RNA Structure from Molecular Dynamics Simulations

Shirley Louise-May, Pascal Auffinger and Eric Westhof
Institut de Biologie Moléculaire et Cellulaire du CNRS
Modélisations et Simulations des Acides Nuc léiques, UPR 9002
15 rue René Descartes
67084 Strasbourg Cedex
FRANCE

Abstract

With the aim of helping to define how molecular dynamics simulation methods can most successfully be applied to RNA structural systems, we have undertaken a series of 100 ps molecular dynamics simulations of the fully neutralized and solvated anticodon and thymine hairpins of tRNAArg totalling 2.8 ns. Specifically, our studies focus on the critical evaluation of MD trajectories with respect to the assessment of MD protocol stability, the accuracy of the electrostatic model and the constancy of MD derived structural parameters, all in the context of RNA tertiary structural topology and dynamics. In each of four independent studies, a multiple trajectory molecular dynamics approach is used, in which multiple independent trajectories starting from the same initial conditions are generated, in order to assess the system stability and to adequately characterize the structural and dynamical interactions of the fragment in the vicinity of the starting configuration. Three independent studies of the anticodon loop were conducted using successively more accurate models of the long-range electrostatic solvent interactions, with all solute-solute interactions explicitly calculated in each case. It was found that short solute-solvent and solvent-solvent truncation distances (8 Å) do not result in consistent or dynamically stabilized solute structural parameters. Extending these truncation distances (to 16 Å) does result in both system and structural stabilization and the contributions of site specific hydration and weaker C-H...O hydrogen bonds in maintenance of the tertiary conformation begin to emerge. Preliminary results on the Particle Mesh Ewald (PME) method applied to the anticodon system indicate yet further stabilization. In a study conducted on the thymine hairpin fragment, using a 16 Å truncation distance for interactions involving the solvent, interactions of the D-loop residues and the T54-A58 reverse-Hoogsteen base pair stability are seen to strongly affect the thymine loop tertiary conformation. Although the solvation properties for this system are in accord with other theoretical estimates, the counterion behavior indicates that protocol improvements are necessary before further structural interpretation of the thymine hairpin is warranted.

Introduction

The structural and functional richness found in RNA systems provides a wealth of questions for structural chemistry and biology researchers trying to understand the local interactions and structural rules which govern the folding process, recognition, binding and cat-
alytic activity of RNA molecules. As more detailed information on RNA systems is gained through X-ray crystallography, NMR, solution probing and footprinting studies, the potential of theoretical methods applied to studies of RNA structure, such as modelling, molecular mechanics and molecular dynamics, begins to be realized. We are currently at the threshold of true integration of experimental and theoretical techniques used to study the structure of biological macromolecules, such as RNA. Like a bridge being built from both shores, experimental techniques are relying more heavily on computational means for structural analysis and refinement (1) while theoretical models and techniques are approaching the experimental conditions of the laboratory (2-5). MD methodology applied to nucleic acid systems has come a long way from its inception over a decade ago (6, 7), however, the structural heterogeneity and cumulative effect of subtle tertiary interactions in RNA systems provides a stringent test for present MD methodologies.

Unlike the linear helical functional structure of DNA, functional RNA structure is more similar to that of proteins; rich in secondary structural motifs and tertiary structural interactions determined by a complex folding process into its native conformation. Although canonical A-form helices constitute the larger part of the structural skeleton in RNAs, noncanonical base pairings between standard bases, as well as base pairings involving modified bases are often found to have functional as well as structural consequences (8). Specific secondary structural motifs of RNAs include double stranded helices, bulges (one or more unpaired residues on one strand), internal loops (several unpaired residues on both strands of an otherwise helical region), hairpin loops (several unpaired residues at the end of a helical region formed by one strand doubling back on itself) and single strands. Tertiary interactions between secondary structural motifs include loop-loop interactions, loop-helix interactions, stranded-helix interactions (triple helices, pseudoknots) and coaxial stacking of helices across loop regions. The folding of an RNA molecule into its native conformation is dependent on the binding of divalent cations (9), as is chemical activity in catalytic RNAs (10). The resulting RNA structures contain a delicate balance of subtle tertiary interactions capable of accommodating the conformational adjustments necessary for the functional activity of RNA molecules.

The debut of nucleic acid studies conducted via the theoretical technique of molecular dynamics came in 1983 with RNA simulations reported by Prabhakaran, Harvey, Mao and McCammon (7), and DNA simulations reported by Levitt (6). Molecular dynamics simulations of a phenylalanine tRNA starting from a 2.5 Å resolution crystal structure (11) were run for 32 ps under vacuum conditions using a modified van der Waals interaction potential, in order to mimic solvent effects. The native tRNA tertiary interactions were not maintained, however, global parameters such as the radius of gyration and accessible polar surface area remained within 5% of the experimental values calculated over the last 12 ps of the trajectory.

