Structural and Dynamical Characterization of Nucleic Acid Water and Ion Binding Sites

Pascal Auffinger, Benoit Masquida, and Eric Westhof

Institut de Biologie Moléculaire et Cellulaire du CNRS Modélisations et Simulations des Acides Nucléiques, UPR 9002 15 rue René Descartes, F-67084 Strasbourg Cedex, France
{P.Auffinger, B.Masquida, E.Westhof}@ibmc.u-strasbg.fr
http://www-ibmc.u-strasbg.fr/upr9002/westhof

Abstract. Recent methodological developments led to longer and more accurate molecular dynamics (MD) simulations. In parallel, methods have been designed with the purpose of characterizing water and ion binding features. Here, we give an outline of some of the methods we used in order to extract structural and dynamical information concerning the first water and ion coordination shell, from MD simulations conducted on RNA and DNA structures. Coordinates for the water and ion binding sites located in the first coordination shell of r(G=C), d(G=C), r(A-U), and d(A-T) base-pairs are provided, along with calculated "pseudo" thermal factors.

1 Introduction

Most, if not all, of the properties of nucleic acids are linked to the presence of a well defined hydration shell surrounding each nucleotide. Consequently, a large number of experimental and theoretical studies have been devoted to the characterization of its structural and dynamical properties [1-3]. Yet, this is not a straightforward task. Many available experimental data are incomplete. For example, it is rare that all the water molecules surrounding a nucleotide embedded in a three dimensional structure can be detected by crystallography. By NMR, generally only the water molecules located in the vicinity of non-exchangeable protons and with long residence times can be observed [4-6]. By theoretical methods, two types of limits have been encountered. The very accurate quantum chemical calculations can only deal with small molecules surrounded by clusters of water molecules and environmental effects are often neglected. On the other hand, molecular dynamics simulations are hampered by the approximations inherent to empirical force-fields such as difficulties in taking explicitly into account polarization and charge transfer effects [7,8]. Nevertheless, at the current level of understanding and for RNA molecules, it has been shown that experimentally determined hydration sites [9] are in good agreement with those calculated by MD methods [10,11]. Similar results for DNA have been reviewed in [12].

Here, we will describe in detail the methodology we used to characterize the hydration shell as well as the ion binding features around RNA and DNA

Watson-Crick base-pairs from nanosecond MD simulations. Naturally, these techniques are applicable to any type of chemical or biochemical structure.

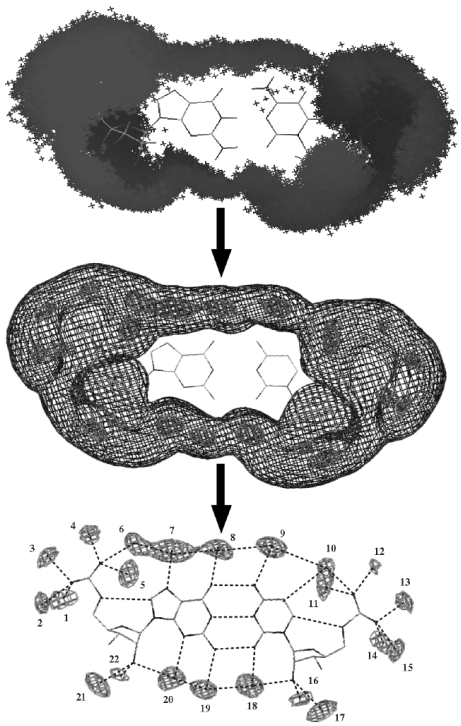

Fig. 1. Schematic representation of the process leading to the characterization of hydration sites around a r(C=G) base-pair [10]. (Top) The cloud of water molecules located at less than 3.5 Å from any heavy atom of the solute is shown. (Middle) The low and high level pseudoelectron densities of water molecules, calculated from the water molecule distribution by using a Fourier transformation, are represented. (Bottom) From the preceding map contoured at high level, hydration sites are characterized.

