RNA Solvation: A Molecular Dynamics Simulation Perspective

Pascal Auffinger
Eric Westhof
Institut de Biologie Moléculaire et Cellulaire du CNRS, Modélisations et Simulations des Acides Nucléiques, UPR 9002, 15 rue René Descartes 67084 Strasbourg Cedex, France

Received 21 February 2001; accepted 15 March 2001

Abstract: With the availability of accurate methods to treat the electrostatic long-range interactions, molecular dynamics simulations have resulted in refined dynamical models of the structure of the hydration shell around RNA motifs. The models reviewed here range from basic Watson–Crick to more specific noncanonical base pairs, from “simple” double helices to RNA molecules displaying more complex tertiary folds, and from DNA/RNA hybrid double helices to RNA hybrids formed with a chemically modified strand. © 2001 John Wiley & Sons, Inc. Biopoly (Nucleic Acid Sci) 56: 266–274, 2001

Keywords: RNA solvation; molecular dynamics; electrostatic long-range interactions; hydration shell

INTRODUCTION

Due to the continuous developments in experimental techniques, a large array of data describing the macroscopic and microscopic hydration features of biomolecular systems give shape to our current solute/solvent interaction model.\(^1\)\(^-\)\(^6\) Clearly, all the newly generated data need to be integrated in a coherent way. Molecular dynamics (MD) simulation techniques contribute to this process. Here, we will concentrate on the main contributions of MD simulations with explicit representations of the solvent to the study of the hydration of RNA molecules. An outline of several common methods used to characterize the structural and dynamical features of the hydration shell, followed by an account of results from MD simulations on RNA fragments describing water as well as ion binding characteristics, will be given. Additional information on nucleic acids,\(^5\)\(^-\)\(^12\) and more generally, on biopolymer\(^13\),\(^14\) hydration features, can be found in several recent articles and reviews.

METHODS

Structure: Characterization of Hydration Sites

Crystallography has unambiguously established that some water molecules occupy spatially well-defined hydration
sites in the vicinity of nucleic acid atoms.\textsuperscript{1,15–18} Unfortunately, even for high resolution structures, not all of these hydration sites can be observed. In order to understand completely the physicochemical properties of these systems and given the symbiotic solvent/nucleic acid relationship, it is essential to be able to determine with a good precision the location of these hydration sites. For that purpose, several techniques have been developed.

One widely used method is based on three-dimensional (3D) grids.\textsuperscript{19} The space around the average structure is divided into small volume elements (usually close to 0.1 Å\textsuperscript{3}). Then, structures extracted from the simulation at regular time intervals are fitted onto the average model and water molecules are binned into the above defined volume elements. Drawing the calculated densities at different contour levels allows to visualize the hydration sites. This method gives generally a good picture of the hydration shell of molecules that are not too big and not subject to markedly large motions. For flexible molecules, an obvious drawback of this approach is that, due to the fitting procedure, some regions like the extremities of long DNA duplexes are less well defined than the center regions. Consequently, their hydration patterns are similarly less well defined.\textsuperscript{19}

A more local approach has been developed in order to deal with this issue.\textsuperscript{20–22} This method consists in overlaying specific structural elements, such as Watson–Crick base pairs surrounded by water molecules belonging to their first hydration shell, in order to form “hydrated building blocks” similar to those derived from statistical analysis of x-ray diffraction data.\textsuperscript{23–25} These hydrated building blocks are Fourier transformed into pseudo-electron density maps that can be drawn at different contour levels in order to delineate the best defined hydration sites (Figure 1). At the base-pair step level, this method should be able to help in the visualization of extended hydration patterns such as the water spines observed in some DNA duplexes.

In parallel, radial distribution function approaches are very common and 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 integrate radial and angular coordinates, was developed.\textsuperscript{26} Alternatively, a computationally efficient method to describe the organization of water around biomolecules has been proposed.\textsuperscript{27,28} It is based on a statistical mechanical expression of the water-density distribution in terms of particle correlation functions.