The evolution of MD methodology applied to nucleic acid systems between 1984 and 1994 proceeded, in the published literature, solely by its application to studies of linear double stranded DNA structures (12). Early primitive MD protocols which did not explicitly include solvent were found insufficient to generate stable trajectories of RNA fragments, although there was a dedicated unpublished effort towards this aim (13). In general, MD methodological developments for the highly charged nucleic acid systems were
mainly concerned with overcoming the approximations made in the calculation of electrostatic contributions (14, 15), although the gain in computational power over the same period has also allowed for simulations of larger systems, such as the crystallographic unit of a system (4), and longer MD trajectories (16-18). The presence of a pairwise interaction potential in standard MD simulation software packages (19-22) make truncation of the nonbonded interactions of the system necessary to adjust the number of interactions that practically can be evaluated in an MD trajectory of reasonable length in a reasonable amount of time with the available computational means. However, the artifactual effects introduced by straight truncation, switching and shifting functions have been found to affect, in some cases quite dramatically (3, 23-27), the solute conformation and dynamics. Recent implementation of the Particle Mesh Ewald (PME) method (28, 29) for calculating the electrostatic contributions of the system in the AMBER MD software package (30) avoids the issue of truncation of long-ranged interactions and is currently being tested and reported on RNA systems (17).

We report on a series of 100 ps molecular dynamics simulations on the fully neutralized and solvated anticodon loop of tRNA\(^{\text{Amp}}\) totalling 2.2 ns to address issues concerning MD protocol stability (31) and the effect of long ranged hydration forces (32). We have also conducted 0.6 ns of molecular dynamics trajectories on the fully neutralized and solvated thymine loop of tRNA\(^{\text{Amp}}\) to investigate its structural topology and dynamics in the presence and absence of D-loop interactions (33). In all of these molecular dynamics studies, solute-solute interactions, involving the RNA fragment and NH\(_4^+\) counterions, were calculated without truncation in order to account accurately for all long-ranged solute-solute interactions. Solute-solvent and solvent-solvent interactions were calculated out to a specified truncation distance for most of the simulations reported here; to 8 Å for a set of ten 100 ps trajectories of the anticodon loop (31) and extended to 16 Å, using a 12-16 Å twin-range truncation scheme, to include the effects of long ranged hydration forces, for two six simulation sets of the anticodon (32) and thymine loops (33). Preliminary results on six simulations of the anticodon loop using a Particle Mesh Ewald (PME) method for the calculation of electrostatic interactions are also presented. The global RMS deviations, the extent of preservation of tertiary interactions, the solvation parameters and the counterion behavior have been analyzed for each individual trajectory and compared as an ensemble to allow for a critical evaluation of the MD studies. Selected analyses which highlight the information gained from each MD study are presented.

**Computational Procedure**

The AMBER4.1 package (30) was used to run each series of simulations at a constant temperature of 298 K and a constant pressure of 1 atm. with a time step of 2 fs, while the nonbonded pair list was updated every 10 time steps (20 fs). The starting system coordinates were kept the same for each simulation within a given study and were extracted from the crystal structure of tRNA\(^{\text{Amp}}\) (34) (Figure 1). Each system was fully neutralized by NH\(_4^+\) counterions and solvated by SPC/E water molecules (35) filling a rectangular box (Table I). In the anticodon hairpin simulations, the counterions were placed at 5 Å from the phosphorus atom along the OPO bisector. Because of topological constraints in the thymine loop, the counterions were placed using the GROMOS subroutine PROION (19) based on the electrostatic potential of the solvated fragment such that no counterion was placed closer than
4.5 Å from the solute and 3.5 Å to any other ion. The charges for the NH₄⁺ ions were extracted from the work of Singh et al. (36). For all simulations except one, the Pearlman & Kim set of charges derived from low temperature X-ray data of isolated nucleotides was used (37), and the charges of the Ψ32, Ψ55, m¹G37 and ribose T54 residues were adapted from those of the standard bases. For one simulation of the ten in the 8 Å solvent truncation distance anticodon study, the AMBER 4.0 charge set was used (38).

