
Studies of base pair sequence effects on DNA solvation based on all-atom molecular dynamics simulations

SURJIT B DIXIT¹, MIHALY MEZEI² and DAVID L BEVERIDGE^{1,*}

¹*Chemistry Department and Molecular Biophysics Program, Wesleyan University, Middletown, CT 06457, USA*

²*Department of Structural and Chemical Biology, The Mount Sinai School of Medicine, New York, NY 10029, USA*

*Corresponding author (Fax, 860-685-2211; Email, dbeveridge@wesleyan.edu)

Detailed analyses of the sequence-dependent solvation and ion atmosphere of DNA are presented based on molecular dynamics (MD) simulations on all the 136 unique tetranucleotide steps obtained by the ABC consortium using the AMBER suite of programs. Significant sequence effects on solvation and ion localization were observed in these simulations. The results were compared to essentially all known experimental data on the subject. Proximity analysis was employed to highlight the sequence dependent differences in solvation and ion localization properties in the grooves of DNA. Comparison of the MD-calculated DNA structure with canonical A- and B-forms supports the idea that the G/C-rich sequences are closer to canonical A- than B-form structures, while the reverse is true for the poly A sequences, with the exception of the alternating ATAT sequence. Analysis of hydration density maps reveals that the flexibility of solute molecule has a significant effect on the nature of observed hydration. Energetic analysis of solute–solvent interactions based on proximity analysis of solvent reveals that the GC or CG base pairs interact more strongly with water molecules in the minor groove of DNA than the AT or TA base pairs, while the interactions of the AT or TA pairs in the major groove are stronger than those of the GC or CG pairs. Computation of solvent-accessible surface area of the nucleotide units in the simulated trajectories reveals that the similarity with results derived from analysis of a database of crystallographic structures is excellent. The MD trajectories tend to follow Manning's counterion condensation theory, presenting a region of condensed counterions within a radius of about 17 Å from the DNA surface independent of sequence. The GC and CG pairs tend to associate with cations in the major groove of the DNA structure to a greater extent than the AT and TA pairs. Cation association is more frequent in the minor groove of AT than the GC pairs. In general, the observed water and ion atmosphere around the DNA sequences in the MD simulation is in good agreement with experimental observations.

[Dixit SB, Mezei M and Beveridge DL 2012 Studies of base pair sequence effects on DNA solvation based on all-atom molecular dynamics simulations. *J. Biosci.* **37** 399–421] DOI 10.1007/s12038-012-9223-5

1. Introduction

Solvation plays an integral role in stabilizing the structure of the DNA molecule in condensed phase. DNA–ligand affinity and specificity in solution is sensitive to solvation since solvent reorganization and release on binding make important thermodynamic contributions. The solvation of DNA at the most elementary level consists of water molecules and mobile ions, some of which are known to be ‘structured’ by

the interactions with the ionic, hydrophilic and hydrophobic sites on the sugar–phosphate backbone and in the major and minor groove of the double helix. Beyond this, the detailed molecular structure of DNA solvation depends on nucleotide base pair sequence, both via direct interactions and indirectly via sequence preferences for various DNA conformations. Thus, knowledge of the magnitude and extent of sequence effects on solvation is important in fully understanding DNA structure and function at the molecular level.

Keywords. ABC simulations; hydration of DNA; ion atmosphere; proximity criterion

Supplementary materials pertaining to this article are available on the *Journal of Biosciences* Website at <http://www.ias.ac.in/jbiosci/jul2012/supp/mezei.pdf>

The four canonical nucleotide bases in DNA give rise to 10 unique dinucleotide base pair steps. The structure of each base pair step may depend as well on sequence context, as in the case of the ubiquitous A-tracts implicated in DNA bending. Considering only first neighbour bases of each step, the minimum length of sequence required for detailed study of sequence-dependent DNA structure is a contiguous sequence of four nucleotide base pairs (tetranucleotide steps). Thus, a comprehensive description of sequence effects on DNA requires consideration of at least all 136 unique tetranucleotide steps, i.e. the 10 unique nucleotide base pair steps in all unique sequence contexts.

The available experimental data on DNA structure is fragmentary with respect to tetranucleotide steps, and does not yet encompass all unique dinucleotide steps, much less all sequence contexts. With the latest advances in high-performance computing, the problem of sequence effects on DNA structure and solvation can be addressed by molecular simulation. The method of choice is molecular dynamics (MD) simulation in which an all-atom model can be treated including explicit consideration of water and ions and provides a description of structure as a function of time. This article describes a study of sequence-resolved B-DNA solvation based on MD simulations on all 136 oligonucleotide tetramers. MD predictions of sequence-dependent DNA hydration and ion solvation are compared with available experimental data from diverse biophysical measures and from crystallography. The general features of DNA solvation are obtained, and a quite plausible description of the structures adopted by water molecules and ions in the vicinity of the B-form duplex is provided. Further analysis of the results reveals considerable new knowledge of the sequence-dependent nature of DNA solvation.

2. Background

2.1 DNA hydration

The importance of hydration on the structure of DNA was recognized quite early in the history of structural biology (Franklin and Gosling 1953). However, due to the complexity of the problem with respect to DNA structure and function, the nature of DNA solvation at the molecular level remains an open question even today. The literature on this topic is extensively described in texts (Saenger 1984; Westhof and Beveridge 1989; Berman and Schneider 1999) and articles (Leikin *et al.* 1993; Eisenstein and Shakked 1995; Umrana *et al.* 1995; Chalikian and Breslauer 1998; Egli *et al.* 1998). There is experimental evidence that the stability and conformational preferences of the DNA double helix depends on the relative humidity of fibres (Leslie *et al.* 1980) and water activity in solution (Malenkov *et al.* 1975). In fibres, the A-form of DNA is preferred under conditions of lower relative humidity (~76%), while increasing the relative humidity to ~86% stabilizes the B-form

(Saenger 1984). Osmotic stress experiments confirm that the water activity around DNA is directly correlated to the free energy required to deform the double helix (Ruggiero *et al.* 2001). A conformational transition from B-DNA to the A-form can also be induced by increased ionic strength (Nishimura *et al.* 1986). Excess salt in GC-rich sequences may cause a transition to the left-handed Z-form (Jovin *et al.* 1987). We restrict our attention here to the right-handed forms of DNA, essentially the family of B-form structures characteristic of right-handed duplexes. However, in B-form sequences of DNA, structures may locally adopt an A-form or be situated along the continuum of structures of right-handed DNA from A to B, and this is believed to have functional significance in DNA recognition. Thus, the hydration of A-form structures as well as B-form is relevant to this study. The terminology used to discuss DNA hydration is not consistent from one paper to another in the literature, and we take care herein to formally define our terms with the hope of more uniformity on this matter in subsequent research.

The basic idea that solvent water in the immediate vicinity of DNA adopts a structure incompatible to crystallization with bulk water was first reported by Kuntz (Kuntz *et al.* 1969), and this result subsequently posed questions about the nature and extent of water structures around DNA. There is now extensive experimental data on DNA hydration from molar volumes, densities and compressibilities (Chalikian and Breslauer 1998), dielectric relaxation (Umehara *et al.* 1990), neutron diffraction (Shotton *et al.* 1997) and molecular spectroscopy (Halle and Denisov 1998). The essential picture that emerges from this data is that the anionic, hydrophilic and hydrophobic constituents of DNA interact strongly with waters in the neighbourhood of the duplex and result in local water structures distinct from that of bulk water. This local structure extends a considerable distance from the surface of a DNA duplex and consists of regions in which the local density is somewhat higher than that of bulk water (Chalikian *et al.* 1994). The nature of DNA hydration is also reflected in the mobilities of the molecules and ions involved. Spectroscopic studies of ultrafast dynamics of DNA hydration indicate two time constants, one at ~1 ps behaving bulk-like, and the second at ~12 ps indicative of water molecules with reduced mobility (Pal *et al.* 2003). An issue of interpretation is that various experiments probe different regions and aspects of the structured hydration. In this article, we adopt the terminology that the structured hydration consists of solvation shells demarcated in principle by the peaks and valleys in DNA–water radial distribution functions. Within the boundary of the first solvation shell, some fraction (but now all) of the solvent water molecules and ions are in direct contact with DNA atoms, and referred to as ‘bound’.

The concept of structured hydration applies of course to both the A-form and B-form of DNA, but at a higher level of resolution there are significant differences. Spectroscopic,

gravimetric and volumetric studies indicate that the local hydration of a B-form duplex consists of ~30 water molecules per nucleotide base pair (bp) and about 28.7 structured waters per nucleotide in the A-form DNA. Dielectric relaxation measurements (Umehara *et al.* 1990) indicate at least 18–19 waters in the B-form and ~13–15 waters in the A-form. X-ray diffraction and optical microscopy measurements on fibre DNA indicate 27 water molecules per GC base pair and 44 waters per AT pair (Albiser *et al.* 2001). In these studies about 11 water molecules per nucleotide pair are observed in A-form structures irrespective of the sequence composition. Variations in the reported number of water molecules in the hydration layer are anticipated because of the differences in experimental settings and the fact that the experiments might be looking at different aspects of solvation properties. Buoyant density experiments on DNA fibres indicate that AT pairs involve 2 more water molecules than GC pairs (Tunis and Hearst 1968). This is in contrast of apparent molar volume and compressibility measurements (Chalikian *et al.* 1994), which indicate that GC base pairs are solvated more strongly than the AT base pairs in DNA sequences. Mixed AT/GC sequences while favouring B form DNA have been noted to be ‘less hydrated’ than homopolymers (Chalikian *et al.* 1994). DNA molecules are always hydrated to some extent, and thus the ‘extent’ or ‘degree’ of hydration need to be made more precise and quantitative to be useful. There is a well-established propensity of AT-rich sequences for B-form DNA and GC-rich sequences for A-form structures (Pilet and Brahms 1972). The correlation between the base sequence composition of DNA and preferential stability has been focus of a number of experimental and theoretical studies, and there is a considerable literature on A- and B-philicity scales (Mazur *et al.* 1989; Hunter 1993; Ivanov and Minchenkova 1994; Basham *et al.* 1995; Tolstorukov *et al.* 2001), which have proved to be useful in this study for validating the accuracy of MD on DNA.

The most detailed experimental data on DNA hydration at the molecular level comes of course from crystallography. Analysis of residual electron densities obtained from diffraction experiments yields information about waters which are crystallographically ordered, typically those which make direct non-covalent contacts with the DNA. One problem in establishing the significance of these observations is that crystallographically ordered water is a fraction of the total structured water around DNA but thermodynamic stability involves the entire consort. Schneider and Berman (Schneider *et al.* 1992, 1993, 1998; Schneider *et al.* 1993; Schneider and Berman 1995) have carried out a comprehensive statistical survey of crystallographic hydration of DNA based on optimally aligned electron density distributions for all available B-form oligonucleotide crystal structures. They have investigated the extent to which hydration sites show a high degree of transferability, and demonstrated that a

generic description of hydration localized at the base pair level could successfully predict the hydration of a dodeca-nucleotide sequence. Well-defined hydration sites are delineated for the hydrophilic donor and acceptors sites in the major and minor grooves of the duplex, cones of hydration around anionic phosphates.

Specifically, crystallographic hydration sites for B-form DNA were found in the minor groove at purine N3, pyrimidine O2 atoms and at purine N6/O6 and N7, and pyrimidine N4/O4 atoms in the major groove. Other major groove sites are purine C6 and G N2, which may be a shared site with C O2. These hydration sites also provide the critical functional groups required to make direct recognition at the protein-nucleic interface (Seeman *et al.* 1976), thereby linking the energetics of desolvation to the thermodynamics of binding and recognition (Privalov *et al.* 2007). While our focus in this article is on B-form DNA, within B-form structures, some steps, especially those with high GC content, may exhibit A-form-like local structures or structures between B-form and A, depending on sequence. Hydration and transition between different DNA forms affects the sequence-specific local deformability of DNA, which is another factor contributing to specificity during ligand recognition (Olson *et al.* 1998). Water molecules in A-DNA crystals reveal a highly ordered network with some performing the role of a stereospecific ‘handle’ in the recognition by cognate protein (Eisenstein and Shakked 1995; Schwabe 1997; Reddy *et al.* 2001). The crystallographic hydration sites are similar but more complex in A-form DNA, and in addition G O6 and C C5 are mentioned, the latter in alternating C/G sequences (Egli *et al.* 1998). In A-form structures, the typical inter-phosphate distance is optimal for accommodating a single water molecule bridging the two PO_2^- groups and it has been proposed that this so-called ‘economic’ use of a water molecule may be significant in relative stability of the A-form at lower water activity (Saenger *et al.* 1986). In case of B-DNA structures, the phosphate groups are separated to the extent that only the second shell water molecules can bridge the groups (Schneider *et al.* 1998). It is more common to observe water bridges between PO_2^- groups and hydration sites on nucleotide bases. Anionic phosphates dominate the hydration of the sugar-phosphate backbone at the expense of ester oxygen hydration.

In a number of cases, there is crystallographic evidence for hydration structures in which the sites form networks such as spines, filaments, polygons and even more elaborate two- and three-dimensional networked structures. The minor groove spine in AT-rich tracts of B-form DNA is the most celebrated of these hydration structures, and has been directly linked to thermodynamic stability (Chen and Prohofsky 1993), supported by the results of a novel grand canonical MC simulation (Guarnieri and Mezei 1996). Beyond this, there has been considerable speculation about the role of

higher-order solvation structures in DNA stability, but this is a complicated matter since the definition of these structures is not unique and also must be defined with respect to statistical weights to make a convincing argument. As a consequence, it has not generally been possible to link higher-order solvation structures unequivocally to thermodynamic stability, but this remains a desirable outcome.

