phi = 2$, $n = 3$, and $\gamma = 0$

$$V_{tor} = \sum_{\text{torsions}} \frac{K_\phi}{2} [1 + \cos(n\phi - \gamma)] \quad (2.4)$$

where ϕ is the angle between the planes formed by the first and the last three of the four atoms; γ is phase angle; n in equation is the multiplicity and its value gives the number of minimum points in the function as the bond is rotated through its minimum value; the value of K_ϕ gives a qualitative indication of the relative barrier to the rotation. For example, K_ϕ for an amide bond, C–N, where C is any sp^2 carbon and N is sp^2 nitrogen in amide, is larger than for CT–CT. The effects of varying K_ϕ , n , and γ are illustrated in Figure 2.3 for commonly occurring torsional potentials.

The torsional parameters in the AMBER force field are divided in three types: general, specific, and improper. A general torsional parameter is a four-atom parameter based on the proper dihedral torsion (Figure 2.1) along an axis defined by the middle pair of atoms. Its energy profile depends solely upon the atom types of the middle pair, and not upon the atom types of the terminal atoms. For example, all torsion angles, where the central bond is between two sp^3 carbon atoms (such as HC–CT–CT–HC, CT–CT–CT–CT, HC–CT–CT–CT, where HC is explicit hydrogen attached to carbon) are described by the same torsional parameters, that can be identified with X–CT–CT–X, where X can be any atom. The specific torsional parameter is also based on proper dihedral torsion, but its energy profile depends not only upon the atom types of the

middle pair but also the terminal atoms and its value overrides any general parameter. For example, the torsional parameter for X–CT–CT–X is not the same for a special case such as OS–CT–CT–OS. The improper torsional parameter is a four-atom parameter based on the plane bending, in which the four atoms are not bonded in sequence (Figure 2.1), and it is a measure of the chirality / planarity of the structure at a specific atom.

The torsional parameter initially comes from X-ray structural data for a torsional angle, ϕ , and from IR or Raman spectroscopic data for K_ϕ and γ . However, since non-bonded and torsional terms are highly coupled, many are modified by generating the energy profile for the molecular mechanical non-bonded potential as is done for the quantum potential, subtract this curve from the quantum curve, and fit the torsional potential to the difference potential. Before these calculations can be done, atomic charges need to be calculated, also by fitting to quantum mechanical results. The difference potential is then deconvoluted into Fourier series terms which give the force field parameters.

2.1.3 van der Waals parameters

The van der Waals interaction between two atoms arises from a balance between repulsive and attractive forces, which can be described using an empirical function using Lennard-Jones (LJ) 12-6 function, which takes the following form for the interaction between atom i and j :

$$V_{vdw} = \frac{A_{ij}}{R^{12}} - \frac{B_{ij}}{R^6} \quad (2.5)$$

where:

$$A_{ij} = \varepsilon R^{*12}, \quad B_{ij} = 2\varepsilon R^{*6} \quad (2.6)$$

V_{vdw} is the van der Waals energy; R_{ij} is the distance between atom i and j ; ε is the LJ well-depth; R^* is minimum interaction radius. The LJ potential is characterized by an attractive part that varies as R_{ij}^{-6} and repulsive part that varies as R_{ij}^{-12} as shown in Figure 2.4. This potential contains only two adjustable parameters: the separation at which the energy passes through minimum R^* and the well-depth ε . The LJ parameters and partial atomic charges are highly correlated. Once partial atomic charges were

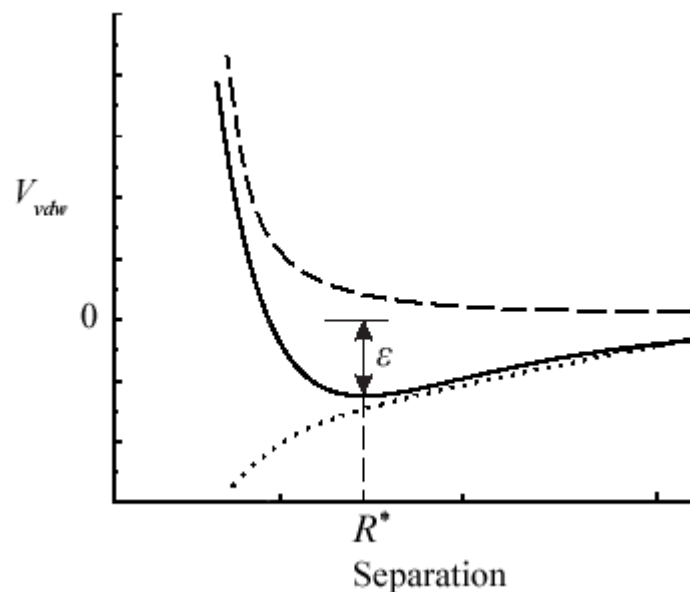


Figure 2.4 The potential V_{vdw} is described by the LJ potential as function of the internuclear separation. The solid line is the total LJ potential which consists of the repulsive part (dashed line) and the attractive component (dotted line). R^* and ϵ are illustrated.

assigned, the LJ parameters for a model compound were optimized for reproducing liquid properties, such as density or enthalpy.

2.1.4 Electrostatic parameters

The electrostatic interaction between two charge particles is given by Coulomb equation:

$$V_{elec} = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r R_{ij}} \quad (2.7)$$

where V_{elec} is electrostatic potential in the presence of dielectric medium; the constant ϵ_0 is the permittivity of free space; ϵ_r is the relative permittivity or relative dielectric constant; q_i and q_j are partial charges of atom i and j respectively.

The determination of the partial atomic charges in AMBER force field is based on quantum mechanics/restrained electrostatic potential approach (Bayly, 1993) called RESP approach. The REPS approach is consistent in a least-square fit of the charges to the electrostatic potential of the molecule with an additional restraint on charges of non-hydrogen atoms. These restraints serve to reduce the magnitude of charges without

impacting the fit, such as buried carbons. For example, the charges on the methyl atoms in methanol can be significantly reduced without impacting the fit, while the charges on the hydroxyl O and H have well defined values (Bayly, 1993).

2.2 Classical Molecular Dynamics Simulations

Computer simulation of molecular system has become an invaluable tool in academic and industrial research (van Gunsteren, 2006; Lamberti, 2002; van Gunsteren, 1998). The continuing growth of computing power makes it possible to analyze, compare, and characterize large and complex data sets that are obtained from experiment. The interaction between molecules, predictions of macroscopic properties, and dynamical properties can be achieved by using computer simulation. Such studies may lead not only to improved understanding and insight, but also to practical results such as engineered proteins or materials with properties optimized for particular application.

Two different kinds of computer experiments based upon statistical mechanical are especially useful for simulating atoms and molecules. The first method is Monte Carlo (MC) technique which is stochastic (Frenkel, 2002). In each step of an MC simulation, a randomly chosen particle is moved to a new randomly chosen location. If the new configuration has lower energy than the previous one, the move is immediately accepted; otherwise the new configuration is subjected to further statistical tests. If the move is ultimately rejected, the system is returned to its previous state. The second method is known as molecular dynamics (MD) simulation which calculates the time dependent behavior of a molecular system. An advantage of MD simulations over Monte Carlo simulations is that each successive iteration of the system is connected to the previous states of the system, which allows us to consider a property of the system, in this case inter-atom distances, as a function of time, so that a time correlation coefficient is able to be produced.

Molecular dynamics (MD) simulations provide evolution of the system in time. The position and velocities for each time step are determined by solving Newton's equations of motion for the atoms as function of time (Frenkel, 2002; Leach, 1999):

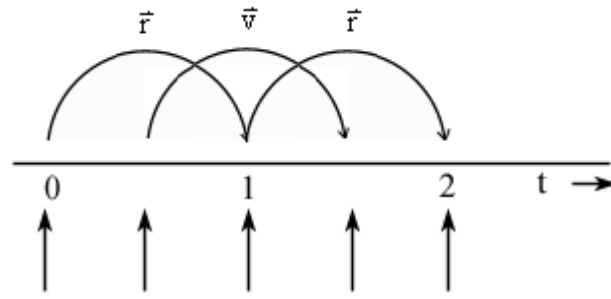


Figure 2.5 The leap-frog integration method. The velocities \vec{v} are first calculated at time $t + \frac{\Delta t}{2}$. In this way, the velocities leap over the positions \vec{r} , then the position \vec{r} at $t + \Delta t$ is calculated and leap over the velocities.

(2.8)

where m_i is the mass of atom i ; \vec{r}_i its position; \vec{F}_i is the total force acting on atom i . The forces \vec{F}_i acting on the atom i of the system, which are related to the first derivatives of the potential V with respect to the atom positions:

$$m_i d^2 \frac{\vec{r}_i}{dt^2} = \vec{F}_i \quad (2.9)$$

The potential energy is a function of the atomic positions of all the atoms in the system. Due to the complicated nature of this function, there is no analytical solution to the equations of motion; thus they must be solved numerically. Several algorithms are algorithms of Runge-Kutta, Gear, Verlet, Beeman and leap-frog. Here the leap-frog algorithm is described (Snyman, 1982). The leap-frog algorithm is a modified version of the *Verlet* algorithm (Verlet, 1967). It updates positions and velocities of each atom at time $t + \Delta t$ for positions and $t + \frac{\Delta t}{2}$ for velocities using the forces \vec{F} determined by the positions at time t :

$$\vec{v}(t + \Delta \frac{t}{2}) = \vec{v}(t - \Delta \frac{t}{2}) + F \frac{(t)}{m} \Delta t \quad (2.10)$$

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \vec{v}(t + \Delta \frac{t}{2}) \Delta t \quad (2.11)$$

The algorithm is visualized in Figure 2.5.

MD simulations can briefly described as follows: An initial set of atomic coordinates and velocities and a description of the interaction between atoms in the system is

required to begin a MD simulation. The coordinates can be obtained from X-ray crystallographic or NMR structure data, or by a manually built structure. Integration of equation of motion generates the coordinates and velocities of the atoms as function of time. The interactions of atoms and molecules are evaluated according to a force field. The resulting forces are then applied to the atoms using Newton's law of motion, thus obtaining slightly different atom positions. The outcome of the simulation consists in the trajectory of all the atoms during the time covered by the simulation. This trajectory, which is a series of molecular configurations as a function of time, can then be analyzed to reach a new understanding of the system based on the atomistic description (Beveridge, 2000).

Furthermore, computer simulation of molecular system requires software to calculate the interatomic interactions and to integrate the equations of motion. The software must often manipulate, simulate and analyze thousands or even tens thousands of atoms. This situation has led to the development of simulation software packages (van Gunsteren, 1998) AMBER, BRUGEL, CEDAR, CHARMM, EGO, ENCAD, FOCUS, GROMACS, GROMOS, MOIL, NAMD, POLARIS, UHBD, X-PLOR, and YASP and three of the most popular molecular dynamics packages are AMBER (Weiner, 1981), CHARMM (Brooks, 1983), and GROMACS (Berendsen, 1995a; Lindahl, 2001; van der Spoel, 2006; van der Spoel, 2005). The simulation presented in this thesis have been performed using the software package GROMACS.

The highly optimized code makes GROMACS the fastest program for molecular simulations to date. Besides, the support of different force field and the open source (GPL) character make GROMACS very flexible. A notable use of GROMACS is in the distributed computing project Folding@home (<http://www.folding.stanford.edu>) where it is used extensively in the simulation of protein folding.

2.2.1 Periodic boundary conditions

Periodic boundary conditions enable a simulation to be performed using a relatively small number of particles in such a way that the particles experience forces as though they are in a bulk solution in order to minimize edge effects in a finite system. The atoms of the system are put into a space-filling space, which is surrounded by translated copies

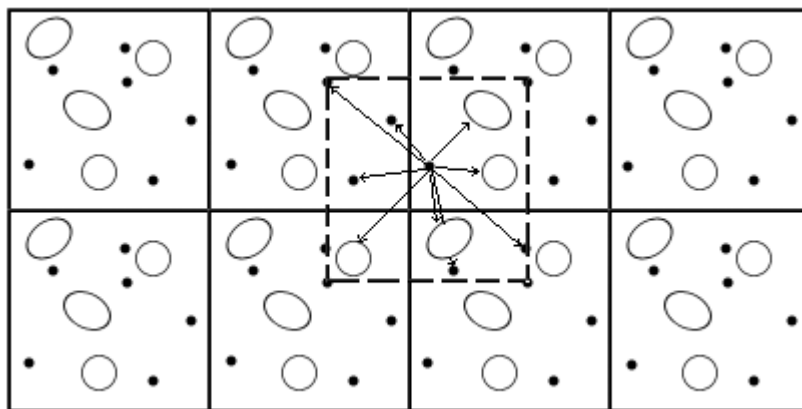


Figure 2.6 Schematic presentation of periodic boundary conditions

(Frenkel, 2002; Gunsteren 1999; Leach, 1999). The black atom in the central box in Figure 2.6 interacts with atoms or images of atoms that lie within the dashed line. The several possible shapes for space-filling unit cells in GROMACS are cubic, rhombic dodecahedron, truncated octahedron, and the one that rhombic dodecahedron is used in the Thesis.

2.2.2 Temperature and pressure

The most common ensemble is the isothermal-isobaric (NPT) ensembles, where the number of particles (N), the pressure (P), and the temperature (T) are constant. Performing MD with similar condition is wanted. Several methods can be used to keep constant temperature and pressure. Among others are Berendsen (Berendsen, 1984), Langevin (Adelman, 1976) and Nosé-Hoover (Nosé, 1990; Hoover, 1985) thermostats. In the thesis Berendsen's weak coupling scheme is used to keep temperature and pressure constant (Berendsen, 1984).

The temperature is kept constant by coupling the system to an external bath at the desired temperature. The bath acts as a source of the thermal energy, supplying or removing heat from the system as appropriate. The rate of change of temperature is proportional to the difference in temperature between the bath and the system:

$$\frac{dT(t)}{dt} = \frac{1}{\tau} (T_{bath} - T(t)) \quad (2.12)$$

where τ is a coupling parameter whose magnitude determines how tightly the bath and the system are coupled together; T_{bath} is the temperature of external bath; $T(t)$ is the actual temperature of system at time t . This method provides an exponential decay of the system towards the desired temperature. The change in temperature between successive time steps is:

$$\Delta T = \frac{\Delta t}{\tau} (T_{bath} - T(t)) \quad (2.13)$$

where Δt is time step. The velocities at actual time t are scaled by factor λ to the velocity of the reference T_{bath} :

$$\lambda = \left[1 + \frac{t}{\tau} \left(\frac{T_{bath}}{T(t)} - 1 \right) \right]^{\frac{1}{2}} \quad (2.14)$$

If τ is large, then the coupling will be weak, and if τ is small, the coupling will be strong.

The pressure can be maintained at a constant value by simply scaling the volume. One can couple the system to a ‘pressure bath’, analogous to a temperature bath (Berendsen, 1984). The rate of change of pressure is given by:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_p} (P_{bath} - P(t)) \quad (2.15)$$

where τ_p is the coupling parameter, P_{bath} is the pressure of the bath, and $P(t)$ is the actual pressure at time t . The volume of the simulation box is scaled by a factor μ , which is equivalent to scaling the atomic coordinates by a factor μ :

$$\mu = \left[1 + \frac{\beta \Delta t}{\tau_p} (P - P_{bath}) \right]^{\frac{1}{3}} \quad (2.16)$$

where β is the isothermal compressibility of the system, Δt is the time difference between the time before the coupling and the time at the coupling. A pressure change can be accomplished by scaling the new position given by:

$$\vec{r}_i' = \mu \vec{r}_i \quad (2.17)$$

where \vec{r}_i' is the new position after the coupling and \vec{r}_i is the position before the coupling.

2.2.3 Constraint dynamics

Intermolecular bond vibrations are typically the highest frequencies in the system and therefore determine how large time step that can be used. If bonds are constrained, a larger time step can be used, which speeds up the computation. It can be applied usually to X–H bonds, which exhibit fastest vibrations. This still allows 2 fs time steps, compared to 0.5-1 fs without bond constraints dynamics. Several constraint algorithms are SHAKE (Ryckaert, 1977), RATTLE (Andersen, 1983), SETTLE (Miyamoto, 1992), LINCS (Hess, 1997), and multiple-step algorithm (Tuckerman, 2000). SHAKE may run into convergence problems for large coordinate displacement, and it does not parallelize well because of its iterative nature. The constraint method LINCS which reset the constraints by calculating the new constrained position from the old positions and the new unconstrained positions, can achieve a speed up of 3-4 times compared to SHAKE. RATTLE and SETTLE allow velocities to be constrained. In GROMACS, LINCS and SETTLE is applied to constrain covalent bond lengths.

2.3 Long-Range Interactions

The most time consuming part of a molecular dynamics simulation is the calculation of the non-bonded terms in the potential energy function. Those are the electrostatic and van der Waals forces. In principle, the non-bonded energy terms between every pair of atoms should be evaluated; in this case, the number of increases as the square of the number of atoms for a pairwise model. To speed up the computation, the interactions between two atoms separated by a distance greater than a pre-defined distance, the cutoff distance, are ignored and the potential is truncated in cutoff distance. But it gives rise to serious inaccuracies for the electrostatic interactions because the potential energy decays slower than r^{-3} , where r is distance between two atoms, because contributions from these long-range interactions are large and do not converge at the cutoff distance (Norberg, 2003; Frenkel, 2002; Harvey, 1989).

The contribution of electrostatic interactions at long-range are not negligible, and particularly difficult to evaluate in computer simulations due to their long-range nature (Róg, 2003; Frenkel, 2002; Sagui, 1999). The methods used for the treatment of electrostatic interactions typically interpret the force on each charge into short- and long-range contribution. The interactions within a short-range cutoff are calculated according to Coulomb interactions. Beyond a cutoff distance or the long-range interaction can be treated in different ways among others by modeling the solvent screening effect and by numerical approach (Gargalo, 2003; Walser, 2001; Hünenberger, 1998; Harver, 1989). A correct treatment of long-range electrostatic interactions is essential for obtaining meaningful simulation results. In this section two approximations used to treat the long-range interactions in this thesis are discussed, which are (1) reaction field (van der Spoel, 2004; Tironi, 1995) and (2) particle mesh Ewald (PME) method (Frenkel, 2002; Darden, 1993).

2.3.1 Reaction field

An approach to treat the long-range contribution is the generalized reaction field method based on Poisson-Boltzmann approach (Tironi, 1995). In the reaction field method, each charge is individually considered as the origin of a spherical coordinate system. A spherical coordinate system represents a natural choice because of the symmetry. A charge is surrounded by two regions I and II. The inner region I within the cutoff sphere contains explicit neighboring particles and has a dielectric permittivity ϵ_1 . The sphere is enclosed by the outer region II, a continuum characterized by a dielectric permittivity ϵ_2 .

The electrostatic interaction between two charge particles given by equation 2.7 can be modified for homogenous systems, by assuming a constant dielectric environment beyond a cutoff distance r_c with a dielectric constant of ϵ_{rf} . The interaction becomes:

$$V_{elec} = \frac{1}{4\pi\epsilon_o\epsilon_r} \frac{q_i q_j}{r_{ij}} \left[1 + \frac{\epsilon_{rf} - 1}{2\epsilon_{rf} + 1} \frac{r_{ij}^3}{r_c^3} \right] - \frac{1}{4\pi\epsilon_o\epsilon_r} \frac{q_i q_j}{r_c} \frac{3r_{rf}}{2\epsilon_{rf} + 1} \quad (2.18)$$

The electrostatic potential is zero at the cutoff r_c (Figure 2.6):

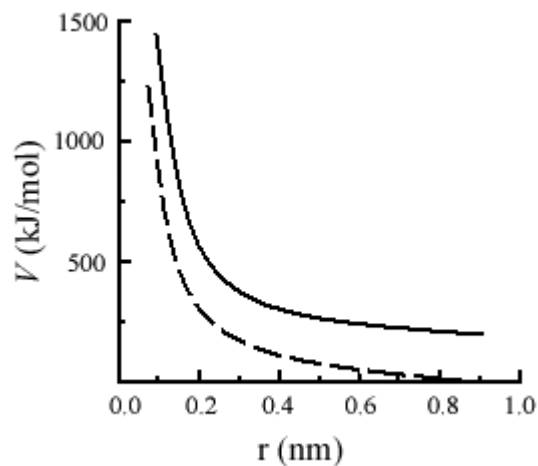


Figure 2.7 The electrostatic interaction with cutoff (line) and with cutoff and reaction field (dashed line)

2.3.2 Particle mesh Ewald

The long-range treatment can be approached from a different point of view by the Ewald Sum method (Sagui, 1999), which uses an artificial set of Gaussian charges to mask the real charges of the system in the direct space lattice, ensuring a neutral non-dipolar surface for the real space sum (Figure 2.8). Charge distribution in lattice of positive and negative ions may be considered as real space sum, which converges since Gaussians mask charges at long-range, and a reciprocal space used to correct masking charges. The Ewald Sum is a technique to sum the long-range interactions between particles and all their infinite periodic images efficiently. It modifies the electrostatic potential energy in equation 2.7 into the sum of two series plus a constant term. The Ewald Sum is written as the sum of three part, namely, the real (direct) sum (V_{real}), the reciprocal (imaginary, or Fourier) sum (V_{rec}), and the constant term (V_{self}). Each term is discussed below.

$$V_{elec} = V_{real} + V_{rec} + V_{self} \quad (2.19)$$

The particle mesh Ewald (PME) method is a modification of Ewald Sum method for intermediate size systems. In PME the reciprocal sum is approximated using fast Fourier transformation with convolutions on a grid where charges are interpolated to the grid points (Darden, 1993) in order to reduce the computational cost.

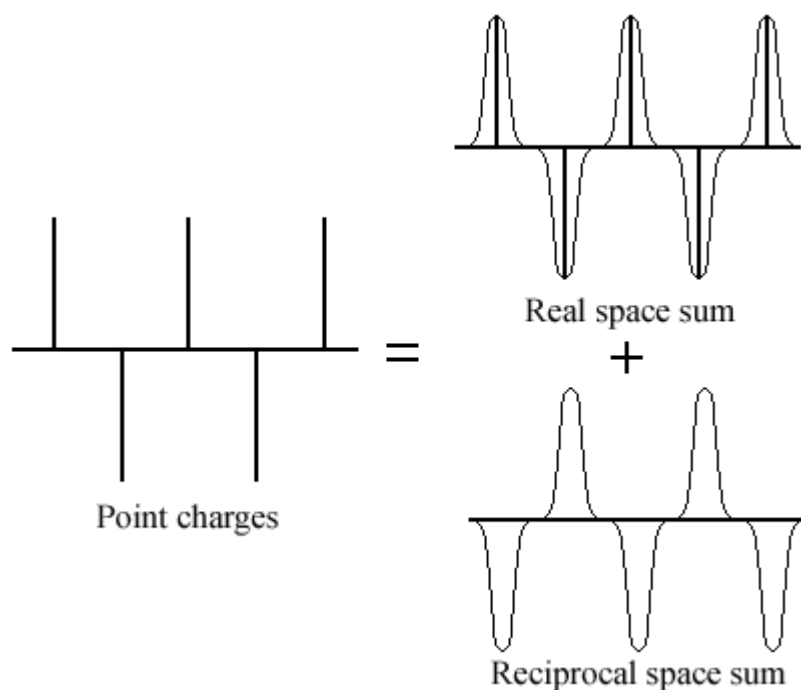


Figure 2.8 The Ewald sum of point-charge system consists of a real space described by original point charges (vertical line) plus screening distribution (curved line) and a reciprocal space. The vertical lines are positive (upward) or negative (downward) unit charge, and the curve lines represent the Gaussian charge cloud normalized to unity.

Ewald Sum : real space terms

The particles are assumed to be located in a cube with a certain diameter within the periodic boundary conditions. The electrostatic energy for real space of the Ewald sum is short-ranged and considered as energy due to point charges screened by oppositely charged Gaussian.

Ewald Sum: reciprocal space terms

Because a screening charge is summed to the point charge in the real space, an exactly compensating charge distribution must be added so that the overall potential is identical to the original one. This cancelling distribution can be obtained from Poisson's equation and is solved as a Fourier series. This long-range contribution is evaluated in reciprocal space.

Ewald Sum: self interaction terms

The complete Ewald sum requires an additional correction, known as the self energy correction, which arises from a Gaussian acting on its own site, and is constant.

2.4 Replica Exchange Molecular Dynamics

Conventional MD simulations provide insufficient conformational sampling for big biomolecules, since systems might be trapped in local energy minima due to the nature of the complex potential energy landscapes (Sugita, 1999). To circumvent these sampling problems, the replica exchange method is developed based on non-Boltzmann probability weight factor.

In replica exchange molecular dynamics (REMD) technique, several independent trajectories, called replicas, are generated at different temperatures, and stochastic exchanges between neighboring trajectories are attempted with predetermined intervals during the simulation (Sugita, 1999). If the exchange is accepted, the bath temperatures of these replicas will be swapped, and the velocities will be scaled accordingly. The trajectory exchanges between the replicas allow the system to escape from the local energy minima easily, exploring a broad range of the potential energy surface.

2.4.1 Replica exchange method

The method constructs a generalized ensemble from M non-interacting trajectories with temperatures T_m ($m = 1, \dots, M$). In generalized ensemble each state is weighted by a non-Boltzmann probability weight factor so that a random walk in potential energy space may be realized. The random walk allows the simulation to escape from any energy barrier and to sample much more space state than conventional methods (Mitsutake, 2001). A state of this generalized ensemble is characterized by $X = \{\dots, x_{(m)}^i, \dots\}$, where $x_m^{(i)}$ represents the coordinates and momenta of atoms of the i th replica at the temperature T_m . The method consists of 2 consecutive steps:

- (1) independent constant-temperature simulations of each replica

(2)exchange of the trajectories (the coordinates) between replicas according to the probability

The probability for randomly exchanging of two replicas at regular intervals for canonical ensemble is:

$$P(1 \leftrightarrow 2) = \min \left(1, \exp \left[\left(\frac{1}{k_B T_1} - \frac{1}{k_B T_2} \right) (V_1 - V_2) \right] \right) \quad (2.20)$$

where T_1 and T_2 are the reference temperatures and V_1 and V_2 are the instantaneous potential energies of replicas 1 and 2 respectively; k_B is Boltzmann constant and $\min(1, x)$ means $P = 1$ if $x \leq 0$, otherwise $P = x$. After exchange the velocities are scaled to get the appropriate velocities to the correspond temperature by $(T_1/ T_2)^{\pm 0.5}$ (derived in equation 2.21) and a neighbor search is performed the next step.

This combines the fast sampling and frequent barrier-crossing of the highest temperature with correct Boltzmann sampling at all the different temperatures. In GROMACS, for all ‘odd’ pairs were attempted to exchange on ‘odd’ attempts, and for all ‘even’ pairs on ‘even’ attempts. Considering, there are four replicas: 0, 1, 2, 3, ordered in temperature and are attempted to exchange every 1000 step. Pair 0-1 and 2-3 will be tried at steps 1000, 3000 etc. and pair 1-2 at steps 2000, 4000 etc.

An extension to the REMD for the isobaric-isothermal ensemble was proposed by Okabe co-workers (Okabe, 2001). In this work the exchange probability is modified to:

$$P(1 \leftrightarrow 2) = \min \left(1, \exp \left[\left(\frac{1}{k_B T_1} - \frac{1}{k_B T_2} \right) (V_1 - V_2) + \left(\frac{P_1}{k_B T_1} - \frac{P_2}{k_B T_2} \right) (Vol_1 - Vol_2) \right] \right) \quad (2.21)$$

where P_1 and P_2 are the respective reference pressures and Vol_1 and Vol_2 are the respective instantaneous volumes in the simulations. In most cases the differences in volume are so small that the second term is negligible. It only plays a role when the difference between P_1 and P_2 is large or in phase transitions.

To achieve the optimal performance of the algorithm for a given system, the temperatures of replica are chosen such that:

- (1) the lowest temperature is small enough to sufficiently sample the states of low energy,

- (2) the highest temperature is large enough to overcome energy barriers of the system,
- (3) the acceptance ratio (P) should be uniform and large enough ($> 10\%$).

To calculate the acceptance ratio the energy difference can be written as:

$$V_1 - V_2 = N_{df} \frac{c}{2} k_B (T_1 - T_2) \quad (2.35)$$

where N_{df} is the total number of degrees of freedom of one replica and c is 1 for harmonic potentials and around 2 for protein/water systems. Assuming $T_2 = (1 + \varepsilon)T_1$ the probability becomes:

$$P(1 \leftrightarrow 2) = \exp\left(-\frac{\varepsilon^2 c N_{df}}{2(1 + \varepsilon)}\right) \approx \exp\left(-\varepsilon^2 \frac{c}{2} N_{df}\right) \quad (2.36)$$

In order to get a probability of $e^{-2} \approx 0.135$ one obtains ε should be $2/\sqrt{c N_{df}}$.

2.5 Trajectory Analysis

The behavior of the RNA hairpins during MD simulation is analyzed by evaluating the following structural and dynamic parameters.

2.5.1 Root mean square deviation

Root mean square deviation from the starting structure is a commonly used criterion for validation of MD simulations. Because the MD protocol is expected to maintain the starting structures in native conformations, a good model system should give small *RMSD* values that become constant during the trajectory. The *RMSD* is a simple measure of the difference between the simulated and the native structures (Maiorov, 1995).

The *RMSD* of certain atoms in a molecule with respect to a reference structure is calculated in two steps, by least-squares fitting of the structure to a reference structure (t_0) and subsequently calculating:

$$RMSD(t, t_0) = \left[\frac{1}{N} \sum_{i=1}^N (r_i(t) - r_i(t_0))^2 \right]^{1/2} \quad (2.37)$$

where N is the number of atoms, $r_i(t)$ and $r_i(t_0)$ are the atomic position after fitting at time t and t_0 , respectively.

2.5.2 Root mean square fluctuation

Calculating the root mean square fluctuation (*RMSF*) of atoms (i.e. standard deviation of atomic positions) averaged over a period of simulation time allows comparisons of atomic mobility (Hünenberger, 1995; Fujita, 1986):

$$RMSF = \left[\frac{1}{n} \sum_{k=0}^n (r_{i,k} - \bar{r}_i)^2 \right]^{1/2} \quad (2.38)$$

where $r_{i,k}$ and \bar{r}_i are the position of atom i at step k and the averaged position of atom i , respectively; n the total number of integration steps.

2.5.3 Radius of gyration.

Radius of gyration (R_g) is a measure of compactness of a molecule. The R_g of a group of atoms is defined as the rms distance from each atom of the molecule to their centroid (Huang 2001; Kuszewski, 1999):

$$R_g = \left\{ \sum_{j=1}^N \frac{\left[r_j - \sum_{i=1}^N \frac{r_i}{N} \right]^2}{N} \right\}^{1/2} \quad (2.39)$$

where r_i and r_j are the position vectors of atoms i and j , and N is the number of atoms. R_g can provide information regarding the global conformation of a molecule.

2.5.4 Solvent accessible surface

The concept of the solvent accessible surface of a protein molecule was originally introduced by Lee and Richards (Lee, 1971), as a way of quantifying hydrophobic burial. The terms “accessible surface area”, A , of an atom is the area on the surface of a sphere of radius R , on each point of which the center of a solvent molecule can be placed in contact with this atom without penetrating any other atoms of the molecule. The radius R is given by the sum of the van der Waal’s radius of the atom and the chosen radius of the solvent molecule. An approximation to this area is computed using the formula:

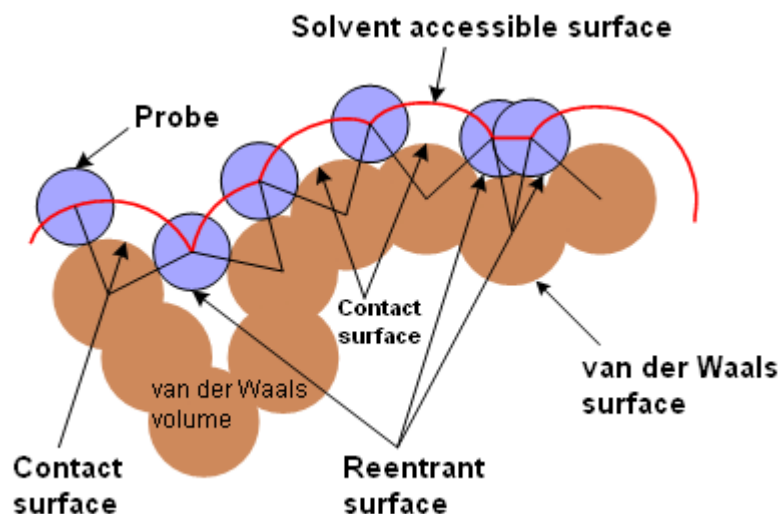


Figure 2.9 Solvent accessible surface of the molecule (in brown) is traced out by the probe sphere center (in light blue) as it rolls over the molecule. The accessible surface is shown by the red line. The contact surface, reentrant surface, and van der Waals surface are also introduced.

$$\text{accessible surface area} = A = \sum \left(\frac{R}{\sqrt{R^2 - Z_i^2}} \right) \cdot D \cdot L_i \quad (2.40)$$

$$D = \Delta \frac{Z}{2} + \Delta' Z$$

where L_i is the length of the arc drawn on a given section i , Z_i is the perpendicular distance from the center of the sphere to the section i , ΔZ is the spacing between the sections, and $\Delta' Z$ is $\Delta Z/2$ or $R - Z_i$, whichever is smaller. Summation is over all of the arcs drawn for the given atom.

For practical purpose, an alternative solvent-accessible surface definition, proposed by Richards (Richard, 1977), is appropriate. Unlike the original surface of Lee and Richards (Lee, 1971), this alternative molecular surface is not displaced from the van der Waals surface. Instead, it consists of the part of the van der Waals surface of the atoms that are accessible to the probe sphere (contact surface), connected by a network of concave and saddle-shaped surfaces (reentrant surface) that smoothes over the crevices and pits between the atoms shown in Figure 2.9. This surface is the boundary of the volume from which a probe sphere is excluded if it is not to experience van der Waals overlap with the atoms. An efficient computer algorithm for deriving these surfaces was described by Connolly (Connolly, 1983). In this thesis, the Connolly's method is used to

calculate the solvent accessible surface. The surface was computed numerically by using atomic radii of 0.16 nm for carbon, 0.13 nm for oxygen, 0.14 nm for nitrogen, 0.20 nm for phosphor and 0.10 nm for hydrogen atoms. The atomic radius of the water was 0.14 nm. An atom was recognized as hydrophobic if the absolute value of its partial charge was less than 0.2 e (Eisenhaber, 1995).

2.5.5. Radial distribution function

The radial distribution function, also known as RDF, $g(r)$, or the pair distribution function, is a measure to determine the correlation between particles within a system. The RDF provides structural information about the system. Specifically, it is a measure of, on average, the probability of finding a particle at a distance of r away from a given reference particle (Allen, 1987; McQuarrie, 1976). At a more detailed level, a RDF can give information about the environment of individual atoms; it can be used to study hydrogen-bonding and salt bridges and liquid structure in general. The RDF is defined as the local density of a certain group of particles around a central particle, and it may be averaged over multiple central particles:

$$\begin{aligned}
 g_{AB}(r) &= \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle_{\text{local}}} \\
 &= \frac{1}{\langle \rho_B \rangle_{\text{local}}} \frac{1}{N_A} \sum_{i \in A} \sum_{j \in B} \frac{\delta(r_{ij} - r)}{4\pi r^2}
 \end{aligned}
 \tag{2.41}$$

In Fig. 2.10 the RDF of a surrounding particle (colored spheres) around a reference atom (a black sphere) is plotted. For short distances, this is related to how the particles are packed together. The spheres can't overlap, so the closest distance two centers can be is equal to the diameter of the spheres, and at very short r the RDF must be zero. Then $g_{AB}(r)$ displays a first strong first peak which indicates a local accumulation of atoms. The peaks in the RDF show where the atoms spend the most time. For a completely homogeneous system of non-interacting particles, $g_{AB}(r)$ is unity for all r within the system. For real solutions where particles interact, $g_{AB}(r)$ will be complicated function having peaks that indicate ordered “coordination shells” or “solvation shell” (York, 1992; Soper, 1986).

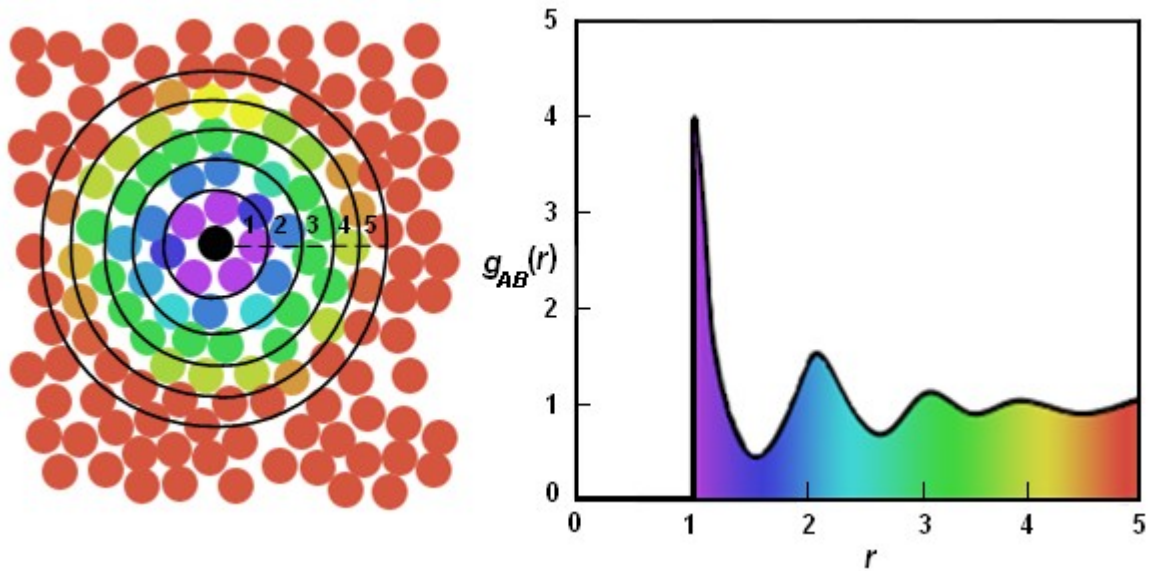


Figure 2.10 Radial distribution function $g_{AB}(r)$. The function $g_{AB}(r)$ (right picture) is calculated based on all pairs of a reference particle of type “B” (a black sphere) and surrounding particles of type “A” (colored spheres). The surrounding particles are colored based on their distance from the black particle

The RDF is dimensionless quantity and useful, not only because it provides insight into the liquid structure, but also because it can be used to calculate several quantities of interest, such as (York, 1992; Allen, 1987):

$$\begin{aligned} \text{CN}^{(n)} &= \rho_A \int_{r_m^{(n-1)}}^{r_m^{(n)}} g_{AB}(r') 4\pi r'^2 dr' \\ &= n_{AB}(r_m^{(n)}) - n_{AB}(r_m^{(n-1)}) \end{aligned} \quad (2.42)$$

and

$$\langle r^{(n)} \rangle = \frac{\int_{r_m^{(n-1)}}^{r_m^{(n)}} r' \frac{g_{AB}(r') r'^2}{\int_{r_m^{(n-1)}}^{r_m^{(n)}} g_{AB}(r'') r''^2 dr''} dr'}{\int_{r_m^{(n-1)}}^{r_m^{(n)}} g_{AB}(r'') r''^2 dr''}, \quad (2.43)$$

where $\text{CN}^{(n)}$ is the coordination number of the n -th peak of $g_{AB}(r)$, ρ_A the density of A particle, $n_{AB}(r_m^{(n)})$ the average number of average number of A particles around B particle within the sphere radius of the n -th peak where $g_{AB}(r)$ goes to minimum, and $\langle r^{(n)} \rangle$ the average distance of a particle in the n -th peak.

Radial distribution function approaches give a one dimensional view of the distribution of the solute/solvent distances. In order to get a more accurate three-dimensional view of the distribution of water molecules around solute atoms, the spatial distribution function, or SDF, which integrates radial and angular coordinates, was developed (Auffinger, 2001; Kulińska, 2000; Kusalik, 1994; Svishchev, 1993; Allen, 1987; Haile, 1980).

2.5.6 Principal component analysis

Principal component analysis is an efficient method to represent the motion of a $3N$ -dimensional system in terms of a few “principal” components (Ichiye, 1991). The basic idea is that the correlated internal motions are represented by the covariance matrix

$$\sigma_{ij} = \left\langle \left(q_i - \langle q_i \rangle \right) \left(q_j - \langle q_j \rangle \right) \right\rangle \quad (2.45)$$

where q_1, \dots, q_{3N} are the mass-weighted Cartesian coordinates of the solute molecule and $\langle \dots \rangle$ denotes the average over all sampled conformations. By diagonalizing σ , $3N$ eigenvectors v_n and eigenvalues λ_n , which are rank-ordered in descending order, i.e., λ_1 represents the largest eigenvalues, are obtained. We may expand the MD trajectory $q(t) = \{q(t)\}$ in the basis of the eigenvectors v_n according to

$$x^{(n)}(t) = \sum_i^n [v_i \cdot q(t)] v_i \quad (2.46)$$

While for $n = 3N$ the expansion becomes exact $[x^{(3N)}(t) = q(t)]$, for small n (in practice, $n = 1-5$) $x^{(n)}(t)$ approximates the motion of the system in terms of a few principal components representing the “essential” dynamics of the system (Amadei, 1993).