In all simulations reported, no cutoffs were applied to the solute-solute interactions, and truncation distances were only used for solute-solvent and solvent-solvent interactions. Nevertheless, in the following, we will refer to these simulations as 8 Å truncation or 16 Å truncation distance sets. On a SGI R4400 machine, when an 8 Å cutoff was used for the solute-solvent and solvent-solvent interactions, 1 ps of MD for the anticodon hairpin system took approximately one hour of CPU time. By contrast, 1 ps of MD was found to take almost eight hours with a full 16 Å truncation distance applied to all nonbonded solvent interactions. In order to reduce the CPU time required for such calculations, we employed a 12-16 Å twin-range residue based cut-off for the solute-solvent and solvent-solvent interactions, in which the long-range electrostatic forces between 12 Å and 16 Å were updated every 10 time steps (20 fs). By this method, shorter ranged interactions (<12 Å) are calculated at each step of the simulation whereas longer ranged interactions (between 12 Å and 16 Å) are only calculated at each update of the nonbonded pair list and kept constant until the next update, since long-range interactions change in magnitude and gradient more slowly than do shorter ranged interactions (39, 40). The reduction in CPU time spent by using this method for our system over the straight 16 Å truncation distance was significant, taking approximately four hours, i.e. half the CPU time per ps. The Particle Mesh Ewald (PME) method (28, 29) applied to the same anticodon fragment solvated by 1143 SPC/E water molecules, using a 9 Å truncation distance for the Lennard-Jones interactions, a charge grid spacing close to 1.0 Å and a cubic

Figure 1: Two dimensional representations of the tRNA<sup>Ant</sup> anticodon and thymine hairpin loop interacting with a two residue fragment of the D-loop of tRNA<sup>Ant</sup>. Ψ and m¹G notations refer to the nonstandard pseudouridine and 1-methylguanine residues. Solid lines indicate standard Watson-Crick type base pairs and dashed lines indicate noncanonical base pairing schemes.
RNA Structure from M. D. Dynamics Simulations

Table I
Details of the molecular dynamics simulations of the tRNA^{asp} fragments.

<table>
<thead>
<tr>
<th>tRNA^{asp} fragment</th>
<th>no. of trajectories</th>
<th>no. of waters</th>
<th>no. of counterions</th>
<th>truncation method</th>
<th>box dimensions Å³</th>
<th>CPU/ps (SGI R4400)</th>
<th>remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>anticodon loop</td>
<td>10</td>
<td>2856</td>
<td>16</td>
<td>8 Å cutoff*</td>
<td>55.7x43.2 x42.7</td>
<td>= 1 Hr</td>
<td>different equilibration procedures</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anticodon loop</td>
<td>6</td>
<td>3765</td>
<td>16</td>
<td>12-16 Å twin range cutoff?</td>
<td>59.7x47.2 x46.7</td>
<td>= 4 Hrs</td>
<td>random seed change</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anticodon loop</td>
<td>6</td>
<td>1143</td>
<td>16</td>
<td>Particle Mesh Ewald (PME)?</td>
<td>47.3x33.7 x32.7</td>
<td>= 1 Hr</td>
<td>random seed change</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>thymine loop (+ D frag.)</td>
<td>4 (2)</td>
<td>4183 (4220)</td>
<td>16 (17)</td>
<td>12-16 Å twin range cutoff?</td>
<td>58.7x49.5 x50.0 (58.9x49.5 x49.9)</td>
<td>= 4 Hrs (= 4.8 Hrs)</td>
<td>trajectories w/ &amp; w/o D fragment; random seed change</td>
</tr>
</tbody>
</table>

* Cut-off applied only to solute-solvent and solvent-solvent interactions, all solute-solute interactions are calculated explicitly in all simulations.


The equilibration protocol was kept consistent between all our calculations, with some exceptions for simulations of the anticodon hairpin using the 8 Å truncation scheme details are reported in ref. (31)), and consisted of 100 steps of steepest descent minimization applied to the solvent molecules with fixed solute, followed by 5 ps of molecular dynamics at 300 K on the water with fixed RNA and NH₄⁺ counterions. Next ensued molecular dynamics of mobile counterions and water molecules at temperatures of 100 K (1 ps), 200 K (1 ps) and 300 K (5 ps) with again fixed RNA atomic positions. In subsequent steps, no position constraints were applied to the system. The temperature was progressively increased to 298 K in steps of 50 K with 1 ps of MD at each step. Finally, at 298 K, 5 ps of dynamics were run in order to allow the system to equilibrate at room temperature. The thermalization and equilibration phase of the MD protocol thus lasted 22 ps (Figure 2). For each simulation data was collected over the subsequent 100 ps of trajectory. At each temperature increase, the velocities were reassigned randomly according to a Boltzmann distribution. In order to generate uncorrelated trajectories starting from the exact same initial configurations, different values of the random seed were used at the 50 K restart (t = 12 ps, Figure 2) for the simulations using the 16 Å truncation scheme and the PME method, with again some exceptions for simulations of the anticodon hairpin using the 8 Å solvent truncation scheme (for details see reference (31)). The SHAKE algorithm (41) was used at all stages of the simulations to constrain each X-H bond of the system. A 10 kcal/mol⁻¹ harmonic distance constraint was applied to the three hydrogen bonds of the first GC base pair of the helix of both tRNA^{asp} fragments throughout the simulations in order to prevent any stem fraying. The MDdraw program (42) was used to visualize the generated trajectories on a Silicon Graphics workstation.
Results and Discussion

