CHEMISTRY
A EUROPEAN JOURNAL

DOI: 10.1002/chem.200501239

Nature of Base Stacking: Reference Quantum-Chemical Stacking Energies in Ten Unique B-DNA Base-Pair Steps

Jiří Šponer,*[a] Petr Jurečka,[a, b] Ivan Marchan,[c, d] F. Javier Luque,[e] Modesto Orozco,[c, d] and Pavel Hobza[b]

Abstract: Base-stacking energies in ten unique B-DNA base-pair steps and some other arrangements were evaluated by the second-order Møller–Plesset (MP2) method, complete basis set (CBS) extrapolation, and correction for triple (T) electron-correlation contributions. The CBS(T) calculations were compared with decade-old MP2/6-31G*(0.25) reference data and AMBER force field. The new calculations show modest increases in stacking stabilization compared to the MP2/6-31G*(0.25) data and surprisingly large sequence-dependent variation of stacking energies. The absolute force-field values are in better agreement with the new reference data, while relative discrepancies between quantum-chemical (QM) and force-field values increase modestly. Nevertheless, the force field provides good qualitative description of stacking, and there is no need to introduce additional pair-additive electrostatic terms, such as distributed multipoles or out-of-plane charges. There is a rather surprising difference of about 0.1 Å between the vertical separation of base pairs predicted by quantum chemistry and derived from crystal structures. Evaluations of different local arrangements of the 5′-CG-3′ step indicate a sensitivity of the relative stacking energies to the level of calculation. Thus, describing quantitative relations between local DNA geometrical variations and stacking may be more complicated than usually assumed. The reference calculations are complemented by continuum-solvent assessment of solvent-screening effects.

Keywords: ab initio calculations · DNA structures · nucleobases · quantum mechanics · solvent effects

[a] Prof. J. Šponer, Dr. P. Jurečka
Intitute of Biophysics
Academy of Sciences of the Czech Republic
Královopolská 135, 612 65 Brno (Czech Republic)
Fax: (+420) 5412-12179
E-mail: sponer@ncbr.chemi.muni.cz

[b] Dr. P. Jurečka, Prof. P. Hobza
Institute of Organic Chemistry and Biochemistry
Academy of Sciences of the Czech Republic
Flemingovo nám. 2, 166 10 Prague (Czech Republic)

[c] I. Marchan, Prof. M. Orozco
Institut de Recerca Biomèdica
Parc Científic de Barcelona, Josep Samitier 1-5
08028 Barcelona (Spain)

[d] I. Marchan, Prof. M. Orozco
Departament de Bioquímica i Biologia Molecular
Facultat de Biologia, Universitat de Barcelona
Martí i Franqués 1, 08028 Barcelona (Spain)

[e] Prof. F. J. Luque
Departament de Fisicoquímica
Facultat de Farmàcia. Universitat de Barcelona
Avda. Diagonal 643, 08028 Barcelona (Spain)

Supporting information for this article is available on the WWW under http://www.chemeurj.org/ or from the author: Geometries of all the studied systems, overview of the empirical potential data, and additional QM results.

Introduction

Stacking of nucleobases is a key interaction in nucleic acids (NAs) that modulates their structure, stability, dynamics, sequence-dependent properties, and folding.[1–25] Thus, base-stacking forces and their role in nucleic acids were studied by techniques ranging from solution thermodynamic experiments to gas-phase quantum-chemical investigations.

The stacking phenomenon is more complex than often assumed. On the one hand, we can describe base stacking using the intrinsic (gas-phase) interaction energies, which reflect the direct forces between the stacked bases. On the other hand, stacking can be characterized by thermodynamic parameters in solution experiments and considering various types of nucleic acids. The intrinsic stacking energies do not directly correlate with the thermodynamic measurements, as the outcomes of the experiments are affected by many other contributions, mainly solvation effects. Thus, thermodynamic data cannot be unambiguously derived from direct base-stacking interactions, and vice versa. A proper description of the intrinsic interactions nevertheless helps to understand the basic principles governing formation of

three-dimensional NA architectures and is of key importance in all molecular-modeling approaches in which the intrinsic energy terms appear explicitly. Distinct experiments may actually indicate different stacking stabilities and nature of base-stacking forces, because the interplay between the intrinsic terms (i.e., direct base–base-stacking forces) and all other contributions may vary depending on the type of nucleic acid structure and type of measurement.[8,20,22,23] In some cases, a given interaction may stabilize one type of nucleic acid architecture while at the same time it destabilizes another, as is the case for stacking between protonated cytosines in i-DNA and DNA triplexes.[22,23] The intrinsic stacking energy is best described by ab initio quantum chemical (QM) theory, which allows reliable energies to be assigned to any base–base configuration.[18,22] Calculations, however, should be performed at an adequate level of theory. For example, density functional, Hartree–Fock, and semiempirical methods fail for base stacking due to their inability to properly capture dispersion effects.[18,25]