2.2 DNA ion atmosphere

DNA is a polyanion in solution at neutral and physiological pH, and counterions are required to achieve electroneutrality. Typical experimental conditions and cellular environments involve even higher levels of ionic strength. Positively charged counterions associate with the negatively charged phosphates on the DNA backbone and, analogous to the structured hydration, exhibit a structured distribution encompassing the DNA surface and are generally referred to as the ‘ion atmosphere.’ The ion atmosphere was predicted theoretically (Manning 1978; Anderson and Record Jr 1982) and shown to generally account for a considerable amount of observed behaviour of DNA in solution. An important result from these studies is the idea of ‘counterion condensation’. The idea is that no matter how dilute the solution, a number of counterions remain in close proximity to the DNA as opposed to being dissociated as in the case of small ion pairs. Condensed counterions were predicted to compensate for a large fraction of phosphate charges, predicted by Manning to be ~76% and independent of concentration. This follows from a simple argument based on molecular thermodynamics (Jayaram and Beveridge 1990, 1991; Jayaram *et al.* 1991). Parallel studies of this problem carried out based on Poisson–Boltzmann calculations are in substantial accord with the basic ideas (Jayaram *et al.* 1989). The initial predictions about the nature and extent of the ion atmosphere of DNA were subsequently confirmed by computer simulation (Jayaram and Beveridge 1996; Young *et al.* 1997a, b) on specific sequence cases. The extent to which these results can be generalized and are sequence dependent is investigated in the present study.

Crystal structures provide evidence for site-bound ions (Tereshko *et al.* 1999) especially in the case of divalents, and in some structures charged organic co-crystallization agents such as spermine have been found in major groove regions (Robinson and Wang 1996). The possibility of some fractional population of monatomic inorganic ions in the grooves of DNA was initially predicted by MD simulation (Young *et al.* 1997a, b). This idea led to a lively debate in the literature (Shui *et al.* 1998; Chiu *et al.* 1999) since a significant groove population of ions would affect binding affinities. At this point in time, the fractional occupancies of ion involved seem to be small and sequence dependent (McConnell and Beveridge 2000; Rueda *et al.* 2004). The level that would

be significant with respect to conformational stability and ligand binding has not yet been established. We revisit the idea of groove populations of ions in the context of the results we have acquired in this study (*vide infra*).

2.3 Molecular simulation studies

Reviews are available of the earlier literature on computer simulation studies of DNA hydration (Feig and Pettitt 1998) and ion atmosphere (Jayaram and Beveridge 1996). Beyond the observations noted above, obtaining a comprehensive molecular model of DNA hydration and ion atmosphere from thermodynamic and spectroscopic measurements is a highly underdetermined problem, particularly since DNA solvation at a temperature must be viewed as a Boltzmann ensemble of thermally accessible structural forms. While computationally intensive, a complete description of the structure of DNA including solvation consistent with Boltzmann statistical thermodynamics can now be obtained theoretically from Monte Carlo (MC) and molecular dynamics (MD) computer simulations, subject of course to approximations intrinsic to the assumed force field and simulation methodologies and computational protocols. In studies aimed at validating all atom models of MD on DNA including explicit water and ions, comparisons of calculated DNA conformations with those observed from crystal structures (Young *et al.* 1997a, b) and from the NMR structures in solution (Arthanari *et al.* 2003) have been in remarkably close accord when the best force fields are used (Cheatham III and Young 2000; Perez *et al.* 2007a, b). A comparison of MD on DNA with time-resolved Stokes-shift experiments focuses specifically on the dynamics of water and ions near DNA (Sen *et al.* 2009). Extensive MD studies have been reported on DNA bending (Beveridge *et al.* 2004a, b) and flexibility (Orozco *et al.* 2008). The A to B transition is uniquely sensitive to water activity, and has been successfully described by molecular dynamics simulations (Cheatham III and Kollman 1996; Sprous *et al.* 1998; Knee *et al.* 2008). Most importantly in this study, the water molecules and ions included in these trajectories can be used to obtain calculated water and ion density distributions and thus extract a computational model of DNA solvation (Ponomarev *et al.* 2004). A number of recent articles have addressed some specific issues of DNA solvation based on MD results, including the minor groove spine of hydration (Subramanian *et al.* 1988; Guarnieri and Mezei 1996) and possible ion insertion into grooves of DNA (Young *et al.* 1997a, b). As noted above, a grand canonical MC simulation study (Jayaram and Beveridge 1991) and MD simulations of the ion atmosphere of DNA (Young *et al.* 1997a, b) generally supported the structure and energetics of counterion condensation theory and provided some predictions about sequence effects.

The use of Ewald summation method for the treatment of long-range electrostatics remarkably improved the quality of MD simulations on DNA (York *et al.* 1995). Subsequently, multiple simulations have addressed the issue of DNA hydration to various extents (Cheatham III and Kollman 1997; Duan *et al.* 1997). A particularly thorough study of the hydration of A- and B-forms of the pentameric duplex of sequence d(C₅T₅) in explicit water and 0.8 M excess NaCl was reported by Feig and Pettitt (1999). These simulations used both the AMBER and CHARMM force field, and the analysis was based on the shell structure of hydration obtained from a variation of quasi-component distribution functions using the proximity criterion (Mehrotra and Beveridge 1980; Mezei and Beveridge 1986). Their calculations indicate some 29 water molecules in the primary hydration shell in agreement with densitometric measurements, and found ~20 water hydrogen bonded directly to the DNA in agreement with dielectric relaxation, resolving the discrepancies between the two sets of experiments. The number of bound waters is decreased in A-form compared with B-form, and less for AT compared with GC base pairs. The calculated hydration densities generally match crystallographic hydration sites and provide evidence for up to three shells of hydration. Water molecules in the primary hydration layer are seen to be shifted and rearranged, but do not dehydrate the DNA. The effect of ions reduced the mobility but not the number of waters partially immobilized by the presence of ions. The agreement between MD results specifically on dC₅T₅ obtained by Feig and Pettitt and general experimental results is quite encouraging, and effectively points out the additional information about solvation of DNA at the molecular level that can be obtained from computer simulation.

Molecular dynamics simulations of canonical A-form DNA structure in water using the AMBER force field leads to a rapid transition of the solute structure to a qualitatively B-like structure (Cheatham III and Kollman 1996). Detailed analysis of the water structure around one such transitioning DNA structure has revealed that considerable delay, to the extent of 5–7 ns, is observed before the solvent structure reorganizes to the new DNA conformation (Stefl and Koca 2000). The convergence of ions in the course of a simulation can be an even slower process. Recent analysis of ion distribution around a palindromic DNA sequence indicates that simulations extending up to 100 ns might be required to achieve a symmetric ion distribution that such a sequence would call for (Ponomarev *et al.* 2004). Varnai and Zakrzewska (2004) have shown that Na⁺ and K⁺ counterions exhibit different groove binding propensities in the course of MD simulations with the Na⁺ binding to both grooves in a sequence-dependent manner while the K⁺ ion mainly visits the major groove. Rueda *et al.* (2004) have observed that the insertion of Na⁺ ions in the minor groove of DNA is a very rare event because of

thermodynamic reasons. They note that in the situation where ions bind inside the minor groove, they remain trapped for fairly long times by current routine simulation lengths (~10 ns) but do not dramatically influence the duplex structure. While the size of Na⁺ and K⁺ ions differ only slightly, their solvent shell display quite different behaviour. The g(r) between Na⁺ and water drops to near zero between the first and second shell, while the lowest value of the g(r) between K⁺ and water is of the order of 0.5. This behaviour was observed already in early simulations (Mezei and Beveridge 1981) and confirmed recently employing the force field used in this study (Mezei, unpublished). This difference translates into a fairly rigid solvation shell for Na⁺, since the 0.01 observed for the g(r) minimum translates into a barrier between the two shells of kT ln(0.01)~2.8 kcal/mol. At this point in time, MD simulations on a B-DNA of up to 1 μs have been dynamically stable (Perez *et al.* 2007a, b).

2.4 Continuum and integral equation solvation models

Recently, there have been attempts to model solvation in MD and MC simulations of DNA using dielectric continuum models based on Poisson–Boltzmann (PB) theory and/or the generalized Born (GB) approximation. While preliminary results were encouraging (Tsui and Case 2000), other calculations of this type show various degrees of pathology (Dixit *et al.* 2006) and efforts are ongoing to improve the quality of such simulations (Chocholousova and Feig 2006). An alternate application of the continuum solvation approach referred to as the MM PB/GB SA employs trajectories of the DNA derived from simulation employing explicit solvent to carry out *post facto* free energy analysis of the solvation thermodynamics (Jayaram *et al.* 1998; Srinivasan *et al.* 1998). Of particular interest to the present work is a study of the electrostatic component of solvation free energies of the 64 trinucleotides using numerical solution to the PB equation (Elcock and McCammon 1995). While based on canonical structures in B-form, the calculated solvation free energies were shown to be more favourable for GC relative to AT pairs. Recently, Howard *et al.* (2011) described an integral equation theory that replicates MD and the experimental results for the base-specific hydration patterns in both the major and the minor grooves and were able to discern characteristic structural pattern differences between Na(+) and K(+) ions.

2.5 MD on DNA: The ABC consortium

As noted above, studies of sequence effects on DNA structure including solvation have been limited by the necessity of considering all unique base pair steps in all unique sequence contexts. The computational task of obtaining MD

trajectories on all of these cases has recently been addressed by a broad-based collaboration of research groups specializing in MD on DNA, the Ascona B-DNA Consortium (ABC). An up-to-date review of the ABC initiative is provided elsewhere in this issue. The emphasis in the ABC project was to generate a database suitable for understanding the effect of flanking base pairs on the structural properties of all 10 unique dinucleotide steps (Beveridge *et al.* 2004a, b; Dixit *et al.* 2005; Lavery *et al.* 2010)). To thoroughly examine context effects, the ABC project generated MD trajectories for at least two copies of all 136 unique tetrameric steps. All simulations were carried out using the AMBER suite of programs and the *parm94* force field of Cornell *et al.* (1995), the first explicit solvent force field to provide a reasonable model of B-form DNA (Cheatham III *et al.* 1995). A relational data base was constructed to deal with the informatics aspects of analysis (Dixit and Beveridge 2006). The results of the ABC consortium indicated that while some specific steps are sensitive to sequence context effects, most steps are relatively unaffected. Specifically, while YpR steps widely implicated in DNA bending (Calladine 1982) are intrinsically the most flexible, they also are least affected by neighbouring base pairs. On the other hand, purine–purine and purine–pyrimidine steps, which are found to be relatively less flexible and in some cases such as A-tracts essentially straight and rigid, are the most sensitive to context effects. The ABC results also revealed a tendency in the *parm94* MD to undergo an irreversible conformational transition in the sugar–phosphate backbone parameters α and γ that traps the MD in a non-canonical conformational sub-state. Such α/γ transitions are observed in A-DNA (Gao *et al.* 1999) and are common in protein DNA crystal structures (Djurjanovic and Hartmann 2003; Sims and Kim 2003), but should be reversible. The AMBER force field for MD on DNA has subsequently been corrected (Perez *et al.* 2007a, b). In the study reported herein, the MD trajectories from the ABC project to date form the basis for analysis of the general nature and sequence dependence of the hydration and ion atmosphere of B-form DNA, including sequence context effects.

3. Methods

All simulations analysed in this study were carried out using the AMBER 6 or AMBER 7 suite of programs (Case *et al.* 1999). The details of these simulations have been reported previously by the ABC Consortium (Beveridge *et al.* 2004a, b; Dixit *et al.* 2005; Lavery *et al.* 2010). MD simulations were carried out on all oligomers with repeating tetranucleotide sequences (ABCDABCDABCD...). Moving a 4-base ‘reading frame’ along the oligomer, we located successively ABCD, BCDA, CDAB and DABC tetranucleotides. The length of the oligomers was chosen to be 15 base pairs, a

compromise between the necessity to avoid end-effects and the computational expense of the simulations. This strategy enables all 136 tetranucleotides to be studied based on only 39 independent MD simulations. The ends of each oligomer were capped with a single GC pair to avoid fraying. A given 15-base-pair oligomer thus contains 3½ tetranucleotide repeats 5'-G-D-ABCD-ABCD-ABCD-G-3', where A, B, C and D are one of A, T, G or C

In all individual simulations the solute molecule is a 15-base-pair oligonucleotide with 28 K⁺ (Aqvist 1990) ions added to achieve system electroneutrality. The DNA with its counterions was simulated in a truncated octahedral box having a face-to-face dimension of ~70 Å, which allows for a solvent shell extending for at least 10 Å around the DNA. The starting configuration has the oligomer in a canonical B-form structure. The ions are randomly placed around the oligomer, and located at least 5 Å from any atom of the solute and at least 3.5 Å from one another in the initial structure. The neutral ion–oligomer complex was solvated with TIP3P water molecules (Jorgensen 1981). Simulations were performed with periodic boundary conditions in which the central cell contained ~8000 water molecules. Considering the DNA, counterions and solvent water, the total system consisted of ~24000 atoms.

The preparations for MD simulations consisted of an initial minimization followed by slow heating to 300 K at constant volume over a period of 100 ps using harmonic restraints of 25 kcal mol⁻¹ Å⁻² on the solute atoms. These restraints were slowly relaxed from 5 to 1 kcal mol⁻¹ Å⁻² during a series of five segments of 1000 steps of energy minimization and 50 ps equilibration using constant temperature (300 K) and pressure (1 bar) conditions via the Berendsen algorithm (Berendsen *et al.* 1984) with a coupling constant of 0.2 ps for both parameters. The final segment consisted of 50 ps equilibration with a restraint of 0.5 kcal mol⁻¹ Å⁻² and 50 ps unrestrained equilibration. The simulations were then continued for a total of 15 ns at constant temperature and pressure conditions, using the Berendsen algorithm (Berendsen *et al.* 1984) with a coupling constant of 5 ps for both parameters. Electrostatic interactions were treated using the Particle Mesh Ewald (PME) algorithm (Essmann *et al.* 1995) with a real space cut-off of 9 Å, cubic B-spline interpolation onto the charge grid with a spacing of ~1 Å. SHAKE constraints (Ryckaert *et al.* 1977) were applied to all bonds involving hydrogen atoms. The integration time step was 2 fs. Centre-of-mass translational motion was removed every 5000 MD steps to avoid the methodological problems described by Harvey *et al.* (1998). The trajectories were extended to 15 ns for each oligomer and conformations of the system were saved every 1 ps for further analysis.

