Toward a Full Characterization of Nucleic Acid Components in Aqueous Solution: Simulations of Nucleosides

Nicolas Foloppe[†] and Lennart Nilsson*

Center for Structural Biochemistry, Department of Biosciences, Karolinska Institutet, S-141 57 Huddinge, Sweden

Received: December 2, 2004; In Final Form: February 16, 2005

The eight nucleoside constituents of nucleic acids were simulated for 50 ns in explicit water with molecular dynamics. This provides equilibrium populations of the torsional degrees of freedom, their kinetics of interconversion, their couplings, and how they are influenced by water. This is important, given that a full and quantitative characterization of the nucleosides in aqueous solution by experimental means has been elusive, despite immense efforts in that direction. It is with the anti/syn equilibrium that the simulations are most complementary to experiment, by accessing directly the influence of the sugar type, sugar pucker, and base on the anti/syn populations. The glycosidic torsion distributions in the anti conformation are strongly affected by water and depart from the corresponding X-ray modal values and the associated energy minima in vacuo. Water also preferentially stabilizes some sugar conformations, showing that potential energies in vacuo are not sufficient to understand the nucleosides. Deoxythymidine (but not other pyrimidines) significantly populates the syn orientation. Guanine favors the syn orientation more than adenine. The ribose favors the syn orientation significantly more than the deoxyribose. The NORTH pucker coexists with the syn conformers. A hydrogen bond is frequently formed between the 5′-OH group and the syn bases, despite competition by water. The rate of the anti/syn transitions with purines is on the nanosecond time scale, confirming a long held assumption underpinning the interpretation of ultrasonic relaxation studies. Therefore, our knowledge of the structure and dynamics of nucleosides in solvent is only limited by the accuracy of the potential used to simulate them, and it is shown that such simulations provide a distinct and unique test of nucleic acid force fields. This confirmed that the widely distributed CHARMM27 force field is, overall, well-balanced with a particularly good representation of the ribose. Specific improvements, however, are suggested for the deoxyribose and torsion γ.

I. Introduction

In recent years, major progress has been achieved regarding the development, testing, and validation of nucleic acid force fields. This led, among others, to the BMS,[1] the modified Cornell et al. AMBER,[2,3] and the CHARMM27[4] nucleic acid force fields. A growing number of studies suggest that this new generation of force fields has led to major improvements in modeling and simulation of nucleic acid properties in explicit solvent.[5−13] These new tools have been applied primarily to oligonucleotides, with little attention for nucleosides or nucleotides. This may reflect the superficial perception that these nucleic acid components are already well-characterized, especially the smaller nucleosides.

Certainly, immense experimental efforts have been devoted to study the standard nucleosides by crystallography,[14−18] NMR,[19−31] circular dichroism,[32] ultrasonic relaxation,[33,34] and vibrational spectroscopies.[35] In addition, quantum mechanical calculations give access to the potential energies of nucleosides in vacuo.[36−43] These efforts have yielded invaluable information, in particular the precise geometries of the nucleosides in crystals. A careful analysis of the literature, however, reveals that a complete and quantitative characterization of nucleoside conformations in solution has been elusive (see Experimental Reference Data subsection, below). This is largely because nucleosides are quite complex molecules with several degrees of freedom. The fast exchange between various conformers can confound the interpretation of experimental data, which reflect ensemble and time averages. Also, many experimental reports investigated only a few nucleosides and did not systematically include the eight standard nucleosides. That has hindered the emergence of a clear, consistent, and detailed picture of the influence of the sugar, individual bases, and solvent on the conformational properties beyond the usual categorization in terms of purines and pyrimidines. For these reasons, experimental studies of nucleosides and nucleotides in solution are still an active area of research.[27,29,44−46]

In the present work, a modern nucleic acid force field, CHARMM27,[4] is used to simulate each of the eight standard nucleosides in water on a time scale of at least 50 ns. Simulations using the modified Cornell et al. AMBER[2,3] are also presented for comparison. The aim is twofold. First, we use these simulations as an additional and distinct test of the force field, by comparing experimentally derived properties to their simulated counterparts. The performance of the CHARMM27 force field is discussed in the context of the underlying potential energies, some of which are obtained as part of the present work. A summary of the reference experimental properties is provided, which should facilitate their

* Author to whom correspondence should be addressed. Phone: 46-8-608-9228. Fax:46-8-608-9290. E-mail: Lennart.Nilsson@biosci.ki.se

[†] Present address: Vernalis (R&D) Ltd., Granta Park, Abington, Cambridge CB1 6GB, U. K.

10.1021/jp044513u CCC: $30.25
Published on Web 04/05/2005

general use for testing other force fields. This summary also delineates the limitations in these experimental data, pointing to the areas where the simulations can best supplement them. This leads to the second and main aim of this work, which is to revisit the conformational properties of nucleosides in solution.

Nucleoside simulations to evaluate a nucleic acid force field have already been touched upon,[3] but we systematically broaden the scope and elaborate further the aims of this test. The following explains why this test is unique and distinct. The simplest test of a nucleic acid force field uses model compounds in vacuo,[47,48] to reproduce the geometries and torsional energy surfaces, using X-ray crystallography and quantum mechanical data as a guide.[3,4,49] This in vacuo calibration with model compounds, however, suffers from (i) the fact that it does not treat explicitly the influence of the solvent, (ii) the dependence of the ab initio data on the level of theory at which they are obtained, and (iii) the limited way in which the coupling between degrees of freedom can be treated. This is why the force field initially obtained for these model compounds is frequently adjusted empirically, typically to reproduce condensed phase data pertaining to oligonucleotides. These data, however, offer only a partial window on nucleic acid properties. The pitfalls involved with testing force fields via simulations of oligonucleotides in condensed phase have already been illustrated and discussed in detail.[48] To summarize, there are two main issues with this approach. First, we do not yet have a precise enough experimental picture of the structure and dynamics of a diverse enough set of nucleic acid sequences in solution. Second, the test simulations are probably not converged enough to provide a definitive comparison between simulated and measured properties of oligonucleotides, in particular because the equilibrium between major substates or even structural families may not yet be adequately reproduced by computation. The present work shows that a key reason why simulations of nucleosides provide a unique test is that the much smaller size of the nucleosides allows the derivation of true equilibrium populations for these systems, as shown by the frequent interconversion between the various conformers.

By giving direct access to the kinetics of interconversions between the conformational states, their exact populations, their structural spread, the coupling between degrees of freedom, and the influence of water, the present simulations complement years of experimental efforts. This is particularly helpful with respect to the anti/syn equilibrium and the influence of water, which has proven difficult to tackle experimentally.[23] Validating simulations with the standard nucleosides should raise awareness regarding the application of this technique to the many modified nucleosides synthesized for therapeutic purposes[50] and other small molecules of biochemical interest.

II. Methods

Preparation of Initial Nucleoside Structures for Molecular Dynamics Simulations. All coordinate manipulations, energy minimizations, and molecular dynamics (MD) simulations were performed with the program CHARMM,[51] a dielectric constant of 1.0, and atom-based nonbonded interactions truncated beyond 12 Å with a force shift,[52] which is reliable for nucleic acids.[53] Nonbonded lists were maintained to 14 Å and updated whenever an atom had moved >1 Å since the last update. Calculations were performed either with the CHARMM27 nucleic acid force field[4] or with the modified Cornell et al. AMBER [2,3] force field, as ported to the CHARMM program (T. Cheatham, unpublished work). Graphical operations were performed with the program VMD.[54]

The four standard deoxyribonucleosides (dA, dC, dG, dT) and the four standard ribonucleosides (rA, rC, rG, rU) were initially built and energy-minimized in vacuo with the dihedral angles β, γ, ϵ, and χ in trans, g^+, trans, and anti, respectively. The initial furanose conformation was C2′ endo for deoxynucleosides and C3′ endo for ribonucleosides.

Molecular Dynamics Protocols. Each nucleoside was overlaid with a $25 \times 25 \times 25$ Å^3 box of preequilibrated CHARMM TIP3P water.[55,56] The water molecules that had their oxygen atom closer than 2.7 Å from any nucleoside atom were removed. Periodic boundary conditions were applied, and all covalent bonds involving a hydrogen were constrained with SHAKE.[57] The solvent was then energy-minimized, keeping the nucleoside fixed. Both the solvent and the nucleoside were allowed to relax in a second round of energy minimization, which was carried out to an energy gradient $\leq 10^{-3}$ kcal/(mol Å). Each system was then submitted to MD simulations, using the leapfrog integrator and a 0.002 ps time step. Heating was performed from 0 to 300 K in 6 ps by 5 K increments, with every nucleoside atom harmonically constrained to its initial position with a force constant of 2.0 kcal/(mol Å^2). The simulation was then pursued at 300 K in the *NVT* ensemble for 40 ps, with the constraints on the nucleoside kept only during the first 20 ps. This was followed by 20 ps of MD in the *NPT* ensemble, to monitor possible volume changes. The primary box volume adjusted only very slightly upon simulation at constant pressure, ensuring that the number of water molecules present was reasonable. The average volume during the last 10 ps of the *NPT* simulation was kept for the remainder of the MD simulations, which were continued in the *NVT* ensemble. Coordinates were saved for analysis every picosecond. The first nanosecond of the simulations was considered equilibration and excluded from the analysis. Each nanosecond of simulation took 5 h of elapsed time to perform in parallel on four 3.06 GHz Intel Xeon CPUs in a GNU/Linux cluster using a gigabit ethernet interconnect.

