Nucleic acids are polyanionic molecules that were historically considered to be solely surrounded by a shell of water molecules and a neutralizing cloud of monovalent and divalent cations. In this respect, recent experimental and theoretical reports demonstrate that water molecules within complex nucleic acid structures can display very long residency times, and assist drug binding and catalytic reactions. Finally, anions can also bind to these polyanionic systems. Many of these recent insights are provided by state-of-the-art molecular dynamics simulations of nucleic acid systems, which will be described together with relevant methodological issues.

**Introduction**

A thorough description of the structure of the solvent shells surrounding nucleic acid systems is important for understanding most molecular recognition processes, ranging from folding and assembly of proteins and nucleic acids to the binding of small ligands. To investigate these subtle aspects, molecular dynamics (MD) simulations are a technique of choice that confirms the inspiring, experimental statement that ‘water is an integral part of nucleic acids’ [1,2]. Hence, recent experimental and theoretical perspectives have recently been published. Several reviews and a book [7] on MD simulations [8] and on the solvation of nucleic acids [10–12] and are also likely to be of interest to the reader.

**Solvent: a structure-stabilizing component and a guide to folding**

**Water**

It is now well accepted that solvent plays a key role in the stabilization of biomolecular systems in general and nucleic acids in particular. This has been very nicely exemplified by some of the first nucleic acid MD simulations that explicitly took into account the solvent. For instance, an MD simulation of a DNA dodecamer in in vacuo conditions (absence of solvent) revealed rapid unstructuring of the duplex on a 100 ps timescale. However, when explicit solvent particles were taken into account, the stability of the system increased significantly, emphasizing the structural role of the solvent [13]. Hence, in silico experiments in which interactions associated with important components of natural systems are switched off (in vacuo) or on (in aquo) illustrate the usefulness of MD techniques in evaluating the effects linked to the presence or absence of the solvent.

Similar conclusions could be drawn from in silico experiments in which the effects related to long-range electrostatic interactions were evaluated. These simulations demonstrated that the inclusion of long-range interactions involving the solvent contributed decisively to the stability of nucleic acid systems. Neglecting them leads to rapid unfolding of key structural motifs, such as tRNA hairpin loops [14,15]. These interactions, also called hydration or solvation forces [16], play a key role in the statics and dynamics of biomolecular systems.

Recent studies exemplified the role of the solvent in the folding of proteins [17] and nucleic acids [18], proposed that “biomolecules have evolved to use water to help guide folding”, and suggested that long-range water-mediated potentials assist folding by smoothing the underlying folding funnel [19,20]. For instance, ensemble MD simulations, with nearly 500 µs of aggregate simulation time, using an explicit representation of the ionic solvent suggested that it is necessary to account for...
for water-mediated interactions to accurately characterize the free energy surface and stochastic nature of folding. At the same time, the simulations helped to precisely define the limitations of the GB/SA (generalized Born/surface area) continuum solvent models [18].

In a similar manner, molecular recognition is often viewed from a static point of view, whereby structures are already folded and complexes formed. However, an often overlooked and not well understood aspect of molecular recognition is related to long-range interactions. A recent study suggested the existence of a solvent site-dipole field connecting a DNA oligomer and a small peptide even at a distance exceeding 10 Å [21]. Similar dipole fields were computationally observed around spherical hydrophilic solutes [22] and might, at least partially, be at the origin of some of the measurable long-range forces occurring in biomolecular association processes [16].

**Monovalent and divalent cations**

Besides water molecules, monovalent (e.g. Na⁺, K⁺) and divalent (e.g. Mg²⁺) ions play a significant role in the folding of biomolecular systems and, especially, in the folding and stabilization of RNA systems [23]. In this regard, MD simulations have helped to understand the mechanism through which hydrated Mg²⁺ cations stabilize RNA molecules (Figure 1) by way of specific water-mediated contacts [24]. Furthermore, MD simulations have led to a more intricate and precise view of the role cations play in nucleic acid systems [25]. Indeed, a few electronegative sites enable specific binding of monovalent or divalent cations [23]. Other electronegative sites, of a ‘non-specific’ type, get saturated by divalent species, such as hexahydrated magnesium cations, at concentrations well above physiological levels, as often observed in experimental conditions (e.g. crystallography). In vivo, by contrast, such sites are likely to be ‘saturated’ by the more easily dehydrated monovalent cations. Hence, ‘mixed’ models including monovalent and divalent cations have been proposed [25].