The program *ptraj* in AMBER contains procedures for generating many of the indices required for analysis of an

MD trajectory, and we utilized this wherever possible. Beyond this, calculations carried out in our own 'MD Toolchest' (Ravishanker 1998), and solvent-accessible surface area calculations were performed using the program Surface Racer 3.0 (Tsodikov *et al.* 2002). The van der Waals radii value reported by Richards (1977) was used for the solute atoms and solvent radius is set to 1.4 Å for all the surface area calculations. To calculate the solvent density about the DNA, we included the 2000 water molecules closest to the solute DNA atoms from the approximately 8000 water molecules surrounding the DNA in the simulation box. Following least square fitting of the DNA, the water oxygen positions were mapped onto a 3D grid at a resolution of 0.5 Å around the solute. The frequency of water occupancy at each of the grid points is a measure of the solvent density. The solvation density analysis was performed using the *ptraj* module of AMBER 8 (Case *et al.* 2004) and the graphics were rendered using Pymol (DeLano 2002). Further analysis of the MD-calculated solvation was carried out on the basis of the proximity criterion (Mehrotra and Beveridge 1980), which permits us to uniquely define the solution environment of each identifiable substructure-atom, functional group, or residue of any polyfunctional solute molecule or macromolecule (Mezei and Beveridge 1986). Specifically, the set of solvent molecules closer to a solute atom *A* than any other solute atom is referred to as the total primary solvation shell of *A*. In geometrical terms, this assignment is equivalent to partitioning the space around the solute molecule into the Voronoi polyhedra and assigning the solvent on the basis of its presence in the Voronoi polyhedra of the individual atoms (figure 1). While the geometric bisector plane between solute atoms is usually used for calculating the Voronoi polyhedra, we employed a charge-based radical plane for partitioning the space around each atom in these calculations. Such a charge-based radical plane partitioning has been shown to relate well to chemical intuition while calculating the solute–solvent energetics (Mezei 1988). Higher orders in the total solvation can further be defined, such as assigning the set of solvent molecules for which the solute atom *A* is the second nearest solute atom to its secondary solvation shell and so on. With the proximity indices thus defined for the solute and solvent molecules, one may develop an analysis of the solvation of a solute molecule atom by atom. All the proximity analysis calculations presented here were performed using the MMC program (Mezei 2006).

4. Results

The results presented in this article are based on the DNA simulation trajectories reported earlier by the Ascona B-DNA Consortium (Beveridge *et al.* 2004a, b; Dixit *et al.* 2005). The simulated data set involves 15 ns MD simulation

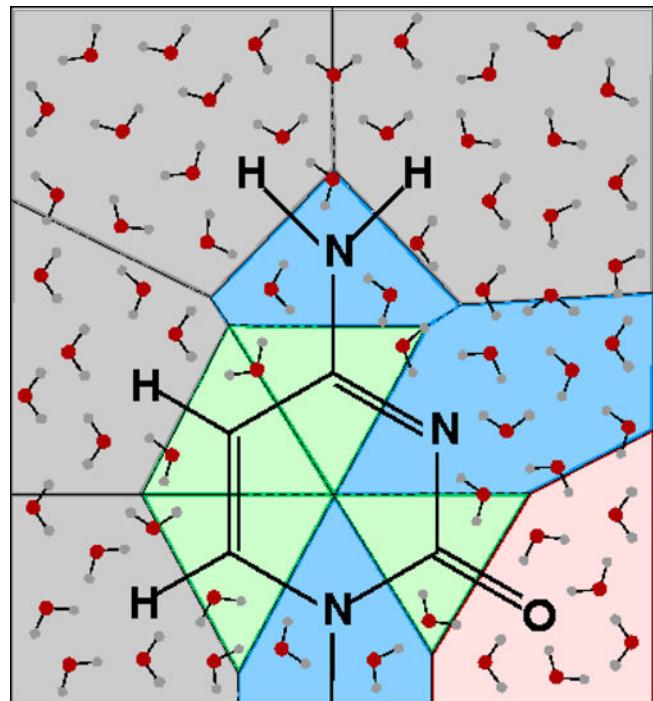


Figure 1. A schematic representation of the proximity region around the cytosine molecules in 2D. The proximity region around carbon is shown in green, nitrogen in blue, oxygen in red and hydrogen in grey. Note that in this figure the radical plane describing the Voronoi polyhedra of the individual atoms is represented by a bisector of the connection between two atoms. Alternately, radical planes on the basis of atomic radii or charge of the atoms involved can be described to generate chemically intuitive description of the proximity regions.

trajectories on 39 different 15-base-pair-long DNA sequences which comprise all the 136 unique tetranucleotides. Here, we first present a general analysis of the 39 simulations that highlights the fact that simulations perform fairly well in presenting the A/B-philicity of various DNA sequences correctly, then we present more detailed analysis of sequence-dependent solvation properties and finally describe the analysis of counter ion atmosphere around the DNA in these 39 simulations.

The RMSD analysis of the 39 ABC trajectories with reference to the canonical A- and B-form structures is shown in figure 2. The RMSD between any 15-base-pair-long canonical A- and B-form structure is about 7.5 Å. The observed RMSD values from the MD simulations signify that the structures are intermediate between the canonical A- and B-type structures with a sequence dependent preference for A- or B-like structures. The G/C-rich sequences are closer to canonical A- than B-form structures while the reverse is true for the A-rich sequences, with the exception of the alternating ATAT sequence. The poly A sequence is the most B-like, while the alternating ATAT sequence is most A-like. In

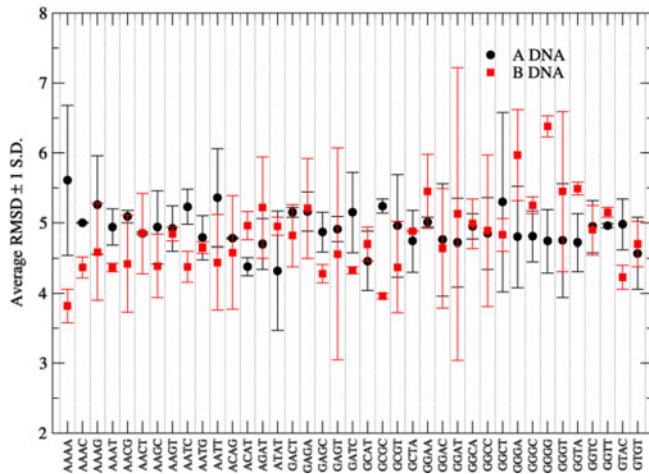


Figure 2. The mass-weighted all-atom root-mean-squared deviation of the 39 15-mer DNA sequences of the form GDABCDABC-DABCDG, where ABCD is the teranucleotide sequence presented on the X-axis. Mean RMSD values over the 15 ns trajectory with reference to the corresponding canonical A- and B-form structures are shown in black and red along with 1 standard deviation.

solution, the poly A sequence has been noted to abstain from transitioning to the A-form structure even under extreme conditions (Saenger 1984) and is known to form the linear and rigid B'-form of DNA structure. The repeating ATAT sequence, which also comprises the TATA element, is known to form A-like structure (Guzikovich-Guerstein and Shakked 1996). The poly G and other G-rich sequences also appear to prefer A-like conformations. This trend is in good accord with the known general preference of G/C-rich sequences for the A-form DNA and the A-rich sequences for the B-form (Ivanov and Minchenkova 1994; Basham *et al.* 1995). Given the fact that all the 39 simulations studied here have been carried out in 100% aqueous environment, these results indicate that apart from the water activity in the environment, the differential B and A preferences is strongly dependent on the sequence composition of the DNA as well. RMSD fit with reference to just the bases of the canonical DNA structure, neglecting the sugar and phosphodiester atoms, maintains this trend in the observed RMSD distribution (data not shown), implying that the structure and orientation of the bases are significantly sequence dependent, irrespective of the sugar conformation. As reported earlier, the sugar pucker of the DNA in all these trajectories are largely in the B-like conformation, i.e. C3' endo (south) (Dixit *et al.* 2005).

Zhurkin and coworkers (Tolstorukov *et al.* 2001) have recently deduced dimeric and trimeric A/B-philicity scale calibrated against experimental data on alcohol-induced B-A transitions in oligomeric DNA duplexes. Figure 3 presents a comparison of the calculated RMSD values of the 39 trajectories against the analytical free energy estimate

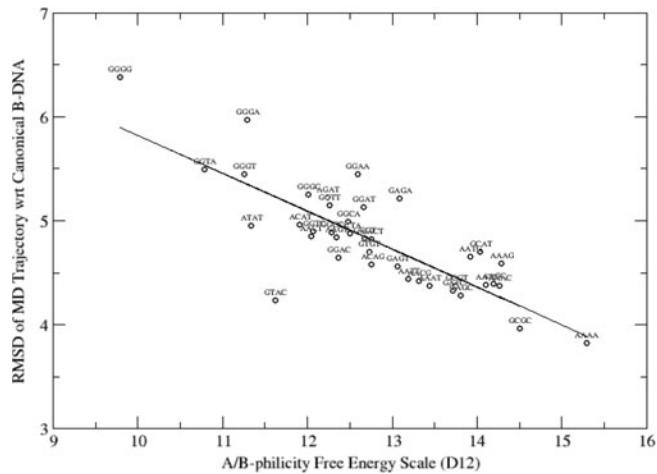


Figure 3. The observed mean RMSD value with reference to the canonical B form structure is plotted against the calculated A/B-philicity free energy scale based on the dimeric model (D12) derived by Tolstorukov and coworkers (Tolstorukov and Maleev 2000). The correlation coefficient between the two datasets is -0.81 .

derived using a sequence composition based dimeric model (D12). The correlation coefficient between the RMSD values and the corresponding D12 A/B-philicity free energy value is -0.81 , underlining the fairly strong correlation between the two datasets and attesting the trends observed in figure 2. Other dinucleotide models have also been reported in the past (Ivanov and Minchenkova 1994; Basham *et al.* 1995), the D12 model being compared here being the latest. The correlation with the trimeric model T-32 presented by Tolstorukov *et al.* (2001) is much weaker with a correlation coefficient of -0.53 , indicating that other factors which are missed in this RMSD comparison might also be contributing to the free energy scale.

Analysis of solvent distribution around the various DNA sequences in detail reveals a great degree of sequence-dependent solvation pattern in the MD trajectories. Figure 4 presents the hydration density maps around the poly A, poly G, poly (ApT) and poly (GpC) sequences as representative of the 39 sequences. Stereo images of these plots are presented in the supplementary material. The hydration density maps are derived by sampling the solvent distribution every 10 ps in the complete 15 ns trajectories. For clarity the maps are based on 2000 water molecules closest to the DNA surface. Strong hydration densities close to the phosphate groups is observed in all these sequences. In case of the poly A sequence we observe an exquisite spine of hydration running along the minor groove of the DNA. The patches of high solvent density are positioned midway between two adjacent base pairs and appear to be shared between the O2 of T and N3 of A from base pairs above and below the water position. This is in very good accord with the spine of hydration observed in the minor groove of central AATT

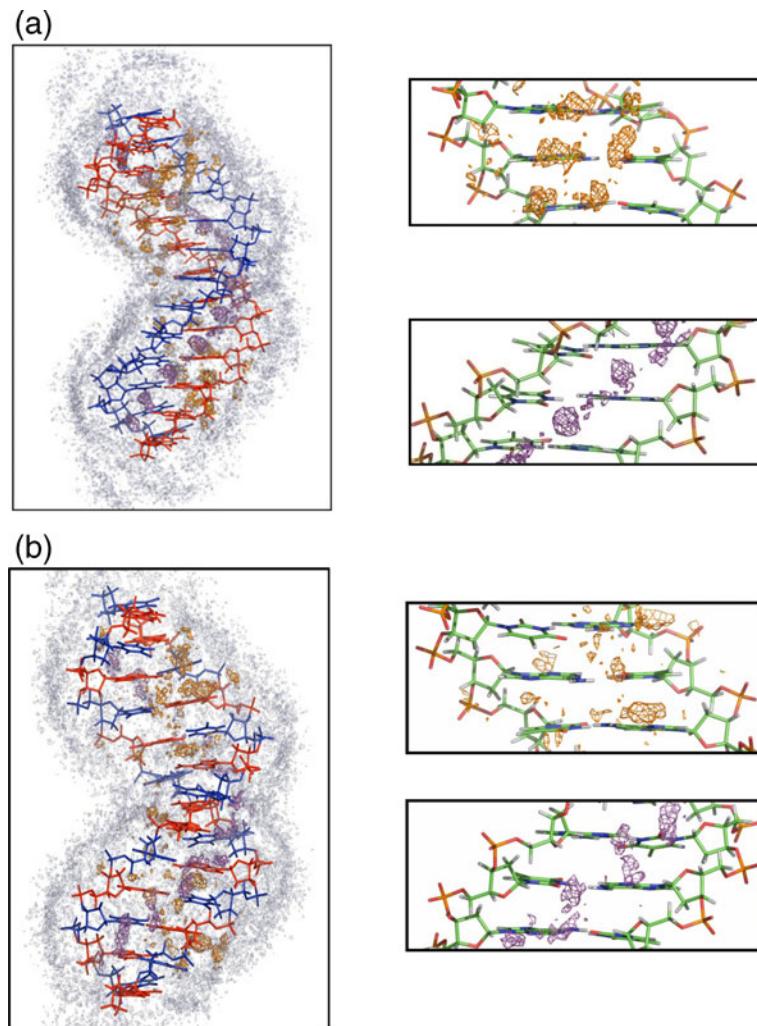


Figure 4. Solvent density around the (a) poly A, (b) poly (ApT), (c) poly G and (d) poly (GpC) sequences. The DNA structures are derived from the average structure of the 15 ns long trajectories of each sequence. The central 13 base pairs of the 15-base-pair-long DNA sequence is shown. The water distribution about the complete DNA is contoured at a density of 25 hits per 0.5 \AA^3 and is shown as light grey grid. The solvent densities within 3.5 \AA of the polar atoms are shown as the orange grid in the major groove and violet grid in the minor groove. The density of groove bound water molecules is plotted at the level of 30 hits per 0.5 \AA^3 . The figures to the right present detailed view of the solvent density in the major (top) and minor (bottom) groove in each case.

