Roles of Hydration on the Structure and Dynamics of Nucleic Acids

P. AUFFINGER and E. WESTHOF

in: Water management in the design and distribution of quality foods
Ross, Y.H., Leslie, R.B., and Lillford, P.J. - editors

ABSTRACT

Nucleic acids, especially DNA molecules, are structurally strongly dependent on the amount of available water molecules and ions present in their environment. Thus, depending on the humidity level and ionic strength, various helical forms of DNA are observed. Similarly, structured RNA molecules adopt functional folds only in the presence of divalent ions. In short, both water activity and ionic strength control structure-function relationships of nucleic acids so much that water molecules and ions are considered integral parts of the nucleic acid structure.

We will review some observations extracted from x-ray and NMR experiments about nucleic acid hydration and ion binding. The tetrafunctional water molecule, with an equal number of basic and acidic sites, has many ways of interacting with the hydrophilic atoms of nucleic acids. In all forms of DNA as well as in structured RNAs, intra-residual water bridges fulfill the hydrogen bonding capacity of the polar atoms, forming strings, spines, or filaments of linked water molecules with exchangeable mobilities. At the same time, the nucleic acids, through variations in torsion angles of the sugar-phosphate backbone and through reorientations of the bases, adapt their structures so that their polar hydrophilic atoms form three-dimensional networks able to interact with the solvent.

We will also discuss molecular dynamics simulations which, when performed in aqueous solution and in the presence of counterions with full computations of electrostatic interactions, lead to reproducible and stable trajectories. From these trajectories, the solvent accessibility of the hydrophilic sites, but also of C-H groups could be estimated, and C-H...O_w hydrogen bonds similar to those observed in crystal structures could be characterized. Among others, the positions and dynamics of the water molecules surrounding the RNA 2'-hydroxyl group, for which no precise experimental data are available, could be described in great details. Further, molecular dynamics simulations allow to estimate residence times of structurally important water molecules.

Pascal Auffinger 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.
INTRODUCTION

In this review, we will first briefly describe some basic facts concerning the chemical and three-dimensional structures of nucleic acids. We will especially stress the high structural polymorphism of nucleic acids in their dependence on the aqueous and ionic environments. The description of the hydration of nucleic acids will be achieved from two view points: a static and a dynamic one. The static view results from the analysis by x-ray diffraction of crystalline nucleic acid fragments. X-ray crystallography reveals “hydration sites” around the crystallized molecules. Those hydration sites are well defined, if the residence time of water molecules occupying those sites are high. The residence times are expected to be proportional to the number of contacts made by water molecules bound in the site. Thus, the electrostatic environment, the hydrophilicity of the macromolecular atoms, and the geometry and size of the binding sites (hole, wedge, groove...) will constitute the relevant parameters.

The dynamic viewpoint will be given by a short review of the NMR analysis of hydration around nucleic acids and by molecular dynamics simulations of nucleic acid fragments. These approaches emphasize the microheterogeneity in the local hydration structure and in residence times.

STRUCTURE OF NUCLEIC ACIDS

Nucleic acid biopolymers comprise DNA and RNA molecules. These two types of molecules possess very different functional roles. In brief, DNA molecules (deoxyribonucleic acids) contain the genetic code, whereas the more versatile RNA molecules (ribonucleic acids) are involved in almost all crucial life processes and especially in the translation of the genetic code into proteins [1,2].

The four deoxynucleosides, adenine (A), thymine (T), guanine (G), and cytosine (C), constitute the basic building blocks of the DNA polymeric chain (Figure 1). These nucleosides comprise a deoxyribose sugar ring and a purine (A, G) or a pyrimidine (C, T) base. They are connected together by a phosphodiester linkage. The nucleoside and its phosphodiester unit are called a nucleotide. In order to perform their functions, the bases of a DNA polymeric chain associate with those of a complementary DNA chain by forming purine-pyrimidine, A-T or G-C, Watson-Crick base pairs (Figure 3). The association of two self-complementary strands results in the formation of a right or more rarely a left handed double helical structure (Figure 2).

Although DNA and RNA molecules possess very distinct functions, chemically RNA differs from DNA only in two aspects: (1) the absence of the methyl group at position 5 of the uridine (U); and (2) the presence of the 2'-hydroxyl group of the RNA ribose sugar (Figure 1). These two small chemical modifications account for the profound functional and structural differences observed between DNA and RNA molecules. In contrast to DNA, but similar to proteins, single stranded polynucleotide RNA chains can fold in a large variety of three-dimensional structures. This ability to form complex folds is exemplified by transfer RNA
Figure 1. Chemical structure of a [...ATGC...] DNA sequence drawn from the 5' to the 3' end. In the inset, the two chemical differences distinguishing DNA from RNA are marked by an arrow: 1) the DNA thymine bases (T) are replaced, in RNA, by uridines (U), which lack the methyl group at position 5, 2) in the sugar ring, the RNA 2'-hydroxyl group is replaced by a hydrogen atom in DNA.