Chapter 3

Effect of Electrostatic Treatment and Force Field on the Conformations of Small RNA Systems

Nowadays, molecular dynamic (MD) simulations are regularly used in studies of large and complex biological systems (Norberg, 2003). The relative stability of conformations of the systems is determined by a variety of forces. Electrostatic effects are among the important factors in determining the conformations of the systems (Róg, 2003; Harvey, 1989). In this chapter, different treatments of long-range electrostatic interactions (RF and PME) in MD simulations are investigated. Two ribonucleic acid (RNA) duplexes, 5'-(UAU)-3'·5'-(GUA)-3' and 5'-(CGC)-3'·5'-(GCG)-3', and the uCACGg RNA hairpin are used as test systems (Figure 3.1). In the first section, 10 ns MD simulations on the small RNA dimers give a general idea of the effects of the electrostatic treatment on the stability of the stems. Then we continue the investigation of electrostatic treatment on an uCACGg hairpin, 5'-(gguauCACGguacc)-3' (see Section

3.3). Ion-effects on this hairpin in different ion concentrations and two different ion types are described. Besides the electrostatic treatment and the ion effect, the ability of the three versions of AMBER force field (AMBER94, AMBER98, and AMBER99) to reproduce the structural feature of the uCACGg hairpin is discussed in Section 3.2

3.1 Simulation of two RNA duplexes

Two RNA duplexes are chosen as model systems (Figure 3.1) to study the effect electrostatic treatment. Duplex 1 contains one G·U wobble base-pair, while Duplex 2 includes only the Watson-Crick base-pairs. The G·U wobble base-pair has comparable thermodynamic stability to Watson-Crick base-pairs and also has unique chemical, structural, dynamic and ligand-binding properties (Varani, 2000). The G and U base are

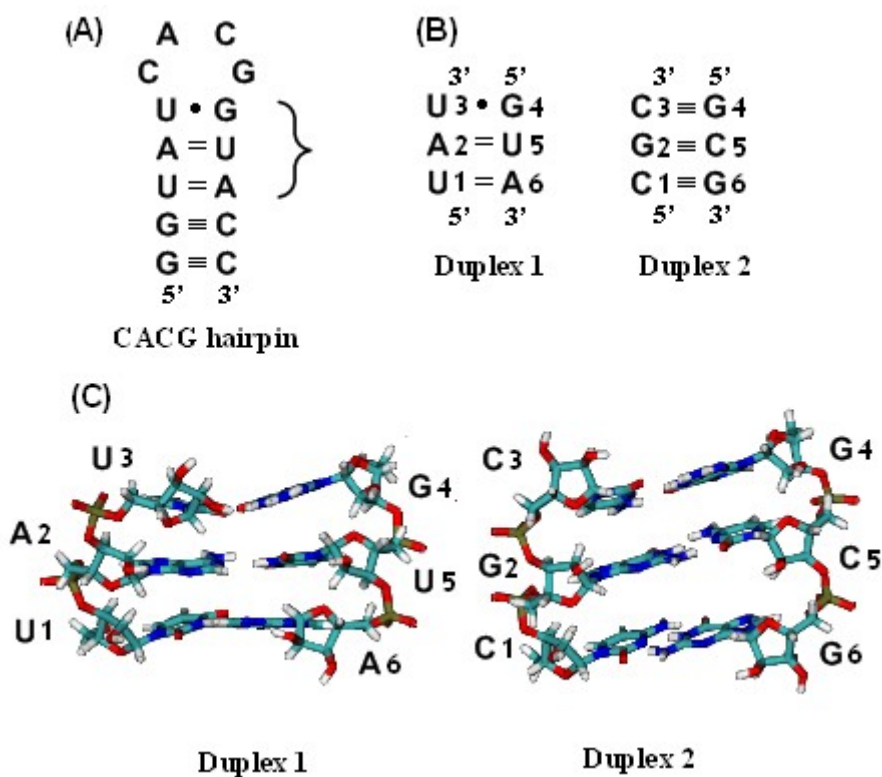


Figure 3.1 The model systems. (A) Secondary structure of uCACGg hairpin and (B) the two duplexes, that are 5'-UAU-3'-5'GUA-' (Duplex 1) and 5'-CGC-3'-5'-CGC-3' (Duplex 2). The numbering scheme corresponds to the respective nucleotide sequences and the hydrogen bonds of the Watson-Crick base-pair are shown with solid lines, which connect the base-pairs, and the wobble base-pair with a point. (C) Energy-minimized RNA duplex starting structures (in stereo) of Duplex 1 (PDB entry 1RFR) and Duplex 2 (Koplin, 2005).

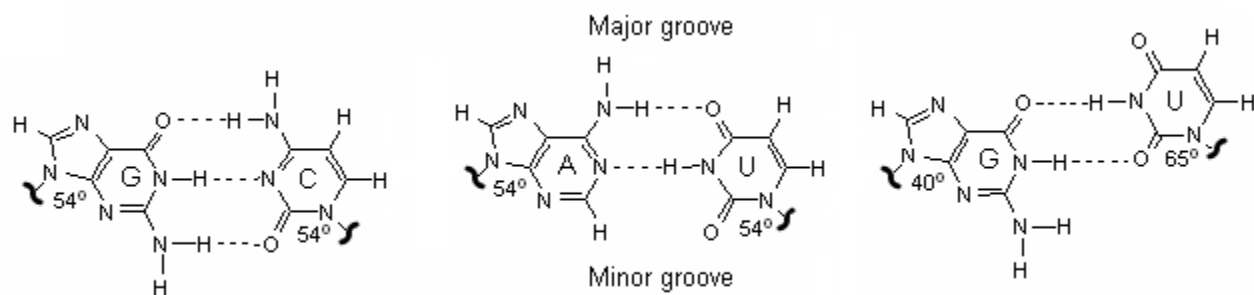


Figure 3.2 Comparison Watson-Crick base-pairs, G·C and A·U, with wobble base-pair. The dashed lines show the hydrogen bonds between the base-pairs. The glycosidic angles, major and minor groove are also illustrated.

able to form two hydrogen bonds by interacting through the same face of base involve in Watson-Crick pairing (Figure 3.2). The Watson-Crick base-pairs differ from the G·U wobble base-pair in type and location of functional groups that are projected into major and minor grooves. Their glycosidic angles also different; while the glycosidic angles of the Watson-Crick base-pairs are similar (about 54°), both angles for G and U differ in the wobble pair.

To our knowledge, there is no experimental value available for two duplexes that are used as model system to study the effect of the electrostatic treatment. This also means that we do not know if these trimers in nature display any duplex formation or not. However, the thermodynamic analysis on the stability of oligonucleotide duplex formation is an interesting subject, and a large number of papers have been published on this aspect (Wu, 2000; Sinclair, 1984; Uesugi, 1984; Westerink, 1984; Uesugi 1982; Borer, 1973). In general, the tetramers and hexamers can form perfect duplexes, and the trimers are observed as dangling (unpaired base in a stem) duplexes.

3.1.1 Computational details

The starting structure of Duplex 1 was taken from model 1 of the NMR structure of the 30-mer stem-loop D of Cocksackieviral RNA, entry 1RFR of the Brookhaven Protein Data Bank, whereas the starting structure of Duplex 2 from Koplín and co workers (Koplín, 2005). The structures were visualized using VMD program (Humphrey, 1990).

The model systems involved explicit all-atom and the interactions between atoms were described using the AMBER94 (Cornel, 1995) and AMBER99 (Wang, 2000) force

Table 3.1 Detail of simulations. RF and PME is defined for reaction field and particle mesh Ewald methods respectively

RNA	Force field	Method	Sodium counterions	r_{elec}	r_{vdw}
Duplex 1	AMBER94	RF	Yes	1.4	1.4
	AMBER99	RF	Yes	1.4	1.4
	AMBER94	PME	Yes	1.0	1.4
	AMBER99	PME	Yes	1.0	1.4
	AMBER94	RF	No	1.4	1.4
Duplex 2	AMBER94	RF	Yes	1.4	1.4
	AMBER99	RF	yes	1.4	1.4
	AMBER94	PME	yes	1.0	1.2
	AMBER94	PME	yes	1.0	1.4
	AMBER94	PME	yes	1.0	1.6
	AMBER94	PME	yes	1.0	1.8
	AMBER94	PME	yes	1.2	1.2
	AMBER94	PME	yes	1.2	1.4
	AMBER94	PME	yes	1.4	1.4
	AMBER94	PME	yes	1.2	1.6
	AMBER94	PME	yes	1.4	1.6
	AMBER99	PME	yes	1.0	1.4
	AMBER94	RF	no	1.4	1.4
	AMBER99	RF	no	1.4	1.4

field. The GROMACS suite of programs (Berendsen, 1995b; Lindahl, 2001) was used to simulate the model systems for 10 ns. The simulations were carried out in the NPT ensemble with periodic boundary conditions, a temperature of 300 K and at pressure 1 atm. The temperature and the pressure were kept constant by Berendsen's weak coupling scheme (Berendsen, 1984) with coupling parameter of 0.1 ps for temperature, 1.0 ps for pressure, and the isothermal compressibility of $4.5 \cdot 10^{-5} \text{ bar}^{-1}$. The duplex was solvated with about 2700 water molecules used water model TIP3P (Jorgensen, 1983) in a rhombic dodecahedron box with minimum periodic distance of about 2.85 nm. Four sodium counterions were added to the system to compensate for the four negative charges on the phosphate groups, except where notified for the simulations which did not require electroneutrality. The equations of motion were integrated using leapfrog scheme (Snyman, 1982) with a time step of 2 fs. Constraint were used for bond lengths using the

LINCS algorithm for the duplex (Hess, 1997), and SETTLE for the water (Miyamoto, 1992).

A twin-range cut off is used for van der Waals interactions. The interactions within the short cutoff of 1.0 nm are evaluated at every time step, while the interactions between atoms separated by a distance ranging between the short and the long cutoff (1.2, 1.4, 1.6, 1.8 nm) were evaluated every tenth step. The electrostatic interactions were evaluated by using two different methods: the cutoff including reaction field correction and particle mesh Ewald (PME) method. The cutoff used with reaction field (RF) correction was 1.4 nm. The settings for the PME method were as follows: a real space cutoff of 1.0, 1.2, and 1.4 nm, a grid spacing of 0.12 nm. The details of simulations are shown in Table 3.1.

3.1.2 Effect of electrostatic treatments on duplexes

The objective of this study is to investigate the influence of two electrostatic treatments on the structural stability of two duplexes. The stability of the duplexes was estimated from their ability to form a duplex (double strands) during the time of simulation. To probe the stability of duplexes during simulation runs, we calculated the hydrogen bonds between pairs for each duplex and the result of 19 simulations are summarized in Table 3.2. A hydrogen bond is considered to exist if the distance between the acceptor and donor atom is less than 0.35 nm and the angle donor-hydrogen-acceptor is larger than 135° (Figure 3.3).

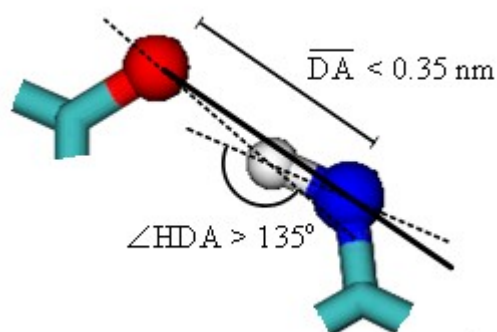


Figure 3.3 Hydrogen bond criteria. Red sphere represents acceptor atom (A), white sphere hydrogen atom (H), and blue sphere donor atom (D).

Table 3.2 The structural stability of Duplex 1 and Duplex 2 during 10 ns.

Model	Type	RF	PME	RF without counterions
Duplex 1	Total number of simulations	2	2	1
	% duplex	0	0	0
	% single strand	50	100	100
Duplex 2	Total number of simulations	2	10	2
	% duplex	0	40	50
	% single strand	50	0	0

Both Duplex 1 and 2 were built up by three base-pairs between two strands. If the populations of all hydrogen bonds of the two strands were more than 80%, we assumed that the two strands remain a duplex until the end of the simulation. Moreover, if the populations of all hydrogen bonds of the two strands were less than 80%, the duplex built up two single strands and was not a duplex anymore and we called it single strand. If the summation of percentage of the duplex and the single strand was not 100 %, that meant another possibilities could be observed; for example, only one or two base-pair of three base-pairs are staying together during the simulations.

As shown in Table 3.2 almost all the simulations of Duplex 1 are broken, except one simulation using reaction field, where the hydrogen bond of the Hoogsteen base-pair between residue U1 and A6 are still observed. The Duplex 2 showed significantly higher stability than Duplex 1. Five simulations of 14 were showing a duplex; four using PME and the RF simulation without ions. The difference between the two electrostatic treatments is observed through the calculation of their radius of gyration (in Figure 3.4). The radii of gyration of the RF simulations without counterions have always larger values than those with PME. This indicates an expansion of the duplex in absence of ions.

Counterions influence the stabilization of a duplex by reducing the repulsion between negative charged RNA phosphates.

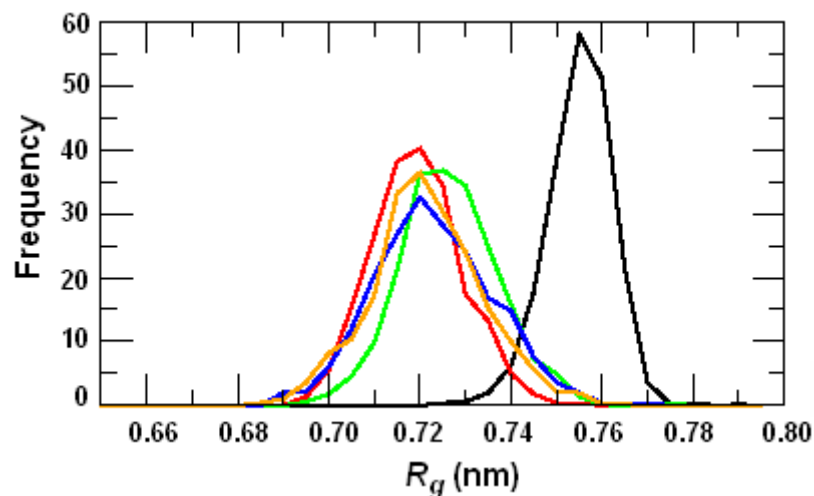


Figure 3.4 Radius of gyration (R_g) distributions for Duplex 2 are given over a 10 ns molecular dynamics simulation. Black line is the simulation that has no counterions using reaction field, the others (red, blue, green, orange) are the simulations using PME.

3.1.3 Conclusion

The effect of electrostatic treatments on the structural stability of RNA is studied by MD simulations of two model systems namely Duplex 1 and 2. The results of this study show that both treatments of electrostatic interactions (reaction field and PME method) indicate a more pronounced destabilization of Duplex 1 than of Duplex 2. In the case of Duplex 2 a stronger stabilization effect is observed using PME method. Because the simulated systems contain a relatively large number of charged residues, the use of the PME method is expected to give a better representation of the electrostatic interactions. However, it has to be taken into consideration that there is no available experiment for our model system.

3.2 Different AMBER force fields in the investigation of uCACGg hairpin

A number of force fields can simulate RNAs as discussed in Chapter 2. The one used in this thesis is the AMBER force field. Continuing development of the AMBER force field is needed to be adjusted to improve an agreement between experiment and simulation. In this thesis, the performance of different AMBER force fields are compared in reproducing the structure of uCACGg hairpin (Figure 3.5) in 60 ns MD simulations. These force fields, based on the Cornell and co workers, force field are the AMBER94 (Cornell, 1995), AMBER98 (Cheatham, 1999), and AMBER99 (Wang, 2000) force fields.

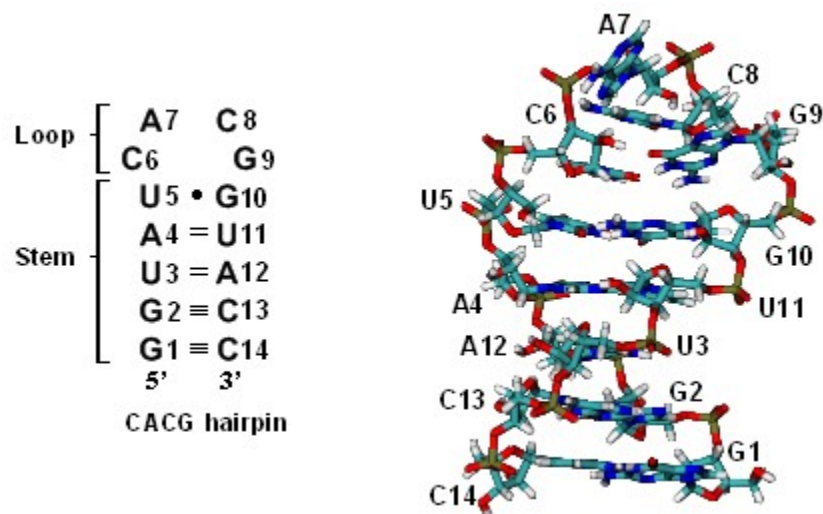


Figure 3.5 The 14-mer uCACGg hairpin which consists of a stem and a loop region with numbering. (A) Secondary and (B) tertiary structure. The lines, which connect two bases, represent the hydrogen bonds between Watson-Crick base-pairs, and the black point wobble base-pair.

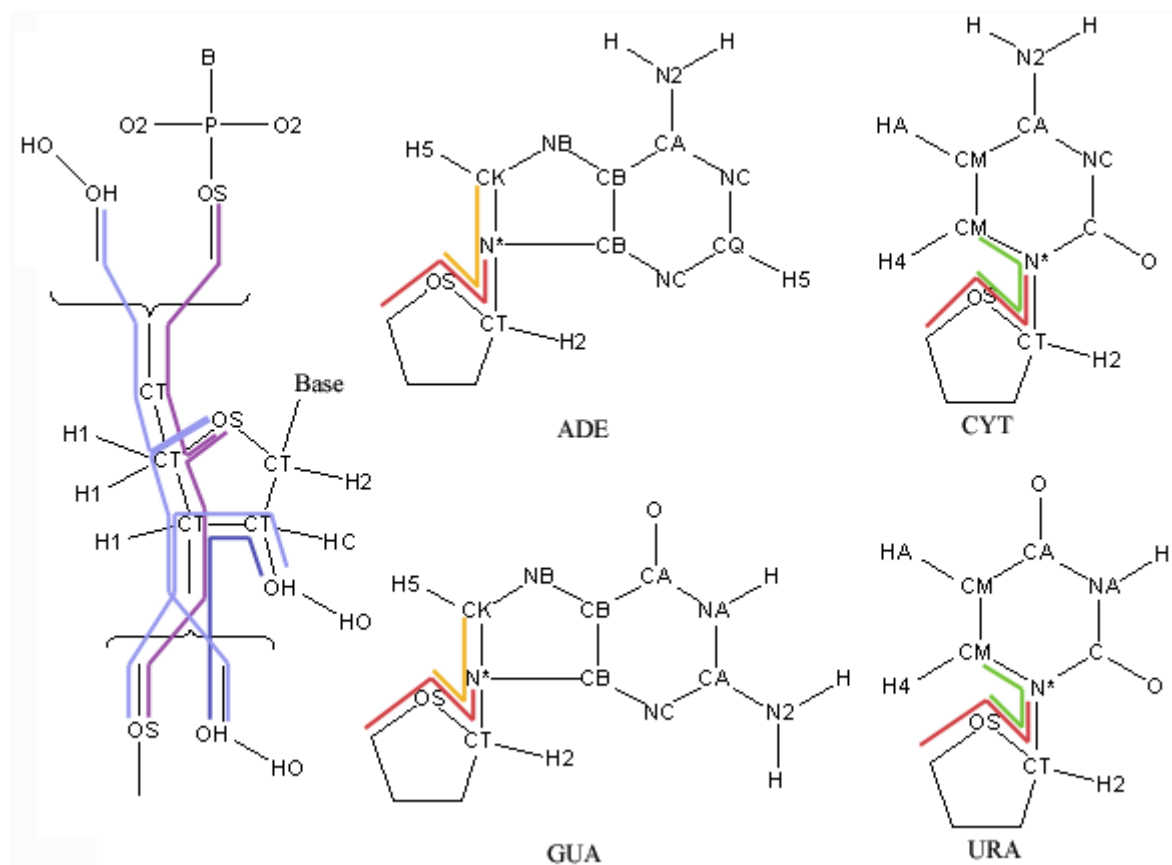


Figure 3.6 Comparison between the AMBER94 and AMBER98 force fields for RNA. All atoms are described in atom types. The colorful lines represent the differences between two force field; \color{purple} : OS-CT-CT-OS; \color{blue} : OH-CT-CT-OH; $\color{lightblue}$: OH-CT-CT-OS; \color{red} : CT-OS-CT-N*; \color{yellow} : OS-CT-N*-CK; \color{green} : OS-CT-N*-CM.

The AMBER94 force field was modified to the AMBER98 force field by using of high level *ab initio* calculations on entire nucleosides (Cheatham, 1999) in contrast to smaller model systems necessitated in 1994-95 by computer limitation (Cornell, 1995). The only differences between the two force fields are the torsion parameters for the sugar phosphate backbone and the glycosidic χ angle (Figure 3.6).

Another new version of the AMBER force field, which is developed also from AMBER94, is the AMBER99 force field. We can see the differences between the force field versions in Figure 3.7.

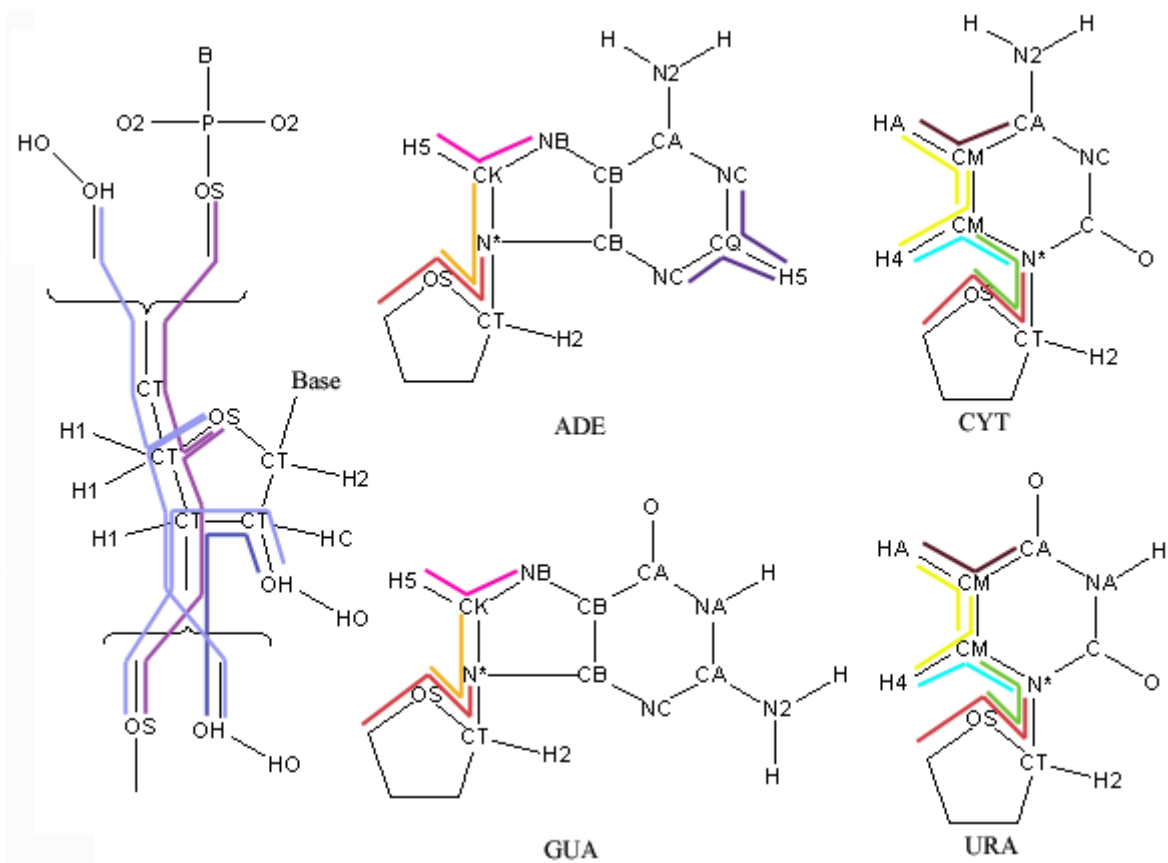


Figure 3.7 Comparison between the AMBER94 and AMBER99 force fields for RNA. All atoms are described in atom types. The colorful lines represent the differences between two force field; ●: OS-CT-CT-OS; ●: OH-CT-CT-OH; ●: OH-CT-CT-OS; ●: CT-OS-CT-N*; ●: OS-CT-N*-CK; ●: OS-CT-N*-CM; ●: H5-CK-NB; ●: HA-CM-CA; ●: HA-CM-CM; ●: H4-CM-N*; ●: NC-CQ-H5.

AMBER98 and AMBER99 use the atom types and topologies (except for torsional parameters) from AMBER94. AMBER98 is similar to AMBER99 except to additional changes in the torsion angle parameter. The comparison of the differences between AMBER94, AMBER98, and AMBER99 are summarized in Table 3.3 and 3.4.

In this study, the reliability of simulations are evaluated by analyzing the NOE (nuclear Overhauser effect) distances, RMSD, torsional angles of backbones, glycosidic torsional angles, sugar pucker, radius of gyration and total solvent accessible surface area. The results are compared with experimental data (Ohlenschläger, 2004). Another ways to obtain more insight into the structural characteristic are calculating the fraction of the hydrogen bonds and the base stacking.

Table 3.3 Comparison of differences in the angle potential between AMBER94, AMBER98 and AMBER99. Here we give the parameter values based on equation 2.3

	AMBER94		AMBER98		AMBER99	
	K_θ	θ_{eq}	K_θ	θ_{eq}	K_θ	θ_{eq}
H5-CK-NB	35.0	123.050	35.0	123.050	50.0	123.05
H5-CK-N*	35.0	123.050	35.0	123.050	50.0	123.05
HA-CM-CA	35.0	123.300	35.0	123.300	50.0	123.300
HA-CM-CM	35.0	119.700	35.0	119.700	50.0	119.700
H4-CM-N*	35.0	119.100	35.0	119.100	50.0	119.100
NC-CQ-H5	35.0	115.45	35.0	115.45	50.0	115.45

Table 3.4 Comparison of differences in the torsional potential between AMBER94, AMBER98 and AMBER99. Here we give the parameter values based on equation 2.4

	n	AMBER94		AMBER98		AMBER99	
		K_ϕ (kcal/mol)	γ (°)	K_ϕ (kcal/mol)	γ (°)	K_ϕ (kcal/mol)	γ (°)
OS-CT-CT-OS	3	0.144	0.0	0.144	0.0	0.144	0.0
	2	1.0	0.0	1.5	0.0	1.175	0.0
OH-CT-CT-OH	3	0.144	0.0	0.144	0.0	0.144	0.0
	2	1.0	0.0	1.5	0.0	1.175	0.0
OH-CT-CT-OS	3	0.144	0.0	0.144	0.0	0.144	0.0
	2	1.0	0.0	1.5	0.0	1.175	0.0
CT-OS-CT-N*	3	1.15	0.0	0.383	0.0	0.383	0.0
	2	0.0	0.0	1.0	0.0	0.65	0.0
OS-CT-N*-CK	2	0.5	180.0	0.0	0.0	0.0	0.0
	1	2.5	180.0	2.5	0.0	2.5	0.0
OS-CT-N*-CM	2	0.5	180.0	0.0	0.0	0.0	0.0
	1	2.5	180.0	2.5	0.0	2.5	0.0

The N-H \cdots N hydrogen bond of Watson-Crick base-pair, the hydrogen bonds of wobble base-pair, and the hydrogen bond (G9) N₁-H₁ \cdots O₂(C6) of the loop region were taken as descriptors for the structure definition of the hairpin. A total number of seven hydrogen bonds were used for the uCACGg hairpin. Beside hydrogen bonds, the base stacking interactions in the successive strand, the U5-C6 base stacking in the stem-loop junction, and the C6-C8 base stacking in the loop region, or a total number of 10 base

stacking, were taken into account for structural information of the uCACGg hairpin. Two bases were considered as stacked when their center of mass separation within 0.15 nm of that seen in the simulated structure at 300 K. In addition, the flexibility of the hairpin was assessed by calculating RMS fluctuations (RMSFs) about the average structure.

3.2.1 Computational details

The starting structure of the investigation on uCACGg hairpin with AMBER94, AMBER98, and AMBER99 is the upper part of the NMR structure of the 30-mer stemloop-D of coxsackieviral RNA of PDB 1RFR in Brookhaven Protein Data Bank (Ohlenschläger, 2004). The hairpin was placed in a rhombic dodecahedron box with minimum periodic distance of about 2.94 nm and filled up with about 4995 TIP3P water molecules (Jorgensen, 1983) and 13 Na⁺ counterions. All MD simulations were performed using the GROMACS package of programs (Berendsen, 1995b; Lindahl, 2001) and force field AMBER94 (Cornell, 1995), AMBER98 (Cheatham, 1999) and AMBER99 (Wang, 2000) in the NPT ensemble with periodic boundary condition up to 60 ns simulation time at a temperature of 300 K. The temperature was maintained close to the intended value by weak coupling of the hairpin, water and counterions separately to an external temperature bath (Berendsen, 1984) with coupling parameter of 0.1 ps. The pressure of the system was weakly coupled to the pressure bath of 1 bar (Berendsen, 1984) with a relaxation time of 0.5 ps and an isothermal compressibility of $4.5 \cdot 10^{-5} \text{ bar}^{-1}$. Bond lengths were constrained by using the LINCS algorithm (Hess, 1997) and SETTLE (Miyamoto, 1992) for the hairpin and water respectively. The time step for the leapfrog integration scheme was set to 2 fs. The van der Waals interactions were evaluated by means of a twin-range method; the short-range of the van der Waals interactions were evaluated at every time step by using a charge-group pair list that was generated with a short-range cutoff radius of 1.0 nm. Long-range van der Waals interactions (between charge groups at a distance longer than the short-range cutoff and shorter than a long-range cutoff of 1.4 nm) were evaluated every tenth step. For electrostatic treatment using PME, the real-space cutoff was set to 1.0 nm and the grid spacing 0.12 nm. The 13 sodium counterions were placed by replacing randomly water molecules to obtain a neutralized system.

3.2.2 AMBER94, AMBER98 and AMBER99 versus experimental data

Table 3.5 shows some structural characteristics (NOE violation, RMSD, fraction of hydrogen bonds, and fraction of base stacking) of uCACGg hairpin obtained from the MD simulations using AMBER94, AMBER98, and AMBER99 force fields. The interproton distances corresponding to the 521 NOE interaction were calculated from the 60 ns trajectory and averaged as $\langle r^{-6} \rangle^{-1/6}$ for comparison with experimentally determined distance upper limits of NOE's (Ohlenschläger, 2004; the NOE data are given kindly by Jens Wöhnert).

The NOE violations were those, whose computed range values (average \pm standard deviation) are larger than NOE upper limit. It is to note that the experiments are carried out for 30-mer uCACGg hairpin at 283 K and the MD simulations were performed for 14-mer uCACGg hairpin at 300 K. The MD simulations using the three AMBER force fields are in overall agreement with the experimental structure (Ohlenschläger, 2004). Specific structural features of NOE violations are shown in Table 3.6 and Figure 3.7

Table 3.5 Structural characteristic of MD simulations using AMBER94, AMBER98, and AMBER99 force fields

	AMBER94	AMBER98	AMBER99
NOE violations (%)	7.2	6.6	6.4
RMSD (nm)	0.22	0.20	0.23
P_H **	0.993	0.994	0.954
P_S ***	0.999	0.998	0.997

**) P_H = fraction of hydrogen bonds between stem base-pairs, plus loop base-pair (N₁/G9:O₂/C6)

***) P_S = fraction of intrastrand base-base stacking interaction

Table 3.6 NOE violations using AMBER94, AMBER98, AMBER99 for uCACGg hairpin

	Total	Intraresidual		Intrastrand		Interstrand	
		Stem	Loop	Stem	Loop	Stem	Loop
Number of NOE distance constraints	521	64	111	147	119	53	27
Number of violations							
AMBER94	37 (7.1 %)	2	3	6	19	1	6
AMBER98	34 (6.5 %)	1	3	7	17	1	5
AMBER99	33 (6.3 %)	0	2	12	14	1	4

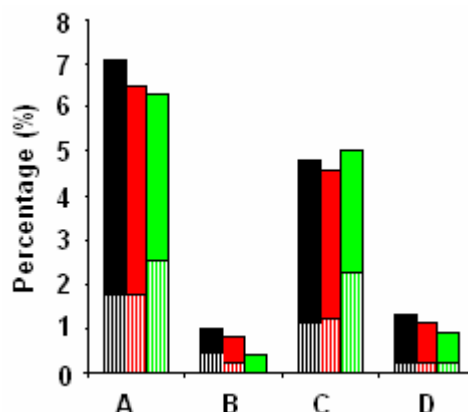


Figure 3.8 NOE violations in percentage using AMBER94, AMBER98, AMBER99 for uCACGg hairpin are shown in black, red, and green columns respectively; A is the total violation of the hairpin, B the violations of intraresidual, C the violations of intrastrand and D the violations of interstrand. The lower parts of the histogram represented by vertical lines show the violations in the stem.

Table 3.7 Average RMSDs of the MD trajectory using AMBER94, AMBER98, and AMBER99 relative to the initial structure and to the average structures of NMR structures

RMSD	MD to starting structure (nm)			MD to NMR structure ^a (nm)			MD to NMR structure ^b (nm)		
	RNA	Stem	Loop	RNA	Stem	Loop	RN A	Stem	Loop
AMBER94	0.22	0.19	0.14	0.22	0.19	0.13	0.20	0.17	0.16
AMBER98	0.20	0.18	0.14	0.21	0.17	0.13	0.18	0.14	0.16
AMBER99	0.23	0.20	0.15	0.24	0.21	0.14	0.21	0.17	0.18

^aRMSD of the trajectory to the average structure of 20 NMR structures from PDB code: 1RFR (Ohlenschläger, 2004).

^bRMSD of the trajectory to the average structure of 10 NMR structures from PDB code: 1ROQ (Du, 2003).

In Table 3.7, the average all-atom RMSDs of trajectory from the starting structure are calculated after fitting on the backbone atoms. During the equilibration period the simulated molecule deviates modestly from the starting NMR structure (model 1 of PDB 1RFR from Ohlenschläger, 2004). We obtain similar deviation from experimental structure using all of the three AMBER force fields. The average RMSDs fitted only on backbone atoms of stem and loop show low values (Table 3.7). The average RMSDs of the fitted backbone atoms of the sub region (loop and stem region) are lower than the average RMSDs of the whole hairpin. This indicates that the local structures are better maintained over the course of the simulation than the global structure.

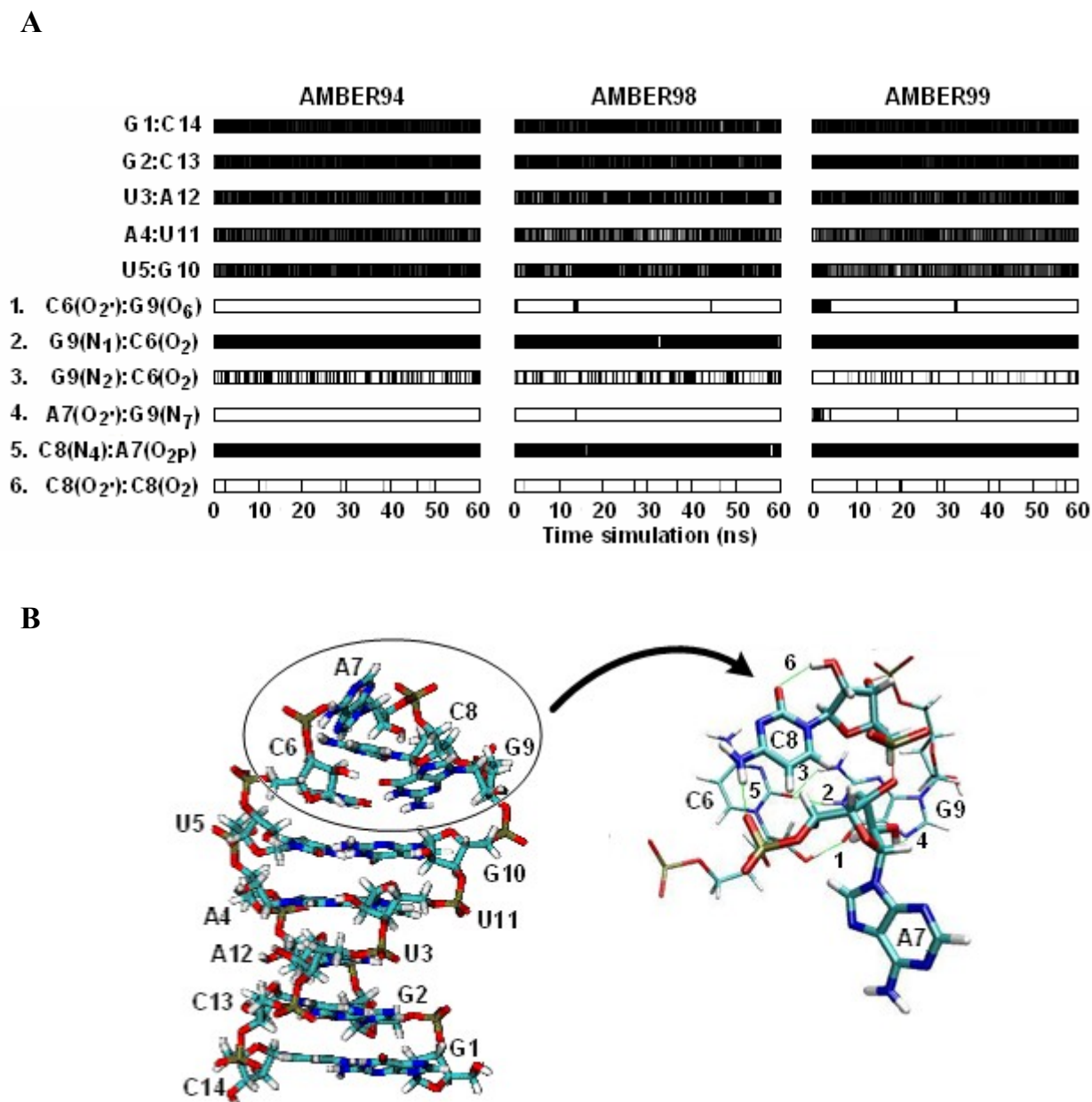


Figure 3.9 Hydrogen bonds of uCACGg hairpin. (A) Time evolution of the occurrence of base-pairing in stem and hydrogen bonds in the loop of uCACGg hairpin as a function of time using AMBER94 (left), AMBER98 (middle), and AMBER99 (right) force fields. The base-pairing in the stem (the first five graphs) may be complete (●), partial (◐) and not present (○). The hydrogen bonding described in donor-acceptor distances in the loop and labeled by a number (the last six graphs) may be present (●), and not present (○) and are labeled. (B) The tertiary structure of the simulated uCACGg hairpin. The hydrogen bonds in the loop region are depicted and labeled according to (A).

Table 3.8 Comparison of hydrogen bonds in the loop of uCACGg hairpin for the experimental data (NMR structure) and MD simulations using three AMBER force fields.

H-bonds	Experimental Data		MD simulations		
	1RFR ^a	1ROQ ^b	AMBER94	AMBER98	AMBER99
1. C6(O _{2'}):G9(O ₆)	yes	yes	no	no	no
2. G9(N ₁):C6(O ₂)	yes	yes	yes	yes	yes
3. G9(N ₂):C6(O ₂)	yes	yes	weak	weak	weak
4. A7(O _{2'}):G9(N ₇)	no	yes	no	no	no
5. C8(N ₄):A7(O _{2P})	yes	yes	yes	yes	yes
6. C8(O _{2'}):C8(O ₂)	no	yes	weak	weak	weak

^aOhlenschläger, 2004

^bDu, 2003

The trend is similar for all three force fields. One can notice that even though the flexibility of the nucleotide A7 looped out into the solvent, the RMSDs of the loop region are lower than the RMSDs of the stem, except for the RMSDs relative to 1ROQ average structure.

More details about the hydrogen bonds as a function of the simulation time are shown in Figure 3.9. Hydrogen bonds between stem base-pairs were observed over the course of the whole simulations, but this is not the case for the hydrogen bonds in the loop. Oliver Ohlenschläger and co workers (Ohlenschläger, 2004) have proposed four hydrogen bonds observed in the loop and Zhihua Du and co workers (Du, 2003) six hydrogen bonds. However, we found from our simulations two strong hydrogen bonds and two weak ones (Figure 3.9 and Table 3.8).

Hydrogen bond interactions involving the ribose 2'-hydroxyl group are supposed to play important role in tetraloop. In fact in NMR structures of CACG loop (Ohlenschläger, 2004; Du, 2003) one/three of the loop hydrogen bonds are involving the hydroxyl group. In the simulations, only two of these hydrogen bonds were observed over the course of the 60 ns simulations. Furthermore, the results from the simulations indicate that the 2'-hydroxyl of C6 could possibly form a hydrogen bond with O_{5q} of A7 and the 2'-hydroxyl of A7 with O_{3q} of A7.

