An Improved Nucleic Acid Parameter Set for the GROMOS Force Field

THEREZA A. SOARES, PHILIPPE H. HÜNENBERGER, MIKA A. KASTENHOLZ, VINCENT KRÄUTLER, THOMAS LENZ, ROBERTO D. LINS, CHRIS OOSTENBRINK, WILFRED F. VAN GUNSTEREN

Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH-Hönggerberg, 8093 Zurich, Switzerland

Received 28 September 2004; Accepted 6 December 2004
DOI 10.1002/jcc.20193
Published online in Wiley InterScience (www.interscience.wiley.com).

Abstract: Over the past decades, the GROMOS force field for biomolecular simulation has primarily been developed for performing molecular dynamics (MD) simulations of polypeptides and, to a lesser extent, sugars. When applied to DNA, the 43A1 and 45A3 parameter sets of the years 1996 and 2001 produced rather flexible double-helical structures, in which the Watson–Crick hydrogen-bonding content was more limited than expected. To improve on the currently available parameter sets, the nucleotide backbone torsional-angle parameters and the charge distribution of the nucleotide bases are reconsidered based on quantum-chemical data. The new 45A4 parameter set resulting from this refinement appears to perform well in terms of reproducing solution NMR data and canonical hydrogen bonding. The deviation between simulated and experimental observables is now of the same order of magnitude as the uncertainty in the experimental values themselves.


Key words: molecular dynamics simulation; GROMOS force field; nucleic acids; A-DNA; B-DNA

Introduction

The structural polymorphism and flexibility of deoxyribonucleic acid (DNA) chains play a key role in fundamental biological processes such as transcription and replication. This conformational variability is modulated by the local nucleotide composition and the environment, in particular, through specific ion binding and hydration. The understanding of the relationship between polymorphism, flexibility, and function in nucleic acids requires the characterization of the structural dynamics of these compounds in their natural solution environment. To this purpose, explicit solvent molecular dynamics (MD) simulations have become an essential tool because they provide an atomistic description of the structural motions occurring within biomolecular systems in solution. However, the accurate modeling of nucleic acids involves two major challenges related to their exceptionally high density of charges: the necessity of an accurate treatment of long-range electrostatic interactions and the requirement of high-quality force-field parameters.

Because the polymorphism and flexibility of nucleic acids is intimately related to their polyelectrolyte nature at physiological pH, electrostatic interactions are expected to play a major role in determining their structure and dynamics. Because charge–charge interactions fall off slowly with distance (Coulomb’s law), these interactions are very susceptible to errors when numerical cutoff truncation is applied. The introduction in recent years of fast lattice-sum methods (P3M or PME) has led to stable trajectories over multinanosecond time scales in cases where the plain cutoff method (without reaction-field forces) clearly failed. It has also allowed for the development of second-generation force fields parameterized specifically to be used with lattice-sum methods. However, Ewald summation and the related mesh methods are only strictly applicable to the treatment of crystalline solids, and it is still a matter of debate whether their use in the context of solutions may not result in periodicity induced artifacts. The main alternative to lattice-sum methods is the application of cutoff truncation complemented by a reaction-field correction. The reaction-field method should not necessarily be viewed as less accurate than lattice-sum methods. It rather results from a different approximation, namely that the medium...
of nucleic acids. The performance of the force field is evaluated by the GROMOS parameter set, 45A4, for the explicit-solvent simulation of scarcity of long-range NOE crosspeaks.

Proton density and the absence of a globular fold, leading to the main effect due to experimental difficulties resulting from the low thermal stability generally observed for nucleic acids. However, its application to elucidate nucleic acid structures has been restricted to very few sequences up to now. This is mainly due to the low sequence specificity of NMR, which makes it difficult to assign the observed NOEs to specific structures.

Methods

Reparameterization of the GROMOS Nucleic-Acid Force Field

The earlier MD simulations of DNA in aqueous solution based on the GROMOS force field had made use of the parameter sets 37C4, 43A1, and 45A3 that had mainly been derived for polypeptides. Long-time (nanosecond) nucleic-acid simulations using the latest GROMOS 45A3 parameter set indicated the need and room for improvement of this set with respect to nucleic acid simulations. The resulting parameter set, labeled 45A4, is reported in the present article. The parameter set 45A4 differs from the 45A3 parameter set in three respects: (1) the inclusion of additional backbone dihedral-angle potential-energy terms and the refinement of the existing ones (Table 1); (2) the addition of explicit aromatic hydrogen atoms on the nucleotide bases; (3) the adaptation of the partial charges of the atoms belonging to the nucleotide bases (Fig. 1). Apart from these
differences, the two sets share their compatibility with the existing GROMOS force-field parameters for other classes of molecules and different types of solvents (identical van der Waals, bond-stretching, bond-angle bending, and improper dihedral-angle interaction parameters).

The refinement of the dihedral-angle potential-energy terms relied on fitting the force-field or molecular-mechanical (MM) energy profiles for the rotation around specific torsional angles against the corresponding quantum-mechanical (QM) profiles. These profiles were calculated for each of the six types of DNA backbone dihedral angles based on Hartree–Fock level of theory with a 6-31\(^*\) basis set using the program Gaussian,\(^{44}\) at a 30-degree resolution (Fig. 2). The compounds considered in these calculations were defined by the four atoms involved in each type of dihedral angle, with their covalently bound neighbors replaced by hydrogen atoms. The corresponding MM torsional profiles were obtained by energy minimization of the same model molecules using the conjugate-gradient minimizer of GROMOS,\(^{43}\) and restraining the relevant torsional dihedral angle at the required values with a force constant of 1.0 kJ mol\(^{-1}\)/H\(^{1}\)/degree\(^{2}\). Figure 2 shows a comparison of the QM- and MM-derived torsional profiles, the latter for both the 45A3 and 45A4 force fields. This comparison reveals significant differences in the energy barriers for the dihedral angles O\(^{-}\)O\(^{-}\)P\(^{-}\)O\(^{-}\)P\(^{-}\)O\(^{-}\)C\(^{-}\) and O\(^{5}\)O\(^{5}\)O\(^{4}\)O\(^{5}\)O\(^{4}\)O\(^{4}\) when comparing the QM and the 45A4 MM with the 45A3 MM data. The corresponding dihedral angle potential energy terms for the parameter set 45A4 were obtained by fitting cosine series to the differences between the QM and MM profiles (the latter excluding the torsional contributions) resulting in the parameters listed in Table 1. The MM torsional profiles corresponding to the new parameter set 45A4 were obtained by fitting cosine terms of multiplicity one stabilize the trans conformation while terms of multiplicity three stabilize staggered conformations.

