Summary
Following several methodological developments, molecular dynamics simulations are now able to reproduce essential features of the solvation shell of biological molecules deduced from X-ray crystallography. Here we how molecular dynamics simulations can complement experimental data by providing clues about the position and orientation of the mobile hydrogen atoms found in RNA systems, namely those belonging to the 2'-hydroxyl groups and to the water molecules.

Introduction
Hydrogen is the lightest element in the periodic classification. Nonetheless, about half of the atoms in proteins as well as in nucleic acids are hydrogen atoms. These are involved in the most important non-covalent interaction structuring biomolecular systems, namely the regular N/O-H···N/O hydrogen bonds\(^1\). The task of experimentally deriving hydrogen atom positions relative to neighboring atoms is therefore of particular importance for our understanding of biomolecular systems. Indeed, much effort is devoted to the development of such techniques. These include mainly neutron diffraction\(^2\)\(^-\)\(^5\) and high-resolution X-ray crystallographic\(^6\) studies through which many significant results have been recently obtained for proteins. However, for nucleic acids experimental data are scarcely available. This is essentially due to the difficulties associated with the growth of the large (> 1 mm\(^3\)) well ordered and impurity free crystals actually needed for neutron diffraction studies\(^7\). Yet, with a few notable exceptions, it is possible to determine unambiguously the position of almost all nucleic acid hydrogen atoms of these biomolecules by using straightforward stereochemical rules. One typical exceptions is related to the determination of the position of the hydrogen atom of the 2'-OH groups in RNA. These hydrogen atoms, as well as those from the methyl group hydrogens of the thymines in DNA, possess a single internal rotational degree of freedom. Another exception is linked to the protonation states of some important nucleic acid residues that cannot be directly observed but, in the best cases, only tentatively inferred from the chemical environment in which they are embedded. Besides, nucleic acids often interact with a large variety of natural and synthetic ligands involving chemical groups (-NH\(_2\), -OH, -SH, ···) whose protonation states are generally unknown\(^8\). Among these interacting molecules, one of the smallest but most widespread ligands is
Water. Water molecules have two hydrogen atoms that could almost never be directly observed in crystal structures of nucleic acids. Indeed, it is well appreciated that water plays a very important and subtle structural role through the formation of important hydrogen bond networks. Thus, a precise and complete knowledge of the position and dynamics of all the atoms, heavy and light, constituting solvent molecules is of great importance.

In this perspective, molecular dynamics (MD) simulations can be very useful, as they possess the ability to explicitly model the dynamical behavior of all the atoms constituting a biomolecular system including the hydrogen atoms that can rarely be observed through experimental methods. In this review, some results related to the determination of the position of important hydrogen atoms in RNA molecules obtained through MD simulation techniques will be detailed.

1. Molecular dynamics simulation techniques

Molecular dynamics (MD) simulations represent a technique that allows calculating the “possible” evolution of a physical, chemical or biochemical system over a given period of time starting from an empirical description of the potential energy of the system. The principles of MD reside in the numerical integration of the Newtonian equations of motion. The resulting force acting on a specific atom is calculated iteratively by taking the derivatives of the potential energy function with respect to its position. The potential energy function contains several terms that account for: (i) covalent bond stretching; (ii) bond angle bending; (iii) harmonic dihedral angle bending; and non-bonded interactions including (iv) van der Waals and (v) Coulomb terms. A common expression of this potential energy function used in the successive AMBER molecular dynamics simulations packages is given by:

\[ E_{\text{total}} = \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{1}{2} V_n \left[ 1 + \cos \left( \eta \frac{\Phi - \Phi_{eq}}{\Phi_{eq}} \right) \right] + \sum_{\text{atoms}} \sum_{ij} \left( \frac{A_{ij}}{R_{ij}^6} - \frac{B_{ij}}{R_{ij}^{12}} \right) + \sum_{\text{atoms}} \sum_{ij} \left( \frac{C_{ij}^1}{R_{ij}^2} - \frac{C_{ij}^2}{R_{ij}^6} \right) \]

The different parameters used by the empirical function are generally obtained from experimental and quantum mechanical studies. Other MD packages that use slightly different potential energy functions and force fields such as CHARMM or GROMOS are also available. It can be noted that in the early versions of these force fields, a special “10-12” term that is now obsolete was specifically designed for hydrogen bonds.