2 Water and ion binding sites

The methods used to determine water binding sites around specific structural elements extracted from crystal structures have been developed by the group of Schneider and Berman [13-16] on the basis of the work of Murray-Rust and Glusker [17]. They have been used in order to characterize water binding sites around DNA bases [13-15], phosphate groups [16], and RNA base-pairs

[9] from high resolution crystal structures as well as water and ion binding sites around RNA and DNA base-pairs from MD simulations [10,11]. Many details concerning these methods can be found in the related publications and especially in [15]. Here, an outline of the methods we have used is provided for a r(G=C) base-pair on the basis of a 2.4 ns MD simulation of an $r(CpG)_{12}$ duplex [10].

A first step consists of locating all the water binding sites surrounding a nucleic acid base-pair. This is done by saving at each step of the MD run all the water molecules that are located at less than 3.5 Å from any heavy atom of the base-pair (Fig. 1). As a result, "Hydrated building blocks" are formed [15]. The 3.5 Å water-to-base distance criteria correspond to the first minimum in the corresponding calculated radial distribution functions [18]. Additionally, in order to exclude water molecules belonging to the first hydration shell of neighboring base-pairs, an out-of-plane boundary of 2.5 Å is set (note that the average interbase distance is close to 3.4 Å). This 2.5 Å criterion is not applied to atoms belonging to the sugar-phosphate backbone. For water molecules that are nevertheless common to the hydration shell of adjacent pairs, an occupancy factor of 0.5 instead of one is assigned to each of them in order not to count them twice.

In a second step, by using the SFALL program of the CCP4 library (http://www.dl.ac.uk/CCP/CCP4), a Fourier transformation is applied to the cloud of water molecules surrounding the base-pair in order to generate a "pseudoelectron" density map (Fig. 1). The O program [19] is used to display the calculated "pseudoelectron" densities (Fig. 1). Specific contour levels of the "pseudoelectron" densities can be drawn in order to emphasize the relative weights of the different binding sites (Fig. 1). For the G=C base-pair, the cell dimensions have been set to a = b = c = 40 Å and $\alpha = \beta = \gamma = 90°$ in space group P1. The principles for determining ion binding sites are the same as those described for water (Fig. 2).

By using the O program [19], water molecules or ions are then fitted manually in the calculated densities and their positions are refined by using the SHEL-97 program [20]. For the refinement process, in order to exclude from the computation regions of the density map which are not associated with the water peaks considered before, a sigma value of 4 Å is chosen. During the refinement process, isotropic or anisotropic "pseudo" thermal factors for the base-pair atoms and solvent molecules [15] can be calculated [21] (Fig. 3). Although these calculated "pseudo" thermal factors give a precise idea of the delocalization of water molecules occupying well defined hydration sites in a liquid environment, they are not equivalent to experimental thermal factors determined by crystallography. Furthermore, in the crystal, many water molecules are involved in contacts which limit their motions, while some hydration sites are occupied by solute atoms involved in crystal packing contacts. Four coordinate files in the Protein Data Bank format corresponding to the r(G=C), d(G=C), r(A-U), and d(A-T) base-pairs with their complete hydration shell and positions for some well defined hydration sites along

with "pseudoelectron" density files in the CCP4 format are accessible at the http://www-ibmc.u-strasbg.fr/upr9002/westhof site [11].

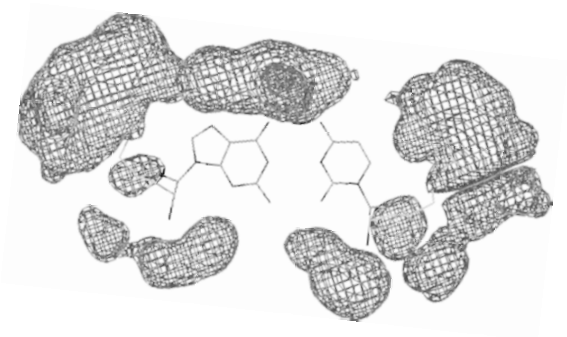

Fig. 2. Ion density calculated around the r(G=C) base-pair [10]. High level densities pointing to the favored ion binding sites in the vicinity of the (G)O6 and (C)N4 atoms.