In addition to the derivation of new dihedral angle parameters, explicit hydrogen atoms were added to the nucleotide bases. The introduction of these aromatic hydrogens is motivated by the observation that united-atom models, although computationally efficient and adequate for the simulation of many properties in condensed-phase systems, do not correctly reproduce the optimal interaction geometry of aromatic groups.\(^{19,45}\) For instance, in benzene, the relative energies involved in face-to-face and T-shape packing can only be reproduced properly by the inclusion of explicit hydrogen atoms bearing small charges.\(^{45,46}\) For this reason, aromatic hydrogen atoms were already represented explicitly in the GROMOS force field for proteins. Because aromatic base stacking plays an important energetic role in stabilizing DNA duplexes, the explicit treatment of all aromatic hydrogens had to be consistently extended to the nucleotide bases adenine, guanine, cytosine, thymine, and uracil.

With the inclusion of explicit aromatic hydrogens, all partial charges on the bases were reconsidered as well. Analogous to the recent reparameterization of polar moieties in proteins\(^{23}\) the charges of the hydrogen-bond donor and acceptor atoms were reassigned (Fig. 1). The partial charge of the aromatic hydrogen atoms attached to a carbon was taken to be 0.1 e.\(^{23}\) The partial charges on most of the nonhydrogen atoms were slightly enhanced to increase the electrostatic interaction between bases.

**Molecular System**

The DNA oligonucleotide used in the present study to validate the new 45A4 force field has the sequence 5’-[CGCGAATTCGCG]-3' and contained the ECO RI endonuclease recognition sequence 5’-[GAATTC]-3’. It is commonly known as the Dickerson–Drew dodecamer (DDD). Initial coordinates corresponding to this sequence in a B-DNA conformation were taken from the crystallographic structure solved at 1.4-Å resolution with entry code 355D.
in the PDB databank. Initial coordinates corresponding to the same sequence in a canonical A-DNA conformation were generated using the program InsightII (Accelrys Inc., San Diego, CA). The coordinates of the explicitly-treated hydrogen atoms were generated using the PROGCH module of the GROMOS program, based on protonation states assigned according to a pH of 7. The oligonucleotide was placed in a cubic box of edge length 7.4 nm. The box was filled with SPC water molecules, respecting a minimum distance of 0.23 nm between water oxygen atoms and any nonhydrogen solute atom. Two salt concentrations were considered, namely 0.0 M NaCl (22 Na\(^+\)/11001 and 13465 or 12487 H\(_2\)O) and 0.1 M NaCl (24 Cl\(^-\)/11002, 46 Na\(^+\)/11001 and 13417 or 12439 H\(_2\)O), where the effective molarities were estimated with exclusion of 22 neutralizing Na\(^+\) ions. The ions were included by replacing in turn water molecules with the most positive (Cl\(^-\)) or negative (Na\(^+\)) electrostatic potential by an ion. The systems were energy-minimized by steepest descent with the solute atoms positionally restrained (force constant of 25.0 \(\times 10^4\) kJ mol\(^{-1}\) nm\(^{-2}\)), followed by an additional minimization without restraints. A summary of the molecular systems and force-field parameter sets corresponding to the different simulations performed is given in Table 2.

**Molecular Dynamics Simulations**

The equilibration of the solute–ion–solvent systems consisted of 100 ps MD simulation with all solute atoms positionally restrained using a force constant initially equal to 25.0 \(\times 10^4\) kJ mol\(^{-1}\) nm\(^{-2}\), and subsequently decreased to 5.0, 3.75, 2.50, 1.25, and 0.0 \(\times 10^4\) kJ mol\(^{-1}\) nm\(^{-2}\) after successive 20-ps intervals. The equilibration was followed by 2- or 3-ns data production (Table 2). For all simulations (equilibration and production) the temperatures of the solute and solvent degrees of freedom were separately coupled to a Berendsen thermostat\(^{48}\) at 298 K with a relaxation time of 0.1 ps. The pressure was maintained through a Berendsen manostat\(^{48}\) at 1.013 \(\times 10^5\) Pa (1 atm) by means of isotropic coordinate scaling with a relaxation time of 0.5 ps. Nonbonded interactions were handled using a triple-range approach,\(^{49}\) with short- and long-range cutoff distances of 0.8 and 1.4 nm, respectively. The pair-list for short-range nonbonded interactions was recomputed every 10 fs simultaneously with the update of the intermediate-range nonbonded interactions. The mean effect of electrostatic interactions beyond the long-range cutoff distance was approximated by the inclusion of a reaction-field force\(^{50}\) using a solvent dielectric constant \(\varepsilon = 66.51\). Bond lengths were constrained using the SHAKE algorithm,\(^{52}\) with a relative geometric tolerance of \(10^{-4}\). A time step of 2 fs was used to integrate the equations of motion based on the leapfrog algorithm.\(^{53}\) Configurations along the trajectories were stored every 0.4 ps for analysis.

