

Balance of Attraction and Repulsion in Nucleic-Acid Base Stacking: CCSD(T)/Complete-Basis-Set-Limit Calculations on Uracil Dimer and a Comparison with the Force-Field Description

Claudio A. Morgado,^{†,‡} Petr Jurečka,^{†,§} Daniel Svozil,[†] Pavel Hobza,^{‡,§} and Jiří Šponer^{*,†,‡}

Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, 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 Department of Physical Chemistry, Palacky University, tr. Svobody 26, 771 46, Olomouc, Czech Republic

Received January 7, 2009

Abstract: We have carried out reference quantum-chemical calculations for about 100 geometries of the uracil dimer in stacked conformations. The calculations have been specifically aimed at geometries with unoptimized distances between the monomers including geometries with mutually tilted monomers. Such geometries are characterized by a delicate balance between local steric clashes and local unstacking and had until now not been investigated using reference quantum-mechanics (QM) methods. Nonparallel stacking geometries often occur in nucleic acids and are of decisive importance, for example, for local conformational variations in B-DNA. Errors in the short-range repulsion region would have a major impact on potential energy scans which were often used in the past to investigate local geometry variations in DNA. An incorrect description of such geometries may also partially affect molecular dynamics (MD) simulations in applications where quantitative accuracy is required. The reference QM calculations have been carried out using the MP2 method extrapolated to the complete basis-set limit and corrected for higher-order electron-correlation contributions using CCSD(T) calculations with a medium-sized basis set. These reference calculations have been used as benchmark data to test the performance of the DFT-D, SCS(MI)-MP2, and DFT-SAPT QM methods and of the AMBER molecular-mechanics (MM) force field. The QM methods show close to quantitative agreement with the reference data, albeit the DFT-D method tends to modestly exaggerate the repulsion of steric clashes. The force field in general also provides a good description of base stacking for the systems studied here. However, for geometries with close interatomic contacts and clashes, the repulsion effects are rather severely exaggerated. The discrepancy reported here should not affect the overall stability of MD simulations and qualitative applications of the force field. However, it may affect the description of subtle quantitative effects such as the local conformational variations in B-DNA. Preliminary calculations for two H-bonded uracil base pairs, including one with a C–H…O H-bond, indicate excellent performance of the tested QM methods for all intermonomer distances. The force field, on the other hand, is less satisfactory, especially in the repulsive regions.

Introduction

Base stacking interactions provide a substantial part of the thermodynamics stability of nucleic acids, shape their

structure, and contribute to their dynamics.^{1–7} QM calculations with the inclusion of electron-correlation effects belong to important tools that help to understand the role of stacking interactions in nucleic acids.^{8–18} Although the gas-phase QM calculations do not directly correlate with the thermodynamics stability of nucleic acids due to the complex interplay between molecular forces in solvated nucleic acids, they reveal direct and accurate structure-energy relationships which allow for an exhaustive description of the potential

* Corresponding author e-mail: sponer@ncbr.chemi.muni.cz.

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

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

[§] Palacky University.

energy surfaces (PES) of stacked structures of dimers of nucleic-acid bases.¹⁹ Such calculations characterize the intrinsic forces between the stacked bases, are indispensable to reveal the nature of base stacking, and provide reference data to parametrize/verify other computational methods. Such calculations can also help to partially rationalize experimental data on nucleic acids, albeit direct transferability of the QM data to nucleic acids would require, e.g., inclusion of solvent screening of the gas-phase electrostatics which is specific for distinct nucleic-acid architectures.^{20,21} The QM description of base stacking in general requires application of electron-correlation techniques and diffuse-polarization basis sets of atomic orbitals. The present golden standard for base-stacking computations is the CBS(T) method,¹⁹ which is based on MP2²² calculations with complete basis-set (CBS) extrapolation corrected for higher-order electron-correlation contributions using the CCSD(T)²³ Cizek method²⁴ with a medium-size basis set. In the present paper we have carried out an extensive QM characterization of stacked conformations of the uracil dimer, which is the smallest base stacking system found in nucleic acids. For the purpose of comparison we have also investigated two H-bonded dimers.

Uracil-Dimer Stacking in Nucleic-Acid Structures. The stacking of uracil bases can be seen in a wide variety of RNA structures, spanning a large range of different conformations. Besides the intrastrand and interstrand uracil dimers occurring in canonical 5'-UU-3', 5'-AU-3', and 5'-UA-3' A-RNA base-pair steps, many diverse stacking arrangements between two uracil bases are observed in the experimental structures of RNA hairpins,^{25,26} pseudoknots,^{27,28} dimeric RNA quadruplexes,²⁹ SRE RNA,³⁰ U2 snRNA stem I from *S. Cerevisiae*,³¹ a satellite tobacco mosaic virus/RNA complex,³² and many other RNA systems including obviously the ribosomal structures.^{33,34} Nevertheless, the main aim of this paper is not to investigate the stacking of uracil bases in particular RNAs but to better describe its physicochemical nature and assess the ability of other methods to evaluate stacking interactions, including a standard MM force field.

Preceding Theoretical Studies. The first electron-correlation characterization of the uracil dimer was reported by Šponer et al.³⁵ using the MP2/6-31G*(0.25)^{36,37} method. The study concluded that the stability of stacking originates in the dispersion attraction and that the orientation dependence stems from the electrostatic term. The calculations ruled out several incorrect models of stacking and showed a surprisingly good performance of common molecular-mechanics force fields³⁸ combining Lennard-Jones potentials with Coulombic terms with atom-centered point charges. Kratochvíl et al.³⁹ investigated the potential- and free-energy surfaces of the uracil dimer in the gas phase, employing a combination of *ab initio*, empirical potential, computer simulations, and statistical thermodynamics techniques. They reported eleven low-energy minima in the PES of the uracil dimer: seven H-bonded, one T-shaped, and three stacked structures. The global gas-phase free-energy minimum was predicted to be an H-bonded structure, while stacked structures were found to be less populated than H-bonded ones. N1-methylation shifted the free energy balance in favor of the stacked structures, due to the increased dispersion

energy and elimination of several stable H-bonded structures.⁴⁰ Hobza and Šponer⁴¹ carried out gas-phase gradient optimizations of a stacked uracil dimer that revealed substantial deformations of the monomers in the gas-phase complexes. The free energy of stacking of the uracil dimer in water was estimated by Florian et al.⁴² using MP2/6-31G*(0.25) and Langevin-dipole calculations, concluding that the electrostatic component of stacking, which determines the mutual orientation of bases in the gas phase, is eliminated by solvent screening effects. Later, Leininger et al.⁴³ and Hobza and Šponer⁴⁴ reported the first large-scale MP2 calculations on the uracil dimer supplemented with CCSD(T) corrections, which were in meaningful agreement with the preceding MP2/6-31G*(0.25) data. Very recently Czyżnikowska et al.⁴⁵ studied the rise and twist dependence of the interaction energy components for the undisplaced (the centers of mass stacked one above the other) uracil dimer at the MP2/aug-cc-pVDZ^{46,47} level of theory, using a variational-perturbational decomposition scheme. These calculations confirmed that the second-order dispersion term is basically independent of the twist angle, while the first-order electrostatic term shows a strong angular dependence. Cybulski and Sadlej⁴⁸ characterized an H-bonded and a stacked uracil dimer using the SAPT⁴⁹ and SAPT(DFT)⁵⁰ methods and demonstrated that the ratio of the dispersion term to the total interaction energy clearly differentiates between H-bonding and stacking interactions. So far the most comprehensive methodological study on the interaction energy of the uracil dimer is the work of Pitoňák et al.,⁵¹ who employed a number of QM methods to evaluate the intermolecular interaction on the H-bonded and stacked structures of the dimer reported in the S22 set⁵² of reference geometries. The predicted stacking energy of $-9.77 \text{ kcal mol}^{-1}$, obtained by combining MP2/[aug-cc-pVTZ → aug-cc-pVQZ] and CCSD(T)/[aug-cc-pVDZ → aug-cc-pVTZ] CBS extrapolations, is in very good agreement with the CBS(T) energy reported for the same stacked uracil dimer in the S22 set. This suggests convergence of the computations. Their study also showed that a good estimate of the $\Delta\text{CCSD}(T)$ term can already be obtained with relatively small basis sets, such as 6-31+G**, and that the DFT-D method of Jurečka et al.⁵³ agrees well with estimated CCSD(T)/aug-cc-pVTZ data along the entire potential energy curve (PEC) for both the H-bonded and stacked structures.

Available Experimental Data. The only experimental data on the stabilization energy of the uracil dimer in vacuo was obtained almost thirty years ago by Yanson et al.⁵⁴ The authors reported for the 1-methyluracil dimer a stabilization enthalpy of $9.5 \text{ kcal mol}^{-1}$, measured in a range of temperature of about 295–318 K using mass-field spectrometry. The structures that were present in these unique experiments, however, are not known. Standard QM computations cannot be directly compared with such experiments since they calculate 0 K interaction energies as the difference between pure electronic energies. The experimental stabilization enthalpy would have to be compared to the weighted average of stabilization enthalpies of all populated structures at the experimental conditions.⁵⁵ Another experimental work relevant to the uracil dimer was done by Casaes et al.,⁵⁶ who

measured the gas-phase infrared spectra of jet-cooled uracil clusters, thymine clusters, and uracil•water clusters. The authors found evidence for the presence of several double hydrogen-bonded uracil dimers and for the formation of a larger highly symmetrical cluster. Interestingly, no evidence was seen for T-shaped and stacked structures, which does not support the predictions made by Kratochvíl et al.³⁹ Regarding experimental data in the condensed phase, an estimate of vertical stacking interactions of uridine in aqueous solution can be found in the work of Ts'o et al.,⁵⁷ who reported for this nucleoside a free energy of association of 290 cal mol⁻¹. This number cannot be directly compared to gas-phase calculations either, because it includes the entropy cost of bringing the monomers together and many other terms related to solvation.

The Scope of the Present Study. The present study is aimed differently than the preceding ones. We have shown that AMBER, the leading molecular-mechanics force field for nucleic acids, provides a surprisingly good description of base stacking.³⁵ However, this does not mean that the force field provides an exact description of stacking. In fact, the MM force fields, albeit often providing a very insightful description of nucleic-acid structure and dynamics, are known to have limitations. Besides the overall topologies, nucleic-acid structures and functions are affected by subtle structural details. In canonical double helices these variations are known as local conformational variations, i.e., modest deviations from the average helices that are determined by base sequence and other molecular interactions, such as those caused by protein or drug binding, or crystal packing forces in X-ray experiments.^{58–61} Accurate local positioning of bases is also important elsewhere, for example in the catalytic centers of ribozymes.^{62,63} Local conformational variations are assumed to be primarily caused by base stacking forces and are of primary importance for indirect readout of proteins, sequence-dependent DNA elasticity, etc. Studies, experiments as well as theory, of local conformational variations turned out to be very difficult, as local conformational variations are associated with very subtle energy changes. Computational studies ranging from extensive analysis of the base-stacking PESs up to full-scale MD simulations could provide insights into the sequence-dependence of NA structure.^{64–72} However, such quantitative studies would require an exceptionally high accuracy of the energy description of direct base–base interactions, solvent screening effects, and the conformational space of the sugar–phosphate backbone. Regarding base stacking, local conformational variations are often associated with interactions involving nonparallel (mutually tilted) bases, where close contacts (steric clashes) between nucleobase edges or exocyclic functional groups are combined with local unstacking in other parts of the stacked base-pair steps. Steric effects associated with the amino groups of guanine in the minor groove of CpG B-DNA steps^{59,69} or helical-twist/base-pair-roll redistribution in alternating pyrimidine-purine A-RNA sequences belong to the best documented examples.^{68,73,74} Nonparallel stacked bases are obviously very common in complex noncanonical RNA regions. When the bases are not parallel, the stacking geometry typically reflects

a competition between a segment of the dimer where the monomers are locally unstacked and another segment where the monomers are sterically clashing. An accurate description of this local compression (clash) is important for a proper description of the whole stacked system. We have recently investigated several geometries of the CpG B-DNA steps using the CBS(T) reference method, and these calculations indeed suggested that once geometries with nonparallel bases are considered, the differences between the benchmark calculations and the force field (as well as other methods) can be significant.⁷⁵ For a proper description of local conformational variations the accuracy of the van der Waals (vdW) term of the force field is critical, as it determines the balance between the steric clashes and the partially unstacked regions. The electrostatic part of the stacking energy with its r^{-1} dependence is not contributing significantly to the energy changes associated with the steric contacts. In addition, the electrostatic components of stacking in DNA are in general effectively attenuated due to solvent screening. The solvent screening actually limits the direct applicability of gas-phase QM calculations in studies of DNA local conformational variations, because the dominant role of electrostatics for the gas-phase stacking-energy dependence on helical twist vanishes in solvated nucleic acids.

Herein we report an interaction-energy analysis of the PES of the uracil dimer in stacked conformations, covering a wide range of about 100 structures with specific emphasis given to geometries of dimers in tilted conformations and with unoptimized vertical separation. The reference points are obtained with the CBS(T) method. We first scan a four-dimensional space considering twist between parallel bases, displacements in x and y directions, and the vertical separation between the monomers. Then, we investigate several geometries with tilted bases, i.e., structures with competing clashing and unstacking. While purely compressed or extended dimers with parallel bases cannot occur in real nucleic-acid structures, structures with nonparallel bases can be accompanied with close interatomic contacts (steric clashes) and local unstacking even upon the overall optimization of the vertical separation between the bases or base-pair steps.⁷⁶ The reference CBS(T) calculations serve as a benchmark for several other computational methods: the AMBER force field,³⁸ DFT-D,⁵³ SCS(MI)-MP2,⁷⁷ and DFT-SAPT.⁷⁸ The choice of the AMBER force field is dictated by the need to evaluate the performance of its nonbonded empirical potential—whose vdW term is very similar to the vdW term of most other MM force fields—in a wider region of the PES of nucleic-acid base dimers. The DFT-D and the SCS(MI)-MP2 methods have been chosen because they have proven a cost-efficient way for obtaining interaction energies in good agreement with CCSD(T) or CBS(T) benchmarks. However, both methods have been parametrized against the S22 training set, which only contains noncovalent complexes optimized to a minimum, and their performance in regions other than the minimum has not been as widely tested. We have performed DFT-SAPT energy calculations and decompositions for all of the structures in order to gain some insight into the nature of the interaction.