Structural Analysis. The atom names and dihedral angle nomenclature are as in Saenger.[58] Sugar puckering pseudorotation angles (P) and amplitude (τ) have been determined following Altona and Sundaralingam [15] using the same reference state for $P = 0.0°$. Two types of analyses were performed for the sugar pseudorotation angle. The first uses a four-state model in which the pseudorotation space was divided into four equally sized quadrants centered around $P = 0.0°$, $P = 90.0°$, $P = 180.0°$, and $P = 270.0°$ that are referred to as the north, east, south, and west quadrants, respectively. To facilitate comparison with the NMR studies, a two-state model is also used, in which the pseudorotation space is partitioned into NORTH ($270° \leq P < 90.0°$) and SOUTH ($90° \leq P < 270.0°$) only. Lowercase and uppercase are used when referring to the four-state and two-state models, respectively.

Given the values adopted by χ in crystal structures of nucleosides[17] and the corresponding potential energy surfaces,[40] it is meaningful to define the anti conformation of χ as the 170−300° range. Likewise, the syn conformation is defined as the 30−90° range. Values of χ falling outside the anti or syn ranges are referred to as other.

Torsion γ was measured with respect to O5′−C5′−C4′−C3′ and categorized as g^+ ($60 \pm 30°$), g^- ($-60 \pm 30°$), and trans ($180 \pm 30°$). Values falling out of these ranges are referred to as other.

A hydrogen bond was considered formed between O5′−H5′ and N3 (purines) or O2 (pyrimidines) of the base if (i) the O5′···N3 (O5′···O2) distance was ≤3.5 Å and (ii) the O5′−H5′···N3 (O5′−H5′···O2) angle was ≥120.0°. The same geometric

TABLE 1: Selected Potential Energies (kcal/mol) Relevant to the Puckering Energetics of the Furanose in Deoxyribonucleosides

	ΔE_{n-s}[a]		B[b]	
	ab initio[c]	CHARMM27[d]	ab initio[c]	CHARMM27[d]
dA	0.4	0.6	4.2	2.6
dC	−0.3	−0.2	4.0	1.9
dG	0.7	1.0	4.3	2.9
dT	0.9	0.2	4.0	2.2

[a] Energy of the north minimum minus that of the south minimum with χ = anti and $\gamma = g^+$ (the south minimum is more stable for all nucleosides except dC). [b] East energy barrier, obtained as the energy in the O4′ endo conformation minus that of the global energy minimum (north or south). [c] Energies obtained at the MP2/6-31G* level.[39] [d] CHARMM27 energies from ref 4.

criteria were used to determine if a hydrogen bond was formed between O2′ and O3′ in the ribose, with each oxygen being alternatively considered as the hydrogen bond donor or acceptor.

Potential Energies in Vacuo. All quantum mechanical potential energies in vacuo are from previous reports.[39,40] When indicated, their CHARMM27 counterpart was obtained as part of the present work, after energy minimization in vacuo to an energy gradient $\leq 10^{-6}$ kcal/(mol Å). The potential energies are reported relative to the global energy minimum, which is for a south (north for dC) furanose, χ = anti, and $\gamma = g^+$.

Interaction with Water. Total interaction energies between the nucleoside and water were calculated every picosecond using the same cutoffs as during the simulations. This was used to obtain average interaction energies.

III. Results and Discussion

This work is primarily concerned with an application and analysis of the CHARMM27 force field. For comparison, however, the modified Cornell et al. AMBER[2,3] force field was also used to simulate dA, rA, and dC for 50 ns each. For simplicity, the AMBER results are only discussed in a final separate section, after presentation of the CHARMM27 results.

The following is not concerned with bond lengths and valence angles, because it has already been shown that the CHARMM27 force field treats these hard degrees of freedom satisfactorily.[4] Instead, we concentrate on the softer torsional degrees of freedom because they govern the overall structure, and they can be probed experimentally in aqueous solution. These degrees of freedom are the pseudorotational angle P, the glycosidic torsion χ, and the torsions β and γ. To characterize these torsions very thoroughly, the simulations were pursued for at least 50 ns for every nucleoside. The CHARMM27 results obtained for the glycosidic torsion with pyrimidines demonstrate that even with small organic molecules such long simulations are needed to fully characterize the conformational space.

We first provide a summary, which cannot be a comprehensive review, of the main conclusions from the large body of published experimental work on the conformational properties of the nucleosides, to facilitate comparison with the simulations. To put the simulation results in the context of the underlying potential energies, we report the previously obtained relative energies for the relevant conformations of the deoxyribonucleosides (Table 1), obtained in vacuo using ab initio calculations or the CHARMM27 force field [3,4,39]. We refrain from presenting potential energies for the ribonucleosides, because they are very sensitive to the relative orientations of the 2′-OH and 3′-OH groups, which would unnecessarily clutter the discussion.

Experimental Reference Data. Unless otherwise noted, we refer only to data obtained with nucleosides and not nucleotides.

The results from the most recent and comprehensive survey of X-ray data of nucleic acid monomers[17] did not separate nucleosides from nucleotides but included significantly more nucleosides than nucleotides. Of the 36 deoxyribonucleosides/tides in this survey, 29 (∼80%) and 7 (∼20%) are in the south and north conformations, respectively, whereas 49 (∼67%) of the ribose compounds are in the south conformation and 24 (∼33%) in the north. Most of these structures had χ in anti, with χ in syn for only 23 (∼21%) of them. Of these 23, 20 were with purines, confirming previous analyses that had suggested that χ in syn is particularly unfavorable for pyrimidines.[14,20] A large majority (∼73%) of the X-ray structures had γ in g^+, and only the two other staggered conformers were also populated, with the trans orientation (∼17%) more populated than g^- (∼9%).

Analysis of the NMR data is less straightforward than their X-ray counterpart, in part because one has to elucidate the coupling between various degrees of freedom, such as P and χ. In addition, the interpretation of NMR data tends to be guided by predetermined structural models, such as a two-state (SOUTH/NORTH) model for the furanose conformation.[23,27,29,30] Only NMR work in aqueous solution is considered here. The information obtained by NMR was reviewed in great detail by Davies,[23] and we do not reexamine individual studies presented prior to this review. In terms of pseudorotation angle populations, this review concluded that purine ribonucleosides (χ in anti) prefer the SOUTH conformer (∼60%) over the NORTH (∼40%). The reverse situation (SOUTH, 40%; NORTH, 60%) was found with pyrimidine ribonucleosides (χ in anti). With deoxyribonucleosides, purines were also found to favor the SOUTH (∼70%) more than pyrimidines (∼60%), but both base categories populated the SOUTH more than the NORTH.

These results are consistent with those obtained later, also by NMR, when investigating specifically the influence of 2′-substituents on the SOUTH/NORTH equilibrium of the sugar in nucleosides with uracil[24] and adenine.[25] In the natural ribouridine and deoxyribouridine, the NORTH populations were found to be 58% and 40%, respectively, the remaining being attributed to the SOUTH state.[24] With an adenine, the NORTH populations were found to be smaller, 36% in riboadenosine and 19% in deoxyriboadenosine.[25] Although the SOUTH/NORTH equilibrium was found to be sensitive to the sugar substituents, a significant population of either state was observed in each derivative. An interpretation of the furanose NMR coupling constants in terms of a three-state model (north/east/south) is possible[31] and shows that the east is not significantly populated for the ribose. With the deoxyribose, the best fit to the experimental data is also obtained with a negligible east population, although an alternative model with up to 11% east pucker cannot be ruled out. The same study confirmed that the south is more prevalent with the deoxyribose than the ribose. Since then, the interest in using NMR to dissect the factors that influence the pseudorotational equilibrium in nucleosides has remained strong and has led to other series of measurements of the SOUTH and NORTH populations.[26−30] ^{13}C-enriched ribonucleosides[30] and deoxyribonucleosides[29] led to more complete measurements, which excluded the presence of the west conformation.[30] The proportions of NORTH and SOUTH derived from the work of Plavec et al.[27] are similar to those from Bandyopadhyay et al.[29] and are reported in Table 2 to facilitate comparison with the MD results. These more recent NMR studies confirm, to a large extent, previous conclusions regarding the relative pseudorotational populations but do not seem to support the view that pyrimidine deoxyribonucleosides

TABLE 2: Distribution (%) of Simulated Pseudorotational Angle *P* in Conformational Ranges, According to a Four-State or a Two-State Model

		dA	rA	dC	rC	dG	rG	dT	rU
				Four-State Model[a]					
CHARMM27[b]	north	1.2	28.2	1.4	76.4	1.1	37.8	0.9	64.3
	east	47.6	12.6	52.6	7.2	36.2	11.2	44.5	11.2
	south	51.2	59.2	46.0	14.4	62.7	51.0	54.6	24.5
	west	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AMBER[b]	north	5.8	31.7					2.9	
	east	14.1	4.5					18.0	
	south	77.2	60.1					78.9	
	west	2.9	3.7					0.2	
				Two-State Model[a]					
CHARMM27[b]	NORTH	3.7	33.9	4.2	81.6	3.2	44.1	4.8	72.2
	SOUTH	96.3	66.1	95.8	18.4	96.8	55.9	95.2	27.8
	transitions[c]	3934	1771	4312	547	1730	988	2864	797
AMBER[b]	NORTH	7.6	35.0					4.8	
	SOUTH	92.4	65.0					95.2	
	transitions[c]	974	350					1168	
NMR[d]	NORTH	27	30	37	65	32	33	34	54
	SOUTH	73	70	63	35	68	67	66	46

[a] See Methods section for the definition of these models. [b] Obtained from molecular dynamics. [c] Number of transitions between NORTH and SOUTH during the simulations. [d] NMR data reported by Plavec et al.,[27] obtained at 278 K in aqueous solution.

populate the NORTH pucker significantly more than those with purines. Such a difference between purine and pyrimidine, however, seems confirmed for ribonucleosides.[27,30]

Regarding the syn/anti populations in aqueous solution, the conclusions from NMR studies are more elusive than for the sugar puckers,[23] even with ^{13}C labeling.[29,30] There is a vast amount of literature on the topic, in part because several NMR techniques were used to study the syn/anti equilibrium. None of these approaches, however, gives unequivocal results for the standard nucleosides.[23,29,30] In addition, there are discrepancies between the results from different approaches. Most methods suggest a predominant anti conformation for the standard nucleosides, with a higher syn population with purines than that for pyrimidines.[23] To our knowledge, however, an unambiguous quantitative estimate of these populations for every base is not yet available.

