JCTC Journal of Chemical Theory and Computation

Conformational Energies of DNA Sugar−Phosphate Backbone: Reference QM Calculations and a Comparison with Density Functional Theory and Molecular Mechanics

Arnošt Mládek,[†] Judit E. Šponer,[†] Petr Jurečka,[‡] Pavel Banáš,[‡] Michal Otyepka,[‡] Daniel Svozil,*[,†,§] and Jiří Šponer*[,†,∥,⊥]

Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic, Department of Physical Chemistry, Faculty of Science, Palacký University, 771 46 Olomouc, Czech Republic, Laboratory of Informatics and Chemistry, Faculty of Chemical Technology, Institute of Chemical Technology, Technická 3, 166 28 Prague 6, Czech Republic, Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo náměstí 2, 166 10 Prague 6, Czech Republic, and National Centre for Biomolecular Research, Faculty of Science, Masaryk University, 611 37 Brno, Czech Republic

Received August 17, 2010

Abstract: The study investigates electronic structure and gas-phase energetics of the DNA sugar−phosphate backbone via advanced quantum chemical (QM) methods. The analysis has been carried out on biologically relevant backbone conformations composed of 11 canonical BI-DNA structures, 8 pathological structures with α/γ torsion angles in the g+/t region, and 3 real noncanonical γ-trans structures occurring in the loop region of guanine quadruplex DNA. The influence of backbone conformation on the intrinsic energetics was primarily studied using a model system consisting of two sugar moieties linked together via a phosphodiester bond (SPSOM model). To get the conformation of the studied system fully under control, for each calculation we have frozen majority of the dihedral angles to their target values. CCSD(T) energies extrapolated to the complete basis set were utilized as reference values. However, the calculations show that inclusion of higher-order electron correlation effects for this system is not crucial and complete basis set second-order perturbation calculations are sufficiently accurate. The reference QM data are used to assess performance of 10 contemporary density functionals with the best performance delivered by the PBE-D/TZVPP combination along with the Grimme's dispersion correction, and by the TPSS-D/6-311++G(3df,3pd) augmented by Jurečka's dispersion term. In addition, the QM calculations are compared to molecular mechanics (MM) model based on the Cornell et al. force field. The destabilization of the pathological g+/t conformers with respect to the reference canonical structure and the network of intramolecular CH···O interactions were investigated by means of natural bond orbital analysis (NBO) and atoms-in-molecules (AIM) Bader analysis. Finally, four additional model systems of different sizes were assessed by comparing their energetics to that of the SPSOM system. Energetics of smaller MOSPM model consisting of a sugar moiety linked to a phosphate group and capped with methyl and methoxy group on the 5′- and 3′-ends, respectively, is fairly similar to that of SPSOM, while the role of undesired intramolecular interactions is diminished.

10.1021/ct1004593
Published on Web 11/16/2010

Introduction

Nucleic acids (NA; DNA and RNA) consist of linear chains of covalently bound sugar−phosphate units to which aromatic nucleic acid bases are attached. Nucleic acids form an astonishing variability of tertiary structures (ranging from simple double helices to complex ribonucleoprotein particles) that determine their function. Structural dynamics of nucleic acids result from delicate balance of numerous contributions. Among them, conformational preferences of the sugar−phosphate backbone belong to the most important ones. The sugar−phosphate backbone is chemically monotonous (sequence-independent). It contains a number of consecutive single bonds, which allow a substantial freedom for dihedral rotations. Thus, it has often been assumed that the backbone plays a rather passive role in structuring nucleic acids, while interactions involving the nucleobases are decisive (the base-centered view of NA structure).[1,2] On the other hand, there have also been suggestions that the internal backbone conformational preferences are decisively important.[3−5]

The conformation of the backbone is defined by a number of torsion angles called α, β, γ, δ, ε, and ζ (Figure 1). In addition to these, the precise conformation of inherently nonplanar (puckered) five-membered deoxyribose sugar ring is fully specified by five internal torsions $\tau_0-\tau_4$, the set of which can be simplified to only two internal degrees of freedom (pseudorotation P and amplitude τ_{max}). As the individual bases in nucleic acids are considered to be flat, the last degree of freedom in nucleotide is represented by the glycosidic bond linking a deoxyribose sugar and a base, the rotation around which is characterized by the torsion angle χ. Steric restrictions confine the values of these structural descriptors to discrete ranges.[6] A common convention for describing these backbone angles is to term values of ∼60° as gauche+ (g+), ∼ −60° as gauche− (g−), and ∼180° as trans (t). The typical (i.e., average) values of torsion angles of individual conformers can be obtained by careful analysis of NMR or crystal structures.[7] Such studies identify not only ranges individual torsions can adopt, but also describe numerous correlations involving pairs of the backbone torsion angles, as well as sugar pucker and glycosidic angle. The X-ray database contains hundreds of high-resolution DNA X-ray structures revealing thousands of individual dinucleotide backbone topologies. Using advanced bioinformatics tools, it is possible to cluster the backbone topologies into typical conformation families and to determine their representative (i.e., averaged) geometries with a high degree of confidence.[7] The existence of correlations is important as it means that the atomic motions in nucleotides follow concerted pattern of interdependence. Between the most important correlations belong the correlation between sugar pucker and glycosidic angle χ, the correlation between γ and α torsions, and the correlation between sugar pucker and δ angle (this correlation is rather strong, as it is given by the fact that one of the sugar internal torsions represents the rotation around the same bond as does δ).

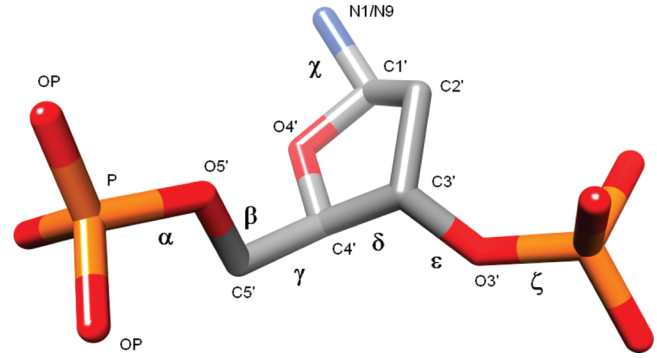

Figure 1. Atomic numbering and definition of the deoxyribonucleotide backbone torsion angles. The nucleotide backbone is described by the P−O5′−C5′−C4′−C3′−O3′ linkage. The torsion angles represent the rotation around the given bond. It is conventional to describe the backbone torsion angles of ∼60° as gauche+ (g+), of ∼300° as gauche− (g−), and of ∼180° as trans (t). The standard progression of NA chain is the 5′→3′ direction, which is from the left to the right in this particular figure.

Because of the inherent conformational flexibility of the polynucleotide backbone, there exists a wide range of different double helical conformations. The most common form is the antiparallel right-handed B-DNA double helix.[8] This conformation, often referred to as canonical one, is characterized by the following set of typical torsion angles: α = 299° (g−), β = 179° (t), γ = 48° (g+), δ = 133°, ε = 182° (t), and ζ = 263°. Another possible right-handed form, A-DNA, is similar to B-DNA, but with different sugar conformation leading to different base position with respect to the helical axis. Z-DNA, a left-handed form, was also prepared, although its biological relevance is still the subject of investigation.[9] The structural variability of DNA is critical for recognition between DNA and proteins, which plays a crucial role in such essential processes as replication or transcription. Characterization of the backbone conformational space is therefore highly important for understanding of DNA recognition.

However, while intrinsic energetics of interbase interactions has been widely studied,[10,11] very little is known about the backbone electronic structure and energetics. Their study is considerably more difficult[12−22] due to the high flexibility and various correlations between the individual torsion angles. The ability to uniquely assign and compare energies of individual biomolecular conformers is indispensable to clarify their conformational preferences. The intrinsic conformational preferences are established by analysis of the relation between molecular structures and molecular energies. Using computational methods, we can derive potential energy surfaces (PES) by assigning corresponding electronic energy to each single geometry, and thus evaluating energy as an unambiguous function of the molecular structure. The resultant potential function that drives a biomolecular system arrangement in natural environment can be generally ex-

* Corresponding author e-mail: daniel.svozil@gmail.com (D.S.), sponer@ncbr.chemi.muni.cz (J.S.).
† Institute of Biophysics, Academy of Sciences of the Czech Republic.
‡ Palacký University.
§ Institute of Chemical Technology.
|| Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic.
⊥ Masaryk University.

pressed as a sum of three distinct terms. The first term, intrinsic electronic energy of the system in vacuum, can be acquired exclusively via a highly accurate ab initio QM treatment. The intrinsic energy component can also be approximated by means of balanced MM force fields, where the electronic effects are mimicked by effective classical potential functions. The second contribution to the potential energy function arises from the interaction of the studied molecular entity with its natural environment, that is, mainly the solvation effects and the overall context of the NA molecule. The last term represents coupling between the intrinsic and environmental contributions.

The present work is focused on the first energy component, for it influences the resulting conformational space occupation to a considerable extent. We provide an extensive QM characterization of the intrinsic conformational preferences of the sugar−phosphate unit of DNA. It is, however, important to bear in mind that the environment and coupling terms significantly affect the resulting energetics too, and thus should not be omitted when predicting structural preferences in real environments. This is the main limitation of our study.

The present study has several key features. First, we derive reference QM data; that is, we push the theoretical calculations to the highest limits achievable by contemporary computational tools. These calculations are then used for comparison with a wide range of less expensive QM methods and also variants of the Cornell et al. MM force field,[23] to assess their performance and to obtain basic physical chemistry insights into the systems under study.

We compare five model systems of different complexity, to establish sensitivity of the results to the choice of the model system. The very first critical step in a theoretical investigation of biologically relevant macromolecules is to select a convenient model system. The size and the overall structural complexity of biological units to be described inherently delimit the set of appropriate model systems. The model should be large enough to capture all important conformational and electronic characteristics of the studied biomolecule. Too large model can make rigorous ab initio high-level QM investigation intractable, while too small model system may be chemically irrelevant. As the backbone conformation is unambiguously determined by six strongly coupled torsion angles α, β, γ, δ, ε, and ζ (Figure 1),[7] the energetic assessment of the complete conformational space constitutes a complex six-dimensional problem. Furthermore, the sugar pucker defined by two independent variables and coupled to a certain degree with the δ torsion angle also influences PES of the model compound. In addition, the correct description of the electronic distribution along the negatively charged backbone is a rather demanding task. To properly analyze the polarizable anionic nature of the backbone, at least moderate size basis set with diffuse functions must be utilized. This makes high-level QM computations difficult even for quite small (e.g., one nucleotide) DNA fragments. Besides, the diffuse electron density due to anionic character causes slower density matrix convergence. However, there is yet another reason that complicates investigations of larger model systems, even when they are computationally tractable per se. The larger is the model system, the more prone it is to adopt geometries that are biased, for example, by intramolecular H-bonds and other interactions that are not relevant to complete solvated biomolecules. The complexity of the conformational spaces increases dramatically with the size of the system. Additional issues that preclude studies of larger systems are uncompensated charges of multiple phosphate groups that would dominate the gas-phase electrostatics and also an artifact known as intramolecular basis set superposition error (BSSE; for more details, see the QM calculations paragraph below).

In contrast to our preceding study,[12] we modified the computations in such a way that basically we always keep all backbone dihedral angles frozen at predefined values. This has been necessitated by the fact that when freezing only very few dihedral angles the remaining free dihedral angles can adopt numerous combinations as local minima. This substantially biases conformational scans and often leads to unrealistic geometries.

We concentrate our efforts to three DNA backbone geometrical substates: canonical B-DNA conformation and two conformations with γ angle in trans and α angle in gauche+ region. The first one corresponds to α/γ topology that normally should not occur in free B-DNA;[24] however, it accumulated in longer simulations with an earlier version of the Cornell et al. MM force field. With the original parametrization (ff94[23] and ff99[25] force fields), irreversible sequence-independent α/γ flips occur in long molecular dynamics (MD) simulations of B-DNA duplexes.[26−28] The accumulation of α/γ transitions then causes entire B-DNA structure degradation.[29] The backbone γ torsional profile was recently reparametrized in the parmbsc0 force field,[29] which allows for long time scale simulations of B-DNA duplexes without any significant loss of helical structure. Although the parmbsc0 force field represents a decisive progress in MD simulations of DNA, further refinement would be still useful as not all imbalances are yet treated sufficiently.[30−32] It can be illustrated with the second γ-trans geometry investigated in this study, which has been experimentally observed in the first nucleotide of single-stranded loops of human telomere guanine quadruplex (G-DNA) and thus corresponds to real substate of DNA backbone.[33,34] However, it has been destabilized by the parmbsc0 force field.[31]

Methods and Model Systems

Model Systems. To fully capture the conformational behavior of the sugar−phosphate backbone, at least the dinucleotide unit must be utilized. However, the complexity of the potential model systems is limited by the fact that, to avoid strong electrostatic repulsion between two anionic phosphate groups, which is otherwise screened by the solvent and counterions in the physiological environment, only one phosphate group is conceivable. In addition, stacking interactions between nucleobases in dinucleotide model systems would result in a significant conformational bias, and the bases were thus excluded from the model systems (see the discussion below). The model systems used in the present study can be divided into two groups.

