Comparison of Intrinsic Stacking Energies of Ten Unique Dinucleotide Steps in A-RNA and B-DNA Duplexes. Can We Determine Correct Order of Stability by Quantum-Chemical Calculations?

Daniel Svozil,*,† Pavel Hobza,^{‡,§} and Jiří Šponer*,‡,||

Faculty of Chemical Technology, Laboratory of Informatics and Chemistry, Institute of Chemical Technology, Technická 3, 166 28, Prague 6, Czech Republic, Institute of Organic Chemistry and Biochemistry & Center for Biomolecules and Complex Molecular Systems, Academy of Sciences of the Czech Republic, Flemingovo náměstí 2, 166 10 Prague 6, Czech Republic, Department of Physical Chemistry, Palacký University, Třída Svobody 26, 771 46 Olomouc, Czech Republic, and Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65, Brno, Czech Republic

Received: November 12, 2009

High level ab initio methods have been used to study stacking interactions in ten unique base pair steps both in A-RNA and in B-DNA duplexes. The protocol for selection of geometries based on molecular dynamics (MD) simulations is proposed, and its suitability is demonstrated by comparison with stacking in steps at fiber diffraction geometries. It is shown that fiber diffraction geometries are not sufficiently accurate for interaction energy calculations. In addition, the protocol for selection of geometries based on MD simulations allows for the evaluation of the variability of the intrinsic stacking energies along the MD trajectories. The uncertainty in stacking energies (difference between the most and least stable geometry) due to the dynamical nature of systems can be, in some cases, as large as 3.0 kcal·mol⁻¹, which is almost 50% of the actual sequence dependence of base stacking energies (the energy difference between the most and least stable sequences). Thus, assessing the relative magnitude of the gas phase stacking energy using a single geometry for each sequence is insufficient to obtain an unambiguous order of gas phase stacking energies in canonical double helices. Though the ordering of ten unique dinucleotide steps cannot be definitive, some general conclusions were drawn. The stacking energies of base pair steps in A-RNA are more evenly separated compared to B-DNA, and their ordering is less sensitive to the dynamics of the system compared to be B-DNA. The most stable step both in B-DNA and A-RNA is the CG/CG step that is well separated from the second most stable step GC/GC. Also the least stable step (the CC/GG step) is well separated from the rest of the structures. The calculations further show that B-DNA stacking is favorable only marginally (on average by 1.14 kcal·mol⁻¹ per base pair step) over A-RNA stacking, and this difference vanishes after subtracting the stabilizing van der Waals effect of the thymine 5-methyl group that is absent in RNA. Basically, no correlation between the sequence dependence of gas phase stacking energies and the sequence dependence of ΔG°_{37} free energies used in nearest-neighbor models was found either for B-DNA or for A-RNA. This reflects the complexity of the balance of forces that are responsible for the sequence dependence of thermodynamics stability of nucleic acids, which masks the effect of the intrinsic interactions between the stacked base pairs.

Introduction

The basic structural form of DNA is a right handed double helix, with about ten nucleotide pairs per helical turn. Each helical strand, composed of a sugar phosphate backbone and attached bases, is connected to a complementary strand by hydrogen bonding between paired bases: adenine (A) with thymine (T) and guanine (G) with cytosine (C). Due to the inherent conformational flexibility of the polynucleotide backbone there exist a wide range of different double helical conformations. The most common form is the antiparallel right-handed B-DNA double helix. Another possible right-handed form (A-DNA) is similar to B-DNA, but with a shorter and

* Corresponding authors. E-mail: daniel.svozil@gmail.com (D.S.),

more compact helical structure. The A-form is the structure adopted by double-stranded regions in RNA. The uniform A-RNA and B-DNA structures that are widely used for the model building purposes, the so-called canonical structures, were derived from fiber diffraction studies.¹

Each particular architecture of nucleic acids (NA) is a result of a balance of a number of interactions of the individual NA building blocks. The following interactions are especially important: (i) hydrogen bonding between aromatic nucleic acid bases, with the major extension to ribose interactions in RNAs, (ii) stacking interactions between nucleic acid bases; (iii) interactions between charged phosphate groups, the neutralizing cations and water molecules; (iv) environmental effects such as salt effects, presence of divalent metal cations, specific hydration, and hydrophobic interactions. All these interactions play an important role in providing stability and conformational variability of nucleic acids and a wide range of experimental and theoretical techniques is available for their investigation.

sponer@ncbr.chemi.muni.cz (J.S.).

† Institute of Chemical Technology.

[‡] Institute of Organic Chemistry and Biochemistry & Center for Biomolecules and Complex Molecular Systems, Academy of Sciences of the Czech Republic.

[§] Palacký Úniversity.

Institute of Biophysics, Academy of Sciences of the Czech Republic.

Proper understanding of the role of molecular interactions in nucleic acids belongs to important goals of modern physical chemistry, biochemistry, and structural molecular biology. It is a difficult task, considering the astonishing complexity of biomolecular interactions and the tiny balance of all the contributing noncovalent forces. This requires studying the interactions at different levels simultaneously, ranging from in depth understanding of the physicochemical origin and magnitude of the intrinsic (direct) molecular interactions and forces, through description of their mutual interplay and interactions with environment up to correct interpretation of their ultimate appearance and exact roles in all the different structural and functional contexts.

Experimental studies of the intrinsic interactions of NA bases and other building blocks require gas phase experiments, in which the effects of the environment are eliminated. The spectroscopy experiments²⁻⁶ of the selected base pairs provide unique evidence about their IR spectra but do not allow extracting information about their structures or stabilization energies. On the other hand, theoretical quantum chemical studies provide a full description of molecular interactions including data on geometry, stabilization energy, electric properties, IR, visible and UV spectra, selected NMR parameters, etc. However, to obtain reliable data, the calculations should be carried out at very high levels of quantum chemical ab initio theory, minimizing problems with, e.g., the size of the AO (atomic orbitals) basis set or incompleteness of the inclusion of electron correlation effects, 7,8 as demonstrated in several recent studies of π - π stacking interactions. ⁹⁻²⁹ Such calculations became available only in recent years.

The first medium quality correlated ab initio calculations for DNA base pairs were performed in the mid-1990s. Calculations performed with the second-order Møller—Plesset method using medium-sized AO basis set containing diffuse polarization functions provided meaningful results for base pairing,³⁰ stacking,³¹ for interactions of DNA bases and base pairs with metal cations^{32,33} as well as for interactions of DNA base pairs with drugs.^{34,35} With the advent of fast and reliable quantum chemical methods, such as RI-DFT-D (resolution of identity density functional theory augmented with dispersion term),^{36,37} the size of the system can be considerably increased, representing thus a major progress in ab initio description of nucleic acid systems.

In this paper, a systematic QM study of the stacking of nucleic acid bases in B-DNA and A-RNA is presented. The protocol for selection of geometries based on MD simulations is proposed, and its suitability is demonstrated by comparison with stacking in steps at fiber diffraction geometries. The problem of choice of proper geometries for stacking calculations is thoroughly studied, and the inferiority of often used fiber diffraction geometries is demonstrated. The uncertainty in stacking energies due to the dynamical nature of the system is assessed, and the problem of stacking energy ordering of ten unique dinucleotide steps both in A-RNA and in B-DNA is discussed. The stacking energies are compared to experimentally derived ΔG°_{37} free energies (1 M NaCl, pH 7) used^{38,39} in nearest-neighbor models. The results provide a valuable picture of stacking interactions in B-DNA and in A-RNA duplexes.

Methods

The essential requirement for obtaining meaningful ab initio interaction energies is to use a set of well-defined geometries that are relevant to the situation in real NA systems. One common way⁴⁰ is to use geometries derived from fiber diffraction data.¹ Such geometries are available in packages such as

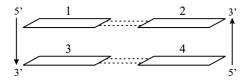
nucgen, 41 3DNA, 42 or NAB. 43 While these geometries are correct globally, they have values of certain backbone torsion angles (e.g., $\delta=143^\circ$, $\varepsilon=219^\circ$, $\zeta=199^\circ$ in B-DNA) that are at odds with atomic-resolution crystal structures of nucleic acids 44 (the values obtained from B-DNA X-ray structures analysis are $\delta=128^\circ$, $\varepsilon=184^\circ$, $\zeta=265^\circ$). Likewise, the stacking geometries may be imperfect. Moreover, fiber geometries are sequence-independent, i.e., identical for each base pair step sequence 45,46 (by base pair step we mean two consecutive base pairs along the double helix). This may produce substantial errors in energy calculations for many base pair steps and thus fiber diffraction data are unsuitable for stacking energy calculations.

Another option is to use experimental X-ray coordinates of selected structures. 47-49 This is definitely a viable approach for the highest resolution and carefully refined structures (such as those of B-DNA decamers), provided the experimental nucleobase geometries are overlaid by QM-optimized monomers. 50,51 However, a small experimental error in interbase positions (mainly incorrect values of interbase angles) leading to close interatomic contacts may introduce large errors in energy calculations.⁵² Thus, lower resolution X-ray structures are not accurate enough for subsequent interaction energy calculations. Moreover, the statistically significant underrepresentation of certain dinucleotide sequences (such as CG and CA) in high resolution canonical B-DNA duplexes⁵³ and the relatively low number of available crystals of A-RNA duplexes does not permit the construction of high quality geometries for all ten unique base pair step sequences for B-DNA and A-RNA, respectively.

The atomic coordinates can also be obtained by averaging over selected sets of atomic-resolution X-ray structures while respecting the sequence dependence of the structures. 54,55 However, this approach must be used with great care, as it can fail miserably. 56 The possible pitfalls with averaged X-ray geometries are exemplified by the study of Fiethen et al., 55 who calculated that the AT step is the least stable from all possible ten B-DNA dinucleotide steps. However, the poor stacking stabilization in this step in their calculations is primarily due to the repulsive 2.9 Å N6–N6 adenine amino group steric clash, which results from the geometry averaging. The AT steps indeed tend to form close amino group contacts (likely stabilized by asymmetric amino group pyramidalization) in the major groove, but 2.9 Å is certainly an unrealistic distance associated with a large energy penalty. 57,58

A more viable option is to construct idealized geometries with partial inclusion of geometrical parameters known from X-ray studies, while utilizing the empirical potential conformational search to eliminate ill-based geometries. Such geometries can be rather easily constructed for B-DNA. ^{23,59–62} However, preparation of equivalent structures of A-RNA base pair steps is much less straightforward. To do so, one would need to set up and vary more conformational parameters than in B-DNA, considering, for example, sequence-dependent redistribution of base pair roll and variations of displacement of the base pairs from the helix axis. ^{63,64} An absolute requirement is optimization of the vertical separation between base pairs for each combination of the other parameters. ⁵²

Another possibility that was utilized in the present study is to use molecular dynamics (MD) simulations to describe the dynamics of the system, and to develop a protocol of selecting appropriate geometries. Short explicit solvent MD simulations using the AMBER force fields have been shown to yield qualitatively correct structures of various DNA and RNA systems. 65-70 Though the use of MD trajectories for selection



CG/CG, TG/CA, TA/TA, CT/AG, CC/GG, AA/TT, TC/GA, AT/AT, AC/GT, GC/GC

Figure 1. Numbering scheme of one base pair step and the list of ten unique dinucleotide steps in DNA. Interactions 12 and 34 correspond to hydrogen bonding, interactions 13 and 42 (5' \rightarrow 3' convention is used) correspond to intrastrand stacking energies, and interactions 14 and 23 correspond to interstrand stacking. The bases within one base pair step are numbered as 13/42 (thus for the CT/AG step, CT and AG are the two intrastrand base stacks while CG and TA are two hydrogenbonded base pairs). RNA contains the same ten unique dinucleotide steps, only thymine (T) is replaced by uracil (U).

of appropriate geometries appears to be straightforward, several issues arising due to the force field's empirical character must be addressed. It is, for example, well established that simulations of nucleic acids modestly underestimate the helical twist. It is also known that not all aspects of the sequence-dependent local conformational variability of double helices are perfectly reproduced, and the structures simulated with AMBER force fields even appear to be subtly shifted from the typical B-DNA structure toward the A-DNA form.⁷¹ Recently, a particularly serious problem (an accumulation of incorrect α/γ backbone substate occurring⁷²⁻⁷⁴ in longer B-DNA simulations with parm99 (or older)⁷⁵ AMBER force field versions) was identified, and usable force field correction (called parmbsc0) of this behavior was introduced. ⁷⁶ However, despite all the limitations, the simulations at least partly reflect the effect of stacking on the fine double-helix structure, allow for monitoring eventual conformational substates, and neatly eliminate any pathological steric clashes. The structures are also not affected by crystal packing forces, which may perturb geometries seen in high resolution X-ray structures.