**Figure 2.** Energy profiles corresponding to rotations around the dihedral angles O–P–O–C (top), P–O5'–C5'–C4 (middle) and O5–C5'–C4'–O4' (bottom). The calculations were performed on molecules including the four atoms defining each type of dihedral angle, with their covalently bound neighbors replaced by hydrogen atoms. The quantum-mechanical torsional profiles were calculated at the HF/6-31G\(^*\) level (solid line and squares). The molecular-mechanical (torsional and nonbonded terms) profiles were calculated using the GROMOS parameter set 45A3 (dot-dashed line and circles) or the new GROMOS parameter set 45A4 (dashed line and triangles).
Results

The atom-positional RMSD, calculated along the different MD trajectories for all atoms belonging to the central 10 base pairs with respect to either a canonical A-DNA (dashed lines) or the crystallographic B-DNA (solid lines) conformation, is displayed in Figure 3 as function of time (including the preceding 0.1-ns equilibration period). In nearly all the cases, the RMSD values are essentially converged after at most 0.5 ns, with the exception of simulation A employing the 45A3 force field at 0.0 M salt concentration. The corresponding RMSD values, averaged over the 0.5–2.0-ns interval of each simulation, are reported in Table 3.

The atom-positional root-mean-square fluctuations (RMSF), calculated for all atoms over the same period, are displayed in Figure 4. All RMSF profiles are characterized by large fluctuations at the four termini of the DNA strands as well as (to a lesser extent) at the phosphate groups along the backbone. The corresponding RMSF values averaged over all atoms are reported in Table 3.

The average occurrence of canonical (Watson–Crick) hydrogen bonds over the same time interval is displayed in Figure 5. The corresponding values averaged over the 32 canonical hydrogen bonds are reported in Table 3.

The probability distributions of the six (H9251/H9256) backbone dihedral angles and glycosidic dihedral angles (H11033), averaged over the same time period and over all nucleotides (distinguishing purine and pyrimidine bases for the latter), are shown in Figure 6. The corresponding average values are reported in Table 4.

The conformations of the deoxyribose rings and of the phosphodiester linkages were evaluated by comparison between experimental and simulated 3J-coupling constants for the proton pairs H1/H11032, H2/H11032, H1/H11032, H2/H11033, H2/H11033, and the H3/H11032 pair, respectively. The corresponding RMSD between the simulation averages (over the 0.5–2.0-ns interval of each simulation) and the experimental coupling constants are reported per (proton) pair in Table 3. Geometrical parameters characterizing the spatial arrangement and conformation of the nucleotides along the DNA helix, averaged over the same time period, are reported in Table 5.

The time evolution of the average distances between successive phosphorus atoms along the same DNA strand is displayed in Figure 7 for the different simulations. In all cases, this average distance remains closer to values characteristic of the canonical B-DNA conformation (0.69 nm) rather than canonical A-DNA conformation (0.59 nm).

Finally, the compatibility of the simulations with experimental data in solution was assessed by comparison with a set of 160 proton–proton distances derived from (inter- and intraresidue) NOE intensities for the dodecamer. The corresponding average proton–proton distances were calculated from the different trajectories using r^6^-averaging. The difference between these r^6^-averaged distances, calculated over a 0.5–2.0-ns interval for each simulation, and the corresponding NMR-derived upper bounds (including appropriate pseudoatom corrections) is presented in Figure 8. A positive distance-bound violation (average distance in the simulation larger than the corresponding NMR-derived upper-bound distance) represents a disagreement with experiment. On the
other hand, a negative violation may be explained by a weakening of the NOE intensity through various effects and need not indicate a discrepancy between simulation and experiment. The corresponding average and maximal positive violations over the whole set are reported in Table 3 for the different simulations. The average positive violations are all lower than 0.05 nm and are highest for the simulations using the 45A3 parameter set (A and

<table>
<thead>
<tr>
<th>Simulations</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atom-positional RMSD (nm)</td>
<td>from A-DNA</td>
<td>0.42</td>
<td>0.33</td>
<td>0.31</td>
<td>0.32</td>
<td>0.31</td>
</tr>
<tr>
<td>Atom-positional RMSD (nm)</td>
<td>from B-DNA</td>
<td>0.38</td>
<td>0.36</td>
<td>0.35</td>
<td>0.51</td>
<td>0.33</td>
</tr>
<tr>
<td>NOE bound violations</td>
<td>average</td>
<td>0.048</td>
<td>0.032</td>
<td>0.020</td>
<td>0.028</td>
<td>0.021</td>
</tr>
<tr>
<td>NOE bound violations</td>
<td>maximal</td>
<td>0.030</td>
<td>0.025</td>
<td>0.016</td>
<td>0.032</td>
<td>0.028</td>
</tr>
<tr>
<td>Number (&gt;0.0 nm)</td>
<td>104</td>
<td>87</td>
<td>74</td>
<td>80</td>
<td>78</td>
<td>80</td>
</tr>
<tr>
<td>Number (&gt;0.1 nm)</td>
<td>23</td>
<td>13</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Number (&gt;0.2 nm)</td>
<td>10</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$^3$J-value RMSD from experimental value</td>
<td>$^3$J$_{H1'H2'}'$</td>
<td>0.9</td>
<td>0.8</td>
<td>1.3</td>
<td>2.0</td>
<td>1.2</td>
</tr>
<tr>
<td>$^3$J-value RMSD from experimental value</td>
<td>$^3$J$_{H1'H2}''$</td>
<td>0.7</td>
<td>0.7</td>
<td>0.5</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>$^3$J-value RMSD from experimental value</td>
<td>$^3$J$_{H2'H3'}$</td>
<td>0.8</td>
<td>1.0</td>
<td>1.2</td>
<td>1.0</td>
<td>1.1</td>
</tr>
<tr>
<td>$^3$J-value RMSD from experimental value</td>
<td>$^3$J$_{H2'H3}''$</td>
<td>1.6</td>
<td>1.6</td>
<td>1.3</td>
<td>2.0</td>
<td>1.3</td>
</tr>
<tr>
<td>$^3$J-value RMSD from experimental value</td>
<td>$^3$J$_{H3''P}$</td>
<td>1.9</td>
<td>1.9</td>
<td>2.1</td>
<td>2.0</td>
<td>2.1</td>
</tr>
<tr>
<td>Canonical hydrogen bonds (%)</td>
<td>40</td>
<td>65</td>
<td>80</td>
<td>67</td>
<td>86</td>
<td>80</td>
</tr>
</tbody>
</table>

Structural properties and NMR observables averaged over the 0.5–2.0-ns time interval of each simulation. The simulation codes refer to the entries of Table 2. (a) Averaged over all atoms. (b) Average, maximum value or number of positive deviations over all proton pairs. (c) $^3$J-coupling constants were calculated from MD trajectories using the Karplus equation\cite{92} $^3$J$_{HH}$ = $a$ cos$^2$($\theta$) + $b$ cos($\theta$) + $c$ with the parameters $a$ = 10.2, $b$ = -0.8, $c$ = 0.0 Hz (ribose) and $a$ = 15.3, $b$ = -6.2, $c$ = 1.5 Hz (backbone), and are compared to the experimental coupling constants per proton pair\cite{54,83} in terms of root-mean-square deviation. (d) Averaged over all canonical hydrogen bonds.

