

# Evaluation of DNA Force Fields in Implicit Solvation

Thomas Gaillard<sup>†,‡</sup> and David A. Case<sup>\*,†</sup>

<sup>†</sup>BioMaPS Institute for Quantitative Biology, Rutgers—The State University of New Jersey, Piscataway, New Jersey 08854-8087, United States

<sup>‡</sup>Laboratoire de Biochimie (CNRS UMR7654), Department of Biology, Ecole Polytechnique, 91128 Palaiseau, France

**ABSTRACT:** DNA structural deformations and dynamics are crucial to its interactions in the cell. Theoretical simulations are essential tools used to explore the structure, dynamics, and thermodynamics of biomolecules in a systematic way. Molecular mechanics force fields for DNA have benefited from constant improvements during the past decades. Several studies have evaluated and compared available force fields when the solvent is modeled by explicit molecules. On the other hand, few systematic studies have assessed the quality of duplex DNA models when implicit solvation is employed. The interest in an implicit modeling of the solvent consists of the important gain in simulation performance and conformational sampling speed. In this study, respective influences of the force field and the implicit solvation model choice on DNA simulation quality are evaluated. To this end, extensive implicit solvent duplex DNA simulations are performed, attempting to reach both conformational and sequence diversity convergence. Structural parameters are extracted from simulations and statistically compared to available experimental and explicit solvation simulation data. Our results quantitatively expose the respective strengths and weaknesses of the different DNA force fields and implicit solvation models studied. This work can lead to the suggestion of improvements to current DNA theoretical models.

## INTRODUCTION

Deoxyribonucleic acid (DNA) is the molecule carrying genetic information, which is translated into proteins by living cells, using a code almost universally conserved among living systems. DNA is involved in many dynamic processes like gene expression and regulation or genetic replication and recombination. DNA molecule structural plasticity is critical in many of these mechanisms, including protein or cofactor recognition, chromatin remodeling, and transcription. Understanding DNA structural deformations and dynamics is thus essential to the study of important components of living systems.

Experimental techniques like X-ray crystallography, NMR spectroscopy, or single molecule experiments greatly contribute to our increasing knowledge of DNA molecule properties. Theoretical simulations are complementary techniques, useful in the investigation of structure, dynamics, and thermodynamics of biomolecules in a systematic way, in particular when experimental data are not available or accessible. The first molecular dynamics simulations of DNA were performed in 1983<sup>1,2</sup> in vacuo and two years later with explicit water molecules and counterions.<sup>3</sup> DNA molecular mechanics models have thus benefited from almost three decades of force field and algorithmic improvements<sup>4–12</sup> and have been applied to a wide variety of questions.<sup>13–21</sup>

A number of systematic studies have attempted to evaluate DNA molecular mechanics models.<sup>22–29</sup> Only the most recent works are described here. In a collaborative work initiated by the “Ascona B-DNA Consortium”, 136 unique tetranucleotide duplexes were studied by molecular dynamics. Simulations were performed over 15 ns, using the AMBER parm94 force field, with an explicit solvation model, on 39 oligomers of 15 bp length containing the tetranucleotides. Methodological questions are addressed in a first paper,<sup>25</sup> and sequence context effects on the DNA structure are discussed in a second paper.<sup>26</sup> The authors observed structural substates distinct from the B form. These states are largely controlled by the backbone conformation,

which in turn is strongly correlated to helicoidal parameters. The effect of flanking base pairs on a dinucleotide was found to be limited for YpR steps, although they are more flexible, and more significant for RpR and RpY steps, which are more rigid. Another study by Fujii et al.<sup>27</sup> concentrated on the sequence dependence of DNA deformability. Molecular dynamics simulations of 10 ns length in an aqueous environment were performed on 12 bp duplexes, comprising the 136 tetramers sandwiched between CGCG sequences. The AMBER parm99 force field was used. The authors showed that the deformability of dimeric steps is consistent with crystal structure data and can be affected by flanking base pairs. Orozco and co-workers<sup>28</sup> conducted a systematic study of B-DNA flexibility with CHARMM all27 and AMBER parmbsc0 force fields. They performed 100 ns molecular dynamics simulations in an explicit solvent environment on four different 18 bp sequences, chosen to contain the 10 unique dinucleotide base pair steps. This study highlighted some differences between parmbsc0 and all27 force fields, yet the authors concluded that today DNA force fields have matured, leading to a consensus view of B-DNA flexibility. A continuation of the Ascona consortium effort was recently published.<sup>29</sup> This work investigated nearest-neighbor effects on B-DNA base pairs and base pair steps. A number of 39 oligomers containing all unique tetranucleotides, each of 18 base pair length, were subjected to 50 or 100 ns molecular dynamics simulations in explicit solvation, with the parmbsc0 force field. The authors concluded that simulations were converged in terms of most B-DNA conformational properties. Sequence effects were found to be small at the base or base pair level, more important at the tri- or tetranucleotide level, and could still be significant beyond the nearest-neighbor level.

**Received:** June 8, 2011

**Published:** August 24, 2011

Solvent modeling is critical in DNA simulations. With an improper solvation model, DNA strand separation can easily be observed. While explicit modeling of solvent molecules provides the most realistic results, the cost of such approaches is high, as an important portion of computer time is spent in simulating bulk solvent molecule dynamics. Implicit solvation models attempt to represent solvent as a continuous medium and capture effects on the solute in an average way. The interest of implicit solvation models lies in reduced computational cost, reduced statistical error, faster conformational sampling, clearer physical insights, and the possibility of modeling big systems like chromatin or systems with large conformational changes. An important class of implicit solvation models includes the methods based on continuum electrostatics where the electrostatic component of the solvation free energy is obtained by solving the Poisson–Boltzmann equation (PB). The generalized Born model<sup>30</sup> (GB) is a computationally efficient approximation of PB, suitable for molecular dynamics simulations. Several GB variants have been implemented in molecular mechanics programs, mostly differing by the way effective Born radii are calculated. Some GB methods are based on the pairwise descreening formulation of Hawkins and co-workers;<sup>31–39</sup> others approximate the solute volume by overlapping Gaussian functions,<sup>40,41</sup> while others calculate Born radii by analytical volume integration.<sup>42–44</sup> Performance comparisons of several GB implementations are available.<sup>45–47</sup>

GB implicit solvation models have been widely applied,<sup>48–51</sup> in particular to proteins and to a lesser extent to nucleic acids.<sup>36,52–56</sup> Srinivasan et al.<sup>52</sup> conducted PB and GB calculations on snapshots extracted from explicit solvent molecular dynamics simulations of DNA, phosphoramidate-modified DNA, and RNA decamers in the A or B form. Free energy estimations confirmed that the B form is preferred for DNA and the A form for RNA and phosphoramidate-modified DNA and that salt inclusion modestly favors the A form. In another work, Srinivasan et al.<sup>53</sup> performed PB and GB calculations on multiple conformations of proteins, 10–12 bp DNA duplexes, and a RNA hairpin. Good agreement is found between GB and PB solvation energies and salt contributions. Tsui and Case<sup>36,54</sup> have simulated 10 bp DNA and RNA duplexes in the A and B forms using molecular dynamics with GB implicit solvation and obtained good agreement with explicit solvent simulations in terms of both structures and energetics. The authors also showed that DNA in the A form converges to the B form about 20 times faster with GB than with explicit solvation. Chocholoušová and Feig<sup>55</sup> have performed molecular dynamics simulations with GB implicit solvation on the DNA Drew–Dickerson dodecamer and a protein–DNA complex. The authors demonstrated that stable and realistic DNA simulations can be obtained, agreement being closer to either explicit solvation results or experimental data, depending on the atomic radii set chosen. Partial transitions to A-DNA could be observed when 1 M salt was added. Molecular dynamics simulations with GB implicit solvation have also been applied to bigger nucleic acid assemblies such as nucleosomal DNA.<sup>56</sup> To our knowledge, there is no systematic evaluation of DNA simulations in implicit solvation published to date.

In this work, we intend to evaluate DNA molecular mechanics models with generalized Born implicit solvents against explicit solvation results and experimental data. Different force fields and implicit solvation methods are implemented and customarily used with different molecular mechanics programs, like CHARMM or AMBER. We need to be able to use any combination of force field and implicit solvation variant to investigate the respective

influence of both on results. We concentrate on two recent and widely used DNA force fields (AMBER parmbsc0 and CHARMM all27) and five generalized Born variants (GBHCT, GBOBC1, GBOBC2, and GBn implemented in AMBER and GBMV implemented in CHARMM). Molecular dynamics simulations are conducted on several DNA duplex sequences, and care is taken that both simulation length and sequence diversity convergence are achieved. Several structural parameters are then measured and statistically compared to both explicit solvation results and experimental data.

## METHODS

**Starting Structures.** DNA duplexes of 12 base pair length with different sequences corresponding to the 5'-CGCGWXYZCGCG-3'/5'-CGCGZYXWCAGC-3' pattern were studied, where WXYZ/ZYXW are tetramer duplexes located at the center and flanked by C–G base pairs. The overline symbol designates the complementary nucleotide. There are 256 combinations of nucleotide tetramers but only 136 unique tetramer duplexes (for example, AAAA/TTTT and TTTT/AAAA cannot be distinguished). Starting coordinates were built with the NAB program distributed with the AMBER suite,<sup>57</sup> using Arnott B-DNA fiber diffraction data.<sup>58</sup>

AMBER prmtop and inpcrd files were generated with the leap program, using the parmbsc0 force field.<sup>12</sup> CHARMM psf and crd files were generated with the CHARMM program<sup>59</sup> using the all27<sup>11</sup> force field. “CHAMBER” (CHARMM force field with the AMBER program) prmtop and inpcrd files were generated with the CHAMBER conversion program<sup>60</sup> distributed with the AMBER suite, from CHARMM psf and crd files, and the all27 rtf and prm files. Differences between CHAMBER and CHARMM energies in vacuo with an infinite cutoff were less than  $5 \times 10^{-4}$  kcal mol<sup>-1</sup>. “AMBARM” (AMBER force field with the CHARMM program) psf and crd files were generated with the CHARMM program compiled with the AMBER keyword to be consistent with AMBER conversion factors and using the parmbsc0 force field converted to CHARMM format by Jeff Klauda (<http://terpconnect.umd.edu/jbklauda/research/download.html>). Differences between AMBARM and AMBER energies in vacuo with an infinite cutoff were less than  $3 \times 10^{-3}$  kcal mol<sup>-1</sup>.

**Molecular Dynamics.** Molecular dynamics simulations with AMBER and CHARMM programs were done with a generalized Born implicit solvation model, using an infinite cutoff for non-bonded interactions, dielectric constants of 78.5 for the solvent and 1 for DNA interior, and an inverse Debye–Hückel length of 0.1 Å<sup>-1</sup> to account for the 0.1 M salt concentration. Minimization was first performed over 200 steps with 5 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic restraints on non-hydrogen atoms. Langevin dynamics was then carried out using a heat bath at 300 K, a collision frequency of 5 ps<sup>-1</sup>, a 1 fs time step, and the SHAKE algorithm was used to restrain the elongation of bonds involving hydrogen atoms. Three equilibration phases of 50 ps were performed with decreasing harmonic restraints on non-hydrogen atoms of 5, 1, and 0.1 kcal mol<sup>-1</sup> Å<sup>-2</sup>, respectively. The production phase of 5 ns was then run without restraints. Simulations with the AMBER program employed GBHCT,<sup>31,52</sup> GBOBC1,<sup>37,38</sup> GBOBC2,<sup>38</sup> or GBn<sup>39</sup> generalized Born model variants. Simulations with the CHARMM program employed the GBMV implicit solvation model,<sup>42,43</sup> with the  $1/r^7$  Coulomb field correction and parameters  $C_0 = -0.1$ ,  $C_1 = 0.9$ ,  $\beta = -12$ , and  $S_0 = 0.65$ , as in Chocholoušová and Feig.<sup>55,61</sup> Recommended atomic radii sets for each GB model variant were used (mbondi<sup>36</sup> for GBHCT,

mbondi2<sup>38</sup> for GBOBC1 and GBOBC2, bondi<sup>62</sup> for GBn, and van der Waals radii for GBMV).

**Analysis.** Structural data were collected from the 5 ns molecular dynamics trajectories of the 136 CGCGWXYZCGCG duplex sequences, each trajectory containing 5,000 frames. Combinations of two force fields (AMBER parmbsc0 and CHARMM all27) and five GB implicit solvation models (AMBER GBHCT, GBOBC1, GBOBC2, GBn, and CHARMM GBMV) led to a total simulation time of 6.8  $\mu$ s and a number of  $6.8 \times 10^6$  frames.

Trajectories were analyzed with the ptraj program of the AMBER suite<sup>57</sup> for root-mean-square deviations (RMSD) calculations, and structural parameters<sup>63</sup> were evaluated with the 3DNA program.<sup>64,65</sup> Parameter measurements were done on the central XY/YX dimer and included base pair step parameters (shift, slide, rise, tilt, roll, twist), groove widths (minor and major), sugar conformational parameters (amplitude and phase angle of ring pseudorotation), backbone and glycosidic torsion angles ( $\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \chi$ ). Only structures with correct Watson–Crick pairing at the central dimer, as defined by 3DNA, were considered.