There exist ten unique DNA (and similarly RNA) base pair step (dinucleotide) sequences (Figure 1). To obtain a welldefined set of their geometries, molecular dynamics simulations of two model canonical B-DNA duplexes (18-mers) were performed: d(GCCTATAAACGCCTATAA) (further denoted as s1; the complementary strand is not shown) and d(CTAG-GTGGATGACTCATT) (further denoted as s2).⁷⁷ The same sequences were also utilized in molecular dynamics of canonical A-RNA duplexes. All ten unique steps are represented at least once in the model sequences, most of them 2-3 times. The initial geometries of model duplexes were built from Arnott fiber diffraction data¹ using the nucgen module of AMBER 8 software package.⁷⁸ All simulations were carried out using the AMBER 8 suite of programs⁷⁸ and the parm99 force field.⁷⁵ The trajectories were monitored for the presence of the α/γ noncanonical (γ -trans) backbone flips, ^{72–74} and all base pair step geometries with such γ -trans flips were discarded. Note that during preparation of this study, the corrected parmbsc0 version of the force field, effectively suppressing the α/γ flips, has been released.⁷⁶ The parm99 and parmbsc0 force fields provide an equivalent description of the B-DNA during the trajectory portions before the irreversible α/γ flips emerge.⁷⁹ The original simulation protocol of the Ascona B-DNA Consortium^{72,80} available from the MDDNA Web site 81 was used.

Influence of the Dynamics of the System on Stacking Energies-Monitoring of Stacking along the Trajectories. One of the purposes of this study was to monitor changes of base stacking energies along the trajectories. For the DNA simulations, one step for each unique dinucleotide sequence where no α/γ backbone flip^{72–74} occurred within the whole 20 ns trajectory was identified. For these steps, fifty structures averaged over 400 ps were generated for each of 20 ns trajectory using the ptraj module of Amber 8.78 Some time-averaging is needed since individual snapshots can be structurally rather deformed, but care must be taken as too long averaging could mix different substates (if present). A 400 ps period appears to be a reasonable compromise. For A-RNA, exactly the same dinucleotide steps as in DNA were used. However, in RNA simulations short-living reversible α/γ flips occur along the whole trajectory, with equilibrium populations of $\sim 10\%$.⁸² In contrast to B-DNA, the γ -trans flips are known to occur in A-RNA X-ray structures to some extent, 83 so their sampling in A-RNA simulations is not considered as pathological. Nevertheless, it is still not clear if their force field description is not associated with some bias of stacking geometries. Thus, only time windows where the backbone was in its canonical state were considered in further analysis of A-RNA duplexes, so the number of evaluated structures could be less than fifty for A-RNA steps (Table 2).

The sugar-phosphate backbone was removed from the generated structures, and the force field nucleobases were replaced by monomers of bases optimized separately at the RI-DFT-D/TPSS/TZVP level of theory at the C_s symmetry, i.e., assuming the planarity of amino groups. The N1/N9-C1' bond, where the base is attached to the sugar, was replaced by the N1/N9—H bond. Except when explicitly stated otherwise, the positions of hydrogens within individual base pairs were not reoptimized to reduce computational costs; i.e., the amino groups remained planar. Note that this is a reasonable approximation, as in real systems the positions of amino hydrogens are primarily determined by the base pair H-bonds, and the amino group nonplanarity is only negligible or modest, and certainly much smaller than for isolated nucleobases.

The base pair step stacking energies $\Delta E(12,34)$ between two base pairs (see Figure 1) were evaluated for all 400 ps-averaged geometries with canonical backbone conformations (i.e., with α/γ flips excluded) using eq 1. This equation does not incorporate the cooperativity effects between hydrogen bonding interactions and between hydrogen bonding and stacking interactions.84-88 Their inclusion would require considerably more calculations, and because these effects can be considered small compared to the base pair step stacking energies, they are neglected for purposes of the present study. The total energy of the tetramer E(1234), as well as total energies of two hydrogen bonded dimers E(12) and E(34), were evaluated at the RI-DFT-D/TPSS/TZVP level of theory (see below), which is fast and quite accurate (RI-DFT-D/TPSS/TZVP base pair step stacking energies are, on average, higher by 1.8 kcal·mol⁻¹ in B-DNA (Table 3), and by 1.6 kcal·mol⁻¹ in A-RNA (Table 4) compared to CBS(T) results)36,89,90 and is thus suitable for a monitoring of the development of stacking energies along the trajectories. The fluctuations of stacking energies in the course of the 20 ns trajectories were characterized by their means, standard deviations, and minimum and maximum values for a given step.

$$\Delta E(12,34) = E(1234) - (E(12) + E(34)) \tag{1}$$

The RI-DFT-D method has been parametrized to match very accurate complete basis set limit CCSD(T) interaction energies.³⁷ The RI-DFT-D interaction energies are thus implicitly corrected for the basis set superposition error (BSSE), and a subsequent correction for this artifact in the actual calculation is not needed. The DFT-D calculations have been performed with the TurboMole 5.8⁹¹ software package, in combination with a local code that computes the empirical dispersion correction.³⁶ For the DFT part of the calculation the RI approximation,^{92–95} also known as density fitting, was used.

Stacking Energies Obtained at Different Levels of Theory—Reference Calculations. Since the DFT-D method used to monitor the base stacking energy variations in the course of the trajectories is approximate, we also carried out reference high quality ab initio calculations for few selected structures. For each of ten unique dinucleotide steps both in DNA and in RNA, one 400 ps average was chosen randomly (with no α/γ flips, see above), these geometries will further be referred to as random-energy geometries (available as Supporting Information). For the reference calculations, positions of hydrogens were adjusted by their RI-DFT-D/TPSS/TZVP optimization separately for each base pair with no symmetry enforced. This additional optimization of hydrogens improves their positioning due to their involvement in the Watson-Crick hydrogen bonding. Then, the intrastrand $\Delta E(13)$, $\Delta E(42)$ and interstrand $\Delta E(14)$, $\Delta E(23)$ stacking energies of the base dimers (Figure 1) were evaluated to obtain the overall stacking energy according to eq 2. While eq 1 evaluates the stacking energy as energy difference between the complete four-base stacked system and the two Watson-Crick base pairs, eq 2 evaluates the stacking energy as a sum of four base - base stacks and a many body correction.

$$\Delta E(12,34) =$$

$$\Delta E(13) + \Delta E(42) + \Delta E(14) + \Delta E(23) + \Delta E^{m}$$
 (2)

The stacking energies were evaluated using RI-MP2 complete basis set (CBS) limit method⁹⁶ augmented with the CCSD(T) correction term with smaller basis set,⁵⁶ which is abbreviated as CBS(T) method (see Supporting Information for further details). The MP2 and CCSD(T) calculations have been performed with the Molpro 2006.1 software package,⁹⁷ applying the frozen-core approximation. The CBS(T) calculations were compared with the RI-DFT-D/TPSS approach either with TZVP or with larger 6-311++G(3df,3pd) (further referred to as LP) basis sets.³⁷

Ordering of DNA and RNA Stacking Energies Based on **Average-Energy Geometries.** In an attempt to order ten unique dinucleotide steps based on their intrinsic gas phase stacking energies, another set of reference geometries was generated. First, RI-DFT-D/TPSS/TZVP stacking energies of ten unique dinucleotide steps identified previously were averaged over the trajectory window with no α/γ flips occurring (for explanation, see above). Then the reference geometry was chosen from the trajectory so that the stacking energy of such geometry was as close as possible to the stacking energy averaged over the whole trajectory (Figure 2). This set of reference geometries will further be referred to as average-energy geometries (available as Supporting Information). Their base pair stacking energies were evaluated at the RI-DFT-D/TPSS/LP level of theory according to eq 1, and the many-body term $\Delta E^{\rm m}$ was obtained by subtracting the sum of intrastrand and interstrand stacking terms (first four terms in eq 2) from the dinucleotide base paired stacking energy $\Delta E(12,34)$ calculated according to eq 1.

Results and Discussion

Influence of the Dynamics of the System on Stacking Energies. The first point that will be addressed is how representative is a single geometry taken from MD trajectory; i.e., what is the variability of the stacking along the trajectories?

Tables 1 and 2 list the mean values, standard deviations, and minimum and maximum values of stacking energies of all ten

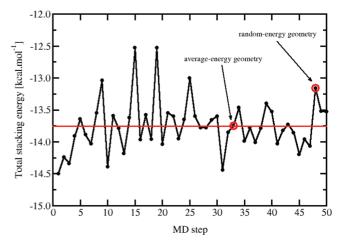


Figure 2. Change in total stacking energy of the CG/CG step in B-DNA s1 duplex along its 20 ns trajectory. The stacking energy was evaluated at the RI-DFT-D/TPSS/TZVP for each of 400 ps-averaged structures, resulting thus in 50 values. The stacking energies of the random-energy geometry and the average-energy geometry are indicated. The stacking energy of the average-energy geometry is as close as possible to the stacking energy averaged over the whole trajectory (-13.75 kcal·mol⁻¹, red line).

unique dinucleotide steps during the 20 ns long parm99 trajectories (see Methods for details how the structures were selected). The stacking energies were evaluated according to eq 1 at the RI-DFT-D/TPSS/TZVP level of theory. Figure 2 shows the time evolution along the 20 ns trajectory in the CG/CG B-DNA step from the s1 duplex. Because all interaction energies were evaluated consistently in the same manner, the standard deviations characterize the changes in stacking energies induced by the dynamical character of the system. Apparently, the variability of stacking energy along the B-DNA trajectory is significant and may reach up to 3 kcal·mol⁻¹, which is ~25% of the average base-pair step stacking energy and almost 50% of the calculated sequence dependence of the stacking energy. These numbers clearly show that evaluating the stacking energy order based on a single geometry is in fact not straightforward.