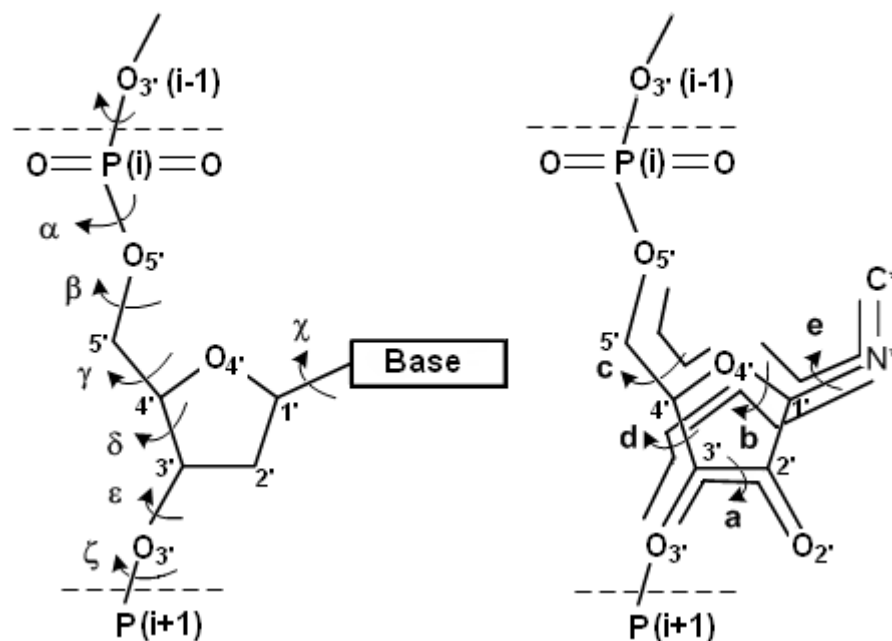
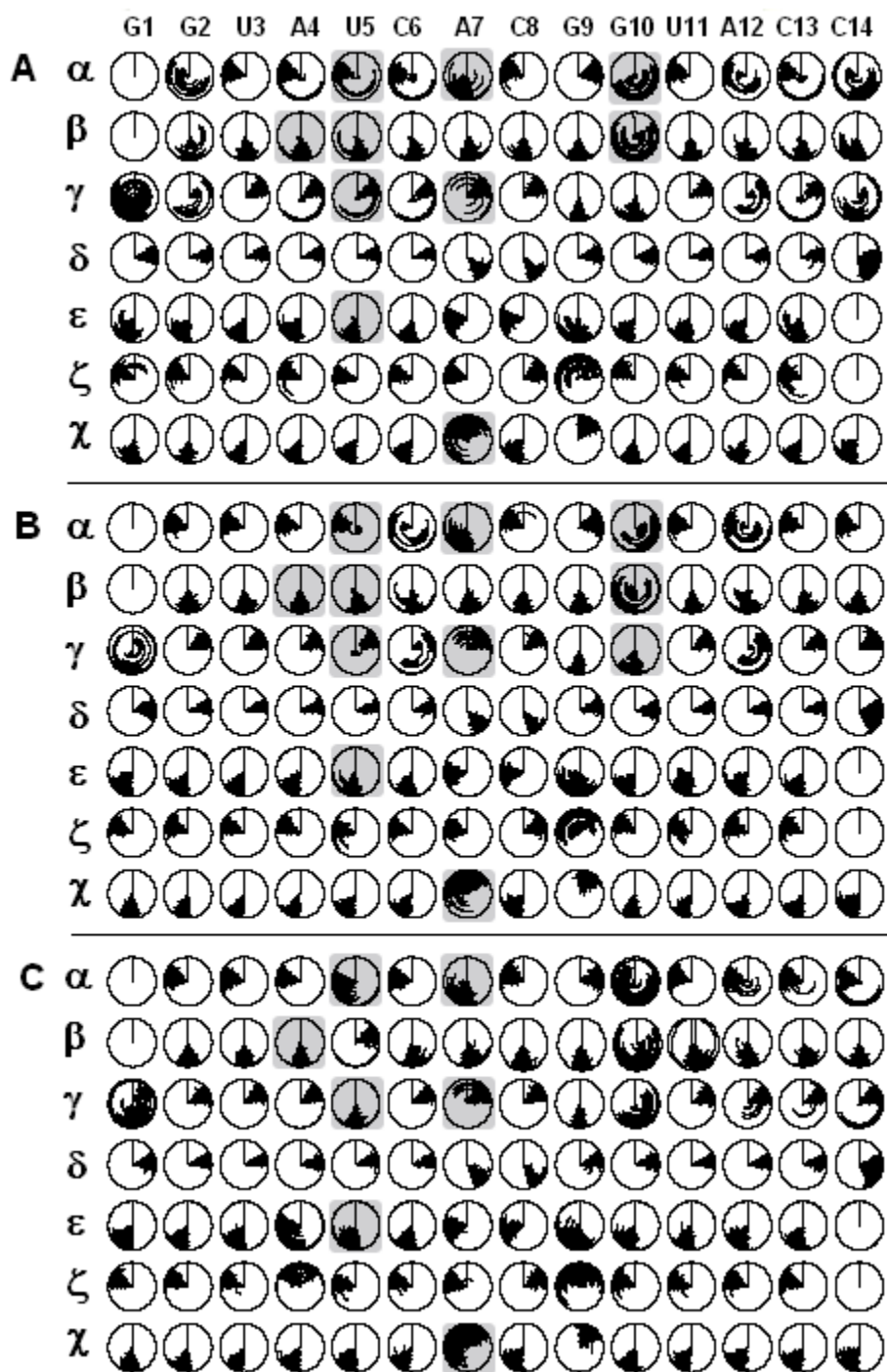


Figure 3.10 Description of backbone (α , β , γ , δ , ϵ , ζ), glycosidic (χ) and certain (a , b , c , d , e) torsional angles for uCACGg hairpin. The equivalent torsionals γ to c , δ to d , and χ to e are shown. The star symbol (*) denotes the atom type $N=N_9$ and $C=C_8$ for purine bases, and $N=N_1$ and $C=C_5$ for pyrimidine bases.

The three AMBER force fields differ in their sugar pucker parameters and the glycosidic torsional angle parameters (Figure 3.10). Hence, it is instructive to compare the dihedral angles of the backbone atoms, and glycosidic torsional angle of the hairpin between the simulations performed using three force fields and between the MD simulations and experiments. The NMR structures of the uCACGg hairpin were determined and deposited as 20 structures of 30-mer and 10 structures of 14-mer RNA hairpin in PDB code 1RFR (Ohlenschläger, 2004) and 1ROQ (Du, 2003) respectively. The two NMR structures have similarity in sugar pucker (δ angle in Figure 3.11 and Table 3.9, and a , b and d angles in Figure 3.12 and Table 3.10), glycosidic torsional angle (χ angles in Figure 3.11 and Table 3.9, and e angles in Figure 3.12 and Table 3.10), and torsional angle of backbone atoms (α , β , γ , ξ , and ζ angles in Figure 3.11 and Table 3.10), except for ζ of the residue C8, and α , β , and γ of the residue G9 and G10.



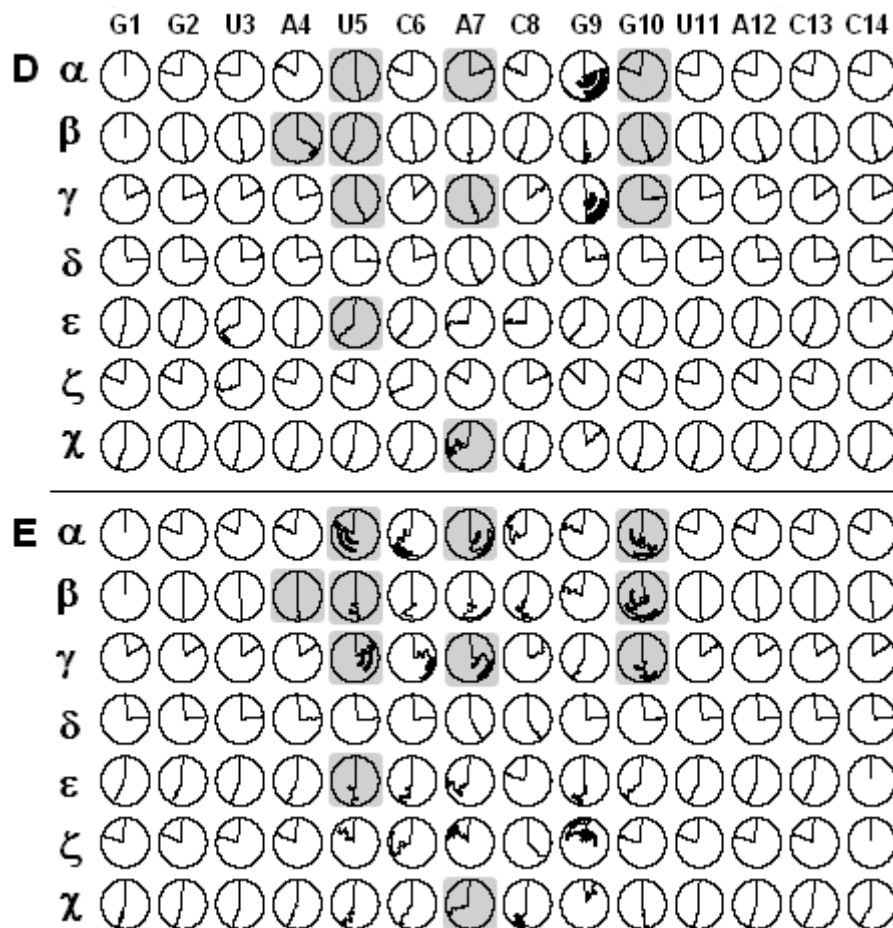


Figure 3.11 Time course of the torsional angles α , β , γ , δ , ϵ , ζ , and the glycosidic angles χ for the 14-mer uCACGg hairpin over 60 ns. The trajectories are displayed in a clockwise direction between 0 and 360° using (A) AMBER94, (B) AMBER98, and (C) AMBER99, and from PDB: (D) 1RFR, and (E) 1ROQ.

The backbone dihedral transitions are observed especially around residues G9 (ϵ , ζ) and G10 (α , β , γ) at the loop/stem junction for the all three force fields. At the other side of the loop/stem junction U5 (ϵ , ζ) to C6 (α , β , γ), the same transitions are observed, except when AMBER99 is used. The agreement of dihedral angles between the simulations and the experimental data in the loop is better with the 1ROQ than with the 1RFR, even if the starting structure is 1RFR (Ohlenschläger, 2004). The glycosidic torsional angles χ of the looped out nucleotide A7 showed the highest fluctuation.

Table 3.9 Backbone and glycosidic torsional angles for the CACG residues from simulation and experiment: average plus standard deviation. The deviations from experiment values are shown in the bold text. A value with a star means the torsional has more than one values.

	α	β	γ	δ	ϵ	ζ	χ
G1 _{AMBER94}			242±273*	80±8	-150±24*	-63±15*	-169±10
G1 _{AMBER98}			55±151*	81±8	-152±11	-66±8	-171±9
G1 _{AMBER99}			109±117*	80±7	-153±10	-65±8	-171±8
G1 _{IRFR}			66±1	83±0	-163±0	-67±0	-158±1
G1 _{IROQ}			54±1	84±0	-150±0	-74±0	-161±2
G2 _{AMBER94}	-103±58*	164±38*	102±58*	79±7	-153±10	-65±8	-167±8
G2 _{AMBER98}	-71±9	175±9	60±8	79±6	-153±9	-65±8	-165±7
G2 _{AMBER99}	-72±9	176±10	62±8	78±6	-153±9	-63±8	-165±7
G2 _{IRFR}	-78±0	169±0	72±1	84±0	-160±0	-65±1	-164±0
G2 _{IROQ}	-70±0	106±142	56±1	83±0	-151±1	-67±0	-160±1
U3 _{AMBER94}	-73±10	177±10	60±8	76±6	-156±10	-67±8	-157±8
U3 _{AMBER98}	-72±9	175±9	59±8	76±6	-156±9	-67±8	-159±7
U3 _{AMBER99}	-73±9	177±9	59±8	76±6	-153±10	-69±8	-157±8
U3 _{IRFR}	-80±1	171±2	67±1	80±1	-118±10	-107±7	-163±1
U3 _{IROQ}	-67±1	174±0	57±0	83±0	-155±0	-73±1	-160±1
A4 _{AMBER94}	-85±38*	177±10	69±33*	78±7	-153±13	-65±10	-156±9
A4 _{AMBER98}	-73±9	178±9	59±8	78±6	-152±11	-64±9	-156±8
A4 _{AMBER99}	-75±17	177±8	58±15	84±6	-102±28*	-39±22*	-155±9
A4 _{IRFR}	-60±2	122±6	74±0	78±1	-174±1	-74±0	-161±0
A4 _{IROQ}	-68±2	141±107	57±0	86±1	-153±2	-69±2	-154±1
U5 _{AMBER94}	-108±63*	176±14	90±53*	78±6	-168±10	-79±9	-152±9
U5 _{AMBER98}	-112±66*	178±13	93±58*	79±6	-168±10	-77±9	-151±10
U5 _{AMBER99}	-97±32	77±33*	178±25*	76±7	-167±10	-79±9	-154±11
U5 _{IRFR}	166±0	-151±0	151±0	89±1	-132±2	-65±1	-156±0
U5 _{IROQ}	-15±93*	104±130*	73±37*	87±0	-137±103	-49±10	-152±8
C6 _{AMBER94}	-97±47*	174±11	80±41*	79±6	-170±8	-73±9	-149±9
C6 _{AMBER98}	-123±67*	179±13	102±59*	80±6	-167±8	-73±8	-148±11
C6 _{AMBER99}	-78±10	176±10	61±8	80±6	-165±9	-73±13	-143±11
C6 _{IRFR}	-69±0	171±0	46±1	74±1	-137±2	-112±1	-153±0
C6 _{IROQ}	-85±87*	-135±99*	87±31*	85±0	-158±9	-105±19	-144±5
A7 _{AMBER94}	-159±15	171±11	61±12	137±9	-96±11	-75±9	-32±40*
A7 _{AMBER98}	-160±15	172±11	58±11	140±8	-96±11	-75±9	-33±41*
A7 _{AMBER99}	-164±15	169±12	59±11	138±9	-94±12	-74±10	-37±44*
A7 _{IRFR}	69±1	179±3	155±4	157±1	-90±4	-59±1	-90±17
A7 _{IROQ}	78±133*	27±159*	99±30*	147±0	-116±20*	-54±14	-107±4

Table 3.9 Backbone and glycosidic torsional angles for the CACG residues from simulation and experiment: average plus standard deviation. The deviations from experiment values are shown in the bold text. A value with a star means the torsional has more than one values. (cont.)

	α	β	γ	δ	ϵ	ζ	χ
C8 _{AMBER94}	-66±10	-177±9	55±8	147±7	-96±9	66±9	-149±11
C8 _{AMBER98}	-64±10	-176±9	54±8	149±6	-96±10	66±10	-148±11
C8 _{AMBER99}	-64±10	-178±10	55±9	146±7	-99±11	65±10	-145±12
C8 _{IRFR}	-67±3	-159±2	53±3	157±3	-88±5	65±1	-162±4
C8 _{IROQ}	-85±26*	-89±131*	63±12	147±2	-74±1	132±2	-124±100*
G9 _{AMBER94}	75±10	-177±9	-177±7	76±7	-168±21*	15±52*	32±11
G9 _{AMBER98}	77±24	179±9	180±23	75±7	-174±25*	3±53*	32±13
G9 _{AMBER99}	75±15	-179±9	-179±14	79±10	-150±37*	0±53*	38±15
G9 _{IRFR}	18±108*	-175±8	130±55*	80±5	-137±2	-48±1	50±2
G9 _{IROQ}	-65±4	-70±5	-148±4	82±0	-62±153*	-6±58*	30±6
G10 _{AMBER94}	-166±62*	142±71*	170±34*	81±7	-150±9	-59±8	-168±8
G10 _{AMBER98}	139±56*	-177±64*	-170±13	84±8	-150±10	-60±8	-166±8
G10 _{AMBER99}	-174±153*	134±58*	150±51*	81±7	-154±11	-65±9	-159±10
G10 _{IRFR}	-66±1	157±2	86±1	84±0	-164±2	-63±0	-159±2
G10 _{IROQ}	35±140*	-64±112*	167±22*	87±0	-129±4	-77±3	169±4
U11 _{AMBER94}	-74±10	174±8	61±8	77±6	-160±11	-70±8	-154±8
U11 _{AMBER98}	-73±10	175±8	59±8	78±6	-161±12	-71±9	-153±8
U11 _{AMBER99}	-74±10	171±8	62±8	78±6	-163±10	-72±9	-150±9
U11 _{IRFR}	-77±0	168±1	74±2	78±0	-150±1	-76±0	-158±1
U11 _{IROQ}	-70±0	174±0	54±2	86±0	-151±0	-72±1	-165±1
A12 _{AMBER94}	-107±59*	177±12	89±51*	80±7	-156±12	-65±9	-156±10
A12 _{AMBER98}	-130±71*	-179±14	107±61*	82±7	-151±11	-66±9	-157±11
A12 _{AMBER99}	-88±39*	174±12	72±32*	81±6	-155±11	-67±9	-144±11
A12 _{IRFR}	-73±0	159±1	66±0	83±1	-163±1	-57±0	-155±1
A12 _{IROQ}	-67±1	141±106	56±0	85±0	-155±1	-73±0	-160±1
C13 _{AMBER94}	-111±63*	179±13	95±56*	79±7	-164±14	-71±11	-156±10
C13 _{AMBER98}	-72±9	172±9	59±8	79±6	-157±9	-67±8	-156±9
C13 _{AMBER99}	-80±27*	171±10	66±24*	80±6	-159±11	-69±9	-149±10
C13 _{IRFR}	-68±1	173±0	58±0	82±0	-153±1	-70±0	-160±0
C13 _{IROQ}	-70±0	-179±0	55±1	84±0	-151±0	-67±0	-152±0
C14 _{AMBER94}	-135±65*	-174±13	122±62*	123±30*	-146±12		
C14 _{AMBER98}	-74±10	180±9	58±9	91±25*	-148±12		
C14 _{AMBER99}	-92±42*	-178±11	76±41*	102±29*	-145±13		
C14 _{IRFR}	-74±0	165±0	67±0	85±0	-163±0		
C14 _{IROQ}	-65±1	177±1	51±1	82±0	-149±0		

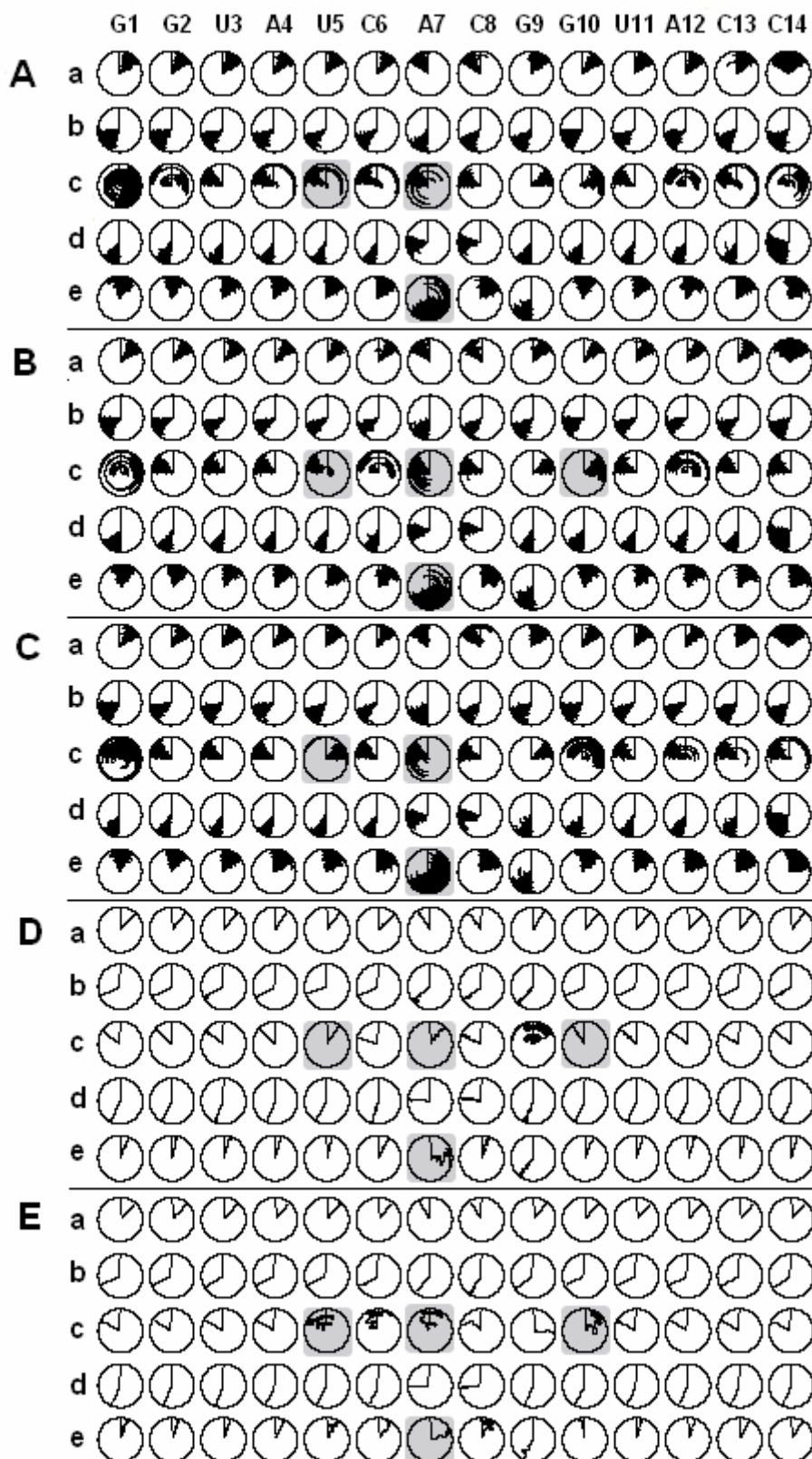


Figure 3.12 Time course of the torsional angles a, b, c, d, and e for the 14-mer uCACGg hairpin over 60 ns. The trajectories are displayed in a clockwise direction between 0 and 360° using (A) AMBER94, (B) AMBER98, (C) AMBER99, and from PDB: (D) 1RFR, (E) 1ROQ.

Table 3.10 Certain torsional angles for the CACG residues from simulation and experiment: average plus standard deviation. The deviations from experiment values are shown in the bold text. A value with a star means the torsional has more than one values.

	a	b	c	d	e		a	b	c	d	e
G1 _{AMBER94}	41±8	-119±13	123±274*	-157±8	9±12	C8 _{AMBER94}	-38±7	-136±8	-65±9	-89±7	33±10
G1 _{AMBER98}	43±6	-110±9	-65±151*	-154±8	6±11	C8 _{AMBER98}	-40±7	-133±7	-67±8	-86±6	33±11
G1 _{AMBER99}	43±6	-112±9	-10±118*	-155±8	6±11	C8 _{AMBER99}	-37±8	-135±8	-66±9	-89±8	36±11
G1 _{IRFR}	42±0	-113±0	-53±1	-153±0	22±1	C8 _{IRFR}	-40±2	-126±3	-65±4	-80±3	18±3
G1 _{IROQ}	43±0	-114±0	-62±1	-156±0	18±2	C8 _{IROQ}	-32±0	-144±3	-54±12	-92±2	19±13*
G2 _{AMBER94}	41±7	-118±10	-16±59*	-157±7	10±10	G9 _{AMBER94}	30±10	-142±10	66±8	-161±6	-154±10
G2 _{AMBER98}	43±6	-112±6	-61±8	-155±6	14±9	G9 _{AMBER98}	36±9	-134±10	62±24*	-162±7	-152±12
G2 _{AMBER99}	42±6	-115±7	-59±8	-156±6	13±9	G9 _{AMBER99}	33±10	-134±12	63±14	-158±10	-147±16
G2 _{IRFR}	39±0	-116±0	-48±1	-152±0	13±1	G9 _{IRFR}	31±0	-137±3	13±57*	-158±3	-140±2
G2 _{IROQ}	44±0	-115±0	-61±1	-156±0	19±1	G9 _{IROQ}	39±0	-125±0	94±4	-157±0	-149±6
U3 _{AMBER94}	41±7	-123±9	-59±9	-159±6	23±9	G10 _{AMBER94}	41±6	-119±11	52±35*	-156±7	7±10
U3 _{AMBER98}	44±6	-117±6	-61±8	-158±6	22±9	G10 _{AMBER98}	42±6	-113±9	71±13	-152±8	10±11
U3 _{AMBER99}	41±7	-121±8	-61±8	-159±6	24±10	G10 _{AMBER99}	40±7	-118±10	29±53*	-155±7	17±13
U3 _{IRFR}	38±0	-116±1	-53±1	-155±1	17±1	G10 _{IRFR}	40±0	-114±1	-34±1	-152±0	22±1
U3 _{IROQ}	41±0	-122±1	-59±0	-157±0	19±1	G10 _{IROQ}	40±0	-115±0	50±22*	-152±0	-10±4
A4 _{AMBER94}	38±8	-123±9	-49±34*	-158±6	22±11	U11 _{AMBER94}	39±7	-123±8	-58±8	-158±6	25±9
A4 _{AMBER98}	40±7	-118±7	-61±8	-156±6	23±10	U11 _{AMBER98}	40±8	-120±7	-61±8	-157±6	27±10
A4 _{AMBER99}	41±7	-118±7	-62±16	-152±6	27±12	U11 _{AMBER99}	38±7	-123±7	-58±8	-157±6	31±10
A4 _{IRFR}	33±0	-117±0	-46±0	-157±0	15±0	U11 _{IRFR}	39±0	-115±0	-47±2	-155±0	17±1
A4 _{IROQ}	41±0	-116±1	-60±0	-154±1	24±1	U11 _{IROQ}	42±0	-115±0	-62±2	-154±0	14±1
U5 _{AMBER94}	38±7	-126±8	-28±55*	-159±6	31±10	A12 _{AMBER94}	39±7	-119±9	-30±52*	-156±7	22±13
U5 _{AMBER98}	39±7	-122±8	-26±59*	-157±6	31±11	A12 _{AMBER98}	42±6	-113±7	-13±62*	-154±7	20±13
U5 _{AMBER99}	37±7	-133±10	61±25*	-161±6	27±13	A12 _{AMBER99}	39±7	-117±7	-48±33*	-154±6	35±13
U5 _{IRFR}	38±1	-108±0	34±0	-149±1	15±0	A12 _{IRFR}	46±0	-110±0	-55±0	-151±1	18±1
U5 _{IROQ}	40±0	-115±0	-43±37*	-152±0	27±8	A12 _{IROQ}	42±0	-116±0	-60±0	-155±0	19±1
C6 _{AMBER94}	40±6	-124±7	-39±42*	-158±6	34±10	C13 _{AMBER94}	37±8	-126±10	-24±57*	-157±6	23±13
C6 _{AMBER98}	40±6	-121±7	-18±60*	-156±6	34±12	C13 _{AMBER98}	42±7	-119±7	-61±8	-156±6	25±11
C6 _{AMBER99}	36±6	-124±6	-58±8	-155±6	39±12	C13 _{AMBER99}	38±8	-123±8	-53±24*	-156±6	32±12
C6 _{IRFR}	46±1	-114±0	-72±0	-162±1	26±0	C13 _{IRFR}	40±0	-113±0	-62±0	-154±0	14±0
C6 _{IROQ}	39±0	-120±1	-29±31*	-155±0	35±5	C13 _{IROQ}	41±0	-119±0	-61±0	-156±0	27±0
A7 _{AMBER94}	-36±8	-155±8	-62±20	-99±8	151±37*	C14 _{AMBER94}	-14±35	-139±12	2±63*	-114±29*	36±14
A7 _{AMBER98}	-38±7	-150±10	-66±14	-96±8	150±38*	C14 _{AMBER98}	24±31	-127±11	-62±9	-143±25*	35±14
A7 _{AMBER99}	-35±8	-151±10	-59±21	-98±9	146±41*	C14 _{AMBER99}	10±34	-132±12	-44±42*	-134±28*	38±14
A7 _{IRFR}	-32±1	-130±3	39±4	-83±1	90±20*	C14 _{IRFR}	38±0	-120±1	-51±0	-153±1	15±1
A7 _{IROQ}	-30±0	-139±1	-18±30	-93±0	72±4	C14 _{IROQ}	40±0	-124±0	-65±1	-157±0	30±0

Table 3.11 Radius of gyration of uCACGg hairpin

	Total (Å)	Stem	Loop
AMBER94	10.2±0.20	5.8±0.08	8.8±0.16
AMBER98	10.1±0.19	5.8±0.09	8.7±0.15
AMBER99	10.1±0.18	5.7±0.10	8.7±0.13

Table 3.12 Total, hydrophobic, and hydrophilic accessible surfaces of uCACGg hairpin

	Total (nm ²)	Hydrophobic (nm ²)	Hydrophilic (nm ²)	Ratio (hydrophobic/hydrophilic)
AMBER94	27.88±0.43	7.14	20.74	0.34
AMBER98	27.70±0.41	7.00	20.70	0.34
AMBER99	27.71±0.43	7.08	20.63	0.34

The calculations of R_g using the three AMBER force fields are summarized in Table 3.11. The overall NMR structure shows a R_g of 9.8 Å. In the simulation, the average R_g was 10.1 Å, representing a 3 % increment due to expansion in the loop region. These results can be due the flexibility of the looped out base A7.

The total solvent accessible surfaces (together with the hydrophobic and hydrophilic part) are calculated and reported in Table 3.12. The calculated values are used to estimate hydrophobic/hydrophilic ratio (Table 3.12). All three simulated hairpins have the same solvent accessible surface that fluctuates around the value of 27.8 nm². The ratio between hydrophilic and hydrophobic surface is constant during the simulation time for all the three calculations. The percentage of solvent accessibility per nucleotide in all three simulations is similar as shown in Figure 3.14. That reveals that only few nucleotides are solvent-exposed by less than 9 %. These nucleotides are the two end-nucleotides (G1 and C14), the looped out nucleotide (A7) and the loop nucleotide G9. All other nucleotides show a similar amount of solvent-exposed surface except C6 that has the lowest value.

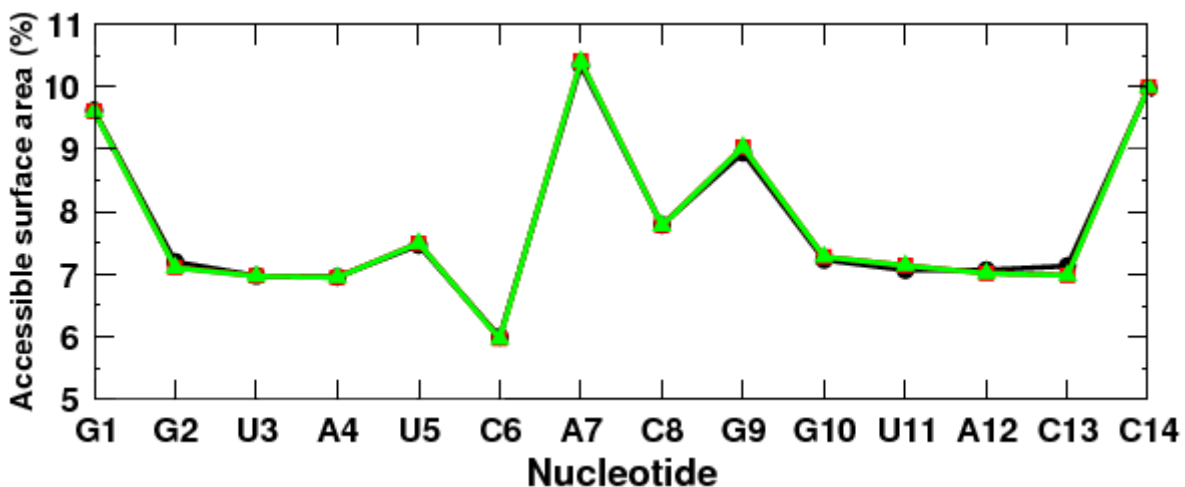


Figure 3.13 Percentage of the solvent accessible surface per nucleotide for uCACGg hairpin using AMBER94 (black), AMBER98 (red), and AMBER99 (green). Values represent average values for the 60 ns MD simulations and the nucleotides are labeled according to the Figure 3.5

Moreover, the solvent accessible surface per atom of residues U5 to G10 is calculated for the starting structure and for the average values over 60 ns simulation time. Figure 3.14B shows that the difference between the force fields is located around the U5. It is to note that also the some dihedral angles (β and γ) of U5 show different values especially in the AMBER99 MD simulation with respect to the reference structure 1RFR, whereas in the structures 1ROQ they show a quite large standard deviation. Interestingly, the atoms that involved in the hydrogen bonding in the NMR structure and in the simulations did not show significant change in the surface except for atom O₆ of G9 from 0.07 nm² (starting structure) to 0.17 nm². This finding suggests that the ribose 2'-hydroxyl group of C6 is still buried in the loop even if not involved in hydrogen bond.

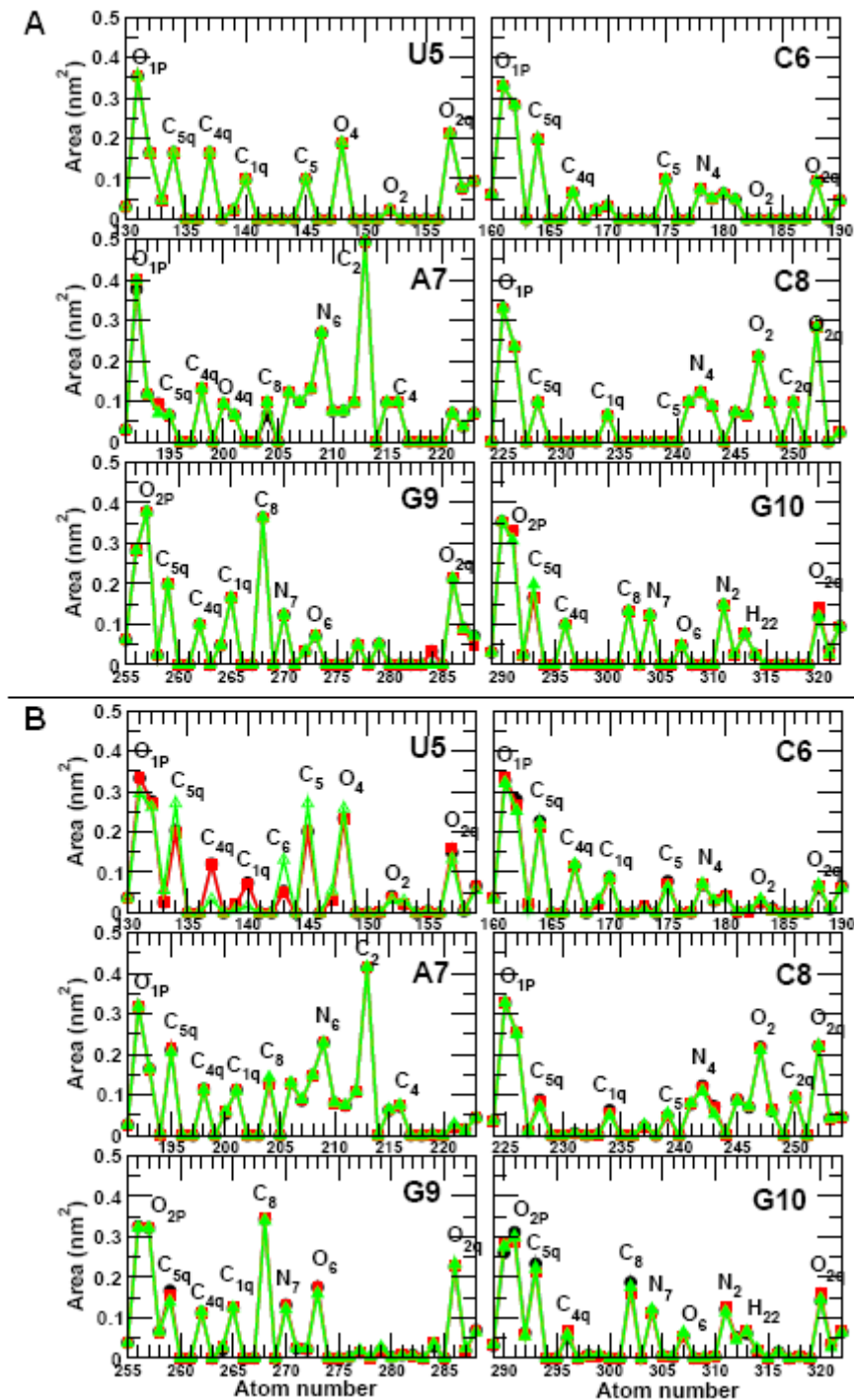


Figure 3.14 Calculation of the solvent accessible surface per atom for uCACGg hairpin using AMBER94 (black), AMBER98 (red), AMBER99 (green). Values represent (A) initial value at 0 ns, and (B) average values for the 60 ns MD simulations. The nucleotides are labeled according to uCACGg hairpin numbering scheme (see Figure 3.1). Only values for the residues U5 to G10 are reported. Peak are labeled with atom names.

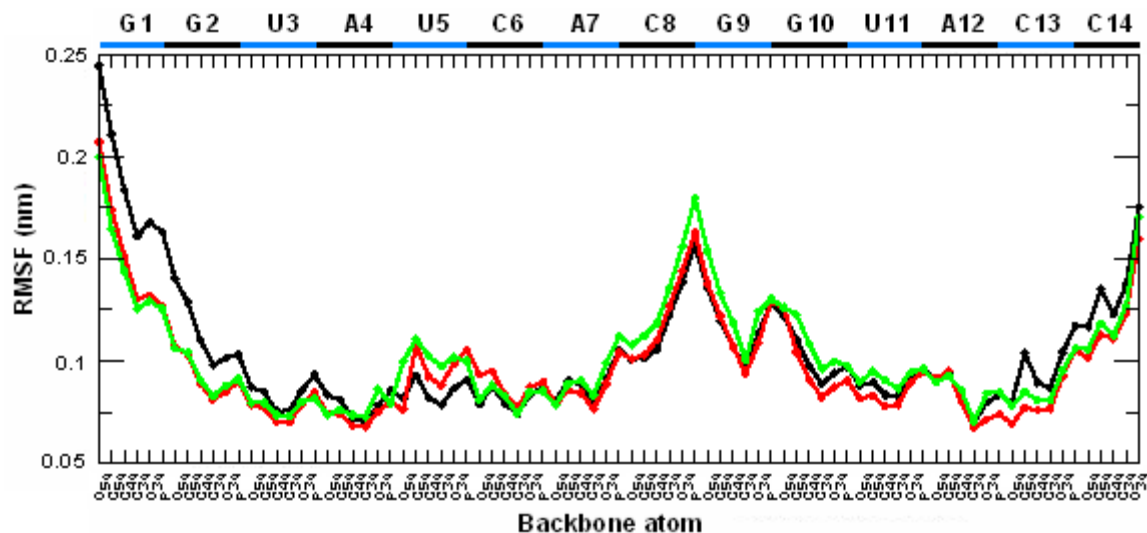


Figure 3.15 RMSFs of all backbone atoms relative to their average positions after fitting on all atoms. The nucleotide concerning to the atoms are presented. The black, red and green lines are for the simulation using AMBER94, AMBER98 and AMBER99 respectively.

Features that assess how good the starting structure is maintained are not enough to tell how good a model is. It is also very important to check if low RMSD values are not the result of damping of important intramolecular motions or freezing out of molecular degrees of freedom. Therefore RMSF of backbone atoms about the average structure were calculated and are shown in Figure 3.15.

As shown in Figure 3.15, the MD simulations using three different AMBER force fields have the same general trends. The backbone atoms between residues C8 and G9 show unexpected highest fluctuation although the dihedral angles around these nucleotides are relative stable over the simulations. A high fluctuation in the backbone atoms is also observed around U5–C6 and G9–G10 in agreement with the dihedral angles transition discussed before. It is also interesting to note, that the backbone atoms of the residue A7 shows low fluctuation even if this residue is looped out in the solvent. The fluctuations of backbone atoms around C8 and G9 are depicted in Figure 3.16.



Figure 3.16 Superposition of backbone atoms of the ten structures uCACGg hairpin using AMBER98 at 50 to 50.1 ns time of simulation. The nucleotides G1, C8, and G9 are visualized

3.2.3 Conclusion

The main outcome of the simulations is generally valid for the MD simulations of uCACGg hairpin using the three AMBER force fields. The calculation of NOE distances, RMSD, dihedral angles, radius gyration, solvent accessible surface do not show any significant drift when different versions of AMBER force field are used. The simulated structure differs from NMR structure in the hydrogen bond network in the loop region. Two of six hydrogen bonds observed in experimental structure were not presented in the simulations. However, the NOE distances around those loop atoms have no significant violations.

The flexibility of the hairpin evaluated by calculating the dihedral angles and atomic fluctuation shows no significant differences between the three versions of AMBER force fields. It is to note that high atomic fluctuations are always observed at stem/loop junction (U5–C6, and G9–G10) and the loop residues (C8 and G9).

In conclusion, all the versions of the three AMBER force fields describe with the same accuracy the small RNA hairpin, uCACGg hairpin.

3.3 Effect of electrostatic treatments on uCACGg hairpin

In the Chapter 1, it is described that one of the most common structural motifs in RNA is the hairpin that consists of double stranded stem and a single stranded loop. Stable RNA tetraloop motifs, whose loop comprises of four nucleotides, are found frequently in biologically active RNAs, and as our research object, the uCACGg hairpin (Figure 3.5) is chosen. In order to investigate the effect of the electrostatic treatment in detail, additional MD simulations are carried out for the uCACGg hairpin using reaction field and PME method.

3.3.1 Computational details

The starting structure, the 14-mer uCACGg hairpin, for simulations is the upper part of the NMR structure of the 30-mer stemloop-D of coxsackieviral RNA of PDB 1RFR of the Brookhaven Protein Data Bank (Ohlenschläger, 2004). The model systems involved explicit all-atom and the interactions between atoms were described using the AMBER98 (Cheatham, 1999) force field. The GROMACS suite of programs (Berendsen, 1995b; Lindahl, 2001) was used to simulate the model systems for 60 ns. All the MD simulations were carried out in the NPT ensemble with periodic boundary conditions, a temperature of 300 K and at pressure 1 atm. The temperature and the pressure were kept constant by Berendsen's weak coupling scheme (Berendsen, 1984) with coupling parameter of 0.1 ps for temperature, 1.0 ps for pressure, and the isothermal compressibility of $4.5 \cdot 10^{-5} \text{ bar}^{-1}$. The hairpin was solvated with around 5000 water molecules used water model TIP3P (Jorgensen, 1983) in a rhombic dodecahedron box with minimum periodic distance of about 2.94 nm. The equations of motion were integrated using leapfrog scheme (Snyman, 1982) with a time step of 2 fs. Constraint were used for bond lengths using the LINCS algorithm for the hairpin (Hess, 1997), and SETTLE for the water (Miyamoto, 1992).