MD Protocol Stability. A systematic search for a stable MD protocol with which to study the structure and dynamics of the anticodon loop of tRNA^{Amp} using an 8 Å truncation scheme was carried out in which minor modifications to the protocol in the equilibration steps of ten 100 ps simulations were made, while starting in all cases from the exact same initial solute configuration (31). Four independent simulations were generated by changing the velocity distribution in the thermalization steps of the equilibration protocol, three simulations explored variations in the length and means of bringing the system to the target temperature during equilibration, and three simulations involved a change in the parameters of the system; the charge set used, the number of water molecules in the system and the version of the AMBER program used (4.0 vs. 4.1). The highest and lowest final structure RMS fit to the crystal starting structure is shown in Figure 3, affording an estimate of the range of structural stability present in the independent MD trajectories generated by using an 8 Å truncation distance for the solute-solvent and solvent-solvent interactions. The sole difference between the two simulation protocols used, resulting in a 2.0 Å RMS deviation discrepancy between them (with respect to the starting structure), concerns the temperature increase during the thermalization period which was reduced to steps of 25 K instead of 50 K, thus smoothing the heating process, in the MD trajectory resulting in the higher RMS deviated final structure.

The entire ensemble of MD simulations revealed structurally divergent trajectories that could not be correlated to the type or magnitude of modifications made to the MD protocol. For example, the change in the charge set used produced an MD trajectory statistically indistinct from the set of four trajectories where only the velocity redistribution in the equilibration phase was changed. Systematic attempts to bring the system to an equilibrated state more slowly or gently also yielded nonsystematic results. As a set, the spread in structural and dynamical characteristics of the system, revealed by both time dependent and time averaged global RMS deviations (Figure 7 top panels), indicated that the protocol used, while calculating explicitly all important solute-solute interactions, is

![Figure 2](image_url)  
*Figure 2:* Curves representing the evolution of the temperature with time during the equilibration procedure for all of the six simulations of the anticodon hairpin using a 12-16 Å solvent truncation distance (32). The arrow at \( t = 12 \) ps indicates the change in velocity distribution at the 50 K restart used to generate independent MD trajectories under the same starting conditions and protocol. A similar equilibration procedure was used for all the MD simulations reported here.
insufficiently reliable to warrant interpretation of detailed structural analyses or extension of any of the ten trajectories to longer MD timescales. The advantages of utilizing a multiple trajectory molecular dynamics approach as a means of assessing the protocol stability, however, did emerge from this study.

Effect of Long-ranged Hydration Forces. Extension of the truncation distance to 16 Å for solute-solvent and solvent-solvent interactions was employed next to the anticodon loop system, in order to improve the preceding protocol and assess the effect of long-ranged hydration forces, generally neglected in molecular MD simulations, on the system stability (32). Six 100 ps trajectories generated using the extended truncation distances and the multiple molecular dynamics approach revealed system stabilization with respect to the 8 Å.

Figure 3: Stereoview of the final structures (bold lines) from the MD trajectories of the lowest and highest RMS fit superpositions to the crystal structure (thin lines) from a set of ten MD trajectories calculated using an 8 Å truncation distance for the solute-solvent and solvent-solvent interactions. The RMS deviation values have been calculated over the last 5 ps of each trajectory (figure adapted from ref. (31)).
À truncation scheme studies, as evidenced by a smaller spread and mean value of both the time dependent and time averaged global RMS deviations from the starting structure (Figure 7 middle panels). The effect of the inclusion of long-ranged hydration forces in the MD protocol on both time averaged structural parameters and their dynamical behavior is illustrated in Figure 4 for the lowest final RMS deviation trajectory of each of the 8 À and 16 À sets (1.8 À and 1.6 À final RMS deviations respectively). The time dependent distances of two interactions important to the maintenance and stabilization of the native tertiary conformation of the anticodon loop are shown. The (U33)N3-H3...OA-P(C36) hydrogen bond, well maintained on average in both trajectories, shows potentially disrupting short dynamical fluctuations to alternate geometries in the 8 À simulation. More dramatically, the stacking of the U33 base on the U35 phosphate group is not maintained at all in the lowest final RMS trajectory of set 8 À, while it is both dynamically stable and well maintained on average in the corresponding lowest final RMS trajectory of the 16 À set. Therefore, even the two structures from each set having the lowest RMS deviations, both close to 1.8 À, do not show the same structural stability. Additional analyses revealed a decrease by half in dihedral angle transitions and an increase by three times (from 32% to 98%) of calculated water molecule occupancies involved in some site specific hydration interactions upon the inclusion of long-ranged hydration forces (32).