Stacking Energies Obtained at Different Levels of Theory. Table 3 summarizes the intrinsic base pair step stacking energies for ten unique dinucleotide steps (at random-energy geometries) in B-DNA obtained at the reference CBS(T) level of theory, as well as at the RI-DFT-D/TPSS level of theory using two different basis sets: TZVP and LP. The same calculations were performed also for ten unique dinucleotide steps in A-RNA duplex (Table 4). The details about CBS(T) (Table S1 and Table S3) and RI-DFT-D/TPSS/LP (Table S2 and Table S4) interaction energies, as well as about their differences (Table S5 and Table S6), can be found in Supporting Information.

The CBS(T) base pair step stacking energies of B-DNA (Table 3) lie within the range -15.27 kcal·mol⁻¹ (GC/GC step) to -9.93 kcal·mol⁻¹ (CC/GG step). For A-RNA the dinucle-otide steps have CBS(T) stacking energies (Table 4) within the range -16.75 kcal·mol⁻¹ (GC/GC step) to -8.78 kcal·mol⁻¹ (CC/GG step). Similarly, B-DNA base pair step stacking energies evaluated at the RI-DFT-D/TPSS/LP level of theory (Table 3) lie within the range -14.32 kcal·mol⁻¹ (GC/GC step) to -8.90 kcal·mol⁻¹ (CC/GG step), and A-RNA stacking energies (Table 4) are from the range -16.17 kcal·mol⁻¹ (GC/GC step) to -7.82 kcal·mol⁻¹ (CC/GG step). The very weak gas phase stacking in CC/GG steps reflects quite unfavorable intrastrand electrostatics due to similar orientations of the dipoles of highly polar consecutive guanines in one and cytosines in the other strand.^{51,59}

TABLE 1: RI-DFT-D/TPSS/TZVP Stacking Energies (kcal·mol⁻¹) of Ten Unique Dinucleotide Steps in B-DNA Averaged over the 20 ns parm99 Trajectory^a

st	ер	average	standard deviation	min	max
YR/YR	CG/CG	-13.75 ± 0.11	0.40	-14.50	-12.53
YR/YR	TG/CA	-11.19 ± 0.14	0.48	-12.33	-9.87
YR/YR	TA/TA	-11.84 ± 0.22	0.77	-12.81	-9.98
YY/RR	CT/AG	-10.46 ± 0.07	0.26	-11.18	-9.99
YY/RR	CC/GG	-7.85 ± 0.12	0.42	-9.81	-7.02
RR/YY	AA/TT	-11.07 ± 0.14	0.50	-12.29	-9.93
RR/YY	TC/GA	-10.52 ± 0.08	0.30	-11.23	-9.86
RY/RY	AT/AT	-11.71 ± 0.06	0.23	-12.16	-11.05
RY/RY	AC/GT	-12.21 ± 0.15	0.53	-13.16	-10.77
RY/RY	GC/GC	-14.14 ± 0.16	0.55	-15.20	-12.89

^a The averages and standard deviations were evaluated from 50 structures (N = 50; see Methods). The confidence intervals for the mean values of base pair step stacking energies were computed under the assumption that the energies at the 95% confidence level are distributed normally. Minimum and maximum values of the stacking energies from the 20 ns trajectory are also given. Y and R stand for pyrimidine and purine, respectively.

TABLE 2: RI-DFT-D/TPSS/TZVP Stacking Energies (kcal·mol⁻¹) of Ten Unique Dinucleotide Steps in A-RNA Averaged over the 20 ns parm99 Trajectory^a

S	tep	average	standard deviation	N	min	max
YR/YR	CG/CG	-13.54 ± 0.39	0.56	10	-14.39	-12.78
YR/YR	UG/CA	-10.19 ± 0.12	0.28	23	-10.69	-9.48
YR/YR	UA/UA	-11.87 ± 0.09	0.33	50	-12.34	-11.06
YY/RR	CU/AG	-9.61 ± 0.04	0.15	50	-10.07	-9.31
YY/RR	CC/GG	-6.77 ± 0.14	0.40	35	-7.46	-5.51
RR/YY	AA/UU	-8.49 ± 0.11	0.23	20	-8.90	-8.12
RR/YY	UC/GA	-10.13 ± 0.11	0.41	50	-11.18	-9.43
RY/RY	AU/AU	-9.12 ± 0.09	0.19	50	-9.58	-8.58
RY/RY	AC/GU	-9.98 ± 0.10	0.22	21	-10.32	-9.43
RY/RY	GC/GC	-14.58 ± 0.23	0.50	21	-15.49	-13.58

^a The regions with α/γ flips were excluded from the analysis (see Methods). The number of structures from which averages and standard deviations were evaluated is given in the column "N". The confidence intervals for the mean values of base pair step stacking energies were computed under the assumption that the energies are distributed normally at the 95% confidence level. Minimum and maximum values of the stacking energies from the regions with α/γ flips excluded are also given.

TABLE 3: Stacking Energies (kcal·mol⁻¹) of Ten Unique Dinucleotide Steps (at Random-Energy Geometries) in B-DNA Evaluated at Different Levels of Theory^a

	•	
TPSS/TZVP	TPSS/LP	CBS(T)
-13.22	-13.80	-15.22
-12.18	-12.84	-13.99
-13.24	-13.96	-14.84
-10.14	-11.05	-12.23
-8.06	-8.90	-9.93
-11.16	-12.17	-12.96
-10.56	-11.52	-12.32
-11.10	-12.38	-13.02
-11.93	-12.94	-13.42
-13.53	-14.32	-15.27
	-13.22 -12.18 -13.24 -10.14 -8.06 -11.16 -10.56 -11.10 -11.93	TPSS/TZVP TPSS/LP -13.22 -13.80 -12.18 -12.84 -13.24 -13.96 -10.14 -11.05 -8.06 -8.90 -11.16 -12.17 -10.56 -11.52 -11.10 -12.38 -11.93 -12.94

^a The RI-DTF-D/TPSS energies were calculated according to eq 1. The CBS(T) energies (Supporting Information, Equation S1, Equation S2) were calculated as the sum of CBS(T) inter- and intrastrand stacking contributions and many-body term ΔE^{m} evaluated at the MP2/aug-cc-pVDZ level of theory (eq 2).

Following expectations, the $\Delta CCSD(T)$ corrections for base pair step stacking interactions are always positive (Supporting Information, Table S7 and Table S8), meaning that passing from MP2 to CCSD(T) reduces the stabilization. The higher correlation energy contributions are rather significant, ranging from +2.45 to +3.43 kcal·mol⁻¹ in B-DNA and from +2.57 kcal·mol⁻¹ to +3.23 kcal·mol⁻¹ in A-RNA and can thus be never neglected. 98,99 The differences between Δ CCSD (T) corrections in B-DNA and A-RNA are only marginal, not exceeding 0.35 kcal·mol⁻¹.

The comparison of the RI-DFTD/TPSS method with the high level CBS(T) calculations on the B-DNA system reveals (Table

TABLE 4: Stacking Energies (kcal·mol⁻¹) of Ten Unique Dinucleotide Steps (at Random-Energy Geometries) in A-RNA Evaluated at Different Levels of Theory^a

TPSS/TZVP	TPSS/LP	CBS(T)
-12.36	-12.85	-14.03
-9.95	-10.76	-12.02
-12.10	-12.73	-13.65
-9.17	-10.05	-10.94
-6.98	-7.82	-8.78
-8.53	-9.38	-10.16
-10.94	-11.61	-12.46
-8.65	-9.68	-9.84
-10.01	-10.95	-11.46
-15.61	-16.17	-16.75
	-12.36 -9.95 -12.10 -9.17 -6.98 -8.53 -10.94 -8.65 -10.01	-12.36 -12.85 -9.95 -10.76 -12.10 -12.73 -9.17 -10.05 -6.98 -7.82 -8.53 -9.38 -10.94 -11.61 -8.65 -9.68 -10.01 -10.95

^a The RI-DTF-D/TPSS energies were calculated according to the 1. The CBS(T) energies (Supporting Information, Equation S1, Equation S2) were calculated as the sum of CBS(T) inter- and intrastrand stacking contributions and many-body term ΔE^{m} evaluated at the MP2/aug-cc-pVDZ level of theory (eq 2).

3) that, while the RI-DFTD/TPSS/TZVP calculations yield the base pair step stacking energies on average higher (less stabilizing) by 1.8 kcal·mol⁻¹ compared to CBS(T) results, this error is reduced about 2-fold when using the larger LP basis set. The same effect was also observed for RNA base pair steps (Table 4). On average, CBS(T) yields B-DNA base pair step stacking energies that are on average lower (more stabilizing) by $0.93 \pm 0.20 \, \mathrm{kcal} \cdot \mathrm{mol}^{-1}$ than that calculated at the RI-DFT-D/TPSS/LP level of theory. For A-RNA, the CBS(T) base pair step stacking energies are lower (more stabilizing) by $0.81~\pm$ 0.23 kcal·mol⁻¹ than their RI-DFT-D/TPSS/LP counterparts.

TABLE 5: RI-DFT-D/TPSS/LP Interaction Energies at Average-Energy Geometries of Ten Unique Base Pair Steps in B-DNA (kcal·mol⁻¹)^a

step	intrastrand stacking		interstrand stacking		many-body	total stacking
	$\Delta E(13)$	$\Delta E(42)$	$\Delta E(14)$	$\Delta E(23)$	$\Delta E^{ m m}$	$\Delta E(12,34)$
CG/CG	-5.89	-5.43	0.71	-3.66	0.63	-13.65
TG/CA	-3.28	-3.02	-1.53	-5.08	1.67	-11.50
TA/TA	-5.33	-5.27	0.69	-2.96	0.20	-12.67
CT/AG	-6.07	-6.05	0.15	-1.24	1.74	-11.47
CC/GG	-2.59	-2.26	-3.35	-4.04	3.15	-9.08
AA/TT	-5.25	-3.90	-1.12	-2.13	0.21	-12.19
TC/GA	-4.65	-7.83	0.65	-1.01	0.82	-12.03
AT/AT	-6.32	-6.42	-1.03	0.58	0.69	-12.50
AC/GT	-5.23	-5.60	-3.39	-0.14	1.04	-13.32
GC/GC	-10.86	-10.66	2.57	2.93	0.56	-15.45

^a The bases within one base pair step are numbered as 13/42; see Figure 1.

Large difference between RI-DFT-D/TPSS/LP and CBS(T) stacking in B-DNA comes from the many-body correction term (Supporting Information, Table S1, Table S2, and Table S5), which is lower (i.e., smaller in absolute value) by 0.52 ± 0.16 kcal·mol⁻¹ at the MP2/aug-cc-VDZ level used for the fourbody correction of the CBS(T) base-pair step stacking energies. Similar values were observed also for A-RNA steps (Supporting Information, Table S3, Table S4, and Table S6): the many body correction term is lower by $0.48 \pm 0.07 \text{ kcal} \cdot \text{mol}^{-1}$ at the MP2/ aug-cc-VDZ level compared to the RI-DFT-D/TPSS/LP level. The many body term, obtained at the MP2/aug-cc-pVDZ level of theory, does not cover the dispersion nonadditivity, which is included only at the MP3 or higher theoretical levels (e.g., CCSD(T)). The dispersion nonadditivity is also not captured by the DFT-D method, which uses classical London atom-atom pair additive dispersion formula that is parametrized using only monomers and dimers.³⁷ We thus suggest that the RI-DFT-D method overestimates the energetic contribution of the many body terms due to polarization effects, at least compared with the MP2/aug-cc-pVDZ method.