3 Number of solvent molecules facing hydrophilic atoms

The number of water molecules facing a given hydrophilic solute is calculated by counting the solvent molecules that are located at less than 3.5 Å from each hydrophilic atom with an additional out-of-plane criterion of 2.5 Å for the atoms belonging to the bases as defined above. For Watson-Crick pairs, most of the sites are surrounded by close to one water molecule indicating an occupancy factor close to one (Table 1). Yet, some hydrophilic sites enter into contact simultaneously with more than one water molecule like the OR and OS anionic oxygen atoms and the RNA O2' atoms which are surrounded by close to three water molecules. These three water molecules form well defined hydration cones [16,22,23]. Hence, the simple count of water molecules in the vicinity of a hydrophilic atom (Table 1) is useful for quickly estimating the number of its accessible donor and acceptor sites. It is useful to mention that overlaps between adjacent binding sites cannot be avoided. As a consequence, the number of water molecules surrounding a base-pair is not equivalent to the sum of the water molecules associated with each individual binding site. Thus, individual counts have to be performed in order to calculate the precise number of water molecules surrounding each functional group or whole base-pair.

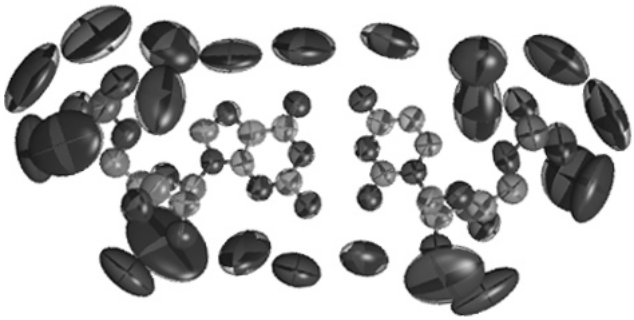

Fig. 3. "Pseudo" thermal ellipsoids, calculated for a r(G=C) base-pair from a 2.4 ns MD simulation of a r(CpG)$_{12}$ duplex [10], are plot at a 50% level probability by using the RASTEP program which is part of the RASTER3D molecular package [20]. (see Plate 2 on page 497.)

4 Hydrogen bonding percentages (HB%)

The hydrogen bonding percentages (Table 1) complete the information given by the numbers defined above. They allow to derive a finer picture of the contacts established between the solute and the solvent atoms. These HB% are defined as the total number of hydrogen bonds established during a single trajectory between a particular solute atom and the surrounding water molecules divided by the total number of configurations analyzed [10,11,24,25]. The hydrogen bonding criteria used are d(A...H)$<$ 2.5 Å and Θ(A...H-D)$>$ 135°, A representing an acceptor atom and D a donor atom. Specifically, although the "occupancy" of the site located in front of the (G)N3 is of 0.8, the (G)N3 atoms appears to be a not so efficient hydrogen bond acceptor (HB%$\approx$56). Similarly, the O4' and O5' atoms have close to one water molecule in their vicinity (i.e., 0.9 and 0.7), and the associated HB% values are low (O4'$\approx$ 14; O5'$\approx$ 18) as they are involved in C2'-H(n)...O4'(n+1) and C6/8-H...O5' intra-molecular contacts [9,23]. The HB% values close to 300 for the OR and OS atoms reflect the fact that there are three water molecules involved in hydration cones forming each one hydrogen bond contacts. Low HB% values reflect also mixed occupancies of the binding sites. For example, the binding site in front of the O6 atom is alternatively occupied by water molecules and ions leading to a HB% value of 71. More generally, the occupancy of each site by solvent molecules is always close to one, although these values may reflect mixed occupancies by water molecules and ions. Hence, the occupancy factors determined experimentally are really meaningful only for sites occupied by more than one type of solvent molecules and typically reflect methodological and conceptual difficulties in the interpretation of experimental data.

Table 1. Average number of solvent/RNA contacts as well as hydrogen bonding percentages (HB%) calculated for a r(G=C) base-pair from a 2.4 ns MD simulation of a r(CpG)$_{12}$ duplex (a complete table is found in [10]).