**Anions**

As stated above, our current views on solvation embrace water, and monovalent and divalent cations, but, despite early reports [26,27], do not generally take into account the possibility that anions (e.g. Cl⁻, SO₄²⁻) might bind to polyanionic nucleic acids. MD simulations of the 5S rRNA loop E motif, at a 1.0 M concentration of excess KCl, revealed numerous binding sites for Cl⁻ ions close to electropositive nucleotide edges [28]. Similar binding sites were subsequently found through an exhaustive survey of crystallographic structures deposited in the Nucleic Acid Database (NDB) and the Cambridge Structural Database (CSD) (Figure 2). This led to the construction of a nucleic acid anion-binding map. Currently, the biological roles of anions in DNA and RNA systems are not understood. Nevertheless, it is important to realize that these species can interact with and are recurrently observed in nucleic acid structures. Moreover, there is a clear correspondence between nucleic acid anion-binding sites and the nucleic acid binding sites for the two negatively charged aspartic and glutamic amino acids, nucleic acid phosphate groups and negatively charged chemical groups belonging to various small ligands. Such data can be of help for a better understanding of folding and molecular recognition phenomena. Moreover, because it has been proposed that “biomolecules have evolved to use water to help guide folding”, it should not be too surprising to ‘discover’ that nucleic acids have evolved to use monovalent and divalent cations (most likely K⁺ and Mg²⁺) as well as physiological anions (such as Cl⁻ and SO₄²⁻) for folding and stabilization purposes.
Nucleic acid systems

DNA

Some large-scale projects related to DNA systems are currently being pursued. For instance, nanoscale and atomic-resolution MD simulations of the nucleosome core particle surrounded by approximately 67,000 water molecules and ions have been reported [29]. At a different level, two groups continue their efforts in characterizing sequence context effects by analyzing the 136 unique tetranucleotide sequences [30–32]. Until now, most of the investigations have focused on conformational aspects. Interesting data related to the solvation of some of these 136 sequences will probably be made available in the near future [33].

Two MD studies were directed at improving our understanding of the B-to-A and B-to-Z transition mechanisms. For the former, MD studies of a 24 base pair double-stranded DNA oligomer were carried out to explore, among other aspects, the evolution of the hydration patterns on the surface of the DNA [34]. The authors reported that, although the total number of hydrogen bonds remained constant during the forced transition from B-to-A, the number of water bridges increased by approximately 40% and Na⁺ condensation was suggested to drive the conformational change. For the latter, targeted MD simulations proposed a physically plausible mechanism from which the impact of the ionic strength on the stability of the DNA conformations could be inferred [35].

Amazingly, earlier relatively short MD studies (~1.5 ns) suggested that ions might intrude into DNA grooves [36]. This proposition prompted researchers to develop new and more accurate crystallographic methods that confirmed, along with numerous MD studies, that monovalent cations bind to hydrophilic atoms located in the grooves. In this respect, the following issues are still of interest: do ions preferentially bind to the major or minor groove; does ion binding induce groove widening or narrowing; and what differences in binding can be expected between Na⁺ and K⁺ cations? This last issue is of importance because it is now widely recognized that K⁺ cations (~100 mM) dominate the cytoplasm of practically all living cells [37,38], whereas the less physiologically relevant Na⁺ cation has been used, mainly for historical reasons, in a majority of experimental and theoretical studies. Nanoscale studies using the CHARMM [39,40,41], AMBER [42–45] and GROMOS [46] force-fields focused on these issues. There is a consensus that these ions intrude into the major and minor grooves in a sequence-dependent manner, and that K⁺ cations are more mobile than Na⁺ cations. Furthermore, the presence of Na⁺ slows down the exchange of water near DNA hydrophilic sites compared with K⁺-containing systems. However, no real consensus seems to emerge regarding finer structural details, most probably because the calculated data are partially affected by force-field parameters associated with monovalent cations, as discussed in the methodology section below.

New methods for estimating residency times in the first shell of water molecules showed, in agreement with earlier studies [37,38], that water dynamics is reduced in both the minor and major grooves [47,48]. Moreover, these authors described techniques to estimate the...
entropy of water associated with DNA [49]. Besides these simulations using classical force-fields, simulations using a polarizable force-field for a DNA decamer in solution and in the crystal cell have been reported [50**].