The differences between CBS(T) and RI-DFT-D/TPSS/LP methods are much larger for the intrastrand stacking in B-DNA than for interstrand (Supporting Information, Table S5), as the overlap of two bases from two opposite strands is very small compared to the intrastrand arrangement; i.e., the interstrand stacking has a much smaller van der Waals component of stacking. However, this effect is much less pronounced in A-RNA (Supporting Information, Table S6), as in A-type duplexes the role of interstrand interaction is generally larger due to the mutual slide of the consecutive base pairs inherent to the A-form.

Not only the differences between RI-DFT-D/TPSS/LP and CBS(T) stacking energies but also the relative orderings of the base pair step stacking stabilities evaluated at different levels of theory were monitored (Supporting Information, Table S9 and Table S10). Though the differences in stacking energies between various levels of theory are rather systematic, they are not entirely uniform; i.e., the relative values of stacking energies remain modestly sensitive to the level of calculations. However, the method-dependent minor variations in the relative base pair step ordering are not frequent, and the main features of the stability patterns (such as the identity of the most and least stable steps) remain preserved (see also the discussion below).

In addition, when the CBS(T) energies are plotted against either RI-DFT-D/TPSS/TZVP or RI-DFT-D/TPSS/LP interaction energies of ten unique dinucleotide steps in B-DNA (Supporting Information, Figure S1) and in A-RNA (Supporting Information, Figure S2), the regression equations clearly il-

lustrate the high correlation between the CBS(T) values and those obtained from RI-DFT-D/TPSS computations, thus giving more confidence to this computational approach. In particular, the regression shows that TPSS/LP values nearly match the CBS(T) data but for a systematic deviation of $-0.89 \, \mathrm{kcal} \cdot \mathrm{mol}^{-1}$ for B-DNA and of $-0.77 \, \mathrm{kcal} \cdot \mathrm{mol}^{-1}$ for A-RNA (as noted earlier).

Therefore it may be concluded that the RI-DFT-R/TPSS/LP method may be used as a cheap method for monitoring the stacking along the trajectories without a significant loss in accuracy. The time savings are approximately on the order of two magnitudes compared to CBS(T).

Decomposition of Stacking Energies into Intra- and Interstrand Terms. As demonstrated in the previous section, RI-DFT-D/TPSS/LP represents very affordable and accurate alternatives allowing for characterization of systems not amenable to high level ab initio calculations. Thus, the RI-DFT-D method was exclusively used for the analysis of inter- and intrastrand stacking contributions. In contrast to the preceding paragraph comparing various levels of theory, the average-energy geometries were utilized for this purpose, as explained in the Methods. The results are summarized in Tables 5 and 6.

The intrastrand stacking brings the major portion of the stabilization, mainly considering the van der Waals component of stacking. Thus intrastrand stacking both in B-DNA (Supporting Information, Table S11), and in A-RNA (Supporting Information, Table S12) were analyzed separately. There are four basic intrastrand base-base stacks: purine-pyrimidine (RY), purine—purine (RR), pyrimidine—pyrimidine (YY), and pyrimidine-purine (YR) stacks, each consisting of four sequences. RY interactions are most favored both in B-DNA $(-6.99 \text{ kcal} \cdot \text{mol}^{-1})$ and in A-RNA $(-5.93 \text{ kcal} \cdot \text{mol}^{-1})$, followed by RR $(-5.35 \text{ kcal} \cdot \text{mol}^{-1} \text{ in B-DNA}, -4.72 \text{ kcal} \cdot \text{mol}^{-1}$ in A-RNA), and by YR interactions (-4.32 kcal·mol⁻¹ in B-DNA, −3.88 kcal·mol⁻¹ in A-RNA). YY interactions are as strong as YR interactions in B-DNA (-4.30 and -4.32 kcal ⋅ mol⁻¹, respectively), and they are slightly weaker than YR interactions in A-RNA (-3.33 and -3.88 kcal·mol⁻¹, respectively). The observed B-DNA ordering RY < RR < YR \sim YY is different from that reported a decade ago using fiber diffraction geometries⁴⁰ (RR < YR \sim RY < YY) utilizing the MP2/6-31G* method. The main source of differences is not in using different sets of geometries (average-energy vs fiber), as these two sets yield similar orderings at the RI-DFT-D/TPSS/ LP level of theory (demonstrated later), but in using the standard 6-31G* basis set with energy optimized d-polarization functions in the MP2 calculations. This basis set is known to qualitatively underestimate the dispersion energy contribution, in contrast to

TABLE 6: RI-DFT-D/TPSS/LP Interaction Energies at Average-Energy Geometries of Ten Unique Base Pair Steps in A-RNA $(\mathbf{kcal \cdot mol}^{-1})^a$

step	intrastrand stacking		interstrand stacking		many-body	total stacking
	$\Delta E(13)$	$\Delta E(42)$	$\Delta E(14)$	$\Delta E(23)$	$\Delta E^{ m m}$	$\Delta E(12,34)$
CG/CG	-4.50	-5.86	0.99	-4.76	0.71	-13.42
UG/CA	-2.25	-3.06	-1.13	-5.65	1.23	-10.86
UA/UA	-5.02	-5.00	0.65	-3.71	0.22	-12.87
CU/AG	-5.46	-5.04	-0.04	-0.97	1.62	-9.89
CC/GG	-1.96	-1.91	-2.86	-4.54	3.15	-8.12
AA/UU	-3.72	-1.66	-1.01	-3.31	0.55	-9.15
UC/GA	-4.24	-8.22	0.75	-0.77	1.01	-11.47
AU/AU	-4.76	-5.19	-1.26	0.41	1.12	-9.68
AC/GU	-3.70	-4.24	-3.23	-1.23	1.35	-11.05
GC/GC	-10.84	-10.80	2.85	2.73	0.14	-15.93

^a The bases within one base pair step are numbered as 13/42; see Figure 1.

modified 6-31G*(0.25) basis set with diffuse (polarizabilityoptimized) d-polarization functions used in numerous stacking calculations by Sponer and Hobza. 100

It could be at first sight surprising that on average the RR stacks provide less intrastrand stacking stabilization than the RY stacks. However, the RR stacks include two homodimers (GG and AA) which obviously have unfavorable electrostatics (mainly the GG stack). This reduces the overall gas phase stacking energy and masks the effect of van der Waals (overlap) stabilization which certainly is the largest for the RR stacks. However, it must be emphasized that regarding thermodynamics stability of DNA or RNA, the solvent screening compensates for the electrostatics interactions in both attractive and repulsive arrangements. 59,101 Thus the gas phase data, albeit important, cannot be directly used to assess the contribution of particular stacking interactions to the thermodynamics stability of nucleic

For the average-energy geometries, the total RI-DFT-D/TPSS/ LP stacking in B-DNA ranges (Table 5) from −15.45 kcal⋅mol⁻¹ (GC/GC step) to −9.08 kcal⋅mol⁻¹ (CC/GG step) and for A-RNA (Table 6) from -15.93 kcal·mol⁻¹ (GC/GC step) to $-8.12 \text{ kcal} \cdot \text{mol}^{-1}$ (CC/GG step). The stabilization is dominated by the intrastrand (Figure 1) contribution ranging from -10.86 (G/C stack in the GC/GC step) to -2.26 (C/C stack in the CC/GG step) kcal·mol⁻¹ in B-DNA (Table 5) and from -10.84 kcal·mol⁻¹ (G/C stack in the GC/GC step) to -1.66 kcal·mol⁻¹ (U/U stack in the AA/UU step) in A-RNA (Table 6). The weak stacking in the CC/GG step (both in B-DNA and A-RNA) is caused by unfavorable dipole—dipole electrostatic energy between the guanines and cytosines within the strands.⁵¹ The interstrand contribution ranges from -5.08 kcal·mol⁻¹ (G/A stack in the TG/CA step) to +2.93 kcal·mol⁻¹ (C/C stack in the GC/GC step) in B-DNA, and from -5.65kcal⋅mol⁻¹ (G/A stack in the UG/CA step) to 2.85 kcal⋅mol⁻¹ (G/G stack in the GC/GC step) in A-RNA. The interstrand and intrastrand contributions mutually compensate each other.^{51,59} For example, in B-DNA, the weakest intrastrand stacks (-2.26and $-2.59 \text{ kcal} \cdot \text{mol}^{-1}$) in the CC/GG step are compensated for by the rather attractive interstrand interactions of -4.04 and $-3.35 \text{ kcal} \cdot \text{mol}^{-1}$ (Table 6). The opposite distribution of values is observed for the GC/GC step, with the most attractive intrastrand contributions (-10.86 and -10.66 kcal·mol⁻¹) and the most repulsive interstrand terms of +2.93 and +2.57kcal ⋅mol⁻¹. Similar behavior was also observed for A-RNA base pair steps. This confirms the anticipated fact that the nature and magnitude of intrinsic stacking in B- and A-forms of nucleic acids double helices do not differ significantly. Considering the base stacking, the energetics of B to A transition is smooth;

i.e., the stacked base pairs can slide between the two forms without overcoming any significant stacking energy barriers. 102

Nonadditivity of Stacking. The many-body term $\Delta E^{\rm m}$, evaluated at the RI-DFT-D/TPSS/LP level of theory using average-energy geometries (Tables 5 and 6), consists of four three-body terms and one four-body term. It is always positive and no larger than 1.7 kcal·mol⁻¹ for most steps with the exception of a CC/GG step, which shows a non-negligible degree of nonadditivity ($\Delta E^{\rm m} = 3.15 \text{ kcal} \cdot \text{mol}^{-1}$) both in B-DNA and in A-RNA. The many-body term increases the energy difference between the most and least stable steps both in B-DNA (from 3.79 to 6.37 kcal·mol⁻¹) and in A-RNA (from 4.56 to 6.04 kcal·mol⁻¹). In B-DNA, the many-body term is rather variable in steps consisting of two GC base pairs, but always small in steps with two AT base pairs⁵¹ (Tables 5 and

The many-body correction term (for random-energy geometries) calculated at the MP2/aug-cc-pVDZ level of theory is not always positive (Supporting Information, Table S1 and Table S3). When calculated by the RI-DFT-D/TPSS/LP method, it is shifted to more positive values on average by 0.5 kcal·mol⁻¹ (Supporting Information, Table S2 and Table S4) and becomes always positive. Note that to include the dispersion nonadditivity that is neglected at the MP2/aug-cc-pVDZ level,⁵⁹ MP3 or higher orders of perturbation expansion, or the CCSD(T) method should be used.⁵⁶ However, such calculations are extremely expensive and out of our computational capabilities, and thus the dispersion nonadditivity is not properly accounted for in this study.