The data set was then symmetrized by duplicating each entry with parameter values that would have been measured if the other strand was chosen as the “leading strand”. This is necessary because some structural parameters (shift and tilt) have their sign dependent on the choice of the duplex leading strand, which is arbitrary. Structural parameter global averages, standard deviations, and probability density functions were then calculated over the whole data set, using a weighting to correct for sequence composition bias and ensure equal representation to each of the 10 unique dimer steps (AA/TT, AT/AT, AG/CT, AC/GT, TA/TA, TG/CA, TC/GA, GG/CC, GC/GC, CG/C). Dimer level averages, standard deviations, and probability density functions were also calculated over subsets corresponding to each of the 10 unique dimer steps. The probability density function was approximated by a kernel density estimation, using a Gaussian function as the kernel. Periodicity was taken into account when calculating statistics of angular parameters.

**Experimental Data.** Experimental structural data were extracted from the Protein Data Bank<sup>66</sup> and Nucleic Acid Database.<sup>67</sup> The 3DNA Landscapes platform<sup>68</sup> (<http://3dnascapes.rutgers.edu>) was employed to obtain PDB/NDB codes of DNA-containing structures. A list of X-ray structures with a resolution better than 2 Å was retrieved as well as a separate list of NMR structures. Structures were then downloaded and analyzed with the 3DNA program. Unwanted structures were discarded according to the following criteria: dimer steps retained in the data set were those containing only A, T, G, or C nucleotides; with correct Watson–Crick pairing; flanked by A, T, G, or C nucleotides; and belonging to a right-handed duplex structure. Further filtering excluded A-DNA ( $z_p > 1.5$  Å) and TA-DNA ( $z_p(h) > 4.0$  Å) structures, as in Lu and Olson.<sup>64</sup> The number of dimer steps available in the experimental data set is given in Table 1. Statistical analysis of experimental data was conducted as for implicit solvation MD data.

**Explicit Solvation Simulation Data.** Structural data were also obtained from molecular dynamics simulations in explicit solvation performed by Orozco and co-workers.<sup>28</sup> The 100 ns trajectories of four different sequences simulated with both CHARMM all27 and AMBER parmbsc0 force fields were downloaded (<http://mmb.pcb.ub.es/raist/CONSENSUS/>) and uncompressed, as explained in the article. Each of the eight trajectories contained about  $10^5$  frames, of which one every 10 frames was extracted and analyzed with the 3DNA program. Every dimer step with correct

**Table 1. Amount of Dimer Steps Available in Experimental Data<sup>a</sup>**

exptl.	NMR	X-ray
all	2788	3018
AA	388	476
AT	200	273
AG	338	270
AC	343	303
TA	172	187
TG	351	341
TC	353	420
GG	249	266
GC	206	238
CG	188	244

<sup>a</sup>NMR and X-ray (resolution better than 2 Å) structures are extracted from the Protein Data Bank. Dimer steps retained in the data set were those containing only A, T, G, or C nucleotides, with correct Watson–Crick pairing, flanked by A, T, G, or C nucleotides, and corresponding to a right-handed B-DNA duplex structure.

Watson–Crick pairing and not at the extremities of the duplex to avoid end effects was kept in the data set. Statistical analysis of the data set was then performed as for implicit solvation MD data.

## ■ RESULTS

DNA structural parameters statistics, obtained from molecular dynamics simulations with different force fields and generalized Born implicit solvation methods, are presented and compared to experimental X-ray and NMR data, as well as to published explicit solvation simulation data. The validity of simulations and methodological robustness are first discussed. A detailed examination of selected structural parameter statistics is then provided.

**Validity of Simulations. Watson–Crick Pairing.** The Watson–Crick pairing percentage was calculated for each simulation as a measure of DNA duplex structure conservation (Table 2). This was obtained as the ratio of frames for which a correct pairing of the central dimer was detected by the 3DNA program to the total number of frames simulated. Percentages are then averaged over the 136 sequences simulated with each force field and GB method combination. Percentages calculated over subsets corresponding to the 10 different central dimer types are also provided. An examination of the trajectories shows that strands get irreversibly separated after about 500 ps of simulation when the GBn method is employed, thus explaining the low percentages (around 10%) found with GBn. This method is indeed not recommended for nucleic acid simulations by its authors<sup>39</sup> and will not be considered in analyses. In all other simulations, Watson–Crick pairing percentages were always higher than 98.5% on average. The AMBER parmbsc0 force field gave slightly higher percentages than CHARMM all27 in general. GBMV gave the best overall pairing conservation among GB methods, while GBOBC1 had the lowest percentages. Limited variations were observed between different types of central dimer.

**Root Mean Square Deviations.** Root mean square deviations of heavy atoms with respect to the initial conformation were calculated for the different simulations, taking into account either the whole duplex or only the central dimer step (Table 3). The RMSD values given are obtained by conformational averaging over the course of simulations and sequential averaging over the

**Table 2.** Watson–Crick Pairing Percentages of the Central Dimer in MD Simulations<sup>a</sup>

FF	parmbsc0					all27				
	solv.	GBHCT	GBOBC1	GBOBC2	GBn	GBMV	GBHCT	GBOBC1	GBOBC2	GBn
all	99.73	99.72	99.89	14.99	99.99	99.42	98.66	99.34	7.95	99.70
AA	99.38	99.76	99.95	4.58	100.00	99.92	99.27	99.71	4.60	99.99
AT	99.95	99.76	99.96	4.15	100.00	99.91	98.78	99.57	4.40	99.92
AG	99.83	99.79	99.87	12.21	99.99	99.56	99.03	99.49	8.56	99.61
AC	99.79	99.60	99.89	12.75	100.00	99.22	98.13	98.99	6.89	99.11
TA	99.63	99.70	99.99	3.29	100.00	99.57	99.29	99.54	4.92	99.73
TG	99.69	99.67	99.77	11.70	99.99	99.50	98.70	99.47	8.19	99.68
TC	99.90	99.78	99.85	11.69	99.99	99.61	99.14	99.67	7.18	99.98
GG	99.74	99.73	99.86	33.50	99.95	98.52	97.40	98.61	10.21	99.39
GC	99.88	99.86	99.97	27.41	99.99	99.33	98.86	99.40	14.32	99.89
CG	99.50	99.58	99.83	30.76	99.99	99.18	98.21	99.01	11.42	99.96

<sup>a</sup> Percentages are obtained as the ratio of frames with correct Watson–Crick pairing at the central dimer, as measured by the 3DNA program, to the total number of trajectory frames. They are calculated for all frames as well as over subsets corresponding to each of the 10 unique dimer steps.

**Table 3.** Heavy-Atom Root Mean Squared Deviation (Å) Averages with Respect to Initial Conformation<sup>a</sup>

FF	GB	duplex	central dimer
parmbsc0	GBHCT	3.45	1.13
parmbsc0	GBOBC1	4.26	1.21
parmbsc0	GBOBC2	4.08	1.17
parmbsc0	GBn	15.89	9.41
parmbsc0	GBMV	2.85	0.92
all27	GBHCT	2.68	0.86
all27	GBOBC1	3.82	0.96
all27	GBOBC2	3.58	0.91
all27	GBn	17.77	11.43
all27	GBMV	2.98	0.77

<sup>a</sup> RMSDs are calculated on the whole 12 bp duplex as well as on the central dimer only and averaged over trajectories of the 136 sequences simulated.

136 sequences simulated with each force field and GB variant. Except GBn, for which strand separation occurs, RMSD calculated on the whole 12 bp duplex ranged from 2.7 to 4.3 Å on average, and from 0.8 to 1.2 Å when only the central dimer was considered. Slightly higher values are in general obtained with the parmbsc0 force field. The GB method choice influences RMSD in a consistent way between both force fields, namely, GBOBC1 > GBOBC2 > GBHCT ≈ GBMV. The highest RMSD was obtained with GBOBC1 and then GBOBC2, and the lowest values were obtained with the GBMV method with both force fields, or with the all27+GBHCT combination. It should be noted that structural measurements in analyses are done exclusively on the central dimer; results will thus not be directly affected by deformations at the extremities of the duplex. Another remark is that initial structures are not necessarily considered as a reference in this work, where we assess the validity of GB simulations with respect to explicit solvation simulation and experimental data. Initial structures are indeed ideal B-DNA structures based on fiber diffraction data, which are not taking into account sequence effects on backbone conformation.

**Convergence and Robustness.** Convergence and robustness of the results were assessed by several preliminary tests, employing energy minimizations or molecular dynamics. It was found

that DNA structural parameters were mostly converged with respect to simulation time (5 ns), sequence diversity (136 unique duplex tetramers), and duplex length (12 bp). Results were in general not significantly sensible to the choice of the starting conformation, salt concentration in the reasonable range, or GB radii set.

**Structural Parameters.** Structural parameter averages are given in Table 4. Base pair step, groove width, backbone and glycosidic torsions, and sugar puckering parameters are discussed in detail, on the basis of a comparison of averages and whole probability distributions. Most of the time, distributions are close to Gaussian shape, and averages are a good approximation of the maximum of probability. As a general pattern, we first compare NMR to X-ray data, then force fields to experimental data, and finally GB implicit solvation methods to explicit solvation.

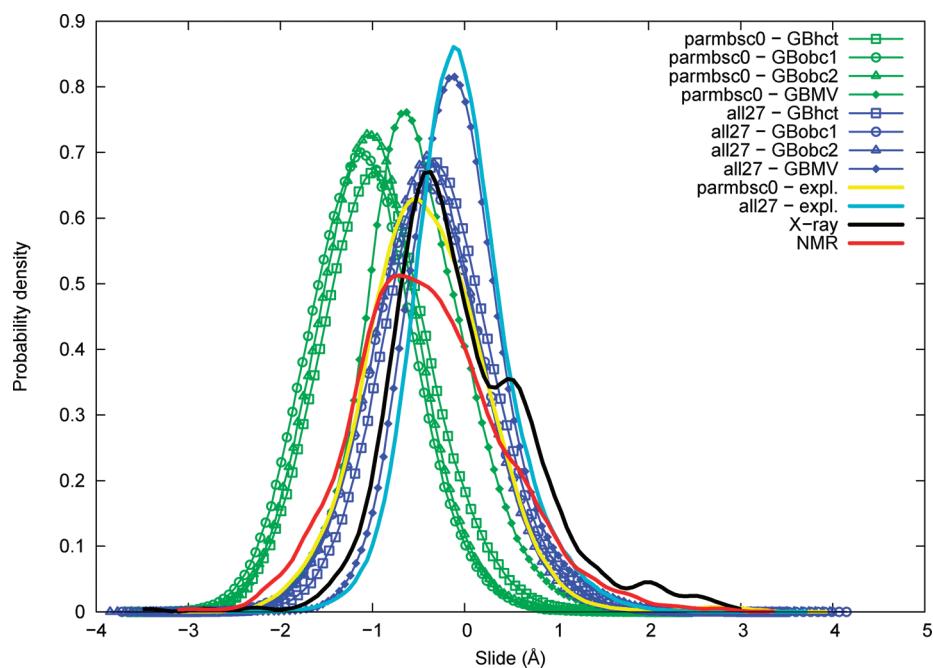
**Base Pair Step Parameters.** Base pair step parameters describe the geometry of one base pair with respect to the next one. There are three translational base pair step parameters (shift, slide, rise) and three rotational parameters (tilt, roll, twist). The shift and tilt parameters, presenting a sign inversion depending on the choice of the leading strand of the duplex, have an average value of zero by construction and will not be discussed in detail here.