The possible contributions of C-H...O hydrogen bonds to the stabilization of the tertiary structure of the anticodon loop became apparent from the frequency and duration of such bonds analyzed in the MD trajectories utilizing the improved 16 À truncation scheme. Such interactions were then sought and identified in the crystal structures of tRNA[A^p] (34) and a literature search revealed possible evidence from neutron diffraction data on small molecules of the structural influence of C-H...O bonds (43-45). C-H...O contacts have been shown to occur for C...O distances between 3.0 À to 4.0 À, angles between 90-180° and having energies (1-2 kcal/mol depending on the acidity of the C-H group) commensurate with other stabilizing tertiary interactions found in macromolecules. The abundance of C-H groups with neighboring protonated nitrogens make nucleic acids systems a good candidate for the possible contributions of C-H...O interactions. Figure 5 shows, at top, two possible C-H...O interactions as seen in the crystal structure, the (U33)O2...H5-C5(C36) and the (U35)C5-H5...O2'(C36) hydrogen bond, and at bottom, the time evolving distance of these interactions for one trajectory of the 16 À truncation distance set where they are well maintained. To varying extents, all possible C-H...O hydrogen bond interactions identified in the crystal structures were, in general, observed in the MD trajectories of the 16 À set and to a lesser extent in the 8 À MD trajectories. This result not only supports the proposal of C-H...O bond contributions to the stabilization of the anticodon tertiary conformation but also highlights the importance of long-ranged hydration forces in stabilizing subtle, yet contributing, tertiary structural interactions.

**Particle Mesh Ewald Method.** Preliminary results of a set of six 100 ps MD trajectories of the anticodon loop of tRNA[A^p] utilizing the Particle Mesh Ewald (PME) method for calculation of electrostatic interactions (28, 29), show an increased set consistency over preceding simulations using truncation schemes as revealed by time dependent and time averaged global RMS deviations from the starting crystal structure (bottom panels of Figure 7). One advantage of the PME model, as an alternative to truncation schemes previously
Figure 4: Time evolution of two interactions maintaining the anticodon loop topology compared for the lowest calculated RMS deviation MD trajectories of each set generated by two different electrostatic models; left: 8 Å truncation distance applied to the solute-solvent and solvent-solvent interactions, right: 12-16 Å twin-range truncation scheme applied to the same interactions (figure adapted from ref. (31)).

Figure 5: At top, the (U33)O2...H5-C5(C36) and the (U35)C5-H5...O2'(C36) hydrogen bonds observed in the crystallographic structure (hydrogen atom positions have been estimated), and their time evolution for one MD trajectory of the anticodon hairpin set of MD trajectories (bottom) generated utilizing a 12-16 Å twin-range truncation scheme applied to the solute-solvent and solvent-solvent interactions.
employed, is that the method results in smooth forces calculated at all distances. All the interactions of the system are calculated with respect to image systems in all directions until a converged sum is obtained. In this sense, its accuracy in calculation of the electrostatic contributions of the atom-atom interactions supersedes that of any method involving truncations. However, there are adjustable parameters controlling convergence of the sum and there is the possibility of long-ranged correlations introduced by the method. Thus, only when detailed analyses of multiple studies using the PME or other Ewald summation methods are available for scrutiny, will the possible advantages and disadvantages of its application to nucleic acid systems be known. Our preliminary results, based on global RMS deviations, reveal yet again better convergence of the structural and dynamical ensemble of results with respect to individual trajectories, as well as with respect to the starting configuration, associated with the inclusion of longer-ranged interactions. A detailed analysis of the structure and dynamics of the solvation of the tRNA fragment is now deemed profitable using extended 500 ps trajectories already conducted and is forthcoming. In the 16 Å truncation distance MD trajectories, the ensemble averaged structural properties were found to remain closer to the starting crystal structure than any of the individual trajectories of the set. A similar result has been observed in a few protein simulations in their crystallographic environment in which the unit cell average structures are found to have better RMS fits to the starting crystallographic structure than any of the individual monomers (3, 4, 39, 46). Figure 6 shows a stereoview of a superposition of the ensemble and time averaged structure of the six 100 ps PME anticodon MD simulations onto the starting crystal structure. The RMS difference between the two is 1.4 Å. The MD structure represented here reflects the computational equivalent to the ensemble and time averaged structures resulting from crystallography and NMR derived methods.