For a system of several thousands of atoms such as a solvated tRNA molecule, a tremendous number of force evaluations has to be performed at each time step. With modern computational means it is possible to generate accurate MD trajectories of complex nucleic acid systems including a complete representation of the environment, e.g. water molecules but also counter-ions (Na⁺, K⁺, Mg²⁺, ...) and co-ions (CT⁻, SO₄²⁻, ...). These simulations currently reach the nanosecond time scale and allow estimating, among other things, the stability of important tertiary interactions as well as some of the hydration characteristics of the investigated systems.

2. RESULTS

One of the major advantages of MD simulations is the ability of the method to follow the position of each single atom that is part of the investigated system as a function of time. This allows describing the dynamical behavior of atoms (hydrogen atoms, ...) or molecules (water, ions, ...) that are often observable only with difficulties by experimental techniques.
2-1. The hydrogen atom of the RNA ribose 2'-hydroxyl group

In the nucleic acid realm, the ribose 2'-OH group is of particular importance since it represents the main chemical discriminant between RNA and DNA molecules. Determining the orientation of this hydroxyl group associated with the only RNA hydrogen atom to possess one internal degree of freedom (with the exception of some hydrogen atoms of modified nucleotides) is therefore essential for our understanding of: (i) the conformational and folding features of RNA molecules and (ii) how these molecules interact with their environment. Unfortunately, experimental methods are unable to locate unambiguously the positions of these mobile hydrogen atoms\(^{22,23}\). Analysis of MD simulations conducted on tRNA fragments have revealed that only three orientations were accessible to the 2'-OH group for sugars in the most frequent \(\text{C}3'\)-endo conformation \(^{24}\). These three domains are: (i) the O3' and (ii) the O4' domains where the O2'-H orientation is stabilized by favorable sugar-backbone or intra-ribose electrostatics interactions, and (iii) the “Base” domain where the hydroxyl group points towards the N3(R) or O2(Y) atoms of the attached base (R stands for purines and Y for pyrimidines; Fig. 1). These orientations are

Figure 1  Stereochemical rules for the orientation of the 2'-OH bond in a ribose ring adopting the typical RNA \(\text{C}3'\)-endo sugar pucker \(^{24}\). A: Conformational wheel outlining the three favored O3', O4' and “Base” domains (marked by black arrows) as well as the three forbidden H1', H2' and H4' domains (marked by dotted arrows). The C2'-O2' axis is perpendicular to the plane of the page. B: Representation of a fragment of an RNA backbone in the most frequent conformation adopted in helical regions and characterized by a \(\text{C}3'\)-endo sugar pucker. The C2'-O2' axis is again perpendicular to the plane of the page. In this conformation the occurrence of a C2'-H(n) \(\cdots\) O4'(n+1) rather than a O2'-H(n) \(\cdots\) O4'(n+1) hydrogen bond is inferred from MD simulations. C and D: View of the hydration of the 2'-OH group\(^{24}\). The three allowed orientations of the 2'-OH groups are displayed in black while the three forbidden orientations intercalate between the former ones. The dots surrounding the 2'-OH groups represent three well-defined clusters of water molecules.
delimited by three “forbidden” orientations involving repulsive steric and electrostatic interactions with the H1', H2', H4' ribose, and H5'1 backbone hydrogen atoms. Moreover the energetically penalized orientations where the H2'-C2'-O2'-H angle is close to zero (eclipsed orientation) are also avoided. Hence, three conformationally favored domains for the orientation of the ribose 2'-OH groups could be detected. These findings have allowed the creation of a conformational map for the 2'-OH groups (Fig. 1) similar in spirit to Ramachandran maps and complementing some conformational maps obtained for nucleotides by Sundaralingam and coworkers. It can be noted that the “Base” orientation is in agreement with NMR data subsequently obtained for an RNA duplex.

One of the most surprising results derived from these simulations is that, no intra strand O2'-H(n) O4'(n+1) hydrogen bond was formed. This result contrasts with what has been suggested, based on the proximity of the O2'(n) and O4'(n+1) atoms in crystal structures, in early studies and textbooks. Our simulations, that explicitly account for solvation effects, indicate that the 2'-OH groups in helical regions prefer to form hydrogen bonds with “stronger” acceptor atoms such as the oxygen atoms of water molecules rather than with the “weaker” O4'(n+1) acceptor atoms, whereas the hydrogen bond acceptor potential of the O4'(n+1) oxygen atoms is used to form “weaker” intra strand C2'-H(n) ...O4'(n+1) hydrogen bonds as suggested by an analysis of crystallographic structures of RNA and A-DNA helices.