The presence of a syn population for purine nucleosides in solution was also detected by ultrasonic relaxation, assuming that it probed anti/syn transitions.[33,34] Pyrimidines (rC and rU) showed at most a trace of relaxation, suggesting that they did not undergo the same conformational exchange. The associated process occurred on a nanosecond time scale, with a time constant of 2.9 ns with dA, and was 2 times slower with rA.[33] More recent studies by ultrasonic relaxation processes confirmed that this technique may be uniquely suited to probe the syn/anti equilibrium,[44−46] but they concentrated on nucleotides rather than nucleosides. These studies claim to access the thermodynamic properties of the syn/anti equilibrium,[46] but it is unlikely that these properties would be the same for nucleosides as for nucleotides, because it is well-known that the presence of a 5′-phosphate strongly influences the relative energies of the syn and anti conformations.[40,59] Indeed, these properties vary substantially between mono-, bi-, and triphosphate 5′-nucleotides.[46] A major assumption in the interpretation of the ultrasonic relaxation studies was that they probed specifically the anti/syn equilibrium, without interferences from other degrees of freedom, such as torsion γ.

In solution, NMR showed [23] that γ is predominantly g^+ but also populates trans and g^-, with a preference for trans over g^-, which echoes X-ray observations. For nucleosides in the anti conformation, the γ populations in g^+, trans, and g^- were reported [23] to be (i) 61%, 25%, and 14%, respectively, with a purine and (ii) 57%, 29%, and 14%, respectively, with a pyrimidine. A significant decrease in the g^+ population and increase in the trans population were found to accompany the syn orientation in some, but not all, studies. The analysis of the influence of the base orientation on the γ populations seems to have been hampered by the difficulties in ascertaining the base orientation itself. For each base, a decrease in the g^+ population and a concomitant increase in the trans and g^- populations were found when going from ribose to deoxyribose.[23] NMR work with pyrimidine nucleosides has also led to the suggestion that the sugar puckering influences the $\gamma = g^+$ population, which was found to be higher with a C3′ endo than that with a C2′ endo sugar, with both ribose and deoxyribose.[19]

Pseudorotation Angle. The simulated pseudorotational populations are given in Figure 1 and Table 2. Detailed conformational analyses of the sugar furanose ring[15,39,60,61] show that it is meaningful to divide the pseudorotation space into four quadrants because this reflects the features of the associated potential energy surface. The two energy minima on this surface belong to the north and south quadrants, separated by energy barriers in the east and west areas. This is why we report the pseudorotational populations in terms of a four-state model, but we also discuss them according to the two-state model (NORTH/SOUTH) typically used in NMR studies (Table 2).

The relative potential energies between the north and south minima as well as the east barrier were previously obtained with high-level quantum mechanical calculations in vacuo,[39] and CHARMM27 reproduces these energetics reasonably well (Table 1). Given these potential energies, it is interesting to examine to which degree they may be mirrored in the simulated populations in water. Clearly, the west quadrant is not populated in solution, reflecting its high potential energy,[58,61] in agreement with NMR. The relative potential energies between the north and south minima are less predictive of the corresponding simulated populations for the deoxyribonucleosides. These energy differences are $\leq$1.0 kcal/mol in CHARMM27, and yet the deoxyribonucleosides populate the south much more than the north in the MD simulations. This is even the case for dC, for which the north potential energy is more stable than that of

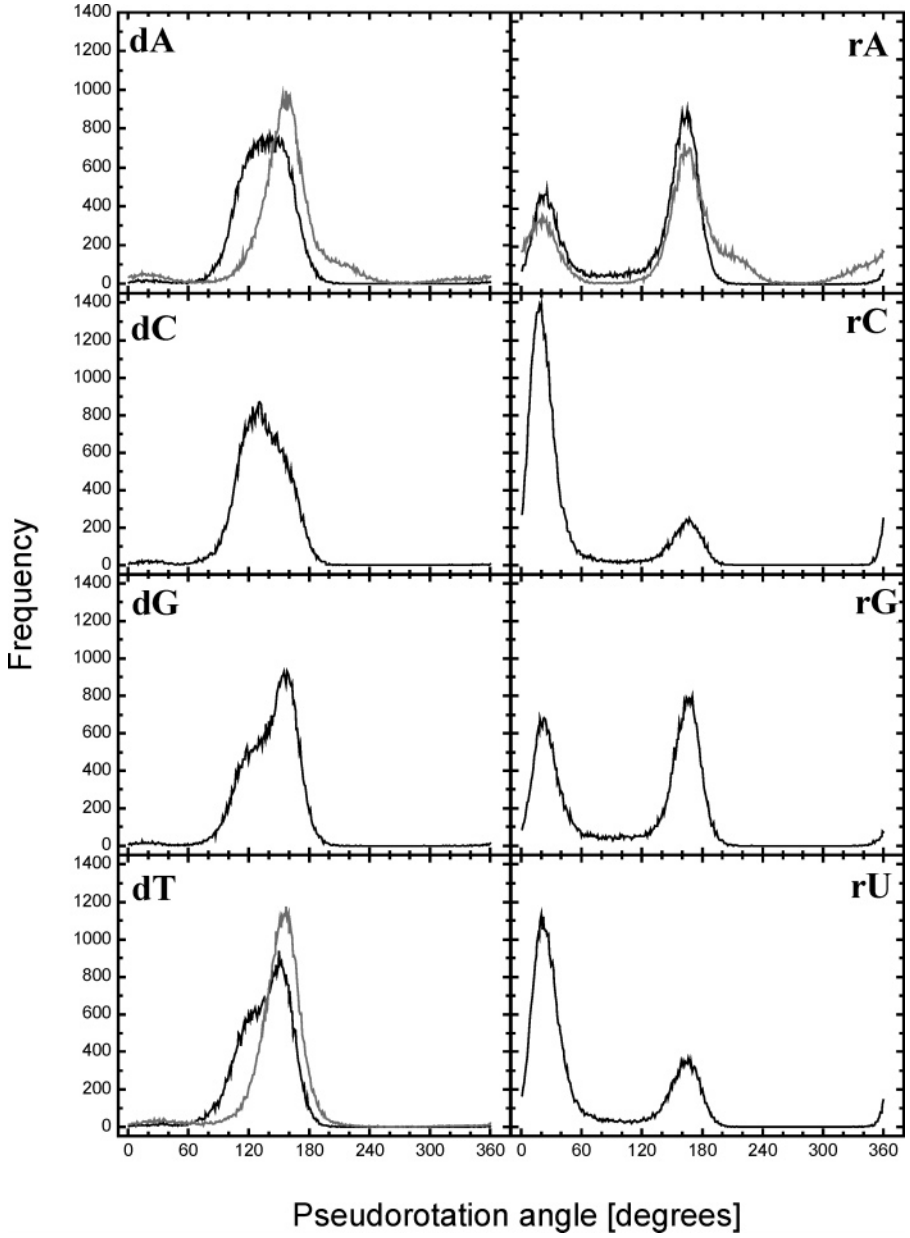

Figure 1. Frequency distribution of the simulated pseudorotation angle for each of the eight standard nucleosides, obtained with CHARMM27 (black) or the modified Cornell et al. AMBER[2,3] force field (gray). The nucleoside corresponding to the shown distributions is indicated in each panel.

the south. This reflects the influence of solvation (see Influence of Water subsection).

The fact that solvent effects can overcome and mask specific aspects of the intrinsic energetics of dC is worth noting because it means that the particularity of dC[39] could not be detected experimentally by probing of its NORTH and SOUTH populations. Therefore, dC populating SOUTH more than NORTH in solution is still compatible with the notion that dC has peculiar intrinsic energetics. It is important because this particularity of dC is likely to have an influence on the equilibrium between the various canonical forms of DNA[39] and possibly on noncanonical structures.[62]

dC is the only nucleoside for which the east is the most populated quadrant. The east is, however, also significantly populated with the other deoxyribonucleosides (Table 2). This was not anticipated given that there is a potential energy barrier in the east region but can be partly explained by stabilization of this barrier by water (see Influence of Water subsection). This shows again that caution should be exercised when comparing directly potential energies in vacuo to the corresponding distributions in the condensed phase. The simulated east populations appear higher than those expected from NMR,[31] indicating that the true east barrier is likely to be higher than that currently implemented in CHARMM27. This is consistent with the CHARMM27 east barrier being less than its ab initio counterpart (Table 1), and it strengthens the validity of this MP2/6-31G* ab initio result versus the lowest barrier obtained with the B3LYP/6-31G* method.[62] Nevertheless, our results suggest that the east region may be more populated in solution than previously anticipated.