Figure 4. Atom-positional root-mean-square fluctuations (RMSF) displayed as a function of the atom sequence number along the two DNA strands. The simulation codes A–F refer to the entries of Table 2. The RMSF values were calculated over the 0.5–2.0-ns interval of each simulation. The configurations were superimposed onto the crystallographic structure using all atoms except those belonging to the 3'- and 5'-terminal residues of the two strands. Atom sequence numbers follow the direction 5' to 3' of strand one, followed by 5' to 3' of strand two, with the border between the two strands marked by a dashed line, based on the GROMOS ordering of atoms.\cite{43} The successive fluctuation peaks are characteristic of the phosphate groups.

Figure 5. Average occurrences of canonical (Watson–Crick) hydrogen bonds along strand one of the oligonucleotide. The graphs for the other strand would be symmetrical (not shown). The simulation codes refer to the entries in Table 2. The occurrences are calculated over the 0.5–2.0-ns interval of each simulation. The canonical hydrogen bonds for the CG base pair are given in the order O2$'$N2, N3$'$N1 and N4$'$O6, and those for the AT base pair in the order N6$'$O4 and N7$'$N3. A hydrogen bond is defined by a hydrogen-acceptor distance smaller than 0.25 nm and a donor-hydrogen-acceptor angle larger than 135°.
Discussion

Influence of the Force-Field Parameters

In the MD simulations using the GROMOS parameter set 45A3, the structures sampled appear to be intermediate between A- and B-DNA conformations with respect to many properties (see Figs. 3 and 7 and Table 3). The simulation at 0.0 M NaCl appears not to have reached equilibrium after 2 ns, as indicated by the steadily increasing atom-positional RMSD with respect to the B-DNA starting structure (Fig. 3). This simulation (A) is also characterized by particularly large atom-positional RMSF values (Fig. 4), the largest being observed for the atoms in the three to four pairs of terminal nucleotides at one end of the double-helix (5'-end of strand one and 3'-end of strand two). Increased motion in this region is correlated with the loss of canonical hydrogen bonding between the corresponding nucleotides (Fig. 5). The other end of the helix is also affected by a partial loss of canonical hydrogen bonds, but only for one base pair. Similar end effects involving the terminal base pairs are also observed in the 45A3 simulation (B) at 0.1-M salt concentration. They have also been observed by Cheatham and others. 58

The probability distributions of five out of six backbone dihedral angles obtained from the simulations performed with the 45A3 parameter set show more than one population peak (Fig. 6, gray lines). In particular, the distribution for the dihedral angle $\alpha$ ($O3'_{n-1}$—P—O5'—C5') displays a sizeable second peak, treated in Figure 9. In this figure, the sites corresponding to occupancies larger than 10% during the interval 1–2 ns are represented by a sphere.

Table 4. Backbone Dihedral Angles

<table>
<thead>
<tr>
<th>Dihedral angle$^a$</th>
<th>A-DNA$^b$</th>
<th>B-DNA$^b$</th>
<th>X-ray (deg)</th>
<th>NMR (deg)</th>
<th>A (deg)</th>
<th>B (deg)</th>
<th>C (deg)</th>
<th>D (deg)</th>
<th>E (deg)</th>
<th>F (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>O3$'_{n-1}$—P—O5'—C5$'$</td>
<td>$-sc$</td>
<td>$-sc$</td>
<td>313</td>
<td>294</td>
<td>217</td>
<td>247</td>
<td>283</td>
<td>278</td>
<td>281</td>
</tr>
<tr>
<td>$\beta$</td>
<td>P—O5—C5—C4$'$</td>
<td>$+ap$</td>
<td>$-ac$</td>
<td>213</td>
<td>184</td>
<td>169</td>
<td>166</td>
<td>160</td>
<td>164</td>
<td>162</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>O5—C5—C4—C3$'$</td>
<td>$+sc$</td>
<td>$+sc$</td>
<td>36</td>
<td>99</td>
<td>125</td>
<td>104</td>
<td>86</td>
<td>77</td>
<td>78</td>
</tr>
<tr>
<td>$\delta$</td>
<td>C5—C4—C3—O3$'$</td>
<td>$+sc$</td>
<td>$+ap$</td>
<td>156</td>
<td>236</td>
<td>122</td>
<td>119</td>
<td>215</td>
<td>107</td>
<td>104</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>C4—C3—O3—P</td>
<td>$-ac$</td>
<td>$+ap$</td>
<td>155</td>
<td>178</td>
<td>210</td>
<td>198</td>
<td>198</td>
<td>198</td>
<td>198</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>C3—O3—P—O5$'_{n-1}$</td>
<td>$-sc$</td>
<td>$-ac$</td>
<td>264</td>
<td>269</td>
<td>248</td>
<td>259</td>
<td>260</td>
<td>265</td>
<td>268</td>
</tr>
<tr>
<td>$\chi$</td>
<td>O4—C1—N9—C4</td>
<td>$-ap$</td>
<td>$-ac$</td>
<td>262</td>
<td>245</td>
<td>239</td>
<td>240</td>
<td>232</td>
<td>230</td>
<td>258</td>
</tr>
<tr>
<td>Sugar pucker (P)</td>
<td>C3'-endo</td>
<td>C2'-endo$^c$</td>
<td>191</td>
<td>154</td>
<td>133</td>
<td>131</td>
<td>178</td>
<td>183</td>
<td>174</td>
<td>180</td>
</tr>
</tbody>
</table>