(tRNA) molecules, which are constituted by a single chain of about 70 nucleotides (Figure 2). The analogy with proteins includes also the fact that some RNA molecules are able to perform catalytic reactions [3].

ENVIRONMENTAL-DEPENDENT POLYMORPHISM

DNA double helices are polymorphic and can adopt three major conformations called A, B, and Z (Figure 1) depending on the relative humidity and the ionic strength of their environment. A more complete description of nucleic acid polymorphism can be found in Saenger [1]). A- and B-form helices are right
Figure 2. Side and top views of the three B-, A-, and Z-forms adopted by DNA double helical structures. In the inset, the three-dimensional structure of a transfer RNA molecule (tRNA) resulting from the complex folding of a single polymer chain comprising 75 ribonucleotides is represented.

handed while Z-form helices are left-handed. B-DNA forms occur at low salt conditions and these are believed to be the conformation adopted most frequently by biologically “active” DNA molecules. In this form, the two DNA grooves, called major and minor, differ in width, but they are comparable in depth. When the salt concentration is raised, the water structure around the nucleic acid is altered and transformation to the A-DNA form takes place. The major groove becomes deeper and narrower and the minor groove becomes wider and shallower. Z-DNA left-handed helices are essentially observed for alternating poly(dG-dC) sequences in very high salt conditions (>4 M NaCl). For Z-DNA, the G-C base pairs are still of the Watson-Crick type, while the conformation of the individual nucleotides is considerably altered.

RNA molecules are less affected by changes in their environment and adopt essentially the A-form in the helical parts of their structures under nearly all
conditions. Many RNA molecules, however, need specific divalent cations in order
to fold properly into biologically active conformation [4]. The absence of such ions
may induce profound structural changes, such as loss of important three-
dimensional interactions.

Although there is a clear correlation between DNA conformation and water
activity (which is related to humidity and ionic strength), the mechanistic roles of
water molecules in DNA conformational transitions are not completely understood
[5,6]. However, large amount of data originating mainly from crystallography are
available. NMR and computer simulations give a good description of the local
hydration of DNA and RNA molecules. Some of the results obtained from these
methods will be summarized. Some earlier reviews have emphasized the integral
roles of water and ions in nucleic acid structures [5,7-10]. Several reviews
describing the hydration of nucleotides [2,5], A-DNA [11,12], B-DNA [13], Z-
DNA [6], and RNA [14] are also available.

Figure 3. Possible hydration sites around G-C and A-T Watson-Crick base pairs and sugar-
phosphate backbone atoms. Hydrogen bond acceptor sites are marked by bold gray arrows,
hydrogen bond donor sites are marked by bold black arrows, weak hydrogen bond donor sites (C-H
groups) are marked by thin arrows. The RNA 2'-hydroxyl group (see inset) display simultaneously a
hydrogen bond donor and acceptor potential.
THE CRYSTALLOGRAPHIC PERSPECTIVE

HYDRATION OF WATSON-CRICK BASE PAIRS

Watson-Crick base pairs constitute the basic unit of interstrand nucleic acid recognition and are essentially found in DNA and RNA helical regions. As a consequence of regular hydrogen bonding interactions between bases, the Watson-Crick hydrophilic sites of standard DNA and RNA helices are not accessible to water. Hence, the sequence specific hydration patterns of nucleic acid helices originates from the accessibility to water of the non-Watson-Crick hydrophilic sites located in the minor and major grooves (Figure 3). Further, the rather regular structures molds hydration patterns with helical periodicity, especially in homopolymer heteropolymers.

The hydration of DNA Watson-Crick bases has been investigated through the analysis of oligonucleotide crystal structures by Schneider and coworkers [15-17]. Despite the fact that the bases are buried to 80% in the nucleic acid structures, as

Figure 4. Scattergram of water molecules found within 3.4 Å of guanine bases extracted from 14 B-DNA decamer structures (top left stereoview). Density of water molecules derived from scattergrams similar to the one presented on the left for the four A, G, C, and T bases in a B-DNA Watson-Crick base pairing scheme and forming "hydrated" building blocks (right top and bottom). Possible explanation for the two distant hydration sites located close to the guanine N2 atoms (bottom left). The HS1 and HS2 sites of both cytosines may contribute to the weak hydration of the guanine amino groups. All figures are adapted with permission from Schneider and Berman [17].
deduced from the total nucleic acid surface accessible to water, the authors found that the DNA bases are still remarkably hydrated in the three, A-, B-, and Z-DNA, forms. They observed that the hydration of the bases in the A- and B-forms is comparable while that of the bases in the Z-form is quite distinct [15,16]. These authors reached the conclusion that water molecules around nucleic acid bases are clustered at specific locations close to the hydrophilic atoms. For B-DNA structures, Schneider and Berman [17] used the results of their analysis to generate hydrated building blocks, which are formed by the nucleic acid bases and their associated hydration sites (Figure 4). Among other observations, they noted that guanine major groove hydration is mostly confined to the 3' side of the base and the minor groove N3 atom has a single hydrogen bonding site slightly out of the base plane in the 5' direction. Surprisingly, the N2 amino group is poorly hydrated. The two density spots observed above and below the guanine base, close to the amino group, may result from the hydration of neighboring pyrimidines, as proposed by Schneider and Berman [17] and could, thus, be part of sequence dependent hydration motif.