**Slide.** Slide distributions are given in Figure 1. The slide average appears noticeably lower in NMR data ( $-0.38 \text{ \AA}$ ) compared to X-ray data ( $-0.02 \text{ \AA}$ ). The difference in average value between NMR and X-ray values is partly explained by a secondary peak around  $0.5 \text{ \AA}$  in the X-ray distribution, which otherwise has a maximum of probability around  $-0.40 \text{ \AA}$ , thus close to the NMR average value. Sequence variability is important in experimental slide distributions, and the X-ray peak at  $0.5 \text{ \AA}$  is mostly contributed by CG and GC steps. The slide average obtained with the AMBER parmbsc0 force field in explicit solvation ( $-0.40 \text{ \AA}$ ) is close to the NMR results in average, whereas the CHARMM all27 force field ( $-0.02 \text{ \AA}$ ) is closer to X-ray results. However, when considering the maximum of probability, slide is lightly overestimated with the all27 force field. We have thus NMR ≈ parmbsc0 < X-ray < all27. The GB method influence is consistent between the two force fields, with a general underestimation of the slide parameter, compared to explicit solvation. GBOBC1 and GBOBC2 are the most affected, then comes GBHCT, and GBMV is the closest to explicit solvation, namely, GBOBC1 ≈ GBOBC2 < GBHCT < GBMV < explicit. Distributions obtained from simulations are close to a Gaussian

**Table 4. Structural Parameter Averages<sup>a</sup>**

FF	parmbsc0							all27					
	solv.	NMR	X-ray	expl.	GBHCT	GBOBC1	GBOBC2	GBMV	expl.	GBHCT	GBOBC1	GBOBC2	GBMV
shift	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
slide	-0.38	-0.02	-0.40	-0.95	-1.12	-1.04	-0.52	-0.02	-0.28	-0.38	-0.40	-0.08	
rise	3.33	3.34	3.34	3.48	3.51	3.42	3.35	3.36	3.50	3.61	3.53	3.39	
tilt	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
roll	2.33	2.45	3.59	3.75	5.53	6.28	4.67	5.60	4.60	6.50	6.97	7.69	
twist	33.87	34.10	32.44	30.96	29.65	30.25	32.70	34.15	33.50	32.09	32.44	34.40	
minor	13.2	12.7	13.0	13.2	14.1	14.0	12.7	13.8	13.3	14.4	14.4	14.4	
major	18.1	18.3	19.3	21.8	22.7	21.9	19.1	17.1	18.8	19.7	19.2	15.7	
$\alpha$	-64.7	-55.6	-73.6	-69.5	-69.9	-69.8	-68.5	-59.9	-60.1	-61.4	-61.7	-60.6	
$\beta$	179.4	172.0	169.7	173.0	172.7	172.2	170.4	168.0	171.7	172.5	173.1	172.1	
$\gamma$	53.5	46.2	55.6	57.4	57.5	57.4	58.3	51.7	52.3	52.0	51.8	50.2	
$\varepsilon$	-177.6	-166.2	-162.1	-171.5	-172.1	-172.5	-170.5	-166.6	-169.7	-168.9	-169.8	-169.6	
$\zeta$	-99.5	-109.0	-102.3	-94.4	-90.8	-90.6	-93.1	-108.0	-106.8	-105.0	-103.4	-100.9	
$\chi$	-114.4	-110.9	-117.2	-123.6	-125.7	-124.7	-122.4	-112.3	-111.8	-111.4	-111.8	-113.9	
P	145.0	145.0	133.2	125.1	120.4	122.0	126.4	143.9	146.1	145.2	146.7	146.5	
$\tau_m$	33.7	36.0	33.9	37.7	37.6	37.3	35.3	35.1	39.4	39.5	38.8	37.6	

<sup>a</sup> Structural parameters are measured with the 3DNA program on the central dimer and averaged over trajectories of the 136 sequences simulated. Global averages are obtained by applying a weighting to ensure equal composition of the 10 unique dimer steps. Angular parameters' periodicity is taken into account. Parameter averages from experimental data and explicit solvation simulations of Orozco and co-workers<sup>28</sup> are calculated following the same protocol. Averages are given in Å for translational step parameters and groove widths and in degrees for angular parameters.



**Figure 1.** Slide probability distribution. Structural parameters are measured with the 3DNA program on the central dimer over trajectories of the 136 sequences simulated. Distributions are obtained from series with a kernel density estimation, using Gaussian functions as a kernel. A weighting is applied to ensure equal composition of the 10 unique dimer steps. Angular parameters' periodicity is taken into account. Parameter distributions from experimental data and explicit solvation simulations of Orozco and co-workers<sup>28</sup> are calculated following the same protocol.

shape. We see here that structural deformations resulting from the GB method choice can be of comparable magnitude to force field differences. Another point is that the GB method effect is here relatively independent of the force field.

**Rise.** Rise distributions are given in Figure 2. The rise average is very close between NMR (3.33 Å) and X-ray (3.34 Å). The

maximum of probability is however slightly smaller in NMR (around 3.23 Å) compared to X-ray (around 3.27 Å). Both force fields in explicit solvation are very close in average to experimental data (3.34 Å for parmbsc0 and 3.36 Å for all27) but slightly higher in terms of the maximum of probability, namely, NMR < X-ray < parmbsc0 ≈ all27. The AMBER GB methods

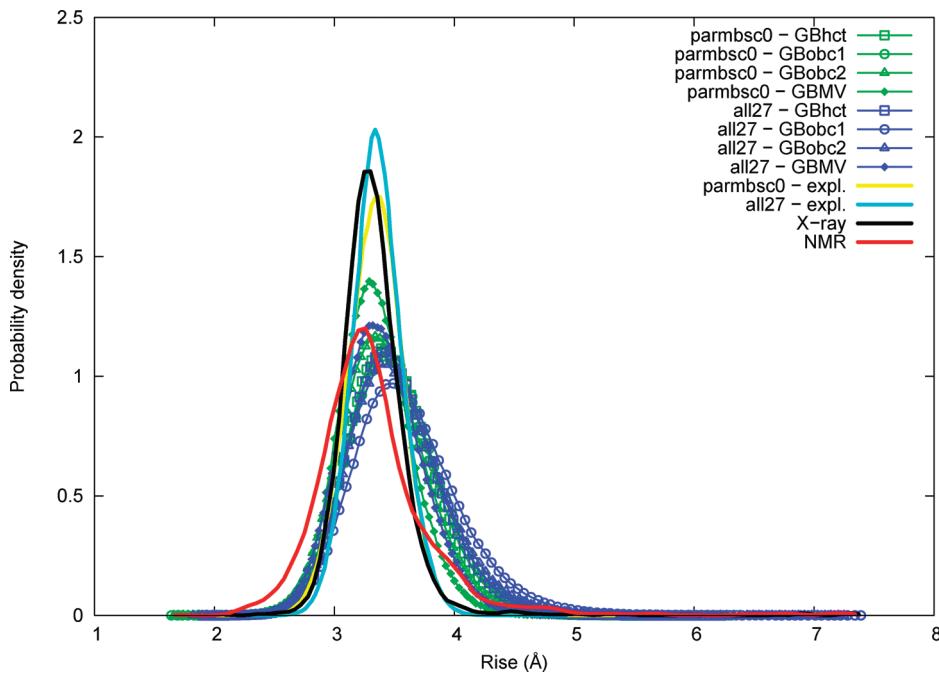


Figure 2. Rise probability distribution. See Figure 1 legend.

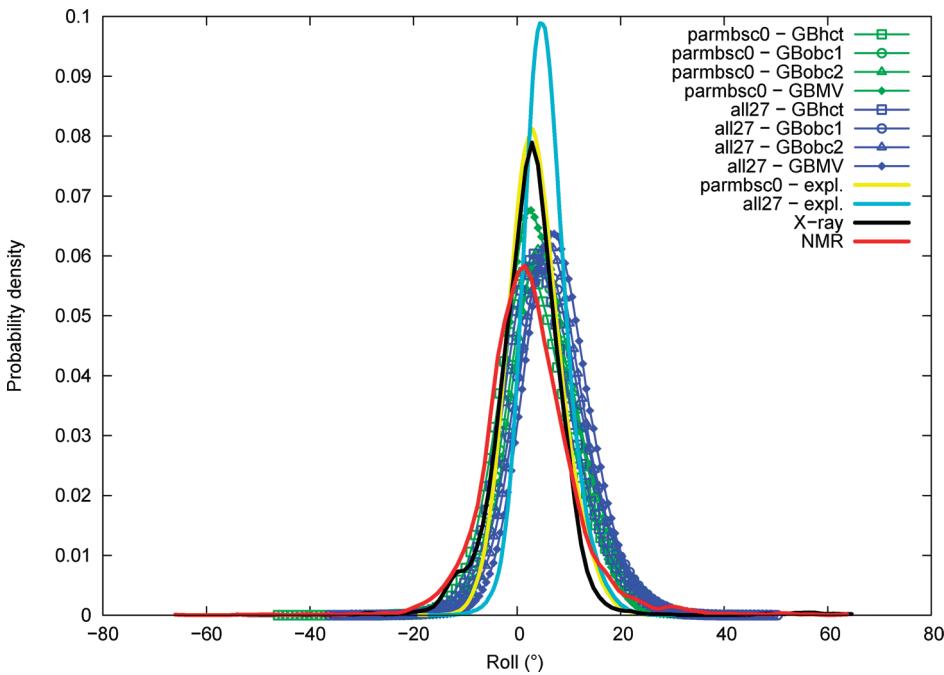
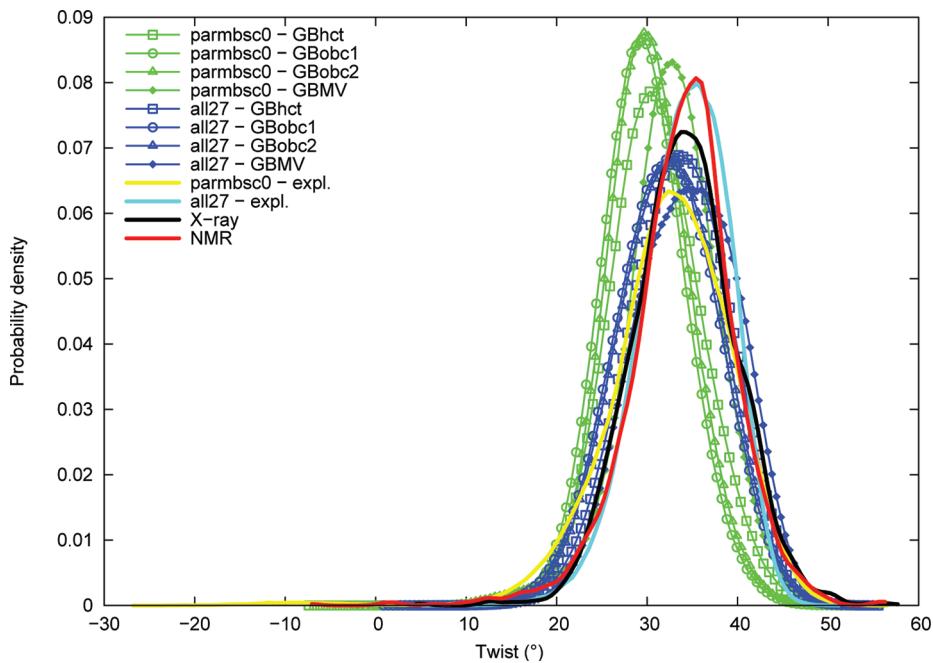


Figure 3. Roll probability distribution. See Figure 1 legend.

(GBHCT, GBOBC1, and GBOBC2) slightly overestimate rise, in particular GBOBC1, whereas the CHARMM GBMV method is close to explicit solvation, namely, explicit  $\approx$  GBMV  $<$  GBHCT  $\approx$  GBOBC2  $<$  GBOBC1. Compared to slide, both the GB method and force field effects are of lower relative importance for rise.

**Roll.** Roll distributions are given in Figure 3. The roll average is slightly lower for NMR ( $2.33^\circ$ ) compared to that for X-ray ( $2.45^\circ$ ). Both force fields appear over-rolled, in particular the all27 force field ( $5.60^\circ$ ), while the parmbsc0 force field ( $3.59^\circ$ ) is closer to experimental values. The distributions indicate that

parmbsc0 is in very good agreement with the X-ray results, with a maximum of probability around  $2.8^\circ$ , we have then NMR  $<$  X-ray  $\approx$  parmbsc0  $<$  all27. The influence of the GB method choice on roll is limited, and GB results stay within the range of force field or experimental data differences. We note however that among the three AMBER implementations, GBOBC1 and GBOBC2 give a slightly higher roll than GBHCT. The effect of GBMV is not clear, as it increases roll with all27 but stays in good agreement with explicit solvation when used with parmbsc0.



**Figure 4.** Twist probability distribution. See Figure 1 legend.

**Twist.** Twist distributions are given in Figure 4. The twist average is close between NMR ( $33.87^\circ$ ) and X-ray ( $34.10^\circ$ ). The parmbsc0 force field is markedly under-twisted ( $32.44^\circ$ ), whereas the all27 force field is closer to experimental values ( $34.15^\circ$ ). The maximum of probability confirms this trend with about  $35.4^\circ$  for NMR,  $34.1^\circ$  for X-ray,  $32.5^\circ$  for parmbsc0, and  $35.4^\circ$  for all27; we have thus  $\text{parmbsc0} < \text{X-ray} \approx \text{NMR} \approx \text{all27}$ . The GB method influence on twist is consistent between the two force fields. AMBER GB methods lead to a significant twist lowering, until modal values around  $29.4^\circ$  for parmbsc0+GBoBC1, whereas GBMV is close to explicit solvation values; namely, we find  $\text{GBoBC1} < \text{GBoBC2} < \text{GBHCT} < \text{explicit} \approx \text{GBMV}$ . Both the force field and GB method effects are thus important for twist.