Although a three-state analysis is possible,[31] NMR data obtained in solution were typically reported using a two-state model, where the east and west populations are implicitly assigned to the NORTH or SOUTH states. A direct comparison of these NMR populations and their MD counterpart (Table 2) should provide a useful assessment of how well-balanced is the force field used for the simulations. Overall, the NMR data yield a population of ~30% for the NORTH with deoxyribonucleo-

sides, while this state is populated almost 10 times less with CHARMM27. This seems to represent an imbalance in CHARMM27, barring difficulties in the interpretation of the NMR data. A similar low percentage of C3′ endo sugars was reported for the modified Cornell et al. AMBER force field when simulating nucleosides.[3] This may reflect the fact that these force fields were primarily parametrized to reproduce properties of oligonucleotides, with an emphasis on the B form of DNA. It was shown that CHARMM27 allows simulation of the equilibrium between the A and B forms of DNA and its dependence on some environmental factors.[63] Therefore, CHARMM27 does not preclude the existence of the north puckers in A DNA. It could be that this force field is well-adjusted for oligonucleotides and that this requires sacrificing the corresponding balance for deoxyribonucleosides. Alternatively, it may be that slight improvements in the sugar energetics would reconcile the simulated properties of DNA with those of its building blocks. This could be achieved without disrupting the results obtained with the ribose moiety, given that some atom types differ between the ribose and deoxyribose in CHARMM27.[4]

There is a notable difference in the simulated pseudorotational populations between deoxyribonucleosides and ribonucleosides. The ribonucleosides populate the east region significantly less than the deoxyribo compounds, probably due to a van der Waals repulsion between the eclipsed 2′- and 3′-hydroxyl groups in the O4′ endo conformation.[64] The north is clearly more populated by the ribose than the deoxyribose compounds, consistent with the trends in hydration energies (see Influence of Water subsection). Comparison of the NMR and CHARMM27 NORTH and SOUTH populations shows an excellent overall agreement. This is particularly interesting because for RNA the sugar energetics cannot be parametrized with respect to equilibria between various forms of RNA. In addition, it is more difficult with the ribose than with the deoxyribose to use in vacuo model compound calculations to calibrate the energetics, because these energetics are very sensitive to the relative orientations of the 2′-OH and 3′-OH groups (Foloppe, unpublished results). It would therefore be possible to inadvertently overstabilize the ribose north conformer, and this imbalance would not be detected by the standard simulations of A RNA. This would be unfortunate, because some of the complex tertiary structures formed by RNA may require that the ribose adopts a south pucker at some specific nucleotides. Therefore, other condensed phase reference data, in addition to A form RNA double helices, are necessary to test the ribose energetics. We propose that simulations of ribonucleosides in solutions are the most systematic and reliable way to perform this test. CHARMM27 is very well-balanced in that respect.

The frequency of the transitions between NORTH and SOUTH is on the order of several tenths per nanosecond (Table 2), with typically more transitions for the deoxyribose than the ribose. This is consistent with a higher O4′ endo barrier in the ribose and a less populated east range with the ribose than that for the deoxyribose.

In the ribose, intramolecular hydrogen bonds between O2′ and O3′ (each oxygen considered alternatively as a donor or acceptor) were formed relatively rarely, typically less than 3% of the time according to the geometric criteria given in the Methods section.

Puckering Amplitude. The average puckering amplitudes (Table 3) were extracted from the simulations for the all of the combinations of nucleoside type, pseudorotation domains (NORTH or SOUTH), and base orientation (anti or syn). The present simulations provide enough data to derive statistically sound averages for all of these combinations. The only experimental technique that gives access to the puckering amplitudes is X-ray crystallography, and the X-ray average amplitudes were derived for the north ($\tau_n = 37.3°$) and south ($\tau_s = 36.2°$) puckers without further categorization.[17] Therefore, for these properties, only indirect comparisons are possible between experiment and simulations in solution.

TABLE 3: Average Sugar Puckering Amplitude (deg) in the Simulated Nucleosides

		dA	rA	dC	rC	dG	rG	dT	rU
					CHARMM27				
anti	NORTH	47.0	39.7	47.7	40.3	47.0	39.6	48.5	39.9
	SOUTH	44.8	41.6	44.6	39.9	44.3	41.0	44.4	40.1
syn	NORTH	51.5	38.9	52.6	na[a]	50.9	38.9	50.7	40.3
	SOUTH	38.5	39.3	37.9	na[a]	36.6	38.8	39.5	40.2
					AMBER				
anti	NORTH	36.6	37.2					36.3	
	SOUTH	40.4	39.7					40.1	
syn	NORTH	35.8	37.0					34.3	
	SOUTH	36.5	36.9					36.5	

[a] Not available (na), because syn is not populated for rC.

The categorized average amplitudes for all deoxyribonucleosides are 47.6° (NORTH/anti), 44.4° (SOUTH/anti), 50.8° (NORTH/syn), and 38.0° (SOUTH/syn). For the ribonucleosides, they are 39.8° (NORTH/anti), 40.7° (SOUTH/anti), 39.4° (NORTH/syn), and 39.5° (SOUTH/syn). Overall, the simulated amplitudes tend to be higher for the deoxyribonucleosides than those for the ribonucleosides, although the opposite trend was observed with X-ray structures.[17] The simulated deoxyribose amplitudes are also significantly higher than their X-ray counterparts. This may reflect limitations of the force field, although there is excellent agreement between the CHARMM27 deoxyribose geometries (bond length and valence angles) and reference data (ab initio or X-ray).[4] Therefore, one cannot rule out that the deoxyribose may be more puckered in solution than when packed in a crystal, especially in view of the significant influence of water on other conformational properties (see Influence of Water subsection). The amplitudes may also depend on the temperature, further complicating the analysis because one would have to take into account the temperature at which the X-ray structures were solved. More work will be needed to clarify these issues.

With the ribose, the simulated amplitudes are essentially within the standard deviation of the average X-ray values,[17] from which they depart only by $\sim 2-3°$. In combination with the good agreement between simulated and experimental pseudorotational populations, this confirms that the ribose is well-represented in CHARMM27.

MD simulations are particularly well-suited to analyze possible correlations between puckering and the glycosyl torsion (χ). Overall, for the simulated SOUTH sugars, there is a systematic decrease of amplitude when χ goes from anti to syn. This indicates that the flexibility of the furanose plays a role in accommodating the syn conformation. There is, however, no such systematic trend for NORTH sugars. The correlation between pseudorotation and χ is addressed in the next section.

Glycosyl Torsion. This torsion controls the base orientation relative to the sugar, in anti or syn. This degree of freedom is key in defining the overall DNA and RNA conformations and their various canonical forms.[58] It is therefore critical to assess how force fields treat the glycosyl torsion. Unfortunately, simulations of oligonucleotides do not allow us to examine the anti/syn equilibrium, because that would require the intercon-

TABLE 4. Relative Potential Energies (kcal/mol) of the Syn and Anti Conformations in Deoxyribonucleosides

		$\Delta E_{syn-anti}$[a]			
		south[b]		north[b]	
		ab initio[c]	CHARMM27[d]	ab initio[c]	CHARMM27[d]
dA	$\gamma = g^+$	3.3	5.2	ne[e]	ne[e]
dC		7.7	8.7	7.1	ne[e]
dG		3.6	4.5	4.2	ne[e]
dT		6.7	7.0	6.4	ne[e]
dA	$\gamma =$ trans	3.8	4.1[f]	3.7	10.1
dC		6.7	5.7[f]	5.1	11.6
dG		5.2	4.7[f]	5.1	10.7
dT		6.9	5.5	5.8	11.6

[a] Potential energy with $\chi =$ syn ($\beta = 180.0°$ and γ in trans or g^+) minus the energy of the global energy minimum (χ in anti and γ in g^+). [b] Furanose pucker. [c] Obtained at the MP2/6-31G* level from ref 40. [d] Calculated with CHARMM27 as a part of this work. [e] Nonexistent (ne); there is no energy minimum corresponding to this conformation. [f] γ switches to g^- in the energy-minimized structure.

version between nucleic acid forms (e.g., B and Z DNA) that are unlikely to be probed even on state-of-the-art MD time scales. The nucleosides, where χ is not constrained by base pairing, present a unique opportunity to address this question.

The relative potential energies of the anti and syn conformations of deoxyribonucleosides were recently probed with high-level quantum mechanical calculations,[40,41,43] but their CHARMM27 counterpart has not been presented. These CHARMM27 potential energies were obtained as part of this work (Table 4). With a south pucker, there is a reasonable agreement between CHARMM27 and quantum mechanical results, the energy difference between the two types of calculations ranging from 0.3 to 1.9 kcal/mol. This is encouraging because these conformations were not included explicitly when parametrizing CHARMM27. With a north pucker, the deviations between molecular mechanics and ab initio are larger. All CHARMM27 energy minimizations initiated with a north pucker and $\gamma = g^+$ and $\chi =$ syn converged to a structure with χ in anti (including when χ was restrained to be in syn in the initial stages of the minimization). Therefore, there is no energy minimum corresponding to pucker = north, $\chi =$ syn, and $\gamma = g^+$ in CHARMM27. This, however, is only a subtle departure between force field and quantum mechanical results, given that these results showed that the energy minimum in question is intrinsically very shallow.[40] That does not prevent the existence of the north, $\chi =$ syn, and $\gamma = g^+$ conformation in the simulations, emphasizing the role of solvation (see below). There is a CHARMM27 energy minimum corresponding to pucker = north, $\chi =$ syn, and $\gamma =$ trans, but it is 5.6−6.5 kcal/mol higher than its ab initio counterpart (Table 4). This is expected to be a limitation for the simulation of Z DNA, where the purine nucleotides adopt this conformation. These high energies for pucker = north and $\chi =$ syn in the nucleosides confirm results obtained with simpler compounds, with both the CHARMM27 and the modified Cornell et al. AMBER force fields.[48] This discrepancy between molecular mechanics and ab initio is likely to reflect the favorable hydrogen bond character of the C3′−H···N3 (purines) and C3′−H···O2 (purines) interactions when treated quantum mechanically. These remarks, however, should not detract from the fact that the χ torsional energetics were parametrized and tested with great care in CHARMM27,[4,48] providing a robust basis to analyze the properties associated with χ in the simulated nucleosides.