The concept of hydrated building blocks appears to be very attractive and will probably gain popularity, as the increasing number of high resolution crystal structures will reinforce the notion of well defined hydration sites around the nucleic acid bases not only for Watson-Crick helices, but also for non-canonical nucleic acid motifs. Besides, this concept may be extended to the dinucleotide or to larger units and help to understand and rationalize sequence specific hydration patterns.

HYDRATION OF NON WATSON-CRICK BASE PAIRS

Non-Watson-Crick base associations are rarely found in DNA but very often in RNA structures, where they serve as specific recognition elements for proteins, nucleic acids, and ligands, or as ion binding sites. Such non-canonical base pairs are linked by at least one interbase hydrogen bond and occasionally involve water mediated base-base interactions. Next, the hydration of representative examples of non-Watson-Crick base pairs will be described.

Wobble Base Pairs

Wobble base pairs are typical of RNA molecules and can be inserted without great distortions into regular Watson-Crick helices. These base pairs display a characteristic shift of their Watson-Crick interaction sites. Two of these sites point respectively toward the minor and major groove and are, thus, completely accessible to the solvent. Wobble base pairs comprise, among others, G-T (DNA) and G-U base pairs (RNA). For these base pairs, the guanine amino group protrudes into the minor groove and the O4 atom of the thymine or uridine base point towards the major groove. Furthermore, G-U (or G-T) base pairs display a deep groove side (or major groove side) lined with three hydrophilic acceptor atoms (N7, O6, O4) and no hydrophilic donor atoms while the shallow (minor)
Figure 5. Different wobble base pairs from crystal structures. Environment around a G-T base pair in the structure of the Z-DNA hexamer d(CGCGTG) [18] (top). In one G-T base pair, a hexahydrated Mg$^{2+}$ ion coordinates to the DNA major groove. Note, the similar hydration pattern around both base pairs. Stereoview of the hydration of an I-T wobble base pair [22] (middle). View of the hydration of two G-U base pairs from the crystal structure of the r(GUGUGUA)dC oligomer with tandem GU/UG wobble base pairs [29] (bottom). The figures are adapted with permission from Ho et al. [18], Cruse et al. [22], and Biswas and Sundaralingam [29].

groove side presents a cavity wide enough for trapping a water molecule. Both grooves present, therefore, very unique recognition patterns.

Specific hydration motifs around G-T or G-U base pairs have been observed in several crystal structures. Figure 5 presents two hydration patterns of a G-T wobble base pair, as observed in the crystal structure of a Z-type d(CGCGTG) oligonucleotide. The first involves only water molecules while the second involves a [Mg(H$_2$O)$_6$]$^{2+}$ cluster [18]. Interestingly, the fact that the guanine and thymine hydration sites are very similar for both G-T base pairs indicates that ion coordination can occur without significantly disturbing the hydration pattern of the wobble base pair. Furthermore, these hydration patterns seem to occur recurrently for different types of wobble base pairs, i.e., G-T [19-21], I-T [22] or I-m$^5$C [23],
in A-form DNA [12] and are very close to those observed for the bases involved in regular Watson-Crick interactions (Figure 5). As for Watson-Crick base pairs, the concept of hydrated building blocks could help to clarify our view of the hydration of such non-regular base associations.

Remarkably, the hydration site located for G-T base pairs between the (G)N\textsubscript{2} and (T)O\textsubscript{2} atoms is also observed in I-T base pairs lacking the N\textsubscript{2} amino group, and between the (G)N\textsubscript{2} and (U)O\textsubscript{2} atoms of G-U base pairs found in RNA [24, 25] (Figure 5). For G-U base pairs, the water molecules occupying this hydration site may form additional contacts with the RNA 2'-hydroxyl group, although these water molecules are probably not long-lived, as indicated by molecular dynamics simulations [26]. It has been often proposed that the water molecules trapped in the hydration site close to the RNA 2'-hydroxyl group are contributing to the stability of G-U base pairs, although water molecules are found at similar locations in G-T and I-T base pairs of DNA lacking the 2'-hydroxyl group. Thus, the hydration site close to the (T)O\textsubscript{2} or (U)O\textsubscript{2} atom seems to be mainly related to the presence of the O\textsubscript{2} carbonyl oxygen atom. Further, from the crystal structure of a r(CCCCGGGG) oligomer, it was noted that, in the majority of base pairs, water molecules link the 2'-hydroxyl groups to the (C)O\textsubscript{2} and (G)N\textsubscript{3} atoms [27, 28], as also inferred from molecular dynamics simulations [26].