The structure of DNA quadruplexes that are known to be intimately linked to monovalent cations is currently being investigated through MD simulations [51,52]. As expected, the results of these simulations are quite sensitive to the parameters used for the cations.

RNA
Currently, the largest MD simulations of nucleic acids with explicit representation of the solvent are focused on multimillion-atom systems, such as the ribosome [7,53**] and the satellite tobacco mosaic virus [54**]. In the latter study, water represented ≈84% of the 1 066 628 atoms forming the complete system and ions (Mg$^{2+}$, Cl$^-$) formed less than 0.1% of it. The viral capsid was fairly permeable to water during the first stages of these simulations. The calculated mobility of Mg$^{2+}$ and Cl$^-$ ions also differed dramatically. Whereas the majority of Cl$^-$ ions moved around freely, the Mg$^{2+}$ cations remained attached to the RNA, as expected.

On a ‘smaller’ scale, the Sponer group has actively contributed to the investigation of RNA systems through MD simulations [7,9]. One of their main goals is to understand the dynamics of key ribosomal fragments, such as the kink-turns [55–57], sarcin–ricin loops [58**], loop-E motifs and complexes [59,60], and of the Hepatitis delta virus ribozyme [61,62]. They showed that all these systems are associated with a unique network of water- and ion-binding sites that are intimately involved in modulating their conformational dynamics. They also described how these solvation patterns adapt to conformational changes observed in the calculated trajectories. For kink-turns, some water molecules with residency times over 5.0 ns occupy very specific binding sites, whereas most other sites were occupied by water molecules that exchange quite frequently and exhibit residency times in the 0.05–0.2 ns range, in agreement with earlier calculated data [37,38]. In the same studies, Na$^+$ cation binding times exceeding 20 ns were reported. In contrast to the large dynamic motions observed for the kink-turns, they reported that the sarcin–ricin loop [58**] was, like the loop-E motif [59], fairly rigid. Interestingly, some extremely unusual sugar flips and irreversible anti/syn flips of the apical base of a GNRA loop were reported. These sarcin–ricin and loop-E motifs were associated with very long-lived water molecules (>1 ns), whereas maximum calculated residency times for standard Watson–Crick helices are much shorter (<25 ps) [37,38]. Yet, compared to the loop-E motif [24,25,59], the sarcin–ricin loop is not a major cation-binding motif. MD simulations of the hairpin ribozyme emphasizing the role of the solvent in the active site were also reported [63].

MD simulations of the anticodon hairpin of tRNA$^{lys,3}$ with modified bases [64] reported relatively short residency times for a water molecule connected to pseudouridine 39 (≈40 ps), in contrast with earlier MD and NMR data showing that the N1-H proton of pseudouridine 32 is protected from exchange by interacting directly with a long-lived water molecule [65]. Conformational transitions in RNA single uridine and adenosine bulges were investigated through the use of MD free energy simulation techniques [66].

Ligand binding and drug design strategies
Besides water and ions, larger ligands also modulate the properties of nucleic acids by forming very specific and strong interatomic contacts. Studies related to the binding of small ligands to nucleic acid systems are quite rare.

DNA
For DNA, a long-standing interest has focused on the affinity and specificity of minor-groove-binding ligands. A courageous attempt was made to calculate the difference in the free energy of binding of netropsin (+2 charge) and distamycin (+1 charge) to a DNA sequence by using perturbation techniques coupled with an explicit representation of the solvent [67]. The authors were able to reproduce the preferential binding of netropsin over distamycin, and to describe their respective hydration patterns while simultaneously showing the limits of the methods used.

The importance of water in mediating the dynamic interactions between dicationic amidine minor groove binders and DNA bases was described [68]. The binding of the anticancer drug anthramycin to the DNA minor groove suggested that the electrostatic field created by the water molecules played a negligible role in the activation of the drug and, subsequently, the alkylation process, as compared with the field generated by the DNA [69].

Long-lasting water molecules observed at the interface of a protein–DNA complex were mainly located in the cleft between key residues and the core bases, whereas some of them occupied deep hydration sites in the DNA major groove [70,71]. It seems quite reasonable to propose that water molecules play roles in mediating sequence-specific recognition and/or enhancing the affinity of proteins for specific sites on DNA.