**Groove Widths.** Groove widths at the central dimer step are calculated as the average of interstrand distances  $P7-p4$  and  $P8-p5$  for the minor groove and as the  $P4-p8$  distance for the major groove, where  $P_i$  is the  $i$ th phosphate atom of strand I counting from the 5' extremity and  $p_j$  is the  $j$ th phosphate of strand II counting from the 3' extremity, as in El Hassan and Calladine.<sup>69</sup>

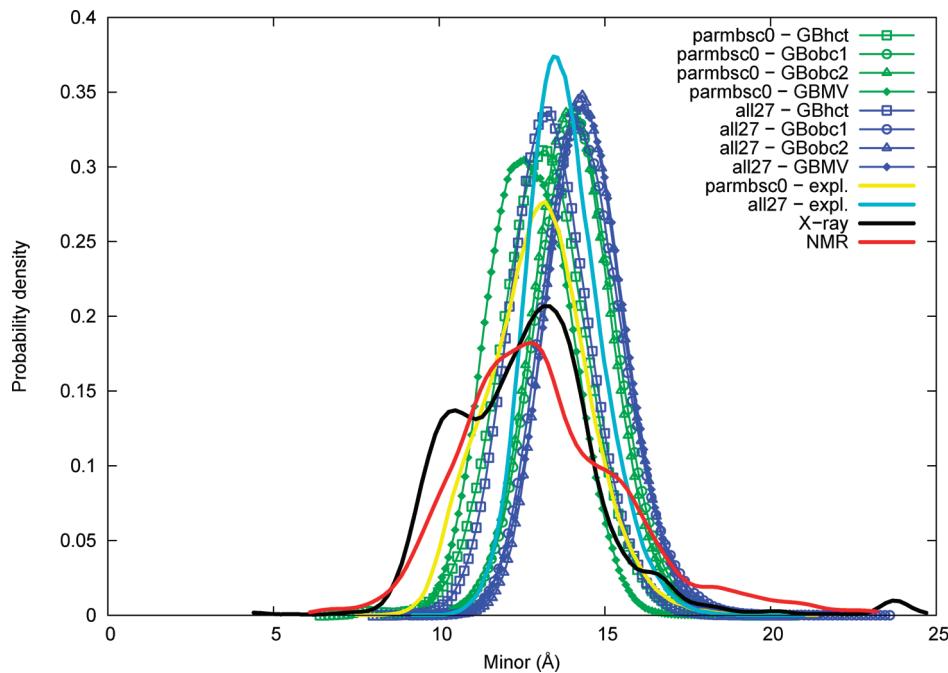
**Minor Groove.** Minor groove width distributions are given in Figure 5. Experimental distributions have a broad shape and present some differences, but averages are on the same order ( $13.2\text{ \AA}$  for NMR,  $12.7\text{ \AA}$  for X-ray), as well as maxima of probability ( $12.8\text{ \AA}$  for NMR,  $13.2\text{ \AA}$  for X-ray). A secondary peak is present in the X-ray data at about  $10.4\text{ \AA}$ , mostly contributed by TG, AT, and AG dimer steps. A shoulder is observed in the NMR data around  $15\text{ \AA}$ , partly contributed by the GC step. These differences in experimental distributions are probably reflective of a lack of data at the underlying dimer level; however, convergence seems quite reasonable at the global level, as indicated by the proximity of NMR and X-ray averages and maxima of probability. The parmbsc0 force field with explicit solvation lies in the experimental range ( $13.0\text{ \AA}$ ), whereas the all27 force field gives a lightly wider minor groove ( $13.8\text{ \AA}$ ). We have thus  $\text{X-ray} \approx \text{parmbsc0} \approx \text{NMR} < \text{all27}$ . The GBoBC1 and GBoBC2 methods

tend to increase the minor groove width with both force fields. GBHCT is the closest to explicit solvation with the two force fields, in good agreement with parmbsc0 and slightly lower with all27. GBMV influence is not consistent between both force fields, leading to slightly lower values than explicit solvation with parmbsc0 and overestimated values with all27.

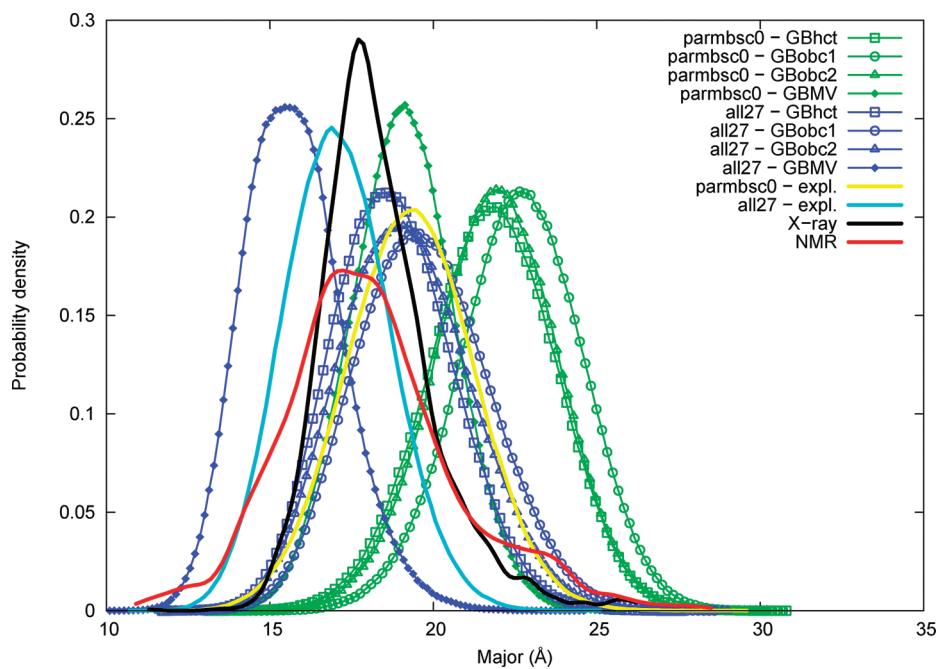
**Major Groove.** Major groove width distributions are given in Figure 6. Major groove width averages are on the same order between NMR ( $18.1\text{ \AA}$ ) and X-ray ( $18.3\text{ \AA}$ ). The maxima of probability are  $17.2\text{ \AA}$  and  $17.7\text{ \AA}$  for NMR and X-ray, respectively. The parmbsc0 force field in explicit solvation has a noticeably overestimated major groove width ( $19.3\text{ \AA}$ ); on the contrary, the all27 force field underestimates major groove width ( $17.1\text{ \AA}$ ). We have thus  $\text{all27} < \text{NMR} < \text{X-ray} < \text{parmbsc0}$ . The GBMV method lowers major groove width, in particular with the all27 force field (average value as low as  $15.7\text{ \AA}$ ). AMBER GB methods, on the other hand, tend to importantly increase major groove width in a consistent order (average value as high as  $22.7\text{ \AA}$  with parmbsc0+GBoBC1); we have thus  $\text{GBMV} < \text{explicit} < \text{GBHCT} < \text{GBoBC2} < \text{GBoBC1}$ . We observe here that the major groove width is very sensible to the GB method choice, and also but to a lesser extent to the force field choice.

**Backbone and Glycosidic Torsion Parameters.** Backbone and glycosidic torsional parameters  $\alpha, \beta, \gamma, \varepsilon, \zeta$ , and  $\chi$ , as canonically defined, will be discussed here. The  $\delta$  torsion angle is already indirectly taken into account in sugar ring parameters. Conformational substates *gauche*–, *gauche*+, and *trans* are defined following a classical 3-fold staggered pattern as  $-60 \pm 60^\circ$ ,  $60 \pm 60^\circ$ , and  $180 \pm 60^\circ$ , respectively.

**Alpha ( $\alpha$ ).** Distributions of the  $\alpha$  backbone torsion angle are given in Figure 7. Experimental distributions of  $\alpha$  are mostly Gaussian and slightly smaller in NMR data (average of  $-64.7^\circ$ , maximum of probability around  $-66^\circ$ ) compared to those in X-ray (average of  $-55.6^\circ$ , maximum of probability around  $-58^\circ$ ), both corresponding to a *gauche*– conformation. A small secondary peak around  $35^\circ$  is present in the X-ray distribution



**Figure 5.** Minor groove width probability distribution. See Figure 1 legend.

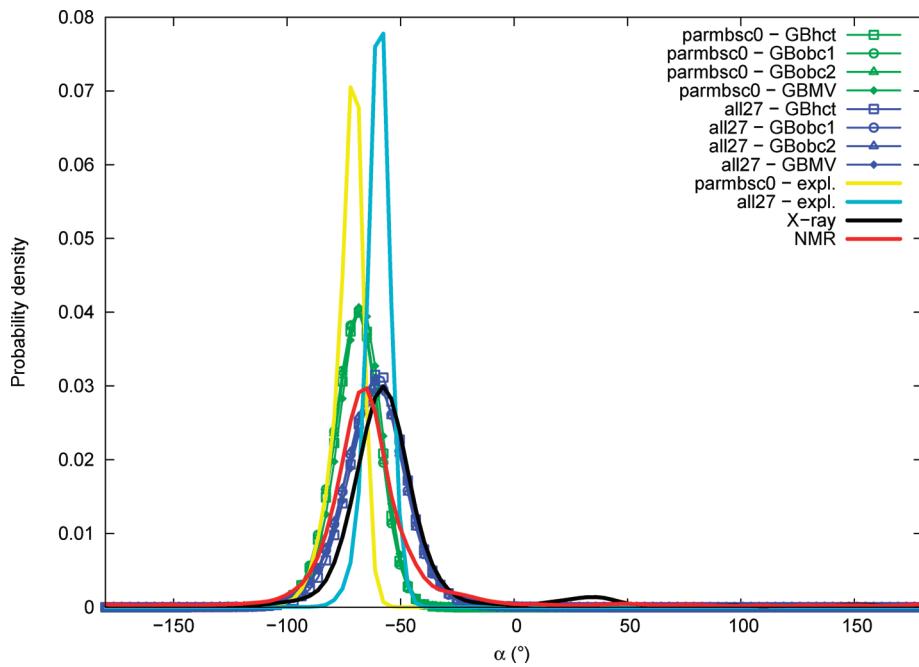


**Figure 6.** Major groove width probability distribution. See Figure 1 legend.

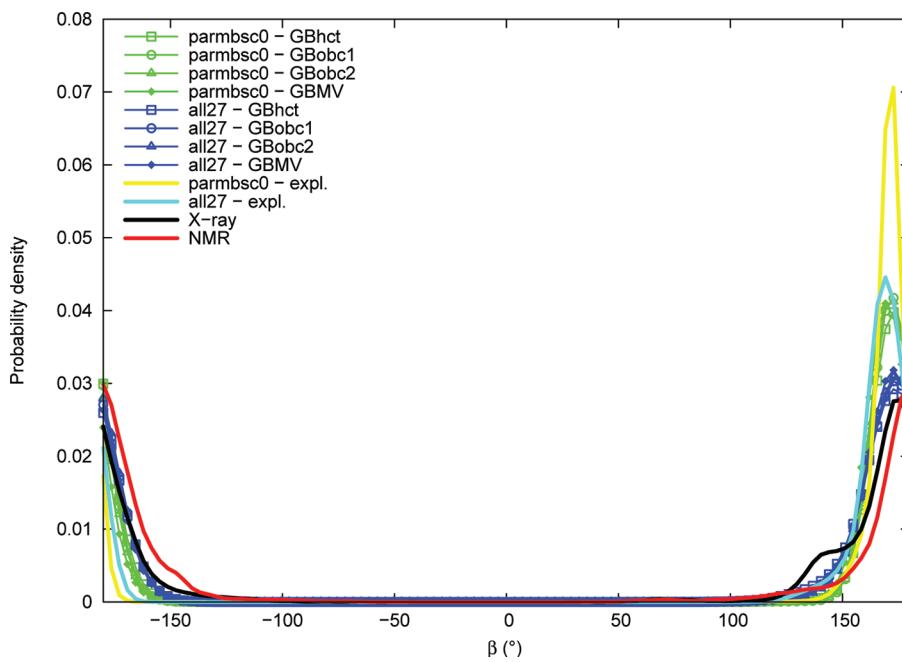
corresponding to a *gauche+* conformation. The  $\alpha$  angle distributions from simulations are all Gaussian-shaped with no significant secondary peaks. The parmbsc0 force field in explicit solvation gives significantly lower  $\alpha$  angles (average of  $-73.6^\circ$ , maximum of probability around  $-70^\circ$ ), whereas the all27 force field lies in the experimental range (average of  $-59.9^\circ$ , maximum of probability around  $-60^\circ$ ). We have thus parmbsc0 < NMR < all27 < X-ray. The use of GB implicit solvation methods leads to a slight increase of the average  $\alpha$  torsion angle compared to explicit

solvation when used with the parmbsc0 force field and is in good agreement or leads to a slight decrease with all27. A widening of distributions is observed for both force fields when GB methods are used. We note that the GB method choice has almost no influence here, the four GB method distributions being almost superimposable.