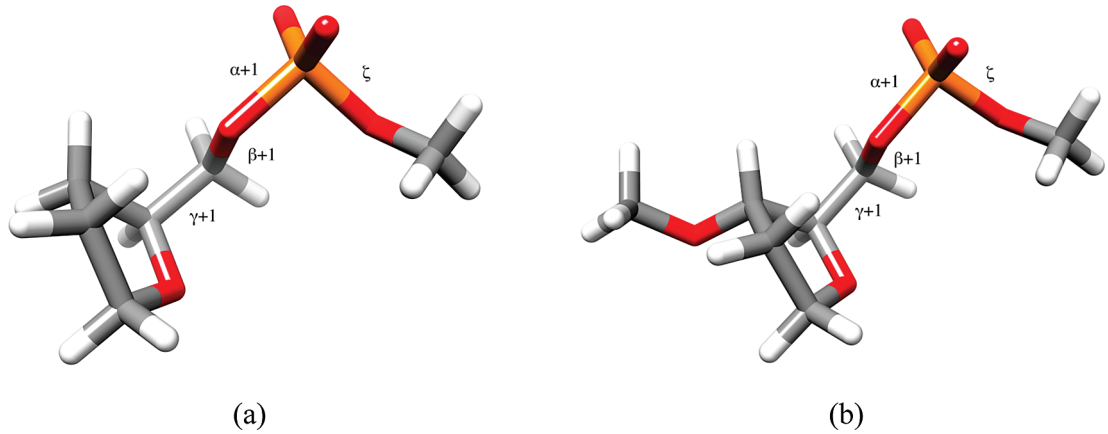

Figure 2. (a) SPM (sugar−phosphate−methyl) and (b) MOSPM (methyl−oxygen−sugar−phosphate−methyl) Group I model compounds. Labeling of the bonds is according to the standard nucleic acid nomenclature. The ζ torsion angle is defined as C3′−O3′−P−O5′(n+1), the α+1 torsion angle as O3′−P−O5′(n+1)−C5′(n+1), the β+1 torsion angle as P−O5′(n+1)−C5′(n+1)−C4′(n+1), and the γ+1 torsion is defined as O5′(n+1)−C5′(n+1)−C4′(n+1)−C3′(n+1).

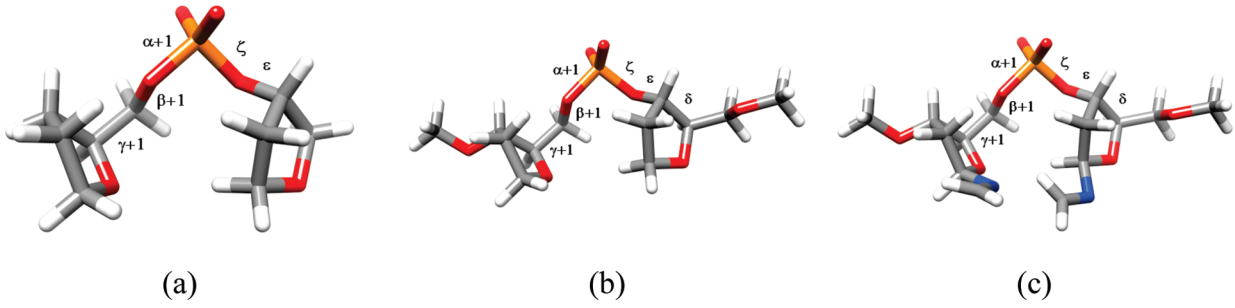

Figure 3. (a) T3PS (tetrahydrofuran with 3′ phosphate with a capping sugar), (b) SPSOM (sugar−phosphate−sugar with capping methoxy groups), and (c) SPSOM-NCH2 (sugar−phosphate−sugar with capping methoxy groups and with $-N{=}CH_2$ groups) Group II models. Labeling of the bonds is according to the standard nucleic acid nomenclature. The δ torsion angle is defined as C5′−C4′−C3′−O3′, the ε torsion angle as C4′−C3′−O3′−P, the ζ torsion as C3′−O3′−P−O5′(n+1), the α+1 torsion angle as O3′−P−O5′(n+1)−C5′(n+1), the β+1 torsion angle as P−O5′(n+1)−C5′(n+1)−C4′(n+1), and the γ+1 torsion is defined as O5′(n+1)−C5′(n+1)−C4′(n+1)−C3′(n+1). The 3′-end (i.e., n+1) and 5′-end sugars are on the left and right, respectively, because the 5′−3′ direction is from right to left in this figure (opposite of Figure 1).

Group I (Figure 2) consists of two organophosphate models with only one sugar residue. The first system, SPM, (the abbreviation stands for sugar−phosphate−methyl; see Figure 2a), consists of a sugar residue and a methyl group linked via a phosphodiester bond as suggested by Orozco et al.[29] The second model compound, MOSPM, (i.e., methyl−oxygen−sugar−phosphate−methyl; see Figure 2b), represents an extended version of the SPM model with the H3′(n+1) atom replaced by a methoxy functional group (−O−CH3). In this system, the sugar moiety is situated in a more natural chemical environment because of the longer backbone fragment. To prevent formation of artificial intramolecular hydrogen bonds, the 3′ and 5′ ends of MOSPM, as well as the 5′ end of SPM, were terminated with methyl groups.

Group II (Figure 3) contains three models with the sugar−phosphate−sugar unit mimicking a dinucleotide step. The simplest system containing no additional groups was designed and used by MacKerell[13] and is further referred to as T3PS (tetrahydrofuran with 3′ phosphate with a capping sugar). Another model compound, SPSOM (sugar−phosphate−sugar−oxygen−methyl), used in our previous study[12] retains all the features characteristic for the DNA dinucleotide building blocks. The last, and the most complex, model[12] referred to as SPSOM-NCH2 was constructed out of SPSOM model by replacing H1′ of both sugar residues with a methylene-imino functional group ($-N{=}CH_2$). This extension was proposed to compensate, at least partially, for the neglect of the nucleobases, leading to a better description of the hyperconjugation effects along the sugar−phosphate backbone.[12]

Starting Geometries of the Conformers. All canonical (BI) DNA backbone geometries (geometries at and around the canonical structure) were labeled with the symbol “a”; that is, the particular conformers in the canonical series are referred to as a1, a2, ..., a11 (Table 1). The typical torsion angles for this most populated substate of free B-DNA duplex (α = 299° (g−), β = 179° (t), γ = 48° (g+), δ = 133°, ε = 182° (t), and ζ = 263°) were obtained by the means of analysis of crystal structures of 1531 dinucleotide steps in DNA.[7] The representative of the “average” BI-DNA and simultaneously the reference structure in this study is labeled a1; the remaining 10 geometries (a2-a11) were prepared to characterize PES in the vicinity of the above “average” BI structure. The a1 structure is the best representative of the BI cluster of geometries.

The first group of noncanonical DNA backbone geometries occupying the less populated α/γ = g+/t conformational

Table 1. Torsion Angle Values for the Canonical "a" and Noncanonical "b" Geometries

structure label	$\alpha+1^c/\gamma+1^a$	$\gamma+1^a$	$\beta+1^b$	$\alpha+1^c$	ζ^d	ε^e	δ^f	targeted perturbation
a1	g−/g+	45	180	300	260	180	136	BI; no perturbation
a2		50	180	280	270	190	136	$(\alpha+1)-20$
a3		40	180	320	260	180	136	$(\alpha+1)+20$
a4		45	170	300	260	190	136	$(\beta+1)-10$
a5		45	190	300	260	180	136	$(\beta+1)+10$
a6		35	180	310	260	185	136	$(\gamma+1)-10$
a7		55	180	295	260	180	136	$(\gamma+1)+10$
a8		50	180	300	260	190	116	δ: C1′-exo
a9		45	180	295	260	190	150	δ: C3′-exo
a10		50	190	310	270	160	136	$\varepsilon-20$
a11		45	170	300	260	200	130	$\varepsilon+20$
b1	g+/t	195	225	65	190	250	145	no perturbation
b2		200	220	50	190	265	150	$(\alpha+1)-15$
b3		190	220	80	190	225	140	$(\alpha+1)+15$
b4		190	215	60	190	260	150	$(\beta+1)-10$
b5		195	235	70	190	240	140	$(\beta+1)+10$
b6		205	225	60	190	250	145	$(\gamma+1)+10$
b7		195	220	75	210	235	145	$\zeta+20$
b8		190	220	65	180	255	145	$\zeta-10$

[a] $\gamma+1$: O5′(i+1)−C5′(i+1)−C4′(i+1)−C3′(i+1). [b] $\beta+1$: P(i+1)−O5′(i+1)−C5′(i+1)−C4′(i+1). [c] $\alpha+1$: O3′(i)−P(i+1)−O5′(i+1)−C5′(i+1). [d] ζ: C3′(i)−O3′(i)−P(i+1)−O5′(i+1). [e] ε: C4′(i)−C3′(i)−O3′(i)−P(i+1). [f] δ: C5′(i)−C4′(i)−C3′(i)−O3′(i).

Table 2. Torsion Angle Values for the Quadruplex Loop "q" Geometries

structure label	$\alpha+1^c/\gamma+1^a$	$\gamma+1^a$	$\beta+1^b$	$\alpha+1^c$	ζ^d	ε^e	δ^f	PDB code/NDB code	residue ID
q1	g+/t	176	189	77	73	230	143	1KF1/UD0017	DT11
q2		183	190	79	61	224	152	1KF1/UD0017	DT05
q3		195	184	63	64	220	147	1KF1/UD0017	DT17

[a] $\gamma+1$: O5′(i+1)−C5′(i+1)−C4′(i+1)−C3′(i+1). [b] $\beta+1$: P(i+1)−O5′(i+1)−C5′(i+1)−C4′(i+1). [c] $\alpha+1$: O3′(i)−P(i+1)−O5′(i+1)−C5′(i+1). [d] ζ: C3′(i)−O3′(i)−P(i+1)−O5′(i+1). [e] ε: C4′(i)−C3′(i)−O3′(i)−P(i+1). [f] δ: C5′(i)−C4′(i)−C3′(i)−O3′(i).

region is analogously labeled with symbol "b", that is, b1, b2, ..., b8 (Table 1). Note that labeling the two series of structures as "a" or "b" has nothing to do with the helix form designation. The second subset of α/γ = g+/t structures from human telomeric quadruplex loops[33,34] is labeled with symbol "q" and consists of three members denoted as q1, q2, and q3 (Table 2).

The geometries in the present Article were derived in the following way. At the beginning, we have taken two "parent" structures of the SPSOM model optimized at the B3LYP/6-31+G(d) level in our earlier paper (Supporting Information of ref 12, g−/g+ (p S13) and g+/t (p S15) structures). These structures (named spsom_a and spsom_b) served as the initial structures for the derivation of the "a" and "b" subsets of structures of the SPSOM system in the present Article. Next, we have modified (using modredundant route section keyword of the Gaussian 03 software[35]) all dihedral angles to the desired values (Table 1) to obtain structures a1−a11 and b1−b8, which were subsequently optimized at the respective theoretical levels (see below). Note that the actual final geometries derived in this Article are not affected by details of the two initial "parent" geometries, as we set up and constrained the dihedral angles upon optimizations.

However, due to the correlations involving pairs of backbone angles (known from the X-ray database study), the shift of each torsion angle from its canonical value introduces also the changes in values of other torsional angles. To cover these changes, each time when the target torsion angle was shifted from the canonical a1 structure, the remaining torsion angle values were adjusted accordingly to reflect correlation of torsion angles suggested by the X-ray database study (Table 1).[7] As this procedure takes the backbone torsion angles correlation into account, we suppose it is better for sampling the PES of the real DNA molecule than just keeping the remaining torsions fixed at their canonical values.

The geometries of the three q-conformers were prepared from the crystal structure of human telomeric quadruplex loops (pdb code: 1KF1, resolved at the 2.10 Å resolution) by extracting the corresponding SPSOM segments (Table 2) from the three independent loop structures. The addition of hydrogen atoms was carried out manually using Accelrys ViewerPro molecular modeling software. Their initial positions were adjusted according to the hybridization state of the linked heavy atom. The structures were then optimized at the respective theoretical levels (see below) with frozen dihedrals.

The other models, T3PS, SPM, MOSPM, and SPSOM-NCH2, were derived in the following manner. We have taken the a1, b1, and qx (x = 1, 2, and 3) MP2-optimized SPSOM geometries (Table 3). These were appropriately chemically modified to get the other model systems. These geometries served as "parent" structures for the T3PS, SPM, MOSPM, and SPSOM-NCH2 models (Table 3; parent structure name). These parent structures are not reoptimized after modification. The corresponding a1−a11 and b1−b8 structures were then derived from these parent structures by setting up the required combination of dihedral angles (Table 1) and subsequent constrained optimizations. The starting geometries, that is, all the above-noted parent structures of all model systems for the canonical "a", noncanonical "b", and quadruplex "q" variant (Table 3), are given in the Supporting

Table 3. List of Parent Geometries with Their Names and Origin[a]

model label	α/γ family	parent structure name	origin of the geometry
SPSOM	g−/g+	spsom_a	g−/g+ SPSOM[b]
	g+/t	spsom_b	g+/t SPSOM[b]
	g+/t G-DNA	spsom_q*x*[d]	quadruplex loop, NDB, UD0017; PDB, 1KF1[c]
T3PS	g−/g+	t3 ps_a	a1 SPSOM||MP2/6-31+G(d), with modifications
	g+/t	t3 ps_b	b1 SPSOM||MP2/6-31+G(d), with modifications
	g+/t G-DNA	t3 ps_q*x*	q*x* SPSOM||MP2/6-31+G(d), with modifications
SPSOM-NCH2	g−/g+	spsom_nch2_a	a1 SPSOM||MP2/6-31+G(d), with modifications
	g+/t	spsom_nch2_b	b1 SPSOM||MP2/6-31+G(d), with modifications
	g+/t G-DNA	spsom_nch2_q*x*	q*x* SPSOM||MP2/6-31+G(d), with modifications
SPM	g−/g+	spm_a	a1 SPSOM||MP2/6-31+G(d), with modifications
	g+/t	spm_b	b1 SPSOM||MP2/6-31+G(d), with modifications
	g+/t G-DNA	spm_q*x*	q*x* SPSOM||MP2/6-31+G(d), with modifications
MOSPM	g−/g+	mospm_a	a1 SPSOM||MP2/6-31+G(d), with modifications
	g+/t	mospm_b	b1 SPSOM||MP2/6-31+G(d), with modifications
	g+/t G-DNA	mospm_q*x*	q*x* SPSOM||MP2/6-31+G(d), with modifications

[a] All parent starting geometries are available in *xyz* format in the Supporting information. The notation "X||Y" denotes system "X" optimized at "Y" level of theory. The corresponding a1−a11 and b1−b8 structures were derived from these parent structures by modifying the dihedrals according to the Table 1 and subsequently using constrained gradient optimization. [b] See the Supporting Information of ref 12, g−/g+ (p S13) and g+/t (p S15) structures for spsom_a and spsom_b, respectively. [c] X-ray structure of the human telomeric quadruplex sequence. [d] "q*x*" represents q1, residue, DT11; q2, residue, DT05; and q3, residue, DT17.