Definitions of the backbone dihedral angles in nucleic acids, together with the values commonly found in crystallographic canonical A- and B-DNA structures, the corresponding values for the (X-ray) crystallographic structure, averaged over a set of five NMR-derived model structures, and averaged over the 0.5–2.0-ns interval of the six MD simulations. The simulation codes refer to the entries of Table 2. The Klyne and Prelog nomenclature for torsional-angle ranges is used as recommended by the IUPAC-IUB commission on nucleotide nomenclature. (a) The subscripts indicate the nucleotide sequence number relative to the reference nucleotide n (omitted when equal to n). (b) $+sc$ (30° to 90°); $-sc$ (90° to 180°); $+ap$ (150° to 180°); $-ap$ (180° to 270°); $-ac$ (210° to 270°); $-sc$ (270° to 330°); C3'-endo (0° to 36°); C2'-endo (144° to 180°). (c) The phase angle of pseudorotation P in crystallographic structures of B-DNA often displays a broad range of values for C1'-exo to C3'-exo (144° to 216°).
Table 5. Geometric Parameters.

<table>
<thead>
<tr>
<th>Helical parametersa</th>
<th>A-DNA</th>
<th>B-DNA</th>
<th>X-ray</th>
<th>NMR</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>x-Displacement (nm)</td>
<td>−0.42</td>
<td>0.05</td>
<td>0.0</td>
<td>−0.01</td>
<td>−0.23</td>
<td>−0.29</td>
<td>−0.33</td>
<td>−0.36</td>
<td>−0.36</td>
<td>−0.37</td>
</tr>
<tr>
<td>y-Displacement (nm)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>−0.06</td>
<td>0.0</td>
</tr>
<tr>
<td>Helical rise (nm)</td>
<td>0.28</td>
<td>0.33</td>
<td>0.34</td>
<td>0.33</td>
<td>0.26</td>
<td>0.26</td>
<td>0.28</td>
<td>0.29</td>
<td>0.28</td>
<td>0.26</td>
</tr>
<tr>
<td>Inclination (degree)</td>
<td>14.7</td>
<td>2.1</td>
<td>−5.9</td>
<td>4.6</td>
<td>19.4</td>
<td>19.9</td>
<td>16.4</td>
<td>18.5</td>
<td>15.2</td>
<td>15.8</td>
</tr>
<tr>
<td>Tip (degree)</td>
<td>−0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>−0.6</td>
<td>−0.2</td>
<td>0.0</td>
<td>0.3</td>
<td>−0.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Helical twist (degree)</td>
<td>32.5</td>
<td>36.5</td>
<td>35.8</td>
<td>35.1</td>
<td>33.9</td>
<td>34.7</td>
<td>34.9</td>
<td>36.5</td>
<td>33.1</td>
<td>32.5</td>
</tr>
<tr>
<td>Shear (nm)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>−0.01</td>
<td>0.0</td>
</tr>
<tr>
<td>Stretch (nm)</td>
<td>−0.02</td>
<td>−0.01</td>
<td>−0.01</td>
<td>0.04</td>
<td>−0.04</td>
<td>−0.02</td>
<td>−0.01</td>
<td>−0.02</td>
<td>−0.02</td>
<td>−0.02</td>
</tr>
<tr>
<td>Stagger (nm)</td>
<td>0.0</td>
<td>0.01</td>
<td>0.0</td>
<td>0.0</td>
<td>0.05</td>
<td>−0.02</td>
<td>−0.18</td>
<td>−0.03</td>
<td>−0.19</td>
<td>−0.03</td>
</tr>
<tr>
<td>Buckle (degree)</td>
<td>−0.1</td>
<td>−0.5</td>
<td>0.0</td>
<td>−0.1</td>
<td>−3.0</td>
<td>1.5</td>
<td>0.4</td>
<td>−4.4</td>
<td>1.4</td>
<td>0.5</td>
</tr>
<tr>
<td>Propeller (degree)</td>
<td>−11.8</td>
<td>−11.4</td>
<td>4.2</td>
<td>−12.0</td>
<td>−10.3</td>
<td>−14.3</td>
<td>−7.4</td>
<td>−5.3</td>
<td>−6.8</td>
<td>−4.9</td>
</tr>
<tr>
<td>Opening (degree)</td>
<td>0.6</td>
<td>0.0</td>
<td>−5.5</td>
<td>1.9</td>
<td>0.8</td>
<td>−2.5</td>
<td>−1.8</td>
<td>2.0</td>
<td>−1.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Shift (nm)</td>
<td>0.0</td>
<td>−0.02</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>−0.01</td>
<td>0.0</td>
</tr>
<tr>
<td>Slide (nm)</td>
<td>−0.15</td>
<td>0.02</td>
<td>−0.03</td>
<td>−0.05</td>
<td>−0.03</td>
<td>−0.07</td>
<td>−0.07</td>
<td>−0.11</td>
<td>−0.11</td>
<td>−0.12</td>
</tr>
<tr>
<td>Rise (nm)</td>
<td>0.33</td>
<td>0.33</td>
<td>0.34</td>
<td>0.34</td>
<td>0.33</td>
<td>0.33</td>
<td>0.31</td>
<td>0.37</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>Tilt (degree)</td>
<td>0.1</td>
<td>−0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>−2.2</td>
<td>0.2</td>
<td>0.5</td>
<td>−0.3</td>
<td>0.0</td>
<td>−0.4</td>
</tr>
<tr>
<td>Roll (degree)</td>
<td>8.0</td>
<td>0.6</td>
<td>−3.6</td>
<td>2.4</td>
<td>12.9</td>
<td>11.8</td>
<td>11.5</td>
<td>11.2</td>
<td>8.1</td>
<td>8.4</td>
</tr>
<tr>
<td>Twist (degree)</td>
<td>31.1</td>
<td>36.0</td>
<td>35.6</td>
<td>34.6</td>
<td>29.3</td>
<td>30.0</td>
<td>28.5</td>
<td>32.4</td>
<td>30.2</td>
<td>29.2</td>
</tr>
<tr>
<td>P−P distance (nm)</td>
<td>0.59</td>
<td>0.69</td>
<td>0.65</td>
<td>0.68</td>
<td>0.69</td>
<td>0.70</td>
<td>0.68</td>
<td>0.67</td>
<td>0.68</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Geometrical parameters characterizing the spatial arrangement and conformation of the nucleotides along the DNA helix, calculated for ideal A-DNA and B-DNA structures, for the X-ray structure,33 for a set of five NMR-derived model structures of the dodecamer,28 and averaged over the 0.5–2.0-ns interval of the six simulations. The parameters include the helicoidal parameters following the definitions used in the program 3DNA,97 and the distance between successive phosphorous atoms along the same DNA strand of the dodecamer. Helical rise and helical twist are measured relative to the helical axis, while the rise and twist are local parameters, that is, relative to the adjacent base pair. (a) The analysis of the spatial arrangement of the base pairs along the DNA helix was performed using the program 3DNA.97 (b) Average values of helical parameters in high-resolution A- and B-DNA crystal structures.28