Among other results we show that the MM force-field calculations are basically capable of providing a satisfactory description of base stacking for the uracil dimer. However, the agreement between the force field and the reference CBS(T) calculations breaks down in the repulsive regions of the PES. The observed differences are caused by the vdW term of the force field and are large enough to substantially affect the description of the fine local conformational variations in nucleic-acid duplexes. Therefore, the correct description of such geometries may represent one of the challenges for future refinements of MM force fields. The DFT-D, SCS(MI)-MP2, DFT-SAPT, and CBS(T) methods are mutually much more consistent in the repulsive region, although the differences between them also increase upon incrementing the repulsion. We have also investigated the dependence of the interaction energy on the intermonomer distance for two H-bonded uracil dimers. The mutual agreement between the QM methods is surprisingly good. On the contrary, the discrepancy between the AMBER force field and the QM methods for short intermonomer distances is much more accentuated, with the force field severely exaggerating the repulsion as the H-bonding distance is decreased. Nevertheless, it is important to point out that, considering the reliability of molecular modeling, the inaccuracy of the force-field description for the H-bonded base pairs is not as painful as the differences reported for stacking. The H-bonded base pairs, in contrast to stacking, usually represent interactions that are well separated from the other interactions. Thus the exaggeration of short-range repulsion in base pairs is likely to lead to mere overestimation of H-bond lengths with no significant effects on the local conformational variations. We would like to note here that our present study focuses on the stacking interactions and that the H-bonded data are more limited. The balance of forces in H-bonded base pairs is very different from that in the stacked structures. This topic requires a thorough analysis, and work in this direction is currently underway. For the information concerning the accuracy of the DFT-D methods for both H-bonded and stacked complexes we refer the reader also to studies that are available in the literature. For instance, Sherrill⁷⁹ showed that Grimme's DFT-D (PBE-D¹⁰¹) performs very well for both H-bonded and stacked complexes. A feeling about the spin-component-scaled methods can be obtained from refs 13 and 80. Finally, the DFT-SAPT method has been shown to give highly accurate results for both H-bonded and π -bonded complexes,¹²⁰ and the accuracy of the intermolecular components has also been studied for both cases.⁴⁸

Computational Details

Geometries. The structures have been obtained in the following way. First, a uracil monomer is optimized using the RI-MP2^{81–83} method along with the cc-pVTZ⁴⁶ basis set. Then, it is placed in the xy plane ($z=0$) with the center of mass coinciding with the origin. The N1–H1 “glycosidic” bond is parallel to the y-axis and is pointing to the direction of negative y values, and the Watson–Crick face of the base is oriented toward the direction of negative x values (Figure 1), with a minor modification for the structures derived to

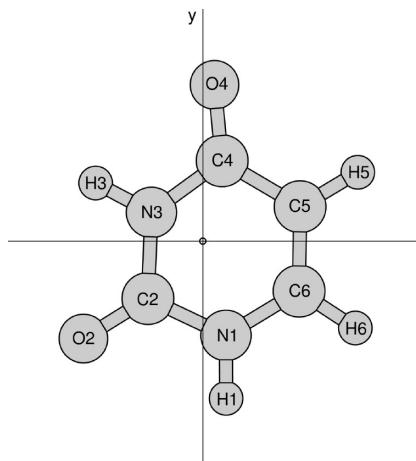


Figure 1. Orientation of the first uracil monomer in the xy plane.

study interatomic clashes (NP1–NP4, see below and Supporting Information). This first monomer is always fixed, and the second monomer is initially superimposed on the first one. Then the position of the second monomer is determined via 6 independent parameters. There are three angles that are applied counter-clockwise and consecutively via rotational matrices: rotation around the x axis (γ), rotation around the y axis (α), and rotation around the z axis (ω). The first two angles are applied to introduce the tilting between bases. When comparing to B-DNA conformational parameters, the α , γ , and ω angles are roughly analogous to propeller twist, buckle, and helical twist angles. Then, the second monomer is shifted along the x and y axes by the parameters Δx and Δy . Finally, the vertical distance is adjusted by Δz . This means that in all cases Δz is equal to the distance between the center of mass of the second monomer and the xy plane, where the first monomer is located (note that $\Delta z \equiv r$ in the figures of the stacked dimers). As the stacking energy is very sensitive to vertical compression or extension of the dimer and the vertical dependence of the stacking energy reflects the balance of vdW interaction terms,⁷⁶ we have scanned the stacking-energy dependence on Δz for most combinations of α , γ , ω , Δx , and Δy .

The structures are grouped into two different sets, P and NP. The set P contains 9 different uracil dimers (P1, P2, ..., P9), in conformations where the planes of the rings are parallel (P) to each other (α , $\gamma = 0$), whereas the set NP contains 4 different dimers (NP1, NP2, NP3, and NP4) in conformations in which one monomer is tilted with respect to the other one, that is, nonparallel (NP) conformations. A dimer is defined by those structural parameters that are kept fixed during the scan, where a free parameter is varied in order to build a potential-energy curve for each dimer. This way we have generated a subset of structures for each dimer, and these subsets add up to a total of 105 distinct geometries. With one exception (P3, for which we vary the twist angle) we have scanned the Δz dependence of the stacking energy for a fixed combination of the other five geometrical parameters.

The four tilted dimers NP1–NP4 have been selected manually based on visual inspection of a number of α , γ , ω , Δx , and Δy combinations in order to obtain four vertical scans with diverse clashes. Obviously, one could imagine

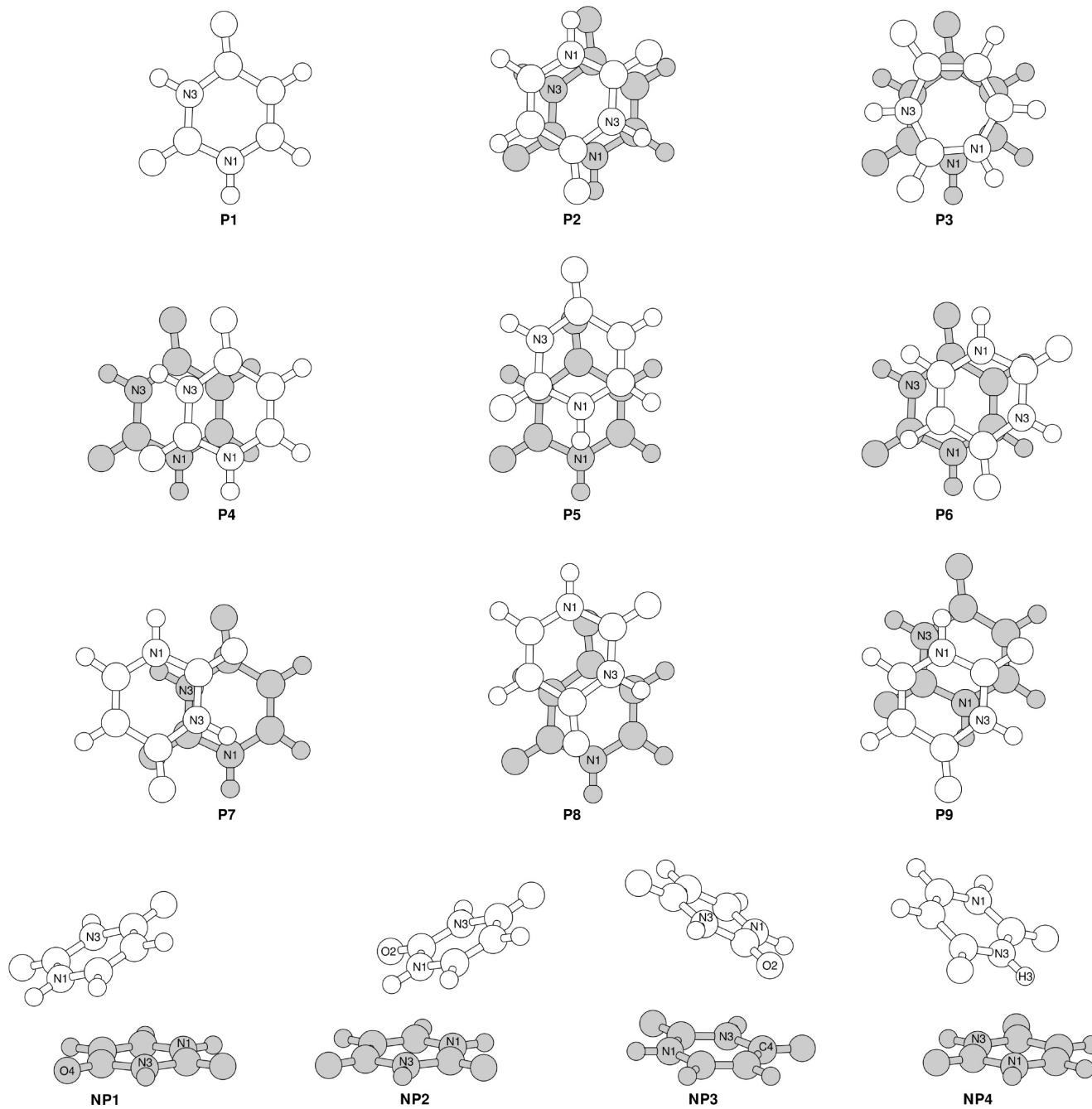


Figure 2. Top and side views of the parallel and nonparallel structures of the uracil dimer, respectively. The first uracil monomer is shown in gray. Some atoms have been labeled in order to aid in the visualization.

many other such geometries that would be worthy of study; however, the calculations remain expensive, and thus we have limited the number of studied structures.

For the H-bonding calculations the WC and “Calcutta” dimers have been selected (denoted HB1 and HB2, respectively), and the scans have been carried out by varying the O₂···N₃ and the O₄···N₃ distances, respectively. Note that the Calcutta base pair contains a C—H···O H-bond. The initial (equilibrium) geometries of both H-bonded dimers are taken from ref 84.

The Cartesian coordinates of all structures are given in the Supporting Information, while Figures 2 and 3 show the geometrical arrangement of the stacked and H-bonded dimers respectively.

Interaction Energies. The interaction energy ΔE^{AB} of a stacked or H-bonded complex is calculated according to the supermolecular approach as the difference in electronic energy between the complex and the isolated monomers (eq 1).^{85,86} Since we have used only rigid monomers, we do not consider monomer deformation energies. In the case of the SCS(MI)-MP2 and CBS(T) methods, all of the interaction energies have been corrected for the basis set superposition error (BSSE) using the counterpoise (CP) technique.^{87,88} This correction is not applied within the DFT-D formalism as this method was parametrized with respect to BSSE-corrected data.⁵³

$$\Delta E^{AB} = E^{AB} - E^A - E^B \quad (1)$$

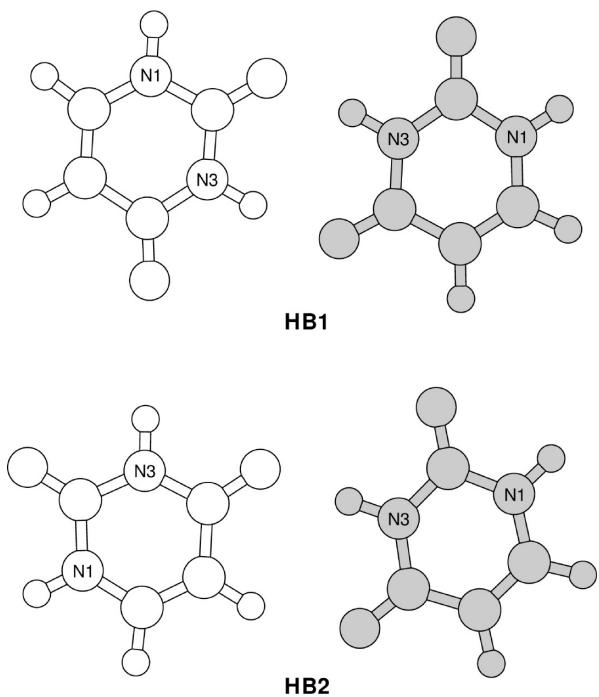


Figure 3. Top views of the H-bonded structures of the uracil dimer. Some atoms have been labeled in order to aid in the visualization.

The supermolecular approach is applied for all of the methods used in the present work, except for DFT-SAPT, where the interaction energy is given by the sum of selected perturbation contributions. In an intermolecular perturbation theory like DFT-SAPT the BSSE takes place only in the δHF term (see below).

Methods

AMBER Nonbonded Empirical Potential. The empirical-potential calculations of the interaction energies have been carried out with a local code, employing the vdW and Coulombic terms of the AMBER force field.³⁸ The atom-centered point charges were derived by means of an electrostatic potential (ESP) fitting at the MP2/aug-cc-pVDZ level of theory using the Merz–Kollman^{89,90} methodology as implemented in the Gaussian 03⁹¹ software package. This nonbonded potential is analogous to that in AMBER, with the only exception that condensed-phase simulations with AMBER are performed with charges derived in a RESP⁹² fitting with the Hartree–Fock (HF) method and the 6-31G* basis set. Given that the HF method overestimates dipole moments, the use of HF charges implicitly accounts for polarization.

The choice of the MP2/aug-cc-pVDZ ESP charges instead of the original AMBER HF/6-31G* RESP charges deserves an explanation. One of the main purposes of this paper is to validate the physical correctness of the nonbonded potential of the AMBER force field. As demonstrated in the literature, meaningful benchmarking of the force field requires that the force-field electrostatic term is derived at the same computational level as used in the reference QM calculations.³⁵ Then the force-field electrostatic parameter set mimics as close as possible the

respective Coulombic term of the full QM calculation. When using standard HF/6-31G* AMBER charges, we would inevitably bias the force field vs QM comparison. It would be often difficult to judge whether the observed differences reflect the basic limitation of the force-field model, the presence of interactions not included in the force-field model, or just the fact that the force-field and the QM calculations consider uncorrelated and correlated electrostatics, respectively. In addition, the present paper is focused on evaluating the capability of the van der Waals term of the force field to describe steric clashes. When introducing steric clashes, the electrostatic energy is basically constant. Note that a comparison of HF and MP2 ESP charges for base stacking is available in the literature.^{93,94} In addition, in one of our earlier stacking papers we have also demonstrated that the use of out-of-plane charges or distributed multipoles does not improve the performance of the force field for base stacking compared with the set of atom-centered ESP-derived point charges.⁹⁵ Thus, the use of atom-centered MP2/aug-cc-pVDZ charges minimizes the difference between the force-field and QM description caused by the ESP electrostatics, allowing unambiguous assessment of the van der Waals part of the force field.

DFT Augmented with an Empirical Dispersion Term (DFT-D). The DFT-D calculations have been performed with the TurboMole 5.8⁹⁶ software package, in combination with a local code that computes the empirical dispersion correction. For the DFT part of the calculation we have used the RI approximation,^{97–100} also known as density fitting. In the DFT-D approach as developed by Jurečka et al.,⁵³ a pairwise additive potential of the form $C_6 r^{-6}$ is used to account for long-range dispersion effects that are poorly described with common density functionals. The dispersion-corrected energy is given by

$$E_{DFT-D} = E_{KS-DFT} + E_{disp} \quad (2)$$

where E_{KS-DFT} is the self-consistent Kohn–Sham energy, and E_{disp} is a term containing the empirical dispersion correction:

$$E_{disp} = - \sum_{ij} f_{damp}(r_{ij}, R_{ij}^0) C_{6,ij} r_{ij}^{-6} \quad (3)$$

$$f_{damp} = \frac{1}{1 + e^{-d(r_{ij}/(s_R R_{ij}^0) - 1)}} \quad (4)$$

In eq 3 r_{ij} is the distance between atoms i and j , R_{ij}^0 is the equilibrium vdW separation between atoms i and j (derived from the atomic vdW radii), and $C_{6,ij}$ is the composite dispersion coefficient for the pair of atoms i and j (calculated from the corresponding atomic C_6 coefficients). The damping function (eq 4) is needed because the r^{-6} form is not valid at short distances, and because some short-range correlation effects are already present in the density functional. In this equation d is a parameter determining the steepness of the damping function, and s_R is a scaling coefficient that adjusts the magnitude of the vdW radius and that has been determined for several density-functional/basis-set combinations. The values of the C_6 atomic coefficients have been taken from the work of Grimme,¹⁰¹ whereas the d and s_R parameters as well as the combination rules for the vdW

radii and the composite dispersion coefficients are the same as those in the work of Jurečka et al.⁵³ Here we use the TPSS¹⁰² functional along with the 6-311++G(3df,3pd)³⁶ basis set. This level of theory complemented with the dispersion correction will hereafter be referred to as DFT-D.