However, water molecules wedged between the base and the sugar are not systematically observed in the shallow groove of wobble base pairs. In several crystal structures [19, 20], it was noticed that some atoms of the mismatches are involved in crystal packing interactions and, consequently, are prevented from interacting with water molecules. Thus, some hydration sites may be occupied by hydrophilic atoms belonging to symmetry related molecules instead of being occupied by solvent molecules. This should be taken into account in the determination of hydrated building blocks from crystal data.

Water Mediated Base-Base Interactions

Specific recognition between two nucleic acid bases is best achieved by pairing involving at least two hydrogen bonds. Yet, the occurrence of base pairs linked by a single hydrogen bond was noted in the first crystallographic structures of transfer RNA molecules [2, 32-34]. However, the low resolution of these structures did not permit the determination of their hydration patterns. A single hydrogen bonded C-U mismatch (Figure 6) associated with a solvent mediated base-base interaction was found in a crystal structure of the r(GCUUCGGC)\textsubscript{dB}U oligomer [24]. The insertion of the water molecule resulted in orientations of the glycosidic bonds of the C-U base pair similar to those of a Watson-Crick base pair. It was concluded that the C-U base pair could be inserted in a standard A-RNA duplex without inducing large distortions. A similar water mediated base-base interaction was reported earlier in a crystal structure of tRNA\textsuperscript{G11} complexed with its cognate synthetase [30].

Recently, the high-resolution crystal structure of the ribosomal 5S loop E structural motif revealed unforeseen types of base pair associations [31], such as the single hydrogen bonded G-U, G-A, and G-G base pairs involving solvent
mediated interactions (Figure 6). Such mismatches indicate that, when water is taken into consideration, the possible arrangements of base pairs expands the structural diversity of RNA molecules as well as the array of their specific recognition elements. In such motifs, water has to be considered as an integral part of the mismatches.

HYDRATION OF THE BACKBONE

While the hydration of the B-form and A-form DNA bases share some common motifs [15,16], the hydration of their sugar-phosphate backbone is clearly different. Taken as an example, water bridges linking anionic oxygen atoms belonging to
adjacent phosphate groups are often observed in A- and Z-DNA, but not in B-DNA, where the distance between the phosphate groups is too large (Figure 7).

It was suggested that the more "economical" hydration in A- and Z-DNA compared to B-DNA, which results from reduced water activity, is the underlying cause of B to A and B to Z transitions [2,35]. Noteworthy, recurrent water bridges are often seen in the deep groove of helical parts of A-DNA [12] and RNA molecules [9] and these have been extensively described [9,36]. Two nice examples of strings of bridging water molecules, which appear to be sequence independent, have been observed in the high resolution structures of the r(CCCCGGGG)\textsubscript{2} [28] and 2'-O-Me(CCGCGG)\textsubscript{2} [37] RNA duplexes (Figure 8). These water molecules participate to recurrent hydration motifs linking the bases to the backbone. Recurrent hydration patterns linking the base to the 2'-hydroxyl groups are also seen in the shallow groove of RNA duplexes [28]. Energetically, water binding to phosphate groups is important since it involves binding to a charged moiety.
r(CCCCCGGGG) 2'-O-Me(CGCGCGC)

Figure 8. Schematic view of the water structure in the deep groove of an r(CCCCCGGGG) duplex (left) [28] and of a 2'-O-Me(CGCGCGC) RNA duplex with methylated hydroxyl sugar groups (right) [37]. The figures are adapted with permission from Egli et al. [28] and Adamiak et al. [37].

SPINES OF HYDRATION

Besides the strings of water molecules observed in the deep groove of RNA, spines of hydration were detected in the minor grooves of Z- and B-DNA (Figure 9). Some of the water molecules constituting these spines are involved in bridging interactions between atoms of successive base pairs. In B-DNA, the water molecules bridging the N3 and O2 atoms and, in Z-DNA, the water molecules bridging the O2 atoms are themselves connected by water molecules and form a spine of hydration running down the minor groove [6,38]. These motifs were found to be sequence dependent and are essentially observed in B-DNA at AATT steps. It was noticed that the amino group of guanines, which occupies part of the minor groove, disrupts the B-DNA spine of hydration while the occurrence of A-T steps in Z-DNA alters the helical conformation and, consequently, interrupts the spine of water molecules [8]. From an historical point of view, the detection of a spine of hydration in the first solved structure of a B-DNA dodecamer [39] gave early evidence of the structural role played by water molecules in nucleic acids.
HYDRATION OF C-H GROUPS