in the crystal structure of the Dickerson-Drew dodecamer sequence (Kopka *et al.* 1983). In MD simulations of the poly G sequence, the minor groove sequence exhibits two peaks in the neighbourhood of each base pair positioned. Unlike the poly A sequence, independent hydration densities are observed for the O2 atom of C and N3 atom of G. In accord with X-ray crystallographic observations (Schneider and Berman 1995; Egli *et al.* 1998), the N2 of G does not seem to make direct contact with waters in the first hydration shell. At most of the base pair step positions in this poly G sequence, the hydration peak around the O2 atom of C merges with the hydration pattern around N3 atom of neighbouring G in the opposite strand on the 3' side in a tandem

orientation. The GGCC sequence presents a similar double peak at each base pair step. This is in contrast to the report of Schneider and Berman (1995), which suggests that a GGCC sequence in the position of an AATT sequence in the Dickerson-Drew dodecamer would also exhibit a similar spine of hydration composed of a single hydration peak between each base pair. The major groove of the poly G sequence presents two well-organized tracks of water density. Note that the solvent distribution patterns are much weaker in the poly (ApT) and poly (GpC) sequences compared to the poly A and poly G sequences. Apart from the sequence composition, the increased level of flexibility and dynamics exhibited by these sequences with the pyrimidine-purine

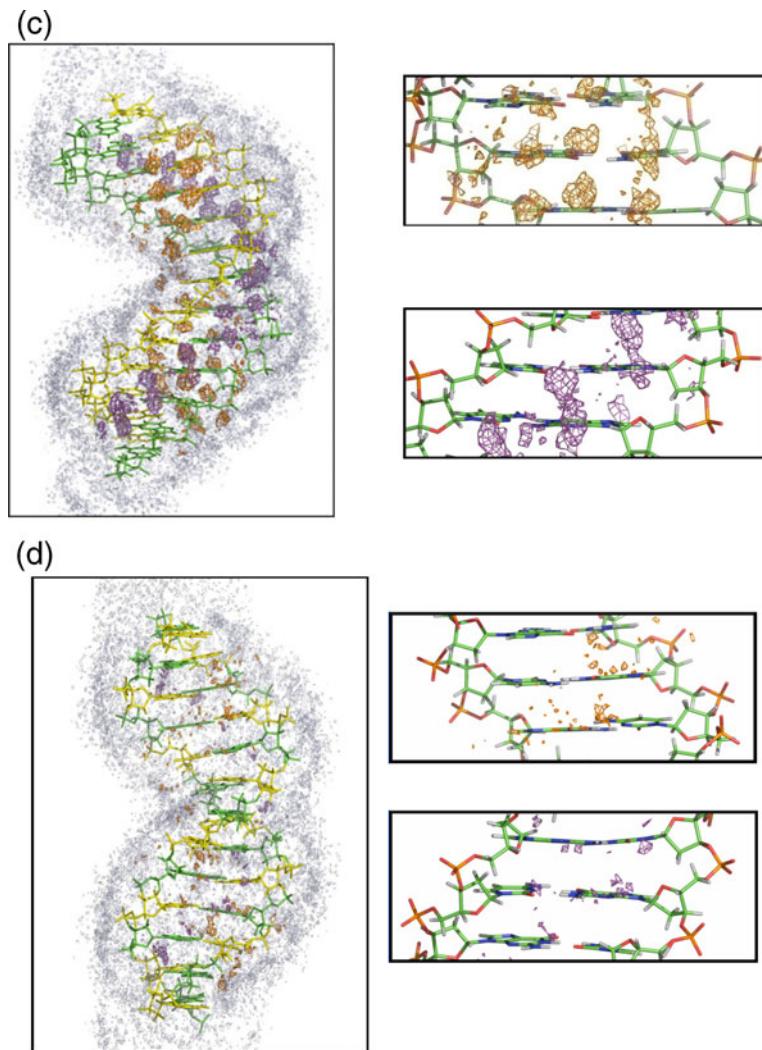


Figure 4. (continued).

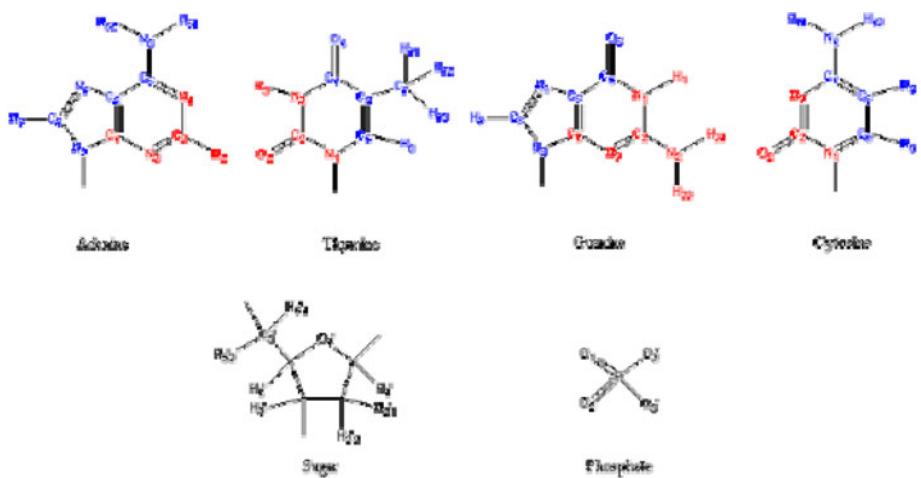


Figure 5. Selection of major and minor groove atoms for proximity analysis. The atoms labeled in blue are considered to lie in the major groove and the one labelled in red lie in the minor groove.

(YpR) steps are responsible for the decreased solvation pattern in these figures.

The proximity analysis (Mehrotra and Beveridge 1980; Mezei and Beveridge 1986) offers a more detailed approach to quantify the general solvent distribution in terms of the coordination numbers, solvation volume of individual atoms in the solute and their corresponding interaction energies. The classification of atoms in the nucleotide as belonging to the major groove, minor groove, sugar and

phosphate groups for these proximity calculations is depicted in figure 5. A summary of the results derived from the proximity analysis of 39 simulation trajectories is provided in table 1. The average values for the A, T, G and C nucleotides and their base pairs from data in all the 39 trajectories is employed to summarize this data. The net coordination number of both AT and GC base pairs in the first solvation shell is not significantly different and both the base pairs are coordinated by approximately 14 water

Table 1. Average results from proximity analysis of the 39 ABC trajectories

	$\langle K \rangle^a$	$\langle 2 K \rangle^b$	$\langle TK \rangle^c$	$\langle VFS \rangle^d$	$\langle SLTBE \rangle^e$	$\langle TSLTBE \rangle^f$	$\langle BEWWT \rangle^g$
Complete Nucleotide							
G	7.4	26.5	74.1	522.1	-101.1	-201.0	-222.1
A	7.0	26.1	71.4	509.6	-95.4	-198.8	-224.6
C	6.8	26.4	72.7	478.1	-92.0	-189.6	-225.4
T	6.8	27.4	71.6	508.8	-95.7	-197.6	-221.0
AT	13.7	53.5	143.1	1018.4	-191.1	-396.4	-445.6
CG	14.3	52.9	146.9	1000.2	-193.1	-390.6	-447.5
Sugar							
G	0.1	6.9	13.4	151.5	-0.7	-36.3	-100.4
A	0.1	6.7	12.8	151.1	-0.7	-36.5	-99.4
C	0.0	6.3	11.5	148.0	-0.4	-32.5	-99.0
T	0.0	5.8	9.4	147.3	-0.3	-30.3	-98.3
AT	0.1	12.5	22.2	298.4	-1.0	-66.8	-197.7
CG	0.1	13.1	24.9	299.5	-1.1	-68.8	-199.4
Phosphate							
G	5.6	15.8	56.7	178.5	-82.2	-135.3	-51.7
A	5.6	15.7	54.5	177.9	-81.8	-135.6	-52.6
C	5.6	15.8	55.9	178.4	-81.4	-134.0	-52.0
T	5.6	15.6	53.5	179.3	-81.7	-135.3	-51.6
AT	11.2	31.2	108.1	357.2	-163.5	-270.9	-104.2
CG	11.3	31.6	112.5	356.9	-163.6	-269.3	-103.6
Major Groove							
G	1.1	2.3	2.6	103.6	-14.1	-19.7	-34.5
A	0.9	3.0	3.4	114.4	-8.5	-19.8	-54.9
C	0.6	3.8	4.8	108.6	-3.6	-16.6	-65.1
T	0.6	5.4	8.1	131.1	-6.7	-25.0	-62.6
AT	1.5	8.4	11.5	245.5	-15.2	-44.8	-117.5
CG	1.8	6.1	7.4	212.2	-17.8	-36.3	-99.6
Minor Groove							
G	0.6	1.4	1.4	88.9	-4.4	-10.1	-35.5
A	0.4	0.6	0.6	66.6	-4.5	-7.0	-17.7
C	0.5	0.5	0.5	42.6	-6.2	-6.2	-9.3
T	0.5	0.5	0.5	50.8	-6.9	-6.9	-8.5
AT	0.9	1.2	1.2	117.4	-11.4	-14.0	-26.2
CG	1.1	1.9	1.9	131.5	-10.6	-16.3	-44.7

a: Number of waters in first hydration shell; b: number of water in the first two hydration shells; c: total number of water molecules in the proximity region; d: volume of first shell (\AA^3); e: interaction energy of solute with waters in first hydration shell (kcal/mol); f: total solute–water interaction energy in the proximity region; g: water–water interaction energy of molecules in the proximity region.

molecules in the first solvation shell. Note that the first solvation shell is defined on the basis of the first minima in the calculated primary radial distribution function. About 80% of the water molecules in the first hydration shell are coordinating the phosphate group, while the major and minor grooves are associated with the remaining 20% of the water molecules. A similar trend has been reported in the X-ray crystallographic analysis of hydration in the A- and B-form structures (Egli *et al.* 1998). Evaluation of the first solvent shell volume for the different regions of the nucleotide against the corresponding first shell coordination number indicates that phosphate groups preferentially attract a larger number of solvent molecules in a comparable volume of space. On considering both the first and second hydration shell, there are about 53 water molecules solvating each nucleotide, with approximately equal number of water molecules observed around both the purines and pyrimidines. In subsequent layers of hydration, the association of water molecules by the phosphate group goes down to about 60% as the water structuring introduces molecules in the proximity region of sugar atoms as well. About 15% of the water molecules around the DNA appear in the proximity region of the sugars. Note that going by the description of proximity region, the minor groove atoms are completely closed from solvent access beyond the first few solvent layers.

Considering the ratio of total solute–solvent interaction energy to the total coordination number in each region of the nucleotide in table 1, the strongest interactions of the solute atoms occur with waters bound to the minor groove, followed by those in the proximity region of the major groove atoms. The last column of table 1 presents the interaction energy of water molecules in a particular proximity region with other water molecules with in a distance of 4 Å. This value is noticeably large for the waters in the proximity of the sugars compared to the phosphate. This explains the fact that the hydrophobic nature of the sugar atoms forces the water molecules in its proximity region to make strong interactions among themselves. On the other hand, the water molecules in the proximity region of the phosphate group align themselves in such a manner that the solute–solvent interactions are much stronger than the solvent–solvent interactions.

As noted earlier the proximity criterion provides a means to analyze solvation effect at the individual nucleotide level for each DNA trajectory independently. A representative plot of the average coordination number and interaction energy at the sequence level for the GTGGATGGATG GATG sequence is shown in figure 6. The results indicate that while it is difficult to discern the differences between sequences on the basis of solvation in the first hydration shell, the sequence-dependent solvation properties become

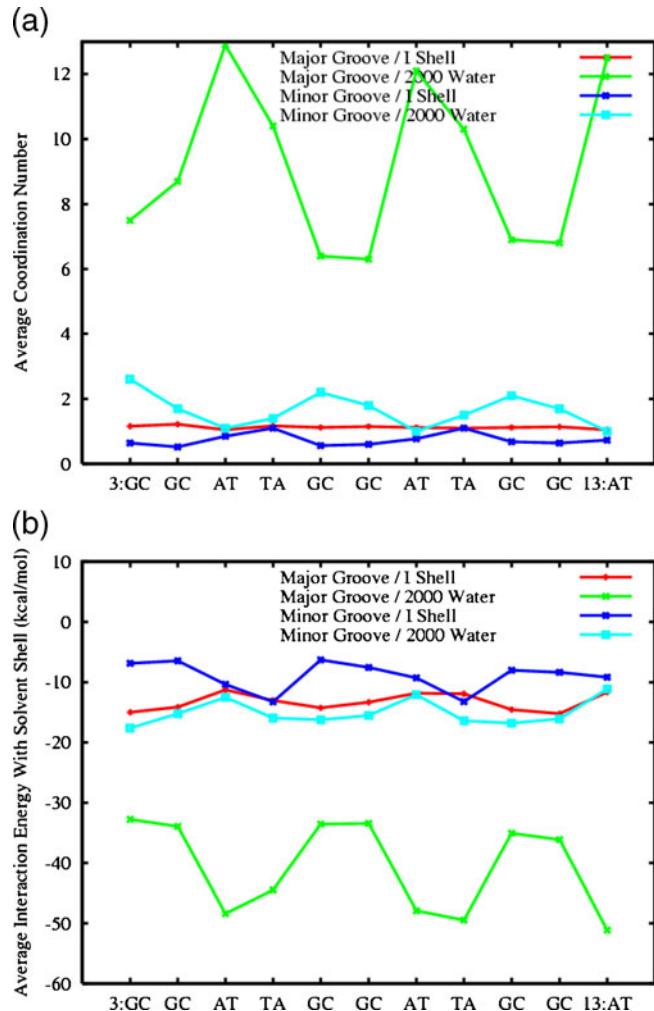


Figure 6. The average (a) coordination number and (b) average interaction energy of d(GTGGATGGATG GATG) duplex simulation in water derived by proximity analysis.

apparent on considering the subsequent layers of solvation. The AT pairs appear to be more solvated in the major groove than the GC base pairs. We notice strong difference between the solvation properties of AT and GC major grooves on including the bulk solvent. While a sequence-dependent trend is observed in the minor groove as well, the sequence effect is not as strong. Neglecting end-effects, the repeating sequence pattern in the DNA sequence causes a symmetrical trend in the calculated solute–solvent interaction energy and the average coordination number. The difference in trend of the number of coordinated water molecules and interaction energy in the first shell and bulk supports the idea that the definition of hydration layer or the solvent shell radius would determine the number of water molecules observed in that layer.