It is of particular interest to analyze the simulated anti/syn populations (Table 5) and their influence on the pseudorotational

TABLE 5: Distribution[a] of the Glycosyl Torsion in the Simulations

	dA	rA	dC	rC	dG	rG	dT	rU
				CHARMM27				
anti	67.6	34.6	95.2	99.2	48.9	27.1	62.5	98.8
syn	14.5	45.7	0.9	0.0	33.9	49.6	26.4	0.6
other[b]	17.9	19.6	3.9	0.8	17.2	23.2	11.1	0.6
transitions[c]	28	141	1	0	56	72	8	2
				AMBER				
anti	10.0	11.0					35.1	
syn	19.3	40.5					53.1	
other[b]	70.7	48.5					11.8	
transitions[c]	228	302					95	

[a] The distribution in the conformational ranges anti, syn, and other is given in percent of the time. [b] Other denotes any conformation that is not anti or syn. [c] Number of transitions between anti and syn.

TABLE 6: Influence of the Glycosyl Conformation on Selected Properties in the Simulations

		dA	rA	dC	rC	dG	rG	dT	rU
					CHARMM27				
anti	NORTH[a]	4.0	40.7	4.2	81.6	4.0	44.7	5.1	72.3
	SOUTH[a]	96.0	59.3	95.8	18.4	96.0	55.3	94.9	27.7
syn	NORTH[a]	2.7	30.7	1.9	na[c]	1.4	43.8	4.7	68.1
	SOUTH[a]	97.3	69.3	98.1	na[c]	98.6	56.2	95.3	31.9
	O5′ Hbond[b]	76.6	73.4	81.4	na[c]	87.8	75.3	71.1	21.5
					AMBER				
anti	NORTH[a]	11.7	53.3					7.2	
	SOUTH[a]	88.3	46.7					92.8	
syn	NORTH[a]	4.1	26.7					3.0	
	SOUTH[a]	95.9	73.3					97.0	
	O5′ Hbond[b]	82.8	78.2					84.0	

[a] Percent of the time the sugar is in NORTH or SOUTH when χ is anti or syn. [b] Percent of the time that a hydrogen bond is formed between O5′ and the base in syn. [c] Not available (na), because syn is not populated by rC.

distributions (Table 6). Indeed, despite tremendous experimental efforts dedicated to the characterization of the populations corresponding to these various equilibria in aqueous solution, a clear and quantitative view of these properties has remained elusive (see Experimental Reference Data subsection). Therefore, the comparison between simulations and experiment for the anti/syn equilibrium is less straightforward and quantitative than that for the sugar pseudorotational populations. Relatively frequent transitions between anti and syn are observed in the present simulations, with the nucleosides that populate the syn conformation (Table 5). This shows that meaningful equilibrium populations can be extracted. Therefore, MD simulations obtained on the present time scale are a valuable complement to experiment, with respect to the quantification of the anti/syn populations, the kinetics of interconversion between them, and the χ distributions within the anti and syn ranges.

A well-established, but not quantitated, experimental result is that the syn conformation is significantly more populated with purines than with pyrimidines. This is largely what is observed in the simulations (Table 5, Figure 2, and Figure 3), with the intriguing exception of dT, which was found to populate syn 21.0% of the time. There is no indication that this derives from a force field artifact because the CHARMM27 parameters that control the energetics of the χ torsion are the same for all of the pyrimidines.[4] With dC, rC, and rU, the syn conformation is virtually not populated in the simulations, which is consistent with the results from X-ray, NMR, and ultrasonic relaxation. The presence of the syn conformation with dT but not dC may reflect, to some degree, the higher potential energy of dC in

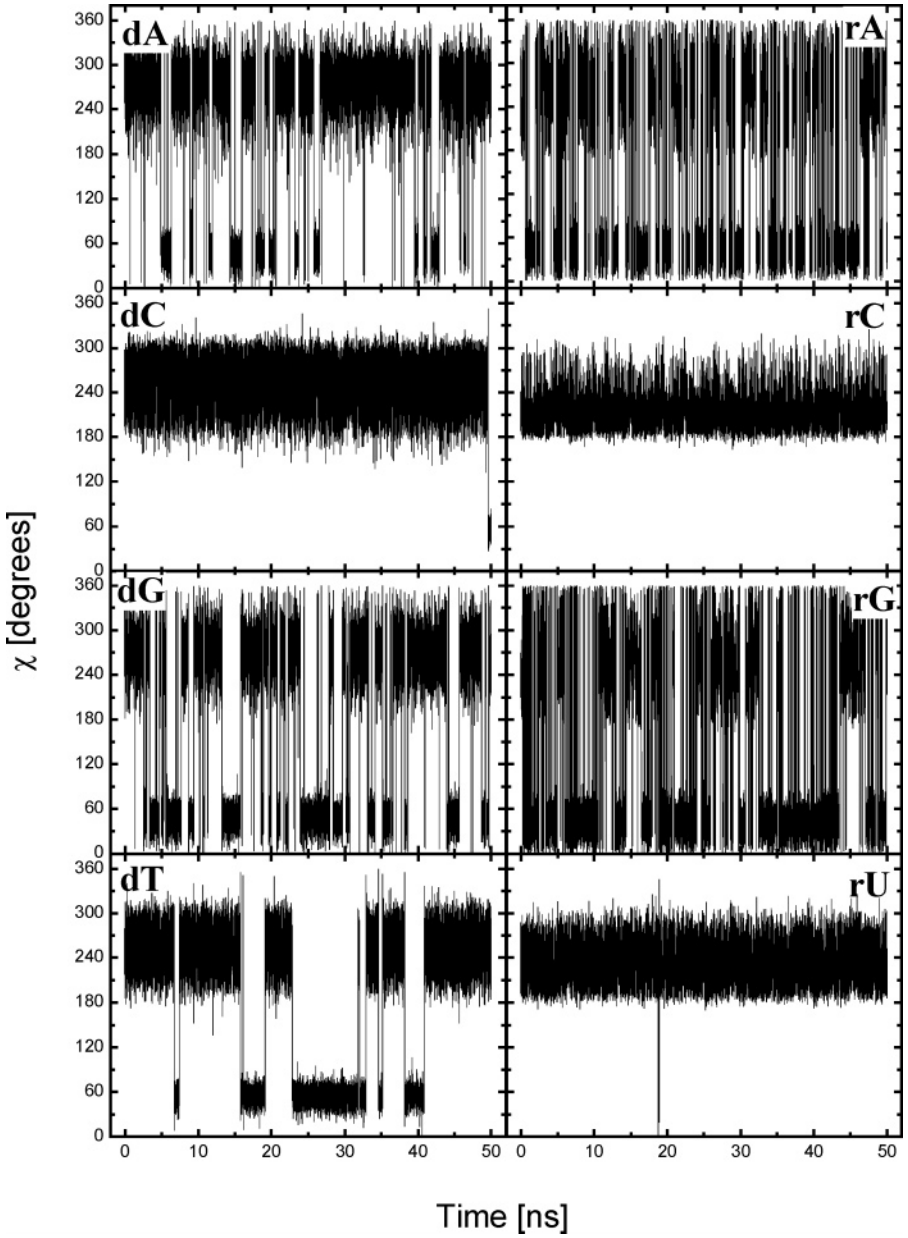

Figure 2. Torsion χ vs time along the simulations for the each of the eight standard nucleosides. The nucleoside corresponding to the shown time series is indicated in each panel.

syn, with a south pucker and $\gamma = g^+$ (Table 4). Differences in the potential energies for this conformation may also contribute to the explanation of why syn is more populated with dG than dA. But given the potential energies of syn relative to anti, it is the fact that the syn state is populated at all that is remarkable. One needs to keep in mind, however, that the potential energies presented in Table 4 were obtained with β constrained at 180.0°, which prevents formation of an intramolecular hydrogen bond between the base and the O5′−H group.

Such a hydrogen bond has been observed in X-ray structures of syn nucleosides[14] and was found to stabilize the syn orientation in quantum mechanical calculations.[42] But one cannot infer automatically from these observations in crystal or in vacuo that this hydrogen bond would be formed in aqueous solution, because of competition by water. To our knowledge, the existence of such a hydrogen bond in water has never been established. The simulations reveal that this hydrogen bond is frequently formed when the base is syn (Table 6), up to 88.1% of the time with dG. It is formed ∼10% of the time more frequently with dG than with dA, consistent with dG being ∼10% of the time more frequently in syn than dA. Therefore, the present results indicate that guanosine has a higher propensity to be in the syn conformation than adenosine. This had long been a matter of debate[16,23,65] and had been questioned based on recent ab initio calculations.[66]

The influence of the base is not limited to the relative anti/syn populations but extends to the shape of the χ distributions, especially within the anti range (Figure 3). The modal values of χ in anti are 287° (dA), 279° (dC), 285° (dG), 259° (dT), 294° (rA), 206° (rC), 285° (rG), and 210° (rU). For the deoxyribonucleosides, these modal values are significantly higher than those from the crystal structures[17] or the anti quantum mechanical energy minima in vacuo.[40] The values of χ in these minimized structures (south pucker and $\gamma = g^+$) were 207° for dC and ∼230° with the other bases. This departure between the energy minimum in vacuo and the simulated modal values illustrates the strong influence of water on the structures of the nucleosides. The most populated values of χ in water, for the deoxyribonucleosides, are similar to their counterpart in B DNA,[67] as opposed to the lower values of χ in these