Comparison of B-DNA and A-RNA Stacking. The differences in stacking energies (RI-DFT-D/TPSS/LP level of theory using average-energy geometries) between A-RNA and B-DNA (Tables 5 and 6) show that for most steps the B-type stacking is more favorable than A-type stacking. The biggest difference is for the AA/TT step (3.05 kcal·mol⁻¹ in favor of B-DNA); other steps significantly more stable in B-DNA are AT/AT (2.83 kcal·mol⁻¹ in favor of B-DNA), AC/GT (2.27 kcal·mol⁻¹ in favor of B-DNA), and CT/AG (1.57 kcal·mol⁻¹ in favor of B-DNA). The only two steps more stable in A-RNA than in B-DNA are UA/UA and GC/GC, though the difference is only small (0.19 and 0.48 kcal·mol⁻¹, respectively). On average, B-DNA is more stable than A-RNA by 1.14 kcal·mol⁻¹, which is much less than suggested previously for B-DNA vs A-DNA stacking.⁴⁰ This discrepancy arises very likely from differences in geometries, since inaccurate geometries (including incorrect vertical separation of base pairs) derived from fiber diffraction data were used previously to study the A- and B-type stacking. 40 In addition, the previous study⁴⁰ utilized small 6-31G* basis set lacking the diffuse polarization functions, which is known

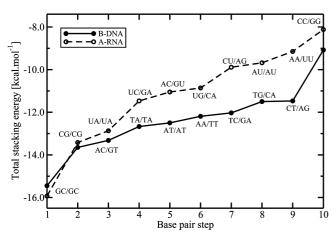


Figure 3. Ordering (in decreasing stability from left to right) of ten unique dinucleotide steps in B-DNA and A-RNA obtained from RI-DFT-D/TPSS/LP interaction energies evaluated at average-energy geometries.

to qualitatively underestimate the dispersion energy contribution, as explained earlier.

Regarding the present study, it is also important to note that DNA and RNA contain thymine and uracil, respectively. The 5-methyl group of thymine increases the van der Waals component of stacking (Table S13). When effects of this methyl group are subtracted, the DNA and RNA gas phase stacking stabilization are on average basically the same. In other words, we do not see any fundamental difference between stacking stability inherent to the B- and A-form helices, in contrast to the earlier MP2/6-31G* fiber geometry calculations, 40 suggesting the B-DNA stacking being by 6.0 kcal mol more stable than the A-DNA stacking.

Note that the reason why uracil is replaced by thymine in DNA is most likely not the improved stacking stabilization. The T vs U difference is important for the repair mechanisms to distinguish thymines from uracils that are quite frequently formed in DNA as a result of spontaneous cytosine deamination. ¹⁰³ The fundamental difference between RNA and DNA is the 2'-OH group of ribose, due to its chemical reactivity and potential for extended base pairing. ^{104,105}

Not only the numerical differences between individual steps but also their relative ordering were investigated both in B-DNA and in A-RNA (Figure 3). As the energy difference between two base pair step sequences that are consecutive on the stacking energy scale is often very small (see Figure 3), the ordering of ten unique dinucleotide steps cannot be definitive. It can vary with the dynamics of the system to a certain extent, and with the level of theory the stacking energies were evaluated at (Supporting Information, Table S9 and Table S10). However, some general conclusions may still be drawn. The stacking energies of base pair steps in A-RNA are more evenly separated compared to B-DNA (the A-RNA curve at the Figure 3 rises more monotonically than that of B-DNA), and their ordering will be less sensitive to the dynamics of the system compared to B-DNA (and the A-RNA in addition appears to be more rigid). The most stable step both in B-DNA and in A-RNA is the CG/CG step, which is very well separated (1.80 kcal·mol⁻¹ in B-DNA and 2.51 kcal·mol⁻¹ in A-RNA) from the second most stable step, which is the CG/CG one. On the other side of the stability ordering lies the least stable step CC/GG, again well separated from the last but one step on the stacking stability scale in B-DNA (2.38 kcal·mol⁻¹) and in A-RNA (1.03 $kcal \cdot mol^{-1}$).

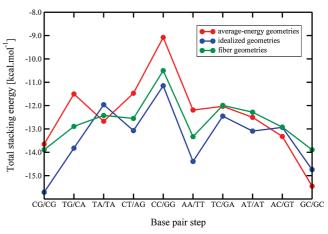


Figure 4. RI-DFT-D/TPSS/LP stacking energies at various sets of geometries of ten unique base pair steps in B-DNA (kcal·mol⁻¹).

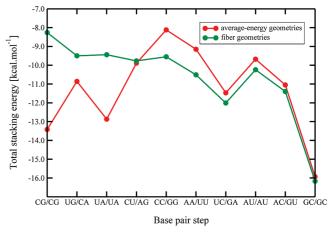


Figure 5. RI-DFT-D/TPSS/LP stacking energies at various sets of geometries of ten unique base pair steps in A-RNA (kcal·mol⁻¹).

Comparison with Data for Fiber Diffraction and Idealized Geometries. The B-DNA stacking energies of average-energy geometries were compared (Figure 4) to the energies in ten unique dinucleotide steps idealized geometries taken from ref 59 (Supporting Information, Table S14, Table S17, and Figure S3) and at fiber diffraction¹ (Supporting Information, Table S15 and Table S17) geometries. For A-RNA, as explained in the Methods, the idealized geometries were not proposed so far, and thus only stacking of average-energy geometries was compared (Figure 5) to that of fiber diffraction geometries (Supporting Information, Table S16 and Table S17). All stacking energies were evaluated at the RI-DFT-D/TPSS/LP level of theory.

In B-DNA, ordered stacking energies for average-energy geometries are generally shifted up (less stable) compared to the idealized geometries (Supporting Information, Figure S4). The most striking difference is the permutation at the position of the most stable dinucleotide step (Supporting Information, Table S18 and Figure S4). While for the average-energy geometries the most stable step is the GC/GC followed by the CG/CG, for idealized geometries the most stable step is the CG/CG succeeded by the GC/GC. Stacking energies for fiber geometries are generally more stable than these for average-energy geometries, but less stable compared to idealized geometries (Supporting Information, Table S18 and Figure S4). For fiber geometries, the first two most stable steps GC/GC and CG/CG have the same energy (Supporting Information, Figure S4). The identity of the least stable step (CC/GG) remains the

same for all three sets of geometries, but the order of other steps differs significantly between the three sets of geometries.

These differences thus again illustrate sensitivity of the data to the choice of the geometry and difficulty to characterize stacking by a single geometry. The fiber geometries¹ were obtained by analyzing fiber diffraction data, they are uniform for all sequences, and certain backbone angles are at odds with the atomic-resolution crystal structures (see Methods). The idealized B-DNA geometries were obtained by fixing the helical twist at 36° and optimizing the rise and the propeller twist independently for each of the ten systems. On the other hand, the average-energy geometries were obtained from force field trajectories. However, simulated Amber B-DNA duplexes show several differences compared to data obtained from X-ray structures analysis. The force field has a tendency to decrease the helical twist from $\sim 36^{\circ}$ found (on average) in B-DNA crystals to the range of 28-32° (Supporting Information, Table S19). Also, while B-DNA has a slide in the range 0.1–0.6 Å and the idealized geometries were constructed assuming slide 0, MD simulated structures have a slide in the negative territory of -0.5 to -1.8 Å (Supporting Information, Table S19). This basically means that there is a modest bias of the simulated structures toward the A-form helix values. Thus, we are far from claiming that the MD geometries are perfect. Interestingly, the reduction of helical twist should lead to the increase of the overlap of bases, and thus to the increase of the van der Waals portion of stacking. Despite this, the MD-generated geometries have on average weaker stacking. Perhaps the electrostatic contribution to stacking or the negative slides are responsible for this behavior. The very good stacking energies obtained for the idealized geometries⁵⁹ nevertheless verify their usefulness for reference stacking energy calculations, albeit the present data clearly show that future studies of stacking should consider a wider set of geometries.

Though it may seem that the A-RNA ordered stacking energies for average-energy geometries are generally shifted down (more stable) compared to fiber diffraction geometries (Supporting Information, Figure S5), the close inspection of Figure S5 shows that for the majority of steps (GC/GC, UC/ GA, AC/GU, AA/UU, AU/AU, CC/GG) the reverse trend is generally found (i.e., these steps are more stable at fiber geometries than at average-energy geometries). The higher stability of certain steps at fiber geometries was also found in B-DNA (Supporting Information, Figure S4).

The ordering of base pair steps is very different for fiber diffraction geometries compared to average-energy geometries. The most striking difference can be found for the CG/CG step, which is the second most stable at average-energy geometry, but the least stable step at fiber geometry. The poor stacking stabilization in this step at fiber geometry is due to the repulsive steric clash of guanine O6-O6 atoms (distance between them is 2.89 Å), which is not present in average-energy geometry, where these atoms are 3.42 Å apart. Such a short distance is associated with a large energy penalty; while the GG interstrand stacking energy at average-energy geometry is -4.76 kcal·mol⁻¹ (see $\Delta E(23)$ energy of CG/CG step in Table 6), it decreases by ca. $3.50 \text{ kcal} \cdot \text{mol}^{-1}$ to the value of -1.23kcal·mol⁻¹ at the fiber geometry (see $\Delta E(23)$ energy of CG/ CG step in Supporting Information, Table S16).

Though for more steps at the fiber diffraction geometries the B-type stacking is, similarly to average-energy geometries, more favorable than A-type stacking, the differences between fiber diffraction and average-energy stacking energies in A-RNA and B-DNA are numerous (Supporting Information, Table S20). Between the most striking distinctions belongs the CG/CG step, which is almost equally (0.23 kcal·mol⁻¹ in favor of B-DNA) stable in B-DNA and in A-RNA at average-energy geometry, but much more stable in B-DNA (by 5.63 kcal·mol⁻¹ in favor of B-DNA) at the fiber diffraction geometry (for explanation, see the preceding paragraph). The TA/TA step, which is, albeit marginally, more stable in A-RNA at average-energy geometries, becomes significantly more stable (by 2.98 kcal·mol⁻¹ in favor of B-DNA) in B-DNA at the fiber diffraction geometry. The GC/GC step, which was found to be more stable (by 0.48 kcal·mol⁻¹ in favor of A-RNA) in A-RNA at average-energy geometries, is much more stabilized in A-RNA at the fiber diffraction geometry (2.29 kcal·mol⁻¹ in favor of A-RNA). Again, these findings emphasize, in our opinion, the importance of proper choice of geometries and shows that fiber diffraction geometries are not sufficiently accurate for interaction energy calculations.