To summarize the data from all the trajectories, we have computed the average solute–solvent interaction energy for each base pair as a function of the flanking sequence derived from the complete database of 39 trajectories. In figure 7 we present a solute–solvent interaction energy map of all the base pairs in the context of the

different trinucleotides. Understandably, the average solute–solvent interaction energies based on proximity analysis indicates that while the sugar and phosphate groups do not exhibit sequence dependence, the major and minor grooves exhibit strong sequence-dependent solvation properties. For both AT and GC base pairs, solute–

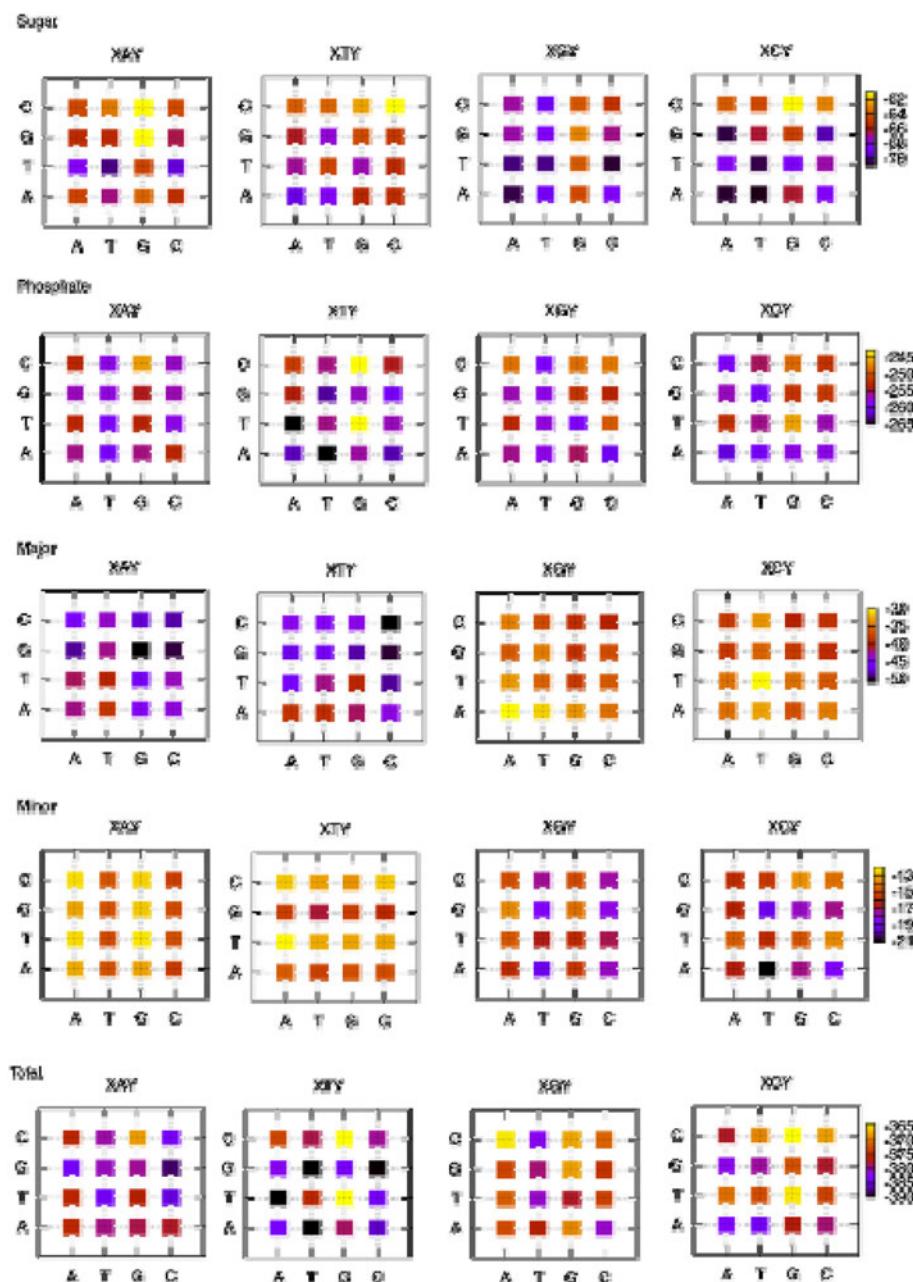


Figure 7. Effect of flanking base pairs on the calculated interaction energy of base pair with the first solvent shell presented for the sugar, phosphate and major and minor grooves. The central base pair of interest is shown at the top of each column. For instance, the top left plot presents the interaction energy of first shell solvent with the sugars of the AT base pair, flanked by the four different base pairs. The interaction energies of the sugar, phosphate and major and minor grooves are significantly different and the colour/energy scale is shown to the right of each row.

solvent interactions are stronger in the major groove than in the minor groove. The AT base pairs bind water molecules much more strongly in the major groove than the GC base pairs. On the other hand, water molecules in the minor groove of GC base pairs bind more strongly than to the AT base pair. While the difference between the AT and GC base pair solvation energies can be as high as 60%, the range of solvent interaction energy for the trinucleotides where the central basepair is AT / TA (or GC / CG) is fairly small. Thus, this result suggests that the solvation of individual base pairs grooves is fairly local in agreement with earlier crystallographic analysis (Schneider and Berman 1995).

Analysis of the solvent radial distribution around the different electronegative atoms in the simulated DNA trajectories reveals that while distinct distribution patterns are observed as a function of the atoms types and the nucleotide, the radial distributions functions (RDF) are not very sensitive to the sequence composition of the DNA. In figure 8 we present the radial distribution function observed for various atom types in the GTGGATGGATGGATG sequence. The O1P and O2P atoms of the phosphate groups are the most clearly solvated atoms with a sharp peak at about 2.8 Å. We observe a second peak about 5.0 Å and a very small third peak at about 7.0 Å. While the O5' and O3' atoms also exhibit the primary solvation peak at about 3.0 Å, these atoms types lack the subsequent solvation shells observed in the case of O1P and O2P. The solvation pattern around the O4' atoms of all the four nucleotides are fairly weak and

indistinguishable. The RDF of N7 atoms in A and G exhibit a strong peak at about 3.0 Å. A secondary peak shows up at 5.0 Å but the sampling in this region appears to be much weaker. The N3 atom of A and G only exhibits the primary peak at about 2.8 Å. Notice that for all RDFs the value at the first minimum is still significantly larger than zero. This means that there is no distinct separation between the first and second shells. This explains the observed sensitivity of coordination numbers to the choice of first shell radius value.

We have also compared the solvent-accessible surface areas (SASA) of the four nucleotides in the crystallographic database and the SASA calculated from MD simulations of diverse nucleic acid sequences in solution. Figure 9 presents a comparison of the solvent accessible surface area derived from the database of 49 DNA crystal structures with resolution higher than 3.0 Å classified as B-DNA (PDB IDs are presented in supplementary table 1) and the 39 MD simulation trajectories. The phosphate groups are the most solvent accessible component of the nucleotide followed by the sugars. The thymine base is most solvent accessible in the major groove, while guanine is most accessible in the minor groove. The net accessibility of the AT and GC base pairs are almost equal, the AT base pairs being more accessible in the major groove and the GC pairs being more accessible in the minor groove.

A recent analysis of the solvation energetics of dinucleotides extracted from ABC trajectories provides a clue to the functional destiny of genomic DNA sequences. Solvation

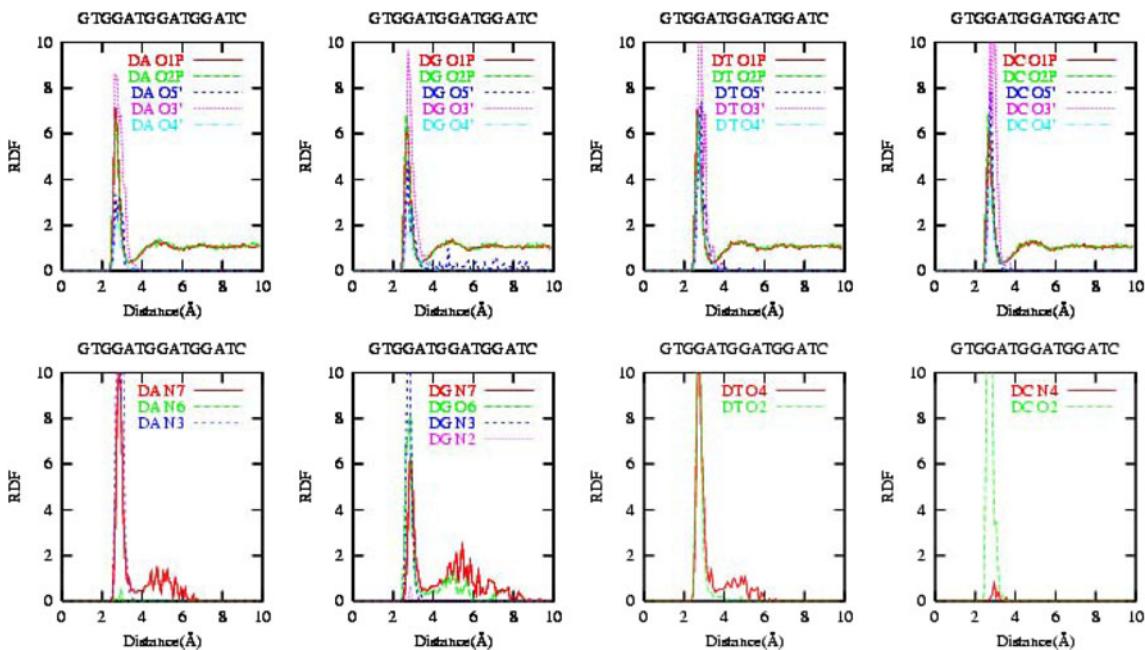


Figure 8. Radial distribution function of waters around electronegative atoms in the GTGGATGGATGGATC sequence.

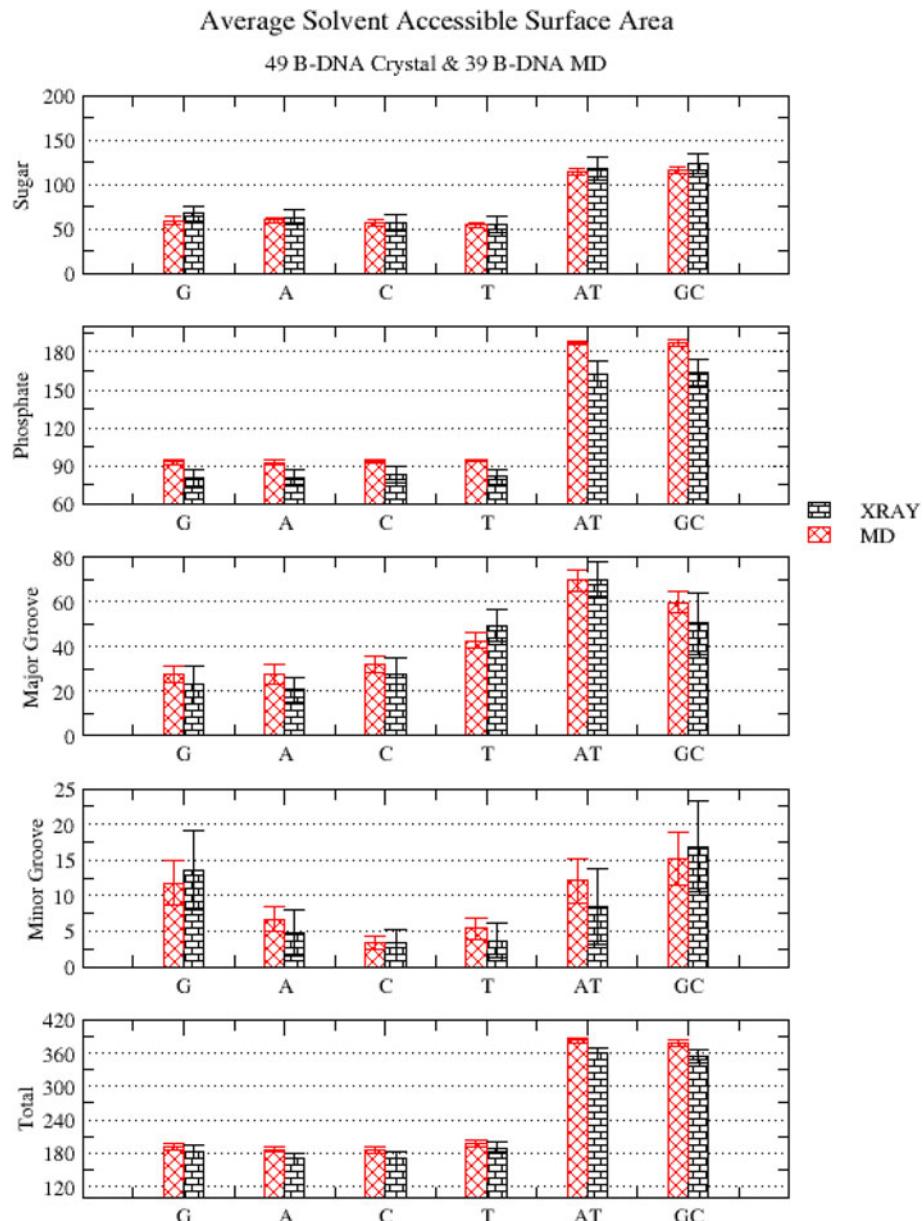


Figure 9. The classification of the average solvent accessible surface area (SASA) of the unique nucleotides in 49 B-DNA crystal structures (in black) compared with average SASA data derived from the 39 15 ns long MD trajectories sampled every 10 ps. The atom-wise SASA has been employed to determine the contribution of the sugar, phosphate and major and minor grooves to the total SASA.

energetics clearly separates DNA coding for mRNA from that coding for tRNA (Khandelwal and Jayaram 2012).