Figure 7. Average distance between successive phosphorous atoms along the same strand, displayed as a function of time for the different simulations. The simulation codes refer to the entries of Table 2. The corresponding distances for canonical A-DNA (0.59 nm) and B-DNA configurations (0.69 nm) are indicated by horizontal dotted lines.

Figure 8. NOE distance bound violations calculated for the six (A–F, simulation codes refer to the entries of Table 2) simulations of the dodecamer, for the crystallographic B-DNA33 (G) and for a canonical A-DNA (H) configuration. A violation is defined as the difference between the proton–proton distance in the given structure or averaged over the given simulation (based on r̄6-averaging) and the corresponding experimental NOE-derived distance (including the appropriate pseudo-atom correction57). The violations are calculated over the 0.5–2.0-ns time interval of the corresponding simulation. The NOE bounds are given in the order defined by the experimental data (PDB entry code 1DUF).28

while the distribution for the dihedral angle γ (O5′—C5′—C4′—C3′) is characterized by two almost equivalently populated peaks. In nonhelical conformations, the rotation
about the P—O ester bond as well as the rotation around the exocyclic C4′—C5′ bond should yield an ethane-like staggered profile with three main conformations (+sc, +ap, −sc) corresponding to all substituents in staggered positions. However, in a right-handed double-helical conformation (either A- or B-DNA), these backbone dihedral angles are sterically restricted, and α and γ should adopt values in the −sc (270° to 330°) and +sc (30° to 90°) ranges, respectively, which indeed matches the ranges observed for the X-ray and NMR-derived model structures (Table 4). The presence of additional peaks in the corresponding distributions from the two simulations thus indicates deviations of the backbone conformation with respect to the ideal double-helical structure. These deviations can be understood by comparing the three MM rotational profiles (upper, middle, lower panels of Fig. 2) associated with these same dihedral angles (α and ζ, β and ϵ; γ) calculated using the 45A3 parameter set to the corresponding QM profiles. The reparameterization of the corresponding dihedral angle potential energy terms, aimed at improving the agreement between the MM and QM dihedral profiles, leads to the disappearance or reduction of the second population peaks in the population distributions for the backbone angles α, β, γ, ϵ, and ζ using the 45A4 parameter set (Fig. 6, black lines). However, it should be stressed that the deviations observed for the 45A3 set are of the same order of magnitude as the variations of these angles among the different experimental structures. In this respect, Table 4 shows that the average dihedral angles in the 45A4 simulations are generally closer to the NMR values (solution state) than to the X-ray values (solid state).

The conformations of the deoxyribose ring and of the phosphodiester linkage during the 45A3 and 45A4 simulations were evaluated by comparison of simulated and experimental 3J-coupling constants for four pairs of protons and one proton-phosphorous pair per nucleotide (Table 3). The corresponding RMS differences between calculated and experimental values are in the range 0.5–2.1 Hz and differ little among the different simulations, less than 0.3 Hz on the average deviation per coupling constant. Because different experimental parameterizations of the Karplus relation connecting 3J-values to the corresponding dihedral angle yield variations of about 1 Hz, a difference of similar magnitude between simulations based on different parameter sets cannot be used to (in)validate one of them.

The ensembles of structures resulting from the simulations using the two parameter sets were compared against the 160 experimental proton–proton distance upper bounds derived from NOE intensities (Fig. 8). The corresponding averages, maximal values, and number of positive violations over all proton pairs are reported in Table 3. The average positive violation for the NMR model structures (derived using the same NOE data) is 0.010 nm, while canonical B-DNA and A-DNA structures lead to values of 0.039 and 0.057 nm, respectively. The corresponding values for the MD simulations are all in the range 0.020–0.032 nm, except for the simulation (A) using the 45A3 parameter set at 0.0-M salt concentration.

The ensembles of structures generated in simulations C and E using the 45A4 parameter set and starting from the X-ray structure in B-DNA conformation exhibit a lower RMSD from this structure compared to those generated in simulations A and B using the older 45A3 parameter set (Fig. 3). The RMSF values in the four-strand termini of the DNA duplex are also significantly lower (Fig. 4), leading to significantly reduced average RMSF values (Table 3) and to more stable hydrogen bonding between complementary bases at the two ends of the helix (Fig. 5). The 45A4 backbone dihedral angle distributions also agree better with the experimental range of values for B-DNA conformations. This is likely to be due to the refinement of the potential-energy terms associated with the dihedral angles O—P—O—C and O5′—C5′—C4′—C3′ (Fig. 2) and the resulting suppression of the corresponding alternative population peaks in the distributions corresponding to the α and γ torsional angles (Fig. 6). The inclusion of all aromatic hydrogens in the bases and the new charge distributions have also led to an improvement of the Watson–Crick

Figure 9. Density of Na+ ions around the (average) dodecamer structure for the different simulations using the 45A4 parameter set. Simulation codes refer to the entries of Table 2. Sites presenting occupancies larger than 10% are represented as spheres. Upper panel: simulation C (left) and D (right) during the 1–2-ns interval of simulation. Middle panel: simulation E during the 1–2-ns (left) and 2–3-ns (right) interval of simulation. Bottom panel: simulation F during the 1–2-ns interval of simulation.
hydrogen bonding between the strands (Fig. 5). Overall, the 45A4 parameter set performs significantly better for DNA than the previous 45A3 set.