SCS-MP2 for Molecular Interactions (SCS(MI)-MP2).

The SCS(MI)-MP2⁷⁷ calculations have been performed with the Molpro 2006.1¹⁰³ software package, applying the frozen-core and density-fitting^{104,105} approximations. This method is a reparameterization of the original SCS-MP2¹⁰⁶ method, in which the same- (SS) and opposite-spin (OS) components of the MP2 energy are empirically scaled ($E_{MP2} = c_{OS}E_{OS} + c_{SS}E_{SS}$) in an attempt to overcome the deficiencies of the MP2 theory, like the overestimation of the dispersion contribution to the correlation energy. The original SCS-MP2 method reduces the overestimation of the dispersion energy for stacked structures and thus provides very good estimates of the stabilization energy for these systems. Unfortunately, the method also reduces the dispersion for H-bonded structures, which results in this case in an underestimation of the stabilization energy. Using multivariate linear least-squares analysis and the CCSD(T) data in the S22 training set,⁵² Distasio and Head-Gordon⁷⁷ found the optimal parameters that minimized the error between SCS-MP2 theory and CCSD(T), for the cc-pVXZ ($X=T, Q$) and extrapolated cc-pV(XY)Z ($XY=DT, TQ$) levels. The resulting method, known as SCS(MI)-MP2, provides very good estimates of stabilization energies for both planar H-bonded and stacked structures. Here we calculate the SCS(MI)-MP2 interaction energies using the cc-pV(DT)Z extrapolation (cc-pVDZ→cc-pVTZ), for which the same- and opposite-spin optimized parameters are 1.46 and 0.29, respectively. The MP2/CBS limit is approximated according to the following extrapolation scheme

$$E_{XY} = E_{SCF,Y} + \frac{X^3 E_{CORR,X} - Y^3 E_{CORR,Y}}{X^3 - Y^3} \quad Y > X \quad (5)$$

where $X = 2$ and $Y = 3$ for the D→T extrapolation used in this work.

DFT-Symmetry Adapted Perturbation Theory (DFT-SAPT). The DFT-SAPT^{107–110} calculations have been performed with the Molpro 2006.1 software package. DFT-SAPT is a method that uses molecular properties from density functional theory in order to calculate intermolecular interaction energies by means of symmetry-adapted perturbation theory (SAPT). In this method the interaction energy is given as the sum of the first- and second-order energies, plus the δHF term. The first-order energy includes the electrostatic and exchange-repulsion contributions, while the second-order energy includes the induction, exchange-induction, dispersion, and exchange-dispersion contributions. The δHF term is an estimate of higher-order Hartree–Fock contributions and is determined as the difference of the HF interaction energy and the sum of the first- and second-order contributions, with the exception of the dispersion and exchange-dispersion energies. Since the HF interaction energy is calculated with BSSE-corrected monomer energies, the δHF term is BSSE dependent. The interaction energy is given by

eq 6, and the electrostatic, induction, dispersion, and exchange contributions are defined in eqs 7–10.

$$E_{\text{int}} = E_{\text{pol}}^{(1)} + E_{\text{ex}}^{(1)} + E_{\text{ind}}^{(2)} + E_{\text{ex-ind}}^{(2)} + E_{\text{disp}}^{(2)} + E_{\text{ex-disp}}^{(2)} + \delta HF \quad (6)$$

$$E_{\text{elec}} = E_{\text{pol}}^{(1)} \quad (7)$$

$$E_{\text{ind}} = E_{\text{ind}}^{(2)} + E_{\text{ex-ind}}^{(2)} \quad (8)$$

$$E_{\text{disp}} = E_{\text{disp}}^{(2)} + E_{\text{ex-disp}}^{(2)} \quad (9)$$

$$E_{\text{exch}} = E_{\text{ex}}^{(1)} \quad (10)$$

For the energy decomposition we have employed the LPBE0AC¹¹⁰ XC potential with the pure ALDA kernel for both the static and dynamic response, along with the aug-cc-pVDZ basis set. We have also utilized the density-fitting¹¹⁰ approximation. The ionization potential (IP) of the monomer and the energy of the highest occupied molecular orbital (HOMO), which are required to evaluate the shift parameter, have been calculated at the PBE0¹¹¹/aug-cc-pVDZ level of theory.

The relatively small aug-cc-pVDZ basis set has been used because our primary goal is to get an idea about the relative magnitude of the interaction energy components. DFT-SAPT interaction energies are usually underestimated when using this basis set, mainly because the dispersion component is underestimated by about 10–20%.¹¹⁰ This underestimation does not change the conclusions regarding the relative importance of the individual contributions to the interaction energy given below. With a larger basis set DFT-SAPT was shown to provide very accurate total interaction energies in good agreement with the most accurate CCSD(T) calculations.¹¹⁰ Using the aug-cc-pVDZ basis set for the IP calculations does not affect the shift values dramatically, and resulting errors should be smaller than the SAPT basis-set-size errors.

MP2/CBS Corrected for Higher-Order Correlation Effects (CBS(T)).

The MP2 and CCSD(T) calculations have been performed with the Molpro 2006.1 software package, applying the frozen-core approximation. In the case of MP2, we have also applied the density-fitting^{104,105} approximation. We have carried out the MP2/CBS calculations by extrapolating the Hartree–Fock and the correlation energies separately, following the Helgaker et al.^{112,113} extrapolation scheme (eqs 11 and 12), in which E_X is the energy for the basis set with the largest angular momentum X , E_{CBS} is the energy for the complete basis set, and α is a parameter that was fitted in their original work. Herein we have used the aug-cc-pVDZ→aug-cc-pVTZ extrapolation.

$$E_X^{HF} = E_{CBS}^{HF} + Ae^{-\alpha X} \quad (11)$$

$$E_X^{CORR} = E_{CBS}^{CORR} + BX^{-3} \quad (12)$$

Given that higher-order correlation-energy contributions cannot be neglected, we have approximated the CCSD(T)/CBS interaction energies according to the following scheme:

$$\Delta E_{CBS}^{CCSD(T)} = \Delta E_{CBS}^{MP2} + (\Delta E^{CCSD(T)} - \Delta E^{MP2}) \Big|_{6-31+G^{**}} \quad (13)$$

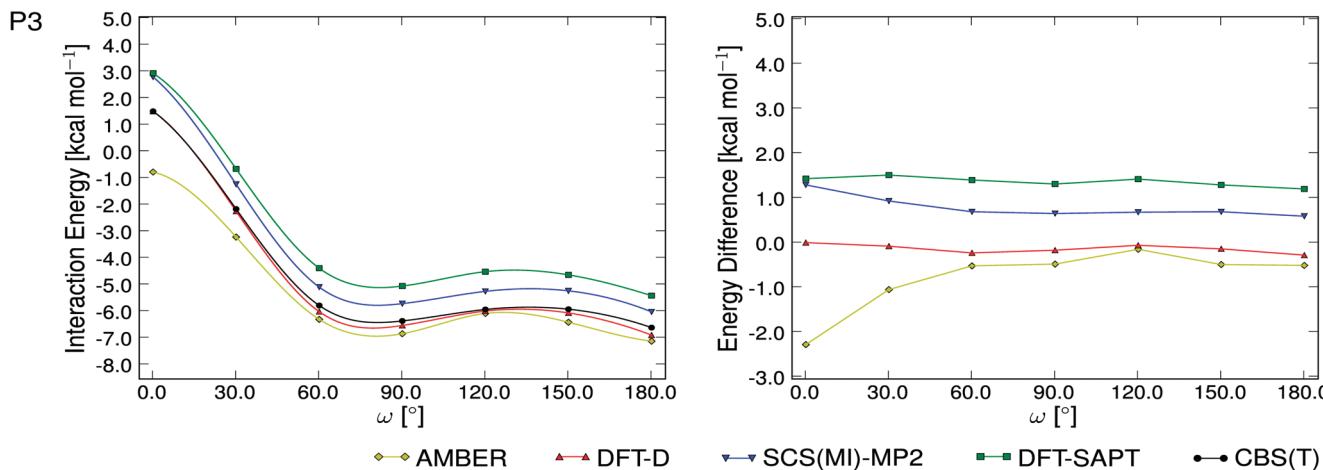


Figure 4. Potential energy curves (left) and energy differences with respect to the CBS(T) data (right) for P3, calculated with the AMBER, DFT-D, SCS(MI)-MP2, DFT-SAPT, and CBS(T) methods. The vertical separation between the monomers is 3.3 Å.

The use of eq 13 is based on the assumption that the difference between the CCSD(T) and MP2 interaction energies is basically basis-set independent and can therefore be evaluated with a small- or medium-size basis set. This assumption was validated for several H-bonded¹¹⁴ and stacked⁴⁴ clusters, including H-bonded and stacked structures of the uracil dimer.¹¹⁵ A CCSD(T)/CBS interaction energy estimated according to eq 13 is also known as CBS(T).¹⁹

Data Analysis. The interaction-energy data have been analyzed in terms of root-mean-squared (rms) errors, maximum absolute deviations (MAX), and XY plots. rms errors are calculated using the standard formula. In this work we define a deviation as the difference between the value obtained with a given method (AMBER, DFT-D, SCS(MI)-MP2, or DFT-SAPT) and the value obtained with CBS(T), and the MAX value for a given dimer simply corresponds to the absolute value of the largest deviation. All of the plots have been generated with the Matplotlib¹¹⁶ graphics package in combination with the SciPy¹¹⁷ and NumPy¹¹⁸ packages. Energies and plots are given in the Supporting Information.

Results and Discussion

The Overall Performance of the Methods. Let us first assess the data qualitatively. Figure 4 shows the dependence of the stacking energy on twist at $r = 3.3$ Å (P3), which is primarily determined by the electrostatic term. The twist energy variation is rather modest due to the lower polarity of uracil compared to cytosine and guanine.³⁵ The DFT-D method achieves a close to exact match to the CBS(T) curve. The AMBER force field is more stabilizing than CBS(T), and also the profile of the twist dependence somewhat deviates from the CBS(T) data. The SCS(MI)-MP2 and DFT-SAPT methods systematically underestimate the binding, while the shapes of the curves basically match the CBS(T) one. In the case of the DFT-SAPT method the underestimation is mainly due to the small basis set used (aug-cc-pVDZ); with a larger basis set yet closer match with the CBS(T) data is expected.

Figure 5 shows the vertical scans for P1 and P2, i.e., the untwisted and antiparallel undisplaced dimers (untwisted and

antiparallel refer to ω angles of 0 and 180°, respectively). When assessing the methods using the vertical scans, the most important region is that $\sim 0.2\text{--}0.3$ Å around the CBS(T) minimum, while the most relevant descriptor is the slope (gradient) of the PEC at a given distance within the repulsive region. This is related to the force associated with the repulsion. Two findings are apparent. First, all methods show very flat energy dependence around the vertical-separation minimum for the untwisted P1 dimer, which has the most repulsive electrostatic arrangement. The AMBER force field gives the shortest optimal vertical separation of the monomers for this arrangement. In contrast, the force field has an excessively steep onset of the repulsion in the short-range repulsion region of the P2 dimer which is also associated with the overestimation of the optimal intermonomer distance. The data for the P2 dimer also suggest that DFT-D is little overestimating the repulsion in the short-separation region, while SCS(MI)-MP2 provides a curve with almost the same slope as the CBS(T) one, albeit the method underestimates the interaction energy. DFT-SAPT also underestimates the interaction energy.

Figure 6 shows the vertical scans for the two untwisted displaced dimers (P4 and P5). These are still in the electrostatically repulsive orientation, but the geometrical overlap of the bases is reduced. The most visible result is the overestimation of the short-range repulsion and of the optimal intermonomer distance by the force field.

Figure 7 summarizes the vertical scans on the four antiparallel displaced dimers (P6–P9). There is again a large exaggeration of the short-range repulsion by the force field. Modest exaggeration of the repulsion by the DFT-D method is also seen. The SCS(MI)-MP2 and DFT-SAPT methods (in particular the latter, due to the relatively small basis set used) are typically shifted toward higher stacking energies for all intermonomer distances. Interestingly, the difference between DFT-SAPT and CBS(T) widens visibly upon reducing the intermonomer separation.

Figure 8 gives the data associated with the vertical scans of the four nonparallel (tilted) dimers with steric clashes (NP1–NP4). The NP1 geometry brings the O4 region of the

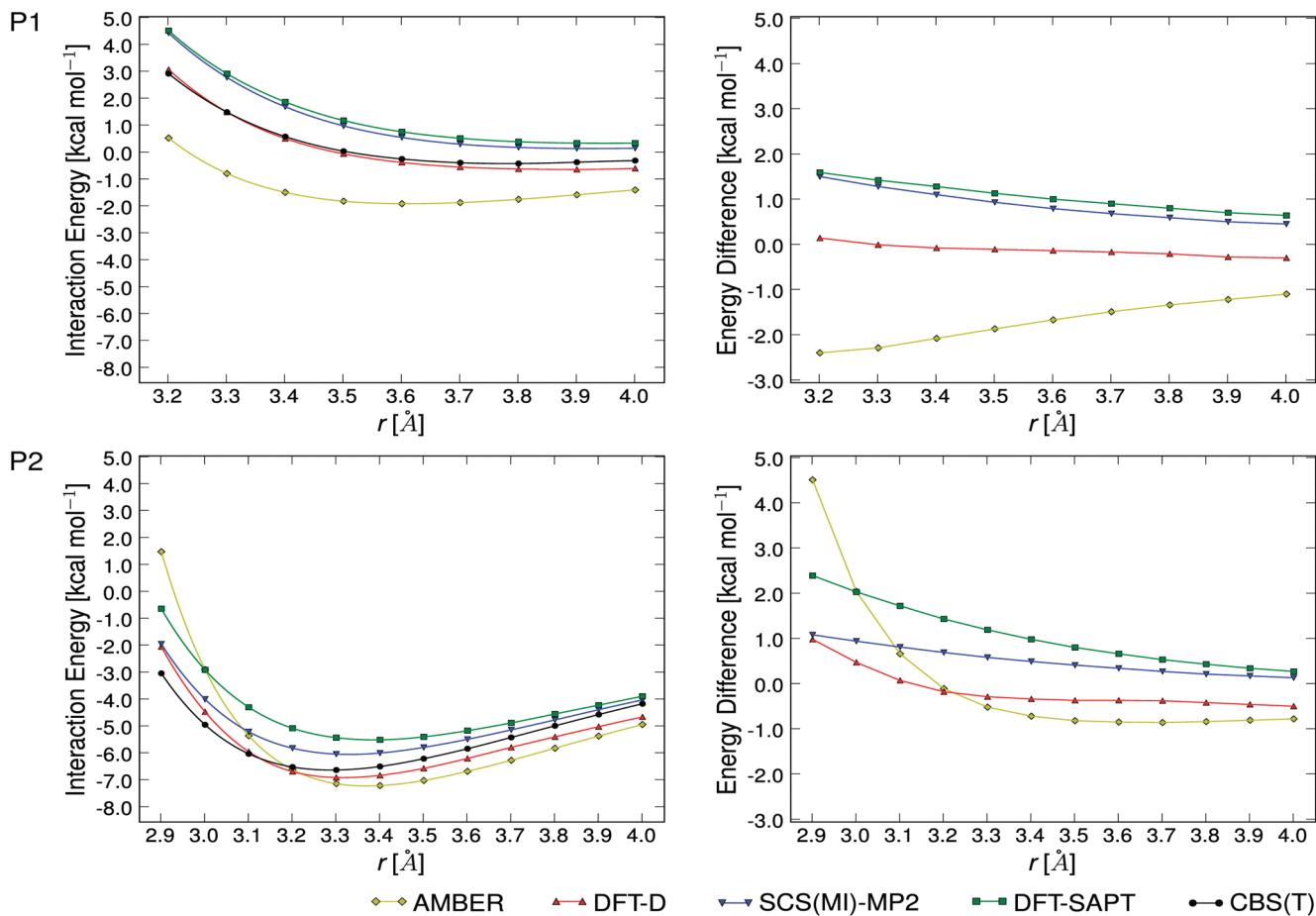


Figure 5. Potential energy curves (left) and energy differences with respect to the CBS(T) data (right) for P1 and P2, calculated with the AMBER, DFT-D, SCS(MI)-MP2, DFT-SAPT, and CBS(T) methods.