A twin-range cutoff is used for van der Waals interactions. The interactions within the short cutoff of 1.0 nm are evaluated at every time step, while the interactions between atoms separated by a distance ranging between the short and the long cutoff were evaluated every tenth step. The electrostatic interactions were evaluated by using two different methods: the reaction field (RF) and particle mesh Ewald (PME) method.

The cutoff using the reaction field correction was 1.4 nm. The settings for the PME method were as follow: a real space cutoff of 1.0 nm, a grid spacing of 0.12 nm. The 13 sodium counterions were placed by replacing randomly water molecules to obtain a neutralized system. In the other simulations, the concentration of sodium ions was modified by 32 and 77 additional Na^+/Cl^- ion pairs, and in another 7 $\text{Mg}^{2+}/1 \text{Cl}^-$ were used. The details of the simulations are in Table 3.13

Table 3.13 Details of ion concentrations in MD simulations of uCACGg hairpin during 60 ns.

	[Na ⁺] (M)		[Mg ²⁺] (M)	
RF	0.144	0.50	1.03	0.077
PME	0.144	0.50	1.03	0.077

3.3.2 Reaction field versus PME

Some structural parameters of the uCACGg hairpin are calculated for 60 ns MD simulations with two different long-range electrostatic treatments. These results are summarized in Table 3.14. Stable conformations of the hairpin are observed mainly when PME method is used for long-range electrostatic. Only one of four simulations performed with reaction field show the structural stability. That is the simulation with Na^+ as counterions at the concentration of 0.144 M. It has been commonly accepted, that the role of the metal ions is to neutralize the polyanionic nucleic acids (Korolev, 2003; Martínez, 2001; Feig, 1999), and the increase of ion concentration (Tan, 2006) should not destabilize the uCACGg hairpin. This phenomenon cannot be accommodated by simulations using reaction field. In contrast, the MD simulations using PME method exhibit reasonable structural stability at different ion concentrations and with multivalent counterion such as Mg^{2+} .

Table 3.14 Structural features for uCACGg hairpin of MD simulations using AMBER98 force field with different ion concentration of Na^+ and one concentration of Mg^{2+}

Structural features	Reaction field	PME
NOE violations (%)		
0.14 M Na^+	5.9	6.5
0.50 M Na^+	11.9	6.9
1.03 M Na^+	11.9	6.7
0.08 M Mg^{2+}	10.6	5.9
RMSD (nm)		
0.14 M Na^+	0.22±0.03	0.20±0.03
0.50 M Na^+	0.88±0.27	0.21±0.03
1.03 M Na^+	0.52±0.06	0.21±0.03
0.08 M Mg^{2+}	0.88±0.10	0.22±0.04
Fraction of hydrogen bonds P_H		
0.14 M Na^+	0.993	0.994
0.50 M Na^+	0.174	0.996
1.03 M Na^+	0.035	0.996
0.08 M Mg^{2+}	0.294	0.994
Fraction of base stacking P_s		
0.14 M Na^+	0.999	0.998
0.50 M Na^+	0.418	0.999
1.03 M Na^+	0.368	1.0
0.08 M Mg^{2+}	0.516	1.0

It is interesting to analyze the distribution of hydrogen bonds in the stem (shown in Figure 3.17). The stabilization of the hairpin by Na^+ ions is shown using both methods (picture top left) at the 0.14 M. As the Na^+ concentration increases from the simulations using reaction field, the distributions of hydrogen bonds significantly decrease mainly for the closing base-pair. Using other counterions such as Mg^{+2} no stabilization of the structure of the hairpin is observed, but only one stable hydrogen bond.

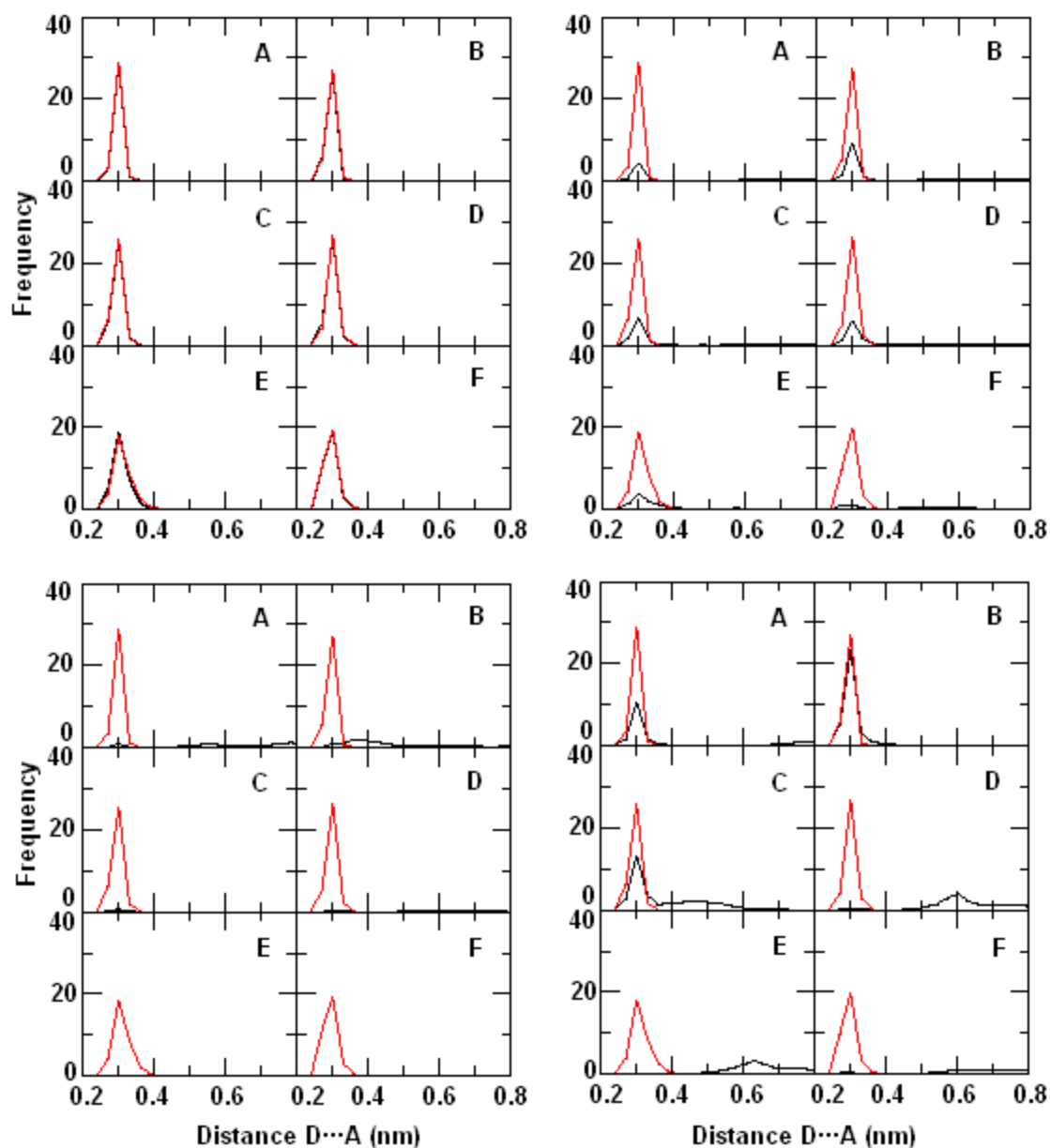


Figure 3.17 Distributions of hydrogen bonds are represented by donor-acceptor ($D \cdots A$) distance of A: $N_1(G1) \cdots N_3(C14)$, B: $N_1(G2) \cdots N_3(C13)$, C: $N_3(U3) \cdots N_1(A12)$, D: $N_3(U11) \cdots N_1(A4)$, E: $N_1(G10) \cdots O_2(U5)$, F: $N_3(U5) \cdots O_6(G10)$ for 0.14 M Na^+ (left top), 0.50 M Na^+ (right top), 0.10 M Na^+ (left bottom), and 0.07 M Mg^{2+} (right bottom). The black lines show the results of the 60 ns MD simulations using reaction field, and the red line using PME method.

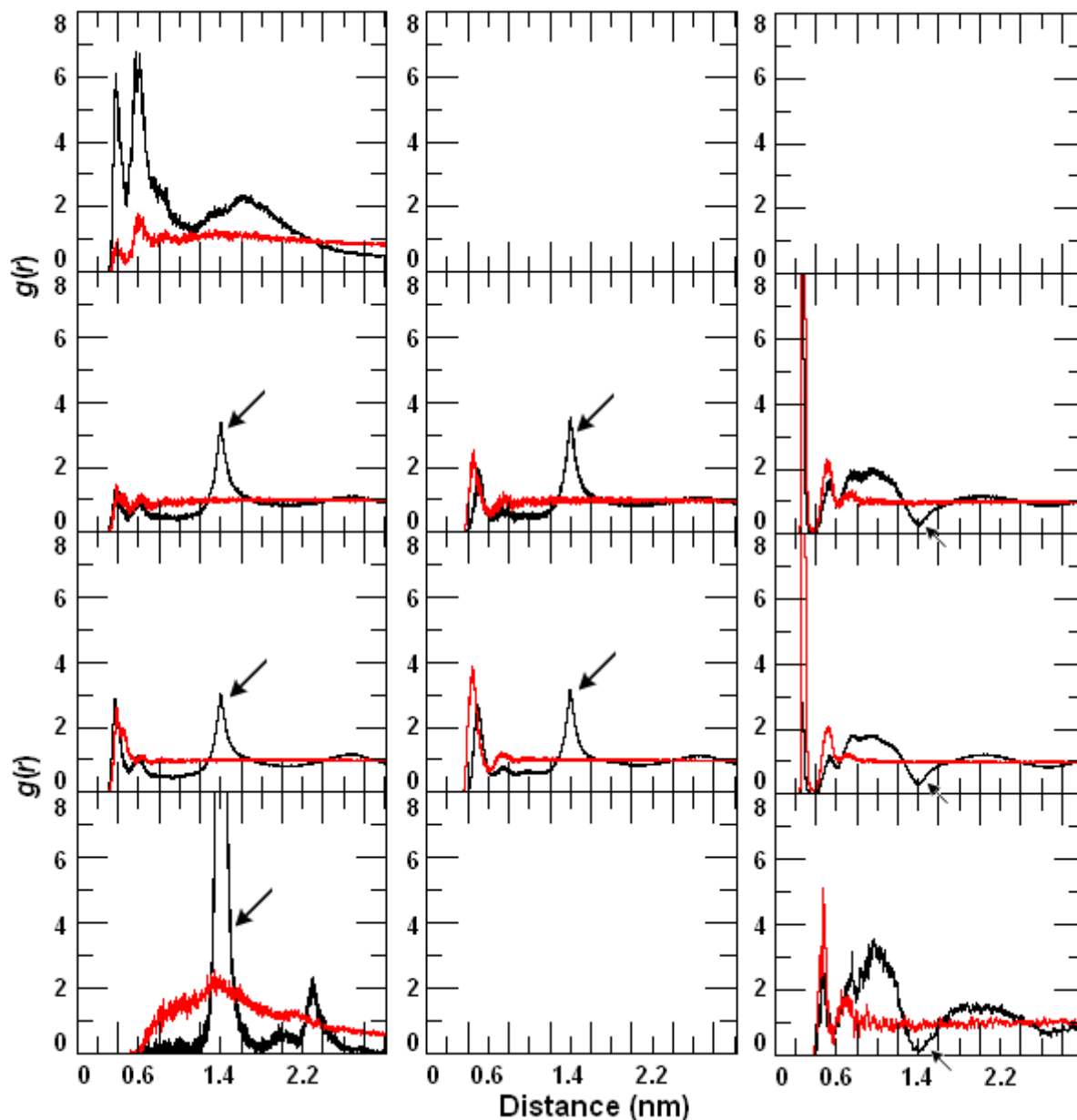


Figure 3.18 Radial distribution function (RDF) for: cation-cation (left), anion-anion (middle), and cation-anion (right); from top to bottom: 0.14 M Na^+ , 0.50 M Na^+ , 1.03 M Na^+ , 0.08 M Mg^{2+} . The $g(r)$ results using reaction field are shown by black curves and using PME red curves. Artifacts at the cutoff radius are depicted by black arrows.

Furthermore, the ion-ion interactions are investigated by calculating the RDF, as shown in Figure 3.18. The previous MD studies on 1 M NaCl water solution showed that $g(r)_{\text{Na-Na}}$ and $g(r)_{\text{Cl-Cl}}$ from MD simulation using cutoff method had a peak around the cutoff radius (Tironi, 1995). The ions tend to accumulate at the cutoff distance to minimize unfavorable interactions. Similar artifact is observed here in $g(r)_{\text{X-X}}$, $g(r)_{\text{Cl-Cl}}$, $g(r)_{\text{X-Cl}}$ ($\text{X} = \text{Na}^+$ or Mg^{2+}). That might indicate that the screening effect of water

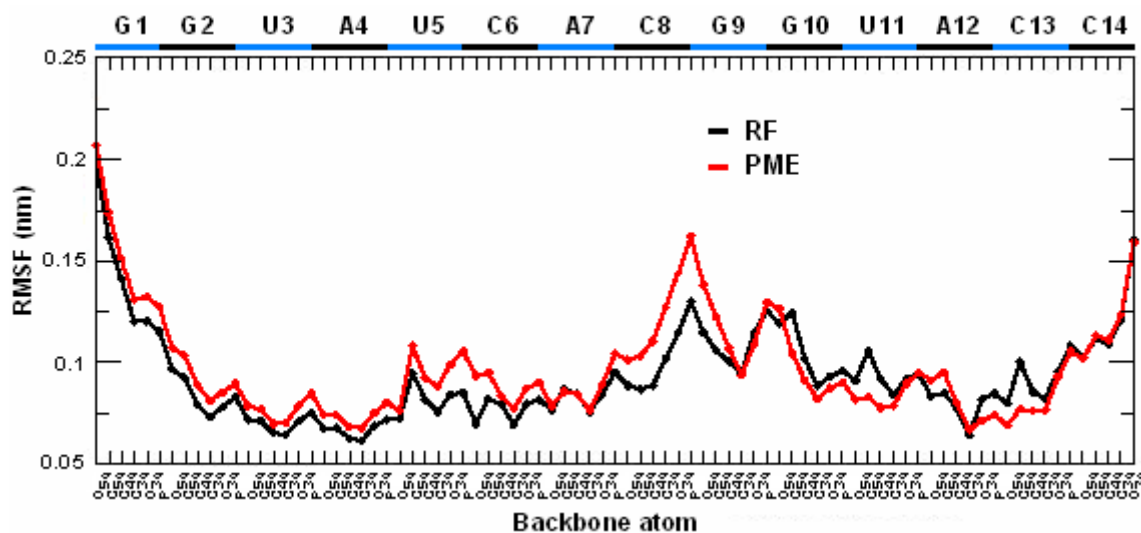


Figure 3.19 RMSFs of all backbone atoms relative to their average positions after fitting on all atoms for uCACGg hairpin of 0.14 M concentration Na^+ . The nucleotide concerning to the atoms are presented. The black and red lines are for the simulation using RF and PME respectively.

molecules also do not improve when the reaction field correction are used. The same reason can be argued for the $g(r)_{\text{Na-Cl}}$ in Figure 3.18 which display a “minimum” at the cutoff radius. On the contrary, due to hydration shell of Mg^{2+} , the calculation of $g(r)_{\text{Mg-Mg}}$ from MD simulations using reaction field suggests an underestimation of screening effect of water, so that the ions accumulate at the cutoff distance. A hydration of Mg^{2+} , whose residence time > 60 ns according to our simulation and 2000 ns to experiment (York, 1992), caused a strong screening effect.

Additionally, the flexibility of the uCACGg hairpin using the two methods is reported in Figure 3.19. The RMS fluctuations of the simulations are calculated using RF and PME methods at 0.14 M of Na^+ concentration. Figure 3.19 provides information concerning atomic fluctuation relative to their average positions after fitting on all atoms for uCACGg hairpin. The both simulations yield the similar average RMS fluctuation, around 0.093 and 0.097 nm for RF and PME method respectively.

3.3.3 Comparison of different ion concentrations

In previous section, the two methods for electrostatic interaction are compared. In this section, the interaction of certain electronegative atoms of the hairpin with cation is studied using the PME method. Occupancies of ions around electronegative atoms are

determined by integration the first maximum of the corresponding RDF. A special treatment was applied to calculate the occupancy of Mg^{2+} due to involvement of $\text{O}_{2\text{P}}$ (C14) in hydration shell of Mg^{2+} . The summary of results is listed in Table 3.15, Figure 3.20, and Figure 3.21 that displays SDF of cations around the uCACGg hairpin.

In the case of Na^+ ions, the cations mainly interact with the $\text{O}_4(\text{U5})$, $\text{O}_4(\text{U11})$, and $\text{O}_2(\text{C8})$. The two first electronegative atoms are located in the major groove of the hairpin. Especially for the first one, $\text{O}_4(\text{U5})$, is not involved in hydrogen bond with its wooble base-pair G10 (shown in Figure 3.2) and therefore can effectively with Na^+ ions interact. It is particular interesting because the experimental information shows that this closing base-pair seems to be considerable in the stabilization of the hairpin. On the other hand, the $\text{O}_4(\text{U11})$ is an acceptor in a hydrogen bonding with $\text{N}_2(\text{A4})$, however, the result strongly suggest its interaction with Na^+ ion. The not involved in hydrogen bonding last atom, $\text{O}_2(\text{C8})$ is placed in the loop region at the major groove, display the high occupancy of Na^+ ions. In addition, the result clearly demonstrates that Na^+ ions reside in relative long time (maximum residence time) although their occupancies are not particular high. Moreover, the occupancies of Na^+ ions around phosphate group appear to be influenced by the ion concentration.

Figure 3.20 and 3.21 exhibit the clear difference between Na^+ and Mg^{2+} interaction with the uCACGg hairpin. Mg^{2+} ions prefer direct binding to phosphate group $\text{O}_{2\text{P}}$ (C14) and have similar interaction with the major groove, although not with the same atom as Na^+ ions and even not with minor groove. This different behavior of Mg^{2+} ions is expected due to their relative more stable hydration shell, so that the cations need more space to interact with the hairpin. However, the both cation show relative higher residence time around the phosphate group $\text{O}_{2\text{P}}$. Interestingly, the affinity of the phosphate groups at the both sides appears different in the interaction with Mg^{2+} ions.

Table 3.15 Interaction of ions with electronegative atoms

Electronegative atoms	Occupancy (%)				Residence Time (ps)			
	Na ⁺			Mg ²⁺	Na ⁺			Mg ²⁺
	0.14 M	0.50 M	1.03 M	0.8 M	0.14 M	0.50 M	1.03 M	0.8 M
O ₆ /G1	1.2	2.6	3.3	31.8	12±5	18±16	23±20	25±23
N ₇ /G1	2.2	5.3	7.4	35.6	18±13	20±19	24±26	30±28
O _{2p} /G2	2.7	3.9	6.8	17.4	66±57	39±45	70±81	49±68
O _{1p} /G2	1.9	3.6	7.1	17.9	40±31	59±65	67±58	116±116
O ₆ /G2	4.1	5.0	5.6	50.9	26±27	27±26	38±34	107±172
N ₇ /G2	3.9	5.5	5.2	51.9	22±19	27±33	22±25	72±80
O _{2p} /U3	3.2	4.9	6.6	34.1	48±55	58±60	55±73	120±205
O _{1p} /U3	2.0	4.2	8.7	28.3	53±40	47±48	61±62	77±89
O ₄ /U3	11.9	8.6	7.7	65.9	45±47	33±32	43±38	86±120
O _{2p} /A4	5.2	7.9	9.8	29.7	77±79	64±88	63±82	112±195
O _{1p} /A4	2.3	3.5	9.5	30.6	41±44	50±49	62±65	96±117
O _{2q} /A4	1.8	1.7	3.2	6.5	18±11	21±18	25±20	47±83
N ₇ /A4	5.1	3.7	8.0	4.5	127±185	79±98	119±97	62±72
O _{2p} /U5	2.9	2.3	3.0	25.5	51±62	29±31	50±60	108±164
O _{1p} /U5	2.2	5.8	9.9	25.2	35±36	53±61	85±107	83±118
O ₄ /U5	20.7	27.9	26.6	29.6	66±80	77±93	69±76	36±38
O _{2p} /C6	1.7	2.9	2.5	17.6	67±71	53±54	42±33	97±145
O _{1p} /C6	3.0	4.6	6.8	25.3	51±69	63±69	52±47	90±102
O _{2q} /C6	0.9	0.8	2.0	3.9	15±11	20±18	27±26	34±38
O _{2p} /A7	0.5	0.7	0.1	16.8	23±19	31±26	18±11	126±161
O _{1p} /A7	1.5	3.6	5.4	18.8	50±57	80±75	80±94	79±106
N ₁ /A7	1.2	1.5	2.9	3.2	33±32	19±17	25±20	12±5
N ₇ /A7	7.6	10.1	9.2	6.8	72±78	58±70	50±59	33±35
O _{2p} /C8	0.6	2.1	3.9	5.4	50±51	44±49	87±158	62±99
O _{1p} /C8	1.6	2.8	7.0	9.1	88±102	53±53	94±114	99±122
O _{2q} /C8	2.4	6.1	7.2	14.1	19±17	24±24	29±27	32±28
O ₂ /C8	7.3	16.8	18.4	16.0	45±46	56±62	55±51	40±53
N ₃ /C8	3.7	8.8	8.6	10.4	61±56	57±82	64±72	33±40
O _{2p} /G9	2.0	5.8	4.3	9.3	38±34	77±76	50±53	67±61
O _{1p} /G9	1.7	5.7	6.0	10.6	40±41	67±90	70±58	72±64
O _{2q} /G9	0.3	1.1	1.7	3.3	12±7	20±18	20±13	17±9
O ₆ /G9	6.7	11.5	9.4	4.8	79±96	80±108	61±69	64±89
N ₇ /G9	1.0	2.5	3.7	3.2	25±25	29±32	21±24	30±22
O _{2p} /G10	1.8	3.3	4.3	6.6	52±55	48±50	52±57	74±89
O _{1p} /G10	1.6	4.6	6.9	8.1	49±34	56±69	60±59	71±65
O _{2q} /G10	0.2	0.8	1.1	4.3	18±11	18±18	25±21	30±25
O ₆ /G10	3.3	1.7	1.3	23.7	16±14	16±13	15±9	60±68
N ₇ /G10	10.4	6.3	6.4	27.2	55±82	36±42	43±44	55±75
O _{2p} /U11	0.8	2.9	4.0	8.9	46±55	48±61	46±49	63±83
O _{1p} /U11	1.6	3.4	8.6	19.8	56±56	34±37	71±85	113±113
O _{2q} /U11	1.1	0.7	1.3	3.2	18±20	15±9	24±14	38±40
O ₄ /U11	15.5	17.1	16.4	54.1	56±69	57±69	71±69	70±93
O _{2p} /A12	1.9	5.1	5.1	8.0	47±55	56±63	50±50	99±134
O _{1p} /A12	3.0	5.5	9.9	9.0	65±65	58±59	80±80	80±106
O _{2p} /C13	0.8	2.3	3.1	1.6	29±40	35±44	52±51	79±89
O _{1p} /C13	2.1	5.9	7.5	1.6	63±72	50±66	66±77	87±128
O _{2q} /C13	0.6	1.4	1.9	1.3	20±10	18±17	21±16	47±52
O _{2p} /C14	0.9	1.2	2.0	0.0	40±44	37±39	35±54	0±0
O _{1p} /C14	2.5	3.6	4.8	0.6	90±76	55±60	69±86	66±77

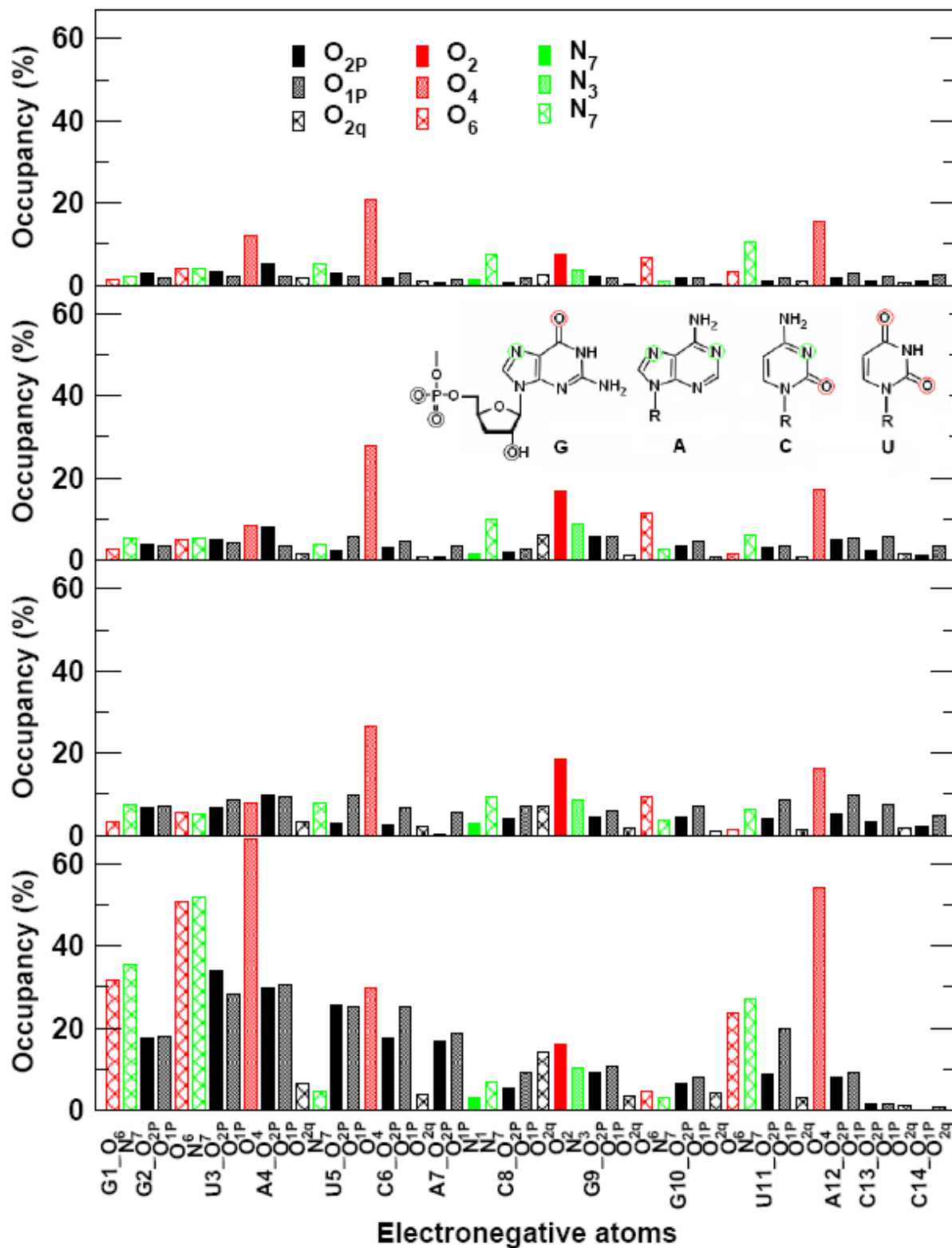


Figure 3.20 Averaged occupancy of 0.14 M Na⁺, 0.50 M Na⁺, 0.50 M Na⁺, and 0.08 M Mg²⁺ (from top to bottom) around electronegative atoms of uCACGg hairpin. The symbols for the electronegative atoms are represented (top). To highlight the certain electronegative atoms, these atoms are showed inside circles in the second box from top.

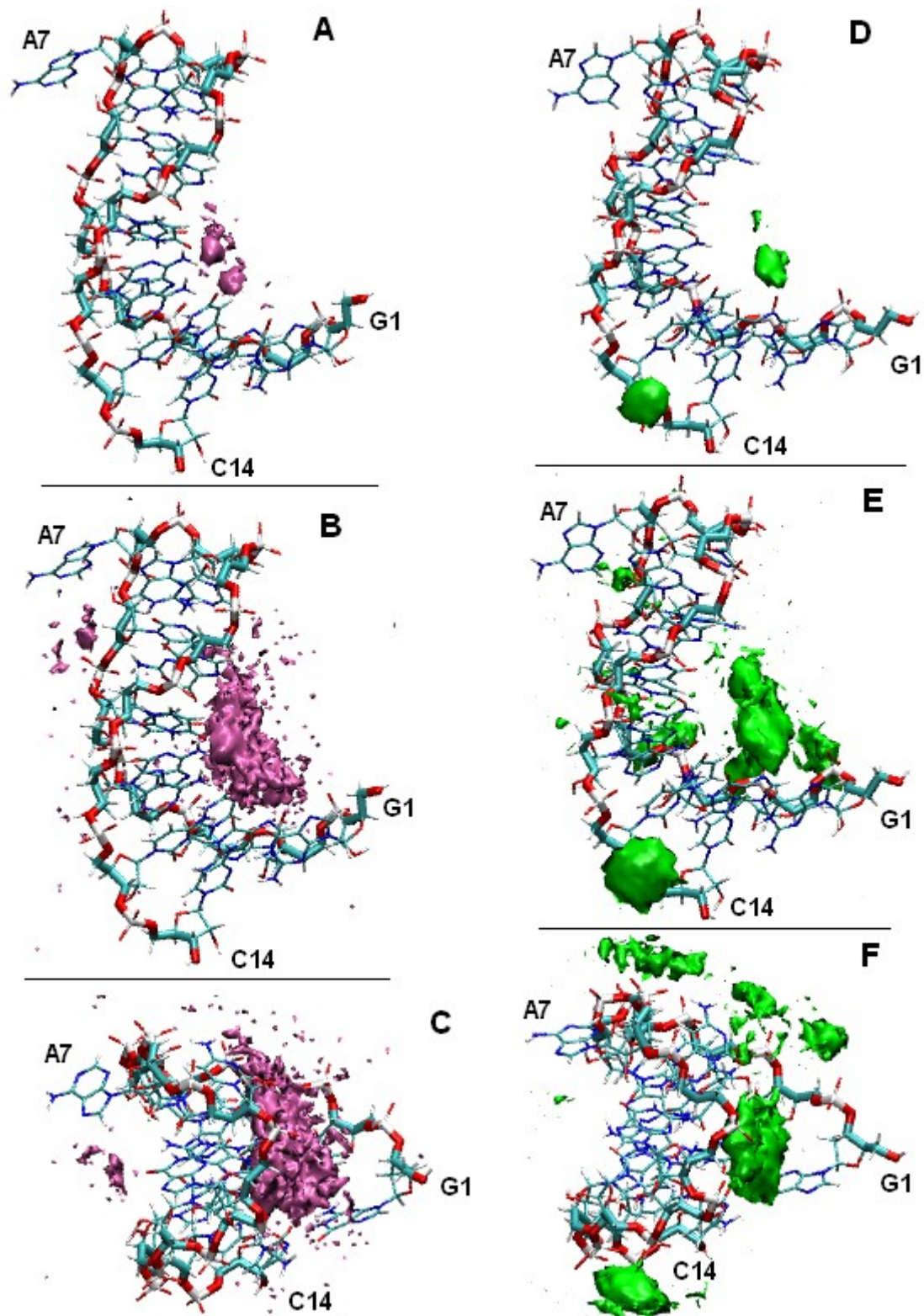


Figure 3.22 RMSF of bFigure 3Figure 3.21 SDFs of Na^+ (violet) and Mg^{2+} (green) around averaged structure of uCACGg hairpin. To highlight the preferential presence of the ions, the backbone atoms have been displayed thicker than the rest atoms of the hairpin, and certain residues are signed. For the best display, the SDFs of ions are displayed in top view (E and F) and side view (A, B, D, and E). The isosurfaces correspond to $g(r, \Omega) = 90$ (A), 32 (B, and C), 350 (D), and 80 (E, and F)

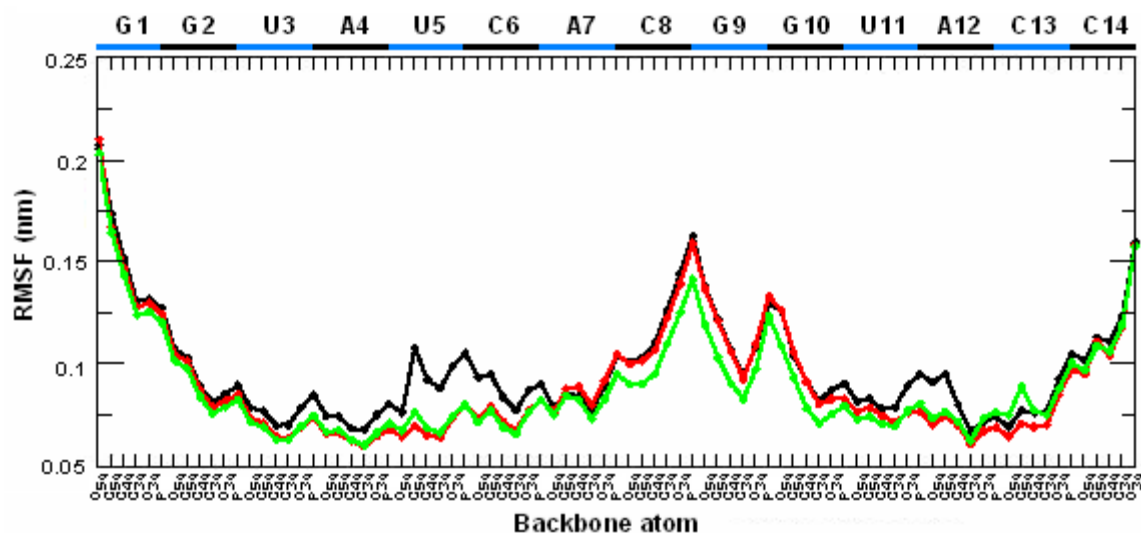


Figure 3.22 RMSF of backbone atoms relative to their average positions after fitting to all atoms for uCACGg hairpin at different concentration of Na⁺. The black, red and green lines show Na⁺ concentration of 0.14 M, 0.50 M, and 1.04 M respectively.

Finally, the increase of Na⁺ ion concentrations influences mainly the maximum residence time and the occupancies of phosphate group. Moreover, the effect of ion concentration on atomic fluctuation is analyzed by calculation the RMS fluctuation of backbone atoms. Although the concentration effect is not really obvious the concentration effect, the average fluctuation decrease 7 % and 10 % with the increase of the Na⁺ ion concentration.

3.3.4 Conclusion

The MD results using the AMBER94, AMBER98, and AMBER99 force fields show no significant difference in the case of structural behavior and atomic fluctuations. The good agreement with the NOE distances and the stability of the hairpin configurations during 60 ns simulations indicates the ability of the model to describe structural properties of RNA hairpins.

The study on the effect of electrostatic treatment on the uCACGg hairpin shows that the PME method is more suitable to treat electrostatic long-range interactions than the reaction field approach in the case of RNA hairpins in water solution. Although at the low concentration of Na^+ (0.14 M) no big effects on the structural features of the hairpin are observed when RF or PME are used, the hairpin loses most of its native contacts using RF, when ion concentration is increased or divalent counterions (Mg^{2+}) are considered. That indicates that effect of cations on the hairpin's stability cannot mimic in simulations using RF. Moreover, in these simulations the ion-ion RDFs clearly show artifacts at the cutoff distance.

The average hairpin's structures are not influenced by increasing of the ion concentration and by changing in the ion types. The main effects of Na^+ concentrations are ¹⁾ to increase the maximum residence time of Na^+ ions around RNA, and ²⁾ to increase the presence of ions around phosphate groups. The ion locations around the hairpin are different if the ion is mono- or divalent. The major interaction sites of Na^+ are the oxygen atoms of bases ($\text{O}_4/\text{U5}$, $\text{O}_4/\text{U11}$) in major groove, and $\text{O}_2/\text{C8}$ in the loop, whereas the interaction sites of Mg^{2+} are the oxygen atoms of bases ($\text{O}_4/\text{U3}$, $\text{O}_4/\text{U11}$) in the major groove, and oxygen atoms of phosphate groups.

Chapter 4

Structure, Dynamics, and Thermostability of the RNA Hairpins uCACGg and cUUCGg

Coxsackieviruses are a genus of RNA viruses associated with several acute and chronic human diseases. These viruses carry the genetic information on a positive-sense single-strand RNA that can be immediately translated by the host cell. The essential step in the assembly of the viral replication process is the formation of a complex between the viral protease and 5'-non-translated region of the virus (Rieder, 2003; Andino, 1990). In this domain the stem-loop D constitutes the major recognition site for the viral protease 3CD^{pro}. In vitro study showed that stem-loop D alone is sufficient to bind the protein 3CD^{pro} (Zell, 2002), and it was suggested (Ihle, 2005; Ohlenschläger, 2004) that the RNA-protein recognition process might be based on structural properties instead of on the specific sequence.

In the last years, nuclear magnetic resonance (NMR) techniques have been used to investigate the structure of the stem-loop D of Coxsackievirus B3 (Ohlenschläger, 2004; Du, 2003). The NMR structure of its apical tetraloop uCACGg is strikingly similar in overall geometry and hydrogen bonding to the canonical cUUCGg tetraloop (Allain, 1995; Ennifar, 2000), despite of the different sequence and closing base-pair of the two loops. In particular, the CACG loop was found to exhibit the interactions, which were thought to explain the unusual stability of the UNCG, loop family (Woese, 1990; Williams, 2000) (where N is any nucleotide). This leads to the suggestion (Proctor, 2002) to extend the UNCG tetraloop family to the motif YNMG (where Y is a pyrimidine and M is an adenine or a cytosine).

Despite of their considerable structural similarity, however, the uCACGg and cUUCGg tetraloops were found to differ in their functionality and thermostability. Mutation studies of the stem-loop D of Coxsackievirus B3 showed that the replacement of uCACGg by cUUCGg did not lead to functional binding (Du, 2003). Furthermore, the melting temperature of the uCACGg loop is about 20 K lower the one of the cUUCGg loop. The melting temperature of a 12-nucleotide cUUCGg hairpin was reported as 346 K (Proctor, 2002). A 38-mer uCACGg loop of Coxsackievirus was observed to melt at about 338 K (Du, 2003), while the broadening of the NMR imino proton resonance of 30-mer uCACGg loop was already recorded at 318 K (Ohlenschläger, 2004). While this difference in stability is in accordance with empirical rules using the base-pair sequence (Saenger, 1988), a microscopic picture of the unfolding of the RNA hairpins has not yet been established.

In this chapter, the classical replica-exchange molecular dynamics (REMD) simulations (Frenkel, 2002; Okabe, 2001; Sugita, 1999) are employed to study the effects of the loop sequence and the closing base-pair on the conformational distribution and on the thermostability of RNA hairpins. Representing the 14-mer hairpins gguauCACGguacc and ggcacUUCGgugcc (see Figure 4.1) by the AMBER all-atom force field (Cheatham, 1999, Cornell, 1995) in explicit water, 30 ns REMD simulations using 48 replicas with temperatures between 297 and 495 K were performed. Combining a microscopic description of RNA hairpins (Villa, 2006; Ihle, 2005; Koplín, 2005; Auffinger, 2003; Nina, 2002; Sorin, 2002; Sarzynska, 2000; Zacharia, 2000; Williams, 2000; Williams, 1999; Hermann, 1998; Miller, 1997; Zichi, 1995) with the enhanced

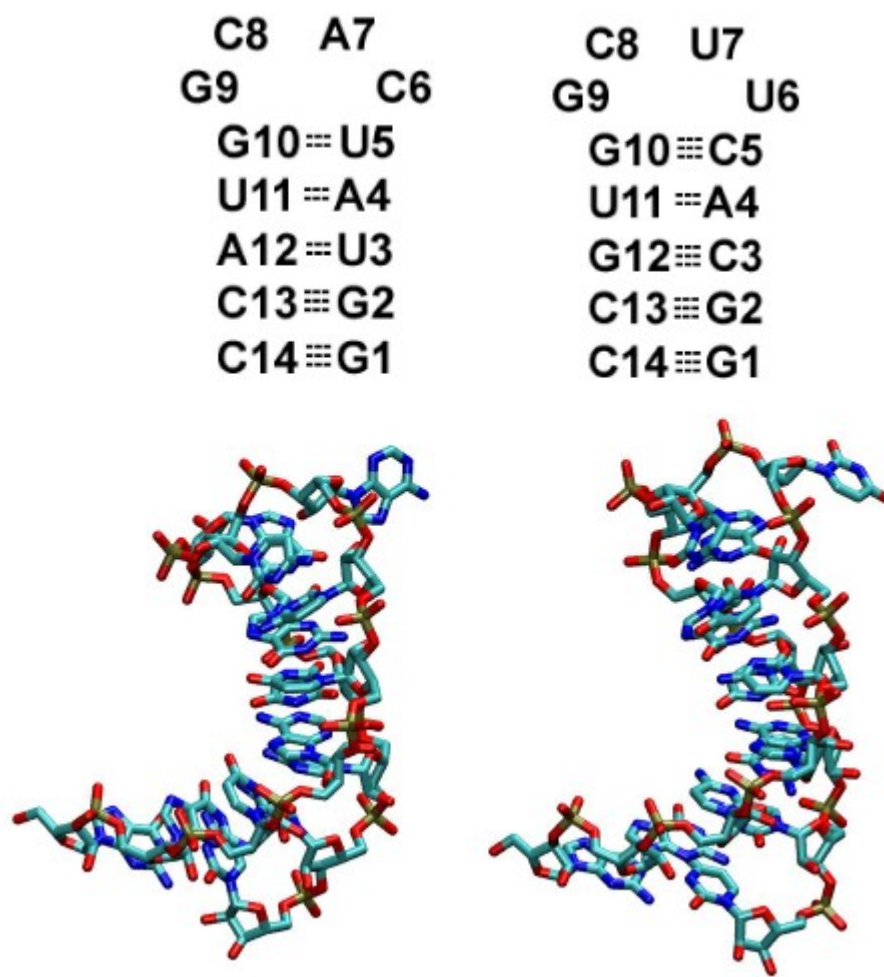


Figure 4.1 The 14-mer RNA hairpins uCACGg (left) and cUUCGg (right). Top: Secondary structures including stem base-pair hydrogen bonds and residue numbering. Bottom: Representative MD snapshots at 20 ns and 300 K. The pictures were performed using the graphical package VMD (Humphrey, 1996)

sampling qualities of REMD (Sanbonmatsu, 2006; Seibert, 2005; García, 2003; Rao, 2003; Rhee, 2003; Zhou, 2003; Sanbonmatsu, 2002; García, 2001; Zhou, 2001), the approach allows the conformational structure and dynamics to be studied as well as the folding and unfolding of RNA hairpins in great detail (Sorin, 2005; Sorin 2004; Sorin, 2003; Sorin, 2002).