**Influence of the Initial Conformation**

As indicated by X-ray crystallography, the two most relevant conformations of DNA in biological systems are the A-DNA and B-DNA forms. The relative stability of the two states is modulated by the DNA sequence and by environmental effects, such as the nature and concentration of ions or cosolvents in the surrounding aqueous medium. In particular, A-DNA is generally the predominant conformation in solutions above 2-M salt concentration whereas B-DNA is prevalent in solutions below 1-M salt concentration. In particular, A-DNA is generally the predominant conformation in solutions above 2-M salt concentration whereas B-DNA is prevalent in solutions below 1-M salt concentration.

The relative stabilities of the A-DNA and B-DNA conformations at 0.0- and 0.1-M salt concentrations were investigated through four simulations (C–F) using the improved parameter set 45A4. Low salt concentrations were selected because of the greater abundance and biological relevance of B-DNA conformations compared to other DNA conformations.

The main differences between crystallographic A- and B-DNA structures are the deoxyribose ring conformation, the width and depth of the minor groove, the rise between base pairs, and the helicoidal parameters defining the orientation of the base pairs relative to the helix axis. Crystal structures of A-DNA exhibit sugar pucker pseudorotation angles that deviate little from C3′-endo, whereas crystallographic B-DNA structures are characterized by a broader distribution of sugar puckers centered on C2′-endo but extending from C4′-exo to C3′-endo. The C3′-endo conformation found in the A-DNA form is associated with a short intrastrand distance of about 0.59 nm between successive phosphorous atoms along the strands, to be compared to 0.69 nm for B-DNA. The helical rise between base pairs, which defines the translation per residue along the helical axis, exhibits average values of 0.28 and 0.30 nm in A- and B-DNA conformations, respectively.

The RMSD values along the MD simulations (C and E) starting from the X-ray B-DNA conformation rapidly converge to a plateau of about 0.35 nm relative to both the B-DNA and A-DNA structures (Fig. 3 and Table 3). The values along the corresponding simulations (D and F) starting from an A-DNA conformation converge to slightly higher values of about 0.45 nm relative to the X-ray B-DNA structure, while they show similar deviations with respect to the A-DNA structure. The simulation (F) at 0.1-M salt concentration starting from the A-DNA structure converges to a conformation closer to B-DNA than the corresponding simulation (D) at 0.0-M salt concentration. Note, however, that RMSD values should be taken cautiously as indicators of (dis)similarity between DNA helices, because small changes in the helical parameters can lead to large changes in RMSD values. The analysis of other structural parameters is therefore a requirement for the proper characterization of DNA conformations.

In the 45A4 simulations starting from either B- or A-DNA conformation at 0.1 M salt concentration (E and F) or from the B-DNA conformation at 0.0 M (C), the double-helical structure is well maintained, as shown by the Watson–Crick hydrogen-bond occupancies (Fig. 5). The simulation started from an A-DNA conformation at 0.0 M NaCl (D) appears to result in somewhat distorted structures, with the base-pair hydrogen bonds disordered in the central (ATTG) region (Fig. 5). Longer simulation times would be required to measure the persistence of this disorder. For the four simulations, the average sugar pucker pseudorotation angle falls in the C2′-endo range (Table 3) and the intrastrand phosphate–phosphate distances converge rapidly to values in the range 0.67–0.68 nm (Fig. 7). These structural parameter values are representative of a B-DNA conformation. In contrast, the rise as well as other structural parameters determining the conformation of the bases show values more typical of the A-DNA conformation (Table 5). In particular, two key helical parameters clearly differ between A-DNA and B-DNA conformations. One is the inclination of the base pairs (average values in DNA structures of different sequences are typically about 15° for A-DNA and around 2° for B-DNA, respectively). The other one is the x-displacement of the base pairs from the helical axis (typically ~0.4 nm for A-DNA and 0.1 nm for B-DNA, respectively). The corresponding values during the simulations are about 15° and ~0.03 nm (simulation C starting from B-DNA) or 15.8° and ~0.31 nm (simulation D starting from A-DNA). The major limitation related to the use of such helical parameters to characterize DNA conformations is that it is unclear how these quantities deviate between structures in solution and in a crystal. Helicoidal parameters cannot be measured directly by means of NMR spectroscopy, but must be derived from interproton distance bounds, and are therefore difficult to obtain in nonglobular systems of elongated shape such as DNA.

Starting from an A-DNA or a B-DNA conformation leads, as expected, to similar conformational ensembles. This also follows from the observation that all simulations using the 45A4 parameter set fulfill the experimental NOE distance bounds equally well (Fig. 8). Only simulation D at 0.0 M NaCl starting from an A-DNA structure shows some deviation for the four central base pairs that display distorted hydrogen bonds.

**Influence of the Counterion Concentration**

The effect of the ionic environment on the conformational ensemble representative of a DNA helix was examined by performing MD simulations at two salt concentrations of 0.0 and 0.1 M. The arrangement of cations along the DNA spine stabilizes the helical conformation by neutralizing the negative phosphate charges. Although cations are predominantly attracted to the phosphate oxygens, a favorable electrostatic potential is also found along the minor and major grooves. The X-ray structures of the dodecamer exhibit ordered hydration sites in the minor groove along the sequence AATT, known as the spine of hydration. It has been shown by means of high-resolution crystallography and magnetic relaxation dispersion techniques that monovalent cations are located primarily within the minor and major grooves and share these sites with the water molecules. Cations actually exhibit the highest affinity for the minor groove, binding to this region with correlation times of 50 ns.

