Geometrical and Electronic Structure Variability of the Sugar–phosphate Backbone in Nucleic Acids

Daniel Svozil,^{*,†} Judit E. Šponer,[‡] Ivan Marchan,^{§,||,⊥} Alberto Pérez,^{§,||,⊥} Thomas E. Cheatham III,[#] F. Forti,[∇] F. Javier Luque,[∇] Modesto Orozco,[○] and Jiří Šponer^{†,‡}

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo náměstí 2, 166 10, Prague 6, Czech Republic, Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic, Molecular Modeling and Bioinformatics Unit, Institut de Recerca Biomèdica, Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028, Spain, Departament de Bioquímica i Biologia Molecular, Universitat de Barcelona, Avda Diagonal 645, Barcelona 08028, Spain, Barcelona Supercomputing Center, Jordi Girona 31, Edifici Torre Girona, Barcelona 08034, Spain, Department of Medicinal Chemistry and Department of Pharmaceuticals and Pharmaceutical Chemistry, College of Pharmacy, University of Utah, Salt Lake City, Utah 84112, Department of Physical Chemistry and Institute of Biomedicine (IBUB), Faculty of Pharmacy, University of Barcelona, Avenida Diagonal 643, 08028 Barcelona, Spain, and Computational Biology Program, Barcelona Supercomputer Center, Jordi Girona 29, Edifici Nexus II, Barcelona 08034, Spain

Received: February 12, 2008; Revised Manuscript Received: April 21, 2008

The anionic sugar–phosphate backbone of nucleic acids substantially contributes to their structural flexibility. To model nucleic acid structure and dynamics correctly, the potentially sampled substates of the sugar–phosphate backbone must be properly described. However, because of the complexity of the electronic distribution in the nucleic acid backbone, its representation by classical force fields is very challenging. In this work, the three-dimensional potential energy surfaces with two independent variables corresponding to rotations around the α and γ backbone torsions are studied by means of high-level *ab initio* methods (B3LYP/6-31+G*, MP2/6-31+G*, and MP2 complete basis set limit levels). The ability of the AMBER ff99 [Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074] and parmbsc0 [Perez, A.; Marchan, I.; Svozil, D.; Šponer, J.; Cheatham, T. E.; Laughten, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829] force fields to describe the various α/γ conformations of the DNA backbone accurately is assessed by comparing the results with those of *ab initio* quantum chemical calculations. Two model systems differing in structural complexity were used to describe the α/γ energetics. The simpler one, SPM, consisting of a sugar and methyl group linked through a phosphodiester bond was used to determine higher-order correlation effects covered by the CCSD(T) method. The second, more complex model system, SPSOM, includes two deoxyribose residues (without the bases) connected via a phosphodiester bond. It has been shown by means of a natural bond orbital analysis that the SPSOM model provides a more realistic representation of the hyperconjugation network along the C5′–O5′–P–O3′–C3′ linkage. However, we have also shown that quantum mechanical investigations of this model system are nontrivial because of the complexity of the SPSOM conformational space. A comparison of the *ab initio* data with the ff99 potential energy surface clearly reveals an incorrect ff99 force-field description in the regions where the γ torsion is in the *trans* conformation. An explanation is proposed for why the α/γ flips are eliminated so successfully when the parmbsc0 force-field modification is used.

Introduction

Molecular dynamics (MD) simulations of nucleic acids (NAs) are especially challenging owing to both the high flexibility of DNA and RNA on the one hand and the crucial role played by long-range electrostatic interactions on the other. This had complicated their simulation for many years.^{3–10} Fortunately,

progress in the development of simulation methods over the past decade, including the availability of balanced molecular mechanical force fields^{11,12} and efficient methods to handle long-range electrostatic interactions,^{13,14} has paved the way toward reliable nano- to microsecond simulations of NAs.^{15,16}

One of the leading force fields for the modeling of NAs has been the Cornell et al. ff94 force field¹¹ and its improved variants ff98¹⁷ and ff99,¹ which are available in the AMBER suite of programs.¹⁸ The ability of the AMBER force field to describe base stacking and base pairing accurately and consistently has been demonstrated by comparing the results with those of *ab initio* calculations.^{19,20} In spite of some known deficiencies, such as the visible under-twisting of the B-DNA double helix,^{6,21} short explicit solvent MD simulations using the AMBER force fields have been rather effectively applied to a variety of systems.^{3–10,22–32} The AMBER force fields have also been

* Corresponding author. E-mail: daniel.svozil@gmail.com.

[†] Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic.

[‡] Institute of Biophysics, Academy of Sciences of the Czech Republic.

[§] Parc Científic de Barcelona.

^{||} Departament de Bioquímica i Biologia Molecular, Universitat de Barcelona.

[⊥] Barcelona Supercomputing Center, Edifici Torre Girona.

[#] University of Utah.

[∇] Faculty of Pharmacy, University of Barcelona.

[○] Barcelona Supercomputer Center, Edifici Nexus II.

Figure 1. Atomic numbering and the definition of the torsion angles for deoxyribonucleotide. The nucleotide backbone is defined by the P–O5′–C5′–C4′–C3′–O3′ linkage. The torsion angles (in red) represent the rotation around the given angle defined by the four atoms along the backbone. It is customary to describe the backbone torsional angles of $\sim 60^\circ$ as *gauche*⁺ (*g*⁺), of $\sim 300^\circ$ as *gauche*[−] (*g*[−]), and of $\sim 180^\circ$ as *trans* (*t*).

successful in studies of RNA molecules, including those that exhibit a wide range of non-Watson–Crick base pairings, tertiary interactions, and complex backbone topologies.^{33–43}

These studies have also emphasized the importance of NAs' sugar–phosphate backbone flexibility, which is delimited by six main-chain torsion angles (designated as α through ζ), in addition to the five internal sugar torsions (denoted as τ_0 through τ_4 , which can be succinctly described by the puckering and the phase angle⁴⁴), and the glycosidic angle χ ⁴⁵ (see Figure 1). Free duplex B-DNA frequently populates two distinct backbone conformational substates, referred to as B_I and B_{II}. They are characterized by the following average torsion angles: $\alpha = 299^\circ$, $\beta = 179^\circ$ (B_{II}: $\beta = 143^\circ$), $\gamma = 48^\circ$, $\delta = 133^\circ$ (B_{II}: $\delta = 143^\circ$), $\epsilon = 182^\circ$ (B_{II}: $\epsilon = 251^\circ$), and $\zeta = 263^\circ$ (B_{II}: $\zeta = 168^\circ$). While the B_I/B_{II} conformational variability involving ϵ and ζ torsion angles was extensively studied both experimentally^{46–49} and theoretically,^{50–52} much less is known about the structural and energetic changes involving other torsion angles, such as the α and γ torsions.^{53–55} However, the variations in these angles were shown to be important in complexes of DNA with ligands and proteins.^{56–63}

In addition, an occurrence of irreversible α/γ backbone flips in B-DNA simulations longer than 10 ns^{21,64,65} was observed when AMBER ff94 or ff99 was used. Their accumulation leads to substantial distortion (unwinding) of the B-DNA double helix. Fortunately, this undesired effect does not occur in RNA simulations of a similar length, where related backbone flips are short-lived, reversible, and compatible with experimental backbone substates, which is interesting.⁶⁶ In addition to this distortion of the DNA helical structure, failures of the force field are responsible also for improper descriptions of the backbone topologies of single-stranded connecting loops in guanine quadruplex DNA molecules.⁶⁷

The DNA backbone conformation reflects a balance between the intrinsic (gas-phase) torsional energetics and other “external” forces, such as the influence of the environment and noncovalent interactions between different parts of the molecule. Steric considerations alone dictate the restriction of backbone angles to discrete ranges, with several backbone torsions being highly correlated.^{46,68} Moreover, additional restraints are imposed by Watson–Crick base pairing and by base stacking. A proper force-field description of the backbone is difficult because of its polarizable anionic nature, resulting in a conformational dependence of atomic charges. This complicates all the efforts to achieve fine-tuning of the force fields. A relevant contribution in this direction is the parmbsc0² force field, which has an improved backbone γ torsional profile, allowing for long time-scale simulations of B-DNA duplexes without any significant loss of helical structure.

Figure 2. The SPM (sugar–phosphate–methyl) and SPSOM (sugar–phosphate–sugar) model systems. The atom numbering follows the standard DNA numbering scheme. The α torsion angle is defined as O3′(*n* − 1)–P–O5′–C5′ and the γ torsion angle as O5′–C5′–C4′–C3′.