![Figure 6: Stereoview of the time averaged MD structure of the ensemble averaged trajectory from six 100 ps MD trajectories generated using the Particle Mesh Ewald (PME) method for calculation of electrostatic interactions (bold lines) superposed on the anticodon hairpin crystal structure (thin lines).](image-url)
RNA Structure from M. D. Dynamics Simulations

Study of the Thymine Loop Topology and Dynamics. The structure and dynamics of the thymine loop of tRNA^Met in the presence and absence of D-loop interactions was investigated in six 100 ps trajectories utilizing a 12-16 Å twin-range truncation scheme (33). Figure 8 reveals the time dependent and time averaged global RMS deviations from the starting crystal structure for the four thymine loop trajectories and the two thymine loop plus D-loop fragment trajectories (bold lines). Note that in all the trajectories, the steady climb of the RMSD values over the course of the trajectory, indicates a structurally non-equilibrated system. Four simulations of the thymine fragment alone resulted in final structures of 2.0 Å, 2.2 Å, 2.3 Å and 3.1 Å RMS deviation from the initial crystal structure. The two trajectories including the D-loop had final RMS deviation values of 2.7 Å and 3.3 Å.

![Graphs showing time dependent global RMS deviations and time averaged per-residue RMS deviation profiles.](image)

**Figure 7:** Time dependent global RMS deviations and time averaged per-residue RMS deviation profiles (averaged over the last 5 ps of each 100 ps trajectory) for each set of MD trajectories on the anticodon hairpin structure depicted here as a function of increased accuracy in the electrostatic model used (descending). The trajectories showing the lowest and highest RMS deviations from the starting crystal structure are in bold. (The 1-2 Å range for high resolution crystallographic structures is included for reference.)
Although the additional D-fragment interactions did not reduce the global RMS deviation value, separate RMS analysis of the stem and loop residues revealed that the RMS deviation contributions are partitioned unevenly with stem values about 3.0 Å (due to unwinding) and loop region values around 1.0 Å in the presence of the D-fragment, but 1.8 Å in its absence. The time averaged per residue RMS deviation profiles reveals high distortion in the stem end base pair for the thymine D-loop fragment simulations, however lower RMS deviation values in the loop region. Nearly the opposite case is found for the simulations of the thymine loop alone.

The presence of six tertiary hydrogen bonds and one stacking interaction which define the tRNA^apo crystal structure loop topology, were in general not well conserved in the four thymine loop trajectories. In two of the four, partial or total disruption of the T54-A58 trans-Hoogsteen base pair was seen to strongly affect the native conformation of the rest of the loop. Figure 9 illustrates two simulations of the thymine loop trajectories; TCY1, in which the T54-A58 trans-Hoogsteen base pair is well maintained, and TCY2 in which this same interaction is ruptured after 30 ps of MD. Figure 9 also displays the time dependent distances of the G56-C61/T54-A58 stacking interaction and the A57-A58 intercalation distance for the two simulations TCY1 and TCY2.

Interdigitation of a residue from the D-loop into the T-loop (position 18 between 57 and 58) and an additional Watson-Crick base pairing (between residues 19 and 56) significantly increases the number of possible hydrogen bond and stacking interactions able to stabilize the native T-loop conformation. To investigate the effect of these additional interactions on the loop structure and dynamics, two additional simulations including a two base fragment from the D-loop were run under the same protocol. Figure 10 shows the nature of the T- and D-loop interactions in the intact tRNA^apo crystal structure (34). The A58(O4')-G18(H22) interaction, 2.9 Å in the crystal structure, is measured at 2.9 Å averaged over the two thymine D-loop fragment simulations (individual trajectory average values of 2.6 Å and 3.3 Å).