The average structures corresponding to simulations (E and F) with the 45A4 parameter set that started from either A- or B-DNA
conformations at 0.1 M NaCl differ by a RMSD of 0.2 nm, whereas the average structures of simulation (C and E) at 0.0 and 0.1 M NaCl starting from a B-DNA conformation differ by 0.07 nm. In comparison, the canonical A- and B-forms of the dodecamer have a RMSD of about 0.50 nm. This structural similarity of the MD ensembles is reflected in the distribution pattern of ordered sodium ions around the A- and B-DNA conformations. The cation binding sites with highest occupancies (larger than 10%) are located in all cases in the groove regions (Fig. 9).

Nature of the A- and B-DNA Conformational States in Solution

X-ray crystallography has made possible the identification and characterization of different conformational families of DNA. The transition between each of these families is modulated by the DNA sequence as well as the solvent and ionic environment. Crystal packing can also influence DNA structure, as has been evidenced by crystal structures of the same sequence showing different conformations depending on the crystal form, or by sequences that adopt a B-DNA conformation in solution but crystallize into A-DNA conformations. The extent to which crystal packing forces constrain the DNA structure remains an open question. Therefore, crystallographic structures must not necessarily be representative of the ensemble of structures of DNA in solution.

Alternatively, NMR spectroscopy is the primary source for DNA structures in solution. It also has the advantage of allowing for structure determination under a wide range of solution conditions, a key factor affecting DNA polymorphism. Yet the NMR structure determination of nucleic acids has turned out to be more challenging than for proteins. Its quality is critically limited by the low proton density and the lack of long-range distance information due to the elongated shape of oligonucleotide complexes. The use of NOE proton–proton distances and 3J-coupling constants for structure determination permits at most an approximate determination of local geometry.

Molecular dynamics simulations can provide a detailed picture of molecular configurations at the atomic level, provided they rely on realistic force-field parameters. Although a rigorous validation of MD models for molecules in solution requires comparison against experimental data in solution, it has remained common practice to evaluate the performance of force-fields by comparison with structures obtained from X-ray crystallography. However, nucleic acids are not the static structures suggested by the crystallographic data, but rather flexible and dynamical systems. For example, the ribose of B-DNA adopts C2'-endo values in the crystalline state, whereas in solution, the average conformation falls into the C1'-exo region due to the rapid equilibrium between C2'-endo and C3'-endo conformations.

The structures generated from the simulations C and E using the improved parameter set 45A4 starting from the X-ray B-DNA conformation satisfy the NOE proton–proton distances better than the crystallographic B-DNA conformation (Fig. 8, G), while the comparison of the simulated structures against this structure results in an atom-positional RMSD value in the range 0.33–0.35 nm. Although it seems possible to distinguish between the canonical A- and B-forms on the basis of the experimental set of NOE-derived distances, the former exhibiting more and larger NOE bound violations, both NOE analyses on single structures show relatively many violations (Fig. 8, G and H). In fact, the data do not distinguish as clearly between B-DNA and A-DNA conformations, which can also be interpreted as A- and B-DNA not being discrete conformations in solution. In other words, rather than pure A- and B-DNA, one should think of a structural continuum of intermediate conformations. In support of this observation, several high-resolution crystal structures of DNA oligomers in intermediate conformations between A- and B-DNA have recently been published. A set of these structures defines a complete pathway from B- to A-DNA through 13 single-crystal structures corresponding to the same sequence. In these intermediate conformations, structural properties that are characteristic of A-DNA coexist with the ones typical of B-DNA. In this regard, the GROMOS parameter set 45A4 reproduces the structural properties of nucleic acids in the solution state.

Conclusion

In this study, we performed six molecular dynamics simulations of the Dickerson–Drew dodecamer sequence 5'-[CGCGAATTCGCG]-3', two using the GROMOS parameter set 45A3 for polypeptides and four using the new set 45A4. The simulations started from either B-DNA (crystallographic) or A-DNA (modeled) conformations, and involved either 0.0- or 0.1-M salt concentrations. Comparison between the simulated and experimental data was made on the basis of the analysis of atom-positional RMSD from the X-ray and model structures, atom-positional RMSF, proton–proton NOE distance-bound violations, J-coupling constants, occurrences of canonical hydrogen bonds, distribution of backbone dihedral angles, and local/global geometrical parameters of DNA.

The previous GROMOS parameter set 45A3 tends to yield ensembles of structures that are characterized by large atom-positional fluctuations, and a loss of canonical hydrogen bonds between the corresponding base pairs, in particular at the strand termini.

The new parameter set 45A4 differs from the 45A3 one by (1) the inclusion of additional backbone dihedral-angle potential energy terms and the refinement of the parameters of the existing ones, (2) the addition of explicit aromatic hydrogen atoms bound to carbon atoms in the nucleotide bases, and (3) a modification of the atomic partial charges. The simulations performed using this new parameter set exhibit lower atom-positional RMSD from the canonical DNA structures than the corresponding simulations using the 45A3 parameter set. The atom-positional fluctuations in the terminal regions of the DNA duplex are reduced, leading to more stable hydrogen bonds between complementary base pairs. Moreover, the corresponding average positive distance-bound violations are smaller than for the X-ray crystal conformation.

The validation of force-field parameter sets for molecular dynamics simulation through MD simulations of biomolecules is an essential step towards the understanding of their behavior. However, this validation is not straightforward in the specific case of nucleic acids, because these are highly charged systems for which condensed-phase data is scarce. Issues such as the susceptibility of...
DNA simulations to errors related to numerical truncation of long-range forces and the extent by which X-ray structures are affected by crystal packing forces are difficult to separate from the one concerning the quality and performance of a given force field. Yet, the new GROMOS parameter set proposed here appears to reproduce recent NMR data on DNA in solution as accurately as the uncertainties associated to these data permit.

References

87. Dickerson, R. E.; Ng, H. L. Proc Natl Acad Sci USA 2001, 98, 6986.
88. Ng, H. L.; Kopka, M. L.; Dickerson, R. E. Proc Natl Acad Sci USA 2000, 97, 2035.