**Beta ( $\beta$ ).** Distributions of the  $\beta$  torsion angle are given in Figure 8. The  $\beta$  angle experimental distributions are mostly Gaussian centered around the *trans* conformation, with slightly



**Figure 7.**  $\alpha$  torsion angle probability distribution. See Figure 1 legend.

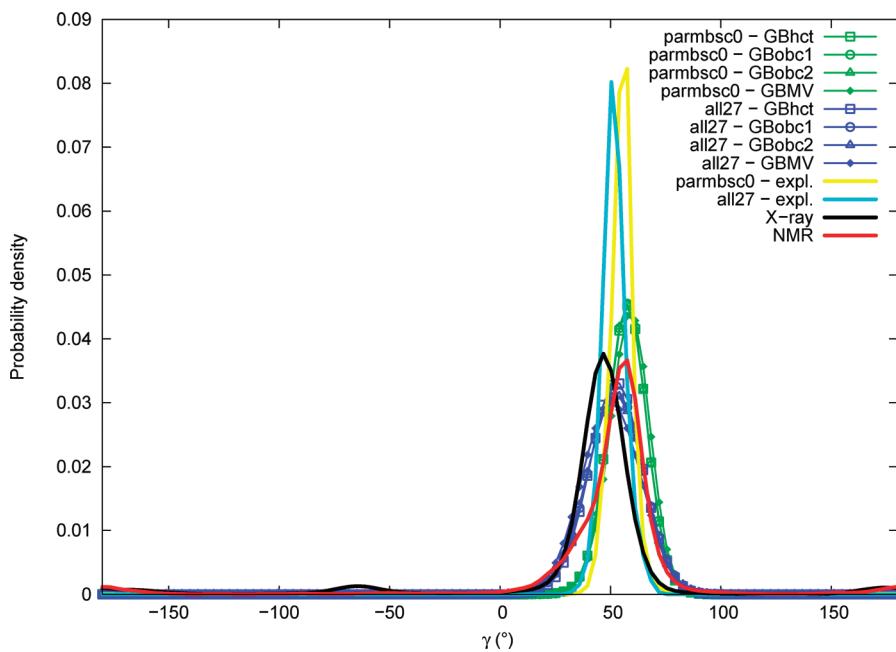


**Figure 8.**  $\beta$  torsion angle probability distribution. See Figure 1 legend.

lower values in X-ray data (average of  $172.0^\circ$ , maximum of probability around  $174^\circ$ ) compared to those in NMR (average of  $179.4^\circ$ , maximum of probability around  $\pm 180^\circ$ ). A small shoulder peak can be noted in the X-ray distribution around  $140^\circ$ . Distributions of  $\beta$  obtained from simulations in explicit solvation are similar to experimental ones but sharper and with slightly lower values for both all27 (average of  $168.0^\circ$ , maximum of probability around  $169^\circ$ ) and parmbsc0 (average of  $169.7^\circ$ , maximum of probability around  $171^\circ$ ) force fields. We have thus all27 < parmbsc0 < X-ray < NMR. The use of GB implicit solvation methods tends to increase the width of distributions

compared to explicit solvation, as well as to slightly shift the main peak toward higher values, in particular with the all27 force field. As for the  $\alpha$  angle, differences between all GB methods are limited.

**Gamma ( $\gamma$ ).** Distributions of the  $\gamma$  torsion angle are given in Figure 9. Experimental distributions of  $\gamma$  comprise a major Gaussian peak in the *gauche+* region, slightly lower in the X-ray data (average of  $46.2^\circ$ , maximum of probability around  $47^\circ$ ) compared to those of the NMR (average of  $53.5^\circ$ , maximum of probability around  $56^\circ$ ). Small secondary peaks can be observed for both distributions in the *trans* conformation region,



**Figure 9.**  $\gamma$  torsion angle probability distribution. See Figure 1 legend.

and in the *gauche*– region only in the X-ray distribution (around  $-65^\circ$ ). The  $\gamma$  angle distributions from simulations are approximately Gaussians centered in the *gauche*+ region, with no significant other peaks. Explicit solvation  $\gamma$  angle values are in the correct experimental range, but slightly lower with the all27 force field (average of  $51.7^\circ$ , maximum of probability around  $52^\circ$ ) than with parmbsc0 (average of  $55.6^\circ$ , maximum of probability around  $56^\circ$ ). We have then X-ray < all27 < NMR < parmbsc0. GB implicit solvation methods tend to widen distributions compared to explicit solvation, as well as to slightly increase the maximum of probability peak value. Differences between GB methods are also limited here.

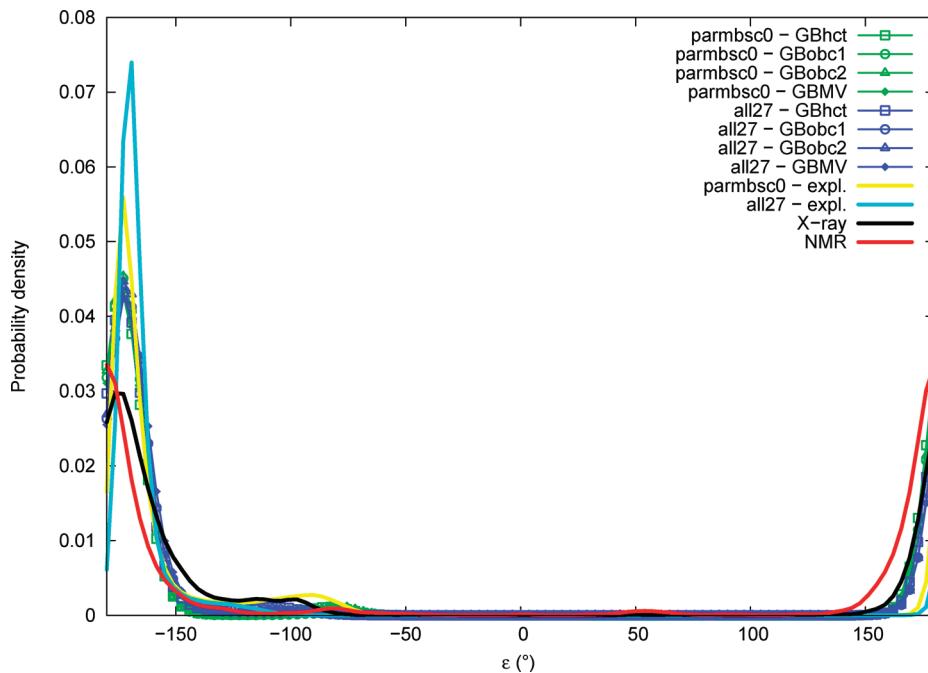
**Epsilon ( $\varepsilon$ ).** Distributions of the  $\varepsilon$  torsion angle are given in Figure 10. Experimental distributions of  $\varepsilon$  are Gaussians centered in the *trans* conformational region, with slightly higher values in X-ray data (average of  $-166.2^\circ$ , maximum of probability around  $-174^\circ$ ) than in NMR (average of  $-177.6^\circ$ , maximum of probability around  $\pm 180^\circ$ ). A small and broad secondary peak is visible in the X-ray distribution around  $-105^\circ$ . Explicit solvation simulations also give  $\varepsilon$  angle distributions with a major peak in the *trans* region, close but slightly above those of X-ray with both all27 (average of  $-166.6^\circ$ , maximum of probability about  $-170^\circ$ ) and parmbsc0 (average of  $-162.1^\circ$ , maximum of probability about  $-172^\circ$ ). A significant secondary peak can be observed in the parmbsc0 distribution around  $-92^\circ$ . Distributions obtained with GB implicit solvation methods are almost superimposable in the *trans* region with a maximum of probability around  $-172^\circ$ ; small secondary peaks can be observed around  $-110^\circ$  with all27 and around  $-80^\circ$  with parmbsc0, with intensities depending on the GB method.

**Zeta ( $\zeta$ ).** Distributions of the  $\zeta$  torsion angle are given in Figure 11. Experimental distributions of  $\zeta$  comprise a major peak located in the *gauche*– region, with values slightly lower in the X-ray data (average of  $-109.0^\circ$ , maximum of probability around  $-98^\circ$ ) compared to those of the NMR (average of  $-99.5^\circ$ , maximum of probability around  $-94^\circ$ ). Notable secondary peaks are present in the *trans* region for both X-ray (significant peak around  $-174^\circ$

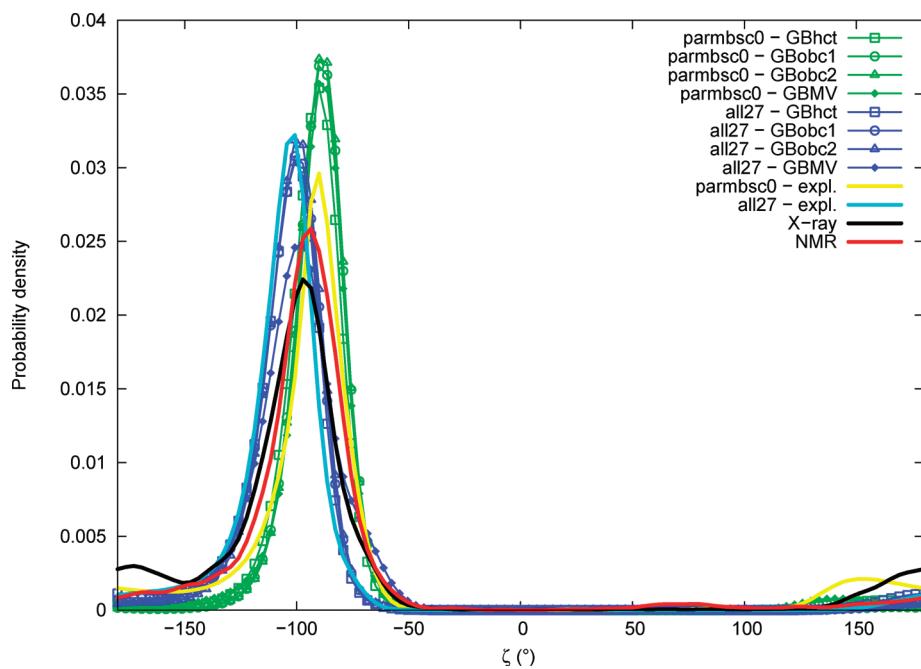
and NMR (small shoulder around  $-170^\circ$ ). Distributions of  $\zeta$  obtained from explicit solvation simulations also contain a major peak in the *gauche*– region, centered at values comparable to the experimental range but slightly lower with the all27 force field (average of  $-108.0^\circ$ , maximum of probability around  $-103^\circ$ ) and slightly superior with parmbsc0 (average of  $-102.3^\circ$ , maximum of probability around  $-91^\circ$ ). Secondary peaks in the *trans* conformational region can be observed with force fields in explicit solvation, in particular with parmbsc0 (significant peak at about  $152^\circ$ ), and also with all27 (small shoulder around  $\pm 180^\circ$ ). Implicit solvation  $\zeta$  angle distributions are in good agreement with explicit solvation results in the *gauche*– region. Small differences between GB methods are observed concerning the precise location and intensity of the secondary peak in the *trans* region.

**Chi ( $\chi$ ).** Distributions of the  $\chi$  torsion angle are given in Figure 12. Experimental distributions of  $\chi$  are mostly Gaussians centered in the *gauche*– region, with slightly lower values in NMR data (average of  $-114.4^\circ$ , maximum of probability around  $-114^\circ$ ) compared to those of the X-ray (average of  $-110.9^\circ$ , maximum of probability around  $-109^\circ$ ). A shoulder is visible in the X-ray distribution around  $-153^\circ$ . Explicit solvation simulations give a  $\chi$  angle in the experimental range for the all27 force field (average of  $-112.3^\circ$ , maximum of probability around  $-109^\circ$ ) and significantly underestimated with parmbsc0 (average of  $-117.2^\circ$ , maximum of probability around  $-119^\circ$ ). The use of GB implicit solvation methods with the parmbsc0 force field further accentuates the shifting of the  $\chi$  angle toward lower values by about  $5^\circ$  (reaching values around  $-125^\circ$ ). When GB methods are employed with the all27 force field, the  $\chi$  angle main peak stays close to explicit solvation results; however, with the GBMV method, an important secondary peak appears around  $-151^\circ$ .