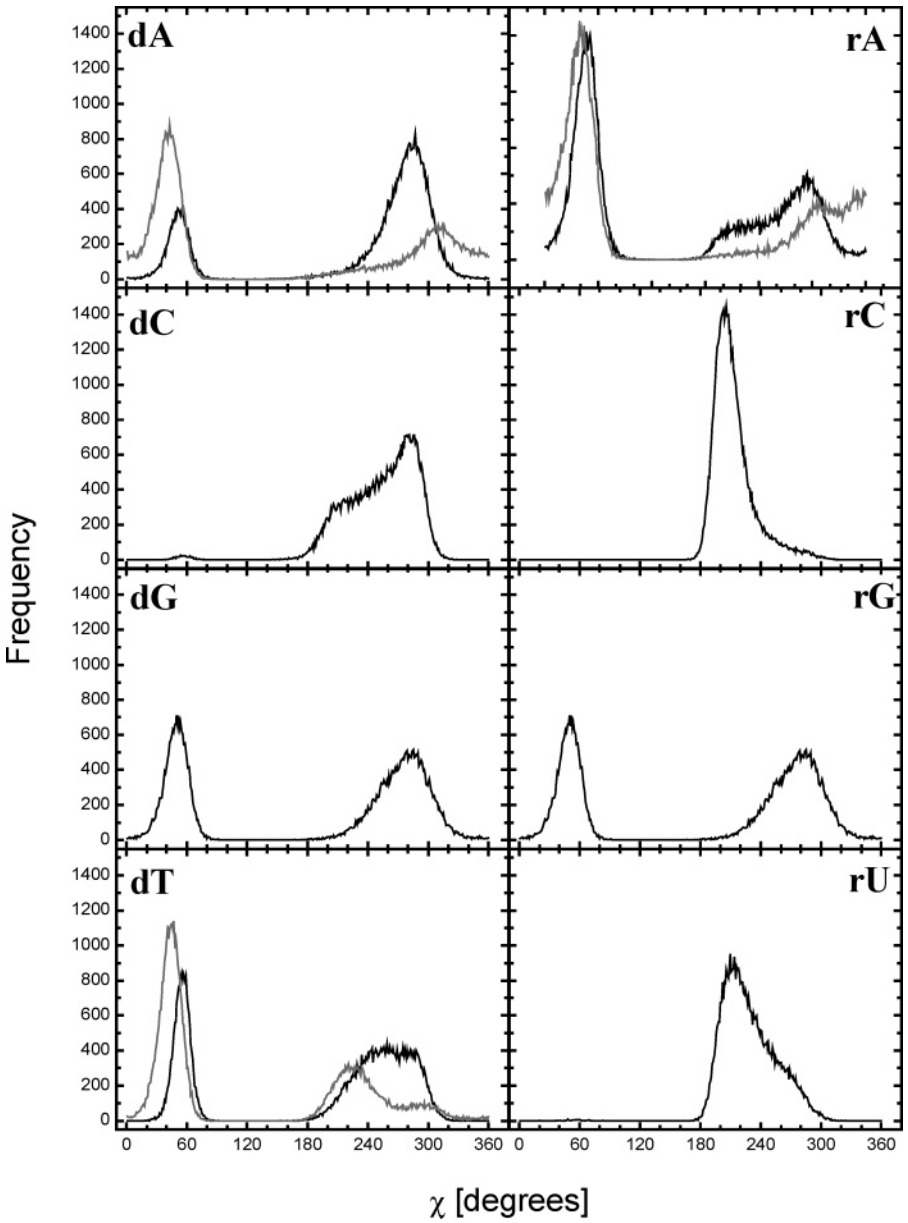

Figure 3. Frequency distribution of the simulated glycosyl torsion angle (χ) for each of the eight standard nucleosides, obtained with CHARMM27 (black) or the modified Cornell et al. AMBER[2,3] force field (gray). The nucleoside corresponding to the shown distributions is indicated in each panel.

nucleosides minimized in vacuo. Therefore, contrary to what could be inferred from these minimizations or the X-ray structures of the nucleosides,[17] the nucleosides in water appear to be preorganized for their incorporation into B DNA, with no conformational strain associated with systematic adjustments in χ. Interestingly, the simulated modal values of χ for rC and rU are much lower than those of dC and dT and are close to the canonical values of χ in A helices formed by RNA. The χ distributions (Figure 3) show that the so-called high anti (χ in the vicinity of 300°) conformation is significantly more populated in water than anticipated from the crystal structures of nucleosides.[16,58] With purines, the high anti populations extend as far as $\chi = 360°$, showing that the syn/anti interconversion takes place preferentially through this pathway rather than via $\chi = 120°$. It is, to our knowledge, the first time that precise information on the values of χ for the nucleosides in solution is presented, given that this information has not yet been captured by experiment.

In addition to the effect of the base, the present work uncovers an influence of the sugar on the anti/syn equilibrium. With purines, the ribose clearly favors the syn state more than the deoxyribose (Table 5). In syn, the ribose populates both the NORTH and SOUTH states, strengthening the notion that a NORTH pucker is compatible with the syn orientation, as suggested by recent ab initio calculations[40,43] and contrary to a long held view.[17,68] In the syn conformation, the NORTH tends to interact more favorably with water than the SOUTH (see below). The NORTH, however, is less populated in syn than in anti, with the only exception of rG.

Another aspect of the anti/syn equilibrium is the rate at which the nucleosides switch from anti to syn (Table 5 and Figure 2). With the purines, the calculated rates of transition are on the order of 1 ns^{-1}. This is essentially in agreement with ultrasonic relaxation results, which suggested an anti/syn transition rate on a nanosecond time scale in purine nucleosides.[33] Although this conclusion was reiterated by additional ultrasonic relaxation studies on nucleotides,[44] it rested on the assumption that the process probed experimentally was the anti/syn transition. The direct observation of these transitions with a frequency on the nanosecond time scale in the simulations significantly strength-

TABLE 7: Conformational Distribution[a] of the Torsional Angle γ in the Simulations

	dA	rA	dC	rC	dG	rG	dT	rU
				CHARMM27				
g^+	95.1	94.7	93.6	94.3	95.2	95.1	94.2	93.8
trans	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1
g^-	0.4	0.3	0.8	0.1	0.7	0.3	1.0	0.1
other	4.4	4.9	5.4	5.5	4.0	4.5	4.7	6.0
				AMBER				
g^+	55.7	57.0					71.0	
trans	42.7	41.0					27.2	
g^-	0.3	0.4					0.5	
other	1.3	1.6					1.3	

[a] In percent of the time. See Methods section for the definition of g^+, trans, g^-, and other.

TABLE 8: Effect of the γ Conformation on Deoxynucleosides Potential Energies (kcal/mol)

		dC		dT	
		ab initio[a]	CHARMM27[b]	ab initio[a]	CHARMM27[b]
$\gamma = g^+$	north	0.0	0.0	0.6	0.2
	south	0.4	0.2	0.0	0.0
γ = trans	north	4.4	9.6	3.9	9.2
	south	2.7	7.9	5.5	7.1

[a] Obtained at the MP2/6-31G*//RHF/6-31G* level, from ref 3. [b] Obtained as part of this work.

ens the interpretation of the ultrasonic relaxation work. This experimental work also suggested that the anti/syn transition rate was 2 times faster in dA than in rA,[33] but the simulations yield the opposite trend, and additional work will be needed to resolve this discrepancy.

Given the direct interaction between the 5′-OH group and a base in syn, the details of the anti/syn equilibrium are also expected to depend on the flexibility of torsion γ.

Torsion γ. The γ conformation is typically analyzed in terms of its populations in the three staggered conformers g^+, trans, and g^- (Table 7). In the simulations, γ populates predominantly the g^+ conformation, in agreement with experiment. Although the γ = trans and $\gamma = g^-$ conformers are visited during the simulations, their populations are significantly less than their experimental counterparts. For γ = trans, this reflects too high potential energies in the force field (Table 8), and we suggest that adjustment in these potential energies would yield a more flexible and realistic torsion γ. The population categorized as other in Table 7 corresponds to values of γ that are just at the periphery of the range $\gamma = g^+$ and could be lumped together with $\gamma = g^+$ if one adopted a broader definition of $\gamma = g^+$ ($60 \pm 40°$ instead of $60 \pm 30°$).

Influence of Water. The present simulations give us the opportunity to analyze how water may preferentially stabilize some nucleoside conformations relative to others. The average total interaction energies between nucleoside and water are given in Table 9, depending on the nucleoside conformation. For completeness, these averages are also presented for the pyrimidines in syn, but these statistics may suffer from the marginal sampling of this conformation with pyrimidines other than thymine.

The interaction energies of the deoxyribonucleosides in the anti conformation with water systematically favor the south pucker over the north (see ΔIE_{n-s}^{anti} in Table 9). These interaction energies tend to be on the order of, but slightly larger than, the intrinsic energy differences between north and south in vacuo (Table 1). In particular, the magnitude of the interaction with the solvent explains why dC populates the south more than the north in water, despite the opposite trend being observed in vacuo. This is compatible with the peculiar intrinsic energetics of dC becoming more influential in conditions of low water activity, such as those that promote A and Z DNA.[39] In contrast, in the anti conformation, water stabilizes the north ribose more than the south (except with rG), consistent with the north being more populated with the ribose than with the deoxyribose compounds.