Since it is virtually impossible to dissect the individual torsional contributions from experimental data alone, quantum mechanical (QM) calculations provide a promising alternative means for deriving the intrinsic torsion properties and energetics of the DNA backbone. Though the early theoretical attempts using semiempirical QM calculations^{69–74} have provided valuable insights into the gross conformational properties of DNA constituents, their limited accuracy precluded the detailed quantitative investigation. However, recently it has become possible to analyze these conformational properties using high-level *ab initio* QM methods.^{55,75–85} In this paper, we attempt to gain insight into the complexity, modeling, and conformational flexibility of the DNA backbone. This is achieved by calculations on two model systems mimicking the basic building block of the DNA phosphodiester backbone. Various QM approaches, in concert with molecular mechanics calculations using the AMBER ff99 force field and its parmbsc0 variant, were utilized throughout the study.

Methods

Model Systems. The choice of an appropriate model system is a critically important step in any QM treatment of biologically relevant macromolecules. The ideal model should be as small as possible (thus making it possible to use high levels of theory) but still large enough to be able to include all relevant chemical moieties. Moreover, the investigations of large model systems in the gas phase lose relevance as their optimal structures, and the balance of forces, are typically far from what is observed in the complete systems within biologically relevant environments. This suggests that smaller model systems be used, specifically systems that retain the essential character of the larger biomolecule yet can be extensively conformationally sampled at high theoretical levels. Accordingly, two models have been considered (Figure 2): (i) a small system (denoted as SPM) consisting of sugar and methyl groups linked via a phosphodiester bond, and (ii) a larger system mimicking a dinucleotide step but without bases (referred to as SPSOM). The SPSOM model system contains two 2′-deoxyribose molecules linked via a phosphodiester bond. Additionally, to prevent unrealistic energies and bumps in the potential energy surface (PES) for certain conformations owing to spurious intramolecular hydrogen bonding, the terminal hydroxyl groups were methylated.

The exclusion of the bases from the SPSOM model avoids a huge conformational bias resulting from the stacking and the omission of the full context of the surrounding double helix. Moreover, it also facilitates theoretical study, because the treatment of the dispersion component of base stacking, whose theoretical description is very demanding,^{86,87} is not necessary. The inclusion of nucleobases would also amplify artifacts arising from the intramolecular basis set superposition error (BSSE; see *Ab Initio* Calculations section below) associated with base stacking. In addition, only one anionic phosphate was included to avoid a dominant contribution of the electrostatic repulsion between phosphates. Consequently, the SPSOM system, which has been tailored to be of minimal complexity, effectively retains all the features characteristic for the DNA building blocks, and it is currently the largest (or electronically most complete) meaningful model for an *ab initio* investigation of the sugar–phosphate segment of DNA.

Amber Potential Energy Surface. The PESs of the SPM and SPSOM models were first investigated using Amber⁸¹⁸ with the ff99 force field.¹ To obtain AMBER-compatible charges, the charges of both model systems were derived with the standard RESP fitting⁸⁸ on the HF/6-31G* electrostatic potential computed for a system optimized at the HF/6-31G* level of theory (Supporting Information, Figure S1).

A two-dimensional PES of SPM was constructed by varying the α and γ torsions independently in 30° increments within the range between 0° and 360°. In addition, the δ angle was fixed (or tightly restrained) to its value derived from fiber diffraction data ($\delta = 143^\circ$),⁸⁹ after which the rest of the system was optimized. The larger step of 30° was selected so as to make the amount of QM calculations computationally tractable. For the SPSOM model system, the α and γ torsion angles were varied independently in 5° increments in molecular mechanics optimizations with the ff99 force field. This resulted in a conformational map consisting of 5184 points. Minimization with the ff99 force field, with the α and γ torsions constrained, allowed all other coordinates to fully relax away from their respective canonical B-DNA values.^{57,90,91} [In Amber, constraints imply strictly fixed values, whereas restraints refer to the usage of an energy penalty function which pushes the given quantity toward a desired target value or region of values. It is impossible to maintain fixed torsion angles within Amber; however, as a high restraint of 3000 kcal·mol⁻¹ was used, the torsions will further be referred to as constrained. On the other hand, the fixed torsions in *ab initio* are true constraints.] Finally, the PES of the SPSOM system was also constructed using the parmbsc0 force-field modifications.²

parmbsc0 represents a reparameterization of the nonpolarizable ff99¹⁷ Amber force field for NA simulations. The new parametrization has been limited to improving the representation of the α/γ conformational space, which is poorly described in very long DNA simulations with current Amber force fields.^{21,64,65} Thus, only the parameters for α and γ torsions were changed compared to the original ff99 version, and no other effects, such as, e.g., the anomeric (*gauche*) effect,^{92–94} were explicitly targeted. The parametrization process was based on the B3LYP and LMP2 PESs of the SPM model system validated by CCSD(T)/CBS calculations. Fitted parameters have been tested by the very extensive set of simulations including a B-DNA all-atom simulation on the microsecond scale.¹⁶

The sampling of the conformational space has several limitations, the most serious of which is the fact that, when the α and γ angles are driven away from their canonical values (g^- for α and g^+ for γ), it often results in substantial changes

in other torsions that correspond to the global PES minima (for a given combination of the α and γ torsions). Without doubt, any given α/γ combination may yield multiple local minima on the PES, and, because of the presence of barriers, the optimization may be trapped in any of these. We have made considerable efforts to construct the α/γ map starting from various noncanonical structures in order to describe different parts of the α/γ space. Despite the fact that we cannot guarantee full relaxation at each point of the α/γ map, we suggest that the resulting map is of sufficient quality for our present purposes.

Ab Initio Calculations. When performing *ab initio* calculations it must be realized that the BSSE is present in all molecular electronics calculations in which orbitals are approximated by expansion in basis set function centered on nuclei that are dependent on the geometry of the studied system. BSSE typically arises in the calculation of intermolecular potential, in which case a supersystem and individual subsystems are described unequally. When the basis set is extended, the BSSE becomes smaller, and it vanishes when an infinite basis set (determined by the extrapolation to the complete basis set (CBS)) is used.⁹⁵ In the case of finite basis set, the intermolecular BSSE is commonly corrected by the counterpoise correction⁹⁶ method.

However, BSSE is present not only in systems with separate molecules, but it always occurs when comparing energies of two different isolated moieties. Such a BSSE, which is usually regarded as a part of the basis set incompleteness error, is called intramolecular BSSE.^{97–99} Hartree–Fock (HF) and density functional theory (DFT) methods yield smaller BSSEs (both inter and intramolecular), while correlated techniques (MP n or coupled cluster methods) are characterized by larger BSSE values.¹⁰⁰ Though not that widely studied as intermolecular BSSE, the magnitude of the intramolecular BSSE for some commonly used basis sets was reported.^{98,99} The influence of intramolecular BSSE on the geometries and energies was described, e.g., for intramolecular hydrogen bonding^{97,101} or for helicenes.^{102,103} Though several schemes for the intramolecular BSSE compensation were proposed,⁹⁷ the most physically justified method to eliminate intramolecular (as well as intermolecular) BSSE is to perform the calculations with large basis set,⁹⁸ ideally going up to the CBS limit.^{95,103}

The PES for the SPM model was explored at the B3LYP/6-31+G* level with a grid spacing of 30°. At each grid point, the model was fully optimized, but for α and γ (fixed at the grid values) and δ (fixed at a fiber diffraction value of 143°), it was started from the corresponding ff99 geometry (see above).⁸⁹ The minima identified on the B3LYP PES were subsequently reoptimized at the MP2/aug-cc-pVDZ level, but with the α and γ fixed at the values corresponding to the minima, and δ fixed at the fiber diffraction value. To eliminate the basis set incompleteness error, the minimum energy conformations were re-evaluated by means of MP2 calculations extrapolated to the complete basis set (CBS) limit.