**Dynamics: Determination of Residence Times**

The determination of the residence times of water molecules is not straightforward and is generally based on the estimation of the lifetime of specific solute-solvent hydrogen bonds.\textsuperscript{20–22,29,30} One possible definition considers that the residence lifetime for any water molecule in the vicinity of a hydrophilic atom equals the amount of time over which a hydrogen bond between both partners exists. Arbitrary time limits are sometimes introduced specifying that, in order to consider a water molecule as resident, the hydrogen bond in which it is engaged should not be broken for more than a few picoseconds (\(~\approx 1\) or \(~2\) ps). However, the use of time limits is probably not absolutely required since it has been observed that a water molecule that escapes from a given hydration site almost never comes back to the same site during a nanosecond MD simulation. Thus, the total time over which a water molecule forms a hydrogen bond with a given solute atom is a good approximation of its residence time. From these data, profiles giving the number of water molecules bound to a specific atom with respect to the lifetime of a given hydrogen bond can be drawn (Figure 2). Although these profiles may slightly change if the simulation time is extended,\textsuperscript{22} they are atom specific. Thus, even, if the values derived from these profiles have to be considered on a semiquantitative basis, especially for the water molecules with long residence times, the comparison between profiles calculated with comparable protocols for different atoms is very useful. Likewise, by recording the time over which water molecules establish hydrogen bonds with two or three neighboring atoms, one can estimate upper limits for water bridge lifetimes, such as the frequent OR\((n)\ldots\text{Ow} \ldots \text{OR}(n + 1)\) hydrogen-bond pattern observed in RNA.\textsuperscript{20,21,29,30}

Other methods rely on the time spent by a water molecule in a 5 Å sphere around a given solute atom\textsuperscript{31} or use coordination correlation function approaches.\textsuperscript{14,32,33} Worth mentioning is that none of the above methods can give statistically exact values for the residence times of the most strongly bound water molecules, since the length of a MD simulation should be many times longer than the longest calculated residence times.

**Force-Field Issues**

Despite real difficulties attached to the inclusion of polarization and charge transfer effects into current two-body force fields,\textsuperscript{34} recent improvements of empirical force fields, reviewed by Cheatham and Kollman,\textsuperscript{12} and of the treatment of long-range electrostatic interactions,\textsuperscript{35,36} have led to more accurate MD simulations. Yet, it has been proposed that simulations of DNA in explicit water are surprisingly insensitive to small perturbations in the force-field or local environment on current time scales.\textsuperscript{37} For DNA, the use of TIP3P\textsuperscript{38} or SPC/E\textsuperscript{39} water models was found to result in comparable hydration patterns. However, the water densities associated with the TIP3P model, which has the highest diffusion rate, appear more “blurred” than those calculated with the SPC/E model.\textsuperscript{39} Besides, a MD simulation of a protein in a crystal environment led to the conclusion that results obtained with the SPC/E model are in much better agreement with neutron-scattering experiment data than those collected with the TIP3P model.\textsuperscript{40} Hence, an estimation of the quality of a water model with respect to dynamical data, which are very scarce, has still to be achieved.
RESULTS

Watson–Crick Base Pairs

Watson–Crick pairs constitute the most basic unit encountered in RNA structures. As such, it is of paramount importance to gain a precise view of their hydration. Besides, data gathered for Watson–Crick pairs are the basis for a sound comparison in the evaluation of modifications of the hydration shell associated with insertions of non-Watson–Crick pairs and modified nucleotides into RNA structures.

A systematic approach for the study of the hydration shells of Watson–Crick pairs has been proposed.\textsuperscript{20,21} This method consists in performing MD simulations of regular RNA duplexes, i.e., r(CpG)\textsubscript{12} and r(UpA)\textsubscript{12}. In order to get statistically accurate data, the 12 central base pairs of each duplex were overlaid and the water molecules surrounding them were counted. On the average, 21.9 and 21.0 solvent molecules (water and ions) were found to enter into contact with the r(G\textsubscript{A}C) and the r(A–U) pairs, respectively. Pseudoelectron density maps were then calculated from the cloud of water molecules surrounding each type of base pairs. High-level contours of the density maps allowed us to delineate 22 well-defined hydration sites for the r(G\textsubscript{A}C) and r(A–U) pairs (Figure 1). The positions of the hydration sites around the bases are in excellent agreement with those derived from a statistical analysis of high-resolution crystallographic data.\textsuperscript{24} Note that both in the crystal and in solution, these sites are better defined for the r(G\textsubscript{A}C) than for the r(A–U) pairs. As expected, the hydration patterns for the r(G\textsubscript{A}C) and r(A–U) pairs are different.