All the 39 simulations studied here have been performed with minimal number of cations required to neutralize the net charge of the system. In particular, 28 K⁺ ions (Aqvist 1990) were included in the simulation box of TIP3P water molecules along with the solute DNA molecule. Figure 10 presents an overlay of the ion atmosphere from 1500 snapshots in a 15 ns long trajectory of the poly A sequence. The ions have been colour-coded according to their distance from

the central axis of the DNA. The axis is defined by the line connecting the centres of the two penultimate base pairs at the ends of the DNA sequence. A distinct variation in the density of ions close to the DNA surface extending to a distance of about 15 to 20 Å and the space beyond is observed. This is in excellent agreement with the Manning's Counterion Condensation Theory (Manning 1978), which describes a layer of 'condensed counterions' around the DNA followed by an extended layer of diffuse Debye-Hückel-like ion atmosphere.

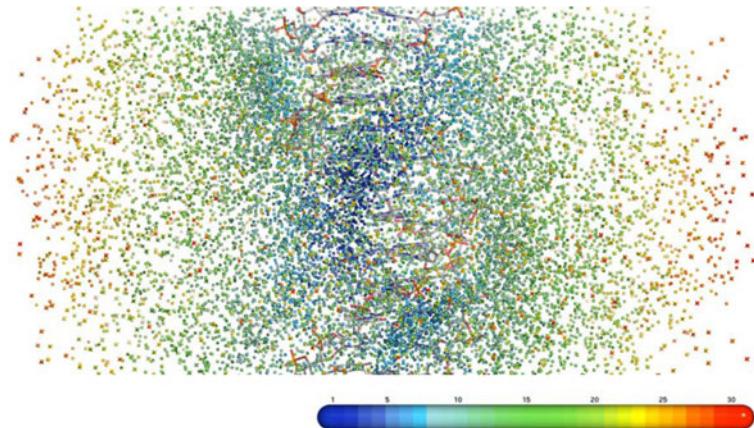


Figure 10. The cumulative counter ion distribution around the poly A sequence in the 15-ns-long molecular dynamics trajectory. Ion distribution in the trajectory has been sampled every 10 ps. The first solute structure in the trajectory that was used as reference is shown. The ions have been colour-coded on the basis of their distance from the helical axis of the DNA.

A plot of the cumulative number of ions as a function of distance from DNA provides further insight into the percentage of counterions present within the region of condensed counterions. Figure 11 shows the radial distribution function and the cumulative number of K^+ ions around the DNA as a function of the distance from the central axis of DNA in all the 39 trajectories. For the radial distribution function $g(r)$ and cumulative ion condensation $N(r)$ calculations, the ions were binned at 1 Å resolution and ions which are located above or below the planes described by the two penultimate base pairs are left out. The volume element for $g(r)$ is calculated from the concentric cylinders around the central axis of the DNA, neglecting the volume correction required for the volume occupied by the solute. Matching the graphical representation in figure 11, the graph of the cumulative counterion condensation data indicates an inflection point near about 20 Å and all the sequences have roughly about 60–70% of the cumulative charge distribution condensed at this distance. Manning's Counterion Condensation Theory proposes that close to 74% of the polyelectrolyte charge could be shielded by the ions present in the condensation region. The height of the first peak in the $g(r)$ plot is different for all the sequences, notably between the A-rich and G-rich sequences. This indicates the sequence-dependent differences in groove binding preference. These ions correspond to the ones that are coloured in shades of dark blue in figure 10 which appear in the grooves (largely the major groove) of the DNA. Beyond this first peak the $g(r)$ is very similar for all the 39 sequences. A consistent second maxima in the $g(r)$ plot is observed at about 13 Å from the DNA surface.

In order to analyse the nature of ion binding to the DNA grooves, figure 12a and b shows, for example, the timeline of ion localization in the vicinity of the DNA molecules for the sequences composed of repeating

tetranucleotides AAAA and AAGG. Each dot in these plots represents an event involving the approach of a cation within 4 Å of the electronegative atoms in the groove or the backbone of the DNA. Only cations that persist in that location for more than 2 ps are shown, and ion approach events shorter than this period of time are assumed to be non-binding diffusion. We observe that ions are consistently present in the vicinity of phosphate oxygen atoms but selectively access the grooves of the DNA in a sequence-dependent fashion. Ion localization in the vicinity of the phosphate groups is altered by the base sequence composition in the DNA. Figure 12c and d presents the cumulative time period over which ions are localized along the length of the DNA. The total extent of ion localization and symmetry in localization pattern becomes apparent in these graphs. Ion approach and localization events are significantly more frequent in the major groove than in the minor groove. The repeating composition of the DNA sequences simulated here provides an opportunity to appreciate the extent of ion convergence around the DNA. Some of the sequences (such as AAGG shown in figure 12) exhibit a sequence-dependent ion localization pattern in the major groove in the form of increased localization at the GG steps. In other cases such as the AAAA sequence, no symmetric patterns can be detected in the ion localization plots. Hence, the degree of ion convergence appears to be a strongly sequence-dependent phenomena.

An analysis of the crystallographic database for the presence of monovalent cations in the vicinity of the AT and GC base pairs is presented in table 2. This data is derived from the Solvation Web Server (<http://tatooine.u-strasbg.fr/~sws/SwS.html>) developed by Auffinger and Hashem (2007). All the B-DNA crystal structures at a resolution of 3.0 Å or higher have been searched for the regular Watson–Crick AT and GC base pairs and monovalent cations within 4 Å

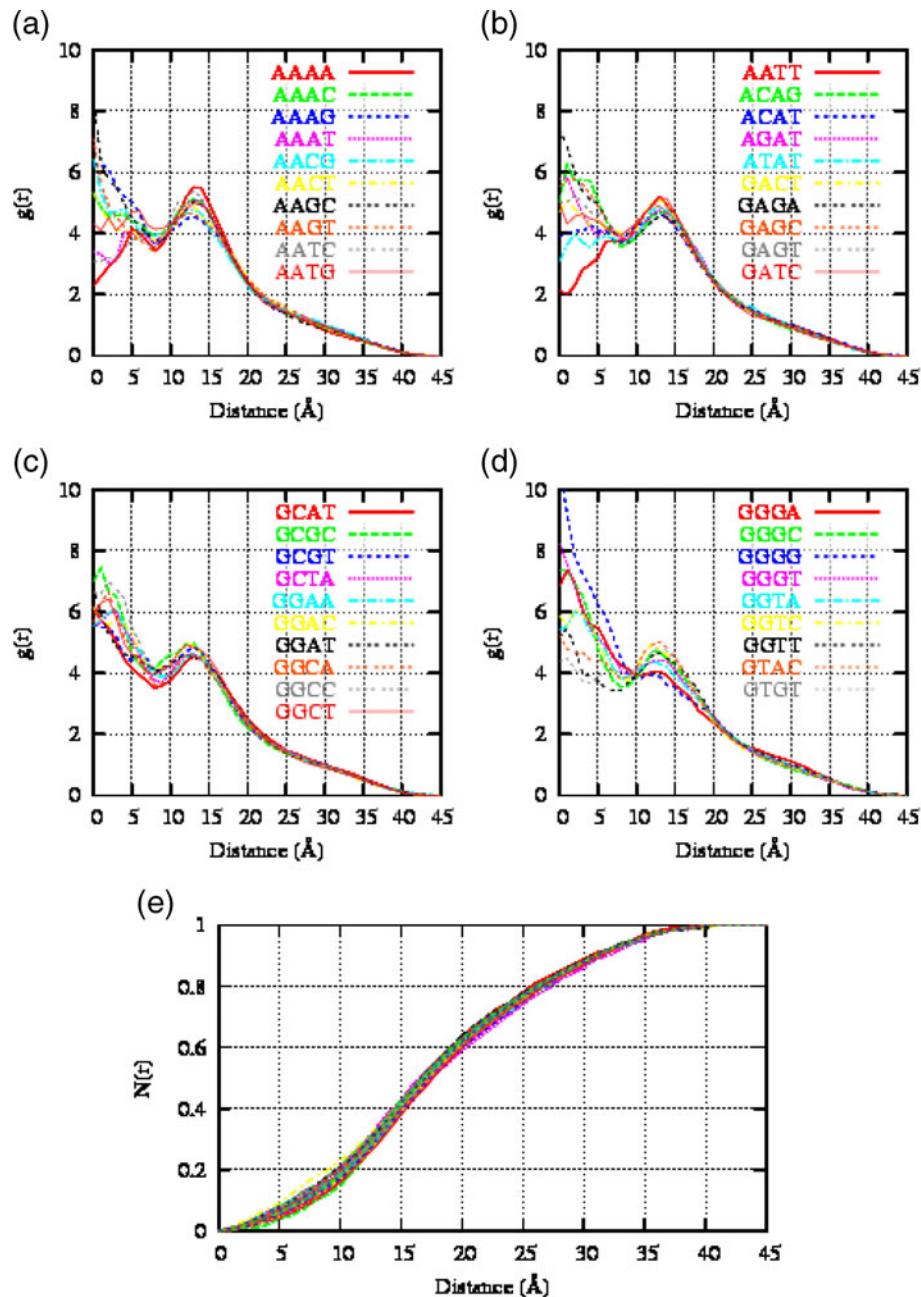
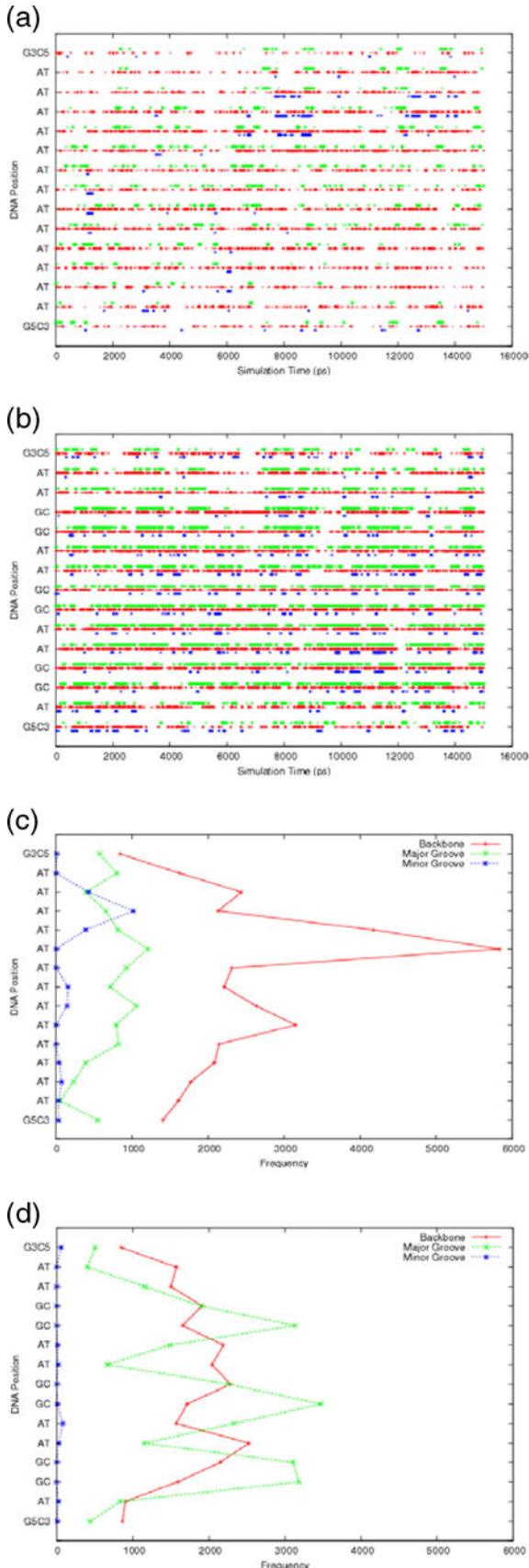


Figure 11. (a-d) radial distribution function $g(r)$ of K^+ ions around the 39 DNA trajectories. The distance r of the ion is calculated with reference to the central axis of the solute DNA. The volume is calculated for concentric cylinders around the DNA axis and the correction in the radial distribution function because of the solute volume has been neglected. (e) The normalized cumulative number of ions condensation as a function of distance from DNA surface for all the 39 trajectories. All the sequences exhibit an inflection point in the range of 15 to 20 Å.

of their electronegative atoms are selected. As of October 2006, the database reveals 7070 GC and 7408 AT base pairs respectively. Of these, only 90 GC pairs and 65 AT pairs are directly associated with some monovalent cations. The bulk of the associated cations are Na^+ followed by a smaller number of K^+ , Tl^+ , Rb^+ or Cs^+ . The

data in table 2 presents a breakup of all these ion–base pair contacts in terms of the grooves and backbone. In the crystal structures the GC pairs make larger number of contacts with the cations in the major groove than the AT base pairs. On the other hand, in the minor groove, the AT pairs make more ion contacts than the GC pairs.