However, since the intramolecular BSSE cannot be accounted for by the counterpoise procedure during optimization,⁹⁶ we were not able to obtain the BSSE-corrected optimized geometries. As a result, the CBS extrapolation eliminates only a part of the overall BSSE artifact, while the geometries remain affected by the BSSE, which may consequently also bias the energies. The MP2/CBS energies were estimated using the extrapolation scheme suggested by Helgaker et al.^{104,105} and Dunning's augmented correlation-consistent basis sets of double- and triple- ζ quality (aug-cc-pVDZ and aug-cc-pVTZ; eq 1):¹⁰⁶

$$E_{\text{CBS}}^{\text{MP2}} = E_{\text{aug-cc-pVDZ}}^{\text{MP2}} + \frac{(E_{\text{aug-cc-pVTZ}}^{\text{MP2}} - E_{\text{aug-cc-pVDZ}}^{\text{MP2}})}{0.703704} \quad (1)$$

Because of the high computational demands of MP2 calculations with the employed basis sets, the resolution of identity MP2 (RI-MP2) method^{107,108} was used throughout the work. In the RI-MP2 method, two-electron, four-center integrals are replaced by linear combinations of two-electron, three-center integrals via the introduction of an auxiliary fitting basis set,^{107–110} which results in an acceleration of around 1 order of magnitude as compared to the standard MP2 calculations.^{109,111,112} Moreover, when the auxiliary fitting basis is selected carefully, the molecular structures and relative energies computed at the MP2 and RI-MP2 levels are virtually identical.^{109,112,113}

The smaller size of the SPM system has also made it possible to evaluate higher energy contributions beyond the MP2 treatment. Assuming that the difference between the CCSD(T) and MP2 energies (denoted as $\Delta\text{CCSD(T)}$) exhibits only a small basis set dependence,^{114,115} the CCSD(T) energies at the complete basis set level (the CBS(T) method) were estimated as

$$E_{\text{CBS}}^{\text{CCSD(T)}} = E_{\text{CBS}}^{\text{MP2}} + (E_{6-31+G^*}^{\text{CCSD(T)}} - E_{6-31+G^*}^{\text{MP2}}) \quad (2)$$

For the SPSOM model system, computational demands currently preclude the full potential energy scan at the *ab initio* levels of theory, which has led to the adoption of a simpler method. On the basis of the shape of the ff99 PES generated, several (typically 3–5) conformations surrounding the individual minima identified on the ff99 PES (with varying α and γ angles) were subjected to a constrained *ab initio* minimization. Assuming a parabolic shape of the PES in the vicinity of the local minima, a quadratic fit was performed on the calculated energies, and the α/γ angles corresponding to the local minima were identified. The local minima were found on the PESs at the B3LYP/6-31+G* and MP2/6-31+G* levels. Approximately 60 points were investigated on each of these PESs. The MP2/CBS energies were subsequently determined using the MP2/6-31+G* geometries. The electronic structure of the SPSOM system was described by the natural bond orbital (NBO) analysis^{116–118} of the electronic density performed with Gaussian 03.¹¹⁹

As solvent effects might introduce significant changes in the torsional profile, the influence of hydration was investigated using the MST continuum model.¹²⁰ The calculations were performed by using the recently parametrized semiempirical RM1 version of the MST model.¹²¹ The conformational preferences in solution were estimated by combining the energy difference in the gas phase and the hydration free energy determined from the MST calculations.

The MST results were compared to the empirical model for calculating the electrostatic component of the solvation free energy based upon the Poisson–Boltzmann (PB) equation. A grid with a grid step of 0.5 Å/grid (a solute occupancy of 10%) was defined, and, for comparison, another computation was performed with a reduced grid step of 0.25 Å/grid (occupancy of the grid box of 25%). PB computations were performed using a permittivity of 78.5 for the solvent, and 1.0 for the interior of the solute.

The RI-MP2 and DFT calculations were performed employing the Turbomole 5.8 package.¹²² The CCSD(T) correction was computed by MOLPRO 2006.1.¹²³ The NBO analysis was performed with the NBO 3.0 program¹¹⁸ implemented in Gaussian 03.¹¹⁹ All empirical force-field calculations were carried out using the Amber 8.0 suite of programs.¹⁸ The MST computations were performed using a locally modified version

Figure 3. The potential energy map for the α and γ torsions in the SPM model system (the energy values are given in kcal·mol^{−1} and angles in degrees) determined from B3LYP/6-31+G* calculations.

of MOPAC 6.0.¹²⁴ The PB calculations were performed with the software package Delphi v. 4, release 1.1.^{125,126}

Results and Discussion

SPM vs SPSOM Model Systems. The SPM model was used to parametrize the parmbsc0 force field,² and our protocol by which the SPS B3LYP/6-31+G* PES was generated followed the procedure adopted in the parmbsc0 parametrization. Therefore, when comparing the SPM model with the extended SPSOM model, which should provide a more realistic electronic structure description of the sugar–phosphate segment, the following points need to be taken into account:

(1) The SPM PES was explored at the B3LYP/6-31+G* level using a 30° grid spacing. While such a grid spacing is necessary to keep the computational requirements within reasonable limits, some of the minima (such as, e.g., g^-/t) may be easily missed.

(2) Whereas for SPSOM only the α and γ torsions were fixed, for SPM δ was also fixed at the fiber diffraction value ($\delta = 143^\circ$),⁸⁹ which is different from the value found in high-resolution crystal structures ($\delta = 133^\circ$). The elimination of this degree of freedom may therefore alter the PES.

(3) While the MP2/6-31+G* minima for SPSOM are true stationary points on the MP2/6-31+G* PES, the MP2/aug-cc-pVDZ “minima” of the SPM system were determined by fixing α , γ , and δ at their B3LYP/6-31+G* minima values and reoptimizing the rest of the system at the MP2/aug-cc-pVDZ level of theory. This procedure may also affect the PES, because the positions of the B3LYP and MP2 minima may differ significantly.

The SPM Model System. The full α/γ potential energy map of the SPM model determined at the B3LYP/6-31+G* level is shown in Figure 3, where four broad minima are perceptible: (1) the canonical g^-/g^+ ($\alpha \sim 270^\circ$, $\gamma \sim 60^\circ$); (2) the g^-/g^- ($\alpha \sim 270^\circ$, $\gamma \sim 300^\circ$); (3) the g^+/g^- ($\alpha \sim 90^\circ$, $\gamma \sim 300^\circ$); and (4) the g^+/g^+ ($\alpha \sim 60^\circ$, $\gamma \sim 60^\circ$) minima. The geometries of the four minima were reoptimized without restrictions (other than the fixed α , γ , and δ torsions) using the MP2/aug-cc-pVDZ method. Single-point calculations at higher levels of theory (MP2/CBS//MP2/aug-cc-pVDZ and MP2/CBS+ $\Delta\text{CCSD(T)}$ //MP2/aug-cc-pVDZ) were then performed. [The // symbol separates the energy-evaluation (left) and geometry-optimization (right) methods.]

TABLE 1: The Relative Energies^a of the Key Minima of the SPM Model System Computed at Various Levels of Theory^b

geometry	energy	g^-/g^-	g^+/g^-	g^+/g^+	g^+/t
B3LYP/6-31+G*	B3LYP/6-31+G*	0.9	0.8	2.4	6.8
ff99	ff99	2.5	3.5	1.9	4.3
MP2/aug-cc-pVDZ	MP2 CBS	2.0	2.4	2.7	
MP2/aug-cc-pVDZ	MP2 CBS+	2.1	2.4	2.8	
	ΔCCSD(T)				

^a In kcal·mol⁻¹, relative to the canonical g^-/g^+ conformation.

^b The pathological g^+/t conformation is not a minimum, and the δ -constrained MP2/aug-cc-pVDZ optimization drives the geometry out of the region.

Figure 4. Depiction of the distribution of the α and γ angles in X-ray crystal structures of DNA and its complexes with ligands and proteins. The data set consists of 447 structures, containing 7750 dinucleotide building blocks at a resolution equal to or better than 1.9 Å (the structures are from the NDB database from July 19, 2005).

The results in Table 1 indicate the intrinsic conformational preference for the canonical g^-/g^+ conformation at the best level of theory (the MP2/CBS+ΔCCSD(T)/MP2/aug-cc-pVDZ), with the g^-/g^- , g^+/g^- , and g^+/g^+ conformations being disfavored by 2.1–2.8 kcal·mol⁻¹. Second, the preference for the g^-/g^+ conformation is also reflected in the ff99 and B3LYP results. However, the relative energies of g^-/g^- , g^+/g^- , and g^+/g^+ conformations are significantly different from those of the MP2/CBS values (Table 1). Finally, the results also demonstrate that the CCSD(T) correction is negligible and can be safely omitted in calculations for the larger, SPSOM system.