While three water molecules are in contact with the deep groove atoms, three (G\textsubscript{A}C) and two (A–U) water molecules are located in their shallow groove. The hydration patterns around the backbone are essentially sequence independent. Well-defined hydration cones around the anionic oxygen atoms and the hydroxyl groups could also be observed. Water molecules with long residence times (∼700–800 ps) are attached to the OR atoms of the phosphate groups (Figure 2). Other water molecules with residence times in the 500 ps range are mainly located in the vicinity of the (G/A)N7, (G)O6, (A)N6, and (U)O4 deep groove atoms. The residence times of water molecules bound to the shallow groove O2′-H groups are especially short (∼100 ps; Figure 2). They display a water-like behavior characterized by a continued formation and breakage of solute/solvent hydrogen bonds. Thus, although hydroxyl groups interact preferentially with water and tend to avoid the formation of intramolecular O2′-H...O4′(n + 1) hydrogen bonds, they do not favor long-lived shallow groove hydration patterns.\textsuperscript{41} A comparison with hydration patterns derived for the d(G\textsubscript{A}C) and d(A–T) pairs extracted from B-DNA duplexes reveals significant differences that will not be discussed here.\textsuperscript{20,21}

Non-Watson–Crick Base Pairs

The non-Watson–Crick pairs play major structural and functional roles in RNA architectures\textsuperscript{42,43} and are associated with specific hydration patterns.\textsuperscript{16,24}
**G-U Pairs.** In order to evaluate the structural effects associated with a conserved G3-U70 base-pair present in the tRNA<sup>Ala</sup> acceptor arm, the hydration of several variants of this motif has been investigated on the basis of seven 2.5 ns of MD simulations. The authors showed that the G-U pair, as found in crystals, induces local deviations from A-form geometry, i.e., the wild-type helix is underwound at the base-pair step above the G-U pair and overwound at the base-pair step below the G-U pair in each case by about 7–9 degrees compared to an A-form helix. They discussed the presence of tightly bound water molecules on the shallow groove side of the G-U pair that display long residence times (∼500 ps), and correlate with water molecules observed by crystallography and by multiple MD simulations for a G-U pair embedded in the tRNA<sup>Asp</sup> anticodon hairpin. These earlier simulations have also revealed that water molecules occupying the shallow groove hydration site in contact with the (U)O2′, (U)O2, and (G)H21 atoms display long residence times (∼450 ps). No long-lived hydrogen bonds involving the (U)O2′ atom were observed, but long-lived hydrogen bonds with the (U)O2 and the (G)N2—H21 atoms were detected. Remarkably, for a U3-G70 pair replacing the G3-U70 pair, a significantly smaller (in size and magnitude) water density peak is observed compared to that found for the wild-type base pair.

**GdU, I-U, and 2AA·IsoC Pairs.** GdU, I-U, and 2AA·IsoC base pairs were inserted in the tRNA<sup>Ala</sup> acceptor stem described above in order to estimate deformations introduced in the hydration shell of the wild-type structure. Removal of the (U)O2′—H group (GdU) or of the (G)NH<sub>2</sub> group (I-U) does not alter significantly the shallow groove water density peak observed for the G-U pair in line with x-ray data showing that a water molecule is systematically connected to the O2 atoms of G-T and I-T pairs. For the 2AA·IsoC pair as well, a region of high water density is seen at the same location. Thus, the occurrence of this hydration site is probably more related to the notch created by the wobble geometry of the base pairs than to the presence and type of hydrophilic groups lining this cavity.

The residence times of water molecules connected to all these sites are different. The longest calculated residence times follow the order G-U (537 ps) > 2AA·IsoC (324 ps) > U-G (224 ps) > GdU (∼180 ps) > I-U (∼140 ps) > G=C (∼50 ps) in line with other data collected from MD simulations on G-U<sup>29</sup> and G=C<sup>20</sup> pairs. Thus, water densities alone are not a sufficient criterion to estimate the strength of a binding site. Residence times have to be taken into account as well.

**C-U Pairs.** MD simulations have been conducted on a r(GGACUUCGGUCC)<sub>2</sub> duplex that contains two G-U pairs framing two water-mediated C-U pairs. It has been observed that the U-C pairs switch during the MD run between two types of conformations involving an interaction between the (C)NH<sub>2</sub> group and either the (U)O2 or the (U)O4 atoms. For the conformations involving a (C)NH<sub>2</sub>…O2(U) interaction, water molecules displaying long residence times (∼300 ps; in one occurrence a 1 ns residence time is observed) are located in the shallow groove. For the conformation involving a (C)NH<sub>2</sub>…O4(U) interac-
tion, stable water molecules (~300 ps) are located in the deep groove of the duplex. These conformations for the C · U pairs are different from the bifurcated hydrogen-bonded C · U pair observed in MD simulations of the rRNAAsp anticodon loop that involve a water-mediated base-to-backbone interaction.46

**GA Pairs.** A transition between a conformation of a G·A pair involving two water molecules and a water-mediated conformation has been observed in a MD simulation of a GCAA tetraloop.47 This water-mediated conformation involves most likely a long-lived water molecule.