Non-bonded interactions involving O-H and N-H donor groups were long considered as the sole kind of structuring hydrogen bonds occurring in biomolecules. Therefore, most of the crystallographic studies focused on the hydration of the O-H and N-H "hydrophilic" groups. Yet, in certain structural contexts, C-H groups were observed to form C-H...O or C-H...N hydrogen bonds. A large body of experimental data points to the significance of these interactions in chemical as well as biochemical systems [42-46]. They occur recurrently in protein tertiary motifs [47,48] and catalytic sites [49], and they play an important role in some inhibitor/enzyme interactions [50] as well as in protein-DNA recognition [51]. In nucleic acids, the base to backbone pyrimidine C$_7$-H$_6$...O$_5^-$ and purine C$_8$-H$_8$...O$_5^-$ hydrogen bonds constitute the most widely recognized C-H...O interactions, as they occur at each nucleotidic step in A-form helical structures [1,2]. Another C-H...O interaction, recently characterized is the intrastrand C$_2$-H$_2(n)$...O$_4(n+1)$ hydrogen bond, which links adjacent riboses in the axial direction.
in A-form nucleic acids [52]. In both cases, the C-H groups are the only hydrogen bond donors in proximity to the O5' or O4' oxygen atoms, which are themselves buried within the structure and can difficulty interact with water molecules, as shown by molecular dynamics simulations [26].

Extensive experimental evidence for the occurrence of C-H...O\textsubscript{W} hydrogen bonds in biomolecules is also available. Neutron diffraction studies revealed C-H...O\textsubscript{W} hydrogen bonds in cyclodextrin inclusion complexes [53]. Similar interactions were observed in a crystal structure of the vitamin B\textsubscript{12} coenzyme [54]. Examples of C-H...O\textsubscript{W} hydrogen bonds are also frequently reported in crystal structures of nucleotides and small molecules [46,55]. In nucleic acid helical motifs, water molecules can interact with the C\textsubscript{5}-H\textsubscript{5} group of pyrimidines especially when they are involved in bridging interactions between anionic oxygen atoms of adjacent phosphate groups (Figure 7). Molecular dynamics simulations have shown that such water molecules, which bridge successive anionic oxygens, can form long-lived hydrogen bonds with the C\textsubscript{5}-H\textsubscript{5} groups of pyrimidines [56].

Therefore, besides the established O-H and N-H donor groups, polarized C-H groups should be considered as potential hydrogen bond donors, which may be involved in the formation of intramolecular and intermolecular interactions as well as C-H...O\textsubscript{W} hydrogen bonds. The inclusion of C-H...O/N hydrogen bonds in refinement algorithms would certainly improve the interpretation of experimental electron densities.

THE NMR PERSPECTIVE

In contrast to crystallography, which allows one to locate and estimate occupancy factors of specific hydration sites, but does not allow one to distinguish between strongly-bound and rapidly exchanging water molecules, NMR can provide information about the residence times of specific solvent molecules [57-59]).

For DNA duplexes, only a limited number of water molecules with long residence times could be detected by NMR. Such water molecules were observed in the minor groove of the well described d(CGCGAATTCCGCG) dodecamer [60,61], but also for the d(GTGGAAATCCAC), the d(GTGGTTAACCAC) dodecamers [62], and the d(AAAAAATTTTT) decamer [61]. These water molecules were proposed to correspond to those which form the spine of hydration observed in high-resolution B-DNA crystal structures. Their residence times were estimated to be significantly longer than one nanosecond while the residence times of water molecules located in the DNA major groove were inferred to be notably shorter than 500ps [61]. For DNA, sequence dependent effects were observed by NMR. A comparison of the dodecamers d(GTGGAAATCCAC) and d(GTGGTTAACCAC) revealed that an ordered spine of hydration with water molecule residence times over 500ps were associated with the sequence 5'-dAATT, whereas the sequence 5'-dTTAA kinetically destabilized minor groove hydration [62]. Subsequent studies emphasized the high sensitivity of water-DNA NMR signals towards small conformational differences [63]. From these data, it was inferred that the hydration
lifetimes were determined by the cumulative effect of interactions between more than just four or five consecutive base pairs. For an acidic form of a short d(TCGA) oligomer, a large number of water molecules, whose residence times were estimated to be longer than 500ps were observed close to the AH$_8$, AH$_2$, CH$_5$, CH$_6$, and T methyl protons as well as near the sugar protons [64]. Noteworthy, NOE between protons of water molecules and those of thymidine methyl groups located in the major groove were generally observable as well.

Using the nuclear magnetic relaxation dispersion (NMRD) method, Denissov et al. [65] proposed a residence time close to one nanosecond for five water molecules located in the minor groove of the d(CGCGAATTCCGCG) B-DNA duplex at 4°C, in agreement with results obtained by nuclear Overhauser effect (NOE) measurements. At 27°C, the residence time of these water molecules was estimated to decrease to 0.2ns. With the NMRD method, it was noticed that no long-lived water molecules were detected in the vicinity of the well-hydrated phosphate groups of the B-DNA structure. This result is remarkable, since water molecules in the vicinity of the phosphate groups cannot be observed by the more classical NOE method given the lack of observable protons close to these groups.

The hydration of DNA triplexes was also investigated by NMR [66-68]. Signals connecting the H$_8$ protons with those of water molecules revealed the presence of an organized pattern of water molecules in two of the three grooves.