The SPSOM Model System: The Amber ff99 and Ab Initio PESs. Several energy minima were located on the Amber ff99 PES of the SPSOM model, including, besides the canonical g^-/g^+ region, the noncanonical g^+/g^+ , g^+/t , g^+/g^- , t/t , t/g^- , t/g^+ , g^-/t , and g^-/g^- conformations. As the t/g^- , t/g^+ , and g^-/g^- conformations have no counterparts in any X-ray (Figure 4) or MD structures, only the canonical g^-/g^+ conformation and the noncanonical g^+/g^+ , g^+/t , t/t , g^-/t and g^+/g^- regions were further investigated. The g^+/t and g^-/t conformations are the problematic substates that tend to accumulate in longer B-DNA simulations with Amber force fields.^{21,64,65}

As explained in the Methods section, because of the complexity of the PES, the lowest-lying minima cannot be guaranteed to be found for all the investigated α/γ regions. The structures were generally obtained by starting from the canonical geometry, subsequently setting up the target α/γ combination while keeping other angles constant (at their canonical values) and finally optimizing the structures with fixed α and γ . We

TABLE 2: The Values of the α/γ Torsions (Degrees) of the Key Minima of the SPSOM Model System Obtained at Various Levels of Theory^a

α/γ	ff99	parmbsc0	B3LYP/6-31+G*	MP2/6-31+G*
g^-/g^+	290/45	285/50	290/43	300/40
g^+/g^-	70/300	70/300	83/280	70/275
t/t	170/195	175/180	174/184	144/180
g^-/t	295/205	295/205	262/196	279/196
g^+/t	75/190	75/175	80/180	75/160
g^+/g^+	35/35	35/35	80/32	80/38

^a g^-/g^+ is the canonical conformation, and g^+/t is the most problematic substate accumulating in long B-DNA simulations with ff99.

nevertheless assume that the PES scan is sufficient for our purposes, because we are primarily interested in the relative comparison between the Amber ff99 and *ab initio* data. Accordingly, the minima on the B3LYP/6-31+G* and MP2/6-31+G* PESs were identified (see Methods) and refined by MP2/CBS//MP2/6-31+G* computations.

In spite of the relatively small size of the SPSOM model, it is not easy to characterize it unambiguously. One of the problems that had not been anticipated was the occurrence of weak C–H···O hydrogen bonds (HBs) in several local minima (Supporting Information, Table S1). Whereas for noncanonical conformers, it is not clear if their presence reflects reality, these HBs are certainly not present in the canonical structures of experimental DNA structures. However, one C–H···O HB was identified in the AMBER ff99 canonical g^-/g^+ conformations, and MP2/6-31+G* yielded even two such HBs. On the other hand, no such HB was identified at the B3LYP/6-31+G* level of theory (Supporting Information, Table S1). The g^-/g^+ minimum thus adopts a visually closed conformation at the MP2/6-31+G* level, while its geometry at the B3LYP/6-31+G* level is more stretched (Supporting Information, Figure S2).

Analogously, three CH···O HBs have been found at the ff99 level for the g^+/g^+ conformation, but only one and two HBs were found at the B3LYP/6-31+G* and MP2/6-31+G* levels, respectively (Supporting Information, Table S1). Apparently, the presence of these HBs depends on the theoretical level of the system's description. Such differences can be caused by several factors. By means of an NBO analysis, it was established that, because of the orbital interaction, the H3'–C3' antibonding orbital within the HB between the O3' of the 3'-sugar and the H3'–C3' of the 5'-sugar is substantially more populated in the g^+/g^+ conformation (0.0490 e) than in the canonical g^-/g^+ (0.0415 e) one, which indicates that the DFT and MP2 procedures differ in their description of the electronic structure, with the MP2 method including the dynamical correlation effects more completely. The occurrence of these HBs may also be, at least partially, attributed to not correcting for the intramolecular BSSE in the geometry optimization, since intramolecular BSSE obviously favors compacted structures. The BSSE error is generally smaller for DFT-based methods but is large for *ab initio* (MP2) methods.¹⁰⁰ However, it does not explain, e.g., the occurrence of one such bond on the ff99 PES of g^-/g^+ conformation.

The presence of these HBs clearly reflects a notable dependence of the geometrical features on the level of theory, which complicates any subsequent comparison of the corresponding minima. The optimal α and γ values determined at different levels of theory (ff99, B3LYP/6-31+G*, and MP2/6-31+G*) for the given substates or regions may hence substantially differ (Table 2). For the canonical g^-/g^+ minima, similar α/γ values

TABLE 3: The Relative Energies^a of the Key Minima of the SPSOM Model System Obtained at Various Levels of Theory^b

α/γ	ff99//ff99	parmbsc0// parmbsc0	B3LYP/6-31+G*// B3LYP/6-31+G*	MP2/6-31+G*// MP2/6-31+G*	MP2/CBS// MP2/6-31+G*	MP2/CBS// ff99	MP2/CBS//MP2/6-31+G* solution +
g^-/g^+	0.0	0.0	0.0	0.0	0.0	0.0	0.0
g^+/g^-	0.3	-1.3	0.7	0.9	0.1	-0.1	3.2
t/t	6.2	8.2	5.9	4.9	3.7	7.0	1.6
g^-/t	-1.6	-1.9	0.8	-1.5	-2.3	-1.2	2.6
g^+/t	1.4	2.7	4.2	3.4	2.9	3.0	0.4
g^+/g^+	5.2	5.6	2.3	1.5	0.3	4.6	2.1

^a In kcal·mol⁻¹, pertaining to the canonical g^-/g^+ conformation. ^b g^+/t (bold) is the most problematic substate accumulating in long B-DNA simulations with ff99.

have been found, irrespective of the level of theory. On the other hand, notable differences have been observed in the α angle of the g^+/g^+ conformation. As this region is scarcely populated in X-ray crystal structures (Figure 4), this conformation will not be further investigated. Other visible differences between the ff99 and B3LYP minima have been found for γ in the g^+/g^- conformation (300° vs 280°) and for α in the g^-/t conformation (295° vs 262°). The differences between the ff99 and MP2 minima are not negligible, namely, for γ in the g^+/g^- (300° vs 275°) and g^+/t (190° vs 160°) conformations, and for α in the t/t (170° vs 144°) and g^-/t (295° vs 279°) conformations (Table 2). Apart from α/γ torsions, the minima may also substantially differ in other torsional angles. For example, MP2/6-31+G* and ff99 minima of the g^+/g^+ conformer differ also in $\delta - 1$ (145° vs 125°), β (232° vs 66°), and δ (143° vs 100°) torsions. Similarly, MP2/6-31+G* and ff99 minima of the t/t conformation differ in $\epsilon - 1$ (276° vs 192°) and β (208° vs 176°) torsions. The differences in torsional angles between individual minima obtained at different levels of theory may be attributed mainly to the presence of CH···O HBs. These bonds belong between “weak hydrogen bonds”,¹²⁷ and their effect on the energies relative to canonical g^-/g^+ minima mostly cancels out. We estimate that the variations in relative energies due to the presence of CH···O HBs is on the order of 1.0 kcal·mol⁻¹. No other correlations, as, e.g., the possible correlation between the RO–P–OR valence angle and α and $\xi - 1$ torsions,¹²⁸ were studied in the present work.

The impact of the geometrical differences on the relative energies of the local minima can be evaluated from a comparison of the MP2/CBS//MP2/6-31+G* and MP2/CBS/ff99 results (see Table 3). Even though the relative stabilities of the g^+/g^- and g^+/t conformations computed at these levels of theory are in close agreement, large differences occur in the relative energies of the g^-/t , t/t, and g^+/g^+ conformers, which amount to 1.1, 3.3, and 4.3 kcal·mol⁻¹, respectively. These findings, therefore, indicate the need to maintain a close correspondence between the molecular geometries in the energetic analysis of the local minima.

The energy ordering relative to the canonical conformation is the same at the MP2/6-31+G* and MP2/CBS//MP2/6-31+G* levels (Table 3). A CBS extrapolation leads to a lowering of relative energies by 0.5–1.3 kcal·mol⁻¹. The orderings of the local minima conformations on the MP2/6-31+G*, MP2/CBS//MP2/6-31+G*, and Amber ff99 surfaces mutually agree, with the only exception being the g^+/t and g^+/g^+ conformations. On the other hand, more significant differences appear between the B3LYP and MP2/CBS//MP2/6-31+G* surfaces, where the only similarity is that the t/t conformation is consistently described as the least favorable one. The results are thus clearly method-dependent, and there are some differences, the origins of which are not completely clear, even between the DFT and MP2 methods.