The chapter is organized as follows. In Section 4.1, the Computational Details, which consists of the simulation condition, replica-exchange molecular dynamics, and trajectory analysis, are described. Afterwards, in Section 4.2 the structural features of

both hairpins at 300 K are discussed and compared to the NMR and crystallographic results (Ohlenschläger, 2004; Du, 2003; Ennifar, 2000; Allain, 1995). Subsequently, the REMD trajectories are used to describe at the microscopic level the thermostability of the hairpins and to give an insight into the folding-unfolding energy landscape. Finally, the conclusion is drawn in Section 4.3.

4.1 Computational Details

4.1.1 Simulation conditions

The AMBER98 force field (Cheatham, 1999; Cornell, 1995) was employed to describe the 14-mer RNA hairpins. The hairpin was placed in a rhombic dodecahedron box (edge length approximately 5 nm), which was subsequently filled with 2713 TIP3P water molecules (Jorgensen, 1983). To neutralize the system, 13 sodium ions were placed randomly in the simulation box. A twin range cutoff was used for the Lennard-Jones interactions, that is, interactions between atoms within 1.0 nm were evaluated every step, while interactions between atoms within 1.4 nm were evaluated every 5 steps. The particle mesh Ewald method (Darden, 1993) was employed to treat Coulomb interactions, using a switching distance of 1.0 nm, a grid of 0.12 nm and a beta value of 3.1 nm^{-1} . Constant pressure p and temperature T were maintained by weakly coupling the system to an external bath at 1 bar and at the chosen temperature, using the Berendsen barostat and thermostat, respectively (Berendsen, 1984). The RNA, the ions, and the solvent were independently coupled to the temperature bath with a coupling time of 0.1 ps. The pressure coupling time was 0.5 ps and the isothermal compressibility $4.5 \cdot 10^{-5} \text{ bar}^{-1}$. The bond distances and the bond angle of the solvent water were constrained using the SETTLE algorithm (Miyamoto, 1992). All other bond distances were constrained using the LINCS algorithm (Hess, 1997). A leapfrog integrator with an integration time step of 2 fs was used. The starting structure of the 14-mer uCACGg hairpin was taken from the Brookhaven Protein Data Bank (PDB structure 1RFR.pdb) (Ohlenschläger, 2004). The cUUCGg hairpin was modeled using the crystallographic structure of Ennifar et al. (Ennifar, 2000) for the loop and a structure built by AMBER6 tools (Case, 1999) for the stem. The structure was then equilibrated at 300 K for 20 ns.

To study the effect of the water model on the unfolding behavior of RNA hairpins, the cUUCGg simulation was repeated using the TIP4P-Ew model (Horn, 2004).

4.1.2 Replica-exchange molecular dynamics

The replica-exchange molecular dynamics (REMD) (Okabe, 2001; Sugita, 1999) simulations were performed using the GROMACS suite of programs (version 3.3) (van der Spoel, 2005; Lindahl, 2001; Berendsen, 1995b). A number of 48 replicas were chosen with temperatures between 297 and 495 K and a 30 ns REMD simulation for each hairpin was performed. Exchange between replicas was attempted every 20 ps. The temperatures have been selected to obtain exchange probabilities between 15 and 33 %. On average, 1.05 folding-unfolding events are observed in 1 ns simulation time. The maximum number of events per nanosecond in one replica is 6.5 and 3.5 for the uCACGg and the cUUCGg hairpin, respectively. In 13 replicas over 48, the systems do not experience a folding-unfolding process during 30 ns REMD simulations. Those replicas performed at temperatures lower than 390 K.

4.1.3 Trajectory analysis

The analysis of the trajectories was performed with tools from the GROMACS package and with modified versions of them. To define the presence of a hydrogen bond, an acceptor-donor distance smaller than 0.35 nm was requested.

To compare the conformational fluctuations of both hairpins, a principal component analysis (Amadei, 1993; García, 1992; Ichiye, 1991) of the merged trajectory of both hairpins was performed at 300 K. Only backbone atoms were considered. Before performing the analysis, each conformation was translated and rotated to give the best fit to the averaged structure. The first three (of in total 249) eigenvectors describe 64 % of the total backbone fluctuations.

To define reaction coordinates for the thermal unfolding of the hairpins, stem base-base hydrogen bonds and base stacking interactions were used. In the first case, the presence of the N-H ··· N hydrogen bond between Watson-Crick base-pair and N-H ··· O hydrogen bonds between wobble base-pair is used as indicator. A total number of 5 hydrogen bonds was used for uCACGg and cUUCGg hairpin (one each stem base-pair).

In the second approach, two bases were considered as stacked when their center of mass separation is within 0.15 nm of that seen in the folded structure (simulation at 300 K). Pairs with larger separations are considered as broken (Sorin, 2002). The 8 stem base stackings, one stem-loop (5:6), and one inside-loop (6:8) base stacking interactions were taken into account. Employing these coordinates, the Gibbs free energy is given by

$$\Delta G(n_H, n_S) = -k_B T \ln [P(n_H, n_S) - \ln P_{\min}] \quad (4.1)$$

where n_H and n_S are the number of selected hydrogen bonds and base stackings, respectively. $P(n_H, n_S)$ denotes the probability of finding a conformation with interactions (n_H, n_S) , and P_{\min} is the population probability of the global minimum with $\Delta G \equiv 0$.

The melting of the uCACGg and cUUCGg hairpins was monitored via the fraction $P_H = \langle n_H \rangle / n_H^{\max}$ of hydrogen bonds between the stem base-pairs and the fraction $P_S = \langle n_S \rangle / n_S^{\max}$ of base stacking interactions. To describe the correlation of two hydrogen bonds n and m , the normalized covariance matrix was calculated

$$\sigma_{nm} = \frac{\langle (P_{H_n} - \langle P_{H_n} \rangle) \cdot (P_{H_m} - \langle P_{H_m} \rangle) \rangle}{\sqrt{\langle (P_{H_n} - \langle P_{H_n} \rangle)^2 \rangle \cdot \langle (P_{H_m} - \langle P_{H_m} \rangle)^2 \rangle}} \quad (4.2)$$

where P_{H_n} is 0 or 1, depending on if the n th hydrogen bond is broken or closed.

4.2 Results and Discussion

4.2.1 Characterization at 300 K

The structures sampled by the 30 ns REMD simulation at 300 K for uCACGg and cUUCGg hairpins are in overall agreement with the corresponding experimental structures (Ohlenschläger, 2004; Du, 2003; Ennifar, 2000; Allain, 1995). The root mean square deviation (RMSD) has been calculated for all backbone atoms of residues 7-11 after performing a least-squares fit of the REMD and NMR average structures. The simulated uCACGg loop shows a deviation of 0.12 and 0.10 nm with respect to the NMR structures of Ohlenschläger and coworkers (Ohlenschläger, 2004) and of Du and coworkers (Du, 2003), respectively. A low RMSD values (0.06 nm) with respect to the

NMR structures of Allain and Varani (Allain, 2000) is likewise observed for the cUUCGg tetraloop.

Similar to experiment, the two hairpins showed a large similarity of the structure. As an illustration, Figure 4.1 shows representative snapshots of each hairpin at 20 ns. The residues forming the stem are all involved in Watson-Crick base-pairs and stacking interactions. In the uCACGg hairpin, the closing residues form a wobble base-pair. Both loops are mainly stabilized by hydrogen bonds between residues C6/U6 and G9 and residues U7/A7 and C8, as well as by the stacking between residue 6 and residues 5 and/or 8. The residue in position 7 is looped out and the residue C8 is unpaired. All bases are in anti-conformation, except for G9 and A7, which show a syn-conformation. Experimentally, the syn-conformation is observed only for the residue G9. Furthermore, C₃'-endo state is observed for all residues, except for the residues in position 7 and 8 which are in C₂'-endo, as also observed in experiment.

It is instructive to compare the hydrogen bond network observed for both loops to the experimental evidence (Ohlenschläger, 2004; Du, 2003; Ennifar, 2000; Allain, 1995). In the uCACGg simulation, the C6 2'-hydroxyl, the C8 amino and the G9 imino hydrogens are all involved in hydrogen bonding, in agreement with NMR experiments. The experiments show indeed that these protons are protected from rapid exchange (Du, 2003). In particular, residues C6 and G9 are involved in base-base hydrogen bonds and there is a hydrogen bond between the C8 base and the A7 phosphate oxygen as observed in the experimental structures. The calculated and NMR refined structures differ only in the C6 2'-hydroxyl hydrogen. This hydrogen is involved in intra-residue hydrogen bond with other hydroxyl groups in the MD trajectory, while it is hydrogen-bonded to the residue G9 in NMR refinement. In the UUCG loop, the residues U6 and G9 are involved in base-base and in base-sugar hydrogen bonds. The U7 phosphate oxygen is hydrogen-bonded as observed in the experimental structures (Ennifar, 2000; Allain, 1995). Moreover, a weak interaction is observed between the 2'-OH group of U7 and the base G9. In the crystallographic structure (Ennifar, 2000), the U7 sugar oxygen is hydrogen-bonded to the G9 base oxygen, while this is not the case in the NMR structures. Calculating the average probabilities of all hydrogen bonds in the loops, 2.5 versus 3.5 hydrogen bonds are indicated for the CACG and the UUCG loop, respectively.

It is interesting to study to what extent the above discussed structural features are reflected in the conformational fluctuations of both hairpins. To this end, a principal component analysis (Amadei, 1993; García, 1992; Ichiye, 1991) of the merged trajectory of both hairpins at 300 K (see Computational Details) has been performed, and the projection of each individual hairpin trajectory has been considered on the first three eigenvectors of the principal component analysis. Figure 4.2 shows this projection as a function of the first and the sum of the second and third eigenvectors. Despite of the differences in the sequence, the trajectories of the two hairpins show large overlapping regions. However, for the cUUCGg hairpin, a separated smaller region, which indicates a second conformation state with a population of 23 %, is also observed. In agreement with previous MD studies (Villa, 2006; Miller 1997), an analysis of the calculated backbone dihedral angles indeed reveals a conformational rearrangement in the UUCG loop region.

To identify the origin of the difference between the conformational distributions of the hairpins, the atomic root mean square fluctuation of the two RNAs with respect to their average structures was considered. Figure 4.3 shows that the main difference of the atomic fluctuations is observed in the loop region, i.e., the UUCG loop is more flexible than the CACG loop. In particular, the atoms around the sixth phosphor atom deviate significantly from the average structure. This is mainly related to the backbone dihedral angle transitions and the C₃-endo/C₂-endo equilibrium of the sugar ring, observed for residues 6 and 7 in cUUCGg hairpin, while the uCACGg shows always single values for these quantities. Recalling that the UUCG loop contains more stabilizing hydrogen bonds, there is the intriguing situation that the loop with more hydrogen bonds is also the one with higher flexibility.

The presence of looped-out residues and potential hydrogen bond donor and acceptor groups are usually regarded as structural features that affect the functionality of an RNA hairpin, such as its attitude to bind other RNA or protein molecules. Indeed, both investigated hairpins have a looped-out residue in position 7, but the base of this residue adopts a different orientation relative to the sugar moiety in the two hairpins. A7 is in syn-conformation where the dihedral angle χ between the sugar and the base moiety holds a value of -51° (72 %) and 29° (28 %) during the simulation. This allows

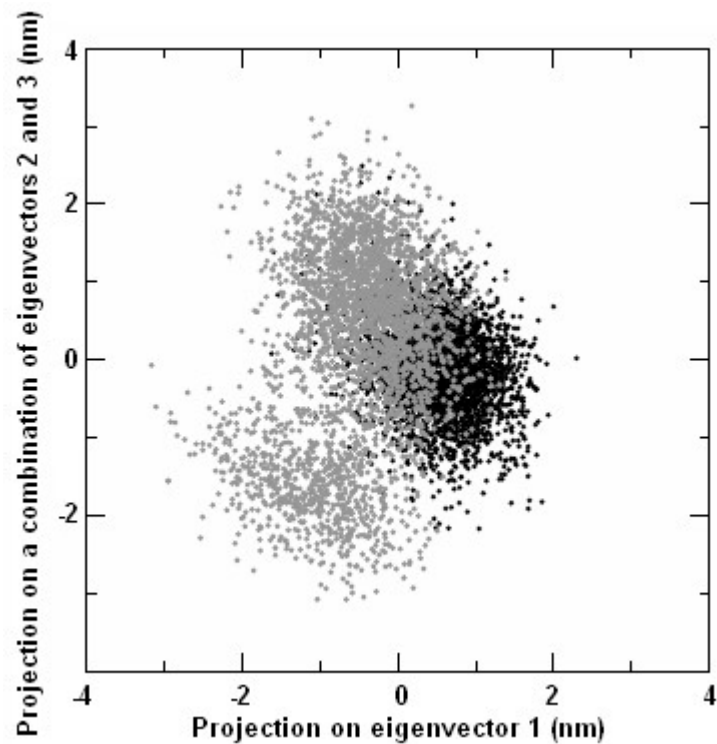


Figure 4.2 Conformations of the uCACGg (black) and cUUCGg (grey) hairpins at 300K, projected on the first eigenvectors of a principal component analysis.

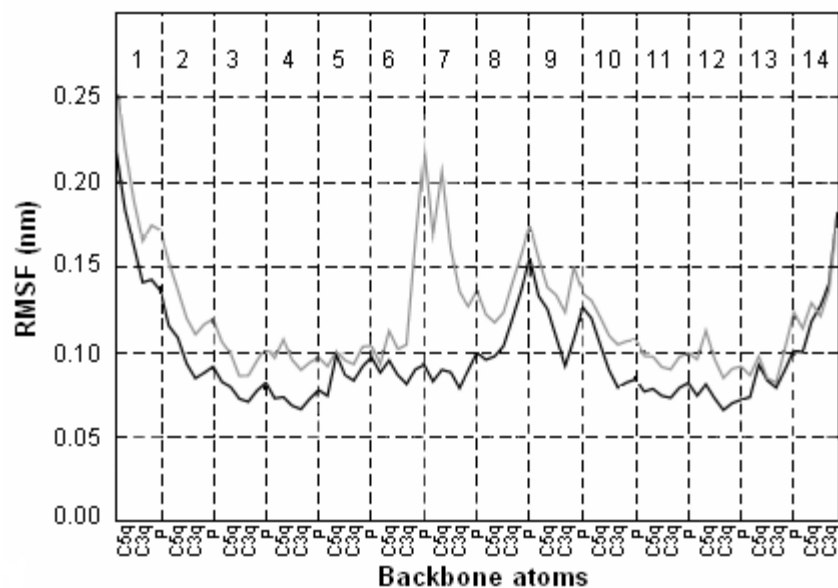


Figure 4.3 Root mean square fluctuations, RMSFs, of the backbone atoms of the uCACGg (black) and cUUCGg (grey) hairpins at 300 K.

the amino group to be fully accessible to the solvent. On the other hand, U7 is in equilibrium between anti (80 %) and syn (20 %) conformations and has two hydrogen bond acceptors on the surface. The total solvent accessible surface of the two RNAs is quite similar, 27.9 nm² and 27.5 nm² for uCACGg and cUUCGg, respectively. In uCACGg, residues 5, 7 and 9 are slightly more exposed and residues 6 and 8 are slightly less exposed than in cUUCGg structures. While the global solvent accessible surface shows only minor difference, the type of functional groups on the surface are quite different. In the simulation of uCACGg, the amino group of A7, the carbonyl groups of C8 and G9, and a C-H group of G9 were fully accessible to the solvent. In the simulation of cUUCGg, the carbonyl groups of U6, U7 and C8, as well as the C-H groups of U7 and G9 were exposed to the water. Hence, the analysis indicates a stronger attitude to donate hydrogens for the uCACGg hairpin than for the cUUCGg hairpin.

4.2.2 Thermal unfolding

Providing the free energy landscape at all temperatures of interest, REMD simulations are an ideal means to study the folding and unfolding of biomolecules (García, 2003; Sorin, 2003; Zhou, 2003; Pande, 1999). One should first study the convergence of the REMD calculations with respect to the simulation time. As a representative example, Figure 4.4 shows the mean potential energy as well as the fraction of base-base hydrogen bonds at six different time intervals of the cUUCGg simulation. Since all replicas start with the equilibrium structure at 300 K, the first few nanoseconds of the high-temperature replicas exhibit clearly too low energies (by up to 100 kJ/mol) compared to later times, when all replicas are equilibrated according to their temperature. Similarly, the fraction of base-base hydrogen bonds is found to shift significantly during the first 10 ns of the simulation. In the discussion below, therefore only the last 20 ns of the REMD simulations of both hairpins are used. On average, one folding-unfolding event per nanosecond simulation time is observed for each replica, resulting in a total number of 10³ such events.

The melting of the uCACGg and cUUCGg hairpins was monitored via the fraction P_H of hydrogen bonds between the stem base-pairs and by the fraction P_S of base stacking interactions, see Computational Details. Plotted as a function of temperature, Figure 4.5 shows these observables which run from 1 (folded state) to 0 (unfolded state).

Defining the melting temperature T_m by $P_H(T_m) = 0.5$, $T_m = 392$ and 425 K are obtained for the uCACGg and the cUUCGg hairpin, respectively. Employing the criterion $P_S = 0.5$, the melting temperature increases by about 14 (uCACGg) and 4 K (cUUCGg). Averaging over both criteria, a computational melting temperature difference of $\Delta T_m = 28$ K is obtained. Although there are no experimental data available for the two 14-mer hairpins under consideration, various experimental studies on similar RNA hairpin suggest that the melting temperatures are ≈ 330 and 350 K for the uCACGg and the cUUCGg hairpin, respectively (Ohlenschläger, 2004; Du, 2003; Proctor, 2002). While the experimental difference in melting temperatures $\Delta T_m \approx 20$ K is reproduced by the REMD simulations within statistical and experimental uncertainties, the absolute melting temperatures are about 20 % too high.

This significant discrepancy between measured and simulated transition temperatures is well known from numerous peptide-folding studies (García, 2003; Sorin, 2003; Zhou, 2003; Pande, 1999). Besides sampling problems (see above), it is most likely related to deficiencies of the force field, since standard biomolecular force fields have been parameterized to reproduce properties at room temperature. Apart from the modeling of the solute, the high-temperature description of the aqueous solvent is a well-known problem. For example, it is known that the isobaric thermal expansion of TIP3P water is a factor of three higher than in experiment and its temperature derivative a factor of two lower (Hess, 2006). This quantity is a measure of the enthalpic contribution to the reorganization of the solvent and is important in the solvation of hydrophobic residues. To study the effect of the water model on the unfolding behavior of RNA hairpins, the cUUCGg simulation using the TIP4P-Ew model (Horn, 2004) was repeated. This water model was parameterized to reproduce both experimental density and enthalpies of vaporization at different temperatures ranging from 235.5 to 400 K, and therefore describes well both kinetic and thermodynamic properties, such as thermal expansion, heat capacity, self-diffusion coefficient and compressibility. Nonetheless, Figure 4.5 reveals that the melting curves obtained for the TIP4P-Ew model are within statistical errors equivalent to the results for the TIP3P model. Apparently, the discrepancy between measured and simulated

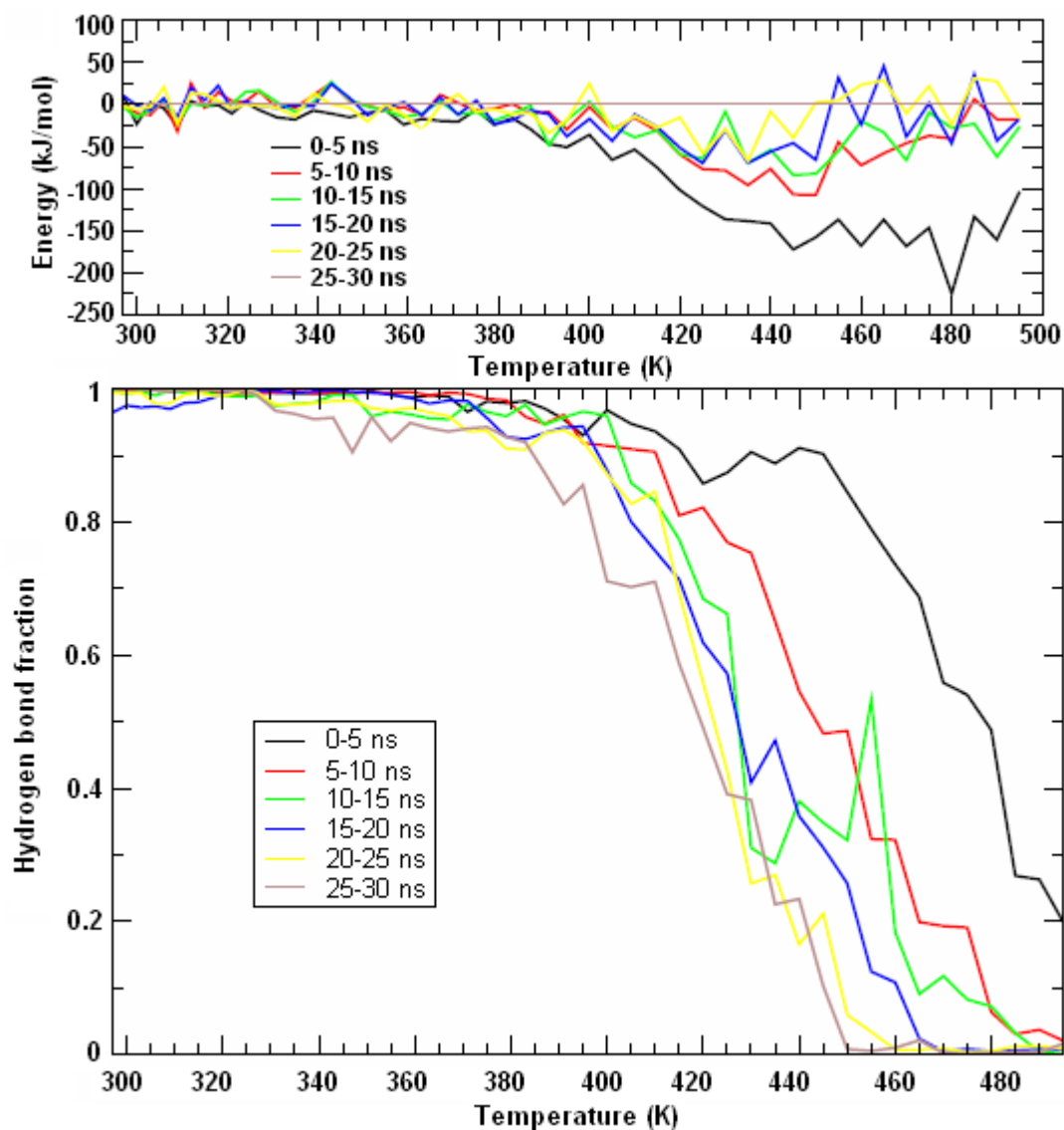


Figure 4.4 Convergence of the REMD simulations of the cUUCGg hairpin, monitored for six different time intervals. Shown are (above) the mean potential energy (in kJ/mol, relative to the energy of the last 5 ns of the simulation) and (below) the fraction of base-base hydrogen bonds.

transition temperatures cannot directly be attributed to the mentioned thermodynamic properties of the water model.

After checking the quality of the force field and the sampling, the unfolding mechanism of the two RNA hairpins is going to be investigated. Apart from the lower transition temperature, Figure 4.5 reveals that the width of the melting curve of the uCACGg hairpin is smaller than for the cUUCGg hairpin. Defining the width as $\Delta T = T(P_H = 0.9) - T(P_H = 0.1)$, $\Delta T = 50$ and 65 K are obtained for the uCACGg and

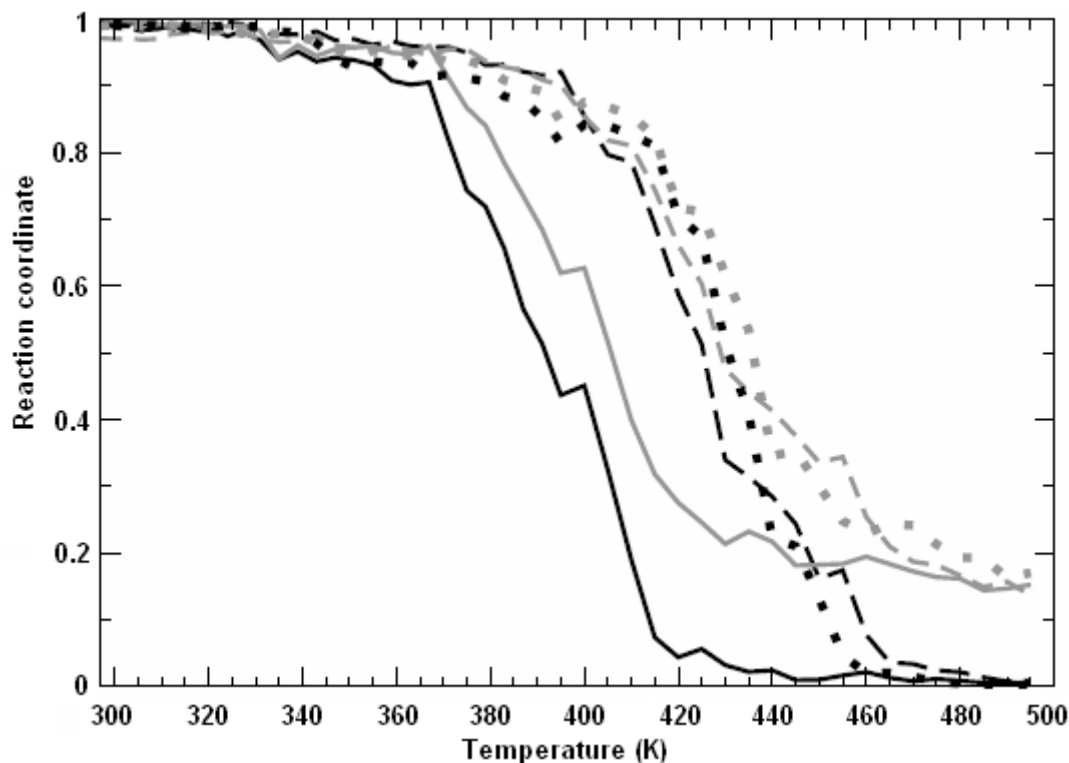


Figure 4.5 Melting curves of the uCACGg (full line) and the cUUCGg (dashed line) hairpin, using the fraction of hydrogen bonds between the stem base-pairs (black) and the fraction of base stacking interactions (grey) as reaction coordinates. Also shown are results for cUUCGg obtained for the TIP4P-Ew water model (dotted line).

the cUUCGg hairpin, respectively. This finding indicates that the folding-unfolding process of the uCACGg hairpin is more cooperative than for the cUUCGg hairpin.

The phenomenon is studied in more detail in Figure 4.6, which shows the temperature-dependent probability of the hydrogen bonds between the individual base-pairs of the stem. Apart from the first base-pair G1:C14, which partially opens already at room temperature, the remaining stem hydrogen bonds are found to open in a concerted manner. A notable exception is the G12:C3 hydrogen bond of cUUCGg, which clearly

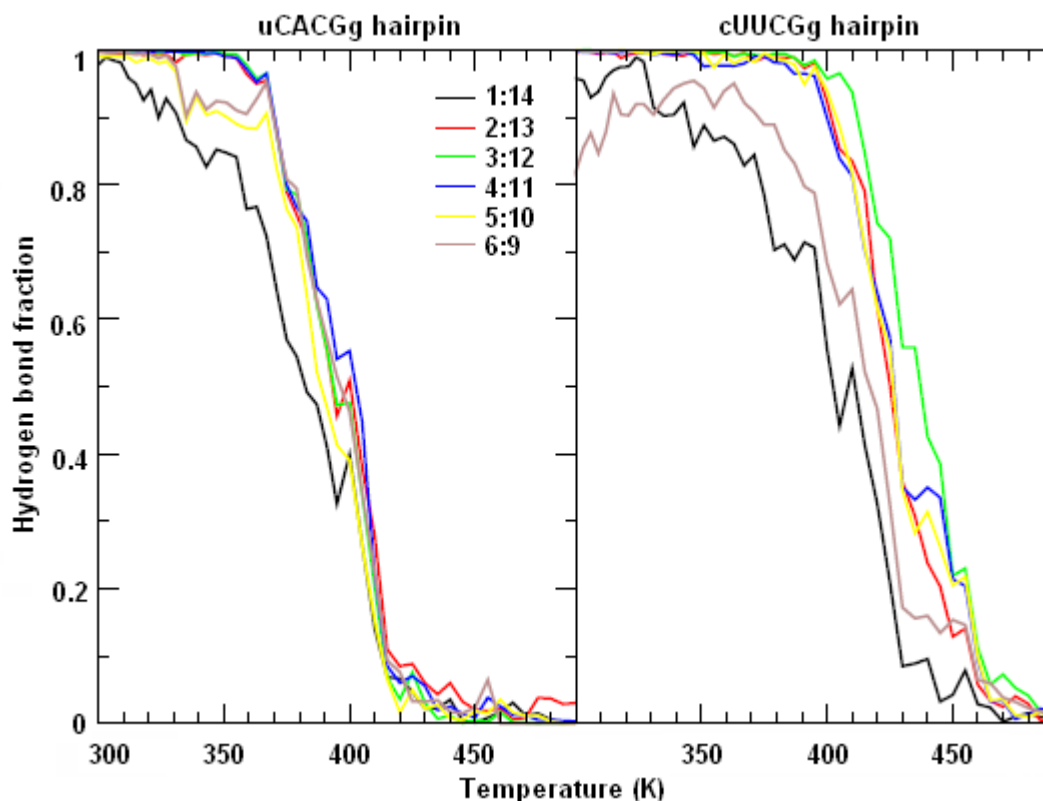


Figure 4.6 Melting curves of the selected hydrogen bonds between individual base-pairs of the two 14-mer RNA hairpins. Shown are temperature dependent hydrogen-bond probabilities of the stem and loop base-pairs.

remains more stable than the other base-pairs of the stem. A further difference between the two hairpins is found in the melting curve of the hydrogen bond between the *loop* residues 6 and 9. While the base-pair C6:G9 of the uCACGg loop appears to open along with the stem, the U6:G9 hydrogen bond in the UUCG loop is less cooperative and also less stable than the stem hydrogen bonds. Moreover, some *premelting* involving the closing (U5:G10) and the loop (C6:G9) base-pairs was observed for the uCACGg hairpin.

To assess the cooperativity in the making and breaking of base-pair hydrogen bonds, the covariance matrix σ_{nm} (equation 4.2) has been calculated, which describes the correlation between two hydrogen bonds n and m . The calculation has been restricted to temperatures in the melting region (370-420 K and 400-465 K for uCACGg and cUUCGg, respectively), in order to focus on the melting process. In agreement with the above observations, Table 4.1 reveals (i) that the hydrogen bonding correlations are generally higher for the uCACGg than for the cUUCGg hairpin, (ii) that the lowest

Table 4.1 Correlation σ_{nm} between two hydrogen bonds n and m , as defined in equation 4.3. Considered are the hydrogen bonds associated with the base-pair (i:j) of the stem of the cUUCGg and the uCACGg hairpin, respectively. Since $\sigma_{nm} = \sigma_{mn}$, only the upper part of the matrix is shown.

n/m	σ_{nm} (cUUCGg)					σ_{nm} (uCACGg)				
	(1:14)	(2:13)	(3:12)	(4:11)	(5:10)	(1:14)	(2:13)	(3:12)	(4:11)	(5:10)
(1:14)	1.00	0.56	0.44	0.51	0.51	1.00	0.69	0.64	0.60	0.58
(2:13)		1.00	0.75	0.68	0.62		1.00	0.87	0.81	0.79
(3:12)			1.00	0.76	0.70			1.00	0.84	0.82
(4:11)				1.00	0.76				1.00	0.78
(5:10)					1.00					1.00

correlation is found between the first (G1:C14) and the remaining base-pairs, and that the making or breaking of the G12:C3 hydrogen bond of cUUCGg is somewhat less correlated with the dynamics of the other base-pairs of the stem.

As an illustration of this cooperativity, the time evolution of the stem hydrogen bonds may be considered along some chosen replica of the REMD simulation. Displaying the number $n_H(t)$ of selected stem hydrogen bonds together with the instantaneous temperature of the replica, Figure 4.7 shows that the rupture of hydrogen bonds occurs collectively in the case of the uCACGg hairpin and via some intermediate states in the case of the cUUCGg hairpin. In both cases, the folded state is characterized by five (or occasionally four) stem hydrogen bonds. With increasing temperature, the uCACGg hairpin unfolds in an all-or-none fashion, while the cUUCGg hairpin exhibits intermediate states characterized by 1 – 3 base-pair hydrogen bonds.

One should finally study the energy landscape obtained for the two RNA hairpins. Figure 4.8 shows the free energy $\Delta G(n_H, n_S)$ of both loops as a function of the number of base-base hydrogen bonds n_H and base stacking interactions n_S , see Computational Details. The energy landscapes are displayed for six different temperatures, ranging from 300 to 490 K. At 300 K, both hairpins are completely in the folded state (with $n_H = 5$, $n_S = 10$) while at 490 K the systems are unfolded (with $n_H = 0$, $n_S = 1$). Along the two coordinates under consideration, the uCACGg hairpin exhibits a simple two intermediate state behavior in the complete temperature range. The energy landscape of cUUCGg hairpin, on the other hand, is more complex and exhibits at least one intermediate state at the transition temperature of 425 K. Located along the diagonal of the $\Delta G(n_H, n_S)$ energy

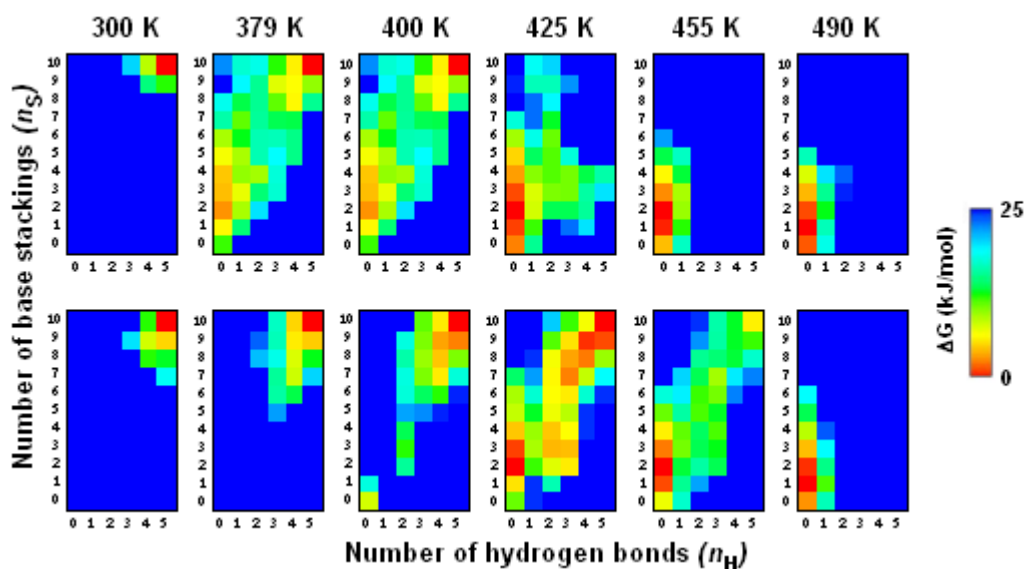


Figure 4.7 Time evolution of the number $n_H(t)$ of selected hydrogen bonds together with the instantaneous temperature, T , of a representative replica, obtained from the REMD simulation of the uCACGg and the cUUcGg hairpin, respectively.

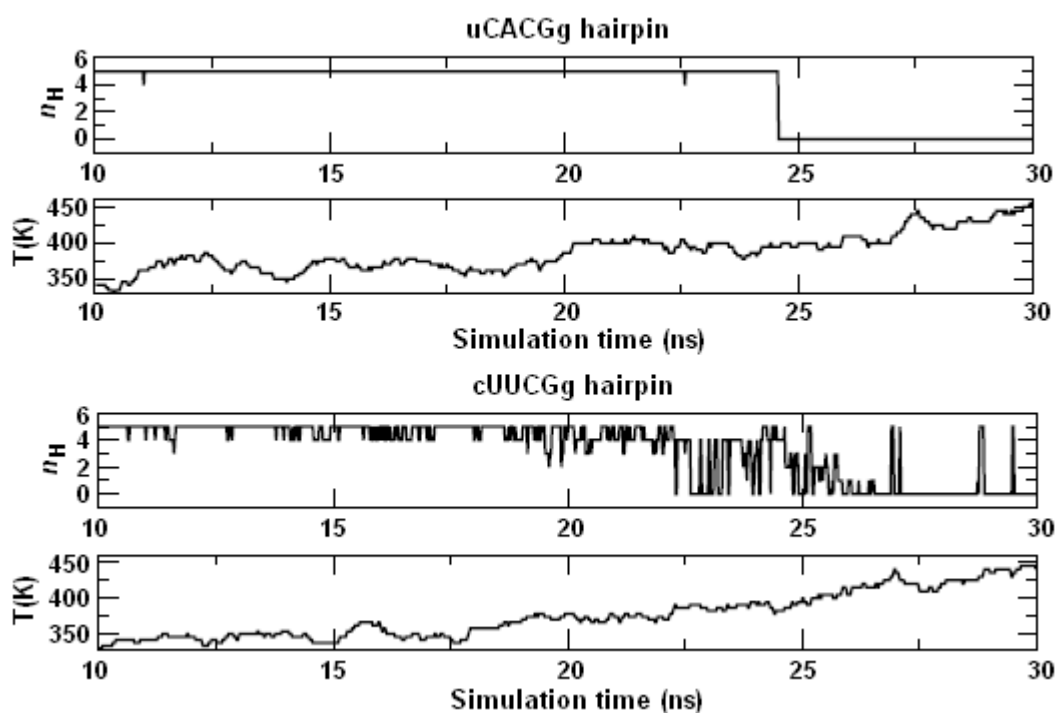


Figure 4.8 Free energy landscape $\Delta G(n_H, n_S)$ (in kJ/mol) of the uCACGg (top) and the cUUcGg (bottom) hairpin at various temperatures, shown as a function of the number of base-base hydrogen bonds n_H and base stacking interactions n_S .

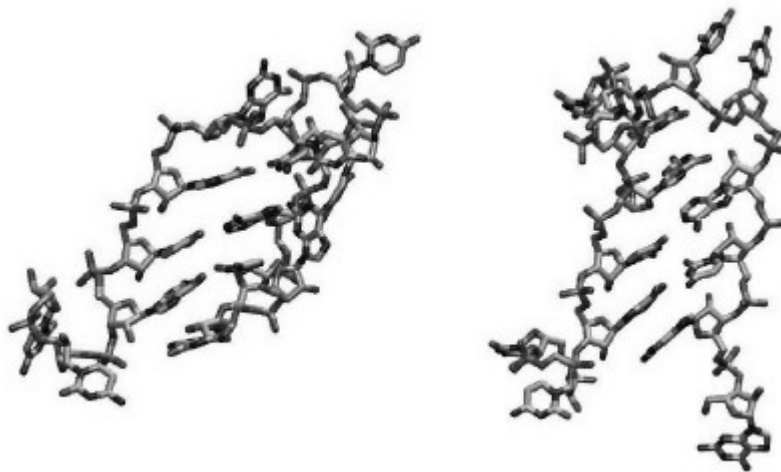


Figure 4.9 Representative structures of the intermediate state of the cUUCGg hairpin at 425 K.

surface, the intermediate states are characterized by $n_H = 2-3$ hydrogen bonds and $n_S = 3-4$ stacking interactions. Figure 4.9 shows representative snapshot of the intermediate-state behavior in the complete temperature range. The energy landscapes of the cUUCGg hairpin, on the other hand, is more complex and exhibits at least one Figure 4.9 shows representative snapshots of the intermediate state at 425 K. In all intermediates, the stem residues C3, C5, G10 and G12 are still involved in native base-pair and/or stacking interactions, while the global helicity of the hairpin is lost.