Considering the greater methodological difficulties, it is not surprising that only one report, which describes the hydration of the RNA r(CGCAAUUUGCG) oligomer using NMR, has been published by Conte et al. [69]. The RNA shallow (minor) groove, despite its increased width, was found to be better hydrated than the DNA minor groove of an analogous DNA sequence probably due to the presence of the RNA hydrophilic 2'-hydroxyl groups, which may help to stabilize some water molecules.

For protein-nucleic acid complexes, it is recognized that water molecules play a major role in the binding specificity and affinity [70]. By NMR, long-lived water molecules were detected in the interface of protein-DNA complexes [71-73]. The NMR study of the protein-DNA interface of the antennapodia homeodomain complex [71] revealed a fluctuating network of hydrogen bonding interactions occurring between protein, DNA and water molecules partially supported by the crystallographic structure of the complex [74].

In particular instances, evidence concerning the presence of long-lived water molecules can be obtained indirectly by NMR from the exchange rate of specific nucleic acid protons. For example, the additional imino proton of the modified base pseudouridine found in RNA molecules displays an exchange rate with the protons of the solvent considerably slower than the exchange rate, which would be expected for a solvent exposed imino group. This result was interpreted in terms of either intramolecular hydrogen bonding or hydrogen bonding to a strongly bound water molecule [75-77]. Crystallography [78] and molecular dynamics simulations [79] confirmed that in certain structural contexts a strongly bound water molecule protects the proton of the additional imino group of the pseudouridine from exchange with water (see Figure 6 and 12).
From the preceding, it appears that NMR yields important results on the kinetics of the hydration of nucleic acids. These results are complementary to those obtained by crystallography. However, NMR methods suffer certain limitations. For example, some potential hydration sites may not be detected by using NOE methods because i) they are too distant from a proton; ii) they are close to rapidly exchanging protons; and iii) water-DNA NOEs are obscured by overlap with other strong NMR signals, such as those originating from terminal hydroxyl groups of DNA. Future developments of NMR techniques should probably result in the detection of new long-lived water molecules in the vicinity of nucleic acids, especially for A-DNA or RNA molecules, which are more strongly hydrated than B-DNA, and to a sharper estimate of their location and residence time. The study by NMR of the hydration of larger RNA molecules, comprising complex tertiary structural motifs, should also provide very useful information about the stabilizing role of water molecules.

**MOLECULAR DYNAMICS SIMULATIONS**

Besides crystallography and NMR, molecular dynamics (MD) simulations allow one to approach the hydration of nucleic acids from the structural and dynamical point of view. One of the advantages of MD simulations resides in its ability to model the dynamical behavior of atoms, which are rarely observable by experimental methods, such as the hydrogen atoms of hydroxyl groups or water molecules. A clear description of the motions of these atoms is of course essential, if one wants to understand the hydration of nucleic acids. Several reviews summarize earlier results obtained by this method [5,80-82] as well as recent developments of the MD methodology applied to nucleic acid systems [83-86].

**MD SIMULATIONS OF DNA**

As for crystallography and NMR, the d(CGCGAATTCCGCG) oligomer can be considered as a paradigm system for nucleic acid MD simulations. The reproduction of the experimentally observed spine of water molecules found in the minor groove of DNA was an interesting test for the MD methodology. This spine was successfully simulated by using static DNA models [87,88]. Recent simulations in which the motions of the DNA molecule were taken into account also reproduced the spine of hydration in the AATT region [89,90]. Solvent densities close to the DNA bases were derived from the simulations [89]. These densities are in partial agreement with those derived from crystallographic data (Figure 4). A detailed analysis of the hydration of the d(CCAACGTGTTGG) oligomer has been reported [91]. Accordingly, the highest density of water molecules was located in the minor groove, where water molecules directly hydrogen bond to the nucleic acid and form a spine of hydration (Figure 10). A lower level of hydration was observed in the major groove suggesting an increased mobility of water molecules and reduced residence times in agreement with NMR data [61].
Figure 10. Stereoview of the average structure and the hydration of a B-DNA d(CCAACGTTGG) oligomer derived from an MD trajectory. The water spine in the minor groove is clearly visible. The contours of the water oxygen density into 0.5 Å³ grid elements are displayed. The contour at a level of 12 hits per 0.5 Å³ and 15 hits per 0.5 Å³ are shown in (a) and (b) respectively. The figures are adapted with permission from Cheatham and Kollman [91].

Yet, one of the most interesting results from these MD simulations suggests that ions may intrude into the grooves of DNA and coordinate with the bases through direct or water mediated interactions [83,90,92]. For the d(CGCGAATTTCGCG) oligomer, an ApT sodium binding pocket was detected. The Na⁺ ion was found to interact favorably with the thymine O₂ atoms and is well articulated with the water molecules, which form the remainder of the minor groove. Ions were also observed in the minor groove of the d(CCAACGTTGG) oligomer [91]. This study revealed further that B-form nucleic acids may favor ion binding in the minor groove, while A-form structures favor ion binding in the deep (major) groove. Support for the fractional occupancy of hydration sites by ions comes from an NMR study of DNA sequences containing A-tracts which demonstrated that divalent cations could bind to sites located in the DNA minor groove [93].
Figure 11. The interaction pattern of GC.T base triple in DNA triple helices (left), as derived from NMR studies [68], and (middle) as observed in a MD simulation [98]. Note the close correspondence between the NMR and the model structure of the GC.T base triple. Hydrogen bonding pattern of a regular GC.G base triple from a model structure (right). Drawings are adapted with permission from Radhakrishnan and Patel [68] and Weersinghe et al. [98].