An examination of the MP2/CBS//MP2/6-31+G* energies and ff99 data reveals that the Amber ff99 force field overstabilizes the pathological g^+/t substate by 1.5 kcal·mol⁻¹, whereas the g^-/t one is destabilized by 0.7 kcal·mol⁻¹ when compared to the MP2/CBS//MP2/6-31+G* results. On the other hand, the t/t region lying on the path between the canonical g^-/g^+ conformation and the flipped g^+/t and g^-/t conformations⁵⁴ (Figure 4) is 2.5 kcal·mol⁻¹ more stable at the MP2/CBS//MP2/6-31+G* level than at the ff99 one. We have further also recomputed the energies using the ff99 geometries. The differences in relative energies between the ff99 and MP2/CBS//ff99 energies are marginal (smaller than 0.9 kcal·mol⁻¹) except for the g^+/t conformer, which is overstabilized by 1.6 kcal·mol⁻¹ at the ff99 level (such an overstabilization almost exactly mimics that found between ff99 and MP2/CBS//MP2/6-31+G*; see above).

The sensitivity of energies on charges was also tested by generating RESP charges for each of ff99 energy minima conformation (see Supporting Information). The relative energies of individual minima were then calculated using either charges for individual conformers, or charges averaged over all conformations. Though there exists a sensitivity to the charge model, the trends are consistent with either the charges chosen or a set of RESP charges averaged over all conformations. The relative ranking of the energies is maintained, and the g^+/t conformer comes out too stable compared to QM in most charge models.

Overall, the results demonstrate that the nature of the molecular geometries and the level of theory greatly influence the energetics of the conformational landscape of the SPSOM model. This makes any quantitative analysis very complicated. Nevertheless, all calculations consistently suggest that g^+/t is intrinsically overstabilized by the ff99, which could be the principal cause of the structural degradation observed in long ff99 B-DNA simulations.

The SPSOM Model System: The parmbsc0 PES. The parmbsc0² PES was also investigated by locating the same set of minima as for the ff99 PES. Whereas the geometries of the individual minima were almost the same as the ff99 ones (Table 2), the relative energies reflect clear differences between the parmbsc0 and ff99 results (Table 3). While the g^-/t and g^+/g^+ minima remained nearly unaltered, the parmbsc0² force field destabilized the g^+/t , g^+/g^- , and t/t conformations by 1.3–2.0 kcal·mol⁻¹. It is worth mentioning that the parmbsc0 relative energies of the problematic g^+/t and g^-/t conformers agree well with the MP2/CBS//MP2/6-31+G* values, whereas the relative energies of the g^+/g^- , t/t, and g^+/g^+ conformations are further destabilized by 1.4, 4.5, and 5.3 kcal·mol⁻¹ at the parmbsc0 level, respectively. These data show that parmbsc0 achieves better agreement with the QM reference data for the problematic g^+/t and g^-/t conformations than ff99. The performance of force fields can evidently be rationalized based on intrinsic gas-phase

TABLE 4: The Solvation Free Energies (kcal·mol⁻¹) of the Key MP2/6-31+G* Minima of the SPSOM Model System^a

	electrostatic MST	electrostatic PB (0.25 Å/grid)	electrostatic PB (0.5 Å/grid)	cavitation	vdW	total MST
g ⁻ /g ⁺	-81.5	-74.9	-75.2	51.4	-39.4	-69.5
g ⁺ /g ⁻	-78.4	-70.8	-70.9	51.6	-39.6	-66.4
t/t	-83.7	-75.0	-75.4	51.7	-39.6	-71.6
g ⁻ /t	-76.6	-68.5	-68.6	51.2	-39.2	-64.6
g ⁺ /t	-83.9	-74.6	-74.4	51.0	-39.1	-72.0
g ⁺ /g ⁺	-79.6	-71.4	-71.6	51.3	-39.4	-67.7

^a The electrostatic component of the solvation free energy was calculated using either the MST continuum model,¹²⁰ or the empirical procedure based upon the PB equation (utilizing two different grid spacings).

properties of the backbone of NAs. Furthermore, the results also indicate that g⁺/t is the decisive substate for tuning the force field.

The SPSOM: The Influence of Hydration. Solvation can play a key role in determining the conformational space of largely charged systems. In order to estimate how hydration can affect the conformational preferences of the SPSOM model, MST computations were performed to determine the hydration free energy of the different conformations (Table 4). The relative hydration free energies of different conformers were determined from gas phase geometries and added to the gas phase energy differences computed at the MP2/CBS//MP2/6-31+G* level to estimate the relative stability between the conformers in an aqueous solution. For the sake of comparison, the electrostatic component of the hydration free energy was also determined from PB computations using RESP charges derived for each MP2/6-31+G* conformation of the SPSOM model.

The results in Table 4 show that there exists a close similarity between the PB results determined using grid steps of 0.25 and 0.5 Å/grid. The largest difference between these two grids amounts only to 0.4 kcal·mol⁻¹ (t/t conformer), which implies a relative error of 0.5%. There exists an agreement between the PB and MST results as far as the solvent-induced destabilization of the g⁺/g⁻, g⁻/t, and g⁺/g⁺ conformers (relative to the g⁻/g⁺ conformer) is concerned. However, PB results fail in predicting the differential stabilization arising upon hydration for t/t and g⁺/t conformations.

The results in Table 3 and in Table 4 (the total MST results) demonstrate that hydration has a dramatic influence on the relative stability of the conformations. The g⁻/t conformer, which is the most stable in the gas phase, has a negligible impact in the conformational space in water. Similarly, the degeneracy between the g⁻/g⁺, g⁺/g⁻ and g⁺/g⁺ states found in the gas phase is broken in solution, where the g⁻/g⁺ conformation is preferred by a factor of 40–100. On the other hand, water greatly stabilizes the problematic g⁺/t conformer, which, according to our best estimates in water, is only 0.4 kcal·mol⁻¹ higher in energy than the canonical g⁻/g⁺ form. Caution is required as hydration effects can be different in a DNA environment that in a pure solvent, but it is clear that the solvent effects are as important as the intrinsic (gas-phase) preferences when determining the α/γ conformational space. It is noteworthy that the results indicate that hydration stabilizes the t/t conformer (i.e., the intermediate for a g⁻/g⁺ $\rightarrow$ g⁺/t transition) by 2.1 kcal·mol⁻¹, which in turn facilitates the understanding of the nature of the artifactual α/γ transitions found in ff99 MD simulations.

Overall, the results demonstrate the complexity of the fitting procedure in the force-field development. The fitting of force-

field parameters can not always be equally accurate in all the regions of the conformational space. Therefore, the developer needs to decide which regions are low-populated, and can thus be considered with lower accuracy during the fitting process. Accordingly, inspection of the gas phase PES could suggest that not much attention is needed to be paid to the g⁺/t region in parametrization. However, solvation changes the energy map, and the g⁺/t conformer becomes important, making mistakes in its fitting a potential major source of errors in simulations.

The B3LYP and MP2 Descriptions of the Phosphate Group. The relative energies obtained at the B3LYP/6-31+G*//B3LYP/6-31+G* and MP2/CBS//MP2/6-31+G* levels reveal notable differences in several cases, such as the g⁻/t, g⁺/t, g⁺/g⁺, and t/t conformations (Table 3). So as to elucidate the origin of those differences, we performed an NBO analysis, which makes it possible to assess the hyperconjugation effects in complex systems.¹²⁹ It was ascertained that MP2 calculations assign ~ 1.8 e more to non-Lewis (i.e., delocalized) orbitals than B3LYP calculations (B3LYP and MP2 render 171.9 and 170.1 e to the classical covalent bonds or core orbitals; for reference, the corresponding HF value is 172.3 e). The B3LYP electron density is closer to that obtained without the inclusion of the electron correlation. Intuitively, for an accurate description of multicentered bonds with their larger spatial dimensions, a method accounting for higher correlation components (e.g., MP2) should be preferred over the DFT approach. We consider this as a very important qualitative result, illustrating the actual complexity of the electronic structure of the SPSOM system. The analysis is, however, complicated by additional differences. The DFT does not include the dispersion energy, while MP2 data are biased by a rather large intermolecular BSSE, which is considerably reduced at the DFT level. The final differences in the geometries and relative energies reflect a complex mixture of all the above-mentioned effects. Analyzing the complex backbone segment is definitely much less straightforward than in studies of base stacking and base pairing.