first monomer into contact with the N1 region of the second one. In the NP2 geometry the N1/O2 region of the second monomer is in close contact with the aromatic ring of the first one. The NP3 geometry leads to the closest contact between the O2 atom of the second monomer and the C4 atom of the first one, and the NP4 geometry has the N3–H3 bond of the second monomer tilted to the center of the aromatic ring of the first one. The twist angle is 180° for all of the four clashed dimers. Obviously, these four dimers do not guarantee an exhaustive sampling of all possible steric contacts between the two uracils; however, we hope that they are diverse enough to allow us to make correct conclusions. The calculations are quite consistent with the results derived from the calculations on the P1–P9 dimers. The large deviation (excessive repulsion) of the AMBER force field in the short-range repulsion region is very clear, and the difference between DFT-D and CBS(T) is also visible. DFT-D overestimates the binding at larger separations, but then it has a faster and steeper onset of the repulsion upon vertical compression of the dimers. SCS(MI)-MP2 and DFT-SAPT underestimate the interaction energy and the difference between either of these methods, and CBS(T) increases as the intermolecular distance decreases. Both effects are more pronounced for DFT-SAPT, mainly because of the basis set chosen for these calculations.

Finally, Figure 9 gives the energy scans for the WC (HB1) and Calcutta (HB2) H-bonded uracil dimers. The problems

of the AMBER force field to describe the repulsive region of the PES are even more visible than for the stacked dimers, in particular for the Calcutta structure. On the QM side there is essentially a very good agreement between the CBS(T), DFT-D, and SCS(MI)-MP2 methods. DFT-SAPT underestimates the strength of the H-bonding interaction at all distances. However, we need to reiterate that DFT-SAPT calculations were done with a relatively small aug-cc-pVQZ basis set. Specifically for H-bonding, very large basis sets with higher-angular momentum functions are vital.¹¹⁹ The repulsion appears to be well captured by DFT-SAPT. We would like to point out again that the H-bonding data presented here is still preliminary. The issue of the description of the short-range repulsion in H-bonding is even more complex than it is for stacking and will be addressed in the future as it is beyond the scope of the present stacking study.

Interaction-Energy Statistics. rms errors with respect to the CBS(T) reference data are given in Table 1 along with the corresponding MAX values. For the twist-angle scan (P3) the best performer is the DFT-D method, with a rms error of only 0.15 kcal mol⁻¹ and a MAX value of 0.23 kcal mol⁻¹. Both SCS(MI)-MP2 and DFT-SAPT underestimate the strength of the interaction across the entire range of twist angles, showing rms errors of 0.74 and 1.39 kcal mol⁻¹ respectively. On the contrary, AMBER overestimates the attraction with a rms error of 0.61 kcal mol⁻¹.

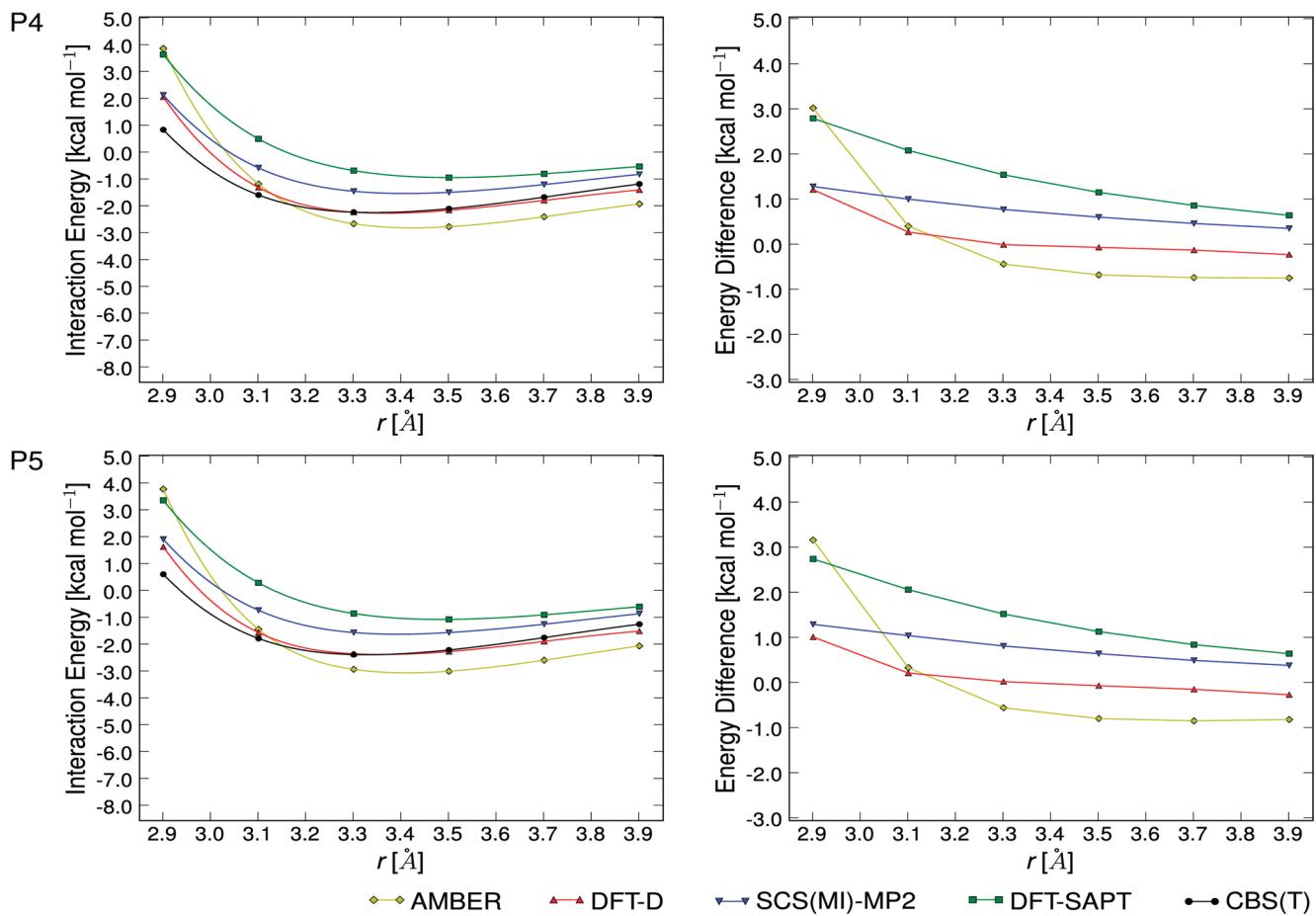


Figure 6. Potential energy curves (left) and energy differences with respect to the CBS(T) data (right) for P4 and P5, calculated with the AMBER, DFT-D, SCS(MI)-MP2, DFT-SAPT, and CBS(T) methods.

Turning now to the performance of the QM methods in the vertical scans of the parallel dimers, we see that the DFT-D method is again the best performer with regards to the chosen statistical descriptors, giving for example the smallest rms deviation ($0.52 \text{ kcal mol}^{-1}$). DFT-D overestimates the strength of the interaction at long distances and underestimates it at short distances (refer to Figures 5–8). The DFT-D deviations around the PEC minimum are typically very small. This is not surprising as the parametrization of DFT-D⁵³ relies mostly on a training set containing noncovalent complexes optimized to a minimum. The fact that DFT-D always underestimates the strength of the interaction at short distances may be rooted in the inability of the density functional to completely account for short-range correlation effects, which are not compensated for by the empirical correction. SCS(MI)-MP2 also shows a small rms deviation with respect to the CBS(T) data ($0.86 \text{ kcal mol}^{-1}$), but, unlike DFT-D, the shape of the SCS(MI)-MP2 curve closely resembles the CBS(T) one for all of the scans. This means that in general this method may produce better energy gradients, in particular in the repulsive regions of the PES. This method, however, systematically underestimates the strength of the interaction. Finally, DFT-SAPT exhibits the largest rms error among the QM methods ($1.37 \text{ kcal mol}^{-1}$), and the method systematically underestimates the strength of the interaction. Note, however, that the DFT-SAPT data were obtained with a relatively small basis set

(aug-cc-pVDZ), while the other methods are calculated (directly or effectively) with a CBS extrapolation.

The AMBER force field shows larger deviations with respect to CBS(T) than the QM methods, with a rms error of $1.53 \text{ kcal mol}^{-1}$. AMBER overestimates the strength of the interaction at long distances and underestimates it at short distances, the only exception being the untwisted undisplaced P1 dimer, where the AMBER energies are lower along the entire range of distances. It partially could be due to increased polarization effects in this particular geometry that would not be captured by a nonpolarizable force field. However, the DFT-SAPT data below do not indicate a significant role of induction. Most likely, the overestimation of the strength of the interaction is due to the underestimation of the repulsion by AMBER, which in turn might be a consequence of an improper description of the anisotropy caused by a particularly unfavorable Pauli repulsion in the P1 conformation. The force field assumes isotropic interactions and spherical atoms, which might not be accurate enough in some specific geometries.⁹⁵

Similar findings are seen in the case of the vertical scans for the nonparallel dimers. With a rms error of $0.55 \text{ kcal mol}^{-1}$ DFT-D is again the best performer with respect to this statistical descriptor. The method remains to overestimate the strength of the interaction at long distances and underestimate it at short distances. The performance of SCS(MI)-MP2 is also good, bearing the smallest MAX value and a

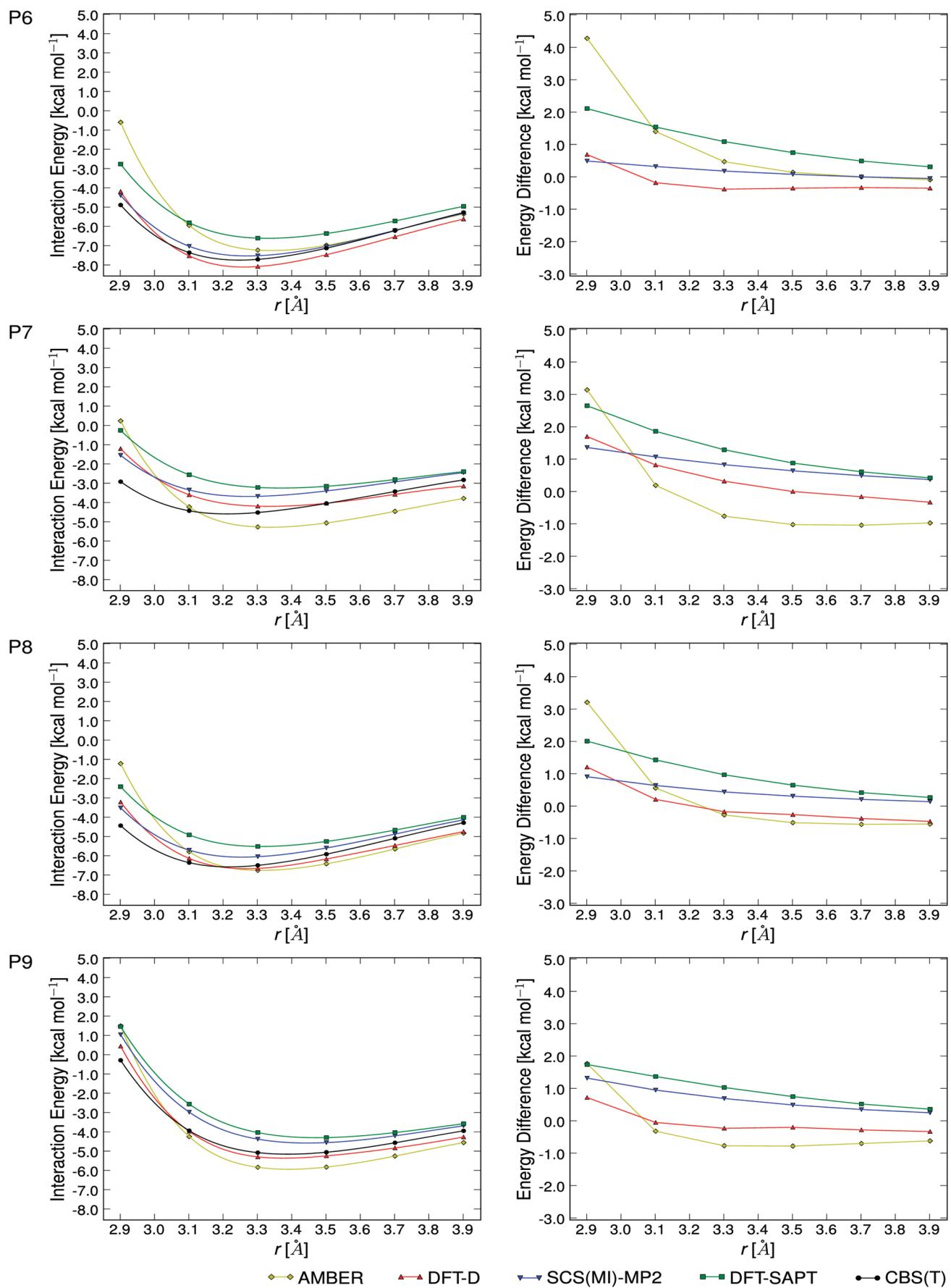


Figure 7. Potential energy curves (left) and energy differences with respect to the CBS(T) data (right) for P6 through P9, calculated with the AMBER, DFT-D, SCS(MI)-MP2, DFT-SAPT, and CBS(T) methods.