Following these observations, the structure of the hydration shell of the 2'-hydroxyl groups has been investigated through MD simulations. It was found that water molecules form three well defined clusters around the 2'-OH groups (Fig. 1) as also observed in crystal structures. Indeed, these water clusters are found in the sectors delimiting the conformationally allowed orientation of the 2'-OH group. Yet, the individual water molecules comprising these clusters are not long lived and form labile hydrogen bond patterns in agreement with the water like behavior of the 2'-OH group described above. Thus, rotation of the 2'-OH group does not involve the formation of new hydration sites but results in a dynamical hydration that involves rapid exchange of water molecules and constant rearrangement of hydrogen-bonding networks.

2-2. How to infer protonation states of RNA nucleotides from MD simulations

In principle the protonation states of chemical groups can only be derived from experimental data. Yet, when no direct experimental evidence is available, MD techniques can be used to solve this issue by performing simulations of protonated and unprotonated states. In practice, given the length of such studies, this can only be done for bases that are suspected to be involved in altered hydrogen bonding patterns. For example, in a MD study of an RNA pseudoknot, a rather stable pattern of hydrogen bonds was observed when a specific cytosine was protonated, while considerable local rearrangements occurred when this same base was left neutral, confirming the N3-protonation of the involved residue inferred from the X-ray structure. Large local conformational changes observed in MD simulations can, thus, suggest an incorrect protonation state of the system. In the same line, multiple simulations of an RNA/drug complex could help to infer the most appropriate protonation state of the ligand.

2-3. Mapping the hydration sites of RNA base pairs

In order to precise the structure of the hydration shell around Watson-Crick and non-Watson-Crick base pairs which constitute the most basic unit encountered in nucleic acid structures, a statistical analysis of high-resolution crystallographic data was performed. The purpose of this analysis was to provide a sound experimental basis against which results from MD simulations could be confronted.

Among all the features derived from this study, eight well-defined hydration sites could be located around the G=C pair: five in the deep and three in the shallow groove (Fig. 2). The hydration around the A-U pair is more diffuse probably as a result of the loss of an
inter base hydrogen bond and because the shallow groove is delimited by three instead of four hydrophilic groups for the G=C pair leading to only two shallow groove hydration sites for the A-U pair. Given the small number of high-resolution structures available at the time this study was performed, no attempt was made to derive hydration patterns around the backbone atoms.

MD simulations conducted on the r(CpG)$_{12}$ and r(ApU)$_{12}$ duplexes were able to reproduce most of these hydration features. The calculated pseudo-electron densities, obtained by analyzing the data from the simulations with crystallographic tools coming from the CCP4 (see references 31,34,35), show a very good agreement with the experimental data (Fig. 2). The experimental and calculated deep and shallow groove hydration patterns are comparable and a more diffuse hydration pattern around the A-U with respect to the G=C pair is seen. The deep groove hydration spot located in the vicinity of the C5 H group points to the formation of a possible C-H ··· Ow hydrogen bond as inferred from early MD simulations 36. In the shallow groove a string of two (A-U) or three (G=C) water molecules connect the two O2'-OH groups. As in DNA crystals 37,38, the hydration sites close to the phosphate groups are well defined and lead to the formation of hydration cones (it was much easier to extract backbone hydration patterns from the MD simulations than from the crystallographic data given the limited number of high-resolution structures available). Hydration cones are also observed around the 2'-OH hydroxyl groups (see Fig. 1C). In addition, it can be noted that MD simulations were able to successfully characterize the ionic atmosphere around Watson-Crick pairs and locate sequence specific ion binding sites in the deep groove of RNA duplexes 33,35. Yet, the main outcome of these studies resides probably in the demonstration, through a comparison between experimental and calculated data, that current MD simulations are now able to reproduce with precision the hydration patterns of regular nucleic acid fragments and can thus be used to deduce other subtler hydration features like the structure of the hydration patterns around RNA systems containing modified nucleotides 21,39 or temperature induced hydration shell structural changes 40.

2-4. Where are the hydrogen atoms of water molecules located?