**Modified Nucleotides**

Naturally modified nucleotides are more than anecdotical features of nucleic acids. They are present in RNAs from all three kingdoms (archaea, bacteria, eukaryote) where they occupy strategic positions.48 Besides the dU, I, 2AA, and isoC nucleotides mentioned before, some study dealt with pseudouridines (Ψ) and 2′O-methylations. The MD simulations described below emphasize important structural modifications of the hydration shell related to the presence of modified nucleotides embedded into RNA structures.

**Pseudouridines (Ψ).** Among other roles, pseudouridines, which display a C1′—C sugar–base linkage with an N1—H group replacing the C5—H group present in uridines, are involved in the stabilization of the structure of the hydration shell of specific structural motifs.49,50 It has been observed in six multiple MD simulations of the yeast rRNAAsp anticodon hairpin that the water molecules that display the longest residence times form very stable hydrogen-bond contacts with the Ψ base and the nucleic acid backbone.29,49 These water molecules establish three hydrogen bonds, i.e., an acceptor hydrogen bond with the (Ψ)N1—H group and two donor hydrogen bonds with anionic oxygen atoms of adjacent phosphate groups. Thus, a strong (Ψ)N1—H...Ow bond replaces a (U)C5—H...Ow interaction and leads to the stabilization of a water molecule with very long residence times.51 This water molecule helps to lock the base in a specific conformation with respect to its backbone, and thus is part of a “nucleotide/water” complex (Figure 3).

**2′O-Methylations.** Methylation of the 2′-hydroxyl groups are very common in RNA molecules. Based on x-ray and NMR structures, MD simulations of the r(CGCGCG)₆ and 2′-O-Me(CGCGCG)₆ duplexes were carried on.62 Interestingly, compared to the nonmodified structure, a more ordered hydration shell around the O2′ atoms of the methylated duplex was observed, and specific shallow groove hydration sites occupied by long lived (>1 ns) water molecules were detected.82

**RNA Duplexes**

MD simulations on regular RNA duplexes formed by Watson–Crick pairs lead to consistent views.19–21,52 RNA duplexes are found to be more rigid than B-DNA duplexes of similar sequence. Interestingly, while sequence dependent long-lived hydration patterns (~1 ns) are found in the minor groove of B-DNA linking the O4′ with the N3 or O2 atoms, no such hydration patterns are found for RNA.21 Instead, long-lived hydration patterns (~800 ps) are located in the deep groove and involve nonsequence-dependent water bridges between adjacent phosphate groups. Indeed, there is clearly a relation between the shape and dynamics of the hydration patterns and the flexibility of the nucleic acid structures. Nevertheless, this relationship is intricate and therefore difficult to quantify.

**Hybrid Duplexes and Chemically Modified Backbones**

**DNA/RNA Hybrids.** Hydration patterns around DNA/RNA hybrids were observed to be less defined
than around an equivalent B-DNA or an RNA duplex. In this study, it is shown that the more flexible hybrid structure has less defined hydration patterns.

**Phosphoramidate Analogs.** Radial distribution functions calculated for a 10 base-pair long DNA/RNA hybrid and two 3′pnDNA–RNA and 5′pnDNA–RNA phosphoramidate analogs were used to propose an explanation for the observed differential stability of these duplexes. Experimentally, it is known that 3′pnDNA–RNA are more stable than corresponding phosphodiester, while 5′pnDNA–RNA do not form duplexes at all.

**PNA/RNA.** Hydration maps were calculated for PNA/DNA and PNA/RNA duplexes by using a grid method and were found to be completely different for both molecules. Interestingly, the PNA/DNA structure exhibits a spine of hydration in the minor groove similar to that found for DNA, while the PNA/RNA helix shows, like RNA duplexes, a better hydration of the deep groove.

**HNA/RNA.** Experimentally, it is known that molecular association between HNA and RNA is more stable than between HNA and DNA and between natural nucleic acids (dsDNA, dsRNA, DNA/RNA). A thorough description of the hydration of hexitol nucleic acid duplexes (HNA/DNA and HNA/RNA) led to the observation that a better hydration of the HNA/RNA vs the HNA/DNA duplex correlates with a reduced flexibility.

**2′-Substituted Analogs.** Nanosecond simulations of a MOE(CCAACGGTTG)–r(CCAACGUGUGG) duplex were MOE stands for 2′-sugar substituted O-(2-methoxyethyl) have been reported. A short description of the hydration of the duplex is given that emphasizes the occurrence of a hydration site between the O3′ oxygen atom of the backbone and the oxygen atom embedded in the MOE chain.