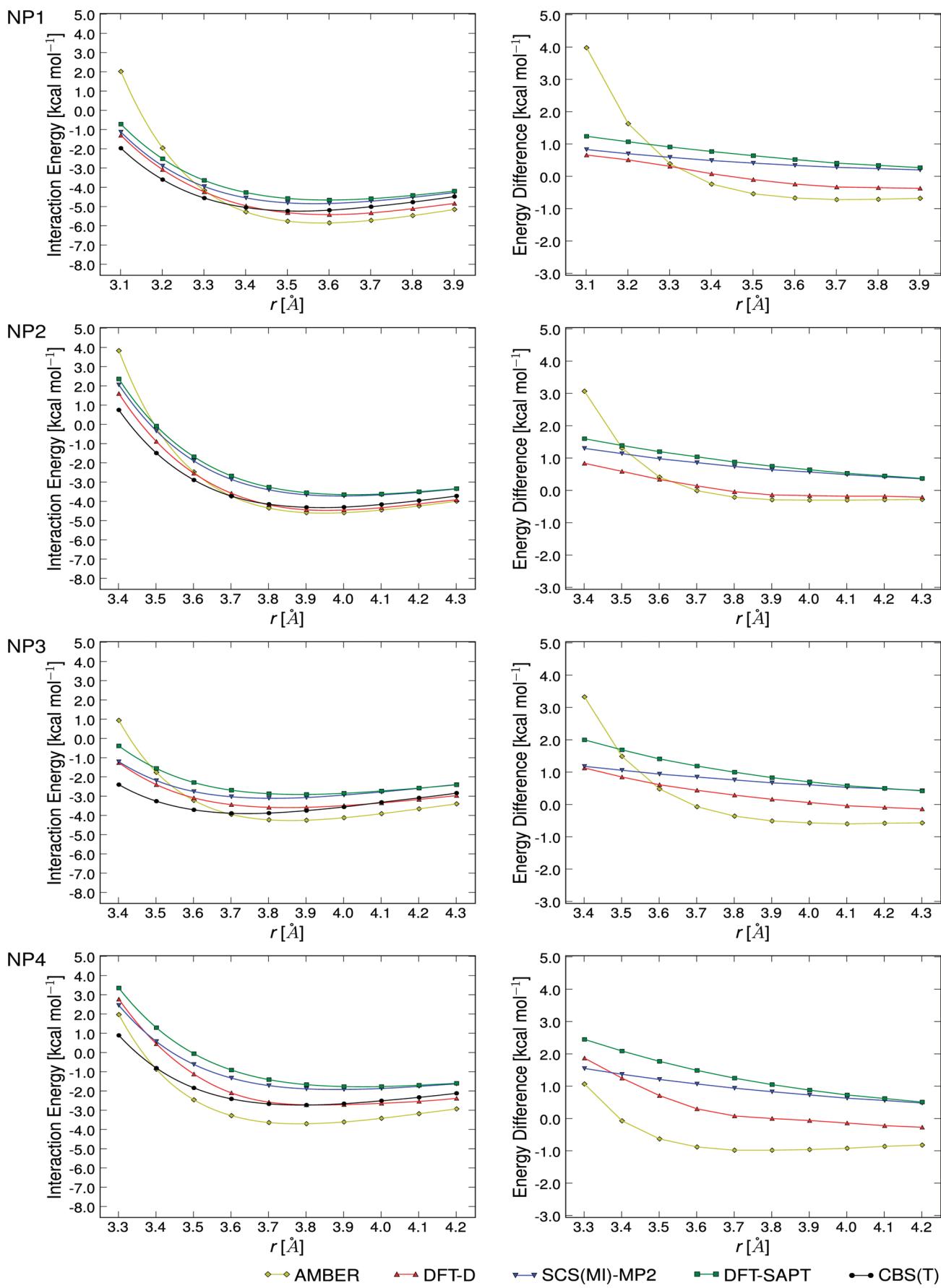


Figure 8. Potential energy curves (left) and energy differences with respect to the CBS(T) data (right) for NP1 through NP4, calculated with the AMBER, DFT-D, SCS(MI)-MP2, DFT-SAPT, and CBS(T) methods.

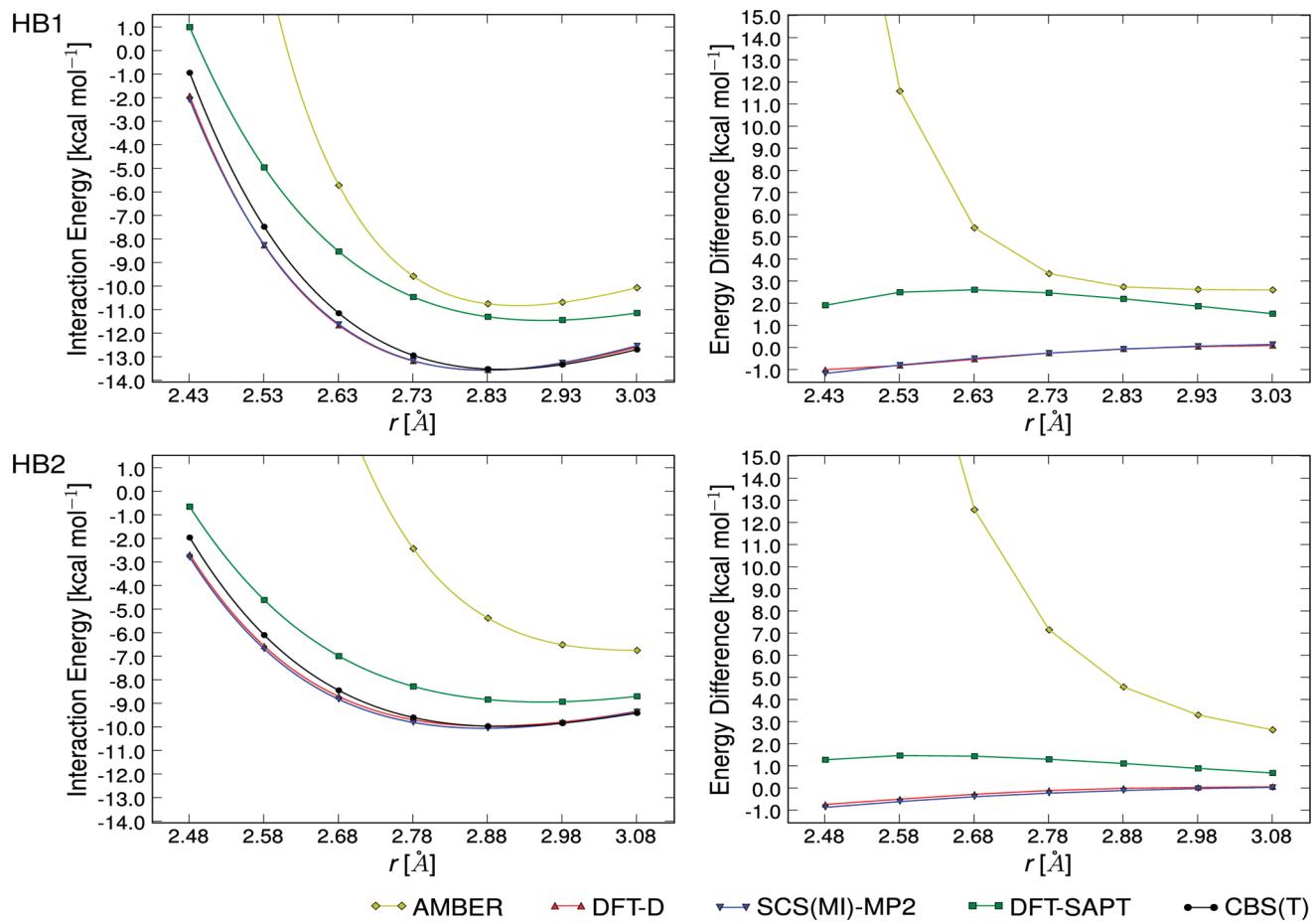


Figure 9. Potential energy curves (left) and energy differences with respect to the CBS(T) data (right) for HB1 and HB2, calculated with the AMBER, DFT-D, SCS(MI)-MP2, DFT-SAPT, and CBS(T) methods.

Table 1. Root-Mean-Squared Deviations (rmsd, kcal mol⁻¹) and Maximum Absolute Deviations (MAX, kcal mol⁻¹) of the AMBER, DFT-D, SCS(MI)-MP2, and DFT-SAPT Interaction Energies with Respect to the CBS(T) Reference Data^a

dimer	AMBER		DFT-D		SCS(MI)-MP2		DFT-SAPT	
	rmsd	MAX	rmsd	MAX	rmsd	MAX	rmsd	MAX
P1	1.74	2.39	0.53	1.50	1.34	2.45	1.36	2.08
P2	1.58	4.52	0.45	0.99	0.60	1.09	1.27	2.40
P3	0.61	1.05	0.15	0.23	0.74	0.93	1.39	1.51
P4	1.36	3.03	0.52	1.22	0.82	1.29	1.69	2.80
P5	1.44	3.17	0.44	1.02	0.84	1.30	1.67	2.75
P6	1.85	4.29	0.41	0.70	0.26	0.50	1.22	2.12
P7	1.50	3.15	0.80	1.71	0.87	1.37	1.51	2.66
P8	1.39	3.22	0.57	1.22	0.52	0.92	1.14	2.02
P9	0.94	1.78	0.36	0.73	0.78	1.33	1.09	1.75
all P ^b	1.53	4.52	0.52	1.71	0.86	2.45	1.37	2.80
NP1	3.03	8.40	0.44	0.83	0.58	1.01	0.86	1.44
NP2	1.09	3.08	0.37	0.85	0.82	1.31	0.98	1.61
NP3	1.24	3.34	0.53	1.14	0.80	1.19	1.16	2.01
NP4	0.85	1.08	0.77	1.88	1.01	1.56	1.43	2.46
all NP	1.77	8.40	0.55	1.88	0.81	1.56	1.13	2.46
HB1	12.09	28.73	0.52	0.98	0.57	1.15	2.21	2.63
HB2	21.91	50.20	0.34	0.72	0.42	0.85	1.22	1.49
all HB	17.70	50.20	0.44	0.98	0.50	1.15	1.78	2.63

^a Some extreme MAX values are not visualized in the figures because they are out of the range of intermolecular distances in those figures. Refer to Tables S2, S3, and S4 in the Supporting Information for the complete data. ^b P3 was not included in the calculation of rmsd and MAX values, i.e., the values in this row correspond to the group of dimers for which vertical scans have been performed.

rms error of 0.81 kcal mol⁻¹. The same error amounts to 1.13 kcal mol⁻¹ in the case of DFT-SAPT. Moreover, the shapes of the SCS(MI)-MP2 curves again resemble the CBS(T) ones, and, particularly when going from the minimum to the repulsive region (see the NP4 curve in Figure

8), this method produces better energy gradients than DFT-D. AMBER shows a larger rms error of 1.77 kcal mol⁻¹, and its deviations are more pronounced in the repulsive region of the PECs. In order to assess the importance of the results, we need to consider which geometries can be

populated in real structures. It is clear that severely compressed structures are not likely to be populated, but structures with shortened intermonomer separations with energies 1–2 kcal mol⁻¹ above the minima on the CBS(T) curves are certainly accessible. For these geometries the force field repulsion is already visibly exaggerated.

With respect to the H-bonding scans the best performance is achieved by the DFT-D and SCS(MI)-MP2 methods, with rmsd values of about ~0.5 kcal mol⁻¹. Both methods overestimate the strength of the interaction at short distances. DFT-SAPT evaluated with the medium-sized basis set underestimates it at all distances, showing a rmsd value of 1.78 kcal mol⁻¹. The large rmsd associated with the AMBER data is due to the huge deviations toward more repulsive values that occur in the regions of short H-bonding distances, in particular for the HB2 structure. The very large repulsion given by the force field for this dimer indicates an excessive repulsion for the C—H···O contact. This can be due to an excessive steepness of the repulsion with the Lennard-Jones 6–12 potential and most likely is also due to a too large atomic radius of that particular hydrogen atom in the AMBER force field. The data show that the force-field performance is better for the stacked dimers than for the H-bonded ones. However, as noted in the introduction, the overestimation of short-range repulsion for H-bonded base pairs may have less serious consequences for the molecular-mechanical studies than errors in the short-range repulsion for stacking. The interaction energy of HB1 at the minimum of the AMBER curve amounts to -10.7 kcal mol⁻¹, whereas for the same geometry Šponer et al. reported an AMBER energy of -12.1 kcal mol⁻¹.⁸⁴ This difference is because the latter value was computed with the original AMBER HF charges, which differ from the MP2 charges used in the present study. The HF charge distribution is more polar, which produces a better stabilization.

Interaction-Energy Gradients. Regions of interatomic contacts are associated with interaction energy gradients, and their inaccurate description may affect the outcome of the calculations, including the population of conformations in MD simulations. Nevertheless, simulations sample also all the additional degrees of freedom which should, in most cases, attenuate the impact of the incorrect description of energy gradients in the clashed regions. In contrast, potential energy scans that were often used to study the local conformational variations can be severely distorted by the incorrect description of the repulsion as in such scans one does not vary enough degrees of freedom to overcome the exaggerated repulsion. It is thus promising to see that the QM methods perform well for the clashes. For instance, for P8 the difference between the interaction energies at $r = 2.9$ and $r = 3.1$ Å (a range that is within 0.3 Å from the CBS(T) minimum) is 1.92, 2.19, 2.50, and 2.92 kcal mol⁻¹, calculated with CBS(T), SCS(MI)-MP2, DFT-SAPT, and DFT-D, respectively. On the contrary, AMBER yields an interaction-energy difference of 4.57 kcal mol⁻¹. Clearly, the SCS(MI)-MP2 value is the closest to the CBS(T) one, and this level of accuracy might be necessary to fully understand the intrinsic interactions in nucleic-acid systems such as those

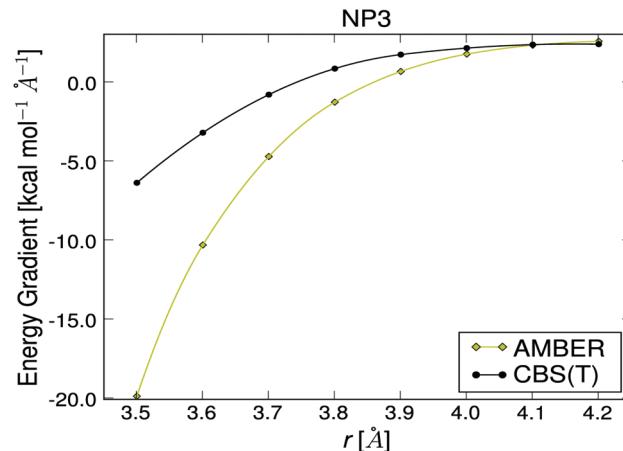


Figure 10. Estimate of the interaction-energy gradient for NP3, calculated with the AMBER and CBS(T) methods.

mentioned above, where variations in the geometry are often accompanied by very subtle energy differences.

Figure 10 presents an estimate of the interaction-energy gradient for the AMBER and CBS(T) PECs of the NP3 dimer. The force field clearly deviates from the CBS(T) data. The NP3 geometry with vertical distance of 3.5 Å is easily accessible on the CBS(T) PEC. At this distance, however, the force field energy gradient is ~3 times larger than the CBS(T) one, so this geometry would be penalized when attempting a force field calculation. The NP3 dimer shows probably the most severe differences among the stacked structures; similar plots for the rest of the stacked dimers can be found in the Supporting Information.