◀ **Figure 12.** (a) Timeline of ion localization within 4 Å of the electronegative atoms in the GAAAAAAAAG sequence. (b) Timeline of ion localization within 4 Å of the electronegative atoms in the GAGGAAGGAAGGAAG sequence. Ions in the vicinity of the major groove, minor groove and backbone are coloured green, blue and red respectively. The separation of ions from the N7 or O6 atoms of G, N4 of C, N7 or N6 atom of A and O4 of T in the major groove are measured. For the minor groove, separation from N9, N3 or N2 of G, N1 or O2 of C, N9 or N3 of A and N1 or O2 of T is measured. Separation from O1P, O2P, O3' and O5' is measure for the backbone. (c–d) Derived from (a) and (b) respectively, these plot shows the cumulative time during which ions are observed in the vicinity of the particular base pair.

To summarize the ion localization results, figure 13 presents the normalized histograms of the observed net ion localization times for the AT (or TA) and GC (or CG) base pairs in the major and minor grooves of the DNA structure in the MD simulations and the crystal structure data. The plot is based on the data of all occurrence of the AT and GC base pairs at the central 13 base pair steps of the 15 base pair sequences, in the complete dataset of 39 trajectories. The extent of ion localization in the major groove is almost an order of magnitude higher than in the minor groove. In the 15-ns-long trajectories, the major grooves of the AT and GC base pairs appear to cumulatively localize ions to the extent of 1 ns and 2.5 ns respectively, while the localization in the minor groove averages to only about a few 100 ps. The complete dataset confirms that the GC pairs assimilate more ions in the major groove than the AT base pairs. In the minor groove, the AT base pairs appear to attract K⁺ ions more frequently compared to the GC pairs.

5. Discussion

In this article we have presented detailed analysis of the solvation and ion distribution around DNA in a database of 39 MD trajectories comprising all the unique tetranucleotide sequences. Each of the trajectories is 15 ns long and the total simulation length amounts to about 0.6 μs. Based on this extensive database of trajectories we have been able to analyse the intrinsic propensity of various DNA sequences to adopt the A- or B-like DNA structures. The trend observed in A/B-philicity in the MD simulations compare well with the empirically based model for predicting A-philicity, confirming that the present simulations model performs very well at predicting the sequence-directed structural properties of DNA. This is a valuable information that can be derived from simulations, which is critical in understanding the properties of reversible A→B transitions critical for developing complete appreciation of DNA structure and its recognition by proteins.

Table 2. Observed frequency of cations in the vicinity of AT and GC base pairs in B-DNA crystal structure at a resolution higher than 3.0 Å available in the NDB

Ions	Major groove	Minor groove	Backbone	Total
GC	55	6	75	136
AT	17	14	53	84
Water	Major groove	Minor groove	Backbone	Total
GC	3649	1489	5217	10355
AT	2824	1425	5494	9743

Monovalent cations within 4.0 Å of the electronegative atoms in the grooves or phosphodiester backbone are considered. Data obtained from the Solvation Web Site (Auffinger and Hashem 2007).

Ion localization close to the DNA structure has important implications for the structural properties of the molecule. In the past MD simulations (Young *et al.* 1997a, b) of the EcoRI endonuclease sequence CGCGAATTGCG had shown that a Na⁺ ion could intrude the minor groove and occupy the position of a water molecule in the spine of hydration. Subsequently, it has been shown with high-resolution X-ray crystallography (Shui *et al.* 1998) and NMR (Denisov and Halle 2000) that ions could replace water molecules in the spine of hydration. An extended 60-ns-long MD simulation of the same sequence had revealed that the ion localization in the minor groove of the central AATT region was not a very frequent event (Ponomarev *et al.* 2004). In the dataset analysed here, which totals to almost 0.6 μs, we noted that localization of ions in the minor groove of ApA or ApT steps is a rare event. In the event such localization occurs, the extent of localization is usually limited to less than 1 ns. We observed that such localization events do not occur symmetrically in all repeating sections of the sequence, indicating that rarity of such binding and the shortness of the 15-ns-long simulations

to normalize ion distribution. It has been shown in earlier simulations that the K⁺ ions diffuse faster and localize less than Na⁺ ions (Varnai and Zakrzewska 2004) in simulations of the DNA. In all the simulations pursued here, K⁺ ions have been the only counterion species in the solution environment and this could be a cause for the lack of localization observed in these simulations. In an earlier study (Varnai and Zakrzewska 2004), it was shown that the presence of Na⁺ as counterions during the DNA simulation led to a greater degree of non-canonical α/γ substates in comparison with simulations employing K⁺ ions. This was attributed to the tendency of Na⁺ ions to sequester in the minor groove of the DNA structure, while K⁺ ions do not.

6. Conclusions

Results from the analysis of solvent and counter ion distribution in 39 different MD simulation trajectories comprising multiple copies of all the 136 unique tetranucleotide steps were presented. Comparison of the DNA structures in the simulated trajectories with the canonical A- and B-form structures reveals that the G/C-rich sequences are closer to canonical A- than B-form structures, while the reverse is true for the poly A sequences with the exception of the alternating ATAT sequence. Analysis of hydration density maps reveals that the flexibility of solute molecule has a significant effect on the nature of observed hydration. While DNA structures of the poly A and poly G sequences, which are fairly stiff, present regular hydration patterns such as the spine of hydration in the case of poly A sequence, the relatively flexible structures such as the poly GpC sequence have very little regular hydration. Energetic analysis of solute-

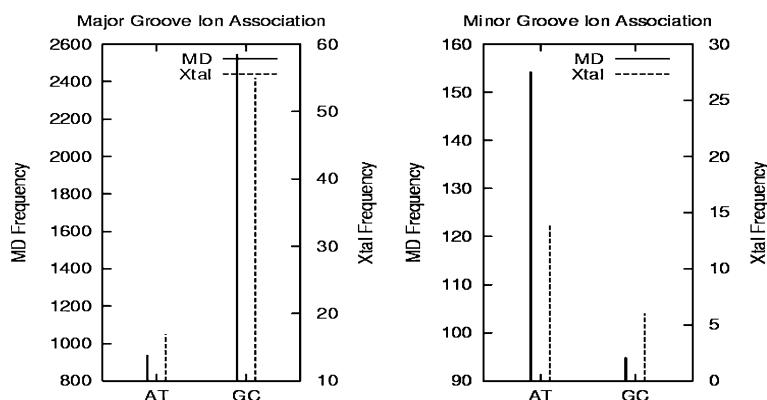


Figure 13. Observed frequency of cation localization in the grooves of the DNA derived from a database of 39 different 15 ns long MD trajectories. The presented data is normalized over the complete dataset which included 210 and 218 copies of the AT and GC base pairs respectively. Only the central 11 base pairs of the 15-mer DNA sequences were included in this analysis to avoid end-effects. The crystal data is based on analysis of DNA structures at a resolution finer than 3.0 Å in the nucleic acid database (NDB) (table 2).

solvent interactions based on proximity analysis of solvent reveals that the GC or CG base pairs interact more strongly with water molecules in the minor groove of DNA than the AT or TA base pairs. On the other hand, the interactions of the AT or TA pairs in the major groove are stronger than those of the GC or CG pairs. Computation of solvent-accessible surface area of the nucleotide units in the simulated trajectories reveals that the similarity with results derived from analysis of a database of crystallographic structures is excellent. Irrespective of the sequence composition, all the MD trajectories tend to obey the Manning's counterion condensation theory, presenting a region of condensed counterions within a radius of about 17 Å from the DNA surface. The GC and CG pairs tend to associate with cations in the major groove of the DNA structure to a greater extent than the AT and TA pairs. Cation association is more frequent in the minor groove of AT than the GC pairs. In general, the observed water and ion atmosphere around the DNA sequences in the MD simulation is in good agreement with experimental observations.

Acknowledgements

We gratefully acknowledge support from MRAC award CHE050040P, SDSC account WEU202, NIH grant GM37909 to DLB, The Keck Center for Integrative Genomics at Wesleyan University and the HHMI grant 52005211. We thank the participants of the Ascona B-DNA Consortium for kindly sharing with us the trajectories of the 39 DNA sequences. DLB acknowledges support from Dr Joshua Boger and the WE Coffman family.

References

Albiser GA, Lamiri A and Premilat S 2001 The A-B transition: temperature and base composition effects on hydration of DNA. *Int. J. Biol. Macromol.* **28** 199–203

Anderson CF and Record Jr MT 1982 Polyelectrolyte theories and their application to DNA. *Annu. Rev. Phys. Chem.* **33** 191–222

Aqvist J 1990 Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* **94** 8021–8024

Arthanari H, McConnell KJ, Beger R, Young MA, Beveridge DL and Bolton PH 2003 Assessment of the molecular dynamics structure of DNA in solution based on calculated and observed NMRNO-ESY volumes and dihedral angles. *Biopolymers* **68** 1097–1028

Auffinger P and Hashem Y 2007 SwS: a solvation web service for nucleic acids. *Bioinformatics* **23** 1035–1037

Basham B, Schroth GP and Ho PS 1995 An A-DNA triplet code: thermodynamic rules for predicting A- and B-DNA. *Proc. Natl. Acad. Sci. USA* **92** 6464–6468

Berendsen HJ, Postma JP, van Gasteren WF, DiNola A and Haak JR 1984 Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81** 3684–3690

Berman HM and Schneider B 1999 Nucleic acid hydration; in *Oxford handbook of nucleic acid structure* (ed) S Neidle (Oxford: Oxford University Press) pp 295–312

Beveridge DL, Barreiro G, Byun KS, Case DA, Cheatham TE III, Dixit SB, Giudice E, Lankas F, et al. 2004 Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys. J.* **87** 3799–3813

Beveridge DL, Dixit SB, Barreiro G and Thayer KM 2004 Molecular dynamics simulations of DNA curvature and flexibility: helix phasing and premelting. *Biopolymers* **73** 380–403

Calladine CR 1982 Mechanics of Sequence-dependent Stacking of Bases in B-DNA. *J. Mol. Biol.* **161** 343–352

Case DA, Pearlman DA, Caldwell JW, Cheatham IIIT E, Ross WS, Simmerling CL, Darden TA, Merz KM, et al. 1999 AMBER 6 (San Francisco: University of California)

Case DA, Darden TA, Cheatham IIIT E, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, et al. 2004 AMBER 8 (San Francisco: University of California)

Chalikian TV and Breslauer KJ 1998 Volumetric properties of nucleic acids. *Biopolymers* **48** 264–280

Chalikian TV, Sarvazyan AP, Plum GE and Breslauer KJ 1994 Influence of base composition, base sequence, and duplex structure on DNA hydration: apparent molar volumes and apparent molar adiabatic compressibilities of synthetic and natural DNA duplexes at 25 degrees C. *Biochemistry* **33** 2394–2401.

Cheatham IIIT E and Kollman PA 1996 Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.* **259** 434–444

Cheatham IIIT E and Kollman PA 1997 Molecular Dynamics Simulations Highlight the structural differences among DNA: DNA, RNA:RNA and DNA:RNA hybrid duplexes journal. *J. Am. Chem. Soc.* **119** 4805–4825

Cheatham IIIT E and Young MA 2000 Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers* **56** 232–256

Cheatham TE III, Miller JL, Fox T, Darden TA and Kollman PA 1995 Molecular dynamics simulations on solvated biomolecular systems: The particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.* **117** 4193–4194

Chen YZ and Prohofsky EW 1993 Synergistic effects in the melting of DNA hydration shell: melting of the minor groove hydration spine in poly(dA)poly(dT) and its effect on base pair stability. *Biophys. J.* **64** 1385–1393

Chiu TK, Kaczor-Grzeskowiak M and Dickerson RE 1999 Absence of minor groove monovalent cations in the crosslinked dodecamer C-G-C-G-A-A-T-T-C-G-C-G. *J. Mol. Biol.* **292** 589–608

Chocholousova J and Feig M 2006 Implicit solvent simulations of DNA and DNA-protein complexes: agreement with explicit solvent vs experiment. *J. Phys. Chem. B* **110** 17240–17251

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, et al. 1995 A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **117** 5179–5197

DeLano WL 2002 The PyMOL molecular graphics system (San Carlos, CA: DeLano Scientific)

Denisov VP and Halle B 2000 Sequence-specific binding of counterions to B-DNA. *Proc. Natl. Acad. Sci. USA* **97** 629–633

Dixit SB and Beveridge DL 2006 Structural bioinformatics of DNA: a web-based tool for the analysis of molecular dynamics results and structure prediction. *Bioinformatics* **22** 1007–1009

Dixit SB, Beveridge DL, Case DA, Cheatham TE III, Giudice E, Lankas F, Lavery R, Maddocks JH, *et al.* 2005 Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides II: Sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys. J.* **89** 3721–3740

Dixit SB, Ponomarev SY and Beveridge DL 2006 Root mean square deviation probability analysis of molecular dynamics trajectories on DNA. *J. Chem. Inf. Model.* **46** 1084–1093

Djurjanovic D and Hartmann B 2003 Conformational characteristics and correlations in crystal structures of nucleic acid oligonucleotides: evidence for sub-states. *J. Biomol. Struct. Dyn.* **20** 771–788

Duan Y, Wilkosz P, Crowley M and Rosenberg JM 1997 Molecular dynamics simulation study of DNA dodecamer d(CGCGAATTCGCG) in solution: conformation and hydration. *J. Mol. Biol.* **272** 553–572

Egli M, Tereshko V, Teplova M, Minasov G, Joachimiak A, Sanishvili R, Weeks CM, Miller R, *et al.* 1998 X-ray crystallographic analysis of the hydration of A- and B-form DNA at atomic resolution. *Biopolymers* **48** 234–252

Eisenstein M and Shakked Z 1995 Hydration patterns and intermolecular interactions in A-DNA crystal structures. Implications for DNA recognition. *J. Mol. Biol.* **248** 662–678

Elcock AH and McCammon JA 1995 Sequence dependent hydration of DNA: Theoretical results. *J. Am. Chem. Soc.* **117** 10161–10162

Essmann U, Perera L, Berkowitz ML, Darden T, Lee H and Pedersen LG 1995 A smooth particle mesh Ewald method. *J. Chem. Phys.* **103** 8577–8593

Feig M and Pettitt BM 1998 A molecular simulation picture of DNA hydration around A- and B-DNA. *Biopolymers* **48** 199–209