In many earlier studies, base stacking was evaluated using the MP2 method with the diffuse-polarized 6-31G*(0.25) basis set. 106 This basis set has modified d-polarization functions with an exponent of 0.25 (compared with the standard exponent of 0.8). The d-polarization functions are capable of filling the gap between the interacting monomers, which significantly improves the description of base stacking interactions compared with the standard 6-31G* basis set. The MP2/6-31G*(0.25) method has been the most widely used ab initio technique to study aromatic stacking in the past literature 31,35,50,51,60,85,100,101,107-120 and has only recently been replaced by CBS(T). The MP2/6-31G*(0.25) method is entirely sufficient to capture the nature of base stacking, and therefore all conclusions reported with this method remain qualitatively valid. However, the MP2/6-31G*(0.25) method does not reach quantitative accuracy. Extensive comparison between the MP2/6-31G*(0.25) method and the CBS(T) data can be found in the literature.^{59,121} For example, for a set of 17 diverse structures of stacked cytosine dimer, the difference between the CBS(T) and MP2/6-31G*(0.25) values is in the range ± 0.3 to -2.1 kcal·mol⁻¹ while the CBS(T) energies vary from +2.5 to -10.0 kcal·mol⁻¹. Similarly, the difference between CBS(T) and MP2/6-31G*(0.25) base-pair step stacking energies varies from -1.1 to -4.6 kcal·mol^{-1.59} The difference is not uniform as it arises due to the two effects: (1) underestimation of the stacking at the MP2 level when using the 6-31G*(0.25) basis set and (2) neglecting of typically repulsive higher order electron correlation by the MP2 method, which depends on the composition of the stacked dimer and its geometry.

Since the modern DFT-D methods are fitted to reproduce the CBS(T) data,³⁶ they achieve better accuracy than the MP2/6-31G*(0.25) calculations for base stacking, although they cannot fully replace the CBS(T) benchmarks (see above). On the other hand, MP2/aug-cc-pVDZ calculations, which give more negative (i.e., more stable) stacking energies than the MP2/6-31G*(0.25) method, can be considered as a reasonable compromise between accuracy and computational demands. 25,47,48

Comment on Comparison with Experiments. The calculations provide intrinsic stacking energies that correspond to the gas phase environment. These energies properly reflect the direct interactions between the stacked base pairs stemming from their electronic structure but cannot be directly used to assess the stability of nucleic acids. As noticed earlier,⁵⁹ there is in fact no correlation between gas phase QM data and ΔG°_{37} free energies for B-DNA³⁸ and A-RNA³⁹ used, e.g., in prediction of RNA secondary structure by free energy minimization. 122,123 The present paper further demonstrates that it is even not straightforward to unambiguously predict the order of gas phase stacking stability due to its sensitivity to the geometry.

Definitely, the most salient feature predicted by the gas phase calculations is the weak stacking of the CC/GG steps both in B-DNA and in A-RNA, which is due to unfavorable intrastrand electrostatics and polarization nonadditivity of stacking. In contrast to this, the CC/GG step appears to belong to the most stable steps in B-DNA and A-RNA duplexes in the physiological salt environment. Nevertheless, the measured free energies depend on the content of the G-C and A-T (A-U) Watson Crick base pairs, while the thermodynamic measurements do not allow separating stacking free energies from other contributions. The G-C base pair provides more thermodynamic stability than the A-T (or A-U) one. As noticed earlier for B-DNA, the CC/GG step appears to be modestly less stable considering enthalpy compared with the CG and GC ones.⁵⁹ This appears to reflect the gas phase trend, though the unfavorable electrostatic interaction is definitely substantially compensated for by solvent screening. Interestingly, a recent study¹²⁴ deriving a nonlinear dynamics model of B-DNA with a sequence-dependent stacking term also indicates clearly reduced base stacking force constants for the CC/GG step compared with the CG and GC ones. Thus, at least for the salient stacking features of the CC/GG step, the traces of the gas phase stacking energies reflected by the thermodynamics experiments and models can be observed. In addition, B-DNA helices with consecutive CC/GG steps are known to possess specific structural-dynamics features that were attributed to this anomalous stacking, such as fast base-pair breathing and substantial shift to A-to-B structural intermediate. 125,126

Yet another correspondence between the gas phase and thermodynamics data appears for the average energy separation between A-rich (AA/TT, AT/AT, TA/TA in B-DNA, AA/UU, AU/AU, UA/UA in A-RNA) and G-rich (CC/GG, GC/GC, CG/ CG) sequences in B-DNA and in A-RNA duplexes. The average separation of ΔG°_{37} free energies (1 M NaCl, pH 7) between A-rich and G-rich sequences (i.e., the difference ΔG°_{37} [A-rich] $-\Delta G^{\circ}_{37}[\text{G-rich}])$ is higher in A-RNA³⁹ (equaling 1.89 kcal • mol⁻¹) than in B-DNA³⁸ (equaling 1.26 kcal • mol⁻¹) due to the absence of the 5-methyl group, which increases the attractive van der Waals component of stacking. This makes the A-rich sequences in B-DNA more stable than A-rich sequences in A-RNA. The same effect was also observed for RI-DFT-D/TPSS/LP gas phase interaction energies (Tables 5 and 6), for which the average separation between A-rich and G-rich sequences is 1.92 kcal·mol⁻¹ in A-RNA and 0.27 kcal·mol⁻¹ in B-DNA. The average separation of ΔG°_{37} free energies (1 M NaCl, pH 7) between A-rich and G-rich sequences is thus always positive not only due to the higher gas phase stability of GC base pairs⁸⁹ but also due to the higher stacking stability of G-rich sequences.

Nevertheless, in general, we have to conclude that there is no quantitative correlation between the QM gas phase stacking data, irrespective of the accuracy of the calculations, and the nucleic acids stability. This is not surprising, as the thermodynamics nearest-neighbor parameters measured at the room temperature cover numerous effects not included in our model, which seemingly entirely reshuffle or mask the gas phase stacking stability order. They obviously include the basic solvent screening, which counterbalances a major part of the sequence-dependent electrostatic contribution to stacking, which is quite dominant in the gas phase. However, there are certainly also other effects modulating the sequence dependence of the thermodynamics that are even more elusive. These might include

specific hydration and its interplay with the base pair step structure and sequence, as different steps may have different capabilities to optimize their specific hydration patterns. It also is important to take into consideration the presence of the sugar-phosphate backbone. Although the backbone in principle is the same for each step, different steps prefer different sequence-dependent stacking geometries and these in turn may have different degrees of compatibility with the intrinsically preferred backbone topologies, resulting in another sequencedependent contribution to the energetics. Other specific contributions that can also be sequence-dependent are the ion binding capabilities, entropy terms, etc. Thus, we should be quite careful when interpreting QM results and refrain from claiming that they evaluate stability of nucleic acids. Likewise, experimentalists should not assume that they measure stacking energies in the sense of base to base direct forces. It is more appropriate to say that the experiments derive overall thermodynamics stability associated with stacked base pairs steps of given sequences in DNA or RNA double helices in solution, which result from a complex mixture of energy contributions. The actual interplay of all these forces is presently entirely unknown and deserves further studies, theoretical as well as experimental. 127 Note that the balance of forces may be very different in different types of nucleic acids and even different in different steps of the same nucleic acid form, and may require to be studied literally case by case.

Conclusions

- (1) The variability of the intrinsic stacking energies along the MD trajectories was assessed by calculating mean values of interaction energies together with their standard deviations for base pairs step structures averaged over 400 ps time windows. We have assessed on total 50 structures for each B-DNA step sequence and 10-50 structures for each A-RNA step sequence (see methods for explanation and Tables 1 and 2). The standard deviations vary for the individual steps within the range 0.23-0.77 kcal·mol⁻¹ in B-DNA and within the range 0.19-0.56 kcal·mol⁻¹ in A-RNA. For some sequences, this corresponds to almost 3.0 kcal·mol⁻¹ uncertainty (difference between the most and least stable geometries) in the calculated stacking energies. This uncertainty is as large as $\sim 50\%$ of the actual sequence dependence of base stacking energies, i.e., the energy difference between the most and least stable sequences. Thus, assessing the relative magnitude of gas phase stacking energy using a single geometry for each sequence may be insufficient to obtain an unambiguous order of gas phase stacking energies.
- (2) The comparison of the RI-DFTD/TPSS/LP method with the high level CBS(T) calculations on DNA (Table 3), as well as on RNA (Table 4) systems (one calculation for each unique dinucleotide sequence) confirms that the RI-DFTD/TPSS/LP calculations yield the base pair stacking energies on average higher (less stabilizing) by 0.9 kcal·mol⁻¹ per base pair step compared to CBS(T) results. The difference is twice as large when using the RI-DFT-D method with the TZVP basis set. A large portion of RI-DFT-D/TPSS/LP and CBS(T) stacking differences in B-DNA, as well as in A-RNA, comes from the many body correction term, which is lower (i.e., less repulsive) by 0.52 ± 0.16 kcal·mol⁻¹ (B-DNA) or by 0.48 ± 0.07 kcal·mol⁻¹ (A-RNA) at the MP2/aug-cc-pVDZ level of theory, which is used to correct the

- CBS(T) data. The differences between CBS(T) and RI-DFT-D/TPSS/LP stacking energies are in B-DNA much bigger for the intrastrand stacking than for interstrand stacking. This is the expected result, as the overlap of two bases from two opposite strands is very small compared to the intrastrand arrangement. However, this effect is less pronounced in A-RNA due to larger interstrand overlap of bases.
- (3) The RI-DFT-D/TPSS/LP method may be used as a cheap alternative to the much more expensive CBS(T) method without any significant loss in accuracy; the time savings are approximately on the order of two magnitudes.
- (4) The stabilization (calculated at the RI-DFT-D/TPSS/LP level of theory) within dinucleotide steps is dominated by the intrastrand term. The interstrand and intrastrand terms mutually compensate each other. 51,59 The manybody term is always positive and not larger than 1.7 kcal·mol⁻¹ with the exception of the CC/GG step (both in A-RNA and in B-DNA) showing a large degree of repulsive nonadditivity (more than 3.0 kcal·mol⁻¹). 51,59 The many-body term significantly increases the gap between the most stable and the least stable steps both in A-RNA and in B-DNA. These results are in line with earlier studies on idealized geometries. 51,59 The RI-DFT-D evaluation appears to overestimate the many body term compared to the MP2 calculation.
- (5) When the individual steps containing either purines (R) or pyrimidines (Y) are ranked on the basis of their base—base intrastrand stacking energies (evaluated at the RI-DFT-D/TPSS/LP level of theory), the stability decreases in the order RY < RR < YR ~ YY. This ordering is in qualitative agreement with ordering (again evaluated at the RI-DFT-D/TPSS/LP level of theory) of base pairs found in idealized B-DNA geometries⁵⁹ and in fiber diffraction B-DNA geometries. However, it is different from the ordering (RR < YR ~ RY < YY) obtained more than a decade ago using the MP2/6-31G* level of theory for fiber diffraction geometries.⁴⁰
- (6) B-type DNA stacking is more favorable than A-type RNA stacking on average by 1.14 kcal·mol⁻¹ per base pair step. The A-form calculations were done for RNA, which contains uracil instead of thymine. The absence of the 5-methyl group, which increases the van der Waals component of stacking, is the main source of the difference in average stacking energies of A-RNA and B-DNA. Thus, stacking stabilization in B-DNA and in A-RNA differs on average only marginally.
- (7) The strongest propensity toward the B-type stacking shows the AA/TT step (favored by 3.05 kcal·mol⁻¹), other steps significantly stable in B-DNA are AT/AT (2.83 kcal·mol⁻¹ in favor of B-DNA), AC/GT (2.27 kcal·mol⁻¹ in favor of B-DNA), and CT/AG (1.57 kcal·mol⁻¹ in favor of B-DNA). The only two steps more stable in A-RNA than in B-DNA are UA/UA and GC/GC, though the difference in favor of the A-form is only marginal (0.19 and 0.48 kcal·mol⁻¹, respectively).
- (8) The most stable step both in B-DNA and in A-RNA is the GC/GC step, which is very well separated (1.80 kcal·mol⁻¹ in B-DNA and 2.51 kcal·mol⁻¹ in A-RNA) from the second stable step, the CG/CG step. On the other side of the stability ordering lies the least stable step CC/GG, again well separated from the previous step both in B-DNA (2.38 kcal·mol⁻¹) and in A-RNA (1.03 kcal·mol⁻¹).