The H-bonding dimers show even more drastic differences between the AMBER and CBS(T) gradients (see Figures S12 and S13), with the same consequences as those described above for the stacked dimers.

DFT-SAPT Interaction-Energy Decomposition. Figures 11 and S14–S23 in the Supporting Information show the DFT-SAPT decompositions for the stacked dimers. Some general trends have emerged from the decomposition analysis. In basically all the stacked structures the dispersion energy is the leading stabilizing contribution, with the electrostatic term being always less stabilizing or even repulsive, as in the case of the P1 dimer at large intermolecular distances. Interestingly, for P1 the electrostatics switches from repulsion to attraction at short distances. This can be easily explained by a competition of the multipolar and overlap electrostatic components at short intermonomer distances. The multipolar part is repulsive here due to the very unfavorable dipole orientation. At short distances, however, the attractive overlap electrostatic interaction prevails. The DFT-SAPT electrostatics thus cannot be directly compared with the force field electrostatics. The latter takes into consideration only the electrostatic potentials (ESP), while the overlap electrostatic effects are effectively included in the vdw term of the force field. Therefore, the utilization of DFT-SAPT calculations in parametrizations of simple biomolecular force fields does not appear to be straightforward. It is to be noted that for biomolecular recognition the ESP part of the electrostatics is the most important term. Note that the overlap electrostatics is always

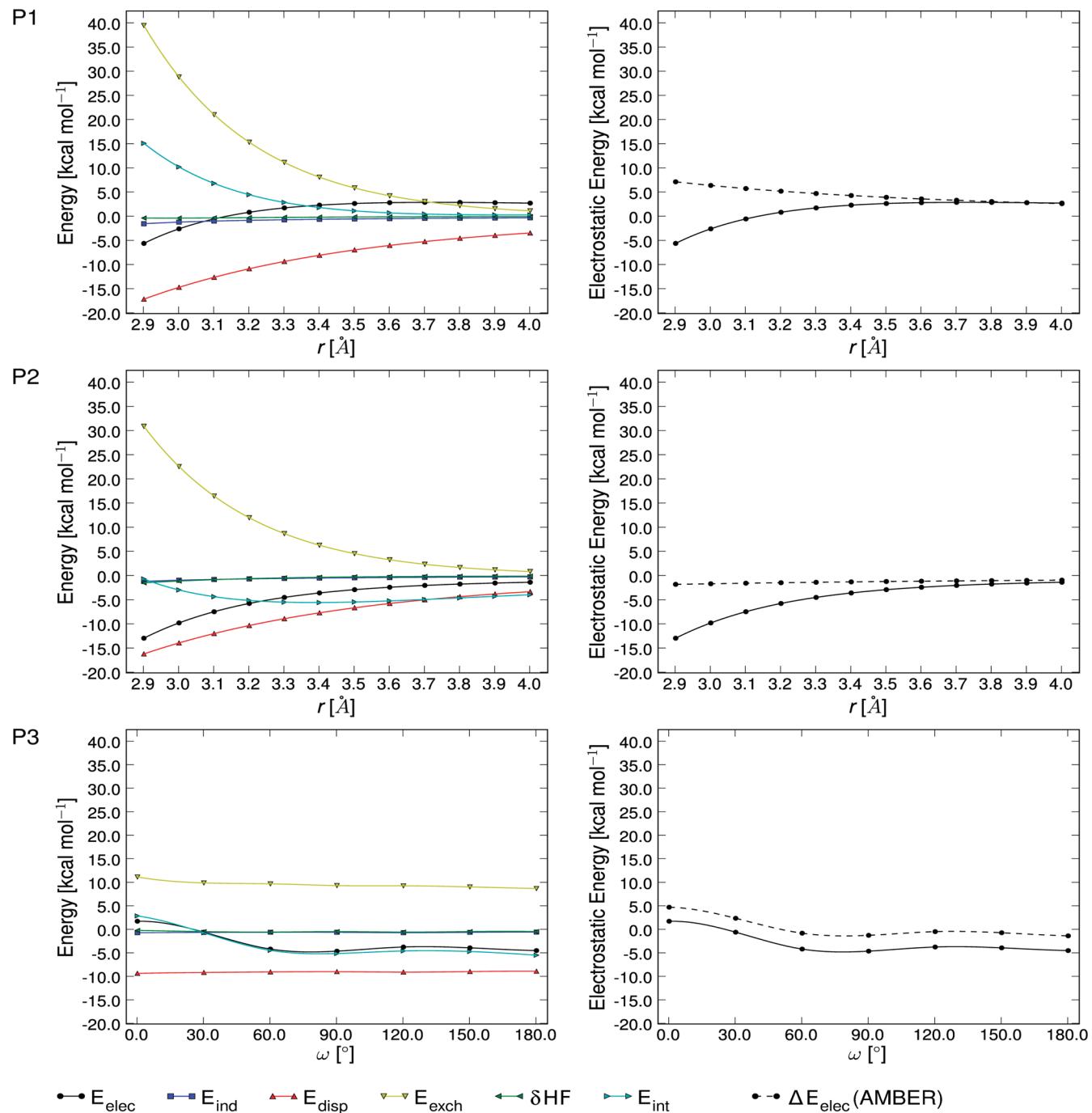


Figure 11. DFT-SAPT energy components (left) and comparison between DFT-SAPT and AMBER electrostatic energies (right) for P1, P2, and P3.

attractive, and the repulsion in the intermolecular complexes is due to Pauli exchange.

For P6, NP1, and NP2 the electrostatic interaction contribution is stabilizing and very significant, with a magnitude that is closer to that of the dispersion term, when compared to the rest of the dimers. For each dimer there is a region where the dispersion and exchange terms tend to cancel each other, usually around the minimum, which is characteristic for vdW complexes. When going into the repulsive part of the PECs the magnitude of the exchange term quickly outweighs that of the dispersion term. It thus should be noted that when assessing the DFT-SAPT decompositions, it is very important to have the dimers at proper

vertical separation. This is straightforward for geometries with parallel bases but may be more complex for geometries with tilted bases. DFT-SAPT decompositions for geometries with unrelaxed vertical separation or with unrealistic clash (e.g., due to errors in the experimental structure) can be biased. Taking the variability due to the intermolecular separation into account our results are consistent with those from previous DFT-SAPT studies.^{48,120,121}

For most of the stacked dimers the magnitude of the δHF term is comparable to that of the induction term, and both terms represent the smallest contributions to the interaction energy. The horizontal twist scan (P3) simply confirms that the electrostatic term exhibits the largest twist dependence

(followed by the exchange term to a lesser extent) and that the dispersion term is basically independent of the twist angle. Also, the dispersion and exchange terms basically cancel each other in the range of twist angles from 90 to 180°.

The DFT-SAPT data of the HB1 and HB2 structures basically confirm that the most relevant stabilizing contribution in the case of H-bonding is the electrostatic term and that the induction and δHF terms play a more important role than they do for the stacked dimers. Although smaller than the electrostatic term, the dispersion contribution is not negligible, as it amounts to about -8 and -6 kcal mol $^{-1}$ at the CBS(T) minimum of the HB1 and HB2 dimers, respectively.

Summary and Conclusions

We have carried out reference CBS(T) calculations for about 100 geometries of the uracil dimer in stacked conformations. The calculations have been specifically aimed at geometries with unoptimized distances between the monomers including geometries with mutually tilted monomers that are characterized by a delicate balance between local steric clashes and local unstacking. Until now such geometries had not been investigated using reference QM methods and were not included in the parametrization of methods such as DFT-D. Such geometries, however, often occur in nucleic acids and are of decisive importance for local conformational variations in B-DNA as well as in some cases for the exact positioning of bases in compactly folded RNAs. Errors in the short-range repulsion region would have a major impact on potential energy scans which were often used in the past to investigate local geometry variations in DNA.^{64–69,76} The incorrect description of such geometries may also partially affect MD simulations in applications when quantitative accuracy is required. For comparison, we have also carried out similar “compression-extension” calculations for two H-bonded systems: the WC and the Calcutta dimers.

The results show that methods like SCS(MI)-MP2 and DFT-SAPT yield a good description of the repulsion when steric clashes are involved in the stacking of uracil dimers. These methods are thus promising to accurately scan the potential energy surfaces of these systems. However, QM studies of DNA local conformational variations are still not straightforward because the electrostatic interactions in nucleic acids are greatly reduced by the solvent screening. Most likely, calculations completely neglecting the electrostatics (equivalent to using a molecular-mechanics force field without the electrostatic term) would give a more relevant description of local B-DNA conformational variations than gas-phase calculations with full inclusion of electrostatics.¹²² This is demonstrated for example in recent QM calculations of dependence of stacking on helical twist, where the optimal values of helical twist are typically outside the experimental range.¹²³ Both above mentioned methods also perform well for the two H-bonded dimers considered here, although for HB1 DFT-SAPT underestimates the strength of the interaction more than it does in the case of the stacked structures. Such underestimation is likely to be rooted in the relatively small basis set used for the DFT-SAPT calculations.

Given the good performance of the aforementioned methods it is natural to ask whether these methods can be further improved. For SCS(MI)-MP2 the use of the cc-pV(TQ)Z extrapolation (cc-pVTZ→cc-pVQZ) might bring a better agreement with the CBS(T) data, but there also seems to be some room for improvement on the formalism side, as it has been proposed in the work of Distasio and Head-Gordon.⁷⁷ Given the MP2 description of bond energies and intermolecular-interaction energies, it is necessary to distinguish between the computation of the two quantities, and these authors have suggested that a distance-dependent scaling could be applied to bridge these two regimes. It remains to be seen if such an approach could also result in an improved description of interaction energies of noncovalent systems in the more repulsive regions of the PES. It should be pointed out that the performance of DFT-SAPT for the systems studied here could be somewhat improved with the use of the nonlocalized PBE0AC/ALDA model as described in ref 124. The performance of DFT-SAPT would also improve with the use of better basis sets as in the present work we have used only the aug-cc-pVDZ one, which is inferior to the CBS extrapolation used for the MP2 method. In general, however, all tested QM methods have a quite reasonable performance and seem to be sufficiently accurate for the computation of intermolecular interactions in uracil dimers. Note that modest systematic shifts of the total energies in the range of ~ 1 kcal/mol are less important than an eventual imbalance in the description of steric clashes and unstacking.

The DFT-D method, at least the one we tested,⁵³ shows somewhat larger deviations in the repulsive region. The DFT-D literature is nowadays very wide, and we do not claim that our calculations represent the full spectrum of the methods.^{101,125–134} These deviations may have a modest impact on the computations, and it therefore would pay for to include some of the clashed geometries into future parametrizations of the DFT-D methods. However, we do not consider the differences as dramatic. Another approach could be a reparametrization of the empirical part of the method to improve the behavior mainly in the short-range repulsion region. However, this may require the introduction of new parameters, and success is not a priori guaranteed. The performance of DFT-D for the two tested H-bonded base pairs was neat.

We need to underline that although we expect that the present conclusions are basically valid also for other base stacking systems, it is not a priori guaranteed that the excellent performance observed here for several methods will fully transfer to all types of stacking interactions in nucleic-acid biopolymers. Thus, further calculations would be still useful.

Finally, the AMBER force field, which currently dominates molecular modeling of nucleic acids, shows large deviations in the repulsive region both for the stacked and for the H-bonded structures. The steric clashes are excessively severe, and the onset of the short-range repulsion upon compressing the interacting systems is too fast. This is expected to have a substantial effect on potential energy scans. On the other hand, we do not expect that this error would have dramatic qualitative impacts on explicit solvent simulations of nucleic acids. The reason is that in the simulations all the other degrees of freedom are also sampled, and these can be efficiently used to dissipate

the excessive clash with not much effect on the geometry. It has been noticed that an excessive repulsion associated with too large hydrogens of the thymine methyl group affects substantially 2D stacking energy scans of propeller twisting in ApT and ApA B-DNA steps. However, the reduction of the hydrogen radius, which markedly improved the quality of the empirical force-field stacking energy scans, had no visible effect in test simulations.¹³⁵ An exaggerated short-range repulsion in the force field description was also reported for cation-solute interactions and affects for example simulations of quadruplex DNA where the cations residing in the channel of the quadruplex stem look oversized. This effect was originally attributed to the lack of polarization in the force field,^{136,137} but it is more likely caused by the short-range repulsion imbalance discussed in the present paper. Also, when using MP2/aug-cc-pVDZ ESP charges the AMBER force field seems to provide a better gas-phase description for stacking than for H-bonding, at least for the uracil dimer. Despite the reported difference we would like to clearly state that the overall description of base stacking and H-bonding by the nonbonded potential of the AMBER force field is good, and, as we pointed out elsewhere, stacking is probably the best approximated term in the force field.^{138,139} The discrepancy reported here should not affect the overall stability of the simulations and qualitative applications of the method. However, it may affect the description of very subtle quantitative effects such as the local conformational variations in B-DNA.

Acknowledgment. This contribution was supported by the Grant Agency of the Academy of Sciences of the Czech Republic, grant nos. IAA400040802 and IAA400550701; the Ministry of Education of the Czech Republic, grant nos. LC06030, LC512, and MSM6198959216 (Petr Jurečka); the Grant Agency of the Czech Republic, 203/09/1476; and by the Academy of Sciences of the Czech Republic, grant nos. AV0Z50040507, AV0Z50040702, and Z40550506. A portion of the research described in this paper was carried out on the high-performance computer system of the Environmental Molecular Sciences Laboratory (EMSL), a scientific user facility at the Pacific Northwest National Laboratory. Pavel Hobza acknowledges the support of Praemium Academiae, Academy of Sciences of the Czech Republic.