Information. The Supporting Information further includes all SPSOM MP2-optimized geometries along with their reference CBS(T) (see below) energies.

QM Calculations. Every molecular electronic calculation with a finite basis set is susceptible to BSSE due to the fact that molecular orbitals are approximated by an expansion in terms of analytical basis functions centered on different points in real space (usually the nuclei) that are dependent on the geometry of the studied system.[36] The magnitude of the BSSE decreases as the size of the basis set increases and becomes zero at the limit of an infinite (also called complete) basis set (CBS).[37] For intermolecular noncovalent interactions, the BSSE can be efficiently corrected by the counterpoise procedure.[38] However, BSSE also affects potential energy surfaces of monomers.[39−44] Although several methods were proposed to eliminate the intramolecular BSSE,[39−41] the most reliable approach is to perform calculations with large basis sets,[40] ideally at the CBS limit. As the intramolecular BSSE is very significant in structures with conjugated aromatic systems,[45] the omission of nucleobases in our model compounds reduces this artifact substantially.

Geometry Optimizations. The gradient geometry optimization of the SPSOM structures using internally determined redundant coordinates was carried out at three levels of theory: (i) B3LYP/6-31+G(d),[46,47] (ii) MP2/6-31+G(d), and (iii) the resolution of identity (RI)[48,49] DFT TPSS meta-GGA functional[50] augmented with the empirical dispersion D-0.96-27 type term[51] (TPSS-D) with the large 6-311++G-(3df,3pd) (LP) basis set.

As we are primarily interested in biologically meaningful conformations rather than in pure and often unnatural gas-phase minima, we have fixed the highly flexible torsion degrees of freedom during the geometry optimization at their specified values (Table 1) via application of constraints to all backbone dihedrals, ranging from $\gamma+1$ to δ (in 3′→5′ direction). Obviously, the actual backbone conformation in real physiological environment is determined by numerous factors we do not take into account in this study, like interactions with solvent and counterions, base−base stacking, and edge-to-edge interactions, etc. It is thus unlikely that the experimentally observed and occupied conformational regions would be obtained by unconstrained (or insufficiently constrained) gas-phase optimizations of small model systems. To preserve the connection to biology, it is imperative to fix the backbone during optimization process and restrict minimization to conformational domains we are interested in. Otherwise, we would end up with intrinsically relaxed but irrelevant geometries, not occurring in real DNA structures. To illustrate the necessity of constrains, we performed several full optimizations, that is, without constraints, the results of which can be found in the Supporting Information (Table S12). The unconstrained optimizations drive the structure energetically downhill and away from initial (and relevant) conformational region.

While the MP2 and B3LYP optimizations were carried out using Gaussian 03 software, which is capable of fixing coordinates at values that differ from those of the input structure (i.e., it is possible to change torsions to desired values prior to constrained optimization within single optimization input), the Turbomole code is able to fix torsions at input values only. Because the module for the dispersion calculation of the TPSS-D optimization run was accessible for Turbomole package only, we used the B3LYP optimized geometries with the backbone torsion angles already set to target values as input structures for the subsequent TPSS-D reoptimization.

The minimum energy geometries of the remaining models (except for SPSOM-NCH2, where the optimization turned out to be problematic, see Results and Discussion) were obtained at the B3LYP/6-31+G(d) and MP2/6-31+G(d) levels of theory. All relevant backbone dihedrals starting from $\gamma+1$ (in 3′→5′ direction) were fixed during the optimization procedure in studied model systems.

Single-Point Calculations. The SPSOM model was chosen as the reference compound, which served to compare number of wave function-based and DFT-based methods with

the most accurate CBS(T) approximation (this abbreviation stands for estimated CCSD(T)/CBS calculations, see below).

Because of computational demand of the CCSD(T) calculations, the method with the highest consensus to CBS(T) was selected to benchmark conformers of the remaining models.

All relative energies are computed with respect to the conformer a1, which is considered as the best representative of the canonical BI-DNA.

Wave Function-Based Single-Point Calculations. The RIMP2/CBS energies were estimated using the extrapolation scheme suggested by Halkier and co-workers[52,53] and Dunning's augmented correlation-consistent basis sets of double-ζ and triple-ζ quality (aug-cc-pVDZ and aug-cc-pVTZ).[54,55] The extrapolation to the CBS effectively eliminates both BSSE and basis set incompleteness errors. Nevertheless, our preceding experience indicates this extrapolation scheme with aug-cc-pVDZ and aug-cc-pVTZ basis sets provides results approaching more likely the MP2/aug-cc-pVQZ calculations than the true MP2/CBS limit. Therefore, some residual BSSE and basis set incompleteness errors are likely to remain.[10,11,51] The Hartree−Fock (HF) energy and the correlation MP2 energy contribution are evaluated independently according to eqs 1 and 2:

$$E_{\mathrm{CBS}}^{\mathrm{HF}} = E_{\text{aug-cc-pVDZ}}^{\mathrm{HF}} + \frac{E_{\text{aug-cc-pVTZ}}^{\mathrm{HF}} - E_{\text{aug-cc-pVDZ}}^{\mathrm{HF}}}{0.760691} \quad (1)$$

$$E_{\mathrm{CBS}}^{\mathrm{MP2}} = E_{\text{aug-cc-pVDZ}}^{\mathrm{MP2}} + \frac{E_{\text{aug-cc-pVTZ}}^{\mathrm{MP2}} - E_{\text{aug-cc-pVDZ}}^{\mathrm{MP2}}}{0.703704} \quad (2)$$

Note that this equation notation for both the HF (eq 1) and the MP2 correlation component (eq 2) has been algebraically derived from the standard Helgaker's formulation (see refs 52 and 53) and is more suitable for practical evaluation of CBS extrapolated energies. The original extrapolation scheme and our formulation are thus equivalent.

To speed up the regular MP2 procedure, the RI-approximation[56−58] was utilized.

To account for higher order correlation effects, coupled cluster (CC) calculations with single, double, and perturbative, noniterative triple excitations utilizing 6-31+G(d) basis set (CCSD(T)/6-31+G(d)) were performed. Providing that the difference between the MP2 and CCSD(T) energies, often abbreviated as ΔCCSD(T), shows only small basis set dependence,[59−63] the CCSD(T)/CBS energies could be estimated as eq 3:

$$E_{\mathrm{CBS}}^{\mathrm{CCSD(T)}} \approx \mathrm{CBS(T)} = E_{\mathrm{CBS}}^{\mathrm{MP2}} + (E_{\text{6-31+G(d)}}^{\mathrm{CCSD(T)}} - E_{\text{6-31+G(d)}}^{\mathrm{MP2}}) = E_{\mathrm{CBS}}^{\mathrm{MP2}} + \Delta\mathrm{CCSD(T)} \quad (3)$$

The CCSD(T) calculations were carried out with the MOLPRO 2006.1 package.[64] The CBS(T) energies represent the reference values with which all other methods were compared. The CBS(T) abbreviation is used to indicate that the CCSD(T) CBS extrapolation is approximated.[10,11]

DFT-Based Single-Point Calculations. The DFT-based methods are computationally less demanding and less affected by the BSSE than conventional wave function-based methods. The older functionals are known to be deficient in the description of the dispersion interaction.[65,66] Thus, much recent effort has been spent on including the dispersion interaction either by adjusting current functionals,[67,68] by developing new functionals,[69−71] or by augmenting the existing functionals with empirical correction dispersion energy term.[51,72−74] In the present work, the performance of several traditional and recent functionals was compared to the reference CBS(T) energies of the SPSOM model system. The density functionals considered in this work fall into the following categories:

(i) The first is pure generalized gradient approximations (GGA) functionals. In the present work, the PBE functional[74] augmented with the Grimme's empirical correction term for long-range dispersion effects,[75] PBE-D, in conjunction with the TZVPP[76] triple-ζ basis set was employed.

(ii) The second is hybrid nonlocal GGA functionals containing a portion of the exact exchange interaction from the HF calculation. The only hybrid GGA functional used in the present study was the B3LYP[46,47] functional with the Pople's 6-31+G(d) basis set.

(iii) The third is hybrid meta-GGA type functionals that include also terms dependent on the kinetic energy density. The nonlocal meta-GGAs used in this work include M06, M06-HF, M06-2X, M08-HX, and M08-SO.[77−79] All listed Minnesota functionals belong to the widely used M06 and M08 suites of functionals and were applied with the 6-31+G(d) basis set.

(iv) The fourth is meta, HF-exchange excluded, GGA functionals with entirely local exchange-correlation description. The local treatment makes these functionals computationally very efficient; they are an order of magnitude faster than methods including HF exchange.[80] The two local meta-GGA functionals employed in the present study are M06-L[81,82] and TPSS[50] in combination with Jurečka's empirical dispersion B-0.96-27 type term (TPSS-D).[51] The 6-31+G(d) and 6-311++G(3df,3pd) "LP" basis sets were used for M06-L and TPSS-D calculations, respectively. The RI approximation was available and used for the TPSS functional calculation only.

(v) The fifth is double-hybrid density functionals, which improve the hybrid GGAs by adding a fraction of MP2 correlation energy on top of the HF exchange interaction. The mPW2-PLYP[83] functional with the Ahlrich's TZVPP[76] triple-ζ quality basis set was used.

All RI-approximated calculations were performed with Turbomole 5.10.[84] The mPW2-PLYP and PBE-D energies were calculated with Orca2.6.[85] All remaining ab initio calculations were carried out using Gaussian 03, revision E.01.[35]

Electronic Structure Analysis. The potential energy surface calculations were complemented by atoms-in-molecules (AIM)[86−88] and natural bond orbital (NBO)[89−95] analyses. The aim of the AIM calculations is to analyze the local electron density curvatures and reveal critical points, which can improve the understanding of the physical principles driving the stabilization of the individual conformers. The charge density topologies were computed from the nonfrozen core approximated MP2/6-31+G(d) converged

wave functions. The basis set contained 6d functions rather than the standard 5d ones. These calculations were performed with the AIMPAC code.[96,97]

NBO analysis was applied to identify orbital interactions leading to electron delocalization in the studied systems. As we have shown elsewhere,[12] both B3LYP and HF methods assign approximately the same fraction of electrons to non-Lewis delocalized orbitals and thus provide similar results. In this study, we performed the analysis for the HF/6-31+G(d) orbitals using MP2-optimized structures and the NBO 3.0 program[91,94] implemented in the Gaussian 03 code.[35]

Force Field Calculations. The force field energies were computed using the nonpolarizable ff99 force field,[25] as well as its reparametrized variant parmbsc0.[29] The poor description of the α/γ energetics in ff99 leading to the serious helix unwinding during long DNA simulations[26−28] is substantially improved in its parmbsc0 reparametrization. Prior to evaluation of both the ff99 and the parmbsc0 energies, the MP2-optimized model system geometries have been relaxed using the respective force field except for the fixed backbone dihedrals. That means that force field energies were derived using force field geometries. The relaxation was carried out to the default tolerances using the steepest descent technique for the first 250 iterations, followed by the conjugate gradient method; no cutoff was applied. To keep the backbone dihedrals at the values given in Tables 1 and 2, tight restraints have been imposed. The penalty function that pushes a given term toward the desired value was set at 3000 kcal mol^{-1}.

The electrostatic contribution to the internal energy is computed as a pairwise interaction between the atom-centered partial charges. Hence, the partial charges must be somehow derived and assigned to each atom with a constraint that the sum of partial charges equals the total charge of the system. QM offers numerous established schemes how to derive atomic charges, such as Mulliken and Voronoi population analysis, NBO analysis, AIM analysis, etc. However, AMBER force field calculations are based on the so-called ESP (electrostatic potential) or RESP (restrained ESP) charges.[98] The (R)ESP charges are used in MM calculations because they allow realistic estimates of molecular interactions and conformational preferences. The (R)ESP charges are determined in the following way: (i) the molecular geometry is optimized to a stable minimum conformation using a convenient QM method, (ii) then the electrostatic potential of the optimized molecule is calculated on a three-dimensional real-space grid, (iii) which is subsequently used to fit the atom-centered charges. So the (R)ESP charges are actually effective charges fitted solely to reproduce the QM-determined electrostatic potential of the system created by the electronic and nuclei distribution. Note that (R)ESP charges have no physical meaning as there is an infinite number of solutions of how to allocate charges among the atoms to reproduce the electrostatic potential to a desired precision. As our model systems do not belong to the standard residues for which the AMBER library contains precomputed partial charges, we used the RESP fitting procedure to obtain new charges for all our model systems. That means that the charges we used in this study are not exactly the same as in the original force field but have been derived with the same conception. This allows a consistent comparison between the QM and MM computations. The charges have been fitted at the HF/6-31G(d) level of theory for the most stable MP2/6-31+G(d) optimized geometry, which is the a2 structure. Note that the HF/6-31G(d) level is intentionally used to derive the Cornell et al. force field charges because the modestly overpolarized HF charge distributions are more compatible with the water models typically used for condensed-phase simulations. The charges are given in the Supporting Information. All force field calculations were performed with the sander module of Amber 10.0 suite of programs.[99]

Results and Discussion

SPSOM Geometries. The conformations corresponding to canonical (a-conformers), noncanonical g+/t (b-conformers), and G-DNA loop g+/t (q-conformers) values of α/γ angles were optimized at B3LYP/6-31+G(d), MP2/6-31+G(d), TPSS-D/LP, parm99, and parmbsc0 levels of theory with the backbone torsion angles kept constant at the values defined in Tables 1 and 2.