TABLE 9: Conformational Dependence of the Average Interaction Energies (kcal/mol) of the Nucleosides with Water[a]

	dA	rA	dC	rC	dG	rG	dT	rU
				CHARMM27				
north/anti	−51.3	−60.1	−57.0	−57.8	−65.0	−69.5	−63.2	−53.1
south/anti	−52.5	−58.1	−59.0	−62.7	−65.7	−72.3	−64.6	−51.9
ΔIE_{n-s}^{anti}	1.2	−2.0	2.0	4.9	0.7	2.8	1.4	−1.2
north/syn	−56.9	−61.2	−65.2	na[b]	−70.2	−68.5	−61.2	−50.4
south/syn	−47.4	−57.1	−56.6	na[b]	−60.0	−63.4	−61.5	−55.4
ΔIE_{n-s}^{syn}	−9.5	−4.1	−8.6	na[b]	−10.2	−5.1	0.3	5.0
east/anti	−54.7	−62.3	−60.1	−60.8	−66.4	−71.9	−64.3	−54.2
$\Delta IE_{e-(n/s)}^{anti}$	−2.2	−2.2	−1.1	1.9	−0.7	0.4	0.3	1.1
				AMBER				
north/anti	−53.7	−53.8					−47.4	
south/anti	−55.0	−56.9					−50.8	
ΔIE_{n-s}^{anti}	1.3	3.1					3.4	
north/syn	−52.5	−50.5					−53.0	
south/syn	−47.6	−49.8					−48.2	
ΔIE_{n-s}^{syn}	−4.9	−0.7					−4.8	
east/anti	−55.4	−53.6					−51.1	
$\Delta IE_{e-(n/s)}^{anti}$	−0.4	3.3					−0.3	

[a] ΔIE_{n-s}^{anti} is the interaction energy in north/anti minus the interaction energy in south/anti. ΔIE_{n-s}^{syn} is the interaction energy in north/syn minus the interaction energy in south/syn. $\Delta IE_{e-(n/s)}^{anti}$ is the interaction energy in east/anti minus the most favorable interaction energy in either north/anti or south/anti. [b] Not available (na), because the syn conformation is not populated for rC.

With most nucleosides in the anti conformation (except rG and dT), however, it is the east puckers that are most stabilized by water (see $\Delta IE_{e-(n/s)}^{anti}$ in Table 9). For half of the nucleosides, this interaction between water and east/anti is more favorable than with the north/anti or south/anti by at least 1.0 kcal/mol. This significant stabilization of the pseudorotation energy barrier between north and south could not be anticipated from the crystal structures of nucleosides, where the east range is virtually not populated.[17] This contributes to the explanation of why the east range is significantly populated in the present simulations of the deoxyribonucleosides (Table 2 and Figure 1), despite this puckering corresponding to an intrinsic potential energy barrier. Therefore, there is a real possibility that the east range may be more populated in aqueous solution than presumed so far. It should be kept in mind, however, that the intrinsic O4′ endo CHARMM27 energy barrier is lower than its ab initio counterpart (Table 1), which also contributes to populating the east range in these simulations. More work will therefore be needed to refine the characterization of the east populations in aqueous solution.

Another striking effect of water is the large stabilization of north/syn relative to south/syn, by more than 4.0 kcal/mol with dA, rA, dC, dG, and rG (see ΔIE_{n-s}^{syn} in Table 9). This would explain why the NORTH range is still observed in syn despite unfavorable intramolecular potential energies in north/syn (Table 6). In addition, solvation tends to favor north/syn more than north/anti or south/anti. dT is an exception to this trend, and solvation does not explain why thymine populates the syn orientation more than other pyrimidines in the simulations.

Overall, these results show that water has a profound influence on the conformational properties of the nucleosides and that it is dangerous to extrapolate from potential energies in vacuo or distributions in crystals to the situation in aqueous solution.

Force Field Dependency. To put in perspective possible dependencies of the above CHARMM27 results on force field specificities, other simulations were performed with the modified Cornell et al. AMBER[2,3] force field, which is also widely used. A full and detailed comparison between the two force fields is beyond the scope of this work, and only selected points of interests are highlighted. Only dA, rA, and dT were simulated with AMBER.

For the deoxyribose, AMBER, like CHARMM27, yields less NORTH puckers than expected from NMR (Table 2). The proportion of NORTH puckers in AMBER is much higher with the ribose, like with CHARMM27. AMBER also leads to a significant population of east pucker for dA and dT. The AMBER pucker distributions extend more into the west quadrant than their CHARMM27 counterparts (Figure 1). The rate of puckering transitions between NORTH and SOUTH is slower with AMBER than that with CHARMM27, and the amplitudes of puckering are systematically lower in AMBER than those in CHARMM27.

It is very interesting to note that, like with CHARMM27, the AMBER χ distributions in the anti conformation are strongly shifted toward the high anti range. That strongly suggests that the high anti region is truly much more populated in aqueous solution than in the crystals[17] or than would be inferred from ab initio energy profiles in vacuo.[40] The trend toward high anti is even more pronounced with AMBER than with CHARMM27, with surprisingly high populations at $\chi \approx 360°$ (Figure 3). This suggests that the χ torsional energy barrier in this region may be underestimated in AMBER, consistent with previously obtained potential energy profiles.[48]

An underestimated energy barrier at $\chi \approx 360°$ is consistent with a rate of interconversion between syn and anti that is significantly higher in AMBER than that in CHARMM27. When compared to the ultrasonic relaxation data,[33] the AMBER rate of interconversion between syn and anti appears overestimated, especially with the pyrimidine. This clearly demonstrates the usefulness of this type of simulations for an assessment of the force field. It may be helpful to revisit the parameterization of torsion χ in AMBER with the model compounds specifically designed to that effect.[47]

With respect to torsion γ, AMBER populates both g^+ and trans (Table 7) and seems better balanced than CHARMM27, even though it is expected that g^- should also be populated.

The simulations with AMBER confirmed the strong influence of water on the nucleoside conformational preferences (Table 9). By and large, the trends in interaction energies between nucleoside and water are similar between AMBER and CHARMM27. In the anti conformation, water stabilizes the deoxyribose in south more than in north, while the opposite is observed in the syn conformation. Importantly, water also interacts particularly favorably with the east deoxyribose in the AMBER simulations.

IV. Conclusions

For each of the eight standard nucleoside constituents of nucleic acids, a molecular dynamics simulation of at least 50 ns in explicit water was obtained, with a modern nucleic acid force field, CHARMM27. Results with the modified Cornell et al. AMBER[2,3] force field were also presented for comparison. This was to complement experimental approaches with respect to the conformations and dynamics of nucleosides in aqueous solution, which are relevant to countless biochemical phenomena. A survey of the experimental literature strongly indicates that this is needed because, despite considerable efforts in this direction, a full and quantitative experimental characterization of nucleosides in solution has remained elusive.

The present work demonstrates that MD simulations can now be obtained on a time scale that yields the equilibrium populations of all of the conformational degrees of freedom of the nucleosides and the associated kinetics of interconversion. It is also straightforward to extract the coupling between the degrees of freedom, which tends to confound the interpretation of the experimental data. Therefore, our knowledge of the properties of the nucleic acid building blocks in solution is arguably only limited by the accuracy of the potential function used to simulate them. This leads to an interplay between complementing experimental observations and using them to test the quality of the force field.

This is why this work dwells on how simulations of nucleosides provide an invaluable test of any nucleic acid force field. This is primarily illustrated with CHARMM27, because it is a widely distributed and already a highly tested and validated force field.[4] The simulations of nucleosides, however, are shown to offer additional and unique, although simple, tests. Some of these tests involve the ribose pseudorotation populations, the rate of the anti/syn interconversion, and the populations of torsion γ, which can hardly be sampled with simulations of oligonucleotides. The simulated results were put in the context of the underlying potential energies, which were obtained when not already available from the literature. A majority, but not all, of these newly presented CHARMM27 potential energies are in close agreement with their ab initio counterparts. Interestingly, we found several examples where the potential energies in vacuo are poor predictors of the corresponding simulated populations in water, stressing the need to perform these simulations for a meaningful comparison to experiment. Clearly, simulations of nucleosides are invaluable to bridge the gap between model compounds in vacuo and simulations of oligonucleotides in explicit solvent, for the purpose of force field development.

Both a four-state and a two-state model were used to analyze the simulated pseudorotation populations. The NMR-derived populations for the sugar puckers provide fairly reliable and quantitative references for comparison with the simulated results. For the ribose pseudorotational populations, we find an excellent agreement between experiment and simulations, including with respect to the influence of the base (purine versus pyrimidine) on the relative NORTH and SOUTH populations. With the deoxyribose, CHARMM27 underpopulates the NORTH state. The relatively high deoxyribose simulated population in the east may also be a reason for concern, although it partly reflects the influence of water. This population is systematically lower with the simulated ribose. Another difference between the two sugars lies in the simulated amplitude of the puckering, which tends to be systematically and significantly higher with the deoxyribose than that with the ribose. Although there is no reference experimental data for this property in solution, the ribose simulated amplitude is closer to its X-ray counterpart than that for the deoxyribose. These results suggest that the ribose is particularly well-represented in CHARMM27 but that additional fine-tuning may be required for the deoxyribose.

The NMR-derived populations for torsion γ in nucleosides mirror their X-ray counterpart and are a useful guide for force

field assessment. This assessment suggests that further refinement of the γ torsional energetics in CHARMM may be required for γ to populate trans and g^+. While γ is in g^+, the 5′-OH group is frequently hydrogen-bonded to the base in syn, despite competition by water. This hydrogen bond had long been suggested to stabilize the syn conformation based on X-ray structures, and the present work shows that this interaction is relevant to the aqueous solution conditions.