Figure 8: Time dependent global RMS deviations and time averaged per-residue RMS deviation profiles (averaged over the last 5 ps of each 100 ps trajectory) for each set of MD trajectories on the thymine hairpin structure. The trajectories including the D loop fragment are in bold. (The 1-2 Å range for high resolution crystallographic structures is included for reference.)
The two G18-Ψ55(O4) hydrogen bonds are well maintained in the MD trajectories, with average values of 2.0 Å, however, in the crystal structure the G18(H1)-Ψ55(O4) interaction is measured at 3.0 Å. Thus the dynamics predicts more stabilization of this interaction than does the crystal structure. The same situation is found for the G19(O4')-A57(H2) bond, measured in the crystal structure at 4.3 Å and in the dynamics at 3.8 and 2.9 Å. Of the two possible A57(H2) interactions with the D-fragment backbone, only one seems to be able to be relatively satisfied at a time (A57(H2)-G18(O2'); XTAL(2.7 Å), A57(H2)-G19(O4'); XTAL(4.3 Å)). The two thymine loop D-fragment dynamics trajectories are found to display the two cases of preference of these two C-H...O interactions (A57(H2)-G18(O2'); TDF1(2.8 Å), TDF2(3.5 Å), A57(H2)-G19(O4'); TDF1(3.8 Å), TDF2(2.9 Å)).

Solvation percentages for phosphate groups were found to be in the range of 2.2-3.3 water molecules per phosphate oxygen, with an average of 2.7, in accord with the theoretical cones of hydration predicted from previous theoretical and MD studies (47). Multiple phosphate group water bridges with individual lifetime greater than 50% of the MD trajectory were consistently observed across the set of 6 trajectories, predominantly in the stem region but also in the loop region, and revealed a preference of bridges across pyrimidine bases over purine bases. Sugar, backbone and base hydrogens available for hydration were also analyzed for their solvation percentages and in general, the percentages calculated were commensurate with the topology of the loop. The 2' hydroxyl hydrogens were solvated 50-70%, except for residue A58 which makes a hydrogen bond with its HO2' hydrogen to the phosphate oxygen of C61. The C5' hydrogens ranged in percentage solvation from 20-75% depending on the local backbone conformation. Sugar proton hydrogens were variably solvated, with percentages around 60% for CH1', CH3' and CH4' positions yet only 0-10% for CH2' sites. Free amino hydrogens in the stem region

Figure 9: The two time dependent hydrogen bond distances of the trans-Hoogsteen T54-A58 interaction of the thymine loop (top), the G53-C61/T54-A58 stacking distance and A57-A58 intercalation distance (bottom), are illustrated for two simulations of the thymine loop MD trajectories, TCY1 and TCY2; preservation of the T54A58 hydrogen bond interactions (TCY1-thin line), loss of the T54-A58 hydrogen bond interactions (TCY2-thick line). Crystal structure reference lines are included for the stacking and intercalation distances.
are well hydrated (90%), free imino hydrogen are not well hydrated (<50%), except in the loop region (>50%), and the methyl hydrogens of the ribo-thymine at position 54 were all equally hydrated by 30%. Thus the localized average solvation properties of the system showed reasonable behavior even though the native tertiary conformation of the thymine hairpin was not entirely conserved.

PO₄(P)-NH₄⁺(N) and NH₄⁺(N)-NH₄⁺(N) atom radial distribution function analysis of the first and last 5 ps of the MD trajectories revealed artifactual counterion behavior not present in the anticodon studies. From Figure 11, the position of the peak situated at 6 Å in the P-N radial distribution plot at the beginning of the simulation has moved out to 11 Å by the end of the 100 ps trajectory, signalling a consistent diffusion of NH₄⁺ ions away from the RNA solute. Further, the growth of the first peak in the N-N radial distribution plot over the course of the simulation illustrated here indicates the formation of small clusters of two to three NH₄⁺ ions over the period of the trajectory. Possible explanations for this effect are currently being sought, however, since the MD protocol and system construction essentially did not differ from that of the anticodon hairpin simulations, where this effect is not observed, the possibility remains that this particular fragment is less stable in solution or may require divalent ion stabilization. Protocol improvements using the PME method are currently being investigated for this system.

**Conclusions**

The conclusions that can be drawn from our molecular dynamics studies of tRNA fragments are many fold. Three generalizations can be made concerning the model used to calculate the electrostatic interactions of the explicitly solvated and neutralized RNA fragments. First, short truncation distances of nonbonded interactions, even applied only to the solute-solvent and solvent-solvent interactions, produce measurable system and structural instabilities (31) compared to longer truncation distances (32) or Ewald summation methods. Second, the inclusion of long-range hydration forces, by extension of the truncation distance of calculated nonbonded interactions, causes both time averaged and dynamical stabilization
of the structural characteristics of the system (32). The inclusion of these longer ranged forces may well be necessary for the accurate description of solvated macromolecular systems, such as RNA fragments studied here, where the interplay of site specific hydration, long-range hydration forces and C-H...O hydrogen bonds, among others, may lend a significant contribution to the stabilization of the stronger tertiary interactions indicated by crystallographic structures. Third, as more accurate means of calculating the electrostatic contributions of the system are employed, the ensembles of structures from the MD trajectories are observed to converge and to remain closer to the vicinity of the starting crystallographic configuration (Figure 7).