**Sugar Puckering Parameters.** Sugar ring conformation is commonly described in terms of puckering or pseudorotation phase angle ( $P$ ) and amplitude ( $\tau_m$ ) parameters, calculated from the five ring dihedral angles.<sup>70</sup> The *endo/exo* notation is sometimes used to describe particular conformations where an atom is



**Figure 10.**  $\varepsilon$  torsion angle probability distribution. See Figure 1 legend.

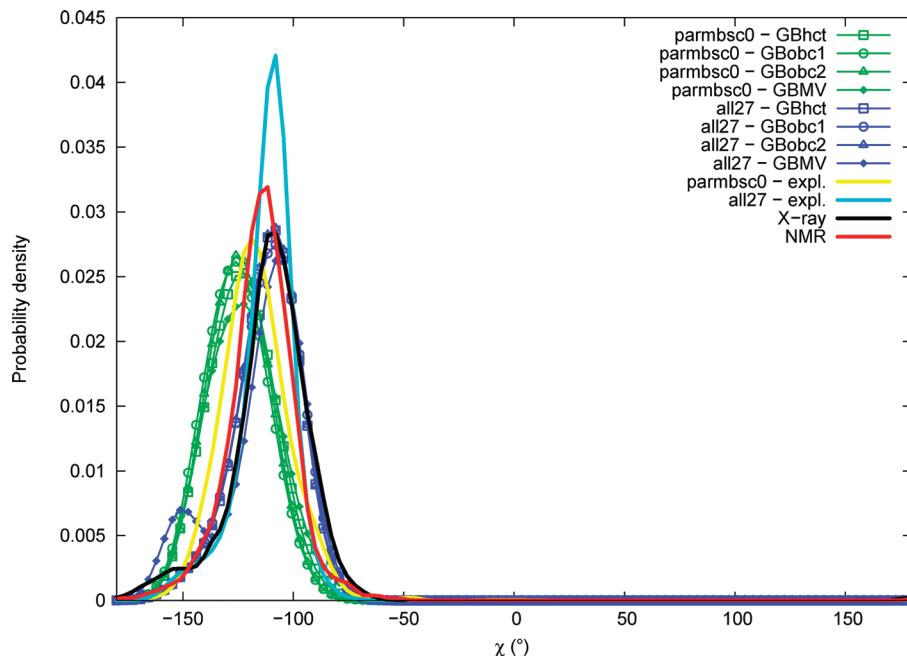


**Figure 11.**  $\zeta$  torsion angle probability distribution. See Figure 1 legend.

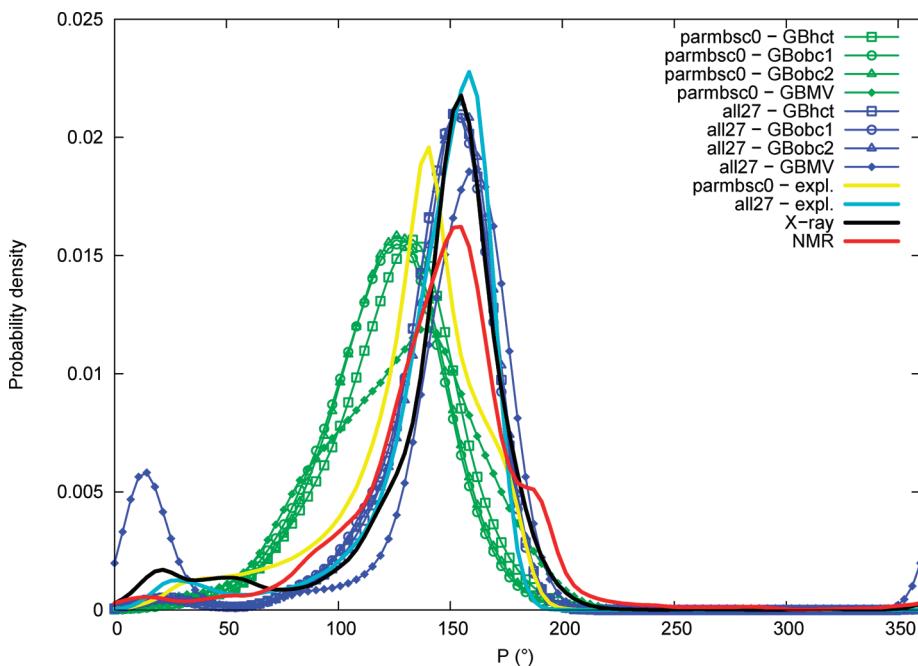
out of the plane formed by the others; these conformations correspond to odd multiples of  $18^\circ$  of the  $P$  angle. DNA in the B form is known to favor a C2'-endo ( $P = 162^\circ$ ) deoxyribose conformation, whereas in the A form, the C3'-endo ( $P = 18^\circ$ ) conformation is preferred. Another way of referring to conformations of the sugar ring is the *North* ( $P = 300\text{--}60^\circ$ ), *East* ( $P = 60\text{--}120^\circ$ ), and *South* ( $P = 120\text{--}220^\circ$ ) notation.<sup>71</sup>

**Puckering Phase.** Sugar puckering phase distributions are given in Figure 13. The sugar puckering phase is the same on

average in NMR and X-ray data ( $145.0^\circ$ ). Both experimental distributions present a main peak centered around  $154^\circ$ , corresponding to a C2'-endo conformation slightly deviated toward C1'-exo. Shoulders are visible around  $85^\circ$  (O4'-endo) and  $185^\circ$  (C2'-endo/C3'-exo) in the NMR data, and small secondary peaks can be seen in the X-ray distribution around  $20^\circ$  (C3'-endo) and  $50^\circ$  (C4'-exo). The parmbsc0 force field in explicit solvation has an average puckering phase markedly underestimated ( $133.2^\circ$ ), whereas the all27 force field is closer to experimental



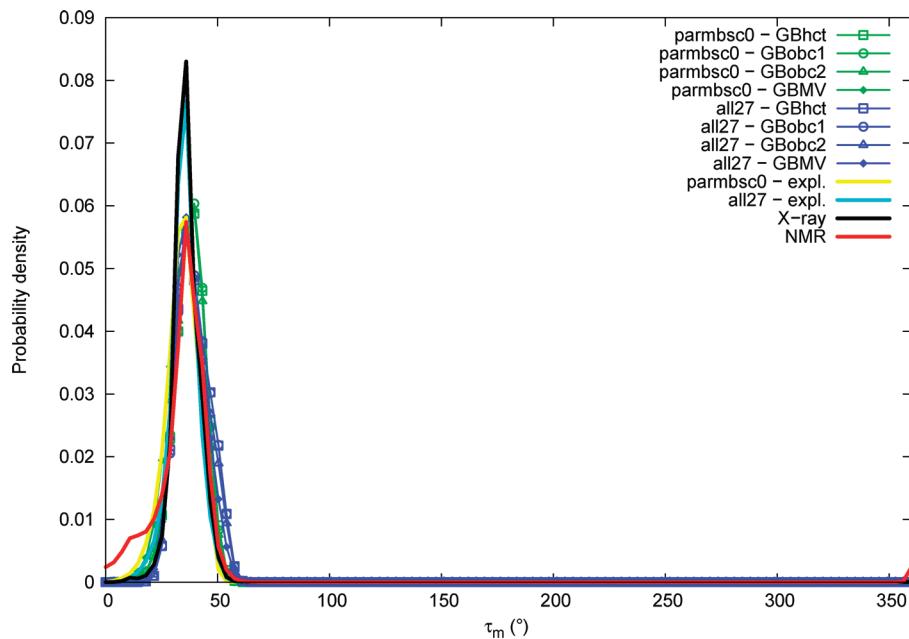
**Figure 12.**  $\chi$  torsion angle probability distribution. See Figure 1 legend.



**Figure 13.** Sugar puckering phase ( $P$ ) probability distribution. See Figure 1 legend.

data ( $143.9^\circ$ ). Distributions indicate that the main peak is indeed shifted on the left with parmbsc0 toward about  $140^\circ$ , indicating that the most populated conformation is closer to C $1'$ -exo than to C $2'$ -endo. The GB implicit solvation effect on the sugar puckering phase compared to that of explicit solvation is not straightforward. AMBER GB methods tend to lower the position of the main peak, in particular with the parmbsc0 force field, where the maximum of probability, already underestimated by parmbsc0 in itself, is further shifted by about  $10^\circ$  toward lower values (reaching about  $125^\circ$ ), corresponding to a C $1'$ -exo conformation. When using AMBER GB methods with the

all27 force field, the major peak is also shifted toward lower values but only by about  $5^\circ$  (at about  $151^\circ$ ). The CHARMM GBMV method effect is quite different between the two force fields. The main peak position is retained around  $140^\circ$  when using GBMV with parmbsc0 but widened by an important shoulder appearing on the left in the C $1'$ -exo region (around  $115^\circ$ ). GBMV employed with the all27 force field slightly shifts the main peak position toward higher values (around  $160^\circ$ ); in addition, a secondary peak around  $13^\circ$  becomes importantly populated. This indicates that transitions from the C $2'$ -endo to the C $3'$ -endo conformation (South to North) have occurred.



**Figure 14.** Sugar puckering amplitude ( $\tau_m$ ) probability distribution. See Figure 1 legend.

**Puckering Amplitude.** Sugar puckering amplitude distributions are given in Figure 14. Sugar puckering amplitude average values are on the same order between NMR ( $33.7^\circ$ ) and X-ray ( $36.0^\circ$ ). Both experimental distributions contain a major peak around  $35^\circ$ . The NMR distribution presents a shoulder on the left of the main peak, with a small peak around  $12^\circ$ . Both force fields in explicit solvation have puckering amplitudes in the experimental range (average values of  $33.9^\circ$  for parmbsc0 and  $35.1^\circ$  for all27). GB methods lead to a systematic but limited increase of the puckering amplitude average value, in a consistent order, namely, explicit < GBMV < GBOBC2 < GBOBC1  $\approx$  GBHCT. When looking at the distributions, it appears that when GB methods are used with the parmbsc0 force field, the main peak around  $35^\circ$  becomes uniformly shifted until about  $40^\circ$ , whereas GB methods with the all27 force field retain the main peak at  $35^\circ$  and populate other values around  $50^\circ$ .

## ■ DISCUSSION

Most of the published work on DNA force field evaluation was done with explicit solvation. In this work, we attempted to evaluate DNA molecular mechanics models when employed with GB implicit solvation, using explicit solvation results and experimental data as a reference. Four GB models implemented in the AMBER suite (GBHCT, GBOBC1, GBOBC2, and GBn), and one implemented in the CHARMM program (GBMV), were tested with two different force fields, AMBER parmbsc0 and CHARMM all27. We performed 5 ns molecular dynamics simulations of 12-bp-length duplexes containing 136 different tetranucleotides, with 10 combinations of force fields and GB methods. The possibility of employing the AMBER parmbsc0 or CHARMM all27 force field with GB implicit solvation methods implemented in either the AMBER or CHARMM program was a technical requirement in this work, allowing us to examine the respective influence of force field and GB method choice on DNA simulations. Structural parameters were extracted from simulations and statistically compared to reference data. To

prevent sequence composition bias between the different data set, a weighting was applied to ensure equal contribution of the 10 different dimer steps. Considerations on the data used as a reference, experimental and explicit solvation, are first presented. The respective validity and defects of force fields and GB models are then discussed.

**Experimental Data.** DNA experimental structural data are mostly known from NMR spectroscopy and X-ray crystallography. It is not obvious which technique is the most reliable experimental reference for the evaluation of our simulations, as they are very different approaches. In NMR experiments, the DNA sample is in solution, as in simulations, whereas in X-ray experiments, the sample is in the solid state, which can lead to structural deformations compared to those in the aqueous phase. On the other hand, NMR data refinement importantly relies on molecular dynamics, which could bring defects from simulation models and lead to circular validation. We therefore decided to present both NMR and X-ray data. Some differences can be observed in structural parameter statistics obtained from X-ray or NMR data, concerning the precise location of the maximum of probability peak, or the detailed shape of the distribution. These differences could be the consequence of insufficient data in particular at the dimer step level but remain limited at the global level. Most of the time, conclusions are qualitatively similar when taking NMR or X-ray as a reference for the evaluation of simulation results. In addition, differences between X-ray and NMR data provide an estimation of the uncertainty expected for structural parameters.

**Explicit Solvation Simulation Data.** Being able to compare implicit solvation results to explicit solvation data obtained with the same force field is essential in this work, in order to assess the quality of GB models. It was necessary to find DNA explicit solvation simulation data available for the two force fields studied (parmbsc0 and all27) and obtained under the same conditions. The work of Orozco and co-workers<sup>28</sup> met these requirements and was therefore chosen as the explicit solvation reference in our study. Although the recent work of the ABC consortium<sup>29</sup> would

have provided a more extensive data set, it is only available for the parmbsc0 force field. A comparison of structural parameter averages obtained with parmbsc0 in these two systematic studies shows close agreement for most parameters. Variations in averages calculated from the same raw data between our study and those published by Orozco and co-workers are small and arise from different technical choices. It was indeed necessary to follow the same protocol and use the same program in the analysis of structural parameters between our implicit solvation results and data chosen as a reference to ensure a fair comparison.

**Force Fields Evaluation.** We recall here the main force field strengths and defects identified by comparing the explicit solvation simulation results of Orozco and co-workers to experimental data. Many of these conclusions have indeed already been reached in recent systematic evaluation studies of DNA simulations in explicit solvation.<sup>27–29</sup> However, we believe that a summary based on our own comparison to up-to-date experimental data will prove useful here and clarify our subsequent evaluation of GB models.

The main force field defects concerning base pair step parameters are under-slide, light over-rise, light over-roll, and under-twist for parmbsc0 and over-slide, light over-rise, over-roll, and light over-twist for all27. Force field respective weaknesses are thus opposite for slide and twist and consistent for rise and roll. Mechanical couplings are known to exist between some dimer step parameters.<sup>72–74</sup> For example, slide and twist are usually positively correlated, which can explain why an under- (parmbsc0) or over- (all27) estimation of one of these parameters is reflected in the other. Slide and twist are also known to be negatively correlated with roll, which is in agreement with the behavior observed with parmbsc0 but not with all27, where roll is overestimated while slide and twist are also lightly overestimated.

Opposite force field effects are found for groove widths: parmbsc0 overestimates the major groove width, whereas all27 underestimates major groove width and overestimates minor groove width. Thus, parmbsc0 increases the differences between groove widths, while all27 decreases it. Groove widths are measured over several dinucleotides; therefore, a simple and direct coupling with local dimer step parameters is not expected. However, it has been shown that increased roll values correlate with wider minor grooves and narrower major grooves.<sup>75</sup> The groove width deformations observed with the all27 force field could thus be linked to its main defect in terms of step parameters, that is, roll overestimation. The opposite groove width deformations observed with the parmbsc0 force field are not explained by this coupling. More complex linear relationships between slide, roll, and twist on one hand and minor and major groove widths on the other hand have been proposed by El Hassan and Calladine.<sup>69</sup> Applying such relations indeed leads to a small overestimation of the major groove width with parmbsc0 contributed by its under-twist defect and an underestimation with all27 contributed by its over-roll defect.