**Complex RNA Structures**

RNA hairpins, which are of higher complexity than double-stranded structures, are more sensitive to the treatment of electrostatic interactions than DNA duplexes. Therefore, the first successful long (200–1000 ps) MD simulations of RNA could only be achieved when accurate treatment of the long-range electrostatic interactions became available. The first published simulation described a water insertion event at the level of the G · A base pair, which closes a GNRA tetraloop. Multiple molecular dynamics simulations conducted on the yeast tRNA\(^{A\text{sp}}\) anticodon hairpin revealed the occurrence of well-defined hydration patterns associated with water molecules displaying long residence times (≈500 ps). The occurrence of long-lived hydration patterns in the loop was proposed to be associated with the presence of modified nucleotides (Ψ). Thus, among other roles, modified nucleotides stiffen the hairpin structure through the stabilization of its hydration shell.

Simulations of more complex RNA systems are still rare and even rarer are descriptions of their hydration shell. Some aspects of the hydration of the flavin–mononucleotide/RNA aptamer complex were described. Several water mediated protein/RNA contacts observed in a 0.6 ns MD simulation of the complex between the human U1A protein and the hairpin II of the U1 small nuclear RNA have been reported. Water mediated contacts between amino-glycosides and the hammerhead ribozyme were also described, and similar contacts were observed in the crystal structure between paromomycin and a RNA oligomer.

**Ion Binding Features**

The literature describing ion binding events to DNA structures both from an experimental and theoretical point of view is growing. For RNA, experimental evidence emphasize the structural involvement of monovalent ions that, therefore, should no longer exclusively be considered as neutral additives. Binding sites for NH\(_4^+\) cations have been characterized in a RNA loop, based on multiple molecular dynamics simulations of the yeast tRNA\(^{A\text{sp}}\) anticodon hairpin. These ions utilize simultaneously about three of their N−H bonds to interact with electronegative atoms of the RNA. Sodium ion binding events to the major groove of a RNA duplex have been described. Potassium ions are found to bind specifically to sites located in the deep groove of r(GpC) and r(ApU) steps but do not penetrate the deep groove of r(CpG) and r(UpA) steps mainly due to steric and electrostatic repulsion constrains emphasizing expected sequence specific effects (Figure 4).

Given the rapid exchange rates of water molecules bound to monovalent ions, such ions represent interesting probes for locating possible binding sites of electropositive chemical groups.

The question of the divalent ions is more difficult to address since the dehydration steps of these ions (≈10 µs for Mg\(^{2+}\)) cannot be simulated on the currently accessible time scales. However, data have been obtained for hexahydrated Mg\(^{2+}\) ions bound to DNA, and Brownian dynamics methods have been
used to locate Mg$^{2+}$ binding sites around static RNA structures.\textsuperscript{75}

**Perspectives**

The field of MD simulations of biological macromolecules is now becoming mature since reliable simulations can be generated on nanosecond time scales. This assertion is reasonable at least for systems around equilibrium or located in conformational “attractors.” As an outcome, the structural hydration characteristics are rather well reproduced. The validity of the derived lifetimes for the various hydration sites is on a slightly less secure basis and has to be considered on a semiquantitative basis. The case of charged ions, especially divalent (Mg$^{2+}$) ions, would still be worse. Will MD simulations ever bridge the gap between macroscopic thermodynamical approaches like the nonlinear Poisson–Boltzmann model\textsuperscript{75,76} and the highly precise and localized crystal structures of complex tRNAs?\textsuperscript{77,78} On one side, it is shown that electrostatics is sufficient for associating Mg$^{2+}$ ions with RNA molecules without involving precise hydrogen-bonding or coordination schemes.\textsuperscript{76} On the other hand, x-ray structures reveal a rich array of contacts between ions, waters, and RNA atoms positioned precisely in space by the three-dimensional (3D) fold. Indeed, the crystal clear perspective of binding sites may induce an overemphasis on the importance of specifically bound water and ions in stabilizing native tertiary folds of structured RNAs in solution. This view of static structures, as derived from x-ray experiments, is mainly “enthalpic” (coordination, hydrogen bonds, van der Waals contacts, . . . ). Yet entropic contributions should not be neglected. For example, in the folding process itself, the entropic contribution of solvent release (water and ions), although only indirectly observed, is primordial. MD simulations, by distinguishing between strongly bound and more labile solvent molecules are central to the evaluation process of the enthalpic/entropic balance and compensation.

**REFERENCES**