The Differences in the Electronic Description of the SPM and SPSOM Models. The effect of replacing the 3'-sugar unit with a methyl group in the SPM model was also investigated by an NBO analysis. A great advantage of the NBO analysis is that it projects the multicentered orbitals into localized bond orbitals, which are more compatible with the chemical view.^{116–118} In addition, it provides a quantitative estimate of the delocalization effects between localized orbitals.¹³⁰ The anomeric effects in the dimethyl-phosphate anion have been successfully explained by employing this approach.^{131,132} The NBO analysis was performed for the HF orbitals using MP2/6-31+G*-optimized geometries.

The NBO analysis was applied to compare the electronic structure of the MP2/6-31+G*-optimized SPM and SPSOM models in three distinct conformations: g⁻/g⁺, g⁺/g⁺, and g⁺/t. The main difference between the two sets of data was found in the orbital interactions between the O3' lone electron pairs and the adjacent C–H and C–C bonds of the 3'-sugar (The off-diagonal Fock matrix elements characterizing these orbital interactions are listed in Supporting Information Table S2). The sum of the Fock-matrix elements corresponding to these orbital interactions was almost the same for both models (ranging from 0.302 to 0.306 for the SPM model and from 0.286 to 0.315 for the SPSOM models). On the other hand, the back-donation from the nonbonding orbitals of the 3'-sugar to O3' in the two conformers (g⁺/g⁺ and g⁺/t) was notably weaker (the sums of the corresponding Fock matrix elements were 0.127 and 0.098 for the g⁺/g⁺ and g⁺/t conformations, respectively) than that

from the methyl group (with the Fock-matrix elements being 0.164 for the g^+/g^+ conformation and 0.124 for the g^+/t conformation). In the g^+/g^- conformation, the back-donation was the same for both models. The electron donation from the C3'–H3' bond to the C4'–H4' antibond (with the corresponding Fock-matrix elements being 0.024 and 0.025 for the g^+/t and g^+/g^+ conformations, respectively) was observed in the g^+/g^+ and g^+/t conformations. Thus in the g^+/g^+ and g^+/t SPSOM conformers, a part of the electron density of the C3'–H3' bonds was transferred to the 3'-sugar rather than to the O3' atom.

The changes in the hyperconjugation network along the sugar–phosphate backbone arising from the omission of the nucleobases in the SPSOM model were also tested. In the g^-/g^+ conformer, the C1'–H bonds of both sugars were replaced with the C1'–N=CH₂ moiety, and the system was optimized at the MP2/6-31+G* level of theory with the α and γ torsion angles fixed (shown in Supporting Information Figure S3). As expected, the addition of the –N=CH₂ unit further extended the hyperconjugation network of both sugars (the main components of the orbital interactions are listed in Supporting Information Table S3) and altered their hyperconjugation network along the C3'–O3'–P–O5'–C5' linkage. In Supporting Information Table S3, the Fock-matrix elements corresponding to the orbital interactions between the O3' lone pairs and the adjacent sugar are also listed. The addition of the –N=CH₂ moiety does not affect the donor–acceptor orbital interactions between the lone pair of O3' and the C2'–C3', C3'–C4', or C3'–H3' antibonding orbitals; the sum of the corresponding Fock-matrix elements is 0.29 for both the C1'-substituted and nonsubstituted forms. In contrast, the back-donation from the C2'–C3', C3'–C4', and C3'–H3' bonds to the Rydberg orbitals of O3' markedly increases in the substituted SPSOM model with respect to its nonsubstituted counterpart (the sum of the corresponding Fock-matrix elements is 0.184 and 0.126 for the models with and without –N=CH₂ groups, respectively; see Supporting Information Tables S2 and S3).

Conclusions

We have examined the intrinsic conformational preferences of two model systems representing the DNA sugar–phosphate backbone by means of QM and force-field calculations. The PESs of the SPM and SPSOM models show surprisingly large differences, which may be attributed not only to the different protocols used to yield them and to the complexity of torsional space but also to the different electronic structures of the two models. At this point, the SPSOM model system appears to reflect the torsional angle variability in NAs more reliably and to provide a better description of the electronic structure of the DNA backbone (including effects such as hyperconjugation).

However, despite its more complete electronic structure, the investigation of the SPSOM system is far from being straightforward. The PES is very complex, and the description depends on the theoretical level of calculations. Our results bring a general cautionary message. Because of continuing hardware and methodological advances, larger and larger fragments can be subjected to QM analyses. However, an increase in the number of atoms in the QM-gas-phase system does not mean that the calculations are gaining relevance and becoming closer to biology. In most cases, the opposite will be the case, and a simple increase in the size of the gas-phase system (dinucleotides, trinucleotides, etc., or dipeptides, tripeptides, etc.) without imposing suitably selected conformational constraints and considering solvent screening will drastically increase the magnitude of model-system incompleteness errors, which will

yield structures that would never occur in realistic biological systems. The SPSOM system is on that edge.

An investigation of the SPSOM PES reveals the problematic nature of the ff99 description of conformations with the γ torsion in the *trans* region as compared to the high-level MP2/CBS//MP2/6-31+G* results. Whereas the ff99 energy (relative to the canonical geometry) of the g^+/t conformation is lower by 1.5 kcal·mol^{−1} than in the case of the MP2/CBS//MP2/6-31+G* PES, the ff99 g^-/t relative energy is higher by 0.7 kcal·mol^{−1}. Also the t/t conformation is more stable (by 2.5 kcal·mol^{−1}) at the MP2/CBS//MP2/6-31+G* PES. Similarly, the ff99 relative energy of the g^+/t conformation is lower by 1.6 kcal·mol^{−1} when compared to its value calculated at the MP2/CBS level of theory using the ff99 geometry. This may partially explain the pathological behavior of the AMBER ff99 (and also the ff94 and ff98) force fields in longer simulations of B-DNA.^{21,64,65}

In parmbsc0², however, the t/t region is elevated relative to the ff99 PES by 2.0 kcal·mol^{−1}, and the MP2/CBS//MP2/6-31+G* PES is elevated by 4.5 kcal·mol^{−1}, which may impede the transition from a canonical g^-/g^+ conformation to a flipped g^+/t or g^-/t conformation. A comparison of the parmbsc0 relative energies and MP2/CBS relative energies evaluated for the g^+/t , g^-/t , and t/t conformations at the ff99 geometries reveals that they are actually very similar, the largest difference being the destabilization of the t/t minima by 1.2 kcal·mol^{−1} in parmbsc0 when compared to its MP2/CBS//ff99 equivalent. These facts suggest that parmbsc0 eliminates the α/γ flips by increasing the energetic difference between the g^-/g^+ and both g^+/t and g^-/t conformations, respectively, when approaching high-level QM estimates.

The hydration of the SPSOM model system also considerably affects the shape of the PES (Table 3). Both the flipped g^+/t and transitional t/t conformations are significantly stabilized upon hydration. Although these results cannot be extrapolated directly to the DNA helix environment, the computations indicate that hydration should favor the transition from the canonical g^-/g^+ conformational state to the g^+/t one. As a result, the force field must pose a substantial intrinsic penalty for the gamma *trans* regions, which is achieved by parmbsc0 but not by parm99.

An operation cannot be explained or the quality of any force field judged merely on the basis of the investigation of a few energetic (e.g., torsional) profiles using the gas-phase PES. Force-field validity can ultimately be demonstrated only through its thorough testing using the MD simulations of various relevant systems. parmbsc0 has already proven its validity in this respect² and is the only choice available for DNA simulations. It performs equally as well as parm99 for RNA, where the α/γ flip problem does not occur, as has recently been demonstrated in our numerous tests (Sponer, J., unpublished results, 2008). However, our QM calculations indicate that there might still be room for improvement in parmbsc0, because some unusual α/γ conformations are not yet well represented in the gas phase. Any further modification will, however, be a formidable task.