Supporting Information Available: Studied geometries (xyz coordinates), tables of interaction energies, plots of energy gradients, and plots of DFT-SAPT energy contributions. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Ng, H.; Kopka, M. L.; Dickerson, R. E. The Structure of a Stable Intermediate in the A↔B DNA Helix Transition. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 2035–2039.
- Mathews, D. H.; Turner, D. H. Prediction of RNA Secondary Structure by Free Energy Minimization. *Curr. Opin. Struct. Biol.* **2006**, *16*, 270–278.
- Petersheim, M.; Turner, D. H. Base-Stacking and Base-Pairing Contributions to Helix Stability: Thermodynamics of Double-Helix Formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp. *Biochemistry* **1983**, *22*, 256–263.
- Guckian, K. M.; Schweitzer, B. A.; Ren, R. X.; Sheils, C. J.; Tahmassebi, D. C.; Kool, E. T. Factors Contributing to Aromatic Stacking in Water: Evaluation in the Context of DNA. *J. Am. Chem. Soc.* **2000**, *122*, 2213–2222.
- Suzuki, M.; Amano, N.; Kakinuma, J.; Tateno, M. Use of a 3D Structure Data Base for Understanding Sequence-Dependent Conformational Aspects of DNA. *J. Mol. Biol.* **1997**, *274*, 421–435.
- Dickerson, R. E.; Drew, H. R. Structure of a B-DNA Dodecamer: II. Influence of Base Sequence on Helix Structure. *J. Mol. Biol.* **1981**, *149*, 761–786.
- Alexandrov, B. S.; Gelev, V.; Monisova, Y.; Alexandrov, L. B.; Bishop, A. R.; Rasmussen, K. Ø.; Usheva, A. A Nonlinear Dynamic Model of DNA with a Sequence-dependent Stacking Term. *Nucleic Acids Res.* **2009**, advance access. (Doi: 10.1093/nar/gkp016).
- Copeland, K. L.; Anderson, J. A.; Farley, A. R.; Cox, J. R.; Tschumper, G. S. Probing Phenylalanine/Adenine p-Stacking Interactions in Protein Complexes with Explicitly Correlated and CCSD(T) Computations. *J. Phys. Chem. B* **2008**, *112*, 14291–14295.
- Rutledge, L. R.; Wetmore, S. D. Remarkably Strong T-Shaped Interactions Between Aromatic Amino Acids and Adenine: Their Increase Upon Nucleobase Methylation and a Comparison to Stacking. *J. Chem. Theory Comput.* **2008**, *4*, 1768–1780.
- Cysewski, P. The Post-SCF Quantum Chemistry Characteristics of the Energetic Heterogeneity of Stacked Guanine-Guanine Pairs Found in B-DNA and A-DNA Crystals. *J. Mol. Struct. (THEOCHEM)* **2008**, *865*, 36–43.
- Lait, L. A.; Rutledge, L. R.; Millen, A. L.; Wetmore, S. D. yDNA Versus xDNA Pyrimidine Nucleobases: Computational Evidence for Dependence of Duplex Stability on Spacer Location. *J. Phys. Chem. B* **2008**, *112*, 12526–12536.
- Cooper, V. R.; Thonhauser, T.; Langreth, D. C. An Application of the Van Der Waals Density Functional: Hydrogen Bonding and Stacking Interactions Between Nucleobases. *J. Chem. Phys.* **2008**, *128*, 204102–4.
- Hill, J. G.; Platts, J. A. Calculating Stacking Interactions in Nucleic Acid Base-Pair Steps Using Spin-Component Scaling and Local Second Order Moller-Plesset Perturbation Theory. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2785–2791.
- Rutledge, L. R.; Durst, H. F.; Wetmore, S. D. Computational Comparison of the Stacking Interactions Between the Aromatic Amino Acids and the Natural or (Cationic) Methylated Nucleobases. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2801–2812.
- Langner, K. M.; Sokalski, W. A.; Leszczynski, J. Intriguing Relations of Interaction Energy Components in Stacked Nucleic Acids. *J. Chem. Phys.* **2007**, *127*, 111102–4.
- Vanommeslaeghe, K.; Mignon, P.; Loverix, S.; Tourwe, D.; Geerlings, P. Influence of Stacking on the Hydrogen Bond Donating Potential of Nucleic Bases. *J. Chem. Theory Comput.* **2006**, *2*, 1444–1452.
- Waller, M. P.; Robertazzi, A.; Platts, J. A.; Hibbs, D. E.; Williams, P. A. Hybrid Density Functional Theory for p-Stacking Interactions: Application to Benzenes, Pyridines, and DNA Bases. *J. Comput. Chem.* **2006**, *27*, 491–504.
- Cysewski, P.; Czyznikowska-Balcerak, Z. The MP2 Quantum Chemistry Study on the Local Minima of Guanine Stacked with All Four Nucleic Acid Bases in Conformations Cor-

- responding to Mean B-DNA. *J. Mol. Struct. (THEOCHEM)* **2005**, *757*, 29–36.
- (19) Sponer, J.; Riley, K. E.; Hobza, P. Nature and Magnitude of Aromatic Stacking of Nucleic Acid Bases. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595–2610.
 - (20) Sponer, J.; Leszczynski, J.; Hobza, P. Electronic Properties, Hydrogen Bonding, Stacking, and Cation Binding of DNA and RNA Bases. *Biopolymers* **2001**, *61*, 3–31.
 - (21) Sponer, J.; Jurecka, P.; Hobza, P. Base Stacking and Base Pairing. In *Computational studies of RNA and DNA*; Sponer, J., Lankas, F., Eds.; Springer: Dordrecht, 2006; Chapter 14, pp 343–388.
 - (22) Moller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.
 - (23) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. Quadratic Configuration Interaction. A General Technique for Determining Electron Correlation Energies. *J. Chem. Phys.* **1987**, *87*, 5968–5975.
 - (24) Paldus, J.; Shavitt, I.; Cizek, J. Correlation Problems in Atomic and Molecular Systems. IV. Extended Coupled-Pair Many-Electron Theory and Its Application To the BH₃ Molecule. *Phys. Rev. A* **1972**, *5*, 50–67.
 - (25) DeJong, E. S.; Marzluff, W. F.; Nikonorowicz, E. P. NMR Structure and Dynamics of the RNA-Binding Site for the Histone mRNA Stem-Loop Binding Protein. *RNA* **2002**, *8*, 83–96.
 - (26) Theimer, C. A.; Finger, L. D.; Trantirek, L.; Feigon, J. Mutations Linked to Dyskeratosis Congenita Cause Changes in the Structural Equilibrium in Telomerase RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 449–454.
 - (27) Kolk, M. H.; Graaf, M. V. D.; Wijmenga, S. S.; Pleij, C. W. A.; Heus, H. A.; Hilbers, C. W. NMR Structure of a Classical Pseudoknot: Interplay of Single- and Double-Stranded RNA. *Science* **1998**, *280*, 434–438.
 - (28) Theimer, C. A.; Blois, C. A.; Feigon, J. Structure of the Human Telomerase RNA Pseudoknot Reveals Conserved Tertiary Interactions Essential for Function. *Mol. Cell* **2005**, *17*, 671–682.
 - (29) Liu, H.; Matsugami, A.; Katahira, M.; Uesugi, S. A Dimeric RNA Quadruplex Architecture Comprised of Two G:G(A):G(G(A) Hexads, G:G:G Tetrad and UUUU Loops. *J. Mol. Biol.* **2002**, *322*, 955–970.
 - (30) Oberstrass, F. C.; Lee, A.; Stefl, R.; Janis, M.; Chanfreau, G.; Allain, F. H. Shape-Specific Recognition in the Structure of the Vts1p SAM Domain with RNA. *Nat. Struct. Mol. Biol.* **2006**, *13*, 160–167.
 - (31) Sashital, D. G.; Venditti, V.; Angers, C. G.; Cornilescu, G.; Butcher, S. E. Structure and Thermodynamics of a Conserved U2 snRNA Domain from Yeast and Human. *RNA* **2007**, *13*, 328–338.
 - (32) Larson, S. B.; Day, J.; Greenwood, A.; McPherson, A. Refined Structure of Satellite Tobacco Mosaic Virus at 1.8 Å Resolution. *J. Mol. Biol.* **1998**, *277*, 37–59.
 - (33) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science* **2000**, *289*, 905–20.
 - (34) Selmer, M.; Dunham, C. M.; Murphy, F. V.; Weixlbaumer, A.; Petry, S.; Kelley, A. C.; Weir, J. R.; Ramakrishnan, V. Structure of the 70S Ribosome Complexed with mRNA and tRNA. *Science* **2006**, *313*, 1935–42.
 - (35) Sponer, J.; Leszczynski, J.; Hobza, P. Nature of Nucleic Acid-Base Stacking: Nonempirical *Ab Initio* and Empirical Potential Characterization of 10 Stacked Base Dimers. Comparison of Stacked and H-Bonded Base Pairs. *J. Phys. Chem.* **1996**, *100*, 5590–5596.
 - (36) Hehre, W. J.; Radom, L.; Schleyer, P. V. R.; Pople, J. A. In *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986; pp 1–548.
 - (37) Hobza, P.; Sponer, J.; Polasek, M. H-Bonded and Stacked DNA Base Pairs: Cytosine Dimer. An *Ab Initio* Second-Order Moeller-Plesset Study. *J. Am. Chem. Soc.* **1995**, *117*, 792–798.
 - (38) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
 - (39) Kratochvil, M.; Engkvist, O.; Sponer, J.; Jungwirth, P.; Hobza, P. Uracil Dimer: Potential Energy and Free Energy Surfaces. *Ab Initio* Beyond Hartree-Fock and Empirical Potential Studies. *J. Phys. Chem. A* **1998**, *102*, 6921–6926.
 - (40) Kratochvil, M.; Engkvist, O.; Vacek, J.; Jungwirth, P.; Hobza, P. Methylated Uracil Dimers: Potential Energy and Free Energy Surfaces. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2419–2424.
 - (41) Hobza, P.; Sponer, J. Significant Structural Deformation of Nucleic Acid Bases in Stacked Base Pairs: An *Ab Initio* Study Beyond Hartree-Fock. *Chem. Phys. Lett.* **1998**, *288*, 7–14.
 - (42) Florian, J.; Sponer, J.; Warshel, A. Thermodynamic Parameters for Stacking and Hydrogen Bonding of Nucleic Acid Bases in Aqueous Solution: *Ab Initio/Langevin Dipoles* Study. *J. Phys. Chem. B* **1999**, *103*, 884–892.
 - (43) Leininger, M. L.; Nielsen, I. M. B.; Colvin, M. E.; Janssen, C. L. Accurate Structures and Binding Energies for Stacked Uracil Dimers. *J. Phys. Chem. A* **2002**, *106*, 3850–3854.
 - (44) Hobza, P.; Sponer, J. Toward True DNA Base-Stacking Energies: MP2, CCSD(T), and Complete Basis Set Calculations. *J. Am. Chem. Soc.* **2002**, *124*, 11802–11808.
 - (45) Czyznikowska, Z.; Zalesny, R.; Ziolkowski, M.; Gora, R. W.; Cysewski, P. The Nature of Interactions in Uracil Dimer: An *Ab Initio* Study. *Chem. Phys. Lett.* **2007**, *450*, 132–137.
 - (46) Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
 - (47) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
 - (48) Cybulski, H.; Sadlej, J. Symmetry-Adapted Perturbation-Theory Interaction-Energy Decomposition for Hydrogen-Bonded and Stacking Structures. *J. Chem. Theory Comput.* **2008**, *4*, 892–897.
 - (49) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of Van Der Waals Complexes. *Chem. Rev.* **1994**, *94*, 1887–1930.
 - (50) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. Intermolecular Potentials Based on Symmetry-Adapted Perturbation Theory with Dispersion Energies from Time-

- Dependent Density-Functional Calculations. *J. Chem. Phys.* **2005**, *123*, 214103–14.
- (51) Pitonak, M.; Riley, K. E.; Neogrady, P.; Hobza, P. Highly Accurate CCSD(T) and DFT-SAPT Stabilization Energies of H-Bonded and Stacked Structures of the Uracil Dimer. *ChemPhysChem* **2008**, *9*, 1636–1644.
 - (52) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. Benchmark Database of Accurate (MP2 and CCSD(T) Complete Basis Set Limit) Interaction Energies of Small Model Complexes, DNA Base Pairs, and Amino Acid Pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.
 - (53) Jurecka, P.; Cerny, J.; Hobza, P.; Salahub, D. R. Density Functional Theory Augmented with an Empirical Dispersion Term. Interaction Energies and Geometries of 80 Noncovalent Complexes Compared with *Ab Initio* Quantum Mechanics Calculations. *J. Comput. Chem.* **2007**, *28*, 555–569.
 - (54) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. Experimental Studies of Molecular Interactions Between Nitrogen Bases of Nucleic Acids. *Biopolymers* **1979**, *18*, 1149–1170.
 - (55) Jurecka, P.; Hobza, P. True Stabilization Energies for the Optimal Planar Hydrogen-Bonded and Stacked Structures of Guanine. .Cytosine, Adenine. .Thymine, and Their 9- and 1-Methyl Derivatives: Complete Basis Set Calculations at the MP2 and CCSD(T) Levels and Comparison with Experiment. *J. Am. Chem. Soc.* **2003**, *125*, 15608–15613.
 - (56) Casaeas, R. N.; Paul, J. B.; McLaughlin, R. P.; Saykally, R. J.; van Mourik, T. Infrared Cavity Ringdown Spectroscopy of Jet-Cooled Nucleotide Base Clusters and Water Complexes. *J. Phys. Chem. A* **2004**, *108*, 10989–10996.
 - (57) Ts'o, P. O. P.; Melvin, I. S.; Olson, A. C. Interaction and Association of Bases and Nucleosides in Aqueous Solutions. *J. Am. Chem. Soc.* **1963**, *85*, 1289–1296.
 - (58) Calladine, C. R. Mechanics of Sequence-Dependent Stacking of Bases in B-DNA. *J. Mol. Biol.* **1982**, *161*, 343–352.
 - (59) Calladine, C. R.; Drew, H. R. A Base-Centred Explanation of the B-to-A Transition in DNA. *J. Mol. Biol.* **1984**, *178*, 773–782.
 - (60) Yanagi, K.; Privé, G. G.; Dickerson, R. E. Analysis of Local Helix Geometry in Three B-DNA Decamers and Eight Dodecamers. *J. Mol. Biol.* **1991**, *217*, 201–214.
 - (61) Dickerson, R. E.; Goodsell, D. S.; Neidle, S. “.the tyranny of the lattice. .”. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 3579–3583.
 - (62) Ferre-D'Amare, A. R.; Zhou, K. H.; Doudna, J. A. Crystal Structure of a Hepatitis Delta Virus Ribozyme. *Nature (London)* **1998**, *395*, 567–574.
 - (63) Krasovska, M. V.; Sefcikova, J.; Reblova, K.; Schneider, B.; Walter, N. G.; Sponer, J. Cations and Hydration in Catalytic RNA: Molecular Dynamics of the Hepatitis Delta Virus Ribozyme. *Biophys. J.* **2006**, *91*, 626–638.
 - (64) Bhattacharyya, D.; Bansal, M. Local Variability and Base Sequence Effects in DNA Crystal Structures. *J. Biomol. Struct. Dyn.* **1990**, *8*, 539–572.
 - (65) Bhattacharyya, D.; Bansal, M. Groove Width and Depth of B-DNA Structures Depend on Local Variation in Slide. *J. Biomol. Struct. Dyn.* **1992**, *10*, 213–226.
 - (66) Sarai, A.; Mazur, J.; Nussinov, R.; Jernigan, R. L. Origin of DNA Helical Structure and Its Sequence Dependence. *Biochemistry* **1988**, *27*, 8498–8502.
 - (67) Srinivasan, A. R.; Torres, R.; Clark, W.; Olson, W. K. Base Sequence Effects in Double Helical DNA. I. Potential Energy Estimates of Local Base Morphology. *J. Biomol. Struct. Dyn.* **1987**, *5*, 459–96.
 - (68) Sponer, J.; Kypr, J. Different Intrastrand and Interstrand Contributions to Stacking Account for Roll Variations at the Alternating Purine-Pyrimidine Sequences in A-DNA and A-RNA. *J. Mol. Biol.* **1991**, *221*, 761–764.
 - (69) Sponer, J.; Kypr, J. Relationships Among Rise, Cup, Roll and Stagger in DNA Suggested by Empirical Potential Studies of Base Stacking. *J. Biomol. Struct. Dyn.* **1993**, *11*, 27–41.
 - (70) Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X.; Neidle, S.; Shakkeb, Z.; Sklenar, H.; Suzuki, M.; Tung, C.; Westhof, E.; Wolberger, C.; Berman, H. M. A Standard Reference Frame for the Description of Nucleic Acid Base-Pair Geometry. *J. Mol. Biol.* **2001**, *313*, 229–237.
 - (71) Reblova, K.; Lankas, F.; Razga, F.; Krasovska, M. V.; Koca, J.; Sponer, J. Structure, Dynamics and Elasticity of free 16S rRNA Helix 44 Studied by Molecular Dynamics Simulations. *Biopolymers* **2006**, *82*, 504–520.
 - (72) Lankas, F.; Sponer, J.; Langowski, J.; Cheatham, T. E. DNA Basepair Step Deformability Inferred from Molecular Dynamics Simulations. *Biophys. J.* **2003**, *85*, 2872–2883.
 - (73) Dock-Bregeon, A. C.; Chevrier, B.; Podjarny, A.; Johnson, J.; de Bear, J. S.; Gough, G. R.; Gilham, P. T.; Moras, D. Crystallographic Structure of an RNA Helix: [U(UA)₆A]₂. *J. Mol. Biol.* **1989**, *209*, 459–474.
 - (74) Wahl, M. C.; Sundaralingam, M. Crystal Structures of A-DNA Duplexes. *Biopolymers* **1997**, *44*, 45–63.
 - (75) Sponer, J.; Jurecka, P.; Marchan, I.; Luque, F. J.; Orozco, M.; Hobza, P. Nature of Base Stacking: Reference Quantum-Chemical Stacking Energies in Ten Unique B-DNA Base-Pair Steps. *Chem.--Eur. J.* **2006**, *12*, 2854–2865.
 - (76) Sponer, J.; Kypr, J. Theoretical Analysis of the Base Stacking in DNA: Choice of the Force Field and a Comparison with the Oligonucleotide Crystal Structures. *J. Biomol. Struct. Dyn.* **1993**, *11*, 277–292.
 - (77) Distasio, R. A.; Head-Gordon, M. Optimized Spin-Component Scaled Second-Order Møller-Plesset Perturbation Theory for Intermolecular Interaction Energies. *Mol. Phys.* **2007**, *105*, 1073–1083.
 - (78) Hesselmann, A.; Jansen, G. The Helium Dimer Potential from a Combined Density Functional Theory and Symmetry-Adapted Perturbation Theory Approach Using an Exact Exchange-Correlation Potential. *Phys. Chem. Chem. Phys.* **2003**, *5*, 5010–5014.
 - (79) Hohenstein, E. G.; Chill, S. T.; Sherrill, C. D. Assessment of the Performance of the M05–2X and M06–2X Exchange-Correlation Functionals for Noncovalent Interactions in Biomolecules. *J. Chem. Theory Comput.* **2008**, *4*, 1996–2000.
 - (80) Bachorz, R. A.; Bischoff, F. A.; Höfener, S.; Klopper, W.; Ottiger, P.; Leist, R.; Frey, J. A.; Leutwyler, S. Scope and Limitations of the SCS-MP2 Method for Stacking and Hydrogen Bonding Interactions. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2758–2766.
 - (81) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. Use of Approximate Integrals in *Ab Initio* Theory. An Application in MP2 Energy Calculations. *Chem. Phys. Lett.* **1993**, *208*, 359–363.