The C1′···C1′ distances within different conformational types increase in the order of a < b < q. The differences in C1′···C1′ distances between MP2 and TPSS-D structures are nearly negligible for a- and b-subsets of conformers. Regarding q-geometries, TPSS-D C1′···C1′ distances are about 0.14−0.28 Å longer than their MP2 equivalents.

The B3LYP geometries are generally more extended than the MP2 and TPSS-D ones with C1′···C1′ distances being on average 0.30 Å longer for all three structure sets. Further relaxation of the MP2-optimized geometries using empirical force fields leads to additional increase in C1′···C1′ distances. The parmbsc0 and parm99 geometries are nearly identical with predicted C1′···C1′ distances being 0.2−0.5 Å longer as compared to MP2 data for a- and q-conformers, respectively. The force field minimized geometries mutually differ more in the noncanonical b-region of the conformational space because parmbsc0 extends the C1′···C1′ distance by 0.2−0.3 Å more than parm99.

The fact that the MP2 and TPSS-D optimized conformations tend to be slightly more packed than the B3LYP and force field ones might be attributed to the formation of weak CH···O hydrogen bonds and van der Waals contacts (see later discussion on AIM analysis). The propensity to form close contacts between C and O atoms follows the subsequent order of methods: MP2 ∼ TPSS-D > B3LYP > force field. The interrelation between the C1′···C1′ distance and formation of CH···O interaction is illustrated in the Supporting Information, Figure S1, which shows that changing the method of calculations is accompanied by variation of the C2′···O5′ distance, while the 3′-sugar is somewhat repuckered. This ultimately affects also the C1′···C1′ distance. Although the majority of detected CH···O interactions are artifacts of optimization process in the absence of solvent environment and DNA context, they do affect both geometries and intrinsic energetics. Note that separation of the specific impact of such interactions on the optimal geometries

from the general propensities of the applied computational levels is far from being straightforward. This underlines the complexity of reference computations on flexible biomolecular fragments. However, we have estimated energy overstabilization contribution in case of significant CH···O interactions (see the AIM analysis below). The capability to establish such contacts also depends on the size of the model system. For a list of potential CH···O contacts, see Table S1.

The C1′···C1′ distances for SPSOM model system are listed in the Supporting Information, Table S2. Note that in real DNA duplex in condensed phase, C1′···C1′ distances represent variables strongly coupled to backbone torsions, glycosidic χ torsion angle, and base pairs stacking parameters (mainly slide, roll, and twist).[3] Although intrastrand C1′···C1′ distances depend on several structural parameters, in B-DNA, the majority of them range between 4.5 and 6.0 Å, which coincides with the present calculations.[3]

For the q-structures, the experimentally determined C1′···C1′ distances are longer than the computed ones with the discrepancy being up to 0.7 Å (Table S2), which is mainly caused by some modest sugar ring adjustments, as illustrated in Figure S1.

Another structural feature worth investigating is the change of the sugar pucker depending on the optimization method employed. While the sugar at the 5′-end has the pucker fixed during the optimization (the δ angle is kept constant; see Tables 1 and 2), the pucker of the 3′-sugar differs for geometries from diverse conformational space regions. While experimentally determined puckers of the a, b, and q-conformations are in the C2′-endo region, 3′-sugar puckers of the optimized a, b, and q-structures are close to C2′-endo, O4′-endo, and C1′-exo, respectively. In both canonical and noncanonical conformational space regions, MP2, TPSS-D, and B3LYP predict similar puckers with only marginal phase angle (P) differences. Sugar puckers predicted by parm99 and parmbsc0 are nearly identical in case of the a- and q-structures with P values below those of QM methods. As in the case of C1′···C1′ distances, parm99 and parmbsc0 3′-sugar puckers differ more significantly in the b-conformational region, for which parmbsc0 pushes the pucker into the O4′-endo domain, while parm99 optimization drives the pucker to the C1′-exo (see the Supporting Information, Table S3). The reason why parm99 and parmbsc0 differ in the 3′-sugar pucker description within the b-conformational space region while they provide virtually the same results for q-conformers is not clear. It may be related to lower ζ values in case of q-conformers leading to relaxation of a strain in the backbone.

Unlike our previous work,[12] the corresponding QM-minimized geometries are qualitatively equal with both B3LYP versus MP2 and TPSS-D versus MP2 RMSD values not exceeding 0.30 Å (q1). This is due to fixation of the torsions during optimizations. The RMSDs between equivalent force field and MP2 optimized geometries are a little bit larger with the maximum value of 0.36 Å for the q1 structure. Because of the extensive backbone fixation, the optimal geometries obtained at different levels of theory differ only marginally (for the largest difference observed between the force field and MP2 geometry, i.e., the q1 structure, see Figure S2).

The maximum difference between the RIMP2/CBS relative energies (all relative energies are calculated with respect to the a1 structure) calculated at the B3LYP and MP2 geometries equals 0.28 kcal mol^{-1} (for the b2 structure), which is ∼3.4% of the relative energy of the respective structure. This value can be considered as the upper limit of uncertainty of the RIMP2/CBS energies introduced by the difference between MP2 and B3LYP geometries. Although the MP2 method accounts better for dispersion interaction than B3LYP, it is more influenced by BSSE. On the other hand, TPSS-D functional combining better description of the dispersion interaction with the small susceptibility to the BSSE (supported by the fact that the dispersion correction is fitted to CBS data[51]) yields geometries similar to those of MP2. Thus, the MP2 geometries optimized using the moderate size 6-31+G(d) basis set are accurate enough, and they are utilized as reference geometries in the present study. Note, however, that the similarity of MP2/6-31+G(d) geometries to the TPSS-D/LP ones still does not guarantee insignificant influence of the intramolecular BSSE. It may also reflect compensation of errors, mainly BSSE versus underestimation of the dispersion energy due to the limited basis set size. Nevertheless, the data suggest that the calculated energetics is not dramatically sensitive to the geometry optimization method.

SPSOM Energies. *Wave Function-Based Methods.* The electronic energies of the SPSOM system conformers were calculated at the MP2/6-31+G(d), CCSD(T)/6-31+G(d), and RIMP2/CBS levels of theory using the MP2/6-31+G(d) minimized geometries. The latter two methods were used to construct the CBS(T) energies, that is, the estimated CCSD(T) energies extrapolated to the CBS limit (see Methods and Model Systems, eq 3). The CBS(T) relative energies represent reference values with which all other methods were compared. The relative energies of the wave function-based methods are given in Table 4 (ordering according to the increasing relative energies of the conformers is listed in the Supporting Information, Table S4).

In the canonical a-region, the MP2/6-31+G(d) results are in a good agreement with the CBS(T) energies with the maximum deviation of 0.21 kcal mol^{-1} for the a10 conformer (Table 4). On the other hand, b-conformers are systematically destabilized at the MP2/6-31+G(d) level of theory by ∼1.50 kcal mol^{-1}. As a result of medium size basis set, this discrepancy might be caused by the intramolecular BSSE, for its magnitude is structure-dependent. Because of slightly more packed geometry of a-conformers as compared to b- and q-structures, it can be anticipated that canonical conformers are more affected by artificially stabilizing BSSE than are b- and q-conformers. This should lead to a systematic destabilization of b- and q-structures with respect to a1, while the relative energetics of a-conformers should remain unaffected (provided that compactness variability within the canonical a-region is marginal). This assumption is supported by the fact that CCSD(T)/6-31+G(d) and MP2/6-31+G(d) relative single-point energies are almost identical to the values of ΔCCSD(T) energy differences being at most

Table 4. Relative Wave Function-Based Energies of the SPSOM Model Conformers Related to the Structure a1[a]

system/ method	MP2/ 6-31+G(d)	CCSD(T)/ 6-31+G(d)	RIMP2/ CBS[b]	CBS(T)[c]
a1	0.00	0.00	0.00	0.00
a2	−0.85	−0.84	−0.82	−0.76
a3	2.71	2.64	2.67	2.61
a4	−0.40	−0.41	−0.34	−0.34
a5	0.18	0.19	0.12	0.14
a6	1.17	1.14	1.18	1.16
a7	0.26	0.24	0.18	0.17
a8	0.69	0.60	0.72	0.66
a9	−0.09	−0.04	−0.10	−0.03
a10	2.40	2.35	2.26	2.19
a11	−0.15	−0.17	0.00	0.02
b1	9.07	8.77	7.83	7.58
b2	9.72	9.48	8.42	8.25
b3	9.49	9.17	8.09	7.82
b4	8.65	8.45	7.38	7.23
b5	9.56	9.19	8.28	7.93
b6	9.71	9.38	8.53	8.25
b7	9.41	9.11	7.98	7.74
b8	8.60	8.34	7.32	7.10
q1	2.34	2.21	1.95	1.87
q2	2.50	2.37	2.01	1.94
q3	2.50	2.34	1.91	1.81

[a] The energies were calculated using the MP2/6-31+G(d) optimized geometries. All energies are given in kcal mol^{-1}. [b] Estimated CBS energies using the extrapolation scheme suggested by Halkier and co-workers[52,53] via aug-cc-pVDZ and aug-cc-pVTZ basis sets (eqs 1 and 2). [c] Estimated CCSD(T) energies extrapolated to CBS according to eq 3.

−0.07, −0.34, and −0.08 kcal mol^{-1} for a, b, and q-conformers, respectively. The effect of inclusion of higher-order CCSD(T) correlation contributions can thus be regarded as negligible.

Among the wave function-based methods, RIMP2/CBS shows the best correlation with the CBS(T) results. The destabilization of b-conformers using the RIMP2/CBS method is increased by 0.1−0.3 kcal mol^{-1} as compared to the CBS(T) reference values. It is of the same order of magnitude as the uncertainty introduced by the choice of level of geometry optimization. This makes RIMP2/CBS a convenient alternative benchmark method, which is much faster than the complete CBS(T) calculation.

DFT-Based Methods. The ability of 10 different functionals to describe energetics of the SPSOM model system was assessed by comparison with the benchmark CBS(T) calculations. Relative energies related to a1 are given in Tables 5, 6, and S7. Ordering of conformers according to their increasing relative CBS(T) energies is listed in the Supporting Information, Tables S5 and S6.

The PBE functional with the triple-ζ quality TZVPP basis set augmented with the Grimme's empirical correction term (PBE-D) gives very good agreement with the CBS(T) calculations. For the a-set of canonical structures, the largest absolute value deviation observed between PBE-D/TZVPP and CBS(T) energies equals 0.68 kcal mol^{-1} for the a10 geometry. Noncanonical b-systems are, as compared to CBS(T), consistently overstabilized on average by 0.18 kcal mol^{-1}. For the q-systems, no significant difference from CBS(T) energies was detected.

The agreement between B3LYP/6-31+G(d) and CBS(T) is significantly worse. Although the largest difference between the B3LYP and CBS(T) relative energies of the a- and q-conformers is −1.09 kcal mol^{-1} (a8), the ordering of these conformers is rather diverse (Tables 5 and S5). The energy separation between a1 and b-conformers on the B3LYP PES is reduced by 1.22 kcal mol^{-1} (b5) to 2.09 kcal mol^{-1} (b3) with respect to the CBS(T) energies.

Table 5. Relative Energies of the SPSOM System Conformers Related to a1[a]

	functional:	PBE-D	B3LYP	mPW2-PLYP	TPSS-D	CBS(T)[b]
system	basis set:	TZVPP	6-31+G(d)	TZVPP	LP[c]	
a1		0.00	0.00	0.00	0.00	0.00
a2		−0.56	−0.99	−0.85	−0.81	−0.76
a3		2.08	2.30	2.53	2.19	2.61
a4		−0.18	−0.07	−0.18	−0.32	−0.34
a5		0.04	−0.32	−0.09	0.07	0.14
a6		0.95	1.03	1.19	0.87	1.16
a7		0.17	0.00	0.01	0.23	0.17
a8		0.63	−0.43	0.25	0.67	0.66
a9		0.13	0.07	0.00	0.17	−0.03
a10		1.51	2.58	2.27	1.86	2.19
a11		0.15	−0.49	−0.15	−0.11	0.02
b1		7.52	6.00	6.78	7.15	7.58
b2		7.88	6.25	7.29	7.80	8.25
b3		7.68	5.73	6.76	7.10	7.82
b4		6.98	5.35	6.27	6.91	7.23
b5		7.90	6.72	7.29	7.45	7.93
b6		8.17	6.46	7.45	7.70	8.25
b7		7.42	6.08	6.92	7.11	7.74
b8		7.12	5.40	6.23	6.77	7.10
q1		1.84	1.80	1.98	1.98	1.87
q2		2.10	2.36	2.37	2.30	1.94
q3		1.49	2.03	2.29	1.81	1.81

[a] Geometries were optimized at MP2/6-31+G(d) level of theory. All energies are given in kcal mol^{-1}. [b] Estimated CCSD(T) energies extrapolated to CBS according to eq 3. [c] LP stands for the 6-311++G(3df,3pd) basis set.