Feig M and Pettitt BM 1999 Modeling high-resolution hydration patterns in correlation with DNA sequence and conformation. *J. Mol. Biol.* **286** 1075–1095

Franklin RE and Gosling RG 1953 The structure of sodium thymonucleate fibers I. The influence of water content. *Acta Cryst.* **6** 673–677

Gao YG, Robinson H and Wang AH 1999 High-resolution A-DNA crystal structures of d(AGGGGGCCCT). An A-DNA model of poly(dG) x poly(dC). *Eur. J. Biochem.* **261** 413–420

Guarnieri F and Mezei M 1996 Simulated annealing of chemical potential: a general procedure for locating bound waters. Application to the study of the differential hydration propensities of the major and minor grooves of DNA. *J. Am. Chem. Soc.* **118** 8493–8494

Guzikovich-Guerstein G and Shakked Z 1996 A novel form of the DNA double helix imposed on the TATA-box by the TATA-binding protein. *Nat. Struct. Biol.* **3** 32–37

Halle B and Denisov VP 1998 Water and monovalent ions in the minor groove of B-DNA oligonucleotides as seen by NMR. *Biopolymers* **48** 210–233

Harvey SC, Tan RK-Z and Cheatham TE III 1998 The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J. Comput. Chem.* **19** 726–740

Howard JJ, Lynch GC and Pettitt BM 2011 Ion and solvent density distributions around canonical B-DNA from integral equations. *J. Phys. Chem. B* **115** 547–556

Hunter CA 1993 sequence-dependent DNA structure: The Role of base stacking interactions. *J. Mol. Biol.* **230** 1025–1054

Ivanov VI and Minchenkova LE 1994 The A-form of DNA: In search of the biological role. *Mol. Biol. (Mosk.)* **28** 1258–1271

Jayaram B and Beveridge DL 1990 Free energy of an arbitrary charge distribution imbedded in coaxial cylindrical dielectric continua: application to conformational preferences of DNA in aqueous solutions. *J. Phys. Chem.* **94** 4666–4671

Jayaram B and Beveridge DL 1991 Grand canonical Monte-Carlo simulations on aqueous solutions of NaCl and NaDNA: Excess chemical potentials and sources of non-ideality in electrolyte and polyelectrolyte solutions. *J. Phys. Chem.* **95** 2506–2516

Jayaram B and Beveridge DL 1996 Modeling DNA in aqueous solution: Theoretical And computer simulation studies on the ion atmosphere of DNA. *Annu. Rev. Biophys. Biomol. Struct.* **25** 367–394

Jayaram B, Sharp K and Honig B 1989 The electrostatic potential of B-DNA. *Biopolymers* **28** 975–993

Jayaram, B, DiCapua FM and Beveridge DL 1991 A theoretical study of polyelectrolyte effects in protein-DNA interactions: Monte Carlo free energy simulations on the ion atmosphere contribution to the thermodynamics of l repressor-operator complex formation. *J. Am. Chem. Soc.* **113** 5211–5215

Jayaram B, Sprous D, Young MA and Beveridge DL 1998 Free energy analysis of the conformational preferences of A and B forms of DNA in solution. *J. Am. Chem. Soc.* **120** 10629–10633

Jorgensen WL 1981 Transferable intermolecular potential functions for water, alcohols and ethers. Application to liquid water. *J. Am. Chem. Soc.* **103** 335–340

Jovin TM, Soumpasis DM and McIntosh LP 1987 The transition between B-DNA and Z-DNA. *Annu. Rev. Phys. Chem.* **38** 521–560

Khandelwal G and Jayaram B 2012 DNA-water interactions distinguish messenger RNA genes from transfer RNA genes. *J. Am. Chem. Soc.* DOI: [10.1021/ja3020956](https://doi.org/10.1021/ja3020956)

Knee KM, Dixit SB, Aitken CE, Ponomarev S, Beveridge DL and Mukerji I 2008 Spectroscopic and molecular dynamics evidence for a sequential mechanism for the A-to-B transition in DNA. *Biophys J* **95** 257–272

Kopka ML, Fratini AV, Drew HR and Dickerson RE 1983 Ordered water structure around a B-DNA dodecamer. A quantitative study. *J. Mol. Biol.* **163** 129–146

Kuntz Jr. ID, Brassfield TS, Law GD and Purcell GV 1969 Hydration of macromolecules. *Science* **163** 1329–1331

Lavery R, Zakrzewska K, Beveridge DL, Bishop TC, Case DA, Cheatham TIII, Dixit SB, Jayaram B, *et al.* 2010 A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.* **38** 299–313

Leikin S, Parsegian VA and Rau DC 1993 Hydration Forces. *Annual Review of Physical Chemistry* **44** 369–395

Leslie AG, Arnott S, Chandrasekaran R and Ratliff RL 1980 Polymorphism of DNA double helices. *J. Mol. Biol.* **143** 49–72

Malenkov G, Minchenkova L, Minyat E, Schyolkina A and Ivanov V 1975 The nature of B-A transition of DNA in solution. *FEBS Letters* **51** 38–42

Manning GS 1978 The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Quart. Rev. Biophys.* **11** 179–246

Mazur J, Sarai A and Jernigan RL 1989 Sequence dependence of the B-A conformational transition of DNA. *Biopolymers* **28** 1223–1233

McConnell KJ and Beveridge DL 2000 DNA structure: what's in charge? *J. Mol. Biol.* **304** 803–820

Mehrotra PK and Beveridge DL 1980 Structural analysis of molecular solutions based on quasi-component distribution functions. application to $[H_2CO]_{aq}$ at 25°C. *J. Am. Chem. Soc.* **102** 4287–4294

Mezei M 1988 Modified proximity criteria for the analysis of the solvation of a polyfunctional solute. *Molecular Simulation*. **1** 327–332

Mezei M 2006 *MMC: Monte Carlo program for computer simulation of molecular solutions* (New York)

Mezei M and Beveridge DL 1981 Monte Carlo studies of the structure of dilute aqueous solutions of Li^+ , Na^+ , K^+ , F^- , and Cl^- . *J. Chem. Phys.* **74** 6902–6910

Mezei M and Beveridge DL 1986 Structural chemistry of biomolecular hydration: the proximity criterion. *Methods Enzymol.* **127** 21–47

Nishimura Y, Torigoe C and Tsuboi M 1986 Salt induced B-A transition of poly(dG)*poly(dC) and the stabilization of A-form by its methylation. *Nucleic Acids Res.* **14** 2737–2748

Olson WK, Gorin AA, Lu XJ, Hock LM and Zhurkin VB 1998 DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA* **95** 11163–11168

Orozco M, Noy A and Perez A 2008 Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr. Opin. Struct. Biol.* **18** 185–193

Pal SK, Zhao L, Xia T and Zewail AH 2003 Site- and sequence-selective ultrafast hydration of DNA. *Proc. Natl. Acad. Sci. USA* **100** 13746–13751

Perez A, Luque FJ and Orozco M 2007 Dynamics of B-DNA on the microsecond time scale. *J. Am. Chem. Soc.* **129** 14739–14745

Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, Laughton CA and Orozco M 2007 Refinement of the amber force field for nucleic acids. Improving the description of $\{\alpha\}/\{\gamma\}$ conformers. *Biophys. J.* **92** 3817–3829

Pilet J and Brahms J 1972 Dependence of B-A conformational change on base composition. *Nature New Biol.* **236** 99–100

Ponomarev SY, Thayer KM and Beveridge DL 2004 Ion motions in molecular dynamics simulations on DNA. *Proc. Natl. Acad. Sci. U.S.A.* **101** 14771–14775

Privalov PL, Dragan AI, Robinson CC, Breslauer KJ, Remeta DP and Minetti CA 2007 What drives proteins into the major or minor grooves of DNA? *J. Mol. Biol.* **365** 1–9

Ravishanker G 1998 Molecular Dynamics Tool Chest 2.0 (Wesleyan: Wesleyan University)

Reddy CK, Das A and Jayaram B 2001 Do water molecules mediate protein-DNA recognition? *J. Mol. Biol.* **314** 619–632

Richards FM 1977 Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **6** 151–176

Robinson H and Wang AH 1996 Neomycin, spermine and hexaamminecobalt (III) share common structural motifs in converting B- to A-DNA. *Nucleic Acids Res.* **24** 676–682

Rueda M, Cubero E, Laughton CA and Orozco M 2004 Exploring the counterion atmosphere around DNA: what can be learned from molecular dynamics simulations?. *Biophys. J.* **87** 800–811

Ruggiero NJ, Pereira de Souza F and Colombo MF 2001 Hydration effects on DNA double helix stability modulates ligand binding to natural DNA in response to changes in water activity. *Cell Mol. Biol. (Noisy-le-grand)* **47** 801–814

Ryckaert J-P, Ciccotti G and Berendsen HJC 1977 Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23** 327–341

Saenger W 1984 *Principles of nucleic acid structure* (New York: Springer Verlag)

Saenger W, Hunter WN and Kennard O 1986 DNA conformation is determined by economics in the hydration of phosphate groups. *Nature* **324** 385–388

Schneider B and Berman HM 1995 Hydration of the DNA bases is local. *Biophys. J.* **69** 2661–2669

Schneider B, Cohen D, and Berman HM 1992 Hydration of DNA bases: analysis of crystallographic data. *Biopolymers* **32** 725–750

Schneider B, Cohen DM, Schleifer L, Srinivasan AR, Olson WK and Berman HM 1993 A systematic method for studying the spatial distribution of water molecules around nucleic acid bases. *Biophys. J.* **65** 2291–2303

Schneider B, Patel K and Berman HM 1998 Hydration of the phosphate group in double-helical DNA. *Biophys. J.* **75** 2422–2434

Schwabe JW 1997 The role of water in protein-DNA interactions. *Curr. Opin. Struct. Biol.* **7** 126–134

Seeman NC, Rosenberg JM and Rich A 1976 Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA* **73** 804–808

Sen S, Andreatta D, Sergei Y, Ponomarev, Beveridge DL and Berg MA 2009 Dynamics of water and ions near DNA: comparison of simulation to time-resolved stokes-shift experiments. *J. Am. Chem. Soc.* **131** 1724–1735

Shotton MW, Pope LH, Forsyth T, Langan P, Denny RC, Giesen U, Dauvergne M-T and Fuller W 1997 A high-angle neutron fibre diffraction study of the hydration of deuterated A-DNA. *Biophys. Chem.* **69** 85–96

Shui X, McFail-Isom L, Hu GG and Williams LD 1998 Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry* **37** 8341–8355

Sims EG and Kim S-H 2003 Global mapping of nucleic acid conformational space: dinucleoside monophosphate conformations and transition pathways among conformational classes. *Nucleic Acids Res.* **31** 5607–5616

Sprous D, Young MA and Beveridge DL 1998 Molecular dynamics studies of the conformational preferences of a DNA double helix in water and in an ethanol/water mixture: Theoretical considerations of the A/B transition. *J. Phys. Chem.* **102** 4658–4667

Srinivasan J, Cheatham IIIT E, Cieplak P, Kollman PA and Case DA 1998 Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *J. Am. Chem. Soc.* **120** 9401–9409

Stefl R and Koca J 2000 Unrestrained molecular dynamics simulations of [d(AT)5]2 duplex in aqueous solution: hydration and binding of sodium ions in the minor groove. *J. Am. Chem. Soc.* **122** 5025–5033

Subramanian PS, Ravishanker G and Beveridge DL 1988 Theoretical considerations on the ‘spine of hydration’ in the minor groove of d(CGCGAATTCGCG) duplex: Monte Carlo computer simulation. *Proc. Natl. Acad. Sci. USA* **85** 1836–1840

Tereshko V, Minasov G and Egli M 1999 The Dickerson-Drew B-DNA dodecamer revisited at atomic resolution. *J. Am. Chem. Soc.* **121** 470–471

Tolstorukov MY and Maleev VY 2000 Conformational transitions of DNA induced by changing water content of the sample: a theoretical study. *J. Biomol. Struct. Dyn.* **17** 913–920

Tolstorukov MY, Ivanov VI, Malenkov GG, Jernigan RL and Zhurkin VB 2001 Sequence-dependent B \leftrightarrow A transition in DNA evaluated with dimeric and trimeric scales. *Biophys. J.* **81** 3409–3421

Tsodikov OV, Record Jr MT J and Sergeev YV 2002 A novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J. Comput. Chem.* **23** 600–609

Tsui V and Case DA 2000 Molecular Dynamics Simulations of Nucleic Acids with a Generalized Born Solvation Model. *J. Am. Chem. Soc.* **122** 2489–2498

Tunis MJ and Hearst JE 1968 On the hydration of DNA. I. Preferential hydration and stability of DNA in concentrated trifluoroacetate solution. *Biopolymers* **6** 1325–1344

Umeshara T, Kuwabara S, Mashimo S and Yagihara S 1990 Dielectric Study on Hydration of B-DNA, A-DNA and Z-DNA. *Biopolymers* **30** 649–656

Umrania Y, Nikjoo H and Goodfellow JM 1995 A knowledge-based model of DNA hydration. *Int. J. Radiat. Biol.* **67** 145–152.

Varnai P and Zakrzewska K 2004 DNA and its counterions: a molecular dynamics study. *Nucleic Acids Res.* **32** 4269–4280

Westhof E and Beveridge DL 1989 *Hydration of nucleic acids. in water science reviews: The molecules of life* (ed) F Franks. (Cambridge: Cambridge University Press) pp 24–136

York DM, Yang W, Lee H, Darden T and Pedersen LG 1995 Toward the accurate modeling of DNA: The importance of long-range electrostatics. *J. Am. Chem. Soc.* **117** 5001–5002

Young MA, Jayaram B and Beveridge DL 1997 Intrusion of Counterions into the spine of hydration in the minor groove of B-DNA: Fractional occupancy of electronegative pockets. *J. Am. Chem. Soc.* **119** 59–69

Young MA, Ravishanker G and Beveridge DL 1997 A 5-nanosecond molecular dynamics trajectory for B-DNA: Analysis of structure, motions and solvation. *Biophys. J.* **73** 2313–2336