Thus, regarding the critical evaluation of the structural integrity of the MD trajectories, preservation of the tertiary interactions as indicated by the crystallographic structure (and, possibly, the NMR reference structure), at least on the 100 ps timescale, should be demonstrated before a detailed interpretation of the dynamics of the system is attempted. However, global parameters, such as time dependent or time averaged RMS deviations cannot, by themselves, give unambiguous indication of the extent of structural stabilization achieved. Therefore, characterization of solvation properties and of counterion behavior of the system, among others, should also be made to detect any possible artifactual behavior. For RNA structures, which experimentally require divalent ions to assume the native folding conformation and/or catalytic activity (9), the inclusion of divalent ions into the theoretical MD model may be necessary to effect stabilization of the native conformation.

The many advantages of the multiple trajectory molecular dynamics approach has been amply demonstrated (31-33). This economizing approach, of generating multiple 100 ps trajectories, was first adopted in order to test systematically for minor protocol variations which would lead to stable extended MD trajectories. It subsequently revealed itself as a means of assessing the stability of the underlying protocol by examination of the spread of global time dependent and time averaged RMS deviations of the solute. When the spread of the global RMS deviations is sufficiently small, the independent trajectories are an adequate means of sampling statistically the local conformational space. Further, the length of the trajectories

Figure 11: The non-normalized P-N and N-N atom radial distribution functions calculated for the first 5 ps (bold line) and last 5 ps (thin line) of the MD trajectory for one of the six thymine loop MD trajectories.
also serves as a means of primarily sampling the conformational dynamics of the starting configuration, that is, before evolution of the time dependent RMS deviation indicates that the system has evolved to a local energy minimum further from the experimental reference. As was found for the six 12-16 Å twin-range simulations of the anticodon loop, the ensemble trajectory averaged properties might better reflect the experimental properties than any of the individual trajectories generated. This is possibly the computational equivalent to the ensemble and time averaging inherent in both NMR and crystallographic methods.

By these studies, we do not claim to be able to describe accurately all relevant structural and dynamical tertiary interactions of our systems based on those observed in the 100 ps trajectories reported here, nor strictly to have demonstrated that the system has stopped evolving structurally from its starting configuration. We have merely used the economy of 100 ps MD trajectories to explore the methodological protocol requirements needed to demonstrate stable MD trajectories of RNA systems and have extracted possibly useful structural information of the RNA fragments studied only to the extent warranted by a critical evaluation of the ensemble of MD trajectories. Further, we would point to the trend implied in Figure 7, that the increased accuracy of the theoretical model of the system used in the MD protocol reduces the amount of equilibration time for the system to reach a stable energy minimum and that this minimum is possibly very close to the experimental reference starting structure.

The most recent major advance in molecular dynamics applied to nucleic acid systems, since the routine inclusion of explicit solvent, has been the extension of MD trajectories to the nanosecond timescale. However, often the MD protocols of the nanosecond trajectory studies still limit the calculation of electrostatic interactions to short ranged contributions, not explicitly taking into account the long-range effects of solute-solute interactions and hydration forces. Further, such costly singular MD trajectories in general do not provide a clear means for evaluating the stability or accuracy of the results. Improvements in theoretical model need to proceed hand in hand with extension of the MD trajectory timescale and a clear assessment of protocol, system and structural stability should be demonstrated before the structure and dynamics of the system, as revealed on the nanosecond timescale, can be reliably interpreted. Future application of the molecular dynamics method to the study of RNA structure and dynamics will potentially yield useful information on the detailed local balance of tertiary interactions which maintain the rich structural complexity of RNA molecules, thereby building and strengthening the bridge between theoretical and experimental structural studies of RNA systems.

Acknowledgements

S.L.-M. is grateful to the French Government for providing a Chateaubriand fellowship and was supported by the C.I.E.S. to attend the 9th Conversation in Biomolecular Stereodynamics. P. A. is supported by a fellowship from ORGANOBIO (28, rue Saint Dominique, Paris, France) in the program CM2AO. The authors acknowledge also the IDRIS computing center which provided computer time and the authors are thankful to the CEE for providing funds through the Biotech contract BIO 2-CT93-0345 to E. W.

References and Footnotes

42. E. Engler, and G. Wipff, 1994, Université de Strasbourg, France.