Table 6. Relative Energies of the SPSOM System Conformers Related to a1[a]

	functional:	M06-L	M06	M06-HF	M06-2X	
system	basis set:		6-31+G(d)			CBS(T)[b]
a1		0.00	0.00	0.00	0.00	0.00
a2		−0.79	−1.08	−1.16	−0.79	−0.76
a3		2.40	2.56	2.83	2.60	2.61
a4		−0.17	−0.68	−0.42	−0.25	−0.34
a5		0.28	0.15	0.13	0.14	0.14
a6		1.15	1.10	1.23	1.27	1.16
a7		0.23	0.06	0.14	0.16	0.17
a8		0.87	0.51	0.62	0.95	0.66
a9		0.21	−0.17	−0.21	−0.01	−0.03
a10		2.31	2.24	2.45	2.19	2.19
a11		0.18	−0.30	−0.12	0.24	0.02
b1		8.64	7.86	8.45	8.62	7.58
b2		9.46	8.85	8.97	9.15	8.25
b3		8.88	7.99	8.81	8.84	7.82
b4		8.24	7.55	7.92	8.02	7.23
b5		9.15	8.37	8.98	9.15	7.93
b6		9.33	8.61	9.01	9.28	8.25
b7		8.75	8.24	8.95	8.86	7.74
b8		8.21	7.34	7.97	8.01	7.10
q1		2.28	1.24	2.24	2.44	1.87
q2		2.68	1.58	2.53	2.63	1.94
q3		2.72	1.98	2.95	2.77	1.81

[a] Geometries were optimized at the MP2/6-31+G(d) level of theory. All energies are given in kcal mol^{-1}. Data for the M08 set of functionals are given in the Supporting information, Table S7. [b] Estimated CCSD(T) energies extrapolated to CBS according to eq 3.

The mPW2-PLYP/TZVPP method decreases the relative energy separation between b-conformers and canonical a1 conformation within the range of 0.64 kcal mol^{-1} (b5) to 1.07 kcal mol^{-1} (b3) (Table 5 and Table S5). mPW2-PLYP functional, as compared to CBS(T) results, subtly destabilizes also q-conformers with respect to a1 by 0.11 (q1), 0.44 (q2), and 0.48 (q3) kcal mol^{-1}. The ordering of the a-conformers according to their mPW2-PLYP relative energies nearly coincides with that of CBS(T).

The TPSS-D/LP energies agree well with the CBS(T) ones. The energetic ordering of a-conformers is identical to the reference CBS(T) data with a9 being the only exception. The energies of q1 and q2 relative to the a1 conformer are shifted upward by 0.11 and 0.36 kcal mol^{-1}, while the relative energy of the q3 system is exactly the same as at the CBS(T) level of theory (Table 5). The b-conformers are all overstabilized with respect to CBS(T) by 0.32 kcal mol^{-1} (b4) up to 0.72 kcal mol^{-1} (b3). Separate consideration of the empirical dispersion term (D) indicates lower degree of dispersion stabilization of b-conformations ($D \approx -17.4$ kcal mol^{-1}) when compared to a-conformations ($D \approx -18.8$ kcal mol^{-1}). Thus, neglect of the dispersion term would result in an artificial overstabilization of b-conformers on average by $\sim$1.4 kcal mol^{-1}. The performance of the TPSS functional alone (i.e., without empirical dispersion correction) along with the LP basis set is comparable to hybrid B3LYP/6-31+G(d).

The common feature of the foregoing functionals is the overstabilization of the b-conformers with respect to a1 as compared to CBS(T) results. Analysis of the PBE and TPSS single-point energies, both with and without dispersion correction, suggests that the overstabilization of b-region conformational subspace is, at least partially, due to the insufficient description of the dispersion interactions within the pure exchange-correlation functionals. Although both PBE-D and TPSS-D also slightly overstabilize b-conformers with respect to CBS(T), the overstabilization is markedly smaller in contrast to B3LYP, mPW2-PLYP, as well as PBE and TPSS without dispersion correction. Thus, for correct description of the potential energy landscape of this model system, functionals including dispersion energy contribution (mPW2-PLYP, PBE-D, and TPSS-D) must be employed. However, the high computational costs of the mPW2-PLYP prohibit its practical use, and PBE-D and TPSS-D functionals, being of at least comparable quality, are generally recommended. To quantitatively assess performance of the foregoing functionals, we present basic statistics in Table 7.

Five hybrid meta-GGA (M06, M06-HF, M06-2X, M08-HX, and M08-SO) and one local meta-GGA (M06-L) Truhlar's functionals yield consistent results. Unlike the previously studied functionals (Table 5), all quoted functionals of M06 and M08 suites destabilize b-systems (with respect to the reference CBS(T) calculations) and, with the exception of M06 functional, also the quadruplex q-conformers. The largest contribution to the destabilization probably comes from the HF repulsion. The higher energetic separation between a/b- and a/q-conformers may be explained in case of M06-2X and M06-HF by the overestimation of nonlocal exchange as M06-2X and M06-HF include 54% and 100% of the HF exchange energy, respectively. The a-systems are very well described by the M06 suite of functionals (Tables 6 and S6). While the M06 functional slightly falls behind the other M06-functionals as far as the a-region is concerned, it is markedly superior for a/b and a/q energy difference estimation and thus also for the overall performance. Regarding M08 functionals, the destabilization of b- and q-conformers with respect to a1 is larger as compared to M06 set of functionals. Moreover, the energetic description of a-conformers diverges from the CBS(T) more as compared to the M06 functionals (Table S7). M06 functionals thus appear to be more appropriate for energetic analyses of this kind of compounds than the M08 set of functionals. For statistical evaluation of the DFT methods, see Table 7.

Table 7. Correlation between the Reference CBS(T) and DFT Energies Computed Using MP2/6-31+G(d) Optimized Geometries[a]

				RSoS·n^{-1} [d]			
functional	basis set	r[b]	q[c]	a	b	q	all[e]
PBE-D	TZVPP	0.9979	**0.9736**	0.082	**0.041**	**0.044**	**0.102**
TPSS with D	LP[f]	0.9986	0.9372	0.039	0.256	0.048	0.263
TPSS without D		0.9864	0.7570	0.145	3.971	0.086	3.975
mPW2-PLYP	TZVPP	0.9963	0.8974	0.029	0.763	0.144	0.777
B3LYP	6-31+G(d)	0.9867	0.7872	0.189	3.086	0.076	3.093
M06-L	6-31+G(d)	0.9982	1.1418	0.021	1.204	0.515	1.310
M06		**0.9989**	1.0423	0.034	0.150	0.184	0.240
M06-HF		0.9977	1.1205	0.030	0.827	0.593	1.018
M06-2X		0.9982	1.1322	**0.014**	1.023	0.571	1.172
M08-HX		0.9965	1.1829	0.076	1.996	0.948	2.211
M08-SO		0.9961	1.2086	0.046	2.558	1.366	2.900

[a] a1 conformer is not included in the statistics as its relative energy equals by definition 0.0 kcal mol^{-1} for all methods. The "best" entry in the given column is highlighted. [b] Pearson's product−moment correlation coefficient detecting linear dependencies between CBS(T) and respective DFT energies defined as

$$\frac{N\sum_i E_i^{\mathrm{CBS(T)}}E_i^{\mathrm{DFT}} - \sum_i E_i^{\mathrm{CBS(T)}}\sum_i E_i^{\mathrm{DFT}}}{\sqrt{N\sum_i E_i^{\mathrm{CBS(T)}^2} - \left(\sum_i E_i^{\mathrm{CBS(T)}}\right)^2}\sqrt{N\sum_i E_i^{\mathrm{DFT}^2} - \left(\sum_i E_i^{\mathrm{DFT}}\right)^2}}$$

where N is the number of conformers (21); the reference a1 conformer is excluded. The summations are over all conformers, a1 excluded. [c] Slope of the linear regression line passing through the origin. The least-squares estimate of the slope (q) is defined as:

$$\sum_i E_i^{\mathrm{CBS(T)}}E_i^{\mathrm{DFT}} \cdot \left(\sum_i E_i^{\mathrm{CBS(T)}^2}\right)^{-1}$$

[d] Residual sum of squares (RSoS) divided by the number of conformers (n; $n_a = 10$, $n_b = 8$, $n_q = 3$) belonging to the respective conformational region (a, b, q):

$$\frac{1}{n}\sum_{i,\mathrm{region}}\left(E_i^{\mathrm{DFT}} - E_i^{\mathrm{CBS(T)}}\right)^2$$

[e] $[(\mathrm{RSoS}\cdot n_a^{-1})^2 + (\mathrm{RSoS}\cdot n_b^{-1})^2 + (\mathrm{RSoS}\cdot n_q^{-1})^2]^{1/2}$. [f] LP is short for 6-311++G(3df,3pd) basis set.

Force Field Energies. The correlation between force field (ff) relative energies calculated at the ff-minimized geometries and the CBS(T) energies is shown in Figure 4. In the canonical a-region, both force fields give similar results, which, with the exception of a8, a9, and a10 conformers (Figure 4), agree well with the CBS(T) energies (Table S8). Destabilization of the a10 structure in both force fields may

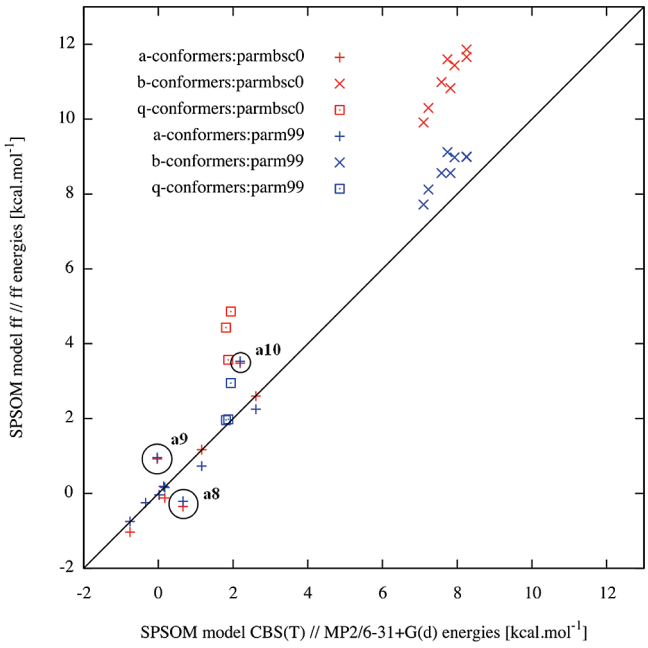

Figure 4. Plot of the correlation between ff//ff and CBS(T)//MP2/6-31+G(d) relative energies (for the corresponding values, see the Supporting Information, Table S8). The RESP charges were fitted to the HF/6-31G(d) electrostatic potential of the most stable conformer a2 (Table S9). The ideal correlation is represented by the black line with the unit slope. Three a-conformation outliers (a8, a9, and a10) are marked with a black circle.

be due to incorrect description of the short-range repulsion by the force field. The structure exhibits contact between H2′ and OP(n+1) atoms, which are as close as 2.4 Å. The sum of O2 and HC atomic types radii is ∼3.15 Å. Another contact occurs between H1′···H5′(n+1) whose distance is about 2.1 Å, while twice a H1 atomic type radius is ∼2.77 Å. Note that the 6-12 Lennard-Jones potential is known to severely exaggerate the short-range interatomic repulsion,[100] while it also has been noticed that the H(C) hydrogens of the Cornell et al. force field are too large.[101] The cause of the energy difference (with respect to CBS(T) results) for the a8 and a9 conformers is not clear. Although both force fields destabilize α/γ = g+/t structures (i.e., b and q-conformers), the parmbsc0 gives at first sight less accurate results. The relative energies of b and q-conformers are shifted upward (i.e., away from the CBS(T) reference values) by 1.6−2.8 kcal mol^{-1} (b-structures) and 1.6−2.5 kcal mol^{-1} (q-structures) with respect to parm99, respectively. It is to be noted, however, that parmbsc0 intentionally penalizes γ-trans geometries as compared to parm99 to prevent their formation in condensed phase molecular simulations.

One of the notorious problems in MM studies is the fact that constant (conformation-independent) atomic charges are used. As the electrostatic potential the charges are fitted on is conformation-dependent, different sets of charges are obtained when derived using different conformers. Thus, to get more insight into the force field performance, we have carried out yet another set of force field calculations, where the charges used for evaluation of the b-conformers were fitted on the HF/6-31G(d) potential of the most stable b-conformer (b8). The charges used to evaluate the a-conformers were kept as before (fitted on the HF/6-31G(d) potential of the a2 system, Table S9). This leads to the a1/b energy separation below ∼6.0 kcal mol^{-1} for parmbsc0 and below ∼3.5 kcal mol^{-1} for parm99, while the reference CBS(T) a1/b separation is 7.1−8.3 kcal mol^{-1}. Fitting charges of b-conformers on the "b-type" electrostatic potential renders parmbsc0 force field superior and supports the basic correctness of the γ-trans penalty of parmbsc0. It also illustrates how sensitive are the force field calculations to the choice of geometry for the derivation of their fixed atomic charges. Our evaluation with different sets of charges roughly corresponds to computations with conformation-dependent charges.

AIM and NBO Analysis. The importance of the weak CH···O hydrogen bonds was assessed by the means of atoms-in-molecules (AIM) Bader analysis. The 6d converged electron density (3;−1) critical points were determined by the AIM analysis of the MP2/6-31+G(d) wave function (see the Supporting Information, Table S10). The electron density and the Laplacian of the electron density threshold for a weak CH···O hydrogen bond was set to 0.01 au. Because of rather small structural differences among structures within the same conformational region (i.e., a, b, and q), we analyzed the first representative out of each region only (a1, b1, and q1). The analysis was done for SPSOM, T3PS, MOSPM, and SPM models (see below for structures of the later three systems).