- (82) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. Integral Approximations for LCAO-SCF Calculations. *Chem. Phys. Lett.* **1993**, *213*, 514–518.
- (83) Bernholdt, D. E.; Harrison, R. J. Large-Scale Correlated Electronic Structure Calculations: The RI-MP2 Method on Parallel Computers. *Chem. Phys. Lett.* **1996**, *250*, 477–484.
- (84) Sponer, J.; Jurecka, P.; Hobza, P. Accurate Interaction Energies of Hydrogen-bonded Nucleic Acid Base Pairs. *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151.
- (85) Van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. State of the Art in Counterpoise Theory. *Chem. Rev.* **1994**, *94*, 1873–1885.
- (86) Szalewicz, K.; Jeziorski, B. Comment on “On the Importance of the Fragment Relaxation Energy Terms in the Estimation of the Basis Set Superposition Error Correction to the Intermolecular Interaction Energy” [J. Chem. Phys. 104, 8821 (1996)]. *J. Chem. Phys.* **1998**, *109*, 1198–1200.
- (87) Jansen, H. B.; Ros, P. Non-Empirical Molecular Orbital Calculations on the Protonation of Carbon Monoxide. *Chem. Phys. Lett.* **1969**, *3*, 140–143.
- (88) Boys, S. F.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553–566.
- (89) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129–145.
- (90) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431–439.
- (91) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.02*; Gaussian, Inc.: Wallingford, CT, 2004.
- (92) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (93) Reha, D.; Kabelac, M.; Ryjacek, F.; Sponer, J.; Sponer, J. E.; Elstner, M.; Suhai, S.; Hobza, P. Intercalators. 1. Nature of Stacking Interactions Between Intercalators (Ethidium, Daunomycin, Ellipticine, and 4', 6-Diaminide-2-phenylindole) and DNA Base Pairs. *Ab initio* Quantum Chemical, Density Functional Theory, and Empirical Potential Study. *J. Am. Chem. Soc.* **2002**, *124*, 3366–3376.
- (94) Jurecka, P.; Sponer, J.; Hobza, P. Potential Energy Surface of the Cytosine Dimer: MP2 Complete Basis Set Limit Interaction Energies, CCSD(T) Correction Term and a Comparison with the AMBER Force Field. *J. Phys. Chem. B* **2004**, *108*, 5466–5471.
- (95) Sponer, J.; Leszczynski, J.; Hobza, P. Base Stacking in Cytosine Dimer. A Comparison of Correlated *Ab Initio* Calculations with Three Empirical Potential Models and Density Functional Theory Calculations. *J. Comput. Chem.* **1996**, *17*, 841–850.
- (96) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. Electronic Structure Calculations on Workstation Computers: The Program System Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165–169.
- (97) Eichkorn, K.; Treutler, O.; Ohm, H.; Haser, M.; Ahlrichs, R. Auxiliary Basis Sets to Approximate Coulomb Potentials. *Chem. Phys. Lett.* **1995**, *240*, 283–289.
- (98) Baerends, E. J.; Ellis, D. E.; Ros, P. Self-Consistent Molecular Hartree–Fock–Slater Calculations I. The Computational Procedure. *Chem. Phys.* **1973**, *2*, 41–51.
- (99) Whitten, J. L. Coulombic Potential Energy Integrals and Approximations. *J. Chem. Phys.* **1973**, *58*, 4496–4501.
- (100) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On Some Approximations in Applications of Xa Theory. *J. Chem. Phys.* **1979**, *71*, 3396–3402.
- (101) Grimme, S. Accurate Description of Van Der Waals Complexes by Density Functional Theory Including Empirical Corrections. *J. Comput. Chem.* **2004**, *25*, 1463–1473.
- (102) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401–1146401–4.
- (103) Werner, H.-J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson T. *MOLPRO, Version 2006.1, A Package of Ab Initio Programs*. See <http://www.molpro.net> (accessed month year).
- (104) Polly, R.; Werner, H.; Manby, F. R.; Knowles, P. J. Fast Hartree-Fock Theory Using Local Density Fitting Approximations. *Mol. Phys.* **2004**, *102*, 2311–2321.
- (105) Werner, H.; Manby, F. R.; Knowles, P. J. Fast Linear Scaling Second-Order Moller-Plesset Perturbation Theory (MP2) Using Local and Density Fitting Approximations. *J. Chem. Phys.* **2003**, *118*, 8149–8160.
- (106) Grimme, S. Improved Second-Order Moller-Plesset Perturbation Theory by Separate Scaling of Parallel- and Antiparallel-Spin Pair Correlation Energies. *J. Chem. Phys.* **2003**, *118*, 9095–9102.
- (107) Hesselmann, A.; Jansen, G. First-Order Intermolecular Interaction Energies from Kohn-Sham Orbitals. *Chem. Phys. Lett.* **2002**, *357*, 464–470.
- (108) Hesselmann, A.; Jansen, G. Intermolecular Induction and Exchange-Induction Energies from Coupled-Perturbed Kohn-Sham Density Functional Theory. *Chem. Phys. Lett.* **2002**, *362*, 319–325.

- (109) Hesselmann, A.; Jansen, G. Intermolecular Dispersion Energies from Time-Dependent Density Functional Theory. *Chem. Phys. Lett.* **2003**, *367*, 778–784.
- (110) Hesselmann, A.; Jansen, G.; Schutz, M. Density-Functional Theory-Symmetry-Adapted Intermolecular Perturbation Theory with Density Fitting: A New Efficient Method to Study Intermolecular Interaction Energies. *J. Chem. Phys.* **2005**, *122*, 014103–17.
- (111) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- (112) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-Set Convergence in Correlated Calculations on Ne, N₂, and H₂O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.
- (113) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Olsen, J. Basis-Set Convergence of the Energy in Molecular Hartree-Fock Calculations. *Chem. Phys. Lett.* **1999**, *302*, 437–446.
- (114) Jurecka, P.; Hobza, P. On the Convergence of the (DECCS-D(T)-DEMP2) Term for Complexes with Multiple H-Bonds. *Chem. Phys. Lett.* **2002**, *365*, 89–94.
- (115) Dabkowska, I.; Jurecka, P.; Hobza, P. On Geometries of Stacked and H-Bonded Nucleic Acid Base Pairs Determined at Various DFT, MP2, and CCSD(T) Levels up to the CCSD(T)/Complete Basis Set Limit Level. *J. Chem. Phys.* **2005**, *122*, 204322–9.
- (116) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comp. Sci. Eng.* **2007**, *9*, 90–95.
- (117) Jones, E.; Oliphant, T.; Peterson, P. and others. SciPy: Open Source Scientific Tools for Python ; 2001. <http://www.scipy.org> (accessed month year).
- (118) Oliphant, T. E. Python for Scientific Computing. *Comp. Sci. Eng.* **2007**, *9*, 10–20.
- (119) Sponer, J.; Hobza, P. Interaction Energies of Hydrogen Bonded Formamide Dimer, Formamidine Dimer, and Selected DNA Base Pairs Obtained with Large Basis Sets of Atomic Orbitals. *J. Phys. Chem. A* **2000**, *104*, 4592–4597.
- (120) Hesselmann, A.; Fitzgerald, G.; Schutz, M. Interaction Energy Contributions of H-Bonded and Stacked Structures of the AT and GC DNA Base Pairs from the Combined Density Functional Theory and Intermolecular Perturbation Theory Approach. *J. Am. Chem. Soc.* **2006**, *128*, 11730–11731.
- (121) Sedlak, R.; Jurecka, P.; Hobza, P. Density Functional Theory-Symmetry Adapted Perturbation Treatment Energy Decomposition of Nucleic Acid Base Pairs Taken from DNA Crystal Geometry. *J. Chem. Phys.* **2007**, *127*, 075104.
- (122) Sponer, J.; Florian, J.; Ng, H.; Sponer, J. E.; Spackova, N. Local Conformational Variations Observed in B-DNA Crystals Do Not Improve Base Stacking: Computational Analysis of Base Stacking in a d(CATGGGCCATG)2 B↔A Intermediate Crystal Structure. *Nucleic Acids Res.* **2000**, *28*, 4893–4902.
- (123) Cooper, V. R.; Thonhauser, T.; Puzder, A.; Schroder, E.; Lundqvist, B. I.; Langreth, D. C. Stacking Interactions and the Twist of DNA. *J. Am. Chem. Soc.* **2008**, *130*, 1304–1308.
- (124) Tekin, A.; Jansen, G. How Accurate Is the Density Functional Theory Combined with Symmetry-Adapted Perturbation Theory Approach for CH-π and π-π Interactions? A Comparison to Supermolecular Calculations for the Acetylene-Benzene Dimer. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1680–1687.
- (125) Ducere, J.; Cavallo, L. Parametrization of an Empirical Correction Term to Density Functional Theory for an Accurate Description of p-Stacking Interactions in Nucleic Acids. *J. Phys. Chem. B* **2007**, *111*, 13124–13134.
- (126) Goursot, A.; Mineva, T.; Kevorkyants, R.; Talbis, D. Interaction Between N-Alkane Chains: Applicability of the Empirically Corrected Density Functional Theory for Van Der Waals Complexes. *J. Chem. Theory Comput.* **2007**, *3*, 755–763.
- (127) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
- (128) Zimmerli, U.; Parrinello, M.; Koumoutsakos, P. Dispersion Corrections to Density Functionals for Water Aromatic Interactions. *J. Chem. Phys.* **2004**, *120*, 2693–2699.
- (129) Wu, Q.; Yang, W. Empirical Correction to Density Functional Theory for Van Der Waals Interactions. *J. Chem. Phys.* **2002**, *116*, 515–524.
- (130) Wu, X.; Vargas, M. C.; Nayak, S.; Lotrich, V.; Scoles, G. Towards Extending the Applicability of Density Functional Theory to Weakly Bound Systems. *J. Chem. Phys.* **2001**, *115*, 8748–8757.
- (131) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. Hydrogen Bonding and Stacking Interactions of Nucleic Acid Base Pairs: A Density-Functional-Theory Based Treatment. *J. Chem. Phys.* **2001**, *114*, 5149–5155.
- (132) Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Eijck, B. P. Transferable *Ab Initio* Intermolecular Potentials. 1. Derivation from Methanol Dimer and Trimer Calculations. *J. Phys. Chem. A* **1999**, *103*, 9872–9882.
- (133) Meijer, E. J.; Sprink, M. A Density-Functional Study of the Intermolecular Interactions of Benzene. *J. Chem. Phys.* **1996**, *105*, 8684–8689.
- (134) Zhechkov, L.; Heine, T.; Patchkovskii, S.; Seifert, G.; Duartes, H. A. An Efficient a Posteriori Treatment for Dispersion Interaction in Density-Functional-Based Tight Binding. *J. Chem. Theory Comput.* **2005**, *1*, 841–847.
- (135) Warmlander, S.; Sponer, J. E.; Sponer, J.; Leijon, M. The Influence of the Thymine C5 Methyl Group on Spontaneous Base Pair Breathing in DNA. *J. Biol. Chem.* **2002**, *277*, 28491–28497.
- (136) Spackova, N.; Berger, I.; Sponer, J. Nanosecond Molecular Dynamics Simulations of Parallel and Antiparallel Guanine Quadruplex DNA Molecules. *J. Am. Chem. Soc.* **1999**, *121*, 5519–5534.
- (137) Fadrna, E.; Spackova, N.; Stefl, R.; Koca, J.; Cheatham, T. E.; Sponer, J. Molecular Dynamics Simulations of Guanine Quadruplex Loops: Advances and Force Field Limitations. *Biophys. J.* **2004**, *87*, 227–242.
- (138) Sponer, J.; Spackova, N. Molecular Dynamics Simulations and Their Application to Four-Stranded DNA. *Methods* **2007**, *43*, 278–290.
- (139) McDowell, S. E.; Spackova, N.; Sponer, J.; Walter, N. G. Molecular Dynamics Simulations of RNA: An *In Silico* Single Molecule Approach. *Biopolymers* **2007**, *85*, 169–184.