Water spines similar to those observed for B-DNA were detected in MD simulations of model-built DNA triple helices and were interpreted as contributing to the stabilization of the DNA architecture [94-97]. The residence times of some of these water molecules were found to exceed 500ps in agreement with NMR studies of DNA triplexes [66,67]. Significant association of Na\(^+\) ions to the N\(_7\) atoms of guanines in one of the grooves of a triple helix has also been reported [96]. An MD simulation of a model pyrimidine-purine-purine DNA triple helix incorporating a GC.T mismatch [98] revealed that a water molecule stabilized the base triple by establishing water mediated base-base interactions (Figure 11). The hydrogen-bonding pattern of the GC.T mismatch, which accommodates a water molecule, is similar to the one observed by NMR [68]. Subsequently, spines of water molecules were also found in both minor grooves of a model d(T\(_{10}\).d(A\(_{10}\).d(T\(_{10}\) triple helix [99].

MD simulations are also beginning to investigate the much larger protein-DNA complexes. Bilester et al. [100] calculated nanosecond residence times for water molecules located at the protein-DNA interface of the antennapodia homeodomain complex. These water molecules may contribute to the binding specificity. Similar results showing immobilized interfacial water molecules were obtained for the binding of an estrogen [101] and a glucocorticoid receptor [102] to DNA.
MD SIMULATIONS OF RNA

RNA molecules, given their ability to form complex tertiary folds which bring charged phosphate groups in close proximity, display an increased sensitivity to simulation protocols compared to DNA molecules and, thus, have been less studied [85,86].

A detailed account of the hydration of the r(CCAACGUUGG) oligomer revealed a specific association of water molecules in the RNA deep groove and a transversal motif of water molecules linking the 2'-hydroxyl groups across the shallow groove [91] consistent with those observed by crystallography [28]. The strong hydration of the deep groove may result from the narrowing of its walls associated with the rotation of the phosphate groups toward each other. In addition, the RNA shallow groove was found to be less hydrated than the corresponding minor groove of B-DNA.

The hydration of the yeast tRNAAsp anticodon hairpin formed by a loop of seven bases connected to a stem of five base pairs was investigated by MD simulations [26]. A detailed description of the hydrogen bonding percentage of water molecules to RNA atoms revealed that each anionic oxygen atom of the phosphate groups, when not involved in intramolecular hydrogen bonding, was interacting with close to three water molecules in agreement with the hydration cones observed by crystallography [5]. The exposed hydrophilic base atoms, not involved in intramolecular contacts were also well hydrated. Besides atoms involved in base-pairing interactions, the hydrophilic atoms, which do not form hydrogen bonds with water molecules due to their location are the backbone O5' the sugar O4', and the purine N3 atoms. From crystallographic studies it is well known that the O5'-atoms are involved in C6-H6...O5' or C8-H8...O5' intranucleotide hydrogen bonds and are, therefore, shielded from the solvent [1,2,46,103]. Similarly, the ribose O4' atoms are shielded from the solvent by forming interresidue C2'-H2'...O4' hydrogen bonds with the C2'-H2' group of the adjacent ribose [52]. More surprisingly, the purine N3 atom does not form hydrogen bond contacts with water, although a strong hydration site is observed in the vicinity of this atom in crystal structures [15,16].

In the MD simulation, long-lived water molecules were essentially found in the vicinity of phosphate groups. They are mainly involved in bridging interactions between anionic oxygen atoms of adjacent phosphate groups, as observed in crystal structures of A-form nucleic acids (Figure 12).

The most stable water molecule found in the MD simulations of the tRNA anticodon hairpin [26,79] forms a bridge between the anionic oxygen atoms of two adjacent phosphate groups and makes a third hydrogen bond with the imino group of a pseudouridine (Figure 12). The residence time of this water molecule was found to largely exceed the 500ps time limit of the MD simulation. Crystallographic [78] and NMR studies [75-77] confirm the participation of such water molecules in the stabilization of RNA motifs involving pseudouridine. Thus, these water molecules are part of a “nucleotide-water” complex. It ensues that the structural implications of the transformation of a uridine into a pseudouridine at particular locations of RNA molecules are to stabilize specific tertiary motifs by
Figure 12. Snapshot extracted from a 500 ps MD simulation of the solvated tRNA$^{Asp}$ showing three long-lived water bridges between anionic oxygen atoms of adjacent phosphate groups [26] (left). Snapshot extracted from a 500ps MD simulation of the solvated tRNA$^{Asp}$ anticodon hairpin showing the water molecule which links $\Psi_{32}$ base to its nucleotidic backbone through a N$_{1}$-H$_{1}$...O$_{W}$ and two O$_{W}$-H$_{W}$...O$_{R}$ hydrogen bonds (right). The time course of the $\langle \Psi_{32} \rangle$ O$_{4}$_{$\bullet$}...N$_{4}$ (C38) contact linking the two bases indicate that this interaction is stable, most probably as a result of the hydration pattern induced by the pseudouridine base. Note that the water molecule bridging the pseudouridine N$_{1}$-H$_{1}$ group to its backbone is also observed in crystal structures (see for example Figure 6).