Deep groove			Shallow groove					Backbone				All atoms
N7	O6	H42	N3	H22	O2	O2'	OR	OS	O3'	O4'	O5'	
A. Average contacts (Water)												
1.1	1.0	1.3	0.8	1.5	0.9	2.6	2.9	3.5	1.3	0.9	0.7	21.3
B. Average contacts (Ion)												
0.06	0.18	0.11	≈0	≈0	≈0	0.04	0.07	0.07	≈0	≈0	≈0	0.6
C. Hydrogen bonding percentages (HB%)												
91	71	75	56	79	83	88	279	287	36	14	18	

5 Residence times

The ability of being able to estimate residence times of solvent molecules in the vicinity of nucleic acids is one of the strength of MD simulations, since such information are up to now very difficult to obtain through experiments. They cannot be accessed by using x-ray methods, while NMR techniques can only provide values for water molecules located in the vicinity of non-exchangeable protons or located in the deep/shallow groove [5,26,27]. Next, details are given for the methods developed by us [10,11,24,25] in order to calculate water molecule and ion residence times.

The estimation of residence times is based on hydrogen bonding contact times (HBct) which represent the time during which an individual water molecule establishes a hydrogen bond contact with a individual solute atom. It has consistently been observed in MD simulations that if a water molecule forming a hydrogen bond with a solute atom moves away, it almost never comes back to the same binding site on the currently investigated time scales, although this may be observed on much longer time scales or for water molecules trapped in hydration pockets. Thus, it is proposed that HBct represent a good estimate of the true residence times of water molecules on the investigated time scales. As an example, the number of water molecules bound to the 24 OR and OS atoms of the 12 central base-pairs of the r(CpG)$_{12}$ duplex in a 2.4 ns MD simulation [10,11] are plotted as a function of their HBct (Fig. 4, note that only the last 1.5 ns of the simulation have been analyzed). The profiles calculated for the OR and the OS atoms are clearly different. The curves indicate that at least one water molecule with a residence time exceeding 500 ps is bound to each of the 24 OR atoms and water molecules are less strongly bound to the more solvent accessible OS atoms. A characteristics of these plots is that the area under the curves remain constant (note that, here, HBct values below 1% of the total simulation time have not been plotted). Thus, it can be observed that less water molecules enter into contact with the atoms for which bound water molecules display longer

residence times (OR≈36) than with the atoms for which the bound water molecules display shorter residence times (OS≈60).

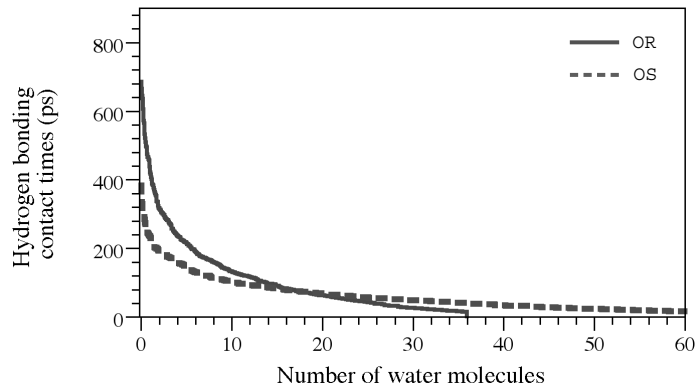

Fig. 4. Hydrogen bonding times (HBct) plotted versus the number of water molecules attached to the OR and OS atoms of a r(G=C) base-pair. The data have been normalized with respect to the number of phosphate groups present in the duplex and, therefore, differ from the non-normalized data presented in [10].

For water molecules bound to acceptor atoms, one must consider the rotational behavior since a water molecule can quickly rotate around his axis. In this respect both hydrogen bonds are equivalent and HBct take indifferently into account Ow-H1...A and Ow-H2...A interactions. Thus, the HBct values reflect more the residence time of a water molecule in a given binding site than the life time of a specific hydrogen bond.

Absolute values for the residence times of solvent molecules or any other calculated properties are very difficult to obtain from MD simulations, since limits are always set by the quality of the force-fields, simulation length, and other methodological approximations. Thus, the profiles presented in Figure 4 will look different if calculated for longer simulations unless the simulation time is much longer than that of the process with the longest relaxation time. This is shown in Figure 5 were data from a 2.4 ns [10] and a 4.4 ns (unpublished results) are compared for the OR atom of a r(G=C) base-pair. The curve corresponding to the longest simulation is shifted toward the longer time-scales as a result of an improved sampling. Yet, the shape of the curve is not significantly altered indicating that these data describe qualitatively the dynamics of the water molecules bound to these atoms (note that the area under the curve corresponding to the 3.0 ns sampling is twice that of the curve corresponding to the 1.5 ns sampling). For much longer simulations, the following changes are expected: (i) very infrequent events with long relaxation times may change the maximum value of the profile; (ii) the slope of the

profile will be altered since more events with long relaxation times will be sampled. Yet, although as noted before absolute values for residence times cannot be accessed with precision, for a given time scale, comparison between data obtained for different set of atoms or between different structures are still extremely informative.