RNA
The binding of paromomycin, an antibiotic composed of four sugar rings from the aminoglycoside family, to the decoding site of the 16S rRNA has been investigated [72] and represents, with two other simulations of RNA–aminoglycoside complexes [73,74], the first set of MD simulations with explicit representation of the solvent around a small RNA-bound ligand (Figure 3). These
simulations suggest that the neamine part (two sugar rings) of the antibiotic represents the main anchor part, whereas the additional sugar rings provide more limited and fragile contacts. Moreover, the simulations suggested that long-residency water molecules present at the drug–RNA interface are involved in the recognition phenomena \[72/C15\]. Information related to the position of such long-residency water molecules, which complements crystallographic data, could improve strategies for nucleic acid docking and drug design \[75\].

**Methodological considerations**

Some of the early methodological issues related to the treatment of long-range electrostatic interactions \[76\] have been solved \[77\] and we are now witnessing the outbreak of nanosecond-scale simulations \[29/C15,53/C15,54/C15\], setting new limits for future investigations of nucleic acid systems. Besides the issues mentioned below, current methodologies and force-fields have been quite successful at solving or, alternatively, helping to grasp details of a few key biomolecular processes \[7,12\]. However, MD simulations are based on empirical force-fields that have to be constantly improved to extend the range of applicability of these methods in terms of both system size and accessible timescales. Thus, it is necessary both to concentrate on the system set-up \[4/C15\] and to track, with great determination, all the major flaws associated with current MD methodologies.

**Force-field developments**

Comparative studies \[78\] set the basis of a force-field-dependent DNA polymorphism \[79\] by showing that the CHARMM22 force-field favors A-form over B-form DNA structures. Differences in the structures of the calculated solvation shells were also analyzed and reported \[27,80\]. A subsequent study assessed that simulations using CHARMM27 did not favor the B-DNA to A-DNA transition, as was observed with the CHARMM22 force-field \[40\]. Recent simulations demonstrated the accumulation of ‘non-canonical’ α/γ backbone torsion angles towards the end of a 50 ns trajectory \[44\] (see also \[31\]), a concern that will be addressed in future versions of the AMBER force-field. Similar issues have been reported for proteins \[81,82\]. Indeed, force-fields evolve under the pressure of uncovered methodological artifacts. Currently, several such MD packages and associated force-fields co-exist and co-evolve \[46,83–85\]. Guidelines for future developments are already in sight and include more refined water models, improved treatment of charge multipoles and inclusion of polarization \[8/C15,50/C15\].

**Ion clusters observed with AMBER monovalent ion parameters**

Although clear progress in refining existing force-fields has been made in recent years, these parameter sets are so large and complex that important problems could remain hidden for a very long time. Indeed, this is the case for the monovalent ion parameters that originated from early free energy perturbation calculations and are distributed with all current versions of AMBER. Recent simulations (Figure 4) reported that, when these parameters are used, ionic clusters start to form and grow rapidly in the vicinity of biomolecules even at physiological concentrations \(\approx 0.25 \text{ M of KCl} \) \[12,45,72/C15\]. When other parameters...
for monovalent cations are used (keeping everything else unchanged), no such ionic clusters are observed even on the nanosecond scale (see e.g. [72,80]). Hence, AMBER users should be especially careful about the quality of parameters associated with these ubiquitous and ‘simple’ ionic species.

**Minimal salt conditions**
The formation of ion clusters has probably been overlooked for a long time, because most AMBER users working on nucleic acid systems agreed to use minimal (charge-neutralizing) salt conditions, contrary to users of other MD packages. There are several rationales for this decision (provided in [61] and associated supplementary material) based on the assumption that the inclusion of Cl\(^-\) anions in simulations could be risky because these anions are highly polarizable species. The data mentioned above, however, point to a significant deficiency in the widely used AMBER monovalent cation parameters. Besides, recent MD simulations successfully reproduced the nucleic acid anion-binding sites observed in crystallographic data [28], suggesting that anion (Cl\(^-\)) parameters, although certainly far from being perfect, can be used in MD simulations to model some aspects of the ionic atmosphere and to provide some hints about their main characteristics. More generally, the question whether the monovalent (and divalent) cation parameters used so far might have influenced our views on cation binding to nucleic acids remains to be clarified through comparative analysis [40,78].

**Molecular dynamics web-based tools**
Some emphasis is placed nowadays on creating web services and designing strategies enabling the storage and analysis of the large amount of digital data generated by MD simulations [86,87] or even multiple molecular dynamics (MMD) simulations [88]. The currently available web analysis tools focus on conformational aspects. Tools designed for characterizing and comparing solvation features in various systems are probably under development. Such databases might also enable efficient comparison of simulations generated using different force-fields and assessment of the progress made over the years in understanding nucleic acid systems.