forming long-lived hydration patterns. In addition, a water molecule interacting with $\langle \Psi_{32} \rangle$ N$_{3}$-H$_{3}$ is often observed in a bridging position between the base 32 and 38 [26].

As the RNA 2'-hydroxyl group is at the origin of the profound structural and dynamical differences between DNA and RNA molecules, the study of its hydration is of great interest. In agreement with crystallographic data [28], water molecules located in the shallow groove were found to be hydrogen bonded to these hydroxyl groups. The hydration of the RNA 2'-hydroxyl groups was investigated by MD simulations [52]. It was found that water molecules form three well defined clusters around these groups (Figure 13), which are similar to those observed in a high resolution crystal structure of an RNA duplex [28]. Yet, these water molecules are not long-lived and form labile hydrogen bond patterns [26]. Thus, the hydration of the RNA shallow groove is more dynamic than that of the deep groove. This result is consistent with the water-like behavior of the 2'-hydroxyl group deduced from MD simulations [52]. When not involved in the stabilization of specific tertiary motifs, a dynamical equilibrium between the three allowed orientations of the 2'-hydroxyl group is observed. Each of these
orientations faces the same three hydration clusters. Thus, rotation of the hydroxyl group does not involve the formation of new hydration sites, a rather energetically unfavorable process, but results in a dynamical hydration, which involves rapid exchange of water molecules and constant rearrangement of the hydrogen bond network. It was inferred that these hydration sites, occupied by water molecules exchanging with the bulk, constitute after dehydration anchor points for specific RNA/RNA or RNA-ligand interactions.

Besides the hydration of hydrophilic RNA atoms, the hydration of so-called "hydrophobic" C-H groups was investigated by MD simulations [56]. As stated above, intramolecular C-H...O as well as C-H...O\textsubscript{w} interactions are systematically
observed in chemical and biochemical systems. From the MD simulations of tRNA fragments [56], water molecules involved in long-lived bridging interactions between anionic oxygen atoms, form also hydrogen bonds with pyrimidine C5-H5 groups in order to complete their tetrahedral coordination (Figure 14). Such interactions are systematically observed in high resolution A-DNA and RNA crystal structures (Figure 7). In addition to these "bona fide" C5-H5...Ow hydrogen bonds, C3'-H3...Ow contacts were described. Water molecules were shown to be involved in a large variety of structurally important interactions. Yet, for correct folding and activity of RNA molecules, monovalent [104-106] and divalent cations [4,107] are required. As shown for DNA, strong binding sites for monovalent cations (NH4+) were found in the loop of the tRNA anticodon hairpin [26]. These sites are specific to NH4+ cations and would probably not accommodate totally or partially hydrated Na+, K+, or Mg2+ cations. The binding abilities of these ions to nucleic acids are currently investigated by several groups using MD simulations.
CONCLUSION

It is now accepted that water molecules and ions form an integral part of nucleic acid structure [10]. Our knowledge of the structural and the dynamical properties of the hydration shell at short and long distance around biopolymers is becoming more general and precise. The increasing amount of high resolution crystal structures of DNA and RNA molecules generates data which, put together, will lead to precise databases storing information on the hydration of specific structural elements. In this perspective, the concept of hydrated building blocks, developed by Schneider and co-workers, is particularly attractive [15-17,108]. The improvement of NMR techniques will lead to a sharper estimate of the residence times of loosely as well as strongly bound water molecules and allow to explore regions of the NMR spectra which are, yet, not well resolved.

The preceding discussions emphasize the fact that MD simulations are able to reproduce, at least qualitatively, essential structural properties and hydration features of nucleic acids. MD simulations are especially successful in distinguishing between the hydration sites occupied by strongly bound water molecules and those occupied by frequently exchanging ones. This ability to visualize, at the atomic scale, the motions of such complex systems is of considerable theoretical and practical interest. Firstly, because it helps to understand the dynamics of the hydration shell which plays a great role in structuring, folding, and association processes. Secondly, because it can help in the interpretation of crystallographic and NMR data. Yet, the accuracy of MD simulations is based on the quality of the empirical force fields on which they are based and, thus, precise experimental results are needed in order to increase the confidence level of long molecular dynamics simulations.
REFERENCES