Evidently, a description of hydration features in biomolecules is not complete if no information related to the position of the hydrogen atoms is given. It is in fact these atoms
that cement the system through the formation of hydrogen bonds. Figure 3 (top) shows an overlay of structures extracted from an MD simulation of the r(CpG)_{12} duplex. It is quite obvious that such a view illustrating the position of all the atoms constituting the water molecules is not very practical although it already indicates that surrounding atoms determine the orientations of first shell water molecules. In order to get a more comprehensive picture of the hydration characteristics of RNA base pairs, we used crystallographic tools to calculate “nuclear” density maps from the simulations (unpublished results; Fig. 3 bottom). In order to do this, we used the SFALL program from the CCP4 library (http://www.dl.ac.uk/CCP/CCP4) and a procedure similar to that described in a previous study for calculating the pseudo-electron density maps shown in Figure 2. The only difference was that the nuclear density maps were obtained by using the appropriate neutron formfactors for the oxygen, hydrogen (H) or eventually deuterium (D) atoms in the SFALL program. Note that maps for natural and deuterated water molecules can be simply obtained by renaming H by D in the coordinate files derived from the MD simulations. This simple procedure does not account for subtle differences associated with solvation of the solute by either H$_2$O or D$_2$O. For that, simulations taking into account explicitly deuterated water molecules should be performed. Nevertheless and in a first approximation, the calculated nuclear density maps are in principle comparable to those derived from neutron diffraction studies.

In Figure 3 (bottom), we show the nuclear densities associated with the non-deuterated water molecules surrounding a G=C pair derived from a MD simulation of a r(CpG)_{12} duplex. For this representation, we took advantage of the fact that hydrogen has a neutron scattering length that is opposite in sign to that of the bound oxygen atom. Thus, it is straightforward to separate in the maps the densities originating from the hydrogen and the oxygen atoms. The calculated map shows several interesting characteristics. In short, the densities of the hydrogen atoms forming hydrogen bonds with the N7(G), O6(G), O4(C) deep groove as well as with the N3(G), N2(G) and O2(C) shallow groove atoms are, as expected, clearly seen. Yet, the most interesting feature of these maps resides in the hydrogen atom densities sandwiched between those associated with the oxygen atoms of the water molecules. Clearly, in the shallow groove, the calculated densities indicate that the water molecules facing the N2(G) and O2(C) atoms are hydrogen bonded. Thus, MD simulations can provide a detailed view, not only of the position of the solvent molecules, but also of the hydrogen bond networks involved in their hydration. A more complete description of these maps will be given in a forthcoming communication.

Conclusions

Structural biology has led to the design of efficient methods for characterizing the position of the heavy atoms constituting biomolecules and it is
now of interest to pursue the design of methods that allow determination of the position of the lightest elements, namely the hydrogen atoms. At present, no experimental method allows the determination of the hydrogen atom positions in nucleic acid systems with ease and precision. It is our hope that we could convey the idea that simulation techniques can now provide useful hints about the location and orientation of solvent molecules and, thus, usefully complement experimental data. In the future, it is also our hope that experimental and simulation techniques will be used jointly in order to shed some light on parts of RNA structures whose conformational and dynamical features are still not well understood.

Acknowledgements

We would like to thank Eric Westhof for its continuing support and encouragements. A. C. V. acknowledges support from the European contract HPRN-CT2002-00190 (CAR-BONA).

References

10) van Gunsteren WF & Berendsen HJC: Angew Chem Int Ed 29, 992-1023, 1990
Curriculum vitae

AUFFINGER, Pascal
Research Scientist, UPR9002, CNRS, Institut de Biolologie Moléculaire et Cellulaire
1961 : Born in Strasbourg, France
1991 : PhD, University of Strasbourg, France
1992 : Postdoctoral Fellow, University of Wesleyan, CT, USA
1994 : Postdoctoral Fellow, University of Strasbourg, France
1997 : Research Scientist, University of Strasbourg, France

VAIANA, Andrea C.
Postdoctoral Fellow, UPR9002, CNRS, Institut de Biolologie Moléculaire et Cellulaire
1967 : Born in USA
1999 : Masters in Physics, University of Palermo, Italy
2004 : PhD, University of Heidelberg, Germany
2004 : Postdoctoral Fellow, University of Strasbourg, France