Acknowledgment. This work was supported by the Ministry of Education of the Czech Republic (Grant Nos. LC512, MSM0021622413, AVOZ40550506, AVOZ50040507, and AVOZ50040702), by the Grant Agency of the Academy of Sciences of the Czech Republic (Grant Nos. IAA400040802, IAA400550701, and IQS500040581) and by the Spanish Ministry of Science and Technology (Grant Nos. CTQ2005-08797 and BIO2006-01602). T.E.C. would like to acknowledge support from the NIH R01-GM079383-01A1, computer time from an NSF LRAC MCA01S027P award, and also

computer resources from the University of Utah Center for High Performance Computing (including those provided by the NIH 1S10RR17214-01 on the Arches metacluster).

Supporting Information Available: The characteristics of the CH $\cdots$ O HBs, solvation free energies, NBO analysis results, SPSOM model-system charges, and geometries (xyz coordinates) of all the investigated conformations on the ff99, parmbsc0, B3LYP/6-31+G*, and MP2/6-31+G* PESs. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (2) Perez, A.; Marchan, I.; Svozil, D.; Spomer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.
- (3) Beveridge, D. L.; McConnell, K. J. *Curr. Opin. Struct. Biol.* **2000**, *10*, 182–196.
- (4) Giudice, E.; Lavery, R. *Acc. Chem. Res.* **2002**, *35*, 350–357.
- (5) Norberg, J.; Nilsson, L. *Acc. Chem. Res.* **2002**, *35*, 465–472.
- (6) Cheatham, T. E. *Curr. Opin. Struct. Biol.* **2004**, *14*, 360–367.
- (7) Auffinger, P.; Westhof, E. *Curr. Opin. Struct. Biol.* **1998**, *8*, 227–236.
- (8) Orozco, M.; Perez, A.; Noy, A.; Luque, F. J. *Chem. Soc. Rev.* **2003**, *32*, 350–364.
- (9) Auffinger, P.; Hashem, Y. *Curr. Opin. Struct. Biol.* **2007**, *17*, 325–333.
- (10) McDowell, S. E.; Spackova, N.; Spomer, J.; Walter, N. G. *Biopolymers* **2007**, *85*, 169–184.
- (11) 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*, 5179–5197.
- (12) MacKerell, A. D.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946–11975.
- (13) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (14) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (15) Spomer, J.; Lankas, F. *Computational Studies of RNA and DNA*; Kluwer Academic Publishers Group: Norwell, MA, 2006.
- (16) Perez, A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2007**, *129*, 14739–14745.
- (17) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845–862.
- (18) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (19) Spomer, J.; Jurecka, P.; Marchan, I.; Luque, F. J.; Orozco, M.; Hobza, P. *Chem.—Eur. J.* **2006**, *12*, 2854–2865.
- (20) Spomer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151.
- (21) Beveridge, D. L.; Barreiro, G.; Byun, K. S.; Case, D. A.; Cheatham, T. E.; Dixit, S. B.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Seibert, E.; Sklenar, H.; Stoll, G.; Thayer, K. M.; Varnai, P.; Young, M. A. *Biophys. J.* **2004**, *87*, 3799–3813.
- (22) Spackova, N.; Cheatham, T. E.; Ryjacek, F.; Lankas, F.; van Meervelt, L.; Hobza, P.; Spomer, J. *J. Am. Chem. Soc.* **2003**, *125*, 1759–1769.
- (23) Spackova, N.; Berger, I.; Spomer, J. *J. Am. Chem. Soc.* **2001**, *123*, 3295–3307.
- (24) Spackova, N.; Berger, I.; Egli, M.; Spomer, J. *J. Am. Chem. Soc.* **1998**, *120*, 6147–6151.
- (25) Sherer, E. C.; Harris, S. A.; Soliva, R.; Orozco, H.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121*, 5981–5991.
- (26) Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **2001**, *123*, 12658–12663.
- (27) Lankas, F.; Spomer, J.; Hobza, P.; Langowski, J. *J. Mol. Biol.* **2000**, *299*, 695–709.
- (28) McConnell, K. J.; Beveridge, D. L. *J. Mol. Biol.* **2001**, *314*, 23–40.
- (29) Strahs, D.; Schlick, T. *J. Mol. Biol.* **2000**, *301*, 643–663.
- (30) Cheatham, T. E.; Kollman, P. A. *J. Mol. Biol.* **1996**, *259*, 434–444.
- (31) Cheatham, T. E.; Crowley, M. F.; Fox, T.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 9626–9630.
- (32) Mazur, A. K. *J. Am. Chem. Soc.* **2003**, *125*, 7849–7859.
- (33) Auffinger, P.; Bielecki, L.; Westhof, E. *J. Mol. Biol.* **2004**, *335*, 555–571.
- (34) Reblova, K.; Spackova, N.; Stefl, R.; Csaszar, K.; Koca, J.; Leontis, N. B.; Spomer, J. *Biophys. J.* **2003**, *84*, 3564–3582.
- (35) Auffinger, P.; Louise-May, S.; Westhof, E. *Biophys. J.* **1999**, *76*, 50–64.
- (36) Villa, A.; Stock, G. *J. Chem. Theory Comput.* **2006**, *2*, 1228–1236.
- (37) Kormos, B. L.; Baranger, A. M.; Beveridge, D. L. *J. Am. Chem. Soc.* **2006**, *128*, 8992–8993.
- (38) Razga, F.; Koca, J.; Spomer, J.; Leontis, N. B. *Biophys. J.* **2005**, *88*, 3466–3485.
- (39) Spackova, N.; Spomer, J. *Nucleic Acids Res.* **2006**, *34*, 697–708.
- (40) Krasovska, M. V.; Sefcikova, J.; Reblova, K.; Schneider, B.; Walter, N. G.; Spomer, J. *Biophys. J.* **2006**, *91*, 626–638.
- (41) Villegas-Diaz, G.; Zacharias, M. *Biophys. J.* **2003**, *85*, 416–425.
- (42) Sanbonmatsu, K. Y. *Biochimie* **2006**, *88*, 1053–1059.
- (43) Zacharias, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 311–317.
- (44) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205–8212.
- (45) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
- (46) Fratini, A. V.; Kopka, M. L.; Drew, H. R.; Dickerson, R. E. *J. Biol. Chem.* **1982**, *257*, 4686–4707.
- (47) Prive, G. G.; Heinemann, U.; Chandrasegaran, S.; Kan, L. S.; Kopka, M. L.; Dickerson, R. E. *Science* **1987**, *238*, 498–504.
- (48) Gorenstein, D. G. *Chem. Rev.* **1994**, *94*, 1315–1338.
- (49) Sklenar, V.; Bax, A. *J. Am. Chem. Soc.* **1987**, *109*, 7525–7526.
- (50) Young, M. A.; Ravishanker, G.; Beveridge, D. L. *Biophys. J.* **1997**, *73*, 2313–2336.
- (51) Cheatham, T. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 4805–4825.
- (52) Trieb, M.; Rauch, C.; Wellenzohn, B.; Wibowo, F.; Loerting, T.; Liedl, K. R. *J. Phys. Chem. B* **2004**, *108*, 2470–2476.
- (53) Pearlman, D. A.; Kim, S. H. *J. Biomol. Struct. Dyn.* **1986**, *4*, 69–98.
- (54) Varnai, P.; Djuranovic, D.; Lavery, R.; Hartmann, B. *Nucleic Acids Res.* **2002**, *30*, 5398–5406.
- (55) Foloppe, N.; MacKerell, A. D. *J. Phys. Chem. B* **1999**, *103*, 10955–10964.
- (56) Djuranovic, D.; Hartmann, B. *J. Biomol. Struct. Dyn.* **2003**, *20*, 771–788.
- (57) Svozil, D.; Kalina, J.; Omelka, M.; Schneider, B. Local DNA conformations and their sequence preferences. *Nucleic Acids Res.*, in press.
- (58) Jones, S.; van Heyningen, P.; Berman, H. M.; Thornton, J. M. *J. Mol. Biol.* **1999**, *287*, 877–896.
- (59) Woda, J.; Schneider, B.; Patel, K.; Mistry, K.; Berman, H. M. *Biophys. J.* **1998**, *75*, 2170–2177.
- (60) Dickerson, R. E. *Nucleic Acids Res.* **1998**, *26*, 1906–1926.
- (61) Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. M.; Zhurkin, V. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163–11168.
- (62) Geierstanger, B. H.; Wemmer, D. E. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 463–493.
- (63) Lu, X. J.; Shakked, Z.; Olson, W. K. *J. Mol. Biol.* **2000**, *300*, 819–840.
- (64) Varnai, P.; Zakrzewska, K. *Nucleic Acids Res.* **2004**, *32*, 4269–4280.
- (65) Barone, F.; Lankas, F.; Spackova, N.; Spomer, J.; Karan, P.; Bignami, M.; Mazzei, F. *Biophys. Chem.* **2005**, *118*, 31–41.
- (66) Reblova, K.; Lankas, F.; Razga, F.; Krasovska, M. V.; Koca, J.; Spomer, J. *Biopolymers* **2006**, *82*, 504–520.
- (67) Fadna, E.; Spackova, N.; Stefl, R.; Koca, J.; Cheatham, T. E.; Spomer, J. *Biophys. J.* **2004**, *87*, 227–242.
- (68) Neidle, S. *Nucleic Acid Structure And Recognition*; Oxford University Press: Oxford, 2002.
- (69) Jordan, F.; Pullman, B. *Theor. Chim. Acta* **1968**, *9*, 242.
- (70) Berthod, H.; Pullman, B. *Biochim. Biophys. Acta* **1971**, *232*, 595.
- (71) Berthod, H.; Pullman, B. *Biochim. Biophys. Acta* **1971**, *246*, 359.
- (72) Saran, A.; Pullman, B.; Perahia, D. *Biochim. Biophys. Acta* **1972**, *287*, 211–231.
- (73) Saran, A.; Perahia, D.; Pullman, B. *Theor. Chim. Acta* **1973**, *30*, 31–44.
- (74) Pullman, B.; Gresh, N.; Berthod, H. *Theor. Chim. Acta* **1975**, *40*, 71–74.
- (75) Foloppe, N.; MacKerell, A. D. *J. Phys. Chem. B* **1998**, *102*, 6669–6678.
- (76) Foloppe, N.; Nilsson, L.; MacKerell, A. D. *Biopolymers* **2001**, *61*, 61–76.
- (77) Foloppe, N.; Hartmann, B.; Nilsson, L.; MacKerell, A. D. *Biophys. J.* **2002**, *82*, 1554–1569.
- (78) Florian, J.; Strajbl, M.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *120*, 7959–7966.
- (79) Shishkin, O. V.; Pelmeshnikov, A.; Hovorun, D. M.; Leszczynski, J. *J. Mol. Struct.* **2000**, *526*, 329–341.