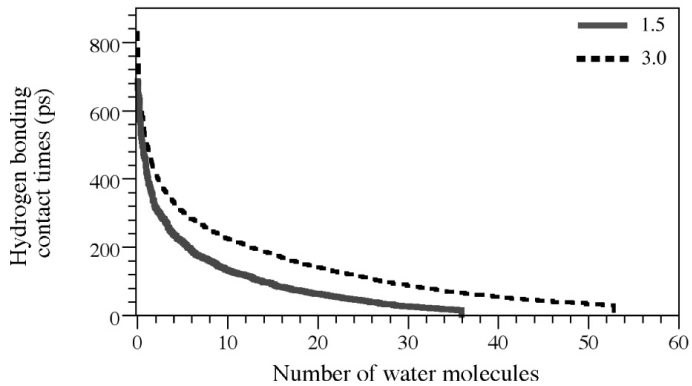

Fig. 5. Comparison between water molecule residence time data extracted from a 1.5 ns and a 3.0 ns part of a 2.4 ns [10] and 4.4 ns (unpublished results) MD simulation, respectively, for the OR atoms. The 1.5 ns curve is identical to that presented in Fig. 4.

6 Conclusions

Creating and adapting tools for extracting information from the huge amount of raw data generated by computer simulations has always been an as important task as producing the data themselves. Given the labile character of the structure of the hydration and ionic shells surrounding nucleic acids, new challenges are faced especially in the determination of the residence times of the solvent molecules. Yet, simulations have already shown their value in reproducing experimental data from which original views could be extrapolated. In view of that, MD simulations will continue to help clarify major sequence dependent properties of nucleic acids by investigating hydration and ion binding features beyond the Watson-Crick base-pair level.

Here, we described the procedures used for characterizing the structural and dynamical properties of the first hydration and ion binding shells surrounding the four r(G=C), r(A-U), d(G=C), and d(A-T) base-pairs. By using crystallographic refinement methods, coordinates for the water molecules proximate to the four base-pairs were generated and anisotropic "pseudo"

thermal factors were calculated. Such coordinate sets will certainly be useful in the refinement of low to medium resolution crystal structures [28] and in the elaboration of nucleic acid hydration shell models [15]. The precise knowledge of the shape adopted by these solvent shells may also help us to deepen our understanding of recognition processes since some of these water and ion binding sites correspond to preferential binding sites for chemical groups belonging to nucleic acids, proteins, and drugs [16,23,29,30].