All identified weak interactions are the so-called CH···O contacts, in which the van der Waals interaction is known to be relatively more important than in standard hydrogen bonds.[102] Because weak CH···O hydrogen bonds were found in canonical a1 system only, we expect their impact exclusively on energetics of the a-conformers. The parameters of critical points found in the remaining b and q-conformers are below the threshold and can thus be regarded as energetically insignificant. To get a basic idea about the extent of stabilization by CH···O hydrogen bonds, interaction energy of a single CH···O contact was also estimated using the AIM analysis. Two different energetic minima for the pucker conformation were localized at the B3LYP/6-31+G(d) and MP2/6-31+G(d) levels of theory in a4, a5, a9, and a11 conformers of the SPM model system (for details, see SPM model system results). The C4′-endo pucker predicted by the MP2/6-31+G(d) calculation in all a-structures allows one to form two almost equally strong (based on AIM) CH···O hydrogen bonds (C2′H···O5′ and C1′H···O5′). The C2′-endo sugar conformation obtained in a4, a5, a9, and a11 structures using B3LYP/6-31+G(d) optimization is stabilized only by the C2′H···O5′ interaction. Calculation of the RIMP2/CBS energies of both optimized geometries allowed one to estimate the energetic contribution (at the RIMP2/CBS level of theory) of one CH···O weak hydrogen bond to be ∼0.6−0.8 kcal mol^{-1}. This difference was indirectly estimated by comparing energies calculated on DFT geometry with C2′-endo pucker and MP2 geometry with C4′-endo. While the former geometry has one CH···O contact, the latter has two.

The only potentially biologically relevant CH···O interaction detected in our model systems is the C2′H···O5′ contact.

Table 8. Off-Diagonal Fock Matrix Elements (*F*, au) Characterizing the Delocalization Effects along the O5′−C5′−C4′−O4′ Bonds in a1 and b1 Structures[a]

	off-diagonal Fockian value			
	a1		b1	
direction	SPSOM	SPM	SPSOM	SPM
n(O4′)→σ*(C4′−C5′)	0.070	0.087	0.033, 0.030	0.035
σ(C4′−C5′)→Ryd(O4′)	0.049	0.046	0.043	0.039
n(O5′)→σ*(C4′−C5′)	0.050	0.049	0.033, 0.061	0.032, 0.062
σ(C4′−C5′)→Ryd(O5′)	0.034	0.039	0.034, 0.032	0.034, 0.031

[a] The results were obtained using the HF-wave functions at the MP2/6-31+G(d) optimized geometries. Two values listed in the same entry refer to two acceptor orbitals centered on the same atom (pair).

The C2′H···O5′ distance is quite frequently around 3.3 Å or even shorter in the B-DNA X-ray structures, including some nucleotides in ultrahigh resolution structures (e.g., ref 103; X-ray structure of a single chain of B double helix resolved at 0.74 Å resolution; the C2′···O5′ distances of DG-4 and DG-9 residues are 3.2 and 3.0 Å, respectively). Occurrence of this interaction in gas-phase computations has been noticed several times.[17,18,21] We nevertheless suppose its rather small impact on conformational preferences of the sugar−phosphate backbone in real environment as the experimental B-DNA C2′···O5′ distances are generally longer than 2.9−3.0 Å seen in gas-phase computations (Table S10). Its effect on the gas-phase energetics should be taken into account.

NBO analysis was carried out to obtain additional insight into destabilization of the b1 structure relative to a1 using the SPSOM and SPM model systems. The unconstrained sugar (i.e., the 3′-sugar of the SPSOM model and the only sugar residue of the SPM model system) in the b1 structure adopts the noncanonical O4′-endo conformation. The same sugar in the a1 structure remains in canonical C2′-endo (SPSOM) or flips to C4′-endo (SPM) conformation. The strongly stabilizing n(O4′)→σ*(C4′−C5′) hyperconjugation in the a1 structure is made impossible in the b1 conformation by O4′ atom pushed out of the C1′,C2′,C3′,C4′ plane. This conformational change is driven by the orbital interactions between n(O5′) and σ*(C4′−C5′), which induces a minor twist of the 5-membered ring along the O4′−C4′ bond due to the repulsion between the electron-rich C4′−C5′ bond and the lone pairs at O4′. The characteristic orbital delocalizations acting in a1 and b1 systems are listed in Table 8. Note that while the introduced stereoelectronic orbital interactions certainly contribute to stabilization of the a1 structure versus b1, we do not suggest that we can in this manner explain the whole energy difference between a- and b-conformers (7.1−8.3 kcal mol^{-1} for the SPSOM model at the CBS(T) level of theory). Note that the empirical force fields, inherently incapable of capturing QM effects, also destabilize b-conformers. In case of the force field, however, we should take into consideration the uncertainty introduced be the fixed atomic charges. When the charges are fitted to reproduce the electrostatic potential in the canonical a-region, they will necessarily introduce error in the b-region description, which could incidentally compensate for the missing electronic structure effects. The results reported above with the charges derived for the b-region geometry indirectly support this possibility. Because of the complex stereoelectronic effects, sugar conformations represent a major challenge for force field derivation. We have recently substantially reparametrized the Cornell et al. force field χ-torsion[104] to prevent ladder-like degradation in long RNA simulations.[105] However, we were still not capable to obtain a fully balanced simultaneous description of pucker and the χ-torsion.

Table 9. List of the Constrained Backbone Torsion Angles[a]

model system	fixed torsion angles
SPSOM-NCH2	γ+1,[b] β+1,[c] α+1,[d] ζ,[e] ε,[f] δ[g]
T3PS	γ+1, β+1, α+1, ζ, ε
MOSPM	γ+1, β+1, α+1, ζ
SPM	γ+1, β+1, α+1, ζ

[a] Values of the fixed torsions are given in Tables 1 and 2. [b] γ+1: O5′(*i*+1)−C5′(*i*+1)−C4′(*i*+1)−C3′(*i*+1). [c] β+1: P(*i*+1)−O5′(*i*+1)−C5′(*i*+1)−C4′(*i*+1). [d] α+1: O3′(*i*)−P(*i*+1)−O5′(*i*+1)−C5′(*i*+1). [e] ζ: C3′(*i*)−O3′(*i*)−P(*i*+1)−O5′(*i*+1). [f] ε: C4′(*i*)−C3′(*i*)−O3′(*i*)−P(*i*+1). [g] δ: C5′(*i*)−C4′(*i*)−C3′(*i*)−O3′(*i*).

Other Model Systems. Geometry optimizations of the remaining model systems (except of SPSOM-NCH2, see below) were carried out at the MP2/6-31+G(d) and B3LYP/6-31+G(d) levels of theory with the backbone torsions (Table 9) fixed at values listed in Tables 1 and 2. Their energies were compared at the RIMP2/CBS//MP2/6-31+G(d) level of theory (Supporting Information, Table S11).

The decision whether to fix the sugar pucker throughout geometry optimization or not is important. Restriction of the pucker to a defined region is advantageous for model comparison and for reduction of the number of potential CH···O interactions. On the other hand, relaxation of the sugar allows one to avoid possible steric conflicts, which may bias PES. To estimate the energetic bias induced by different pucker types, the a1, b1, and q1 conformers of SPSOM, T3PS, MOSPM, and SPM model systems were reoptimized with the sugar(s) kept at the C2′-endo conformation. As the largest change in the relative energies due to the pucker fixation was found to be ∼1.0 kcal mol^{-1} at the MP2/6-31+G(d) level of theory, the bias introduced by not imposing pucker constraints can be considered as acceptable. In the present work, we decided to relax the sugar pucker(s) of the T3PS, MOSPM, SPM, and SPSOM-NCH2 models during the optimization process. We do not claim that letting sugar pucker relax during minimization process is the correct practice as both options have their pros and cons. The estimated error in the relative energies introduced by our decision not to fix the pucker(s) is ∼1.0 kcal mol^{-1} and below.

T3PS Model System. The 3′-sugar (Figure 3) in the T3PS model adopts the C4′-endo conformation in a1, a7, a8, and a10 structures. These four conformers are probably stabilized by C1′H···O5′ contact, in addition to the C2′H···O5′ interaction, which is typical of all a-conformers. The a3 and a6 systems do not adopt C4′-endo pucker due to low value of γ+1 torsion angle (40° for a3 and 35° for a6), preventing formation of the C1′H···O5′ contact (Figure 5). The C4′-endo conformation adoption is also precluded in a2, a4, and a11 conformers because it would likely lead to H1′(*n*+1)/

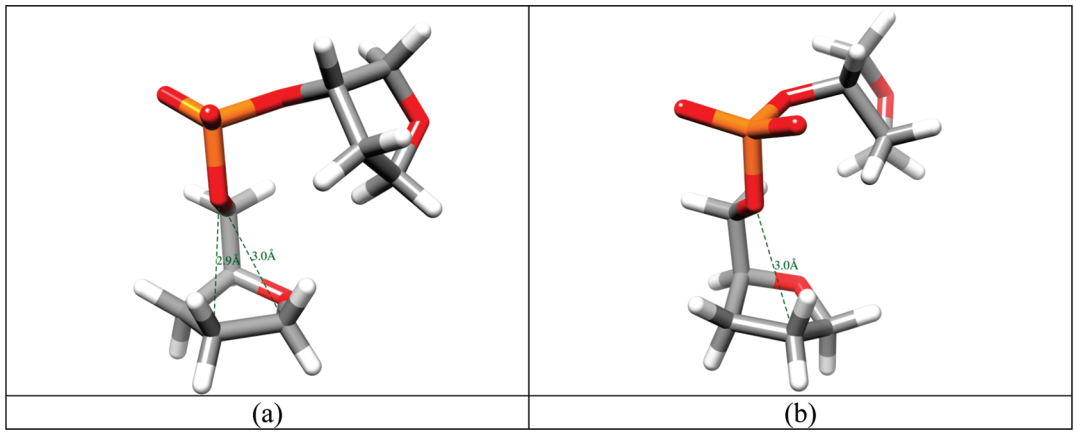

Figure 5. Two 3′-sugar pucker conformations observed in canonical structures of the T3PS model system: (a) C4′-endo pucker of the a1 conformer enabling simultaneous formation of C2′H···O5′ (C···O distance 2.9 Å) and C1′H···O5′ (C···O distance 3.0 Å) contacts. (b) C2′-endo pucker of the a6 conformer enabling a single C2′H···O5′ interaction (C···O distance 3.0 Å).

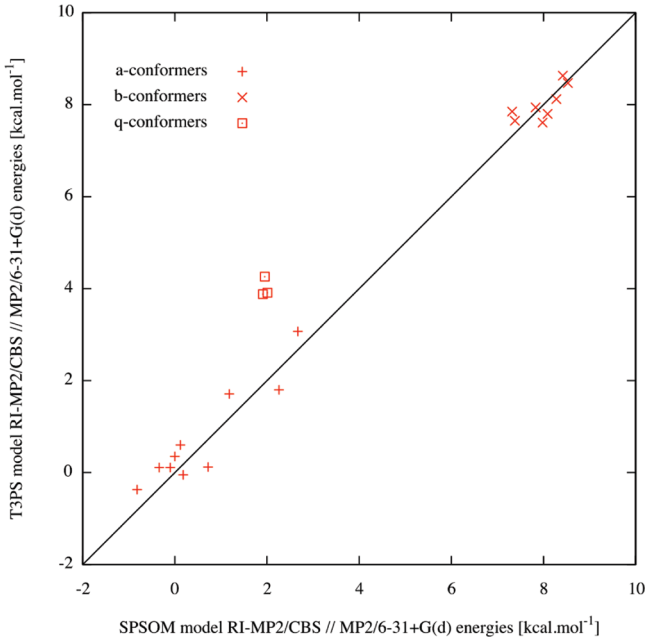

Figure 6. The correlation plot of the RIMP2/CBS//MP2/6-31+G(d) energies between T3PS and SPSOM models (Table S11). The ideal correlation is represented by the black line with the unit slope.

H2′ steric clash. The reason why the 3′-sugar residue of the remaining a-conformers (a5 and a9) do not adopt C4′-endo pucker is not obvious. (Note that the optimizations start from parent structures with C2′-endo arrangement.) The lack of the C1′H···O5′ attractive interaction likely destabilizes (relatively to a1) the respective T3PS a-conformers not adopting the C4′-endo pucker by ∼0.5 kcal mol^{-1} when compared to the SPSOM model (Table S11 and Figure 6). The only overstabilized T3PS conformers with respect to SPSOM (i.e., below diagonal in Figure 6) are those with C4′-endo 3′-sugar conformation, which is clear evidence of the C1′H···O5′ interaction. Moreover, the a1 conformer is affected by C2′H···OP(n+1) interaction whose biological relevancy is arguable. We presume that the alteration of the 3′-sugar pucker within the canonical (i.e., a-conformer) conformational region of T3PS is caused by replacing methoxy group on C3′ present in SPSOM and MOSPM models with a hydrogen atom as no such pucker variations have been observed in SPSOM and MOSPM (see below) models. This observation indicates that the T3PS model system is electronically incomplete and stresses the necessity of a longer backbone fragment at the 3′-end. Because nucleobases are attached via C1′ to the sugar ring, the C1′H···O5′ interaction is a consequence of the lack of nucleobases in our model systems and cannot occur in real DNA.

The increase of energy difference between a1 and q-conformations by ∼1.9−2.3 kcal mol^{-1} in T3PS with respect to SPSOM data (Figure 6 and Table S11) can be, at least partially, explained by the difference in the number and strength of CH···O hydrogen bonds detected in a- and q-conformers.

Although correlation of RIMP2/CBS energies between T3PS and SPSOM models depicted in Figure 6 is rather good, the above discussion clearly shows that the two models are not equivalent.