- (80) Hocquet, A.; Leulliot, N.; Ghomi, M. *J. Phys. Chem. B* **2000**, *104*, 4560–4568.
- (81) Hocquet, A.; Ghomi, M. *Phys. Chem. Chem. Phys.* **2000**, *2*, 5351–5353.
- (82) Louit, G.; Hocquet, A.; Ghomi, M. *Phys. Chem. Chem. Phys.* **2002**, *4*, 3843–3848.
- (83) Shishkin, O. V.; Palamarchuk, G. V.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. B* **2006**, *110*, 4413–4422.
- (84) Gresh, N.; Spomer, J. E.; Spackova, N.; Leszczynski, J.; Spomer, J. *J. Phys. Chem. B* **2003**, *107*, 8669–8681.
- (85) Sychrovsky, V.; Vokacova, Z.; Spomer, J.; Spackova, N.; Schneider, B. *J. Phys. Chem. B* **2006**, *110*, 22894–22902.
- (86) Jurecka, P.; Spomer, J.; Hobza, P. *J. Phys. Chem. B* **2004**, *108*, 5466–5471.
- (87) Spomer, J.; Leszczynski, J.; Hobza, P. *J. Comput. Chem.* **1996**, *17*, 841–850.
- (88) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (89) Arnott, S.; Hukins, D. W. L. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504.
- (90) Schneider, B.; Neidle, S.; Berman, H. M. *Biopolymers* **1997**, *42*, 113–124.
- (91) Gelbin, A.; Schneider, B.; Clowney, L.; Hsieh, S. H.; Olson, W. K.; Berman, H. M. *J. Am. Chem. Soc.* **1996**, *118*, 519–529.
- (92) Lemieux, R. U. Effects of unshared pairs of electrons and their solvation on conformational equilibria. *Pure Appl. Chem.* **1971**, 527–548.
- (93) Wolfe, S. *Acc. Chem. Res.* **1972**, *5*, 102.
- (94) Radom, L.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1972**, *94*, 2371.
- (95) Dunning, T. H. *J. Phys. Chem. A* **2000**, *104*, 9062–9080.
- (96) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
- (97) Reiling, S.; Brickmann, J.; Schlenkrich, M.; Bopp, P. A. *J. Comput. Chem.* **1996**, *17*, 133–147.
- (98) Jensen, F. *Chem. Phys. Lett.* **1996**, *261*, 633–636.
- (99) Senent, M. L.; Wilson, S. *Int. J. Quantum Chem.* **2001**, *82*, 282–292.
- (100) Halasz, G. J.; Vibok, A.; Mayer, I. *J. Comput. Chem.* **1999**, *20*, 274–283.
- (101) Lii, J. H.; Ma, B. Y.; Allinger, N. L. *J. Comput. Chem.* **1999**, *20*, 1593–1603.
- (102) Rulisek, L.; Exner, O.; Cwiklik, L.; Jungwirth, P.; Stary, I.; Pospisil, L.; Havlas, Z. *J. Phys. Chem. C* **2007**, *111*, 14948–14955.
- (103) Valdes, H.; Klusak, V.; Pitonak, M.; Exner, O.; Stary, I.; Hobza, P.; Rulisek, L. *J. Comput. Chem.* **2008**, *29*, 861–870.
- (104) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Olsen, J. *Chem. Phys. Lett.* **1999**, *302*, 437–446.
- (105) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639–9646.
- (106) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- (107) Weigend, F.; Haser, M. *Theor. Chem. Acc.* **1997**, *97*, 331–340.
- (108) Weigend, F.; Haser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143–152.
- (109) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359–363.
- (110) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. *Chem. Phys. Lett.* **1993**, *213*, 514–518.
- (111) Jurecka, P.; Nachtigall, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578–4582.
- (112) Bernholdt, D. E.; Harrison, R. J. *Chem. Phys. Lett.* **1996**, *250*, 477–484.
- (113) Feller, D.; Apra, E.; Nichols, J. A.; Bernholdt, D. E. *J. Chem. Phys.* **1996**, *105*, 1940–1950.
- (114) Koch, H.; Fernandez, B.; Christiansen, O. *J. Chem. Phys.* **1998**, *108*, 2784–2790.
- (115) Jurecka, P.; Hobza, P. *Chem. Phys. Lett.* **2002**, *365*, 89–94.
- (116) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211–7218.
- (117) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735–746.
- (118) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899–926.
- (119) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (120) Curutchet, C.; Orozco, M.; Luque, F. J. *J. Comput. Chem.* **2001**, *22*, 1180–1193.
- (121) Forti, F.; Barril, X.; Orozco, M.; Luque, F. J. *J. Comput. Chem.* **2007**, *29*, 578–587.
- (122) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.
- (123) *MOLPRO, version 2006.1: A Package of Ab Initio Programs*, Cardiff University: Cardiff, U.K., 2006.
- (124) *MOPAC 6.0* - modified version, University of Barcelona: Barcelona, 2007.
- (125) Rocchia, W.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2001**, *105*, 6507–6514.
- (126) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. *J. Comput. Chem.* **2002**, *23*, 128–137.
- (127) Desiraju, G. R. *Steiner T The Weak Hydrogen Bond*; Oxford University Press: Oxford, 2001; Vol. 9.
- (128) Gorenstein, D. G.; Kar, D.; Luxon, B. A.; Momii, R. K. *J. Am. Chem. Soc.* **1976**, *98*, 1668–1673.
- (129) Weinhold, F.; Landis, C. In *Valency and Bonding*, Cambridge University Press: Cambridge, U.K., 2005; pp 215–275.
- (130) Reed, A. E.; Schleyer, P. V. *J. Am. Chem. Soc.* **1987**, *109*, 7362–7373.
- (131) Dupre, D. B.; Vorobyov, I.; Yappert, M. C. *J. Mol. Struct. (THEOCHEM)* **2001**, *544*, 91–109.
- (132) Murashov, V. V.; Leszczynski, J. *J. Phys. Chem. A* **1999**, *103*, 1228–1238.