Backbone  $\alpha/\beta/\gamma$  torsion parameters mostly stay in the  $g-/t/g+$  conformation with both parmbsc0 and all27 force fields in explicit solvation. This conformation is canonically observed in free B-DNA,<sup>71,76–78</sup> while noncanonical  $\alpha/\gamma$  conformers are more abundant in protein/DNA complexes.<sup>71,78</sup> Over-representation of unusual  $g+/t$  conformations of  $\alpha/\gamma$  angles was reported with AMBER parm94 and parm99 force fields,<sup>25,26,79</sup> prompting the parmbsc0 modification. Here, we find no significant  $g+/t$  conformers with all27 and less than 1% with parmbsc0. The  $\varepsilon$  and  $\zeta$  torsion angles are known to adopt two main conformations,

$t/g-$  ( $\varepsilon - \zeta \approx -90^\circ$ ) and  $g-/t$  ( $\varepsilon - \zeta \approx +90^\circ$ ), respectively designated BI and BII.<sup>80–82</sup> It has been shown that the BI/BII proportion is dependent on sequence and has an important influence on helical parameters.<sup>83,84</sup> The sequence averaged experimental ratio of BI/BII substates was reported to be around 80/20 for free B-DNA, from an analysis of crystal structures available in 2002,<sup>78,83</sup> as well as from NMR measurements.<sup>84</sup> The BI/BII ratio is more in favor of BI for protein-bound DNA (around 90/10).<sup>78</sup> The BI/BII ratio estimated from our own analysis of experimental data is 86/14 for X-ray and 94/6 for NMR. It has to be noted that our experimental data sets contain both pure and protein bound DNA; this can be at the origin of differences in BI/BII proportions. The BI/BII ratio calculated by us from Orozco and co-workers' explicit solvation simulation data is 86/14 for parmbsc0 and 94/6 for all27. BII is thus probably under-represented with all27 and closer to the experimental proportion with parmbsc0. In a study on the Jun-Fos oligomer,<sup>85</sup> Hartmann and co-workers concluded that the BII state is under-represented with both parmbsc0 and all27 force fields.

The glycosidic torsion angle  $\chi$  is known from earlier crystal structures surveys to adopt average values of  $-157^\circ$  for A-DNA and  $-108^\circ$  for B-DNA,<sup>86</sup> with a difference between B-DNA purines ( $\chi \approx -102^\circ$ ) and pyrimidines ( $\chi \approx -119^\circ$ ),<sup>76</sup> in conformity with our own analysis of B-DNA experimental data. The  $\chi$  angle modal peak is found in the correct experimental range with the all27 force field and underestimated with parmbsc0. Deficiencies in the description of the glycosidic angle were observed by Banáš et al.<sup>87</sup> in simulations of RNA tetraloops with various force fields (AMBER ff94, ff99, ff99bsc0, and CHARMM all27) in explicit solvation. Reparametrization of the  $\chi$  torsion of AMBER force fields was necessary to reproduce signature interactions in the tetraloops. However, it was noted that fine-tuning of the  $\chi$  angle for DNA and RNA at the same time might not be possible.

In terms of the sugar puckering phase, the major conformation observed in experimental B-DNA is close to South C2'-endo. The all27 force field provides a modal conformation in the correct range, whereas the parmbsc0 force field has an incorrect East leaning behavior, leading to a different major substate still South but closer to East C1'-exo. Hartmann and co-workers<sup>85</sup> indeed observed that sugars attached to pyrimidines oscillate between South and East with the parmbsc0 potential, whereas sugars attached to purines with parmbsc0, as well as all types of sugars with parm98 or all27, present the correct South conformation. Sugar puckering amplitude is found in the correct experimental range for both force fields. There are known correlations between the sugar puckering phase and other parameters, such as a positive coupling with twist<sup>74,88,89</sup> and with the  $\chi$  torsion in the *anti* range.<sup>90,91</sup> The under-twisting and  $\chi$  underestimation by parmbsc0 could thus be linked to its puckering East leaning behavior, and correcting one defect could cure the others.

**Generalized Born Models Evaluation.** We now compare our implicit solvation results to explicit solvation simulation data and present the most important defects of GB methods. We find that, although the generalized Born model is a much simpler treatment of solvation than explicit molecules, deviations observed from explicit solvation results are often on the same order as force field differences or experimental uncertainties. Some of the conclusions presented here for GBMV may have already been qualitatively obtained by Chocholoušová and Feig;<sup>55</sup> however, their study focused on the Drew–Dickerson dodecamer and was

not attempting to obtain converged results in terms of DNA sequence diversity.

Concerning step parameters, the GB implicit solvation method defects are significant under-slide and under-twist with AMBER pairwise GB implementations, observed with both force fields but more pronounced with parmbsc0. Light over-rise and over-roll are also observed. Among pairwise GB methods, the GBHCT method performs better than GBOBC1 and GBOBC2. It has to be noted that pairwise GB methods implemented in AMBER cause structural deformations in the same direction as the parmbsc0 force field, thereby further accentuating its own defects. GBMV, on the contrary, is close to explicit solvation values, except for roll, where light overestimation occurs.

The minor groove width is consistently widened with GBOBC1 and GBOBC2 and correctly modeled or slightly underestimated with GBHCT. GBMV has opposite effects for the two force fields, increasing the minor groove width with all27 and decreasing it with parmbsc0. In all cases, minor groove width deformations remain close to the range of force field and experimental uncertainties. On the other hand, the influence of GB methods on the major groove width is quite spectacular in terms of deformation amplitude. The major groove width is consistently and significantly overestimated with pairwise GB methods, in particular with the parmbsc0 force field, which already inherently widens it, leading to major groove widths as high as about 22 Å with the parmbsc0+pairwise GB combinations. GBMV has an opposite narrowing effect on the major groove width, particularly pronounced with the all27 force field which already underestimates it in itself, reaching a major groove width as low as about 15 Å with the all27+GBMV combination. The best current protocols leading groove widths in the correct explicit solvation range thus appear to be parmbsc0+GBMV or all27+pairwise GB.

Backbone  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\varepsilon$ , and  $\zeta$  torsion angles are overall well modeled with GB solvation methods. Distributions are almost superimposable between the different GB implementations, and maxima of probability stay close to the explicit solvation value. There is no clear performance distinction between GBMV and pairwise GB methods, as is the case with step parameters and groove widths. One can however notice a general tendency of GB methods to slightly increase  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\zeta$  angle average values and decrease  $\varepsilon$ . A widening of distributions can also be observed compared to explicit solvation, standard deviations becoming more comparable to experimental data, possibly reflecting improved conformational sampling of torsional angles around their canonical values with GB. The proportion of noncanonical  $\alpha/\gamma$  conformers is not increased by GB methods, staying under 0.5% in all cases. The BI/BII ratio obtained with GB solvation is more in favor of BI compared to explicit solvation (around 95/5). This could be caused by insufficient sampling of the minor  $\varepsilon/\zeta$  conformer due to limited simulation length.

GB methods present more serious defects in terms of sugar puckering angle and glycosidic torsion  $\chi$ . AMBER pairwise GB methods indeed tend to decrease the puckering and  $\chi$  angles main peak position when used with the parmbsc0 force field, which already underestimates the  $\chi$  torsion and has an intrinsic East leaning puckering defect. Modal angle values as low as 125° for puckering and -126° for  $\chi$  are reached with the parmbsc0 +GBOBC combinations. When AMBER GB methods are used with the all27 force field, the puckering angle underestimation is less pronounced, and the  $\chi$  angle is conformed to explicit solvation. The CHARMM GBMV method effect is quite different between the two force fields. The major peak position of

puckering and  $\chi$  angles is slightly overestimated with the all27 force field and underestimated with parmbsc0 with an important shoulder appearing in the East region of the puckering angle. In addition, an important secondary peak is observed with all27 in the puckering angle North region and around -151° for the  $\chi$  angle. Thus, the all27+GBMV combination appears to abnormally favor the puckering C3'-endo North conformation and a substate of  $\chi$  angle values which are characteristic of the A-DNA form. Sugar puckering amplitude with GB methods is close to the correct range.

An important and perhaps unexpected result of this work is that the GB method and force field choice influences on DNA simulation quality can be of the same magnitude. As a consequence, when FF and GB defects are opposite, they can fortuitously compensate, sometimes leading to even better agreement with experimental results than explicit solvation. For example, there is defect compensation for slide, twist, and major groove width when the all27 force field is used with pairwise GB methods. Another finding is that GB influence on DNA structure is in many cases relatively independent of the force field; that is, we could observe consistent effects of GB methods on DNA structural parameters when using different force fields. For example, pairwise GB methods consistently decrease slide and twist parameters and increase major groove width with both force fields. This facilitates the analysis, as GB and FF effects can to a certain extent be considered additive. An interest of this study is to provide a quantitative estimation of structural deformations expected when simulating DNA with GB solvation methods. Overall, we can conclude that realistic DNA modeling can be obtained with GB implicit solvation methods. In particular, the GBMV method implemented in CHARMM gives the best results, comparable to explicit solvation in most cases. Defects found in AMBER pairwise GB methods are generally less pronounced or not present with the GBMV method. It has however to be noted that the GBMV method is significantly more time-consuming than pairwise GB implementations, by a factor of about 7, although this includes inherent speed differences between AMBER and CHARMM. Among AMBER GB variants, GBHCT leads to better DNA modeling than GBOBC1, GBOBC2, or GBn. In light of our results, the best combinations of force field and GB variants that we can recommend for DNA simulations are parmbsc0+GBMV and all27+GBHCT; in the latter case, compensation of defects contributes to the success of the model. Ironically, this corresponds to the GB implementation of one program with the force field generally used with the other program, that is, the "AMBARMM" and "CHAMBER" approaches. Although all27+GBMV would come next, it features worrying underestimation of the major groove, overestimation of North puckering, and glycosidic angle values characteristics of the A-DNA form. The parmbsc0+pairwise GB protocol is not recommended for DNA simulations, as in this situation, force field and GB method defects add up for slide, twist, major groove width, puckering phase, and glycosidic torsion parameters, leading to important structural deformations.

## CONCLUSION

Implicit treatments of solvation like the generalized Born approach are not expected to be as detailed and realistic as the explicit simulation of solvent molecules, but they present many advantages like improved computational efficiency, enhanced conformational sampling, and the possibility of modeling large

systems and conformational changes. Implicit solvation modeling of DNA is believed to be more challenging than proteins because of the more complicated electrostatics environment. In this study, we performed a systematic evaluation of two DNA force fields with five different GB solvation models. We have shown that past decade developments have brought GB models into reasonable general agreement with explicit solvation and experimental data in terms of DNA structure and dynamics, and we identified particular defects of force fields and GB methods still present. This work provides guidance for the choice of force field and GB variant combinations in DNA simulations. The protocol developed here can be reused to evaluate force field or GB model modifications that could be suggested to improve DNA theoretical models.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: case@biomaps.rutgers.edu.

## ACKNOWLEDGMENT

The authors wish to thank Wilma K. Olson and Guohui Zheng for valuable discussions and help in using the 3DNA program and the 3DNA Landscapes platform. This work was supported by NIH grant GM57513.