- (9) The A-RNA and B-DNA stacking using fiber diffraction geometries is, for some steps, rather different than stacking in MD generated (average-energy) geometries. For example, the CG/CG step is almost equally stable in B-DNA and in A-RNA at MD generated (averageenergy) geometries (by 0.23 kcal·mol⁻¹ in favor of B-DNA), but much more stable in B-DNA at the fiber diffraction geometry (by 5.63 kcal·mol⁻¹ in favor of B-DNA). Similarly, the UA/UA step, which is, albeit marginally, more stable in A-RNA at MD generated (average-energy) geometries, becomes significantly more stable in B-DNA at the fiber diffraction geometry (by 2.98 kcal·mol⁻¹ in favor of B-DNA). These findings emphasize the importance of the proper choice of geometries and show that fiber diffraction geometries are not sufficiently accurate for interaction energy calculations.⁵⁶
- (10) In general, no correlation between the sequence dependence of gas phase stacking energies and the sequence dependence of ΔG°₃₇ free energies (1 M NaCl, pH 7) used in nearest-neighbor models^{38,39} was found either for B-DNA or for A-RNA. This reflects the complexity of the balance of forces that are responsible for the sequence dependence of thermodynamics stability of nucleic acids, which masks the effect of the intrinsic interactions between the stacked base pairs. Nevertheless, as discussed, the salient properties of the CC/GG stacking at least partially influence the properties of nucleic acids.
- (11) The present study shows that stacking geometries generated via MD simulations represent a useful complement of stacking energy calculations using carefully prepared idealized and high resolution X-ray structures. The main advantage of MD geometries is the ability of the simulations to reveal variability of stacking energies due to the dynamics of the system. Nevertheless, the limitations of the MD approach should still be kept in mind. The averaged simulated structures differ somewhat from the target experimental structures (in case of B-form reduced helical twist and negative slide was observed), and the force field also almost certainly does not fully capture all the aspects of the sequence dependent structural variability.

Acknowledgment. This contribution was supported by the Ministry of Education of the Czech Republic [grant numbers MSM6198959216, AVOZ50040507, AVOZ50040702, LC512, and LC06030], by the Grant Agency of the Academy of Sciences of the Czech Republic [IAA400040802], and by the Grant Agency of the Czech Republic [grant number 203/09/1476]. This work was also a part of the research project Z4 055 0506. The support from Praemium Academiae, Academy of Sciences of Czech Republic, dedicated to P.H., is also acknowledged.

Supporting Information Available: Studied geometries (xyz coordinates), explanation of calculation of the CBS(T) stacking energies, tables of interaction energies at different sets of geometries, differences between stacking energies computed at the CBS(T) and RI-DFT-D/TPSS/LP levels opf theory, ΔCCS-D(T) corrections of ten unique base pair steps, orderings of ten unique dinucleotide steps evaluated at different levels of theory or using different sets of geometries, intrastrand RI-DFT-D/TPSS/LP stacking energies at different geometries, RI-DFT-D/TPSS/LP interaction energies of AA/TT, AT/AT, and TA/TA

B-DNA steps with thymine replaced by uracil, values of twist and slide base step parameters in MD simulated B-DNA duplex, differences between B-DNA and A-RNA interaction energies, graphs of CBS(T) stacking energies plotted against RI-DFT-TPSS/TZVP and RI-DFT-TPSS/LP stacking energies, comparison of idealized and average-energy geometries for the TG/CA base pair step, graphs of orderings of ten unique steps in B-DNA and A-RNA at different geometries, and complete refs 54 and 72. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

Soc. 2002, 124, 10887-10893.

- (1) Arnott, S.; Hukins, D. W. L. Biochem. Biophys. Res. Commun. 1972, 47, 1504–1511.
- (2) Nir, E.; Grace, L.; Brauer, B.; de Vries, M. S. *J. Am. Chem. Soc.* **1999**, *121*, 4896–4897.
- (3) Nir, E.; Kleinermanns, K.; de Vries, M. S. *Nature* **2000**, *408*, 949–951
- (4) Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; de Vries, M. S. *Phys. Chem. Chem. Phys.* **2002**, *4*, 740–750.
- (5) Nir, E.; Muller, M.; Grace, L. I.; de Vries, M. S. Chem. Phys. Lett. 2002, 355, 59-64.
- (6) Plutzer, C.; Hunig, I.; Kleinermanns, K.; Nir, E.; de Vries, M. S. *ChemPhysChem* **2003**, *4*, 838–842.
 - (7) Hobza, P.; Sponer, J. J. Am. Chem. Soc. 2002, 124, 11802–11808.
 - (8) Jurecka, P.; Hobza, P. J. Am. Chem. Soc. 2003, 125, 15608–15613.
- (9) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. J. Am. Chem. Soc. **2002**, 124, 104–112.
- J. Am. Chem. Soc. 2002, 124, 104–112.(10) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. J. Am. Chem.
- (11) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. Abstr. Pap. Am. Chem. Soc. 2003, 225, U476.
- (12) Sinnokrot, M. O.; Sherrill, C. D. J. Phys. Chem. A 2004, 108, 10200–10207.
- (13) Tauer, T. P.; Sherrill, C. D. J. Phys. Chem. A **2005**, 109, 10475–10478.
 - (14) Zhao, Y.; Truhlar, D. G. J. Phys. Chem. A 2005, 109, 4209–4212.
- (15) Manojkumar, T. K.; Kim, D.; Kim, K. S. J. Chem. Phys. 2005, 122.
- (16) Piacenza, M.; Grimme, S. J. Am. Chem. Soc. 2005, 127, 14841–14848.
- (17) Tsuzuki, S.; Uchimaru, T.; Mikami, M. J. Phys. Chem. A 2006, 110, 2027–2033.
- (18) Podeszwa, R.; Bukowski, R.; Szalewicz, K. J. Phys. Chem. A **2006**, 110, 10345–10354.
- (19) DiStasio, R. A.; von Helden, G.; Steele, R. P.; Head-Gordon, M. *Chem. Phys. Lett.* **2007**, *437*, 277–283.
- (20) Lee, E. C.; Kim, D.; Jurecka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. J. Phys. Chem. A **2007**, 111, 3446–3457.
- (21) Takatani, T.; Sherrill, C. D. Phys. Chem. Chem. Phys. 2007, 9, 6106-6114
- (22) Pitonak, M.; Neogrady, P.; Rezac, J.; Jurecka, P.; Urban, M.; Hobza, P. *J. Chem. Theory Comput.* **2008**, *4*, 1829–1834.
- (23) Cooper, V. R.; Thonhauser, T.; Puzder, A.; Schroder, E.; Lundqvist, B. I.; Langreth, D. C. *J. Am. Chem. Soc.* **2008**, *130*, 1304–1308.
- (24) Li, S.; Cooper, V. R.; Thonhauser, T.; Lundqvist, B. I.; Langreth, D. C. *J. Phys. Chem. B* **2009**, *113*, 11166–11172.
 - (25) Cysewski, P. New J. Chem. **2009**, 33, 1909–1917.
- (26) Morgado, C. A.; Jurecka, P.; Svozil, D.; Hobza, P.; Sponer, J. *J. Chem. Theory Comput.* **2009**, *5*, 1524–1544.
- (27) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. J. Phys. Chem. A **2009**, 113, 10146–10159.
- (28) Sherrill, C. D.; Sumpter, B. G.; Sinnokrot, M. O.; Marshall, M. S.; Hohenstein, E. G.; Walker, R. C.; Gould, I. R. *J. Comput. Chem.* **2009**, *30*, 2187–2193.
 - (29) Grafenstein, J.; Cremer, D. J. Chem. Phys. 2009, 130.
- (30) Sponer, J.; Leszczynski, J.; Hobza, P. J. Phys. Chem. 1996, 100, 1965–1974.
- (31) Sponer, J.; Leszczynski, J.; Hobza, P. J. Phys. Chem. 1996, 100, 5590–5596.
- (32) Sponer, J.; Burda, J. V.; Sabat, M.; Leszczynski, J.; Hobza, P. J. Phys. Chem. A **1998**, 102, 5951–5957.
- (33) Sponer, J.; Sponer, J. E.; Gorb, L.; Leszczynski, J.; Lippert, B. J. Phys. Chem. A **1999**, 103, 11406–11413.
- (34) Spackova, N.; Cheatham, T. E.; Ryjacek, F.; Lankas, F.; van Meervelt, L.; Hobza, P.; Sponer, J. *J. Am. Chem. Soc.* **2003**, *125*, 1759–1769.
- (35) Reha, D.; Kabelac, M.; Ryjacek, F.; Sponer, J.; Sponer, J. E.; Elstner, M.; Suhai, S.; Hobza, P. J. Am. Chem. Soc. **2002**, 124, 3366–3376.