In summary, the following picture of the unfolding of the two RNA hairpins can be drawn. The REMD study qualitatively reproduces the experimentally measured difference of melting temperatures $\Delta T_m = 20$ K. This effect is not due to differences in sequence, structure, or dynamics of the loop, but appears to be solely caused by the differences in the stem base-pairs (U3:A12 vs. C3:G12) and (U5:G10 vs. C5:G10) of the uCACGg and the cUUCGg hairpin, respectively. In the case of the uCACGg hairpin, the unfolding occurs cooperatively in an all-or-none fashion, thus resulting in a simple two-state behavior (Hyeon, 2005; Sorin 2005; Nivón, 2004; Sorin, 2003; Bonnet, 1998; Wilson, 1995). The cUUCGg hairpin, on the other hand, shows less cooperativity, but exhibits intermediate states in the unfolding process (Ma, 2006; Sorin, 2002; Ansari, 2001; Chen, 2000). While the global helicity of the hairpin is lost in these states, base-pairs C5:G10 and C3:G12 are still involved in native interactions. Interestingly, recent

temperature-jump experiments on an 8-mer cUUCGg hairpin were interpreted by a four-state energy landscape (Ma, 2006).

4.3 Conclusions

Extensive REMD simulations (E 4.3 μs total simulation time) have been performed to study structure, dynamics, and melting of the structurally similar 14-mer RNA hairpins uCACGg and cUUCGg. The simulations have confirmed the experimentally found structural similarities of the two RNA hairpins at room temperature. In particular, the hydrogen bond network and base stacking interactions are quite similar for both systems. Studying the conformational fluctuations, it has been found that the cUUCGg loop is more flexible than the uCACGg loop. Indeed, residues 6 and 7 in cUUCGg hairpin are involved in backbone dihedral angle transitions and C3'-endo/C2'-endo sugar equilibrium. As the cUUCGg loop contains more stabilizing hydrogen bonds, there is the intriguing situation that the loop with more hydrogen bonds is also the one with higher flexibility.

To understand the experimentally found differences in the binding behavior of the RNA hairpins, the relevance of the looped-out residue in position 7 and of potential hydrogen bond donor and acceptor groups has been studied. While the global solvent accessible surface was found to be quite similar for both RNA loops, the type of functional groups on the surface were quite different. In particular, the analysis has indicated a stronger attitude to donate hydrogens for the uCACGg hairpin than for the cUUCGg hairpin.

Providing the free energy landscape at all temperatures of interest, REMD simulations are an ideal means to study the folding and unfolding of biomolecules. Although the calculated absolute melting temperatures are about 20 % too high, the REMD simulations reproduces the experimentally found difference in melting temperatures of $\Delta T_m \approx 20$ K within statistical and experimental uncertainties. Interestingly, this effect is not due to differences in sequence, structure, or dynamics of the loop, but appears to be solely caused by the differences in the stem base-pairs (U3:A12 vs. C3:G12) and (U5:G10 vs. C5:G10) of the uCACGg and the cUUCGg

hairpin, respectively. In the case of the uCACGg hairpin, the unfolding occurs cooperatively in an all-or-none fashion, thus resulting in a simple two-state behavior. The cUUCGg hairpin, on the other hand, shows less cooperativity, but exhibits intermediate states in the unfolding process (Ma, 2006; Sorin, 2002; Ansari, 2001; Chen, 2000). While the global helicity of the hairpin is lost in these states, base-pairs C5:G10 and C3:G12 are still involved in native interactions.

Chapter 5

Internal Motion of RNA Hairpins as Reflected by NMR Relaxation Parameters

Conformational dynamics may play a key role in the function of biomolecules such as proteins, DNA, and RNA. The flexible parts of a protein or ribonucleic acid, such as loop regions, are often involved in mediating specific interactions, for example, between protein and RNA during a binding process (Al-Hashimi, 2005; Leulliot, 2001; Wand, 2001). In the case of RNA, the flexibility of the loop may directly affect both the specificity and the affinity of the binding (Persson, 2002). To account for the function of RNA systems, a site-specific dynamic description is therefore an important complement to static structural information (Zhang, 2006; Koplín, 2005).

Molecular motions in biomolecules occur on a wide range of time scales, from femtoseconds to seconds. Molecular dynamics (MD) simulations and nuclear magnetic resonance (NMR) spin relaxation data are valuable tools to gain access to fast (i.e., subnanosecond) internal motions. MD simulations (Frenkel, 1996) provide directly

information at the atomistic level on inter and intramolecular motions using an empirical force field as molecular model. NMR relaxation measurements (Ernst, 2004) yield the dipolar correlation function, from which dynamical quantities such as generalized order parameters S^2 and effective correlation times τ_e can be extracted (Korzhnev, 2001; Fischer, 1998; Brüschweiler, 1994; Lipari, 1982). In particular, the combination of NMR and MD investigations has been shown to provide a comprehensive description of fast conformational dynamics of proteins (Lange, 2005; Case, 2002; Peter, 2001; Prompers, 2001; Chatfield, 1998; Brüschweiler, 1992; Palmer, 1992; Levy, 1981). On one hand, one may use experimental NMR results as benchmark data to study the accuracy of the MD description, on the other hand, one may employ the MD trajectory to provide a microscopic interpretation of the NMR experiments. While the description of structure and dynamics of proteins is well established, RNA systems have been comparatively little studied using MD simulation (Auffinger, 2001; Auffinger, 2000; Cheatham III, 2000; Zacharias, 2000) and NMR relaxation (Duchardt, 2005; Showalter, 2005; Vallurupalli, 2005; Chiarparin, 2001; Akke, 1997).

In this Chapter the fast dynamics of the 14-mer RNA hairpins (Figure 4.1), cUUCGg and uCACGg, is studied using 50 ns MD simulations and the results are compared with available NMR relaxation data. The cUUCGg and uCACGg hairpins have structural similarities at the room temperature, but their thermal folding-unfolding transition appears to be significantly different (Ohlenschläger, 2004; Du, 2003; Proctor, 2002). Furthermore, the biological roles of the both hairpin in protein recognition seem to be also different (Ohlenschläger, 2004; Du, 2003). Thus the comparison of the dynamical behaviors of the two hairpins at room temperature is of fundamental interest since local flexibility in RNA may facilitate protein recognition (Showalter, 2005). In Chapter 4, the ability of the simulations to describe the 14-mer RNA hairpins at room temperature was deeply discussed.

This Chapter is organized into three major sections. Section 5.1 describes the simulation condition and theory used to back-calculate the NMR relaxation parameters from MD trajectory. In Section 5.2 the results have been discussed in the following way. First, the simulation of the cUUCGg hairpin is used to directly calculate the NMR relaxation rates (Peter, 2001), in order to avoid most of the assumptions usually employed in experimental analysis (Section 5.2.1). Second, various ways are considered

to provide a link between theory and experiment, including motional models such as the Gaussian axial fluctuation model (Brüschweiler, 1994) and the so-called model-free approach developed by Lipari and Szabo (Lipari, 1982) (Section 5.2.2). In particular, it is studied if the underlying assumptions of these approaches (such as the separation of overall and internal motion and the use of mono-exponential Lipari-Szabo fits of the internal correlation function) are satisfied in the case of a flexible RNA hairpin. Subsequently, the relation between NMR order parameters and the underlying internal motion of the cUUCGg hairpin is discussed in detail (Section 5.2.3). Finally, the fast dynamics of the cUUCGg and uCACGg hairpins is compared in Section 5.2.4. The conclusions are in Section 5.3.

5.1 Theory and Computational Details

5.1.1 Molecular dynamics simulations

The MD simulations were performed using the GROMACS suite of programs (version 3.2) (Lindahl, 2001; Berendsen, 1995b). The AMBER98 force field (Cheatham, 1999; Cornell, 1995) was employed to describe the 14-mer cUUCGg and uCACGg RNA hairpins. The hairpin was placed in a rhombic dodecahedron box (edge length approximately 5 nm), which was subsequently filled with 2713 and 2721 TIP3P water molecules (Jorgensen, 1983) for cUUCGg and uCACGg hairpins respectively. To neutralize the system, 13 sodium ions were placed randomly in the simulation box.

A twin range cut-off was used for the Lennard-Jones interactions, that is, interactions between atoms within 1.0 nm were evaluated every step, while interactions between atoms within 1.4 nm were evaluated every 5 steps. The particle mesh Ewald method (Darden, 1993) was employed to treat Coulomb interactions, using a switching distance of 1.0 nm. Constant pressure p and temperature T were maintained by weakly coupling the system to an external bath at 1 bar and 298 K, using the Berendsen barostat and thermostat, respectively (Berendsen, 1984). The RNA, the ions, and the solvent were independently coupled to the temperature bath with a coupling time of 0.1 ps. The pressure coupling time was 0.5 ps and the isothermal compressibility $4.5 \cdot 10^{-5} \text{ bar}^{-1}$. The bond distances and the bond angle of the solvent water were constrained using the

SETTLE algorithm (Miyamoto, 1992). All other bond distances were constrained using the LINCS algorithm (Hess, 1997). A leap-frog integrator with an integration time step of 2 fs was used.

Following 20 ns of equilibration, the systems were simulated for 50 ns. Analysis of the trajectories was performed with tools from the GROMACS package and with modified versions of them. To define the presence of a hydrogen bond, an acceptor-donor distance smaller than 0.35 nm was requested.

5.1.2 NMR Relaxation Parameters

The spin-lattice (R_1), the spin-spin (R_2) relaxation rates, and the nuclear Overhauser enhancement (NOE) are given by (Ernst, 2004)

$$R_1 = \frac{d}{4} \left[J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H + \omega_C) + cJ(\omega_C) \right] \quad (5.1)$$

$$R_2 = \frac{d}{8} \left[4J(0) + J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H) + 6J(\omega_H + \omega_C) \right] \quad (5.2)$$

$$+ \frac{c}{6} \left[4J(0) + 3J(\omega_C) \right]$$

$$\text{NOE} = 1 + \frac{d}{4R_1} \frac{\gamma_C}{\gamma_H} \left[6J(\omega_H + \omega_C) - J(\omega_H - \omega_C) \right] \quad (5.3)$$

where $d = \frac{\mu_o^2}{4\pi^2} \frac{\hbar^2 \gamma_C^2 \gamma_H^2}{r_{CH}^6}$, $c = \frac{\omega_C^2 \Delta\sigma_C^2}{3}$, and $J(\omega)$ represents the spectra density defined in equation 5.5. Here, μ_o is the vacuum permeability, \hbar is Planck's constant divided by 2π , and γ_X is the gyromagnetic ratio of nucleus X. r_{CH} is the intermolecular distance between the two nuclei; for the bond C₁-H₁, distance of 0.109 nm and for C₆-H₆ and C₈-H₈ a distance of 0.108 nm was used. $\Delta\sigma_C$ is the ¹³C chemical shift anisotropy; for C₁, a value of 45 ppm, for C₆ a value of -179 ppm and for C₈ a value of -134 ppm were used. A value of 600.13 MHz was used for ω_H and 150.90 MHz for ω_C . All these values have been chosen in line with the NMR relaxation experiments of Duchardt and Schwalbe (Duchardt, 2005).

5.1.3 Correlation functions

The NMR relaxation due to the dipole-dipole interaction between two nuclei (i.e., carbon and hydrogen) can be described by the correlation function (Ernst, 2004)

$$C(t) = \langle P_2(\boldsymbol{\mu}(0) \cdot \boldsymbol{\mu}(t)) \rangle \quad (5.4)$$

where $\boldsymbol{\mu}$ is a unit vector pointing along the C-H bond, $P_2(x) = \frac{1}{2}(3x^2 - 1)$ is the second Legendre polynomial, and $\langle \cdot \rangle$ denotes an equilibrium average. The spectral density

$$J(\omega) = 2 \int_0^{\infty} C(t) \cos(\omega t) dt \quad (5.5)$$

which determines the relaxation parameters in equations 5.1 – 5.3, is given by the Fourier transform of the correlation function.

Assuming that overall and internal motions of the molecule are independent, the total correlation function $C(t)$ can be factorized in the correlation functions for overall motion, $C_o(t)$, and for internal motion, $C_i(t)$, respectively:

$$C(t) = C_o(t) C_i(t) \quad (5.6)$$

The total correlation functions were calculated for the C₁-H₁, C₆-H₆ and C₈-H₈ dipoles of all residues according to equation 5.4. To obtain the internal correlation functions, each conformation was translated and rotated to give the best fit to a reference structure. Since no large conformational arrangement took place during the 50 ns simulation, the molecule-fixed frame is unambiguously defined by this approach. Subsequently, the correlation functions for overall motion were calculated using equation 5.6. Assuming that the overall motion of the molecule is isotropic, this correlation function is given by

$$C_o(t) = \frac{1}{5} e^{-t/\tau_c} \quad (5.7)$$

where the rotational correlation time τ_c is proportional to the inverse of the rotation diffusion constant.

In the model-free approach of Lipari-Szabo (Lipari, 1982), the internal and overall motions are assumed to be independent and the internal correlation function is given by the following relation:

$$C_I(t) = S^2 + (1 - S^2) e^{-t/\tau_e} \quad (5.8)$$

where S^2 is the order parameter and τ_e is the effective (or internal) correlation time τ_e for the C–H dipole. Insertion of equation 5.8 in equations 5.5 – 5.7 yields the spectral density

$$J(\omega) = \frac{2}{5} \left(\frac{S^2 \tau_c}{1 + \tau_c^2 \omega^2} + \frac{(1 - S^2) \tau}{1 + \tau^2 \omega^2} \right) \quad (5.9)$$

with $\tau^{-1} = \tau_c^{-1} + \tau_e^{-1}$.

5.1.4 Order parameter

Three different approaches have been employed to calculate the order parameters.

Lipari-Szabo fit: Employing the Lipari-Szabo form of the internal correlation function (equation 5.8), S^2 was fitted using the first 100 ps (or the first 1 ns) of the MD internal correlation.

Equilibrium average: Using the general property of correlation functions that $\lim_{t \rightarrow \infty} \langle A(0) B(t) \rangle = \langle A \rangle \langle B \rangle$, the order parameter can be determined by (Ernst, 2004)

$$S^2 = \lim_{t \rightarrow \infty} C_I(t) = \frac{4\pi}{5} \sum_{m=-2}^{m=2} \left| \langle Y_{2m}(\theta, \varphi) \rangle \right|^2 \quad (5.10)$$

where Y_{2m} is the spherical harmonic function of rank 2, $\theta(t)$ and $\varphi(t)$ are the polar angles defining the orientation of the dipole C-H at each snapshot of the trajectory, and $\langle \dots \rangle$ denotes the average over all snapshots. This corresponds to a Lipari-Szabo fit using the full time range of the internal correlation function. It should be stressed that equation 5.10 avoids the cumbersome calculation of time-dependent correlation functions. In particular, this is allowed to use highly efficient Monte-Carlo schemes [e.g., like the popular replica exchange MD (Sugita, 1999)] to calculate the equilibrium average in equation 5.10.

GAF model: Assuming that the nucleobase flexibility monitored by the order parameters of C6/C8 is exclusively caused by base motions along the glycosidic torsional angle χ , the order parameter S^2 can be related to motions around the C₁–N₁ or C₁–N₉ bonds. Assuming furthermore a Gaussian distribution for the dihedral angle χ , the Gaussian axial fluctuation (GAF) model (Brüschweiler, 1994) leads to the following expression for the order parameter:

$$S^2 = 1 - 3 \sin^2 \chi \left\{ \cos^2 \chi \left(1 - e^{-\sigma_\chi^2} \right) + \frac{1}{4} \sin^2 \chi \left(1 - e^{-4\sigma_\chi^2} \right) \right\} \quad (5.11)$$

Here the dihedral angle χ is defined by O₄–C₁–N₁–C₂ in the pyrimidine and by O₄–C₁–N₉–C₄ in the purine, and σ_χ is its standard deviation.

5.2 Results and Discussion

5.2.1 NMR relaxation parameter of the cUUCGg hairpin

The internal and total correlation functions of all C₁–H₁, sugar bonds and C₆–H₆ and C₈–H₈ base bonds of the cUUCGg hairpin have been calculated according to equation 5.4, using the 50 ns trajectory with and without subtracting the overall motion, respectively. The total correlation functions decay on average within a nanosecond (see Figure 5.1), while the internal correlation functions generally show a decay on a time scale of ten picoseconds (see Figure 5.2 for some representative examples). Mono-exponential fits of the internal correlation function are seen to be appropriate in most cases, except for the loop residues U6 – G9. In the latter cases, the internal correlation functions exhibit a multi-exponential decay on pico- and nanosecond time scales.

The validity of the assumption that overall and internal motions are separable has been studied. The separation of the motions leads to the factorization of the total correlation function $C(t)$ into components $C_o(t)$ and $C_i(t)$ describing overall and internal motion, respectively. As a representative example, Figure 5.1 shows the three correlation functions for the sugar and base dipole motions of the relatively rigid stem residue C3 and the most flexible loop residue U7. The overall and internal correlation functions of the stem residue exhibit an obvious separation of time scales (1 ns vs. 10 ps) and are

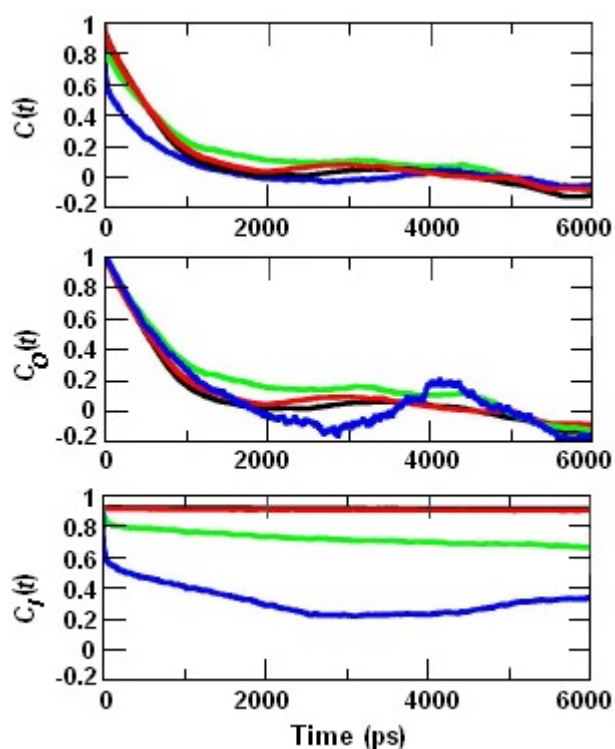


Figure 5.1 Total (top), overall-motion (middle), and internal (bottom) correlation functions of the dipoles $C_1\text{-H}_1$ (black and green line) and $C_6\text{-H}_6$ (red and blue line) for the residues C3 and U7, respectively.

therefore clearly separable. In the case of the loop residue U7, on the other hand, both correlation functions decay on a nanosecond time scale, and one may expect a coupling of overall and internal motions. However, for the relatively short times (0.1 – 1 ns) that are relevant in the analysis of the experimental NMR data, the internal correlation function $C_i(t)$ may be approximated by a 24 ps decay time (see Figure 5.2) and $C_o(t)$ decays just as the other overall-motion correlation functions. That is, for short times the factorization approximation is not expected to change the results of the calculation of NMR data, although the overall and internal motions of the flexible loop residue U7 are not separable in general.

The correlation times τ_c were obtained by fitting the overall-motion correlation function of each dipole to the mono-exponential function e^{-t/τ_c} . The fitted τ_c has an average value of 0.7 ns, which is clearly shorter than values obtained by using a hydrodynamics model (Duchardt, 2005) for the whole hairpin (2.35 ns for C_1 and 2.17 ns for C_6/C_8). The main reason for this deviation appears to be the different

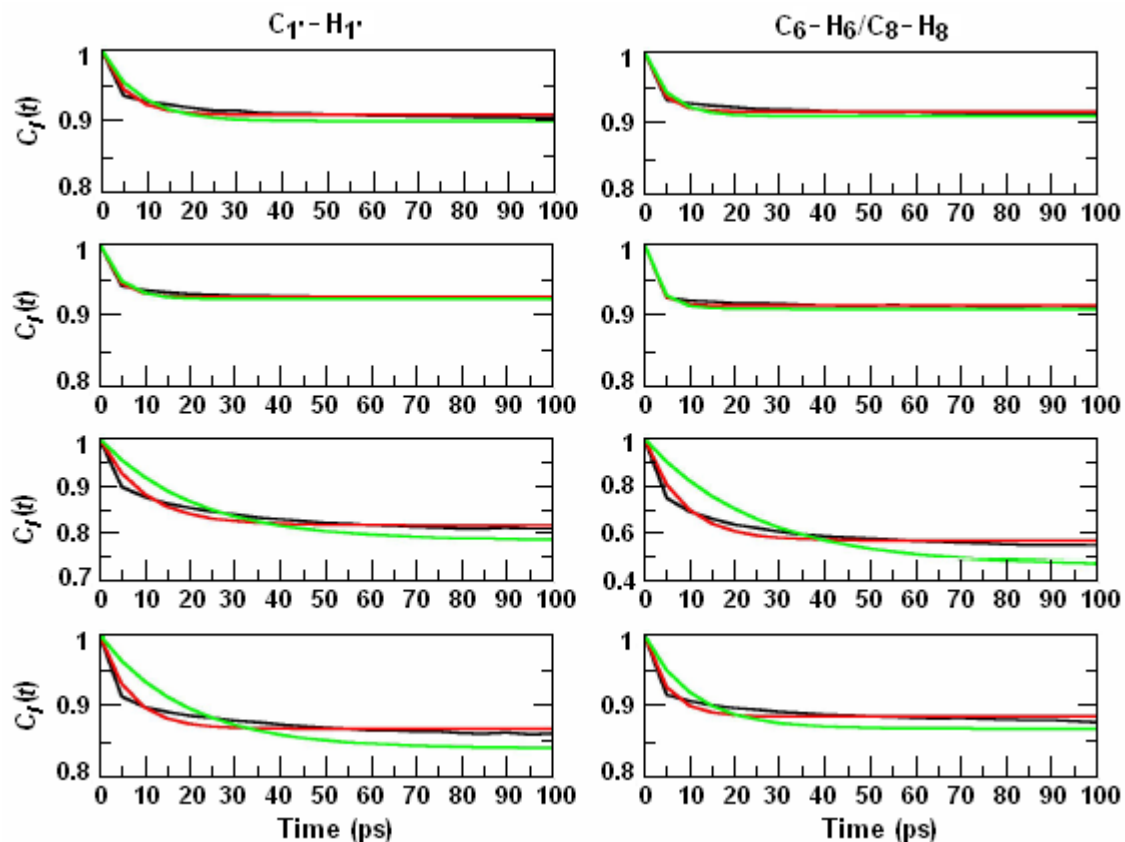


Figure 5.2 Internal correlation functions (black lines) of the dipoles C_1-H_1 (left side) and C_6-H_6 and C_8-H_8 (right side) for the residues G2, C3, U7, and G9 (from top to bottom). Exponential fits using the first 100 ps and the first 1 ns of the correlation function are shown in red and green, respectively.

viscosity of the solvent in experiment and simulation. The diffusion value for TIP3P water is around $5.56-5.70 \cdot 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ (Shirts, 2005; Mark, 2002), that is, about two times larger than the corresponding experimental value. In order to quantitatively calculate NMR observables, the MD correlation functions therefore cannot be taken directly. Instead, it is assumed that the correlation function can be factorized (equation 5.6) and the correlation time τ_c obtained from the NMR analysis (Duchardt, 2005) is used.

Using the experimental correlation time τ_c and the calculated internal correlation functions, equations 5.1 – 5.7 have been employed to compute the relaxation parameters R_1 , R_2 , and NOE for all investigated dipoles. As shown by black circles in Figure 5.3, the calculated values are in good agreement with the experimental results (Duchardt, 2005). The relative error $\sum_i |x_i^{\text{MD}} - x_i^{\text{exp}}| / \sum_i x_i^{\text{exp}}$ is 0.04 for R_1 , 0.14 for R_2 , and 0.03 for

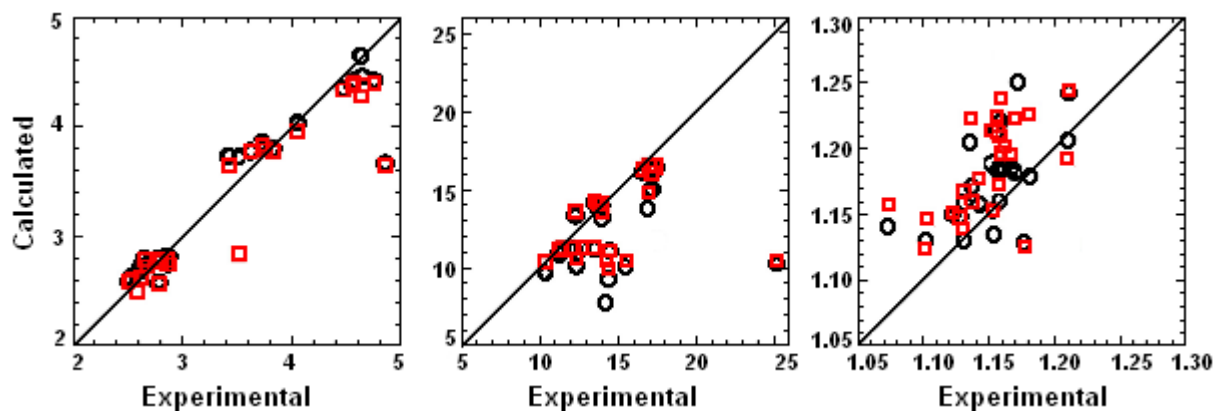


Figure 5.3 Experimental (Duchardt, 2005) vs. calculated values of spin-lattice (left) and spin-spin (middle) relaxation rates (s^{-1}) and NOE (right) for all C-H dipoles. The black circles and red squares correspond to calculations using the spectral densities obtained directly from the correlation function and from a 100 ps Lipari-Szabo fit, respectively.

NOE, respectively. The relatively large discrepancy obtained for the spin-spin relaxation rate R_2 may be related to the fact that in numerous cases the experimental R_2 values have been corrected for conformational exchange contributions during the experimental analysis (Duchardt, 2005). To assess the validity of the Lipari-Szabo approach, the relaxation constants have also been obtained by fitting the first 100 or 1000 ps of $C_I(t)$ to the Lipari-Szabo spectral density (red squares in Figure 5.3). Here, the relative errors for the 100 ps and 1 ns fit are 0.05 and 0.06 for R_1 , 0.12 and 0.13 for R_2 , and 0.02 and 0.03 for NOE, respectively. Although the relative errors of both fits are virtually identical, the 100 ps fit is more appropriate to compare to the experimental data in Ref. (Duchardt, 2005), since it reproduces better the experimental effective correlation times (see below).

The ability of the directly calculated MD data to reproduce the NMR results indicates that the force field and the simulation time scale used in this study are appropriate to describe the relaxation of the C-H bonds monitored in the NMR experiment. Thus, the MD simulation may be employed to reveal the dynamic information included in the experimental NMR data. Furthermore, the MD data may be used to compare and validate various methods to calculate the order parameters of the RNA hairpin.

5.2.2 Order parameters of the cUUCGg hairpin

Experimental and calculated order parameter S^2 for the C₁-H₁, C₆-H₆ and C₈-H₈ dipoles of all residues of cUUCGg hairpin are compared in Figure 5.4 and in Tables 5.1 and 5.2. In the experimental study (Duchardt, 2005), two models have been used to analyze the relaxation data, assuming either isotropic or axially symmetric diffusion. Both models are seen to give quite similar results for the order parameters and show that the loop residues, in particular U₇, exhibit enhanced conformational fluctuations. The experimental effective correlation times τ_e listed in Tables 5.1 and 5.2 are mostly below 10 ps, with the exception of the residues C₅, U₇, and C₁₄.

As detailed in the Section 5.1.4, three different approaches have been used to calculate S^2 from the MD trajectory. In the first approach, the order parameters S^2 and the internal correlation times τ_e were obtained by fitting the first 100 ps (or the first 1 ns) of the MD internal correlation to equation 5.8. Tables 5.1 and 5.2 show that the 100 ps fit is more appropriate to be compared to the experimental data, since it better reproduces the experimental effective correlation times. As shown in Figure 5.4, the 100 ps Lipari-Szabo fit yields an excellent agreement between experimental and calculated order parameters.

In the second approach, the equilibrium average in equation 5.10 has been used to calculate the order parameters. This corresponds to a Lipari-Szabo fit using the full time range of the internal correlation function. Figure 5.4 demonstrates that the resulting order parameters only agree for the relatively rigid stem residues but not for the flexible loop residues. The reason of this discrepancy is that, by using the entire internal correlation function, the order parameters contain information also on internal motions happening on a nanosecond time scale. For example, a loop conformational rearrangement involving residue U6 and U7 and the anti-syn transitions of the U7 base has been observed to occur with the time scale longer than 5 ns. The experimental analysis is limited by the decay of the overall-motion correlation function due to molecular tumbling ($t \approx 2$ ns). As a consequence, possible existing internal motions on a nanosecond time scale can not be extracted from the experimental data. Although the equilibrium-

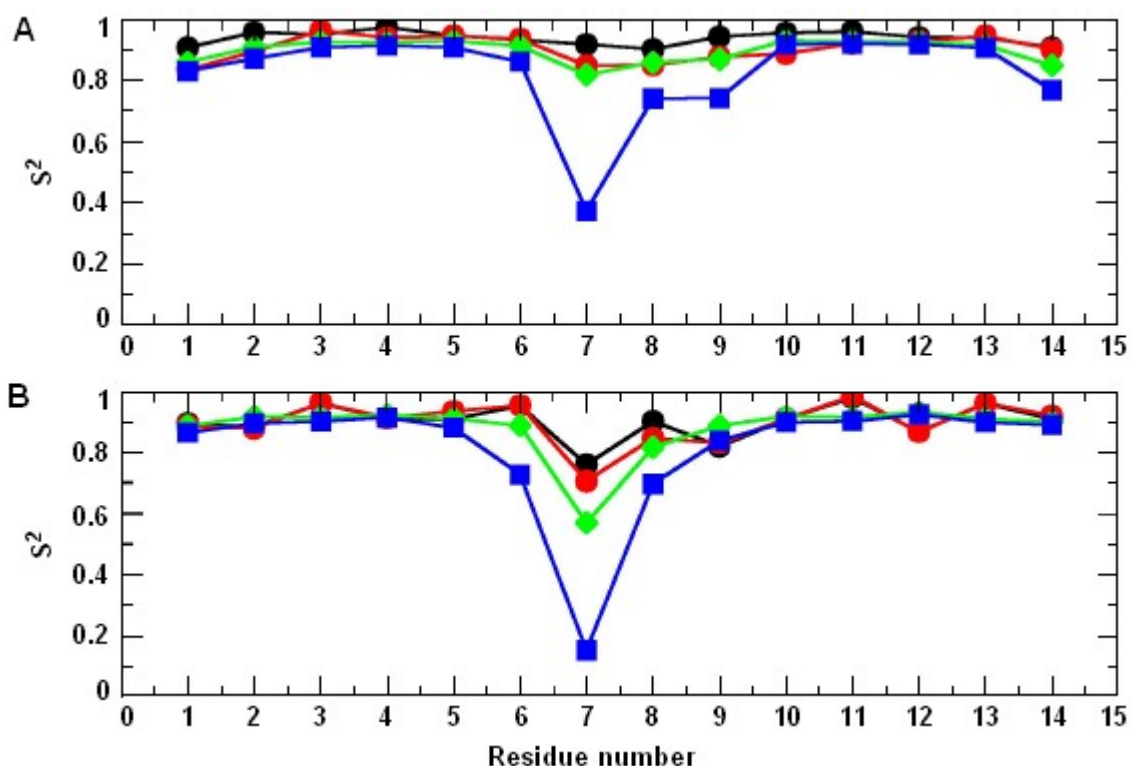


Figure 5.4 Order parameters S^2 for $C_{1'}$ (A) and C_6/C_8 (B) as a function of the residue number of cUUCGg. The experimental values (Duchardt, 2005) are shown in black (using the isotropic diffusion model) and in red (using the axially symmetric diffusion model). Calculated values are shown in green (100 ps fit) and in blue (using equation 5.10).

Table 5.1 Internal correlation times τ_e and order parameters S^2 of $C_{1'}$ for the fourteen residues of the cUUCGg hairpin. Reported values are derived from the MD simulations (by 100 ps fitting, 1 ns fitting, and by using equation 5.10) and from NMR experiment (Duchardt, 2005) (isotropic model).

	MD (100 ps fit)		MD (1 ns fit)		MD (eq 5.10)	Experiment	
	τ_e (ps)	S^2	τ_e (ps)	S^2	S^2	τ_e (ps)	S^2
G1	6.3	0.859	9.1	0.847	0.830	< 10	0.886
G2	5.2	0.909	8.1	0.900	0.871	< 10	0.878
C3	3.6	0.926	4.4	0.922	0.908	< 10	0.961
A4	2.7	0.925	3.2	0.923	0.914	< 10	0.912
C5	3.2	0.928	4.4	0.924	0.908	121.59	0.933
U6	5.4	0.913	14.6	0.897	0.861	< 10	0.953
U7	9.7	0.818	20.6	0.786	0.374	13.38	0.706
C8	9.9	0.858	23.8	0.829	0.741	< 10	0.845
G9	6.6	0.869	19.1	0.841	0.742	< 10	0.830
G10	3.2	0.931	4.3	0.926	0.917	< 10	0.909
U11	2.5	0.927	2.9	0.925	0.919	< 10	0.984
G12	3.1	0.926	3.6	0.924	0.918	< 10	0.866

C13	3.8	0.914	4.7	0.910	0.906	< 10	0.960
C14	8.3	0.849	52.9	0.782	0.767	38.08	0.919

Table 5.2 Internal correlation times τ_e and order parameters S^2 of C_6/C_8 for the fourteen residues of the cUUCGg hairpin. Reported values are derived from the MD simulations (by 100 ps fitting, 1 ns fitting, and by using equation 5.10), from NMR experiment (Duchardt, 2005) (isotropic model), and from GAF model.

	MD (100 ps fit)		MD (1 ns fit)		MD (eq 5.10)	experimental		GAF
	τ_e (ps)	S^2	τ_e (ps)	S^2	S^2	τ_e (ps)	S^2	S^2
G1	4.4	0.886	6.4	0.877	0.864	< 10	0.835	0.998
G2	3.5	0.916	4.9	0.910	0.894	< 10	0.896	0.995
C3	2.5	0.914	3.2	0.909	0.900	< 10	0.963	0.993
A4	2.3	0.921	2.8	0.918	0.913	< 10	0.939	0.987
C5	2.6	0.909	3.5	0.904	0.880	< 10	0.946	0.977
U6	4.1	0.887	19.1	0.858	0.727	< 10	0.936	0.928
U7	8.3	0.571	24.3	0.469	0.153	15.74	0.848	
C8	7.3	0.816	18.7	0.780	0.696	< 10	0.850	0.938
G9	4.9	0.886	10.5	0.869	0.838	< 10	0.877	0.770
G10	2.4	0.917	3.3	0.911	0.897	< 10	0.888	0.991
U11	2.3	0.914	2.7	0.911	0.901	< 10	0.920	0.986
G12	1.9	0.931	2.2	0.929	0.923	< 10	0.928	0.998
C13	2.9	0.909	3.7	0.904	0.898	< 10	0.946	0.993
C14	3.6	0.899	4.6	0.894	0.889	412.14	0.902	0.959

Table 5.3 Calculated and experimental (Duchardt, 2004; Allain, 1995) dihedral angles χ ($O_4-C_1-N_1-C_2$ for pyrimidine and $O_4-C_1-N_9-C_4$ for purine) including their standard deviation, obtained for the fourteen residues of the cUUCGg hairpin. The base of residue U7 adopts both anti (77 %) and syn (23 %) conformations during the simulation.

	χ_{MD} (degree)	χ_{exp}^* (degree)	χ_{exp}^{**} (degree)
G1	-171. \pm 9.		
G2	-163. \pm 8.		-168. \pm 2
C3	-163. \pm 9.		
A4	-156. \pm 9.	-149. \pm 3.	-162. \pm 1.
C5	-153. \pm 11.	-141. \pm 3.	-169. \pm 8.
U6	-147. \pm 17.	-142. \pm 3.	-164. \pm 4
U7	-131. \pm 19.; 48. \pm 12.	-147. \pm 6.	-143. \pm 9.
C8	-137. \pm 12.	-126. \pm 1.	-150. \pm 6.
G9	61. \pm 20	61. \pm 14	44. \pm 4
G10	-161. \pm 9.	-166. \pm 2.	-145. \pm 13.
U11	-153. \pm 8.	-153. \pm 2.	
G12	-171. \pm 8.		-168. \pm 4.
C13	-159. \pm 7.		-171. \pm 3.
C14	-148. \pm 12.		-162. \pm 5.

* (Allain, 1995)

** (Duchardt, 2004)

average calculation of S^2 correctly reflects the fluctuations of the system, it may therefore not be suited for the comparison to NMR experimental results.

Finally, the GAF model (equation 5.11) has been applied, which assumes that the nucleobase flexibility monitored by the order parameters of C_6/C_8 is exclusively caused by base motions along the glycosidic torsional angle χ . To this end, the distribution function of χ , which exhibits a single peak for all residues except U_7 , has been calculated. In the latter case, the base adopts both anti and syn-conformations during the simulation, and the GAF model is not applicable. Table 5.3 lists the mean and the variance of χ for all residues as calculated from the MD trajectory. All results are found to be in good agreement with the experimental data (Duchardt, 2005; Allain, 1995). The order parameters obtained from the GAF model are reported in Table 5.2. Except for the residue G_9 , the GAF values for S^2 are significantly larger (from 0.928 to 0.998) than the ones obtained from experiment and the Lipari-Szabo fit. The failure of the GAF model to correctly reproduce the order parameters of the UUCG hairpin clearly demonstrates that the motion of the base C–H dipole is not only caused by fluctuations of the base but is also due to the flexibility of the sugar ring and the backbone. For the UUCG hairpin under consideration, the GAF model on average accounts for about 20 % of the fluctuations contributing to the order parameter.

5.2.3 Internal dynamics of the cUUCGg hairpin

In practice, the above studied GAF model is not used to calculate S^2 but to rationalize the internal motions described by the experimental order parameter. Having validated the theoretical model in Section 5.2.1, the 50 ns all-atom trajectory is available for this purpose. In what follows, the main motions of the RNA hairpin is characterized first. Then it is analyzed to what extent these motions are reflected in the calculated order parameter.

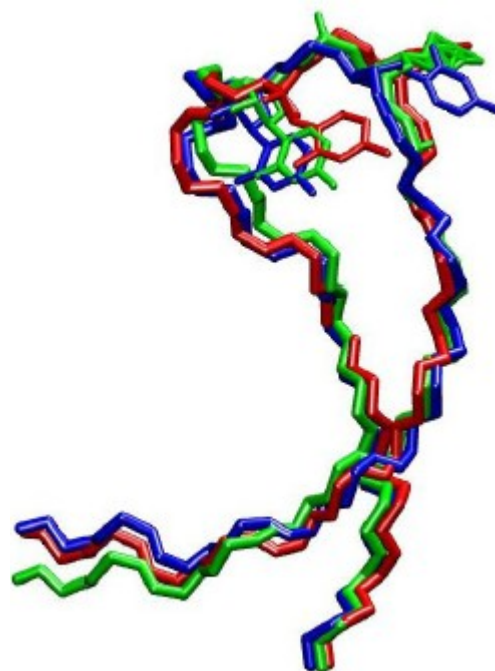


Figure 5.5 The backbone atoms of the 14-mer cUUCGg hairpin and residues U7 and C8 describing the main hairpin motion along the first three principal components of the trajectory. Shown are snapshots at 10 ns (blue), 30 ns (green), and 50 ns (red). The pictures were performed using the graphical package VMD (Humphrey, 1990).

Table 5.4 Calculated and experimental (Allain, 1995) mean values of the backbone dihedral angles (in degrees) for residues A4 – U11 of the UUCG hairpin, including their standard deviations. For the loop residues U6 – G9, the calculated distribution functions showed two peaks. This confirms the presence of at least two conformational states as shown in Figure 5.5. The NMR study also reported two values for some angles of C8 and G9, which however, do not directly correspond to the two conformational states observed in the MD simulations.

		α	β	γ	δ	ϵ	ζ
A4	MD	-72.±9.	178.±10.	60.±8.	78.±6.	-152.±11.	-70.±9.
	NMR	-79.±7.	173.±5.	63.±6.	91.±2.	-157.±3	-73.±3.
C5	MD	-72.±9	171.±10.	59.±8.	78.±6.	-162.±15.	-67.±13.
	NMR	-72.±4	163.±3.	62.±5.	92.±1.	-147.±9.	-68.±10.
U6	MD	-71.±9	173.±9	61.±7.	72.±6.	-168.±10.	-88.±18.
	NMR	158.±11.	-158.±9.	-172.±10	153.±6	-94.±10.	-125.±10.
U7	MD	-63.±10	177.±17.	-171.±12	147.±9.	-83.±10.	-70.±15.
	NMR	-161.±10.		56.±9		-164.±10.	61.±16
C8	MD	-62.±15.	177.±10.	57.±8.	144.±9.	-91.±11.	72.±11.
	NMR	-93.±15.					
G9	MD	-57.±4.	174.±4.	45.±2.	136.±3.	-114.±9.	106.±6
	NMR	30.±8.	-143.±6.	-170.±5.			
G9	MD	67.±12.	-172.±11.	-178.±7.	86.±8.	-84.±10.	-59.±9.
	NMR	108.±8.	135.±6.	-28.±8.	91.±1.	-175.±10.	56.±11.
G10	MD	110.±8.	-130.±12.			-147.±10.	-43.±44.
	NMR						
G10	MD	-69.±9.	79.±11.	176.±15.	82.±7.	-155.±8.	-62.±8.
	NMR	-132.±10.					
U11	MD	-125.±15.	90.±15.	-171.±28.	96.±2.	-151.±2.	-68.±3
	NMR						
U11	MD	-74.±9.	175.±8.	60.±8.	78.±6.	-155.±9.	-70.±13.
	NMR	-66.±3.	173.±4	53.±4.	89.±1.	-156.±7.	-91.±5.