## REFERENCES

- (1) Levitt, M. Computer Simulation of DNA Double-helix Dynamics. *Cold Spring Harb. Symp. Quant. Biol.* **1983**, *47*, 251–262.
- (2) Tidor, B.; Irikura, K. K.; Brooks, B. R.; Karplus, M. Dynamics of DNA oligomers. *J. Biomol. Struct. Dyn.* **1983**, *1*, 231–252.
- (3) Seibel, G. L.; Singh, U. C.; Kollman, P. A. A molecular dynamics simulation of double-helical B-DNA including counterions and water. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 6537–6540.
- (4) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (5) Cheatham, T. E., III; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. 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.* **1995**, *117*, 4193–4194.
- (6) York, D. M.; Yang, W.; Lee, H.; Darden, T.; Pedersen, L. G. Toward the Accurate Modeling of DNA: The Importance of Long-Range Electrostatics. *J. Am. Chem. Soc.* **1995**, *117*, 5001–5002.
- (7) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, S179–S197.
- (8) MacKerell, A. D., Jr.; Wiórkiewicz-Kuczera, J.; Karplus, M. An All-Atom Empirical Energy Function for the Simulation of Nucleic Acids. *J. Am. Chem. Soc.* **1995**, *117*, 11946–11975.
- (9) Langley, D. R. Molecular dynamic simulations of environment and sequence dependent DNA conformations: the development of the BMS nucleic acid force field and comparison with experimental results. *J. Biomol. Struct. Dyn.* **1998**, *16*, 487–509.
- (10) Cheatham, T. E., III; Cieplak, P.; Kollman, P. A. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845–862.
- (11) Foloppe, N.; MacKerell, A. D., Jr. All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* **2000**, *21*, 86–104.
- (12) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92*, 3817–3829.
- (13) Auffinger, P.; Westhoff, E. Simulations of the molecular dynamics of nucleic acids. *Curr. Opin. Struct. Biol.* **1998**, *8*, 227–236.
- (14) Beveridge, D. L.; McConnell, K. J. Nucleic acids: theory and computer simulation, Y2K. *Curr. Opin. Struct. Biol.* **2000**, *10*, 182–196.
- (15) Cheatham, T. E., III; Kollman, P. A. Molecular Dynamics Simulation of Nucleic Acids. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435–471.
- (16) Cheatham, T. E., III; Young, M. A. Molecular Dynamics Simulation of Nucleic Acids: Successes, Limitations, and Promise. *Biopolymers* **2001**, *S6*, 232–256.
- (17) Giudice, E.; Lavery, R. Simulations of Nucleic Acids and Their Complexes. *Acc. Chem. Res.* **2002**, *35*, 350–357.
- (18) Norberg, J.; Nilsson, L. Molecular Dynamics Applied to Nucleic Acids. *Acc. Chem. Res.* **2002**, *35*, 465–472.
- (19) Orozco, M.; Pérez, A.; Noya, A.; Javier Luque, F. Theoretical methods for the simulation of nucleic acids. *Chem. Soc. Rev.* **2003**, *32*, 350–364.
- (20) Cheatham, T. E., III Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr. Opin. Struct. Biol.* **2004**, *14*, 360–367.
- (21) Orozco, M.; Noy, A.; Pérez, A. Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr. Opin. Struct. Biol.* **2008**, *18*, 185–193.
- (22) Feig, M.; Pettitt, B. M. Structural Equilibrium of DNA Represented with Different Force Fields. *Biophys. J.* **1998**, *75*, 134–149.
- (23) MacKerell, A. D., Jr.; Banavali, N. K. All-Atom Empirical Force Field for Nucleic Acids: II. Application to Molecular Dynamics Simulations of DNA and RNA in Solution. *J. Comput. Chem.* **2000**, *21*, 105–120.
- (24) Reddy, S. Y.; Leclerc, F.; Karplus, M. DNA Polymorphism: A Comparison of Force Fields for Nucleic Acids. *Biophys. J.* **2003**, *84*, 1421–1449.
- (25) Beveridge, D. L.; Barreiro, G.; Byun, K. S.; Case, D. A.; Cheatham, T. E., III; Dixit, S. B.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Seibert, E.; Sklenar, H.; Stoll, G.; Thayer, K. M.; et al. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys. J.* **2004**, *87*, 3799–3813.
- (26) Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham, T. E., III; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H.; Thayer, K. M.; Varnai, P. 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.* **2005**, *89*, 3721–3740.
- (27) Fujii, S.; Kono, H.; Takenaka, S.; Go, N.; Sarai, A. Sequence-dependent DNA deformability studied using molecular dynamics simulations. *Nucleic Acids Res.* **2007**, *35*, 6063–6074.
- (28) Perez, A.; Lankas, F.; Luque, F. J.; Orozco, M. Towards a molecular dynamics consensus view of B-DNA flexibility. *Nucleic Acids Res.* **2008**, *36*, 2379–2394.
- (29) Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T., III; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; et al. 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.* **2010**, *38*, 299–313.
- (30) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (31) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **1995**, *246*, 122–129.
- (32) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *J. Phys. Chem.* **1996**, *100*, 19824–19839.

- (33) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.
- (34) Jayaram, B.; Liu, Y.; Beveridge, D. L. A modification of the generalized Born theory for improved estimates of solvation energies and pK shifts. *J. Chem. Phys.* **1998**, *109*, 1465–1471.
- (35) Dominy, B. N.; Brooks, C. L., III Development of a Generalized Born Model Parametrization for Proteins and Nucleic Acids. *J. Phys. Chem. B* **1999**, *103*, 3765–3773.
- (36) Tsui, V.; Case, D. A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* **2001**, *56*, 275–291.
- (37) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.
- (38) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55*, 383–394.
- (39) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *J. Chem. Theory Comput.* **2007**, *3*, 156–169.
- (40) Schaefer, M.; Karplus, M. A Comprehensive Analytical Treatment of Continuum Electrostatics. *J. Phys. Chem.* **1996**, *100*, 1578–1599.
- (41) Schaefer, M.; Bartels, C.; Leclerc, F.; Karplus, M. Effective Atom Volumes for Implicit Solvent Models: Comparison between Voronoi Volumes and Minimum Fluctuation Volumes. *J. Comput. Chem.* **2001**, *22*, 1857–1879.
- (42) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III Novel generalized Born methods. *J. Chem. Phys.* **2002**, *116*, 10606–10614.
- (43) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* **2003**, *24*, 1348–1356.
- (44) Im, W.; Lee, M. S.; Brooks, C. L., III Generalized Born Model with a Simple Smoothing Function. *J. Comput. Chem.* **2003**, *24*, 1691–1702.
- (45) Onufriev, A.; Case, D. A.; Bashford, D. Effective Born radii in the generalized Born approximation: the importance of being perfect. *J. Comput. Chem.* **2002**, *23*, 1297–1304.
- (46) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., III Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **2004**, *25*, 265–284.
- (47) Fan, H.; Mark, A. E.; Zhu, J.; Honig, B. Comparative study of generalized Born models: Protein dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6760–6764.
- (48) Bashford, D.; Case, D. A. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152.
- (49) Simonson, T. Macromolecular electrostatics: continuum models and their growing pains. *Curr. Opin. Struct. Biol.* **2001**, *11*, 243–252.
- (50) Feig, M.; Brooks, C. L., III Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217–224.
- (51) Chen, J.; Brooks, C. L., III; Khandogin, J. Recent advances in implicit solvent-based methods for biomolecular simulations. *Curr. Opin. Struct. Biol.* **2008**, *18*, 140–148.
- (52) Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.
- (53) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* **1999**, *101*, 426–434.
- (54) Tsui, V.; Case, D. A. Molecular Dynamics Simulations of Nucleic Acids with a Generalized Born Solvation Model. *J. Am. Chem. Soc.* **2000**, *122*, 2489–2498.
- (55) Chocholoušová, J.; Feig, M. Implicit Solvent Simulations of DNA and DNA-Protein Complexes: Agreement with Explicit Solvent vs Experiment. *J. Phys. Chem. B* **2006**, *110*, 17240–17251.
- (56) Ruscio, J. Z.; Onufriev, A. A Computational Study of Nucleosomal DNA Flexibility. *Biophys. J.* **2006**, *91*, 4121–4132.
- (57) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (58) Arnott, S.; Smith, P. J. C.; Chandrasekaran, R. In *Handbook of Biochemistry and Molecular Biology*, 3rd ed.; Fasman, G. D., Ed.; CRC Press: Cleveland, OH, 1976; Nucleic Acids, Vol. II, pp 411–422.
- (59) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (60) Crowley, M. F.; Williamson, M. J.; Walker, R. C. CHAMBER: Comprehensive Support for CHARMM Force Fields Within the AMBER Software. *Int. J. Quantum Chem.* **2009**, *109*, 3767–3772.
- (61) Chocholoušová, J.; Feig, M. Balancing an Accurate Representation of the Molecular Surface in Generalized Born Formalisms with Integrator Stability in Molecular Dynamics Simulations. *J. Comput. Chem.* **2006**, *27*, 719–729.
- (62) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441–451.
- (63) Dickerson, R. E. Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res.* **1989**, *17*, 1797–1803.
- (64) Lu, X. J.; Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **2003**, *31*, 5108–5121.
- (65) Lu, X. J.; Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* **2008**, *3*, 1213–1227.
- (66) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (67) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S. H.; Srinivasan, A. R.; Schneider, B. The Nucleic Acid Database: A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **1992**, *63*, 751–759.
- (68) Zheng, G.; Colasanti, A. V.; Lu, X. J.; Olson, W. K. 3DNA Landscapes: a database for exploring the conformational features of DNA. *Nucleic Acids Res.* **2010**, *38*, D267–D274.
- (69) El Hassan, M. A.; Calladine, C. R. Two Distinct Modes of Protein-induced Bending in DNA. *J. Phys. Chem. B* **1998**, *282*, 331–343.
- (70) Altona, C.; Sundaralingam, M. Conformational Analysis of the Sugar Ring in Nucleosides and Nucleotides. A New Description Using the Concept of Pseudorotation. *J. Am. Chem. Soc.* **1972**, *94*, 8205–8212.
- (71) Várnai, P.; Djuranovic, D.; Lavery, R.; Hartmann, B. Alpha/gamma Transitions in the B-DNA backbone. *Nucleic Acids Res.* **2002**, *30*, 5398–5406.
- (72) Gorin, A. A.; Zhurkin, V. B.; Olson, W. K. B-DNA Twisting Correlates with Base-pair Morphology. *J. Mol. Biol.* **1995**, *247*, 34–48.
- (73) Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. M.; Zhurkin, V. B. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163–11168.
- (74) Olson, W. K.; Zhurkin, V. B. Modeling DNA deformations. *Curr. Opin. Struct. Biol.* **2000**, *10*, 286–297.
- (75) Suzuki, M.; Yagi, N. An In-the-Groove View of DNA Structures in Complexes with Proteins. *J. Mol. Biol.* **1996**, *255*, 677–687.
- (76) Schneider, B.; Neidle, S.; Berman, H. M. Conformations of the Sugar-Phosphate Backbone in Helical DNA Crystal Structures. *Biopolymers* **1997**, *42*, 113–124.
- (77) Beckers, M. L. M.; Buydens, L. M. C. Multivariate Analysis of a Data Matrix Containing A-DNA and B-DNA Dinucleoside Monophosphate Steps: Multidimensional Ramachandran Plots for Nucleic Acids. *J. Comput. Chem.* **1998**, *19*, 695–715.
- (78) Djuranovic, D.; Hartmann, B. Conformational characteristics and correlations in crystal structures of nucleic acid oligonucleotides: evidence for sub-states. *J. Biomol. Struct. Dyn.* **2003**, *20*, 771–788.

- (79) Várnai, P.; Zakrzewska, K. DNA and its counterions: a molecular dynamics study. *Nucleic Acids Res.* **2004**, *32*, 4269–4280.
- (80) Gupta, G.; Bansal, M.; Sasisekharan, V. Reversible Bending and Helix Geometry in a B-DNA Dodecamer: CGCGAATT<sup>B<sub>r</sub></sup>CGCG\*. *J. Biol. Chem.* **1982**, *257*, 14686–14707.
- (81) Privé, G. G.; Heinemann, U.; Chandrasegaran, S.; Kan, L.-S.; Kopka, M.; Dickerson, R. E. Helix Geometry, Hydration, and G·A Mismatch in a B-DNA Decamer. *Science* **1987**, *238*, 498–504.
- (82) Hartmann, B.; Piazzola, D.; Lavery, R. BI-BII transitions in B-DNA. *Nucleic Acids Res.* **1993**, *21*, 561–568.
- (83) Djuranovic, D.; Hartmann, B. DNA fine structure and dynamics in crystals and in solution: the impact of BI/BII backbone conformations. *Biopolymers* **2004**, *73*, 356–368.
- (84) Heddi, B.; Foloppe, N.; Bouchemal, N.; Hantz, E.; Hartmann, B. Quantification of DNA BI/BII backbone states in solution. Implications for DNA overall structure and recognition. *J. Am. Chem. Soc.* **2006**, *128*, 9170–9177.
- (85) Heddi, B.; Foloppe, N.; Oguey, C.; Hartmann, B. Importance of Accurate DNA Structures in Solution: The Jun-Fos Model. *J. Mol. Biol.* **2008**, *382*, 956–970.
- (86) Lu, X. J.; Shakked, Z.; Olson, W. K. A-form Conformational Motifs in Ligand-bound DNA Structures. *J. Mol. Biol.* **2000**, *300*, 819–840.
- (87) Banáš, P.; Hollas, D.; Zgarbová, M.; Jurečka, P.; Orozco, M.; Cheatham, T. E., III; Šponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* **2010**, *6*, 3836–3849.
- (88) Zhurkin, V. B.; Lysov, Y. P.; Florentiev, V. L.; Ivanov, V. I. Torsional flexibility of B-DNA as revealed by conformational analysis. *Nucleic Acids Res.* **1982**, *10*, 1811–1830.
- (89) Kollman, P.; Keepers, J. W.; Weiner, P. Molecular-mechanics studies on d(CGCGAATTCTCGCG)<sub>2</sub> and dA<sub>12</sub>·dT<sub>12</sub>: An illustration of the coupling between sugar repuckering and DNA twisting. *Biopolymers* **1982**, *21*, 2345–2376.
- (90) Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. V.; Kopka, M. L. The Anatomy of A-, B-, and Z-DNA. *Science* **1982**, *216*, 475–485.
- (91) Gelbin, A.; Schneider, B.; Clowney, L.; Hsieh, S. H.; Olson, W. K.; Berman, H. M. Geometric Parameters in Nucleic Acids: Sugar and Phosphate Constituents. *J. Am. Chem. Soc.* **1996**, *118*, 519–529.