MOSPM Model System. For the MOSPM system, the B3LYP/6-31+G(d) and MP2/6-31+G(d) geometry optimizations give very similar geometries. The removal of the 5′-sugar moiety in MOSPM model eliminates some CH···O interactions (e.g., C2′H···OP(n+1) observed in a1 of the T3PS model) that are described differently by B3LYP or MP2 methods. Unlike the T3PS (and also SPM, see below) model, the optimization of the parent mospm_x (x = a, b, q1, q2 and q3) structures retains the pucker of the 3′-sugar residue. This is probably due to the methoxy group on C3′, which significantly alters the chemical environment and, as far as a-conformers are concerned, also stiffens the sugar ring while prohibiting the C4′-endo pucker formation.

The approximately same difference in the number and quality of the CH···O interactions in b- and q-conformers with respect to the a1 structure for both MOSPM and SPSOM model systems is responsible for high correlation of RIMP2/CBS MOSPM energies with the SPSOM model for all conformational domains (Figure 7). The a10 conformer constitutes the only outlier and is overstabilized in the MOSPM model (Table S11). The a10 conformer is ∼1.8−2.3 kcal mol^{-1} above a1 in the double-sugar residue model

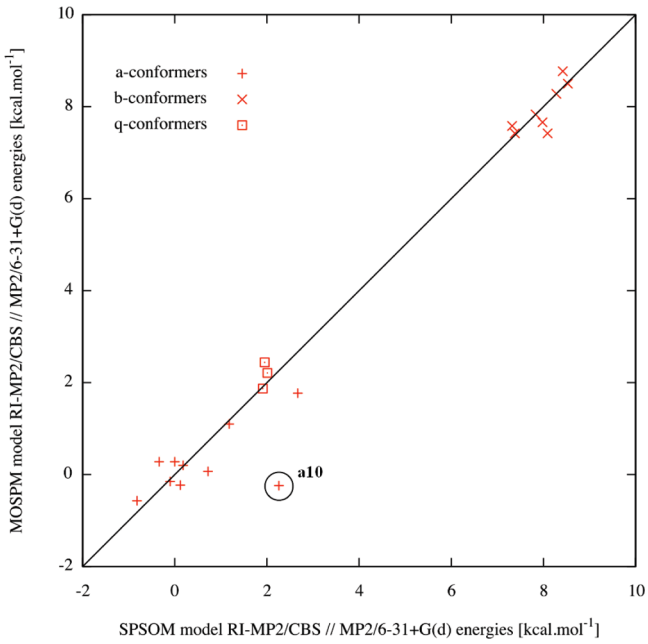

Figure 7. The correlation plot of the RIMP2/CBS//MP2/6-31+G(d) energies between MOSPM and SPSOM models (Table S11). The ideal correlation is represented by the black line with the unit slope. The a10 outlier is marked with a black circle and is discussed in the text.

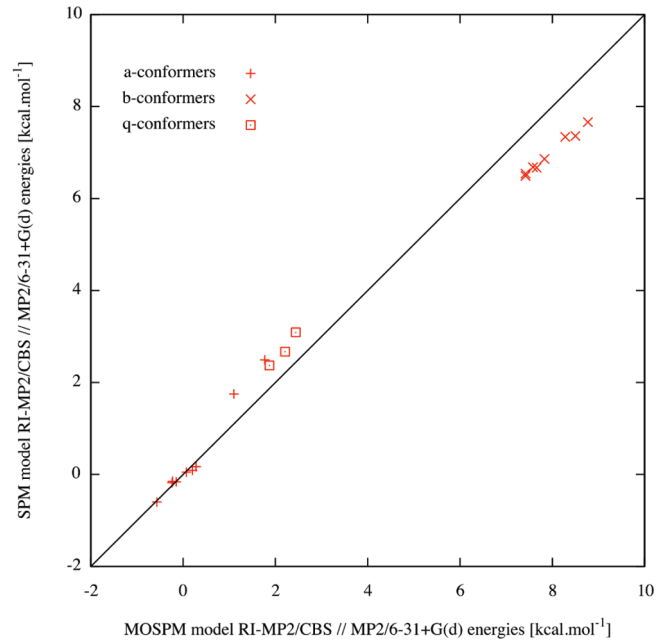

Figure 8. The correlation plot of the RIMP2/CBS//MP2/6-31+G(d) energies between SPM and MOSPM models (Table S11). The ideal correlation is represented by the black line with the unit slope.

systems (SPSOM and T3PS), while in single-sugar models (MOSPM and SPM) it is ∼0.2 kcal mol^{-1} below the a1 conformer. The reason for the destabilization of a10 in SPSOM and T3PS models is the low value of the ε torsion angle (160°, while the canonical value is ∼180°), which puts H2′ and OP(n+1) as close as 2.4 Å and H1′ and H5′(n+1) into a distance of 2.1 Å. As both the H1′ and the H2′ are attached to the 5′-sugar moiety, its substitution by a methyl group releases the repulsion. Thus, the only outlier does not reflect inconsistency of MOSPM and SPSOM model systems because the ε torsion is not defined in the MOSPM model and thus cannot be taken into account in the comparison. We suggest that the SPSOM and MOSPM model systems are equivalent as far as the γ+1, β+1, and α+1 torsions are concerned.

SPM Model System. B3LYP and MP2 optimizations of the SPM model system provide two groups of conformers. While minimum structures of the first group are independent of the used level of theory, B3LYP and MP2 methods yield different conformations differing in the sugar pucker in the second group consisting of a4, a5, a9, a11, and q1 conformers. In the first group, both B3LYP and MP2 minimizations led consistently to either C4′-endo, C2′-endo, or O4′-endo sugar pucker. The C4′-endo pucker (in a1, a2, a7, a8, and a10 systems) arises due to the presence of stabilizing C1′H···O5′ interaction. The low value of γ+1 torsion prevents the formation of this interaction in a3 (γ+1 = 40°) and a6 (γ+1 = 35°) conformers, in which the C2′-endo remains preserved. The O4′-endo pucker of the b- and q-conformers is induced by minimizing the overlap of the lone pair of O4′ with electron-rich σ*(C4′−C5′) bond (see NBO analysis).

The second group consists of a4, a5, a9, a11, and q1 structures. In the case of a-conformers, MP2 yields C4′-endo sugar pucker, while B3LYP optimization preserves the pucker in the C2′-endo region. This can be explained by the change in the fixed β+1 torsion from canonical 180° (a1) to 170° (a4), 190° (a5), and 170° (a11) values, which is probably accompanied by weakening of the C1′H···O5′ interaction that is sensed by the MP2 method but not by the DFT. The reason why B3LYP-optimized a9 structure does not adopt C4′-endo pucker is not clear as the β+1 torsion is fixed at the canonical value of 180°. The sugar pucker of the q1 conformer, which is in O4′-endo conformation at the B3LYP minima, is shifted toward the C1′-exo region at the MP2 PES. However, the reason for this behavior is not obvious.

The effect of the methoxy group was studied by comparing the RIMP2/CBS relative energies of the MP2 optimized conformers of SPM and MOSPM models (Figure 8). The energy ordering relative to the canonical conformation a1 is the same in both models (Table S11). The destabilization of a3, a6, and q-conformers (Table S11) with respect to a1 in the SPM system by ∼0.4−0.7 kcal mol^{-1} is due to the lack of the biologically irrelevant C1′H···O5′ interaction present in a1 SPM but absent in the five specified SPM structures. No such contact has been detected in the a1 MOSPM. Although relative energies of the b-conformers were expected to be shifted upward in the SPM model for the same reason as the a3, a6, and q-conformers, they are overstabilized with respect to MOSPM by ∼0.5 kcal mol^{-1} (Table S11, Figure 8). We ascribe this to the methoxy group, as it is the only difference between these two model systems. Note that apart from the methoxy group, both parent geometries (mospm_x and spm_x) from which optimization was initiated are identical. Both have the sugar pucker in the C2′-endo region. This indicates that the methoxy group effectively prevents the sugar pucker to adopt C4′-endo conformation, and it

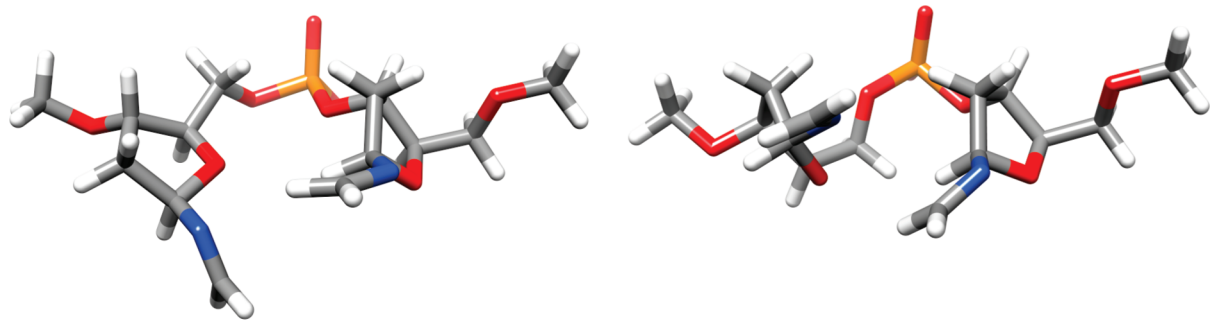

Figure 9. Two interaction modes of the $-N{=}CH_2$ groups: (a) T-shape-like interaction mode, and (b) stack-like interaction mode.

should be an inherent component of any model system for backbone computations.

The correlation of the RIMP2/CBS relative energies of the b- and q-structures between SPM versus MOSPM and SPM versus SPSOM model systems is similar. The q-conformers are destabilized, and the b-conformers are systematically overstabilized in case of SPM model (Table S11). Correlation between MOSPM and SPSOM models is worse. Hence, SPM and SPSOM model systems are not equivalent.

SPSOM-NCH2 Model System. The structures of this model system were optimized at the MP2/6-31+G(d) level of theory only as this model turned out to be inconvenient. The substitution of the C1′ hydrogen atom by the $-N{=}CH_2$ group, which partially mimics the effect of an aromatic nucleobase, leads to the extension of the hyperconjugation network over both sugar residues.[12] However, this system is unsuitable for studying the energetics of the sugar−phosphate backbone. The results of the geometry optimization depend on the initial orientation of the $-N{=}CH_2$ groups. The existence of various interaction modes of the $-N{=}CH_2$ groups (Figure 9) does not allow one to separate the energetic contribution of the methylene-imino groups from that of the sugar−phosphate backbone. It thus illustrates why model systems containing nucleobases (or even their simpler analogues) are not recommended for studying the energetics of the NA backbone. They do not allow to separate the intrinsic backbone preferences from other factors determining their PES.

Conclusions

To gain insight into the intrinsic energetics and electronic structure of the sugar−phosphate backbone, several model systems of 22 relevant DNA backbone conformations from three distinct conformational families were studied in the gas phase by the means of high level ab initio methods. The present study provides a set of accurate structure−energy data for DNA backbone model systems, which can be used as a benchmark database for assessment of other theoretical methods. The most accurate data are obtained at the MP2/CBS level corrected for CCSD(T) term using smaller basis set, that is, using the CBS(T) method. The study leads to the following conclusions:

To maintain the sugar−phosphate backbone to sample relevant conformations and combinations of dihedral angles found in crystal structures of DNA, multiple constraints on the backbone torsion angles have to be imposed. Essentially, it is necessary to fix all dihedral angles at their target values. Albeit fixation of the sugar pucker prevents formation of unnatural CH···O contacts, it may, on the other hand, lead to unnatural tensions biasing energetic analysis. For this reason, fixation of the pucker should be considered and examined from case to case.

The ΔCCSD(T) correction is virtually constant in all studied conformers. Thus, RIMP2 method with sufficiently large basis set (preferably extrapolated to CBS) is adequate for accurate description of nucleic acids backbone.

From the 10 tested DFT approaches, the best results close to the reference CBS(T) calculations are provided by the PBE and TPSS functionals augmented with an empirical dispersion term (PBE-D and TPSS-D), thus stressing the importance of including the dispersion interaction. Very good results were also obtained using the nonlocal meta-GGA M06 functional from the Minnesota M06 suite. The mPW2-PLYP functional also yields reasonable results in accord with CBS(T) reference calculations. Its applicability is, however, limited due to the computational requirements. The remaining M06-type functionals, that is, M06-L, M06-HF, and M06-2X, are of comparable performance. They provide results coinciding with CBS(T) in the canonical a-region but are less accurate in the evaluation of a1 versus b and a1 versus q energy difference. Functionals of the M08 set (M08-HX and M08-SO) are generally inferior to M06 functionals for the energetic analysis of this kind of compounds. The popular B3LYP performs rather unsatisfactorily.

The common attribute of the PBE-D, TPSS-D, and mPW2-PLYP functionals is a slight underestimation of destabilization of b-conformers with respect to the canonical a1 conformation as compared to the CBS(T) data. Neglect of the empirical dispersion terms would result in further stabilization of b-conformers with respect to a1 and thus to a bigger deflection from CBS(T) trend. Performance of pure TPSS/LP (i.e., without dispersion correction) is thus of B3LYP quality.

M06 and M08 functionals, on the other hand, follow the opposite trend as they further destabilize b- and q-conformers with respect to a1 as compared to CBS(T) data.

The intrinsic stability of the noncanonical α/γ = g+/t b-conformers is lower as compared to the canonical a-structures partially due to the deformation of the sugar

conformation, which leads to weakening of the strongly stabilizing n(O4′)→σ*(C4′−C5′) hyperconjugation effect. The conformational change in the 3′-sugar is driven by the orbital interactions between n(O5′) and σ*(C4′−C5′) inducing a minor twist of the sugar ring along the O4′−C4′ bond.