References

1. Westhof E.: Water: an integral part of nucleic acid structure. Annu. Rev. Biophys., Biophys. Chem. 17 (1988) 125-144.
2. Westhof E., Beveridge D.L.: Hydration of nucleic acids. In: Franks F. (ed.) Water Science Reviews, Vol. 5. Cambridge University Press, Cambridge (1990) 24-123.
3. Auffinger P., Westhof E.: Roles of hydration on the structure and dynamics of nucleic acids. In: Ross Y.H., Leslie R.B., Lillford P.J. (eds.): Water Management in the Design and Distribution of Quality Food. Technonic Publishing Co. Inc., Basel (1999) 165-198.
4. Wüthrich K.: NMR-This other method for protein and nucleic acid structure determination. Acta. Cryst. D51 (1995) 249-270.
5. Halle B., Denisov V.P.: Water and monovalent ions in the minor groove of B-DNA oligonucleotides as seen by NMR. Biopolymers 48 (1998) 210-233.
6. Phan A.T., Leroy J.L., Guéron M.: Determination of the residence time for water molecules hydrating B'-DNA and B-DNA, by the one-dimensional zero-enhancement nuclear Overhauser effect. J. Mol. Biol. 286 (1999) 505-519.
7. Hozba P., Sponer J.: Structure, energetics, and dynamics of the nucleic acid base pairs: nonempirical Ab-Initio calculations. Chem. Rev. 99 (1999) 3247-3276.
8. van der Vaart A., Merz K.M.: The role of polarization and charge transfer in the solvation of biomolecules. J. Am. Chem. Soc. 121 (1999) 9182-9190.
9. Auffinger P., Westhof E.: RNA base pair hydration. J. Biomol. Struct. Dyn. 16 (1998) 693-707.
10. Auffinger P., Westhof E.: Water and ion binding around RNA and DNA (C,G)-oligomers. J. Mol. Biol. 300 (2000) 1115-1133.
11. Auffinger P., Westhof E.: Water and ion binding around $r(UpA)_{12}$ and $d(TpA)_{12}$ oligomers - Comparison with RNA and DNA $(CpG)_{12}$ duplexes. J. Mol. Biol. 305 (2001) 1057-1072.
12. Feig M., Pettitt B.M.: A molecular simulation picture of DNA hydration around A- and B-DNA. Biopolymers 48 (1998) 199-209.
13. Schneider B., Cohen D., Berman H.M.: Hydration of DNA bases: analysis of crystallographic data. Biopolymers 32 (1992) 725-750.
14. Schneider B., Cohen D.M., Schleifer L., Srinivasan A.R., Olson W.K., Berman H.M.: A systematic method for studying the spatial distribution of water molecules around nucleic acid bases. Biophys. J. 65 (1993) 2291-2303.
15. Schneider B., Berman H.M.: Hydration of the DNA bases is local. Biophys. J. 69 (1995) 2661-2669.

16. Schneider B., Patel K., Berman H.M.: Hydration of the phosphate group in double-helical DNA. Biophys. J. 75 (1998) 2422-2434.
17. Murray-Rust P., Glusker J.P.: Directional hydrogen bonding to sp^2 and sp^3 hybridized oxygen atoms and its relevance to ligand macromolecules interactions. J. Am. Chem. Soc. 106 (1984) 1018-1025.
18. Feig M., Pettitt B.M.: Modeling high-resolution hydration patterns in correlation with DNA sequence and conformation. J. Mol. Biol. 286 (1999) 1075-1095.
19. Jones T.A., Zou J.Y., Cowan S.W., Kejldgaard M.: Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Cryst. A47 (1991) 110-119.
20. Sheldrick G.M., Schneider T.R.: High resolution refinement. Meth. Enzymol. 276 (1997) 307-326.
21. Merritt E.A., Bacon D.J.: Raster3D photorealistic molecular graphics. Meth. Enzymol. 277 (1997) 505-524.
22. Egli M., Portmann S., Usman N.: RNA hydration: a detailed look. Biochemistry 35 (1996) 8489-8494.
23. Auffinger P., Westhof E.: Rules governing the orientation of the 2'-hydroxyl group in RNA. J. Mol. Biol. 274 (1997) 54-63.
24. Auffinger P., Louise-May S., Westhof E.: Hydration of C-H groups in tRNA. Farad. Discuss. 103 (1996) 151-174.
25. Auffinger P., Westhof E.: RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated $tRNA^{Asp}$ anticodon hairpin. J. Mol. Biol. 269 (1997) 326-341.
26. Denisov V.P., Carlström G., Venu K., Halle B.: Kinetics of DNA hydration. J. Mol. Biol. 268 (1997) 118-136.
27. Johannesson H., Halle B.: Minor groove hydration of DNA in solution: dependence on base composition and sequence. J. Am. Chem. Soc. 120 (1998) 6859-6870.
28. Klosterman P.S., Shah S.A., Steitz T.A.: Crystal structures of two plasmid copy control related RNA duplexes: an 18 base pair duplex at 1.20 Å resolution and a 19 base pair duplex at 1.55 Å resolution. Biochemistry 38 (1999) 14784-14792.
29. Seeman N.C., Rosenberg J.M., Rich A.: Sequence-specific recognition of double helical nucleic acids by proteins. Proc. Natl. Acad. Sci. USA 73 (1976) 804-808.
30. Woda J., Schneider B., Patel K., Mistry K., Berman H.M.: An analysis of the relationship between hydration and protein-DNA interactions. Biophys. J. 75 (1998) 2170-2177.