In contrast to degrees of freedom that can be analyzed with coupling constants, NMR met with particular difficulties with regard to the glycosyl torsion χ and a detailed characterization of the anti/syn equilibrium. This has been vexing because the associated properties are major determinants of nucleoside recognition by proteins and of nucleic acid structural and functional properties. Great care was therefore taken when calibrating CHARMM27 with respect to χ energetics.[4,48] The simulations provide a fresh, more complete, and quantitative picture of the syn/anti equilibrium, giving direct access to the influence of the sugar type, sugar pucker, and base on the anti/syn populations. Surprisingly, dT was found to significantly populate the syn orientation, which challenges the notion that pyrimidines, as a whole, do not adopt the syn conformation. This is worth experimental testing. The χ distributions in the anti conformation are strongly affected by water and depart from the corresponding X-ray modal values and the associated energy minima in vacuo. With the deoxyribonucleosides in water, χ is significantly shifted toward high anti relative to the X-ray distributions, despite the associated potential energy profiles. The deoxyribonucleosides in water are therefore preorganized for incorporation into B DNA, which was not apparent from previous data. Likewise, the χ distributions of rC and rU in water show a preorganization for incorporation into RNA. With both sugars, guanine favors the syn conformation more than adenine. The ribose favors the syn conformation significantly more than the deoxyribose. The NORTH pucker can coexist with syn conformers. The rate of the anti/syn transitions with purines is on the nanosecond time scale. Water has a profound influence on the conformational properties of the nucleosides.

Overall, the present work illustrates the usefulness of MD simulations to explore the structure and dynamics of molecules of biochemical interest. Although much effort is dedicated to apply this technique to large biopolymers, there are many smaller molecules (cofactors, vitamins, second messengers, hormones, and drugs) for which a detailed conformational analysis in solution is needed but missing. This work shows that MD simulations are becoming mature enough to fully characterize the conformational properties of these small molecules in solution. This should prove particularly powerful when trying to elucidate the pharmacophores of drug candidates.

Acknowledgment. This work was supported by the Swedish Research Council.

Abbreviations

dA = deoxyadenosine
dC = deoxycytidine
dG = deoxyguanosine
dT = deoxythymidine
MD = molecular dynamics
NPT = constant number of particles (*N*), constant pressure (*P*), and constant temperature (*T*)
NVT = constant number of particles (*N*), constant volume (*V*), and constant temperature (*T*)
rA = riboadenosine
rC = ribocytidine
rG = riboguanosine
rU = ribouridine

References and Notes

(1) Langley, D. R. *J. Biomol. Struct. Dyn.* **1998**, *16*, 487.
(2) 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. *J. Am. Chem. Soc.* **1995**, *117*, 9.
(3) Cheatham, T. E., III.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.
(4) Foloppe, N.; MacKerell, A. *J. Comput. Chem.* **2000**, *21*, 86.
(5) Cheatham, T. E., III.; Kollman, P. A. *J. Mol. Biol.* **1996**, *259*, 434.
(6) Cheatham, T. E., III.; Crowley, M. F.; Fox, T.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 9626.
(7) Young, M. A.; Ravishanker, G.; Beveridge, D. L. *Biophys. J.* **1997**, *73*, 2313.
(8) Auffinger, P.; Westhof, E. *Curr. Opin. Struct. Biol.* **1998**, *8*, 227.
(9) Norberg, J.; Nilsson, L. *Acc. Chem. Res.* **2002**, *35*, 465.
(10) Huang, N.; Banavali, N. K.; MacKerell, A. D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 68.
(11) Sarzynska, J.; Nilsson, L.; Kulinski, T. *Biophys. J.* **2003**, *85*, 3445.
(12) Cheatham, T. E., III.; Young, M. A. *Biopolymers* **2001**, *56*, 232.
(13) Varnai, P.; Djuranovic, D.; Lavery, R.; Hartmann, B. *Nucleic Acids Res.* **2002**, *30*, 5398.
(14) Rao, S. T.; Sundaralingam, M. *J. Am. Chem. Soc.* **1970**, *92*, 4963.
(15) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205.
(16) Saenger, W. *Angew. Chem., Int. Ed. Engl.* **1973**, *12*, 591.
(17) Gelbin, A.; Schneider, B.; Clowney, L.; Hsieh, S.-H.; Olson, W. K.; Berman, H. M. *J. Am. Chem. Soc.* **1996**, *118*, 519.
(18) Westhof, E.; Sundaralingam, M. *J. Am. Chem. Soc.* **1980**, *102*, 1493.
(19) Hruska, F. E.; Smith, A. A.; Dalton, J. G. *J. Am. Chem. Soc.* **1971**, *25*, 4334.
(20) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1973**, *95*, 2333.
(21) Schweizer, M. P.; Banta, E. B.; Witkowski, J. T.; Robins, R. K. *J. Am. Chem. Soc.* **1973**, *95*, 3370.
(22) Röder, O.; Lündemann, H.-D.; Goldammer, V. *Eur. J. Biochem.* **1975**, *53*, 517.
(23) Davies, D. B. *Prog. Nucl. Magn. Reson. Spectrosc.* **1978**, *12*, 135.
(24) Guschlbauer, W.; Jankowski, K. *Nucleic Acids Res.* **1980**, *8*, 1421.
(25) Uesugi, S.; Miki, H.; Ikehara, M.; Iwahashi, H.; Kyogoku, Y. *Tetrahedron Lett.* **1979**, *42*, 4073.
(26) Plavec, J.; Thibaudeau, C.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **1994**, *116*, 6558.
(27) Plavec, J.; Thibaudeau, C.; Chattopadhyaya, J. *Pure Appl. Chem.* **1996**, *68*, 2137.
(28) Plavec, J.; Tong, W.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **1993**, *115*, 9734.
(29) Bandyopadhyay, T.; Wu, J.; Stripe, W. A.; Carmichael, I.; Serianni, A. S. *J. Am. Chem. Soc.* **1997**, *119*, 1737.
(30) Kline, P. C.; Serianni, A. S. *J. Am. Chem. Soc.* **1990**, *112*, 7373.
(31) Olson, W. K. *Nucleic Acids Res.* **1981**, *9*, 1251.
(32) Ikehara, M.; Uesugi, S.; Yoshida, K. *Biochemistry* **1972**, *11*, 830.
(33) Rhodes, L. M.; Schimmel, P. R. *Biochemistry* **1971**, *10*, 4426.
(34) Hemmes, P. R.; Oppenheimer, L.; Jordan, F. *J. Am. Chem. Soc.* **1974**, *96*, 6023.
(35) Leulliot, N.; Ghomi, M.; Jobic, H.; Bouloussa, O.; Baumruk, V.; Coulombeau, C. *J. Phys. Chem. B* **1999**, *103*, 10934.
(36) Jordan, F.; Pullman, B. *Theor. Chim. Acta* **1968**, *9*, 242.
(37) Saran, A.; Pullman, B.; Perahia, D. *Biochim. Biophys. Acta* **1972**, *287*, 211.
(38) Saran, A.; Perahia, D.; Pullman, B. *Theor. Chim. Acta* **1973**, *30*, 31.
(39) Foloppe, N.; MacKerell, A. *Biophys. J.* **1999**, *76*, 3206.
(40) Foloppe, N.; Hartmann, B.; Nilsson, L.; MacKerell, A. D. *Biophys. J.* **2002**, *82*, 154.
(41) Leulliot, N.; Ghomi, M.; Scalmani, G.; Berthier, G. *J. Phys. Chem. A* **1999**, *103*, 8716.
(42) Shishkin, O. V.; Pelmenshikov, A.; Hovorun, D. M.; Leszczynski, J. *J. Mol. Struct.* **2000**, *526*, 329.
(43) Hocquet, A.; Leulliot, N.; Ghomi, M. *J. Phys. Chem. B* **2000**, *104*, 4560.
(44) Kuramoto, N.; Nishikawa, S.; Jordan, F. *J. Phys. Chem. B* **1998**, *102*, 9181.
(45) Nishikawa, S.; Huang, H.; Jordan, F. *J. Phys. Chem. B* **2000**, *104*, 1391.
(46) Nishikawa, S.; Kuramoto, H.; Huang, H.; Jordan, F. *J. Phys. Chem. B* **1999**, *103*, 3754.
(47) Foloppe, N.; MacKerell, A. *J. Phys. Chem. B* **1999**, *103*, 10955.
(48) Bosh, D.; Foloppe, N.; Pastor, N.; Pardo, L.; Campillo, M. *THEOCHEM* **2001**, *537*, 283.

(49) MacKerell, A. D.; Banavali, N.; Foloppe, N. *Biopolymers* **2001**, *56*, 257.
(50) Marquez, V. E.; Lim, M.-I. *Med. Res. Rev.* **1986**, *6*, 1.
(51) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(52) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667.
(53) Norberg, J.; Nilsson, L. *Biophys. J.* **2000**, *79*, 1537.
(54) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.
(55) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(56) Reiher, W. E., III. Ph.D. Thesis, Harvard University, 1985.
(57) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C.; 327−341. *J. Comput. Phys.* **1977**, *23*, 327.
(58) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
(59) Olson, W. K. *Biopolymers* **1973**, *12*, 1787.
(60) Olson, W. K.; Sussman, J. L. *J. Am. Chem. Soc.* **1982**, *104*, 270.
(61) Foloppe, N.; MacKerell, A. *J. Phys. Chem. B* **1998**, *102*, 6669.
(62) Foloppe, N.; Nilsson, L.; MacKerell, A. D. *Biopolymers* **2001**, *61*, 61.
(63) MacKerell, A. D.; Banavali, N. *J. Comput. Chem.* **2000**, *21*, 105.
(64) Olson, W. K. *J. Am. Chem. Soc.* **1982**, *104*, 278.
(65) Guschlbauer, W. *Jerusalem Symp. Quantum Chem. Biochem.* **1972**, *4*, 297.
(66) Hocquet, A.; Leulliot, N.; Ghomi, M. *J. Phys. Chem. B* **2000**, *104*, 4560.
(67) Schneider, B.; Neidle, S.; Berman, H. M. *Biopolymers* **1997**, *42*.
(68) Haschemeyer, A. E. V.; Rich, A. *J. Mol. Biol.* **1967**, *27*, 369.