The energetics of the studied model systems is biased by the presence of the network of weak CH···O hydrogen bonds, the majority of which can be considered as the gas-phase artifact. Even though their stabilizing effect balances out to a large degree when structures taken from the same conformational region are confronted, their impact should be considered for comparison of structures from different regions of PES. The number of CH···O interactions and their strength are structure-dependent. Description of these interactions is also method-dependent. These usually undesired interactions greatly complicate reference calculations on fragments of DNA backbone. The only such contact that is occasionally seen in high-resolution B-DNA structures is C2′H···O5′.

The simplification of the SPSOM model system to the MOSPM one has only a marginal impact on the relative energies. Thus, we propose MOSPM as the potentially most appropriate model system for the QM studies of the sugar−phosphate backbone preferences in nucleic acids as a function of backbone torsion angles, excluding ε and ζ torsions. It has several advantages over the other studied model systems (SPSOM, T3PS, SPM, and SPSOM-NCH2): (i) It is smaller than SPSOM, T3PS, and SPSOM-NCH2 systems. (ii) The replacement of the 5′-sugar residue by the methyl group significantly reduces the number of CH···O interactions. (iii) In contrast to SPM, it offers (due to the presence of methoxy group) a more complete description of the electronic structure along the backbone. (iv) The addition of the $-N{=}CH_2$ groups in the SPSOM−NCH2 system does not introduce any advantage, as their presence significantly alters (and complicates) the shape of the potential energy surface and also increases the BSSE artifact. Although inherently incapable to model ε and ζ torsion profiles, the MOSPM system could replace the SPSOM model in future reference studies of δ, γ, β, and α torsions of the DNA backbone.

In future work, we plan to extend the present computations in two directions. One of them is inclusion of the solvent effects, and the other is consideration of other dihedral angle combinations.

Acknowledgment. This work was supported by the Grant Agency of the Academy of Sciences of the Czech Republic grant IAA400040802, Grant Agency of the Czech Republic grants 203/09/1476, P208/10/2302, and 203/09/H046, Ministry of Education of the Czech Republic LC06030, MSM0021622413, LC512, MSM6198959216, and MSM6046-137302, and Academy of Sciences of the Czech Republic, grant nos. AV0Z50040507, AV0Z50040702, and Z40550506. The present study was also financially supported by the Masaryk University, project ID: MUNI/A/0134/2009. We would like to thank Prof. D.G. Truhlar for giving us access to the recent DFT functionals developed by his group. Finally, we would like to thank the Brno MetaCentrum staff for the generous allotment of computer time.

Supporting Information Available: Figure of the C2′···O5′ and C1′···C1′ distances dependency on the method of calculation. Table of potential CH···O interactions. Table of C1′−C1′ distances and 3′-sugar pucker phase angles of the SPSOM model system. Figure illustrating the largest difference observed between the respective optimized geometries of SPSOM. Tables with relative energies of wave function-based, DFT-based, and force fields methods. Tables of RESP charges. List of the (3;−1) critical bond points. Energies and resulting backbone torsion angles of unconstrained optimizations of selected SPSOM conformers. Relative energies of conformers represented by T3PS, MOSPM, and SPM model systems. *xyz* coordinates of five parent structures (i.e., starting geometries, see Table 3) labeled as model_x (x = a, b, q1, q2, and q3) for each single model system (model = SPSOM, T3PS, MOSPM, SPM, and SPSOM-NCH2). *xyz* coordinates of all (22) MP2/6-31+G(d) optimized SPSOM conformers along with their reference CBS(T) energies. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Dickerson, R. E. *J. Mol. Biol.* **1983**, *166*, 419–441.

(2) Calladine, C. R.; Drew, H. R. *J. Mol. Biol.* **1984**, *178*, 773–781.

(3) Packer, M. J.; Hunter, C. A. *J. Mol. Biol.* **1998**, *280*, 407–420.

(4) Packer, M. J.; Dauncey, M. P.; Hunter, C. A. *J. Mol. Biol.* **2000**, *295*, 85–103.

(5) Hartmann, B.; Piazzola, D.; Lavery, R. *Nucleic Acid Res.* **1993**, *21*, 561–568.

(6) Sundaralingam, M. *Biopolymers* **1969**, *7*, 821–860.

(7) Svozil, D.; Kalina, J.; Omelka, M.; Schneider, B. *Nucleic Acid Res.* **2008**, *36*, 3690–3706.

(8) Neidle, S. *Nucleic Acid Structure and Recognition*; Oxford University Press: Oxford, 2002.

(9) Schwartz, T.; Rould, M. A.; Lowenhaupt, K.; Herbert, A.; Rich, A. *Science* **1999**, *284*, 1841–1845.

(10) Šponer, J.; Riley, K. E.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595–2610.

(11) Šponer, J.; Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151.

(12) Svozil, D.; Šponer, J. E.; Marchan, I.; Perez, A.; Cheatham, T. E.; Forti, F.; Luque, F. J.; Orozco, M.; Šponer, J. *J. Phys. Chem. B* **2008**, *112*, 8188–8197.

(13) MacKerell, A. D. *J. Phys. Chem. B* **2009**, *113*, 3235–3244.

(14) Foloppe, N.; MacKerell, A. D. *J. Phys. Chem. B* **1999**, *109*, 10955–10964.

(15) Wang, F. F.; Gong, L.-D.; Zhao, D. *J. Mol. Struct. (THEOCHEM)* **2009**, *909*, 49–56.

(16) Leulliot, N.; Ghomi, M.; Scalmani, G.; Berthier, G. *J. Phys. Chem. A* **1999**, *103*, 8716–8724.

(17) Louit, G.; Hocquet, A.; Ghomi, M. *Phys. Chem. Chem. Phys.* **2002**, *4*, 3843–3848.

(18) Shishkin, O. V.; Gorb, L.; Zhikol, O. A.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2004**, *21*, 537–553.

(19) Palamarchuk, G. V.; Shishkin, O. V.; Gorb, L.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2009**, *26*, 653–661.

(20) Millen, A. L.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. B* **2010**, *114*, 4373–4382.

(21) Hocquet, A. *Phys. Chem. Chem. Phys.* **2001**, *3*, 3192–3199.

(22) Poltev, V. I.; Anisimov, V. M.; Danilov, V. I.; Deriabina, A.; Gonzalez, E.; Garcia, D.; Rivas, F.; Jurkiewicz, A.; Les, A.; Polteva, N. *J. Mol. Struct. (THEOCHEM)* **2009**, *912*, 53–59.

(23) 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.

(24) Varnai, P.; Djuranovic, D.; Lavery, R.; Hartmann, B. *Nucleic Acids Res.* **2002**, *30*, 5398–5406.

(25) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(26) Beveridge, D. L.; Barreiro, G.; Byun, K. S.; Case, D. A.; Cheatham, T. E.; Dixit, S. B.; Giudice, E.; Lankaš, 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.

(27) Varnai, P.; Zakrzewska, K. *Nucleic Acids Res.* **2004**, *32*, 4269–4280.

(28) Barone, F.; Lankaš, F.; Špačková, N.; Šponer, J.; Karran, P.; Bignami, M.; Mazzei, F. *Biophys. Chem.* **2005**, *118*, 31–41.

(29) Perez, A.; Marchan, I.; Svozil, D.; Šponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.

(30) Heddi, B.; Foloppe, N.; Oguey, Ch.; Hartmann, B. *J. Mol. Biol.* **2008**, *382*, 956–970.

(31) Fadrná, E.; Špačková, N.; Sarzynska, J.; Koča, J.; Orozco, M.; Cheatham, T. E., III; Kulinski, T.; Šponer, J. *J. Chem. Theory Comput.* **2009**, *5*, 2514–2530.

(32) Beššeová, I.; Otyepka, M.; Réblová, K.; Šponer, J. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10701–10711.

(33) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402–5415.

(34) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876–880.

(35) 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 E.01; Gaussian, Inc.: Wallingford, CT, 2004.

(36) Jensen, F. *Introduction to Computational Chemistry*; Wiley: New York, 2006; Chapter 5, pp 225−227.

(37) Dunning, T. H. *J. Phys. Chem. A* **2000**, *104*, 9062–9080.

(38) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.

(39) Reiling, S.; Brickmann, J.; Schlenkrich, M.; Bopp, P. A. *J. Comput. Chem.* **1996**, *17*, 133–147.

(40) Jensen, F. *Chem. Phys. Lett.* **1996**, *261*, 633–636.

(41) Senent, M. L.; Wilson, S. *Int. J. Quantum Chem.* **2001**, *82*, 282–292.

(42) Balabin, R. M. *J. Chem. Phys.* **2008**, *129*, 164101.

(43) Asturiol, D.; Duran, M.; Salvador, P. *Chem. Phys.* **2008**, *128*, 144108.

(44) Asturiol, D.; Duran, M.; Salvador, P. *J. Chem. Theory Comput.* **2009**, *5*, 2574–2581.

(45) Valdes, H.; Klusak, V.; Pitonak, M.; Exner, O.; Stary, I.; Hobza, P.; Rulisek, L. *J. Comput. Chem.* **2008**, *29*, 861–870.

(46) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(47) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(48) Weigend, F.; Haser, M. *Theor. Chim. Acta* **1997**, *97*, 331–340.

(49) Weigend, F.; Haser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143–152.

(50) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401–146405.

(51) Jurečka, P.; Černý, J.; Hobza, P.; Salahub, D. R. *J. Comput. Chem.* **2007**, *28*, 555–569.

(52) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Olsen, J. *Chem. Phys. Lett.* **1999**, *302*, 437–446.

(53) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639–9646.

(54) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(55) Dunning, T. H., Jr. *J. Phys. Chem. A* **2000**, *104*, 9062–9080.

(56) Kendall, R. A.; Fruüchtl, H. A. *Theor. Chim. Acta* **1997**, *97*, 158–163.

(57) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359–363.

(58) Vahtras, O.; Almlof, J.; Feyereisen, M. W. *Chem. Phys. Lett.* **1993**, *213*, 514–518.

(59) Hobza, P.; Šponer, J. *J. Mol. Struct. (THEOCHEM)* **1996**, *388*, 115–120.

(60) Jurečka, P.; Hobza, P. *Chem. Phys. Lett.* **2002**, *365*, 89–94.

(61) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887–10893.

(62) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200–10207.

(63) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656–10668.

(64) *MOLPRO, version 2006.1: A package of Ab initio Programs*; Cardiff University: Cardiff, U.K., 2006.

(65) Kristyan, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, *229*, 175–180.

(66) Hobza, P.; Šponer, J.; Reschel, T. *J. Comput. Chem.* **1995**, *16*, 1315–1325.

(67) Lacks, D. J.; Gordon, R. G. *Phys. Rev. A* **1993**, *47*, 4681–4690.

(68) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664–675.

(69) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2003**, *367*, 778–784.

(70) Misquitta, A. J.; Jeziorski, B.; Szalewicz, K. *Phys. Rev. Lett.* **2003**, *91*, 033201.

(71) Zhao, Y.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2701–2705.

(72) Hepburn, J.; Scoles, G. *Chem. Phys. Lett.* **1975**, *36*, 451–456.

(73) Ahlrichs, R.; Penco, R.; Scoles, G. *Chem. Phys.* **1977**, *19*, 119–130.

(74) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(75) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

(76) Weigend, F.; Haser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143–152.

(77) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(78) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 13126–13130.

(79) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 1849–1868.

(80) Trindle, C.; Shillady, D. *Electronic Structure Modeling*; CRC Press, Taylor & Francis Group: Boca Raton, FL, 2008.

(81) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.

(82) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157–167.

(83) Grimme, S.; Schwabe, T. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4398–4401.

(84) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

(85) Nesse, F. Orca 2.6: An ab initio, DFT and semiempirical SCF-MO package.

(86) Bader, R. F. W. *Atoms in Molecules. A Quantum Theory*; Oxford University Press: Oxford, U.K., 1990.

(87) Bader, R. F. W. *Chem. Rev.* **1991**, *91*, 893–928.

(88) Bader, R. F. W. *J. Phys. Chem. A* **1999**, *103*, 304–314.

(89) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211–7218.

(90) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1983**, *78*, 4066–4073.

(91) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735–746.

(92) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 1736–1740.

(93) Carpenter, J. E.; Weinhold, F. *J. Mol. Struct. (THEOCHEM)* **1988**, *46*, 41–62.

(94) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899–926.

(95) Weinhold, F.; Carpenter, J. E. In *The Structure of Small Molecules and Ions*; Naaman, R., Vager, Z., Eds.; Plenum: New York, 1988; pp 227−36.

(96) Biegler-Konig, F.; Schonbohm, J.; Bayles, D. *J. Comput. Chem.* **2001**, *22*, 545–559.

(97) Biegler-Konig, F.; Schonbohm, J. *J. Comput. Chem.* **2002**, *23*, 1489–1494.

(98) Bayly, C. I.; Ciepak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(99) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER 10*; University of California: San Francisco, CA, 2008.

(100) Morgado, C. A.; Jurečka, P.; Svozil, D.; Hobza, P.; Šponer, J. *J. Chem. Theory Comput.* **2009**, *5*, 1524–1544.

(101) Warmlander, S.; Šponer, J. E.; Šponer, J.; Leijon, M. *J. Biol. Chem.* **2002**, *277*, 28491–28497.

(102) Desiraju, G. R. *Acc. Chem. Res.* **2002**, *35*, 565–573.

(103) Kielkopf, C. L.; Ding, S.; Kuhn, P.; Rees, D. C. *J. Mol. Biol.* **2000**, *296*, 787–801.

(104) Banáš, P.; Hollas, D.; Zgarbová, M.; Jurečka, P.; Orozco, M.; Cheatham, T. E., III; Šponer, J.; Otyepka, M. *J. Chem. Theory Comput.*, in press.

(105) Mlýnský, V.; Banáš, P.; Hollas, D.; Réblová, K.; Walter, N. G.; Šponer, J.; Otyepka, M. *J. Phys. Chem. B* **2010**, *114*, 6642–6652.

CT1004593