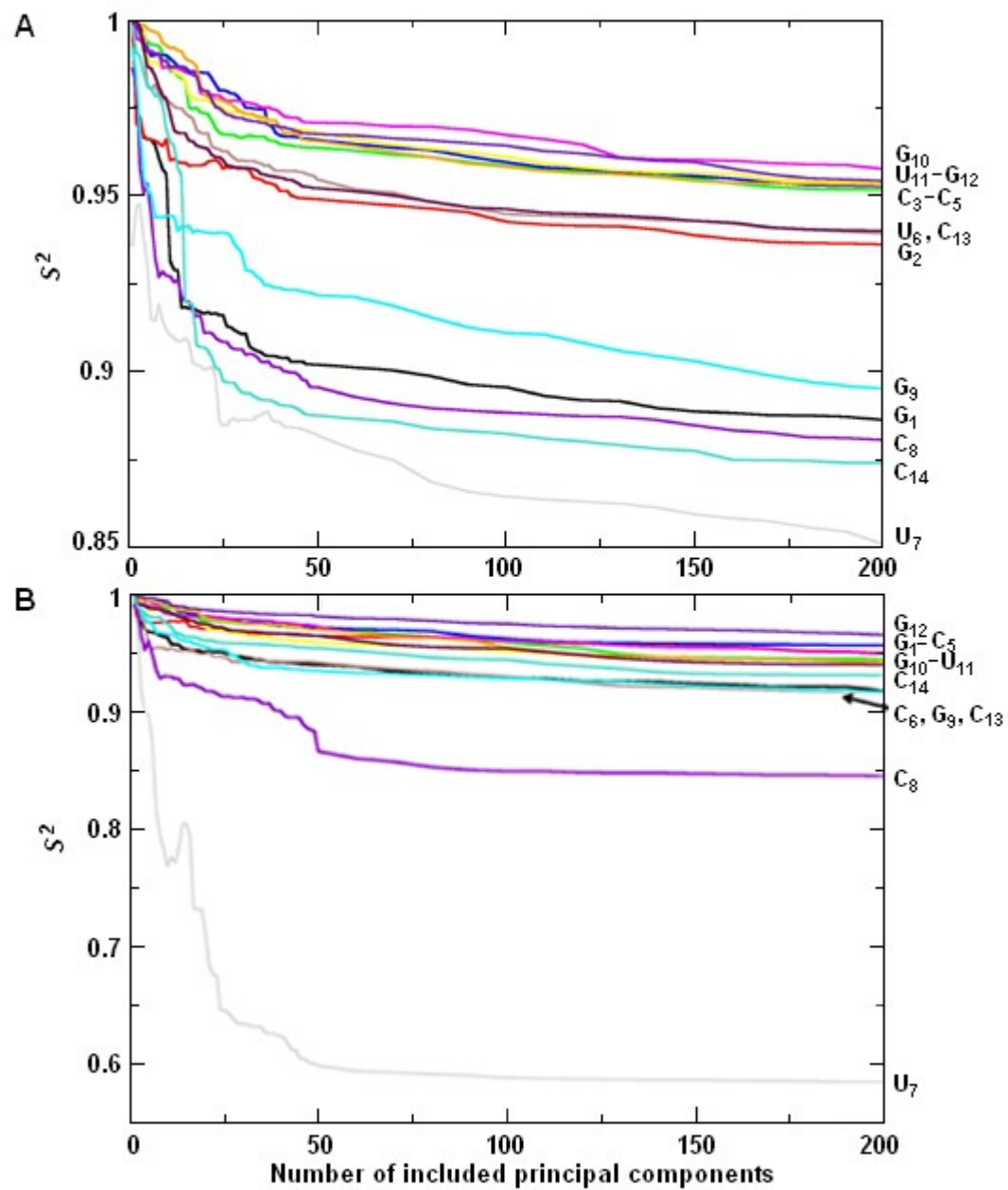


Figure 5.6 Order parameter S^2 (100 ps Lipari-Szabo fit) for C1' (A) and C6/C8 (B) obtained for the fourteen residues of the cUUCGg hairpin, plotted as a function of the number of included principal components

Table 5.5 Results of $1 - S^2$ obtained for the complete trajectory (all) of cUUCGg hairpin and by including the first three (PCA3) and first ten (PCA10) principal components, respectively. The order parameter are calculated either from a 100 ps Lipari-Szabo fit or from an equilibrium average (equation 5.10).

	$1 - S^2$ (100 ps fit)			$1 - S^2$ (eq 5.10)		
	all	PCA3	PCA10	all	PCA3	PCA10
Sugar						
U6	0.087	0.006	0.024	0.139	0.022	0.065
U7	0.182	0.053	0.089	0.626	0.511	0.578
C8	0.142	0.043	0.073	0.259	0.144	0.200
G9	0.131	0.033	0.056	0.258	0.140	0.186
Base						
U6	0.113	0.018	0.047	0.273	0.063	0.191
U7	0.429	0.082	0.232	0.847	0.195	0.680
C8	0.184	0.035	0.070	0.304	0.121	0.192
G9	0.114	0.013	0.035	0.162	0.033	0.074

Principal component analysis (PCA) represents a standard method to identify the “principal” motions of a molecular system (Amadei, 1993; García, 1992; Ichiye, 1991). The approach represents the motion in terms of an orthogonal basis, the principal components (PCs), which are ordered according to their content of root mean square fluctuations. For the 50 ns simulation of the UUCG loop, the first three PCs already contain 60 % of the overall fluctuations of the hairpin, and to cover 80, 90, and 95 % of the fluctuations, only 10, 25, and 50 out of 1442 PCs are required, respectively. In this sense, the first few PCs represent the main motions of the system. As an illustration, Figure 5.5 shows the motion along the first three PCs, which mainly consists of a conformational rearrangement involving the loop region. The presence of (at least) two conformational states is also confirmed by the analysis of the backbone dihedral angles of the loop residues U6 – G9 (see Table 5.4). To assess the influence of this conformational rearrangement on the NMR order parameters, S^2 has been recalculated for the first half of the trajectory. A 100 ps Lipari-Szabo fit yields virtually identical results for S^2 , which confirms that the 50 ns simulation time is enough to investigate the C–H relaxation.

To study to what extent the principal motions of the system account for the order parameters, the MD trajectory has been expanded in its first n PCs, and calculated S^2 from this approximated trajectory, using a 100 ps Lipari-Szabo fit. Figure 5.6 shows the resulting order parameters as a function of the number of included PCs. As may be expected, the first PCs in general make the largest contribution to the decay of S^2 . Compared to the rapid convergence of the overall fluctuations, however, the order parameters converge relatively slowly to their value obtained for the complete trajectory. Focusing on the loop residues U6 – G9, Table 5.5 compares the results of $1 - S^2$ obtained for the complete trajectory (all), for the first three PCs (PCA3), and for the first ten PCs (PCA10). On average, the first three and 10 PCs yield about 20 and 40 % of the value of $1 - S^2$ for the complete trajectory, respectively. Recalling that the first three and ten PCs contain 60 and 80 % of the overall fluctuations, respectively, the order parameter apparently accounts only partially for the principal motions of the system. This is because the motion along the first few PCs may (i) be only weakly correlated with the orientation of the C–H dipoles and (ii) contains slow motion which is not seen by the 100 ps Lipari-Szabo fit. The latter issue can be studied by recalculating $1 - S^2$ via an equilibrium average that covers all time scales of the trajectory. As shown in Table 5.5, in this case the first three and ten PCs on average yield about 40 and 70 % of the value of $1 - S^2$ for the complete trajectory, respectively, that is, quite similar to the values obtained for the overall fluctuations. In particular, the fact is found that the order parameters for the sugars U7, C8, and G9 to a large extent (80 and 55 %) are caused by the motion along the first three PCs shown in Figure 5.5. Hence, if all time scales of the trajectory are taken into account, the order parameters are well described by the principal motions of the system.

5.2.4 Comparison of the fast dynamics of the cUUCGg and uCACGg hairpins

Order parameters S^2 for C_1-H_1 , C_6-H_6 and C_8-H_8 dipoles of all residues of the uCACGg hairpin have calculated from MD trajectory. As detailed in Section 5.1.4, three different approaches have been used to calculate S^2 . Figure 5.7 shows the calculated order parameter S^2 for the C_1-H_1 , C_6-H_6 and C_8-H_8 dipoles of all residues of the uCACGg hairpin and Table 5.6 and 5.7 summarize the calculated order parameters S^2 and effective correlation times τ_e . In the Lipari-Szabo fit, the order parameters S^2 and the internal correlation times τ_e for this hairpin were obtained by fitting the first 100 ps

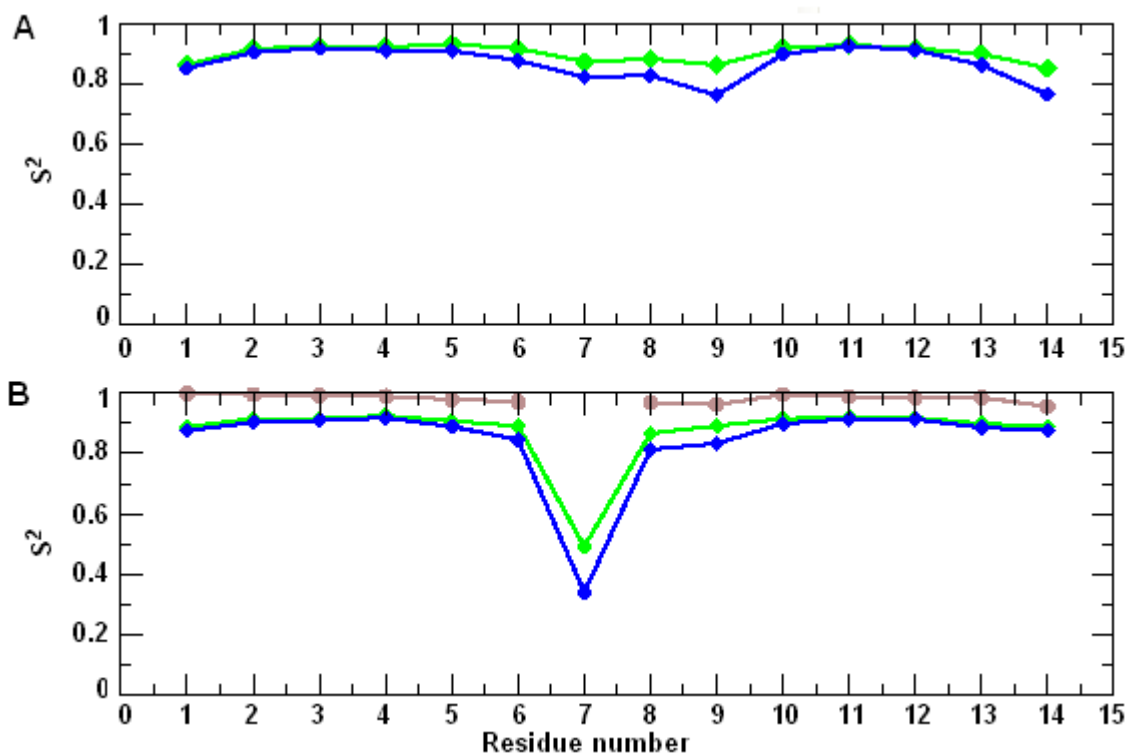


Figure 5.7 Order parameters S^2 for $C_{1'}$ (A) and C_6/C_8 (B) as a function of the residue number of uCACGg hairpin. Calculated values are shown in green (100 ps fit), in blue (using equation 5.10), and in brown (using GAF model).

Table 5.6 Internal correlation times τ_e and order parameters S^2 of $C_{1'}$ for the fourteen residues of the uCACGg hairpin. Reported values are derived from the MD simulation (by 100 ps fitting, 1 ns fitting, and by using equation 5.10)

	MD (100 ps fit)		MD (1 ns fit)		MD (eq 5.10)
	τ_e (ps)	S^2	τ_e (ps)	S^2	S^2
G ₁	5.6	0.866	7.1	0.859	0.854
G ₂	4.0	0.918	5.9	0.912	0.906
U ₃	2.6	0.928	3.1	0.925	0.920
A ₄	2.5	0.925	2.8	0.923	0.914
U ₅	2.5	0.934	3.2	0.930	0.911
C ₆	4.0	0.920	7.1	0.911	0.880
A ₇	4.5	0.876	8.7	0.860	0.821
C ₈	7.4	0.885	19.0	0.862	0.831
G ₉	10.7	0.865	32.3	0.828	0.770
G ₁₀	3.4	0.920	5.7	0.911	0.902
U ₁₁	2.3	0.932	2.7	0.930	0.927
A ₁₂	3.1	0.920	3.6	0.917	0.914
C ₁₃	4.8	0.902	6.9	0.894	0.855
C ₁₄	6.1	0.854	49.8	0.792	0.766

Table 5.7 Internal correlation times τ_e and order parameters S^2 of C_6/C_8 for the fourteen residues of the uCACGg hairpin. Reported values are derived from the MD simulation (by 100 ps fitting, 1 ns fitting, and by using equation 5.10), and from GAF model.

	MD (100 ps fit)		MD (1 ns fit)		MD (eq 5.10)	GAF
	τ_e (ps)	S^2	τ_e (ps)	S^2	S^2	S^2
G1	4.1	0.888	5.5	0.881	0.876	0.998
G2	3.2	0.916	4.7	0.909	0.902	0.996
U3	2.3	0.916	2.7	0.912	0.910	0.993
A4	1.9	0.926	2.1	0.924	0.919	0.991
U5	1.9	0.908	2.4	0.904	0.892	0.980
C6	3.6	0.892	6.0	0.881	0.847	0.973
A7	19.6	0.491	44.7	0.356	0.341	
C8	3.1	0.867	5.3	0.853	0.815	0.970
G9	4.1	0.892	9.9	0.875	0.834	0.962
G10	2.3	0.915	2.8	0.912	0.902	0.996
U11	1.8	0.920	2.0	0.919	0.915	0.990
A12	2.3	0.919	2.7	0.916	0.913	0.986
C13	2.8	0.900	3.6	0.894	0.883	0.986
C14	3.8	0.889	4.7	0.883	0.875	0.965

and the first 1 ns of the MD internal correlation to equation 5.8. The results of these two approaches show no significant differences.

The calculated S^2 values using Lipari-Szabo approach obtained by fitting the first 100 ps of $C_l(t)$ and using equilibrium-average approach (equation 5.10) are similar. The average difference in S^2 values between the two approaches are 0.016 and 0.140 for stem and loop residues, respectively. The largest difference is observed for the S^2 of dipole C_6-H_6 of residue A7, whose values differ around 31 %. In contrast, the values for the residue U7 of the cUUCGg hairpin differ about 73 %. As previously described (Section 5.2.2 and 5.2.3), the cUUCGg hairpin undergoes conformational rearrangement on the time scale longer than 5 ns, while no conformational arrangement on such a time scale is observed for the uCACGg hairpin. This can explain the discrepancy between the differences. Differences in time scale are also observed for the motion along glycosidic torsional angle of the looped out residues (A7 and U7). Figure 5.8 shows the internal correlation function for the base dipole and the glycosidic torsional angle of residue A7

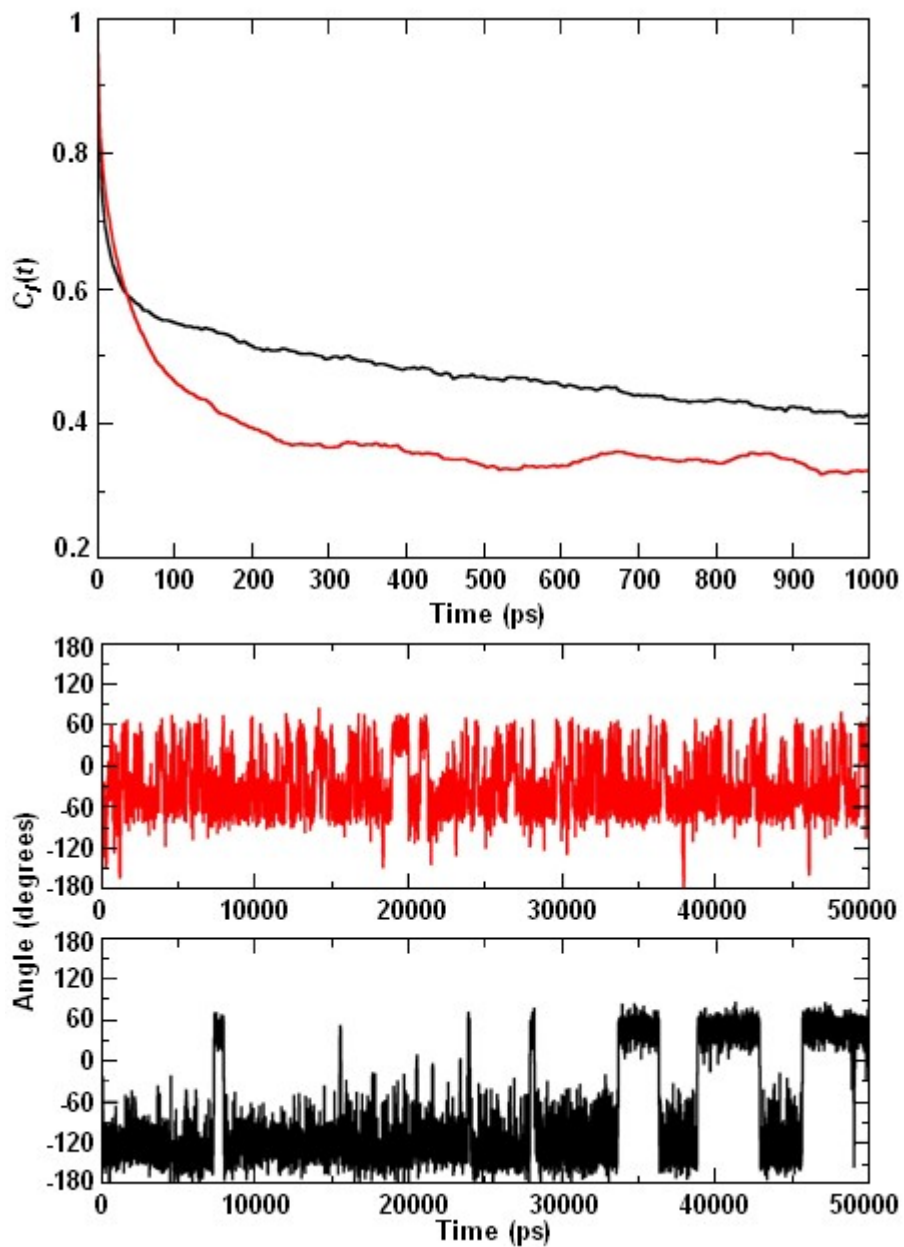


Figure 5.8 Dynamic behavior of the looped out residues U7 (in black) and A7 (in red) at 298 K. Top internal correlational function C_{6-H_6}/C_{8-H_8} , bottom torsion angle χ .

Table 5.8 Calculated and experimental (PDB: 1RFR; Duchardt, 2004) dihedral angles χ ($O_4-C_1-N_1-C_2$ for pyrimidine and $O_4-C_1-N_9-C_4$ for purine) including their standard deviation, obtained for the fourteen residues of the uCACGg hairpin. The base A7 adopts two type syn (72 % and 27 %) conformations during the simulation.

	χ_{MD} (degree)	χ_{exp}^* (degree)	χ_{exp}^{**} (degree)	χ_{exp}^{***} (degree)
G1	-171. \pm 9.	-159. \pm 1.	-162. \pm 2.	
G2	-165. \pm 8.	-164. \pm 1.	-160. \pm 2.	
U3	-158. \pm 8.	-164. \pm 1.	-160. \pm 1.	
A4	-157. \pm 8.	-161. \pm 1	-154. \pm 1.	
U5	-150. \pm 10.	-157. \pm 1.	-152. \pm 8.	
C6	-149. \pm 11.	-154. \pm 1.	-144. \pm 5.	150. \pm 1.
A7	-50. \pm 16.; 31. \pm 17.	-91. \pm 17.	-107. \pm 5.	
C8	-149. \pm 17	-162. \pm 4.	-160. \pm 13	
G9	32. \pm 11.	50. \pm 2.	30. \pm 6.	90. \pm 1.
G10	-166. \pm 8.	-160. \pm 2.	169. \pm 4.	
U11	-155. \pm 8.	-158. \pm 2.	-165. \pm 1.	
A12	-154. \pm 9.	-155. \pm 2.	-160. \pm 1	
C13	-157. \pm 10.	-161. \pm 1.	-153. \pm 1.	-162. \pm 1.
C14	-147. \pm 13.	-163. \pm 1.	-149. \pm 1.	-164. \pm 1.

* (PDB : 1RFR)

** (PDB: 1ROQ)

*** (Duchardt, 2004)

and U7. Both the hairpins have a fast initial decay in internal correlation function, but only in the case of the U7 motion of longer time scale can be observed. The anti-syn equilibrium of the residue U7 takes place at the time scale around 5 ns, while the syn-syn arrangement of the A7 is in the order of picoseconds.

Finally, the GAF model (equation 5.11) has been applied for the uCACGg hairpin as for the previous cUUCGg hairpin, which assumes that the nucleobase flexibility monitored by the order parameters of C_6/C_8 is exclusively caused by base motions along the glycosidic torsional angle χ . These order parameters are listed in Table 5.7 and displayed in Figure 5.7, and the mean and standard deviation of torsional χ for all residues of the uCACGg hairpin is reported in Table 5.8. The simulated angles are in good agreement with the experimental data of Ohlenschläger and co workers [PDB: 1RFR (Ohlenschläger, 2004)] and/or Duchardt and Schwalbe (Duchardt, 2004), except for the residue G9 whose value is closer to the one in the NMR structure of Du and co

workers [PDB: 1ROQ (Du, 2003)]. Again, the GAF model is not applicable for the residue A7 due to the two peaks in the distribution function of the χ angle. All the GAF values for S^2 are significantly larger than the ones obtained from the other approaches. This agrees with what is observed for the cUUCGg hairpin, with exception of the residue 9.

5.3 Conclusion

The 50 ns MD simulations of 14-mer cUUCGg and uCACGg hairpins have been performed in order to study their fast dynamics. In the first part the simulations of the cUUCGg hairpin have been used to validate the procedure to compare MD simulations and NMR experiment in the description of fast dynamics of the RNA system. The main results of this part can be summarized as follows:

Validity of the theoretical model: The ability of the MD data to reproduce the NMR relaxation parameters indicates that the force field and the simulation time scale used in this study are appropriate to describe the relaxation of the C-H bonds monitored in the NMR experiment. Although slow conformation rearrangements of the RNA loop which are not sampled appropriately is found, these motions do not change the calculated NMR results.

Calculation of the NMR order parameters: (i) Except for the most flexible residue the looped out U7, overall and internal motions of the molecule are virtually independent and the factorization approximation of the correlation function holds. Even in the case of U7, the approximation is not expected to change the calculated NMR data, since only the dynamics at short times ($t \approx 1$ ns) is relevant in the calculation. (ii) The calculation of order parameters via an equilibrium average was shown to deteriorate for residues undergoing slow internal dynamics, since the latter can not be seen in NMR relaxation experiment limited by molecular tumbling. Although the equilibrium-average calculation of S^2 correctly reflects the fluctuations of the system, it may therefore not be suited for the comparison to NMR experimental results. (iii) The GAF model only yielded about 20 % of the correct value for $1 - S^2$. This indicates that the motion of the base C-H dipoles is not only caused by fluctuations of the base but is also due to the flexibility of

the sugar ring and the backbone. (iv) Lipari-Szabo fits of the first 100 ps of the internal correlation functions lead to excellent agreement between calculated and experimental order parameters S^2 and internal correlation times τ_e . Being consistent with the analysis of the NMR experiments, this approach was found to represent the best way to calculate order parameters from a MD simulation.

Order parameters vs. internal motion: A principal component analysis of the 50 ns trajectory has shown that a conformational rearrangement involving the loop region represents the main motion of the system. This principal motion, however, accounts only partially for the measured NMR order parameters S^2 , because the latter are not sensitive to internal dynamics on a nanosecond time scale. Calculating the order parameter via an equilibrium average that covers all time scales of the trajectory, a direct correspondence between calculated S^2 and principal motions are obtained.

In the second part of the Chapter the internal dynamics of the uCACGg hairpin has been investigated and the result has been compared with the cUUCGg hairpin. From the comparison of the base and sugar order parameters it results that the stem residues in both the RNA systems show similar rigidity. In fact S^2 of the stem residues have always values bigger than 0.84 in both the hairpins large values are observed. The order parameters calculated for the base loop residues of both hairpins suggest that residues in the two loops are involved in motion with different time scales. The cUUCGg hairpin exhibits a conformational rearrangement on nanosecond time scale. On the other hand, the uCACGg presents no conformational rearrangement during 50 ns simulation time.

Chapter 6

Conclusion

In this thesis molecular dynamics simulations were used to investigate structural, dynamical and thermal properties of RNA hairpins at atomistic level. In particular, the 14-mer RNA hairpins, uCACGg and cUUCGg, have been studied. The available NMR structures of the uCACGg tetraloop are strikingly similar in overall geometry and in hydrogen bonding to the experimental structure of the canonical cUUCGg tetraloop, in spite of the difference sequence enclosing base pairs of two hairpins.

Despite of their considerable structural similarity, however, the uCACGg and cUUCGg tetraloops were found to differ in their functionality and thermostability.

At the beginning, efforts have been oriented to find the best molecular model and the best simulation condition to simulate in appropriate way small RNA hairpins in water solution. First, three versions of the biomolecular AMBER force field have been tested by performing 60 ns simulations on the 14-mer uCACGg hairpin. The simulated structural properties and atomic fluctuations show high similarities among the three

force fields. Moreover, the internuclear distances obtained from the simulations are found to be in good agreement with those experimental distances obtained by NMR experiment. The agreement between simulations and NMR structural data confirms the ability of the AMBER force field to describe the structural characteristic of small RNA hairpins. Then, the effect of methods to describe long-range electrostatic interactions on structural properties has been investigated. In particular, the results using reaction field methods have been compared to PME's ones. The results show that the most appropriate way to treat long-range electrostatic is using PME method, even if simulations at low sodium concentration show similar results in term structural stability and atomic fluctuation when both the approaches are used. Third, the cation's effect on the RNA stability has been investigated. Different ions type (mono- and divalent) and different concentrations have been considered. The simulations show that metal-ion affinity for RNA site differs when Na^+ and/or Mg^{2+} are used as counterions. In particular, Na^+ ions are located in the major groove and near the closing base pair U5-G10, while Mg^{2+} ions prefers residues U3, and A4-U11 and the phosphate groups.

Once found the appropriate model and simulation condition, the simulations have been used to provide other interesting information that may improve insights into system. These include the effects of the loop sequence and the closing base pair on the conformational distribution, on the internal motions and on the thermostability of two RNA tetraloop hairpins that have been investigated in this work.

First, the structural features at room temperature have been analyzed. The observed structural similarities of the simulated hairpins at room temperature and the agreement with NMR structural data confirm the hypothesis that the two tetraloop may belong to the same "extended" family. The two hairpins have similar solvent accessible surface, nevertheless the different functional groups are accessible to the solvent. In particular, a stronger attitude to donate hydrogens is observed for the uCACGg hairpin than for the cUUUCGg hairpin. These results might be related to the difference in binding affinity to viral protease showed by the two hairpins.

The folding and unfolding of uCACGg and cUUUCGg hairpins were monitored using REMD simulations with a purpose to understand better the different thermostability between these hairpins, providing atomistic details of the possible intermediates.

Although the calculate absolute melting temperatures are about 20 % too high, the REMD simulations reproduce the experimentally found difference in melting temperatures of $\Delta T \approx 20$ K. The results indicate that the difference in melting temperature is not due to differences in sequence, structure, or dynamics of the loop, but appears to be solely caused by the differences in the stem base pairs. Moreover, while the uCACGg hairpin unfolds cooperatively, the cUUCGg unfolding occurs less cooperatively.

To study the internal fast dynamics of the uCACGg and cUUCGg, 50 ns MD simulations allow achieving an accurate description of the internal fast motions of the RNA hairpins, even though the complete conformational space accessible to the system cannot be explored. The back-calculation of NMR relaxation parameters have been successfully used to validate the MD simulations. Based on these result, the data can be used to find the best approach to extract the NMR order parameters from MD simulations. The approached tested are include among others the 100 ps Lipari-Szabo fitting, equilibrium average and GAF model. Among those, the last two approaches give only qualitative agreement with experimental data. NMR order parameters can be correctly calculated from MD simulations, only including those motions with a time scale shorter than the overall tumbling. This is achieved when the Lipari-Szabo fitting is performed on first part of the internal correlation functions. The calculated order parameters show the different behavior of the two hairpins in the loop region. A conformational rearrangement observed in the UUCG loop does not occur in the CACG loop.

In summary, the MD simulations have shown to be useful tools to investigate the structural and dynamical properties of RNA systems on atomistic level. The following goals have been achieved in this work. ⁽¹⁾ The MD simulations performed using the three version of AMBER force field (AMBER94, AMBER98, and AMBER99 force fields) in this work are in good agreement with the NOE data and have similar results in terms of structural features and atomic fluctuation. ⁽²⁾ The simulations using two methods to treat electrostatic interactions show that the long-range electrostatic interactions play a major role in MD simulations of RNA systems. The reaction field method may be not accurate enough, and charge-charge interactions should be treated by PME method. ⁽³⁾ Even though ion concentration and ion type show no significant difference in structural

features and atomic fluctuation of the uCACGg hairpin, the specific interactions of Na^+ and Mg^{2+} with the uCACGg hairpin are observed. ⁽⁴⁾ Comparing to the cUUCGg hairpin, the uCACGg hairpin has different functional groups that are accessible to the solvent and different unfolding process. ⁽⁵⁾ The back-calculation of NMR relaxation parameters can validate the MD simulations for an analysis of internal motions. Differences between the uCACGg and cUUCGg hairpins are found for the dynamical behavior in term of fast internal motion and order parameter especially in loop region. Despite of these achievements, it is still questionable why the melting temperature of the hairpin cannot be calculated accurately. Maybe, the long-range electrostatic treatment using PME method could be the main reason for the thermal stabilization of the RNA hairpins. The good agreement of NMR and MD in term of structure and dynamics combined with the capability of the MD study to reveal the underlying internal motions clearly demonstrate the power of a joint NMR/MD study of conformational dynamics of RNA systems. The measured NMR order parameters do not provide a complete description of the motions, and are not sensitive to internal dynamics on a nanoseconds time scale.

Bibliography

- (Adeleman, 1976) S.A. Adelman., J.D. Doll “Generalized Langevin equation approach for atom-solid-surface scattering - general formulation for classical scattering off harmonic solids” *J Chem Phys* (1976), **64**, 2375-2388.
- (Al-Hashimi, 2005) H.M. Al-Hashimi “Dynamics-based amplification of RNA function and its Characterization by Using NMR Spectroscopy” *ChemBioChem* (2005), **40**, 7947-7956.
- (Allain, 1995) F. Allain, G. Varani “Structure of the P1 helix from group I self-splicing intron” *J Mol Biol* (1995), **250**, 333–353.
- (Allen, 1987) M. P. Allen, D. J. Tildesley “Computer Simulations of Liquids” (1987), Oxford, New York, USA: Oxford University Press.
- (Allinger, 1996) N.L. Allinger, K. Chen, J.H. Lii “An improved force field (MM4) for saturated hydrocarbons” *J Comput Chem* (1996), **17**, 642-668.
- (Allinger, 1989) N.L. Allinger, Y.H. Yuh, J.H. Lii “Molecular mechanics. The MM3 force field for hydrocarbons. 1” *J Am Chem Soc* (1989), **23**, 8551-8566.
- (Amadei, 1993) A. Amadei, A. B. M. Linssen, H. J. C. Berendsen “Essential dynamics of proteins” *Proteins: Struct Funct Genet* (1993), **17**, 412–425.
- (Andersen, 1983) H.C. Andersen “RATTLE: a ‘velocity’ version of the SHAKE algorithm for molecular dynamics calculation” *J Comput Phys* (1983), **72**, 8205-8212.
- (Andino, 1990) R. Andino, G. E. Rieckhof, D. Baltimore “A functional ribonucleoprotein complex forms around the 5’ end of the poliovirus RNA (1990), *Cell*, **63**, 369–380.
- (Ansari, 2001) A. Ansari, S. V. Kuznetsov, Y. Shen “Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins” *Proc Natl Acad Sci USA* (2001), **98**, 7771–7776.
- (Antao, 1991) V.P. Antao, S.Y. Lai, I. Tinoco, Jr. “A thermodynamic study of unusually stable RNA and DNA hairpins” *Nucleic Acids Res* (1991), **19**, 5901-5905.
- (Åqvist, 1990) J. Åqvist “Ion-water interaction potentials derived from free energy perturbation simulations” *J Phys Chem* (1990), **94**, 8021-8024.
- (Atkins, 2006) J.F. Atkins, R.F. Gesteland, R.J. Jackson, N.M. Wills “The shapely mRNA: knotting ventured, knotting gained” *The RNA World* (eds. R.F. Gesteland, T.R. Cech, J.F. Atkins) 3rd ed. New York: Cold Spring Harbor (2006) 437-467.
- (Auffinger, 2003) P. Auffinger, L. Bielecki, E. Westhof “The Mg²⁺ binding sites of the 5S rRNA Loop E modif as investigated by molecular dynamics simulations” *Chem Biol* (2003), **10**, 551–561.

- (Auffinger, 1998) P. Auffinger, E. Westhof "Simulations of the molecular dynamics of nucleic acids" *Current Opinion in Structural Biology* (1998), **8**, 227-236.
- (Auffinger, 1995) P. Auffinger, D.L. Beveridge "A simple test for evaluating the truncation effects in simulations of systems involving charged groups" *Chem Phys Lett* (1995), **234**, 413-415.
- (Batey, 1999) R.T. Batey, R.P. Rambo, J.A. Doudna "Tertiary motifs in RNA structure and folding" *Angew Chem Int Ed* (1999), **38**, 2326-2343.
- (Bayly, 1993) C.I. Bayly, P. Cieplak, W.D. Cornell, P.A. Kollman "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model" *J Phys Chem* (1993), **97**, 10269-10280.
- (Berendsen, 1995a) H.J.C. Berendsen "Molecular dynamics simulations : the limits and beyond" Proceedings of the 2nd International Symposium on Algorithms for Macromolecular Modelling, Berlin, May 21-24, 1997.
- (Berendsen, 1995b) H.J.C. Berendsen, D. van der Spoel, R. van Drunen "Gromacs: a message-passing parallel molecular dynamics implementation" *Comp Phys Comm* (1995), **91**, 43-56.
- (Berendsen, 1984) H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak "Molecular dynamics with coupling to an external bath" *J Chem Phys* (1984), **81**, 3684-3690.
- (Beveridge, 2000) D.L. Beveridge, K.J. McConnell "Nucleic acids: theory and computer simulation, Y2K" *Curr Opin Struc Biol* (2000), **10**, 182-196.
- (Bonnet, 1998) G. Bonnet, O. Krichevsky, A. Libchaber "Kinetics of conformational fluctuations in DNA hairpin-loops" *Proc Natl Acad Sci USA* (1998), **95**, 8602-8606.
- (Borer, 1973) P.N. Borer, O.C. Uhlenbeck, B. Dengler, I. Tinoco, Jr. "Circular dichroism of double-helical oligoribonucleotides" *J Mol Biol* (1973), **80**, 759-771.
- (Breslauer, 2006) K.J. Breslauer, R. Frank, H. Blocker, L.A. Marky "Predicting DNA duplex stability from the base sequence" *Proc Natl Acad Sci* (2006), **83**, 3746-3750.
- (Brooks, 1982) B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus "CHARMM: a program for macromolecular energy, minimization, and dynamics calculations" *J Comput Chem* (1982), **4**, 187-217.
- (Brüschweiler, 1994) R. Brüschweiler, D.A. Case "Characterization of biomolecular structure and dynamics by NMR cross relaxation" *Prog Nucl Magn Reson Spectrosc* (1994), **26**, 27-58.
- (Brüschweiler, 1992) R. Brüschweiler, B. Roux, M. Blackledge, C. Griesinger, M. Karplus, R. R. Ernst "Influence of rapid intramolecular motion on NMR cross-relaxation rates. A molecular dynamics study of antamanide in solution" *J Am Chem Soc* (1992), **114**, 2289-2302.

- (Case, 2005) D.A. Case, T.E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K.H. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, R.J. Woods "The Amber biomolecular simulation Programs" *J Comput Chem* (2005), **26**, 1668-1688.
- (Case, 2002) D. Case "Molecular Dynamics and NMR Spin Relaxation in Proteins" *Acc Chem Res* (2002), **35**, 325-331.
- (Case, 1999) D.A. Case, D.A. Pearlman, J.W. Caldwell, T.E. Cheatham, W.S. Ross, C.L. Simmerling, T.A. Darden, K.M. Merz, R.V. Stanton, A.L. Cheng, J.J. Vincent, M. Crowley, V. Tsui, R.J. Radmer, Y. Duan, J. Pitner, I. Massova, G.L. Seibel, U.C. Singh, P.K. Weiner, P.A. Kollman "Amber 6" (1999), University of California, San Francisco CA..
- (Chatfield, 1998) D.C. Chatfield, A. Szabo, A., B.R. Brooks "Molecular dynamics of Staphylococcal nuclease: Comparison of simulation with ^{15}N and ^{13}C NMR relaxation data" *J Am Chem Soc* (1998), **120**, 5301-5311.
- (Cheatham, 2001) T.E. Cheatham, III, M.A. Young "Molecular dynamics simulation of nucleic acids: successes, limitations, and promise" *Biopolymers* (2001), **56**, 232-356.
- (Cheatham, 2000) T.E. Cheatham, III, P.A. Kollman "Molecular dynamics simulation of nucleic acids" *Annu Rev Phys Chem* (2000), **51**, 435-471.
- (Cheatham, 1999) T.E. Cheatham, III, P. Cieplak, P.A. Kollman "A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat" *J Biomol Struct Dyn* (1999), **16**, 845-862.
- (Chen, 2000) S. Chen, K.A. Dill "RNA folding energy landscapes" *Proc Natl Acad Sci USA* (2000), **97**, 646-651.
- (Connolly, 1983) M.L. Connolly "Solvent-accessible surfaces of proteins and nucleic acids" *Science* (1983), **221**, 709-713.
- (Cornell, 1995) W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman "A second generation force field for the simulation of proteins, nucleic acid, and organic molecules" *J Am Chem Soc* (1995), **117**, 5179-5197.
- (Darden, 1993) T. Darden, D. York, L. Pedersen "Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems" *J Chem Phys* (1993), **98**, 10089-10092.
- (Daura, 1998) X. Daura, A.E. Mark, W.F. van Gunsteren "Parameterization of aliphatic CH_n united atoms of GROMOS96 force field" *J Comput Chem* (1998), **19**, 535-547.
- (Dolenc, 2005) J. Dolenc, C. Oostenbrink, J. Koller, W.F. van Gunsteren "Molecular dynamics simulations and free energy calculations of netropsin and distamycin binding to an AAAAA DNA binding site" *Nucleic Acid Res* (2005), **33**, 725-733.
- (Du, 2003) Z. Du, J. Yu, R. Andino, T.L. James "Extending the family of UNCG-like tetraloop motifs: NMR structure of a CACG tetraloop from Coxsackievirus B3" *Biochemistry* (2003), **42**, 4373-4383.

- (Duchardt, 2005) E. Duchardt, H. Schwalbe "Residue specific ribose and nucleobase dynamics of the cUUCGg RNA tetraloop motif by NMR ^{13}C relaxation" *J Biomol NMR* (2005), **32**, 295-308.
- (Duchardt, 2004) E. Duchardt, C. Richter, O. Ohlenschläger, M. Görlach, J. Wöhnert, H. Schwalbe "Determination of the glycosidic bond angle χ in RNA from cross-correlated relaxation of CH dipolar coupling and N chemical shift anisotropy" *J Am Chem Soc* (2004), **126**, 1962-1970.
- (Eisenhaber, 1995) F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, M. Scharf "The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies" *J Comput Chem* (1995), **16**, 273-284.
- (Ennifar, 2000) E. Ennifar, A. Nikulin, S. Tishchenko, A. Serganov, N. Nevskaya, M. Garber, B. Ehresmann, C. Ehresmann, S. Nikonov, P. Dumas "The crystal structure of UUCG tetraloop" *J Mol Biol* (2000), **304**, 35-42.
- (Ernst, 2004) R.R. Ernst, G. Bodenhausen, A. Wokaun "Principles of nuclear magnetic resonance in one and two dimensions" (2004), New York: Oxford University Press Inc.
- (Feenstra, 1999) K.A. Feenstra, B. Hess, H.J.C. Berendsen "Time-scale molecular dynamics simulations of hydrogen-rich systems" *J Comput Chem* (1999), **20**, 786-798.
- (Feig, 1999) M. Feig, B. M. Pettitt "Sodium and chlorine ions as part of the DNA solvation shell" *Biophys J* (1999), **77**, 1769-1781.
- (Fischer, 1998) M.W.F. Fischer, A. Majumdar, E.R.P. Zuiderweg "Protein NMR relaxation: theory, applications and outlook" *Prog Nucl Magn Reson Spectrosc* (1998), **33**, 207-272.
- (Foloppe, 2000) N. Foloppe, A.D. MacKerell, Jr. "All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data" *J Comput Chem* (2000), **21**, 86-104.
- (Frenkel, 2002) D. Frenkel, B. Smit "Understanding molecular simulations: from algorithms to applications" (2002), San Diego, California, USA: Academic Press.
- (Fujita, 1987) T. Fujita "A specific inhibitor design approach by means of molecular dynamics calculation for porcine pancreatic elastase" *J Comput Chem* (1987) **8** 645-650.
- (García, 2003) A.E. García, J.N. Onuchic "Folding a protein in a computer: An atomic description of the folding/unfolding of a protein" *A Proc Natl Acad Sci USA* (2003), **100**, 13898-13903.
- (García, 2001) A.E. García, K.Y. Sanbonmatsu "Exploring the energy landscape of a β hairpin in explicit solvent" *Proteins: Struct Funct Genet* (2001), **42**, 345-354.
- (García, 1992) A.E. García "Large-amplitude nonlinear motions in proteins" *Phys Rev Lett* (1992), **68**, 2696-2699.