**‘Experimental issues’**
**Importance of starting conditions and correction of experimental data**
In most cases, experimental structures (e.g. crystallographic, NMR) are used to initiate MD runs. Unfortunately, some of these structures are not as accurate as one might wish, especially with respect to the interpretation of crystallographic solvent densities. Therefore, it is wise to check them before starting an MD run [4*]. For instance, some Mg\(^{2+}\) cations have been assigned to electron densities located at electropositive sites where Cl\(^-\) or SO\(_4\)^{2-}\) anions would preferentially bind [28]. In other crystallographic structures, water molecules were assigned to sites with very large density-index values [89]. Consequently, these water molecules might have been incorrectly assigned to positions corresponding to metal ions.

Misassignments might significantly alter the quality of subsequent statistical database searches (see below; [28,90**]) and are likely to push MD simulations into frustrating dead-ends [28,90**]. In this respect, it is important to note that the Protein Data Bank (PDB) currently allows the documented deposition of corrected crystallographic structures. For example, the 1JJM and 1JKN structures from the PDB were replaced by the ‘reinterpreted’ 2B8R and 2B8S structures.

**Validation through statistical surveys of structural databases**
Hence, it is necessary to develop tools enabling both the starting structures to be checked [4*,89] and the validation of the generated MD trajectories. We recently developed a ‘Solvation web Service for nucleic acids’ (SwS) [90**], which is based on a statistical survey of crystallographic structures deposited in the NDB. SwS is designed to provide an exhaustive overview of the solvation of nucleic acid structural elements through

![Figure 5](image-url)

Comparison of water-binding sites derived from (a) MD simulations of a solvated r(CG)\(_{12}\) duplex [37] and (b) a statistical analysis of r(G\(\cdot\)C) pairs extracted from all NDB nucleic acid structures with resolutions equal to or below 3.0 Å using the SwS web service ref. [90**].
the generation of three-dimensional solvent density maps. An example of validation of MD data is shown in Figure 5. However, as described in the preceding sections, such databases also contain inaccurate information concerning, for example, misassigned solvent densities. It is our hope that such inaccuracies will not emerge from the noise level of statistical data and will be corrected over time. Ultimately, however, users need to be critical about data provided by this and similar database analysis tools.

Conclusions
The usefulness of MD simulations in providing a better understanding of microscopic biomolecular events through the generation of dynamic models has been assessed in a large number of MD reports. In particular, some landmark studies demonstrated that solvent composed of water molecules, monovalent and divalent cations, and anions is an integral part of nucleic acid systems. Neglecting them generally leads to microscopic ‘catastrophes’, such as severe disruption of three-dimensional structures, which can occur on various timescales. The improvement of the balance between interatomic forces is a prerequisite to obtaining more meaningful data on longer timescales for any type of biomolecular system.

Acknowledgements
The authors wish to thank Neena Grover for reading the manuscript and for interesting discussions, and Eric Westhof for ongoing support.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


4. Auffinger P: Molecular dynamics simulations of RNA systems: importance of the initial conditions. In Computational Studies of DNA and RNA. Edited by Spomer J, Lankas F. Springer Verlag; 2006:293-300. [Lesczynski J (Series Editor): Challenges and advances in computational chemistry and physics, vol III]. This chapter provides a complete list of ‘all’ MD simulations of RNA systems up to October 2005 and ‘details’ important aspects that have to be taken into consideration before starting a nucleic acid MD simulation.


Hydration features of nucleic acids are described in this report and in [48,49], along with interesting methods enabling nucleic acid analysis from a structural and dynamical point of view.


An important contribution describing the use of a polarizable force-field for simulating DNA duplexes in various solution and crystal environments.


A comprehensive review related to large-scale simulations of RNA systems, with a special emphasis on the ribosome particle simulations conducted by this group.


One of the most impressive nucleic acid simulations described to date.


A report extracted from the work of this group, related to the dynamics of ribosomal fragments, describing hydration and ion binding features, as well as intriguing anti/syn flips of apical nucleotides belonging to tetraloop hairpins.


64. McCrate NE, Varner ME, Kim KL, Nagan MC: Molecular dynamics simulations of human tRNA Lys 3UUU: the role of modified


One of the rare MD simulations of a ligand bound to an RNA system. Solvation features of this complex and methodological issues are described.