- (36) Jurecka, P.; Cerny, J.; Hobza, P.; Salahub, D. R. J. Comput. Chem. **2007**, 28, 555–569.
- (37) Cerny, J.; Jurecka, P.; Hobza, P.; Valdes, H. J. Phys. Chem. A 2007, 111, 1146-1154.
- (38) SantaLucia, J.; Hicks, D. Annu. Rev. Biophys. Biomol. Struct. 2004, 33, 415–440.
- (39) Xia, T. B.; SantaLucia, J.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X. Q.; Cox, C.; Turner, D. H. *Biochemistry* 1998, 37, 14719–14735.
- (40) Alhambra, C.; Luque, F. J.; Gago, F.; Orozco, M. J. Phys. Chem. B 1997, 101, 3846–3853.
- (41) Bansal, M.; Bhattacharyya, D.; Ravi, B. Comput. Appl. Biosci. **1995**, 11, 281–287.
 - (42) Lu, X. J.; Olson, W. K. Nucleic Acids Res. 2003, 31, 5108-5121.
- (43) Macke, T.; Case, D. A. Modeling unusual nucleic acid structures. In *Molecular Modeling of Nucleic Acids*; Leontis, N. B., SantaLucia, J., Eds.; American Chemical Society: Washington, DC, 1998; pp 379–393.
- (44) Schneider, B.; Neidle, S.; Berman, H. M. *Biopolymers* **1997**, *42*, 113–124.
- (45) Gorin, A. A.; Zhurkin, V. B.; Olson, W. K. J. Mol. Biol. 1995, 247, 34–48.
- (46) ElHassan, M. A.; Calladine, C. R. Phil. Trans. R. Soc. London Ser. A-Math. Phys. Eng. Sci. 1997, 355, 43–100.
 - (47) Cysewski, P. J. Mol. Struct. (THEOCHEM) 2008, 865, 36-43.
- (48) Cysewski, P.; Czyznikowska, Z.; Zalesny, R.; Czelen, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2665–2672.
- (49) Czyznikowska, Z. J. Mol. Struct. (THEOCHEM) 2009, 895, 161–167
- (50) Sponer, J.; Florian, J.; Ng, H. L.; Sponer, J. E.; Spackova, N. Nucleic Acids Res. 2000, 28, 4893–4902.
- (51) Sponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. *Biophys. J.* **1997**, *73*, 76–87.
 - (52) Sponer, J.; Kypr, J. J. Biomol. Struct. Dyn. 1993, 11, 277–292.
- (53) Svozil, D.; Kalina, J.; Omelka, M.; Schneider, B. *Nucleic Acids Res.* **2008**, *36*, 3690–3706.
 - (54) Olson, W. K.; et al. J. Mol. Biol. 2001, 313, 229-237.
- (55) Fiethen, A.; Jansen, G.; Hesselmann, A.; Schotz, M. J. Am. Chem. Soc. 2008, 130, 1802.
- (56) Sponer, J.; Riley, K. E.; Hobza, P. Phys. Chem. Chem. Phys. 2008, 10, 2595–2610.
 - (57) Sponer, J.; Kypr, J. Int. J. Biol. Macromol. 1994, 16, 3-6.
 - (58) Sponer, J.; Hobza, P. J. Am. Chem. Soc. 1994, 116, 709-714.
- (59) Sponer, J.; Jurecka, P.; Marchan, I.; Luque, F. J.; Orozco, M.; Hobza, P. *Chem.—Eur. J.* **2006**, *12*, 2854–2865.
- (60) Cysewski, P.; Czyznikowska-Balcerak, Z. J. Mol. Struct. (THEOCHEM) 2005, 757, 29–36.
- (61) Langner, K. M.; Sokalski, W. A.; Leszczynski, J. J. Chem. Phys. 2007, 127.
- (62) Hill, J. G.; Platts, J. A. Phys. Chem. Chem. Phys. 2008, 10, 2785–2791.
 - (63) Sponer, J.; Kypr, J. J. Mol. Biol. 1991, 221, 761-764.
- (64) Dockbregeon, A. C.; Chevrier, B.; Podjarny, A.; Johnson, J.; Debear, J. S.; Gough, G. R.; Gilham, P. T.; Moras, D. *J. Mol. Biol.* **1989**, 209, 459–474.
- (65) Beveridge, D. L.; McConnell, K. J. Curr. Opin. Struct. Biol. 2000, 10, 182–196.
 - (66) Giudice, E.; Lavery, R. Acc. Chem. Res. 2002, 35, 350-357.
 - (67) Cheatham, T. E. Curr. Opin. Struct. Biol. 2004, 14, 360-367.
- (68) Auffinger, P.; Westhof, E. Curr. Opin. Struct. Biol. 1998, 8, 227–236.
- (69) Orozco, M.; Perez, A.; Noy, A.; Luque, F. J. Chem. Soc. Rev. **2003**, *32*, 350–364.
- (70) McDowell, S. E.; Spackova, N.; Sponer, J.; Walter, N. G. *Biopolymers* **2007**, *85*, 169–184.
- (71) Olson, W. K.; Colasanti, A.; Li, Y.; Zheng, W. G.; Zhurkin, V. B. DNA simulation benchmarks as revealed by X-ray structures. In *Computational studies of RNA and DNA*, Sponer, J., Lankas, F., Eds.; Springer: Dordrecht, The Netherlands, 2006; pp 235–257.
 - (72) Beveridge, D. L.; et al. Biophys. J. 2004, 87, 3799-3813.
- (73) Barone, F.; Lankas, F.; Spackova, N.; Sponer, J.; Karran, P.; Bignami, M.; Mazzei, F. *Biophys. Chem.* **2005**, *118*, 31–41.
- (74) Varnai, P.; Zakrzewska, K. Nucleic Acids Res. 2004, 32, 4269–4280.
- (75) Wang, J. M.; Cieplak, P.; Kollman, P. A. J. Comput. Chem. 2000, 21, 1049–1074.
- (76) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.
- (77) Lankas, F.; Sponer, J.; Langowski, J.; Cheatham, T. E. *Biophys. J.* **2003**, *85*, 2872–2883.
- (78) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

- (79) Perez, A.; Lankas, F.; Luque, F. J.; Orozco, M. Nucleic Acids Res. 2008, 36, 2379–2394.
- (80) Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham, T. E.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H.; Thayer, K. M.; Varnai, P. *Biophys. J.* **2005**, *89*, 3721–3740.
- (81) Dixit, S. B.; Beveridge, D. L. *Bioinformatics* **2006**, 22, 1007–1009.
- (82) Reblova, K.; Lankas, F.; Razga, F.; Krasovska, M. V.; Koca, J.; Sponer, J. *Biopolymers* **2006**, *82*, 504–520.
- (83) Schneider, B.; Moravek, Z.; Berman, H. M. Nucleic Acids Res. **2004**, *32*, 1666–1677.
- (84) Moody, E. M.; Bevilacqua, P. C. J. Am. Chem. Soc. 2003, 125, 2032–2033.
- (85) Mignon, P.; Loverix, S.; Steyaert, J.; Geerlings, P. Nucleic Acids Res. 2005, 33, 1779–1789.
- (86) Vanommeslaeghe, K.; Mignon, P.; Loverix, S.; Tourwe, D.; Geerlings, P. J. Chem. Theory Comput. 2006, 2, 1444–1452.
- (87) Gil, A.; Branchadell, V.; Bertran, J.; Oliva, A. J. Phys. Chem. B **2009**, 113, 4907–4914.
- (88) Antony, J.; Bruske, B.; Grimme, S. Phys. Chem. Chem. Phys. 2009, 11, 8440–8447.
- (89) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. Phys. Chem. Chem. Phys. **2006**, *8*, 1985–1993.
- (90) Kubar, T.; Jurecka, P.; Cerny, J.; Rezac, J.; Otyepka, M.; Valdes,
- H.; Hobza, P. J. Phys. Chem. A 2007, 111, 5642–5647.
 (91) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. Chem. Phys. Lett. 1989, 162, 165–169.
- (92) Eichkorn, K.; Treutler, O.; Ohm, H.; Haser, M.; Ahlrichs, R. Chem. Phys. Lett. **1995**, 240, 283–289.
 - (93) Baerends, E. J.; Ellis, D. E.; Ros, P. Chem. Phys. 1973, 2, 41-51.
 - (94) Whitten, J. L. J. Chem. Phys. 1973, 58, 4496-4501.
- (95) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. J. Chem. Phys. 1979, 71, 3396–3402.
- (96) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Chem. Phys. Lett. 1998, 286, 243–252.
 - (97) Molpro 2006.1, A Package of Ab Initio Programs, 2006
- (98) Sponer, J.; Hobza, P. Collect. Czech. Chem. Commun. 2003, 68, 2231–2282.
- (99) Leininger, M. L.; Nielsen, I. M. B.; Colvin, M. E.; Janssen, C. L. J. Phys. Chem. A **2002**, 106, 3850–3854.
 - (100) Hobza, P.; Sponer, J. Chem. Rev. 1999, 99, 3247-3276.
- (101) Florian, J.; Sponer, J.; Warshel, A. J. Phys. Chem. B 1999, 103, 884–892
- (102) Noy, A.; Perez, A.; Laughton, C. A.; Orozco, M. Nucleic Acids Res. 2007, 35, 3330–3338.
 - (103) Vertessy, B. G.; Toth, J. Acc. Chem. Res. 2009, 42, 97-106.
- (104) Leontis, N. B.; Stombaugh, J.; Westhof, E. Nucleic Acids Res. 2002, 30, 3497–3531.

- (105) Sponer, J. E.; Leszczynski, J.; Sychrovsky, V.; Sponer, J. J. Phys. Chem. B 2005, 109, 18680–18689.
- (106) Kroonbatenburg, L. M. J.; Vanduijneveldt, F. B. J. Mol. Struct. (THEOCHEM) 1985, 22, 185–199.
- (107) Luo, R.; Gilson, H. S. R.; Potter, M. J.; Gilson, M. K. *Biophys. J.* **2001**, *80*, 140–148.
- (108) Mao, L. S.; Wang, Y. L.; Hu, X. C. J. Phys. Chem. B 2003, 107, 3963–3971.
- (109) Dracinsky, M.; Castano, O. Phys. Chem. Chem. Phys. 2004, 6, 1799–1805.
- (110) Cysewski, P.; Czyznikowska-Balcerak, Z.; Szefler, B. *Pol. J. Chem.* **2003**, *77*, 1287–1300.
- (111) Rutledge, L. R.; Wheaton, C. A.; Wetmore, S. D. *Phys. Chem. Chem. Phys.* **2007**, *9*, 497–509.
- (112) Rutledge, L. R.; Campbell-Verduyn, L. S.; Hunter, K. C.; Wetmore, S. D. *J. Phys. Chem. B* **2006**, *110*, 19652–19663.
- (113) Seio, K.; Ukawa, H.; Shohda, K.; Sekine, M. J. Biomol. Struct. Dyn. 2005, 22, 735–746.
- (114) Rauwolf, C.; Mehlhorn, A.; Fabian, J. Collect. Czech. Chem. Commun. 1998, 63, 1223-1244.
- (115) Alagona, G.; Ghio, C.; Monti, S. J. Phys. Chem. A 1998, 102, 6152-6160.
- (116) Lee, N. K.; Park, S.; Kim, S. K. J. Chem. Phys. **2002**, 116, 7910–7917.
 - (117) Wang, Y. L.; Hu, X. C. J. Chem. Phys. 2002, 117, 1-4.
- (118) Biot, C.; Buisine, E.; Kwasigroch, J. M.; Wintjens, R.; Rooman, M. J. Biol. Chem. **2002**, 277, 40816–40822.
- (119) Wintjens, R.; Biot, C.; Rooman, M.; Lievin, J. J. Phys. Chem. A 2003, 107, 6249–6258.
- (120) McConnell, T. L.; Wetmore, S. D. J. Phys. Chem. B 2007, 111, 2999–3009.
- (121) Jurecka, P.; Sponer, J.; Hobza, P. J. Phys. Chem. B **2004**, 108, 5466–5471.
- (122) Mathews, D. H.; Turner, D. H. Curr. Opin. Struct. Biol. 2006, 16, 270–278.
- (123) Mathews, D. H.; Disney, M. D.; Childs, J. L.; Schroeder, S. J.; Zuker, M.; Turner, D. H. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7287–7292
- (124) Alexandrov, B. S.; Gelev, V.; Monisova, Y.; Alexandrov, L. B.; Bishop, A. R.; Rasmussen, K. O.; Usheva, A. *Nucleic Acids Res.*, in press.
- (125) Warmlander, S.; Sponer, J. E.; Sponer, J.; Leijon, M. J. Biol. Chem. **2002**, 277, 28491–28497.
- (126) Lankas, F.; Cheatham, T. E.; Spackova, N.; Hobza, P.; Langowski, J.; Sponer, J. *Biophys. J.* **2002**, 82, 2592–2609.
- (127) Kopitz, H.; Zivkovic, A.; Engels, J. W.; Gohlke, H. *ChemBioChem* **2008**, *9*, 2619–2622.

JP910788E