- (Gargallo, 2003) R. Gargallo, P.H. Hünenberger, F.X. Avilé, B. Olivia “Molecular dynamics simulations of highly charged proteins: comparison of the particle-particle-mesh and reaction field methods for the calculation of electrostatic interactions” *Protein Science* (2003), **12**, 2161-2172.
- (Haile, 1980) J.M. Haile; C.G. Gray “Spherical harmonic expansions of the angular pair correlation function in molecular fluids” *Chem Phys Lett* (1980), **76**, 583-588.
- (Halgren, 1996a) T.A. Halgren “Merck molecular force field. I. Basis, form, scope, parameterization, and performance of Mmff94” *J Comput Chem* (1996), **17**, 490-519.
- (Halgren, 1996b) T.A. Halgren “Merck molecular force field. II. Mmff94 van der Waals and electrostatic parameters for intermolecular interactions” *J Comput Chem* (1996), **17**, 520-552.
- (Halgren, 1996c) T.A. Halgren “Merck molecular force field. III. Molecular geometries and vibrational frequencies for Mmff94” *J Comput Chem* (1996), **17**, 553-586.
- (Harvey, 1989) S.C. Harvey “Treatment of electrostatic effect in macromolecular modelling” *Proteins: Struct Funct Gen* (1989), **5**, 78-92.
- (Hayryan, 2005) S. Hayryan, C.K. Hu, J. Skřivánek, E. Hayryan, I. Pokorný “A new analytical method for computing solvent-accessible surface area of macromolecules and its gradients” *J Comput Chem* (2005), **26**, 334-343.
- (Hermann, 1998) T. Hermann, P. Auffinger, E. Westhof “Molecular dynamics investigations of hammerhead ribozyme RNA” *Eur Biophys J* (1998), **27**, 153-165.
- (Hess, 2006) B. Hess, N. F. A. van der Vegt “Hydration thermodynamic properties of amino acid analogues: a systematic comparison of biomolecular force fields and water models” *J Phys Chem B* (2006), **110**, 17616-17626.
- (Hess, 1997) B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije “LINCS: a linear constraint solver for molecular simulations” *J Comput Chem* (1997), **18**, 1463-1472.
- (Heus, 1991) H.A. Heus, A. Pardi, “Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops” *Science* (1991), **253**, 191-194.
- (Hobza, 1997) P. Hobza, M. Kabeláč, J. Šponer, P. Mejzlík, J. Vondrášek “Performance of empirical potentials (AMBER, CFF95, CVFF, CHARMM, OPLS, POLTEV), semiempirical quantum chemical methods (AM1, MNDO/M, PM3), and *ab initio* Hartree-Fock method for interaction of DNA bases: comparison with nonempirical beyond Hartree-Fock results” *J Comput Chem* (1997), **18**, 1136-1150.
- (Hoover, 1985) W.G. Hoover, “Canonical dynamics-equilibrium phase-space distributions” *Physical Review A* (1985), **31**, 1695-1697.

- (Horn, 2004) H.W. Horn, W.C. Swope, J.W. Pitera, J.D. Madura, T.J. Dick, G.L. Hura, T. Head-Gordon “Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew” *J Chem Phys* (2004), **120**, 9665–9678.
- (Hougland, 2006) J.L. Hougland, J.A. Piccirilli, M. Forconi, J. Lee, D. Herschlag “How the group I Intron work: a case study of RNA structure and function” *The RNA World* (eds. R.F. Gesteland, T.R. Cech, J.F. Atkins) 3rd ed. New York: Cold Spring Harbor (2006) 133-205.
- (Humphrey, 1996) W. Humphrey, A. Dalke, K. Schulten “VMD: Visual Molecular Dynamics” *J Mol Graph* (1996), **14**, 33-38.
- (Hünenberger, 1998) P.H. Hünenberger, W.F. van Gunsteren “Alternative schemes for the inclusion of a reaction-field correction into molecular dynamics simulations: influence on the simulated energetic, structural, and dielectric properties of liquid water” *J Chem Phys* (1998), **108**, 6117-6134.
- (Hünenberger, 1998) P.H. Hünenberger, A.E. Mark, W.F. van Gunsteren “Fluctuation and Cross-correlation Analysis of Protein Motions Observed in Nanosecond Molecular Dynamics Simulations” *J Mol Biol* (1998), **252**, 492–503.
- (Hyeon, 2005) C. Hyeon, D. Thirumalai “Mechanical unfolding of RNA hairpins” *Proc Natl Acad Sci USA* (2005), **102**, 6789–6794.
- (Ichiye, 1991) T. Ichiye, M. Karplus “Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations” *Proteins: Struct Funct Genet* (1991), **11**, 205–217.
- (Ihle, 2005) Y. Ihle, O. Ohlenschläger, S. Häfne, E. Duchardt, M. Zacharias, S. Seitz, R. Zell, R. Ramachandran, M. Görlach “A novel cGUUAg tetraloop structure with a conserved yYNMGg-type backbone conformation from cloverleaf 1 of bovine enterovirus 1 RNA” *Nucleic Acids Res* (2005), **33**, 2003–2011.
- (Jorgensen, 1988) W.L. Jorgensen, J. Tirado-Rives “The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin” *J Am Chem Soc* (1988), **110**, 1657-1666.
- (Jorgensen, 1983) W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein “Comparison of simple potential functions for simulating liquid water” *J Chem Phys* (1983), **79**, 926-935.
- (Karplus, 1990) M. Karplus, G.A. Petsko “Molecular dynamics simulations in biology” *Nature* (1990), **347**, 631-639.
- (Koplin, 2005) J. Koplin, Y. Mu, C. Richter, H. Schwalbe, G. Stock “Structure and dynamics of an RNA tetraloop: a joint molecular dynamics and NMR study” *Structure* (2005), **13**, 1255-1267.
- (Korzhev, 2001) D. M. Korzhnev, M. Billeter, A. S. Arseniev and V. Y. Orekhov “NMR studies of Brownian tumbling and internal motions in proteins” *Prog Nucl Magn Reson Spectrosc* (2001), **38**, 197-266.

- (Korolev, 2003) N. Korolev, A.P. Lyubartsev, A. Laaksonen, L. Nordenskiöld "A molecular dynamics simulation study of oriented DNA with polyamine and sodium counterions: diffusion and averaged binding of water and cations" *Nucleic Acid Res* (2003), **31**, 5971-5981.
- (Korolev, 1999) N. Korolev, A.P. Lyubartsev, A. Rupprecht, L. Nordenskiöld "Competitive binding of Mg^{2+} , Ca^{2+} , Na^+ , and K^+ ions to DNA in Oriented DNA Fibers: experimental and Monte Carlo simulation results" *Biophysical J* (1999), **77**, 2736-2749.
- (Kulińska, 2000) K. Kulińska, T. Kuliński, A. Lyubartsev, A. Laaksonen, R.W. Adamiak "Spatial distribution functions as a tool in the analysis of ribonucleic acids hydration — molecular dynamics studies" *Comput Chem* (2000), **24**, 451-457.
- (Kusalik, 1994) P.G. Kusalik, I.M. Svishchev "The Spatial Structure in Liquid Water" *Science* (1994), **265**, 1219-1221.
- (Kuszewski, 1999) J. Kuszewski, A.M. Gronenborn, G.M. Clore "Improving the Packing and Accuracy of NMR Structures with a Pseudopotential for the Radius of Gyration" *J Am Chem Soc* (1999), **121**, 2337-2338.
- (Lamberti, 2002) V.E. Lamberti, L.D. Fosdick, E.R. Jessup, C.J.C. Schauble "A hands-on introduction to molecular dynamics" *J Chem Ed* (2002), **79**, 601-606.
- (Lange, 2005) O.F. Lange, H. Grubmüller, B.L. de Groot "Molecular Dynamics simulations of protein G challenge NMR-derived correlated backbone motions" *Angew Chemie Int Ed* (2005), **44**, 3394-3399.
- (Langley, 1998) D.R. Langley "Molecular dynamic simulations of environment and sequence dependent DNA conformations: the development of the BMS nucleic acid force field and comparison with experimental result" *J Biomol Struct Dyn* (1998), **16**, 487-509.
- (Leach, 1999) A.R. Leach "Molecular modelling: principles and applications" (1999), Harlow, England Longman.
- (Lee, 1971) B. Lee, F.M. Richards "The interpretation of protein structures: estimation of static accessibility" *J Mol Biol* (1971), **55**, 379-380.
- (Leulliot, 2001) N. Leulliot, G. Varani "Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture" *Biochemistry* (2001), **40**, 7947-7956.
- (Levy, 1981) R.M. Levy, M. Karplus, P.G. Wolynes "NMR relaxation parameters in molecules with internal motion: exact Langevin trajectory results compared with simplified relaxation models" *J Am Chem Soc* (1981), **103**, 5998-6011.
- (Lii, 1989) J.H. Lii, N.L. Allinger "Molecular mechanics. The MM3 force field for hydrocarbons. 2. Vibrational frequencies and thermodynamics" *J Am Chem Soc* (1989), **111**, 8566-8575.
- (Lindahl, 2001) E. Lindahl, B. Hess, D. van der Spoel "GROMACS 3.0: a package for molecular simulation and trajectory analysis" *J Mol Model* (2001), **7**, 306-317.

- (Ma, 2006) H. Ma, D. J. Proctor, E. Kierzek, R. Kierzek, P. C. Bevilacqua, M. Gruebele “Exploring the energy landscape of a small RNA hairpin” *J Am Chem Soc* (2006), **128**, 1523–1530.
- (MacKerell, 2000) A.D. MacKerell, Jr., N.K. Banavali “All-atom empirical force field for nucleic acids: II. Application to molecular dynamics Simulations of DNA and RNA in solution” *J Comput Chem* (2000), **21**, 105-120.
- (MacKerell, 1995) A.D. MacKerell, Jr., J. Wiórkiewicz-Kuczera, M.Karplus “An all-atom empirical energy function for the simulation of nucleic acid” *J Am Chem Soc* (1995), **117**, 11946-11975.
- (Maiorov, 1995) V.N. Maiorov, G.M. Crippen "Size-independent comparison of protein three-dimensional structures” *Proteins* (1995), **22**, 273-283.
- (Mark, 2002) P. Mark, L. Nilsson “Structure and dynamics of liquid water with different long-range interaction truncation and temperature control methods in molecular dynamics simulations” *J Comput Chem* (2002), **23**, 1211-1219.
- (Martínez, 2001) J. M. Martínez, S.K.C. Elmroth, L. Kloo “Influence of sodium ions on the dynamics and structure of single-stranded DNA oligomers: a molecular dynamics study” *J Am Chem Soc* (2001), **123**, 12279-12289.
- (McQuarrie, 1976) D.A. McQuarrie “Statistical mechanics” (1976), New York: Harper and Row.
- (Miller, 1997) J. Miller, P. Kollman “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* (1997), **270**,436–450.
- (Mitsutake, 2001) A. Mitsutake, Y. Sugita, Y. Okamoto “Generalized-ensemble algorithms for molecular simulations of biopolymers” *Biopolymers (Peptide Science)* (2001), **60**, 96-123.
- (Miyamoto, 1992) S. Miyamoto, P.A. Kollman “SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models” *J Comput Chem* (1992), **13**, 952-962.
- (Moore, 2006) P.B. Moore, T.A. Steitz “The role of RNA in the synthesis of protein” *The RNA World* (eds. R.F. Gesteland, T.R. Cech, J.F. Atkins) 3rd ed. New York: Cold Spring Harbor (2006) 257-285.
- (Murthy, 1999) V.L. Murthy, R. Srinivasan, D.E. Draper, G.D. Rose1 “A Complete Conformational Map for RNA” *J Mol Biol* (1999), **291**, 313-327.
- (Nadassy, 1999) K. Nadassy, S.J. Wodak, J. Janin, “Structural features of protein-nucleic acid recognition sites” *Biochemistry* (1999), **38**, 1999-2017.
- (Nina, 2002) M. Nina, T. Simonson “Molecular dynamics of the tRNA^{Ala} acceptor stem: Comparison between continuum reaction field and particle-mesh Ewald electrostatic treatments” *J Phys Chem B* (2002), **106**, 3696–3705.

- (Nissen, 2000) P. Nissen, J. Hansen, N. Ban, P.B. Moore, T.A. Steitz “The structural basis of ribosome activity in peptide bond synthesis“ *Science* (2000) **289**, 920-930.
- (Nivón, 2004) L.G. Nivón, E.I. Shakhnovich “All-atom Monte Carlo simulation of GCAA RNA folding” *J Mol Biol* (2004), **344**, 29–45.
- (Norberg, 2003) J. Norberg, L. Nilsson “Advances in biomolecular simulations: methodology, and recent applications” *Quartely Reviews of Biophysics* (2003), **36**, 257-306.
- (Nosé, 1996) S. Nosé “Constant-temperature molecular dynamics” *J Phys Condens Matter* (1990), **2**, 115-119.
- (Ohlenschläger, 2004) O. Ohlenschläger, J. Wöhnert, E. Bucci, S. Seifz, S. Häfner, R. Ramachandran, R. Zell, M. Görlach “The structure of the stemloop D subdomain of Cocksackievirus B3 cloverleaf RNA and its interaction with Proteinase 3C” *Structure* (2004), **12**, 237-248.
- (Okabe, 2001) T. Okabe, M. Kawata, Y. Okamoto, M. Mikami “Replica-exchange Monte Carlo for the isobaric-isothermal ensemble” *Chem Phys Lett* (2001), **335**, 435-439.
- (Oostenbrink, 2004) C. Oostenbrink, A. Villa, A.E. Mark, W.F. van Gunsteren “A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6” *J Comput Chem* (2004), **25**, 1656-1676.
- (Orozco, 2003) M. Orozco, A. Pérez, A. Noy, J. Luque “Theoretical methods for the simulation of nucleic acids” *Chem Soc Rev* (2003), **32**, 350-364.
- (Palmer III, 1992) A.G. Palmer, III, D.A. Case “Molecular dynamics analysis of NMR relaxation in a zinc-finger peptide” *J Am Chem Soc* (1992), **114**, zz9059-9067.
- (Pande, 1999) V. Pande, D. S. Rokhsar “Molecular dynamics simulations of unfolding and refolding of beta-hairpin fragment of protein G “ *Proc Natl Acad Sci USA* (1999), **96**, 9062–9067.
- (Persson, 2002) T. Persson, R.K. Hartmann, F. Eckstein “Selection of Hammerhead ribozyme variants with low Mg²⁺ requirement: Importance of stem-loop II” *ChemBioChem* (2002), **3**, 1066-1071.
- (Peter, 2001) C. Peter, X. Daura, W.F. van Gunsteren “Calculation of NMR-relaxation parameters for flexible molecules from molecular dynamics simulations” *J Biomol NMR* (2001) **20**, 297–310.
- (Proctor, 2002) D.J. Pande, J.E. Schaak, J.M. Bevilacqua, C.J. Falzone, P.C. Bevilacqua “Isolation and characterization of a family of stable RNA tetraloops with the motif YNMG that participate in tertiary interactions” *Biochemistry* (2002), **41**, 12062-12075.

- (Prompers, 2001) J.J. Prompers, R. Bruschweiler “Reorientational Eigenmode Dynamics: A Combined MD/NMR Relaxation Analysis Method for Flexible Parts in Globular Proteins” *J Am Chem Soc* (2001), **123**, 7305-7313.
- (Rao, 2003) F. Rao, A. Caflisch ”Replica exchange molecular dynamics simulations of reversible folding” *J Chem Phys* (2003), **119**, 4035–4042.
- (Reddy, 2003) S.Y. Reddy, F. Leclerc, M. Karplus “DNA polymorphism: a comparison of force fields for nucleic acids” *Biophys J* (2003), **84**, 1421-1449.
- (Rhee, 2003) Y. M. Rhee, and V. S. Pande “Multiplexed-replica exchange molecular dynamics method for protein folding simulation” *Biophys J* (2003), **84**, 775–786.
- (Rieder, 2003) E. Rieder, W. Xiang, A. Paul, E. Wimmer “Analysis of the cloverleaf element in a human rhinovirus type 14/poliovirus chimera: correlation of subdomain d structure, ternary protein complex formation and virus replication” (2003), *J Gen Virol*, **84**, 2203–2216.
- (Róg, 2003) T. Róg, K. Murzyn, M. Pasenkiewicz, M. Pasenkiewicz-Gierula “Molecular dynamics simulations of charged and neutral lipid bilayer: treatment of electrostatic interactions” *Acta Biochemica Polonica* (2003), **50**, 789-798.
- (Roy, 2006) H. Roy, M. Ibba “Sticky end in protein synthesis” *Nature* (2006), **443**, 41-42.
- (Ryckaert, 1977) J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen “Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkenes” *J Comput Phys* (1977), **23**, 327-341.
- (Sanbonmatsu, 2006) K.Y. Sanbonmatsu “Energy landscape of the ribosomal decoding center” *Biochimie* (2006), **88**, 1053–1059.
- (Sanbonmatsu, 2002) K.Y. Sanbonmatsu, A. E. García “Structure of metenkephalin in explicit aqueous solution using replica exchange molecular dynamics” *Proteins: Struct Funct Genet* (2002), **46**, 225–234.
- (Saenger, 1988) W. Saenger “Principles of nucleic acid structure” (1988), Springer, New York
- (Sagui, 1999) C. Sagui, T.A. Darden “Molecular dynamics simulations of biomolecules: long-range electrostatic effect” *Annu Rev Biophys Biomol Struct* (1999), **28**, 155-179.
- (Sarzynska, 2000) J. Sarzynska, T. Kulinski, L. Nilsson “Conformational dynamics of a 5s rRNA hairpin domain containing loop D and a single nucleotide bulge” *Biophys J* (2000), **79**, 1213–1227.
- (Seibert, 2005) M.M. Seibert, A. Patriksson, B. Hess, D. van der Spoel “Reproducible polypeptide folding and structure prediction using molecular dynamics simulations” *J Mol Biol* (2005), **354**, 173–183.

- (Serra, 2002) M.J. Serra, J. D. Baird, T. Dale, B. L. Fey, K. Retatagos, E. Westhof “Effects of magnesium ions on the stabilization of RNA oligomers of defined structures” *RNA* (2002), **8**, 307-323.
- (Serra, 1993) M.J.Serra, M.H.Lyttle, T. J.Axenson, C.A.Schadt, D.H.Turner “RNA hairpin loop stability depends on closing base pair” *Nucleic Acids Res* (1993), **21**, 3845-3849.
- (Shirts, 2005) M.R. Shirts, V.S. Pande “Solvation free energies of amino acid side chain analogs for common molecular mechanics water models” *J Chem Phys* (2005), **122**, 134508.
- (Sinclair, 1984) A. Sinclair, D. Alkema, R.A. Bell, J.M. Coddington, D.W. Hughes, T. Neilson, P.J. Romaniuk “Relative stability of guanosine-cytidine diribonucleotide cores: A ¹H NMR assessment” *Biochemistry* (1984), **23**, 2656-2662.
- (Snymann, 1982) J.A. Snymann “A new and dynamic method for unconstrained minimization” *Appl Math Model* (1982), **6**, 449-462.
- (Soares, 2004) T.A. Soares, P.H. Hünenberger, M.A. Kastholz, V. Kräutler, T. Lenz, R.D. Lins, C. Oostenbrink, W.F. van Gunsteren “An improved nucleic acid parameter set for the GROMOS force field” *J Comput Chem* (2004), **26**, 725-737.
- (Soper, 1986) A.K. Soper, M.G. Phillips “A new determination of the structure of water at 25°C” *Chem Phys* (1986), **107**, 47-60.
- (Sorin, 2005) E.J. Sorin, Y.M. Rhee, V.S. Pande “Does water play a structural role in the folding of small nucleic acids?” *Biophys J* (2005), **88**, 2516–2524.
- (Sorin, 2004) E. J. Sorin, B. T. Nakatani, Y. M. Rhee, G. Jayachandran, V. Vishal, V. S. Pande “Does native state topology determine the RNA folding mechanism?” *J Mol Biol* (2004), **337**, 789–797.
- (Sorin, 2003) E.J. Sorin, Y.M. Rhee, B.J. Nakatani, V.S. Pande “Insights into nucleic acid conformational dynamics from massively parallel stochastic simulations” *Biophys J* (2003), **85**, 790–803.
- (Sorin, 2002) E.J. Sorin, M.A. Engelhardt, D. Herschlag, V.S. Pande “RNA simulations: Probing hairpin unfolding and the dynamics of a GNRA tetraloop” *J Mol Biol* (2002), **317**, 493–506.
- (Stone, 1997) A.J. Stone “The theory of intermolecular forces” (1997), New York, USA: Oxford University Press.
- (Storz, 2006) G. Storz, S. Gottesman “Versatile roles of small RNA regulators in bacteria” *The RNA World* (eds. R.F. Gesteland, T.R. Cech, J.F. Atkins) 3rd ed. New York: Cold Spring Harbor (2006) 567-594.
- (Sugita, 1999) Y. Sugita, Y. Okamoto “Replica-exchange molecular dynamics method for protein folding” *Chem Phys Lett* (1999), **314**, 141-151.
- (Svishchev, 1993) I.M. Svishchev, P.G. Kusalik “Structure in liquid water: A study of spatial distribution functions” *J Chem Phys* (1993), **99**, 3049-3058.

- (Tan, 2006) Z.J. Tan, S.J. Chen “Nucleic acid helix stability: effects of salt concentration, cation valence and size, and chain length” *Biophys J* (2006), **90**, 1175–1190.
- (Tironi, 1995) I.G. Tironi, R. Sperb, P.E. Smith, W.F. van Gunsteren “A generalized reaction field method for molecular dynamics simulations” *J Chem Phys* (1995), **102**, 5451-5459.
- (Tuckerman, 2000) M.E. Tuckerman, G.J. Martina “Understanding modern molecular dynamic: techniques and applications” *J Phys Chem* (2000), **104**, 159-178.
- (Uesugi, 1984) S. Uesugi, M. Ohkubo, E. Ohtsuka, M. Ikehara, Y. Kabayashi, Y. Kyogoku, H.P. Westerink, G.A. van der Marel, J.H. van Boom, C.A.G. Haasnoot “Conformation of ribooligonucleotide duplexes containing an alternating C-G sequence which show an unusual circular dichroism spectrum” *J Biol Chem* (1984), **259**, 1390-1393.
- (Uesugi, 1982) S. Uesugi, T. Shida, M. Ikehara “Synthesis and properties of CpG analogues containing an 8-Bromoguanosine residue. Evidence for Z-RNA duplex formation” *Biochemistry* (1982), **21**, 3400-3408.
- (Uhlenbeck, 1990) O.C. Uhlenbeck “Tetraloops and RNA folding” *Nature* (1990), **346**, 613-614.
- (van der Spoel, 2005) D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen “GROMACS: Fast, flexible, and free” *J Comput Chem* (2005), **26**, 1701-1718.
- (van der Spoel, 2004) D. van der Spoel, E. Lindahl, B. Hess, A.R. van Buuren, E. Apol, P.J. Meulenhoff, D.P. Tieleman, A.L.T.M. Sijbers, K.A. Feenstra, R. van Drunen, H.J.C. Berendsen “Gromacs User Manual version 3.3 Nijenborgh 4, 9747 AG Groningen, The Netherlands Internet: <http://www.gromacs.org> (2004)
- (van Gunsteren, 2006) W.F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D.P. Geerke, A. Glättli, P.H. Hünenberger, M.A. Kastholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N.F.A. van der Vergt, H.B. Yu “Biomolecular modelling: goals, problems, perspectives” *Angew Chem Int* (2006), **45**, 4064-4092.
- (van Gunsteren, 1998) W.F. van Gunsteren “Validation of molecular dynamics simulation” *J Chem Phys* (1998), **108**, 6109-6116.
- (van Gunsteren, 1990) W.F. van Gunsteren, H.J.C. Berendsen “Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry” *Angew Chem Int* (1990), **29**, 992-1023.
- (Varani, 2000) G. Varani, W.H. McClain “A fundamental building block of RNA structure crucial to RNA function in diverse biological systems” *EMBO Reports* (2000), **1**, 18-23.

- (Verlet, 1967) L. Verlet “Computer “experiment” on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules” *Phys Rev* (1967), **159**, 98-103.
- (Villa, 2006) A. Villa, G. Stock. “What NMR relaxation can tell us about the internal motion of an RNA hairpin: A molecular dynamics simulation study” *J Chem Theory Comput* (2006), **2**, 1228–1236.
- (Walser, 2001) R. Walser, P.H. Hünenberger, W.F. van Gunsteren “Comparison of different schemes to treat long-range electrostatic interactions in molecular dynamics simulations of a protein crystal” *Proteins* (2001), **44**, 509-519.
- (Wand, 2001) A. J. Wang “Dynamic activation of protein function: A view emerging from NMR spectroscopy” *Nature Struct Biol* (2001), **8**, 926–931.
- (Wang, 2001) W. Wang, O. Donini, C.M. Reyes, P.A. Kollman “Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions” *Annu Rev Biophys Biomol Struct* (2001), **30**, 211-243.
- (Wang, 2000) J. Wang, P. Cieplak, P.A. Kollman “How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?” *J Comput Chem* (2000), **21**, 1049-1074.
- (Weiner, 1981) P.K. Weiner, P.A. Kollman “AMBER: assisted model building with energy refinement. A general program for modelling molecules and their interactions” *J Comput Chem* (1981), **2**, 287-303.
- (Weiner, 1984) S.C. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., P. Weiner “A new force field for molecular mechanic simulation of nucleic acids and proteins” *J Am Chem Soc* (1984), **106**, 765-784.
- (Westerink, 1984) H.P. Westerink, G.A. van der Marel, J.H. van Boom, C.A.G. Haasnoot “Conformational analysis of r(CGCGCG) in aqueous solution: an A-type double helical conformation studied by two-dimensional nuclear Overhauser effect spectroscopy” *Nucleic Acids Res* (1984), **12**, 4323-4338.
- (Westhof, 2003) E. Westhof “RNA solvation: molecular dynamics simulation perspective” *Biopolymers* ((2001) **56**, 266–274.
- (Williams, 2000) J. Williams, K. Hall “Experimental and computational studies of the G(UUCG)C RNA tetraloop” *J Mol Biol* (2000), **297**, 1045–1061.
- (Williams, 1999) J. Williams, K. Hall “Experimental and theoretical studies of the effects of deoxyribose substitutions on the stability of the UUCG tetraloop” *Biophys J* (1999) **76**, 3192–3205.
- (Wilson, 1995) K.S. Wilson, P.H. von Hippel “Transcription termination at intrinsic terminators: the role of the RNA hairpin” *Proc Natl Acad Sci USA* (1995), **92**, 8793–8797.

- (Woese, 1990) C. Woese, S. Winker, R. Gutell “Architecture of ribosomal RNA: Constrains on the RNA tetraloops” *Proc Natl Acad Sci* (1990), **87**, 8467–8471.
- (Wolters, 1992) J. Wolters “The nature of preferred hairpin structures in 16S-like rRNA variable regions” *Nucleic Acids Res* (1992) **20**, 1843-1850
- (Wu, 2000) P. Wu, N. Sugimoto “Transition characteristics and thermodynamic analysis of DNA duplex formation: a quantitative consideration for the extent of duplex association” *Nucleic Acids Res* (2000), **28**, 4762-4768.
- (Zacharias, 2000) M. Zacharias “Simulation of the structure and dynamics of nonhelical RNA motifs” *Curr. Opin. Struct. Biol.* (2000), **10**, 311–317.
- (Zhang, 2006) Q. Zhang, X. Sun, E.D. Watt, H.M. Al-Hashimi “Resolving the motional modes that code for RNA adaptation” *Science* (2006), **311**, 653-656.
- (Zell, 2002) R. Zell, K. Sidigi, A. Stelzner, M. Görlach “Determinants of the recognition of enteroviral cloverleaf RNA by Coxsackievirus B3 proteinase 3C” *RNA* (2002), **8**,188–201.
- (Zhou, 2003) R. Zhou “Trp-cage: Folding free energy landscape in explicit water” *Proc Natl Acad Sci USA* (2003), **100**, 13280–13285.
- (Zhou, 2001) R. Zhou, B. J. Berne, R. Germain “The free energy landscape for β hairpin folding in explicit water” *Proc Natl Acad Sci USA* (2001), **98**, 14931–14936.
- (Zichi, 1995) D.A. Zichi “Molecular dynamics of RNA with the OPLS force field. Aqueous simulation of a hairpin containing a tetranucleotide loop” *J Am Chem Soc* (1995), **117**, 2957-2969.

Zusammenfassung

Das genetische Material der Zellen besteht aus Molekülketten der Desoxyribonukleinsäure (DNA), die ein Träger der Erbinformation ist. In normalen Körperzellen wird die Erbinformation der DNA in eine andere Molekülkette, die sogenannte Ribonukleinsäure (RNA), übersetzt. Die RNA reguliert die Bildung von neuem Protein in der Zelle. Dass die RNA nicht bloß ein „Stempel“ ist, der die Informationen der DNA weitervermittelt, darin sind sich die Experten heute einig. RNA-Moleküle können Informationen speichern, katalytische Aktivitäten entfalten, sich perfekt tarnen, und sie regulieren auch als Produkt ihre eigene Synthese. Manche Viren enthalten ebenfalls RNA (oder DNA) und können so den Produktionsapparat der Zelle täuschen. Erkenntnisse über die Wechselwirkung dieser RNA mit natürlichen und synthetischen Liganden können zur Suche nach potentiellen Wirkstoffen beitragen.

Nukleinsäuren sind lineare Biopolymere von grundlegenden Untereinheiten, die Nukleotide genannt werden und aus Adenin (A), Cytosin (C), Guanin (G), Urazil (U), und Thymin (T) zusammengesetzt sind. Sie sind jedoch in der Lage sich zu falten und so eine Doppel-Helixstruktur auszubilden. Diese besteht größtenteils aus den bekannten "Watson-Crick-Basenpaaren" (G-C und A-U oder A-T), die zur Stabilität der Struktur beitragen, sowie aus den weniger stabilen G-U-Paaren. Durch die Wechselwirkung zwischen verschiedenen Sekundärstrukturelementen entstehen Tertiärstrukturelemente, deren Struktur und Dynamik oft nur schwer experimentell zu bestimmen sind.

Fortschritte in der RNA-Strukturanalyse wurden durch Röntgenkristallographie und Kernresonanzspektroskopie (NMR) möglich. Durch die Röntgenkristallographie wurden viele RNA-Eigenschaften festgestellt. Allerdings besteht keine Kristallstruktur für alle mögliche Einzelfaser-RNA-Haarnadeln, weil diese immer dazu neigen, in eine linearen doppelte Faserform zu kristallisieren, die geringe biologische Bedeutung hat. Außerdem wurde mit Hilfe der NMR-Spektroskopie das dynamische Verhalten von RNA, z.B. Entfaltungsprozesse bei ansteigender Temperatur, beobachtet. Jedoch erlauben diese experimentellen Daten oft keine direkte mikroskopische Beschreibung der molekularen Prozesse. Molekulardynamik (MD)-Simulationen von biologischen Systemen ermöglichen es hingegen, diese Prozesse in atomischem Detail zu untersuchen. Die MD-Simulation beschreibt ein molekulares System auf atomarer Ebene

mit Hilfe der klassischen Mechanik. Kräfte werden von empirischen Potentialen abgeleitet. Sie liefern zeitabhängige Trajektorien, die sich aus den Newton'schen Bewegungsgleichungen ergeben.

Durch verbesserte Computerleistung, bessere Kraftfelder, und neu entwickelte genauere Methoden stimmen heutzutage MD-Simulationen von RNA mit experimentellen Daten immer besser überein. In meiner Doktorarbeit wurden MD-Simulationen durchgeführt um die Dynamik, die Struktur und insbesondere die Stabilität von RNA-Hairpins theoretisch zu beschreiben, um so ein erweitertes Verständnis für die dynamischen Vorgänge zu erhalten. Auch der SFB 579 der Universität Frankfurt beschäftigt sich mit RNA-Systemen. Erforscht wird unter anderem der D-Loop des Coxsackievirus B3 (CVB3), der Virenmyocarditis verursacht. Die Interpretation dieser experimentellen Daten wird durch MD-Simulation möglich.

In dieser Arbeit wurden das GROMACS Software-Paket und das AMBER Kraftfeld verwendet, um das strukturelle, dynamische und thermische Verhalten der RNA-Hairpins mit Hilfe von MD-Simulationen auf atomarer Ebene zu untersuchen. Betrachtet wurden die 14-mer RNA-Hairpins, uCACGg und cUUCGg. Die verfügbaren NMR-Strukturen zeigen, dass das uCACGg-Tetraloop auffallend ähnlich in der gesamten Geometrie und den Wasserstoffbindungen zu der experimentellen Struktur des cUUCGg-Tetraloop ist, obwohl die schließende Basenpaarsequenz der beiden Tetraloops unterschiedlich sind. Trotz beachtlicher struktureller Ähnlichkeit unterscheiden sich allerdings die uCACGg und cUUCGg Tetraloops in Funktionalität und Thermostabilität.

Zunächst orientiert sich unser erstes Bemühen an der Frage nach einem guten Modell für RNA-Hairpins und Simulationsbedingungen, um die zu untersuchenden RNA-Hairpins in Wasser möglichst realitätsnah zu simulieren. Erstens werden drei Versionen des biomolekularen AMBER-Kraftfelds geprüft, indem man die 60 ns Simulationen des 14-mer uCACGg-Hairpins durchführt. Die simulierten strukturellen Eigenschaften und Atomfluktuationen zeigen hohe Ähnlichkeiten in den drei Kraftfeldern. Darüber hinaus stimmen die von MD-Simulationen berechneten Atomkernabstände mit den experimentellen NMR-Daten gut überein. Die gute Übereinstimmung zwischen den Simulationen und den strukturellen NMR Daten belegt die Fähigkeit des AMBER-Kraftfelds zur Beschreibung der strukturellen Eigenschaft von kleinen RNA-Hairpins. Anschließend werden die Einflüsse der Methoden, welche

die langreichweitigen, elektrostatischen Wechselwirkungen beschreiben, auf die strukturellen Eigenschaften untersucht. Insbesondere werden die Ergebnisse der Reaktionfeld-Methode mit denen der Particle Mesh Ewald (PME)-Methode verglichen. Es zeigt sich, dass die PME-Methode die elektrostatischen Wechselwirkungen am besten beschreibt, auch wenn die Simulationen der beiden Methoden Ähnlichkeit in der Struktur-Stabilität und der Atomfluktuation bei niedriger Natriumkonzentration aufweisen. Drittens wird der Kationeffekt auf die RNA-Stabilität untersucht. Betrachtet wurden zwei unterschiedliche Kationen (ein- und zweiwertig) und verschiedene Konzentrationen. Die Simulationen weisen darauf hin, dass sich die Metallionen in der Affinität zum RNA-Hairpin unterscheiden, wenn Na^+ und/oder Mg^{2+} als Gegenionen verwendet werden. Weiterhin wird gezeigt, dass sich die bevorzugten Positionen der Na^+ -Ionen in der großen Furche (*major groove*) des RNA-Hairpins befinden. Insbesondere die Anlagerungsort der Na^+ -Ionen liegen in der Nähe des schließenden Basenpaar U5-G10. Im Vergleich zu Na^+ -Ionen lagern sich Mg^{2+} -Ionen sowohl an die RNA-Basen U3, A4-U11, und die Phosphat-Gruppe, als auch an das schließenden Basenpaar U5-G10 an.

Bestätigt werden die Modelle und Simulationsbedingungen durch den Vergleich von Parametern, die sowohl experimentell als auch durch Simulationen ermittelt werden können. Ferner erlauben MD-Simulationen Einblick in das System, indem sie detaillierte Konformations- und andere Verteilungen liefern. In der vorliegenden Arbeit wurden die Einflüsse der Loopsequenz und des schließenden Basenpaares auf die Verteilung der Konformationen, der internen Bewegungen, und auf die Thermostabilität von zwei RNA-Hairpins mit Hilfe dieser Modelle untersucht.

Zunächst wurden die strukturellen Eigenschaften bei Raumtemperatur ausgewertet. Die starken strukturellen Ähnlichkeiten und die gute Übereinstimmung mit NMR-Daten bestätigen die Hypothese, dass die zwei Tetraloops zur gleichen "erweiterten" RNA-Familie gehören. Diese zwei Hairpins haben ähnliche Lösemittel zugängliche Oberflächen (*solvent accessible surface*), wobei deren Lösemittel zugänglichen funktionellen Gruppen unterschiedlich sind. Weiterhin weist das uCACGg-Hairpin eine stärkere Tendenz auf Wasserstoffe abzugeben als das cUUCGg-Hairpin, was in den unterschiedlichen Bindungsaffinitäten zwischen diesen Hairpins und der viralen Protease begründet liegt.

Darüber hinaus wurde der Faltungs- und Entfaltungsprozess mit Hilfe der Replica-Exchange-Molekulardynamik-Simulationen untersucht. Diese Untersuchung

zielt auf das bessere Verständnis der unterschiedlichen Thermostabilität der Hairpins, indem sie die möglichen Zwischenprodukte im atomaren Detail liefern. Sowohl experimentell als auch von den MD-Simulationen ergibt sich eine Differenz in den Schmelztemperaturen der beiden Hairpins von ungefähr 20 K. Allerdings sind die von MD beobachteten Schmelztemperaturen 20 % höher als die von Experiment zu ansehende Wert. Die Ergebnisse machen deutlich, dass die Schmelztemperaturdifferenz nicht auf die Unterschiede in der Sequenz, in der Struktur, oder in der Dynamik der Loops zurückführen sind, sondern auf die Unterschiede der Basenpaaren in den Stämmen. Weiterhin wird gezeigt, dass sich das uCACGg-Hairpin einerseits kooperativ entfaltet, und die Entfaltung des cCACGg-Hairpins anderseits weniger kooperativ stattfindet.

Um die schnelle interne Dynamik der uCACGg- und cUUCGg-Hairpins zu untersuchen, erlauben die Simulationen von 50 ns eine akurate Beschreibung der schnellen internen Bewegung der RNA-Hairpin, obwohl der den Hairpins zugängliche Konformationsraum nicht vollständig abgedeckt wird. Die NMR-Relaxationsparameter, die mit Hilfe der MD-Simulationen zurückgerechnet wurden, bestätigen das Modell und die Simulationsbedingungen der MD-Simulationen. Im Hinblick auf die Übereinstimmung kann man den besten Ansatz zur Berechnung der NMR-Ordnungsparameter bestimmen.

In dieser Arbeit wurden drei verschiedene Ansätze angewandt, nämlich das Fitting von 100 ps auf modellfreiem Ansatz nach Lipari-Szabo, *equilibrium average*, und das *Gaussian Axial Fluctuation* (GAF)-Modell. Die zwei letzteren können nur qualitativ mit den experimentellen Daten übereinstimmen.

Die NMR-Ordnungsparameter können mit Hilfe des Modells von Lipari-Szabo richtig ermittelt werden, wenn sich die interne Bewegung in kleineren Zeitskalen als zur Gesamtbewegung vollzieht. Voraussetzung für die Berechnung dieses Modells ist aber, dass das Fitting der internen Korrelationsfunktionen nur auf den ersten Teil von 100 ps der Korrelationsfunktionen eingesetzt wird. Die berechneten Ordnungsparameter deuten auf ein unterschiedliches Verhalten der beiden Hairpins besonders im Loop-Bereich hin. Die konformationelle Umordnung, die beim UUCG-Loop beobachtet wurde, tritt beim CACG-Loop nicht ein.

Zusammenfassend lässt sich sagen, dass es durch den Einsatz von MD Simulationen ermöglicht wird, die strukturellen und dynamischen Eigenschaften der

RNA-Systeme auf atomarer Ebene zu untersuchen. Als Schlussfolgerung, zeigt diese Doktorarbeit, dass sich die Studie der konformationell Dynamik der RNA-Systeme durch die Kombination aus MD-Simulation und NMR-Spektroskopie sowie der Leistungsfähigkeit der MD-Simulationen, die die interne Bewegungen deutlich beschreiben können, untersuchen lässt.

Lebenslauf

Persönliche Daten

Name : Catherina Widjajakusuma
Rufname (Taufname) : Elisabeth
Geboren am 05. 04. 1974 in Ujung Pandang (Indonesien)
Ledig, katholisch

Schulbildung

07/1980–06/1986 Besuch der katholischen Grundschule Sankt Joseph Rajawali
07/1986–07/1992 Besuch der katholischen Hauptschule Rajawali

Studium

08/1992–08/1996 Studium der Chemie (Sarjana) an der Airlangga Universität in Surabaya
09/1999-09/2001 Studium der Physikalische Chemie (Magister) an der *Bandung Institute of Technology* in Bandung
10/2004-09/2007 DAAD-Stipendium für die Promotion an der Johann Wolfgang Goethe Universität in Frankfurt. Thema : Investigation of the Structure and Dynamics of RNA Systems by Molecular Dynamics Simulations.

Berufstätigkeit / Zivildienst

August 2005 Volontär des XX. Weltjugendtags in Köln
01/1997- dato Wissenschaftlicher Mitarbeiter am Lehrstuhl für Physikalische Chemie an der Katholischen Universität Widya Mandala Surabaya in Surabaya

Publikationen

1. Alessandra Villa, Elisabeth Widjajakusuma, Gerhard Stock, "Molecular dynamics simulation of the structure, dynamics, and thermostability of the RNA hairpins uCACGg and cUUCGg" (2007). Biophysical J **submitted**.
2. Elisabeth Catherina Widjajakusuma, Alessandra Villa, Gerhard Stock, "Role of electrostatic and ion effect on RNA stability: A molecular dynamics study" (2007). **in preparation**.