Reference QM base-stacking energies were obtained around a decade ago[26] by using medium-quality second-order Møller–Plesset (MP2) method and the 6-31G*(0.25) diffuse-polarized basis set of atomic orbitals.[27] Such MP2/6-31G*(0.25), or closely related, calculations were carried out for a wide range of configurations of stacked nucleobase dimers,[18–20,21a,b,22,26,28] base–intercalator complexes,[29,30] and other aromatic stacking interactions of biological or chemical interest.[31–36] The studies demonstrated that aromatic stacking can be rather well described as a combination of three contributions: short-range exchange repulsion, dispersion attraction, and electrostatic forces. Other hypothetical contributions were ruled out.[26]

Recent developments in computer hardware and software enable major improvements in calculations of aromatic stacking.[37–40] Stabilization energies can be evaluated to the complete basis set (CBS) limit with the MP2 method. The CBS limit is obtained by extrapolation from series of MP2 calculations with large basis sets of atomic orbitals.[41,42] The standard MP2 method is usually substituted by the resolution-of-identity MP2 (RI-MP2) technique,[43–45] which provides essentially identical energies and is quite inexpensive.[46] Higher order contributions to the electron correlation are included by using the coupled-cluster method with noniterative evaluation of the triple electron excitations, abbreviated as CCSD(T). The higher order electron-correlation contributions are typically repulsive for all aromatic stacked clusters.[47] The final stacking energies can be denoted as CBS(T). Recent analysis of the potential energy surface of stacked cytosine dimer revealed that the difference between the CBS(T) values and the MP2/6-31G*(0.25) values varies from $+0.3$ to -2.1 kcal mol^{-1} for the range of stacking energies between $+2.5$ and -10 kcal mol^{-1}.[40]

Here we present CBS(T) stacking energies for all ten unique B-DNA base pair steps. We evaluate almost 100 individual base–base geometries and thus extending the available literature data substantially. Furthermore, our analysis includes the many-body contribution (nonadditivity of stacking) and estimates of solvent effects. New reference values are compared with results from empirical force-field calculations to highlight the nature of base stacking and to estimate the accuracy of force fields used for molecular modeling.

Methods

Selection of geometries: All ten unique combinations of B-DNA base pair steps were considered: 5′-AA-3′(5′-TT-3′), AT(AT), TA(TA), GG(CC), GC(GC), CG(CG), GA(TC), AG(CT), TG(CA) and GT(AC) (A = adenine, C = cytosine, G = guanine, T = thymine; Figure 1).

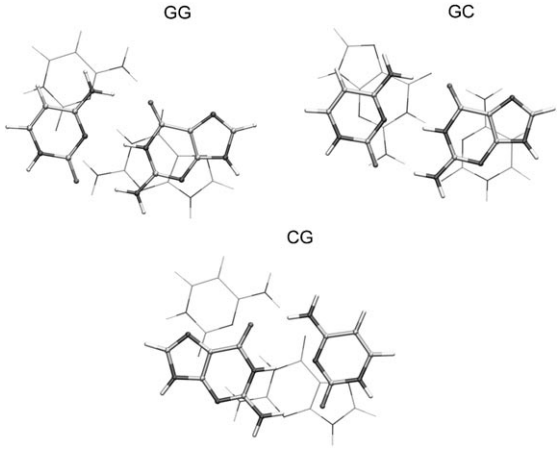

Figure 1. B-DNA base stacking (helical twist 36°) in purine–purine GG (= pyrimidine–pyrimidine CC) step (top left), purine–pyrimidine GC step (top right), and pyrimidine–purine CG step (bottom).

A rather straightforward way to select geometries would be to use oligonucleotide X-ray structures. However, direct use of experimental data can introduce bias into the energy calculations. Due to the nonlinear dependence of energy on interatomic distances, a rather modest structural error in experimental interbase positions (typically close interatomic contacts) may cause major bias in energy calculations.[13] Thus, we construct idealized base-pair steps (stacking of two consecutive base pairs) with QM-optimized geometries of base pairs. An appropriate force field (known from previous studies to reliably estimate stacking energy profiles) is used to map the potential energy surface.[13,16] The QM energies are then evaluated for carefully selected and well-defined geometries. The CG and AT H-bonded base pairs were optimized at the HF/6-31G** level, with hydrogen atoms capping the N1 pyrimidine and N9 purine positions. The long base-pair axis is defined by C8 of purine and C6 of pyrimidine in the optimized base pair. (C6–C8 axis is a common reference axis; note that the exact position of the C6–C8 axis with respect to the other atoms depends on the reference base-pair geometry.)

Initially, the two base pairs are positioned in the *xy* plane with the long base-pair axes and their geometrical centers

coincident (Figure 2). Propeller twist is introduced as a counterrotation of the two bases in the base pairs along the C6–C8 axes (Figure 3). Both base pairs have equal propeller

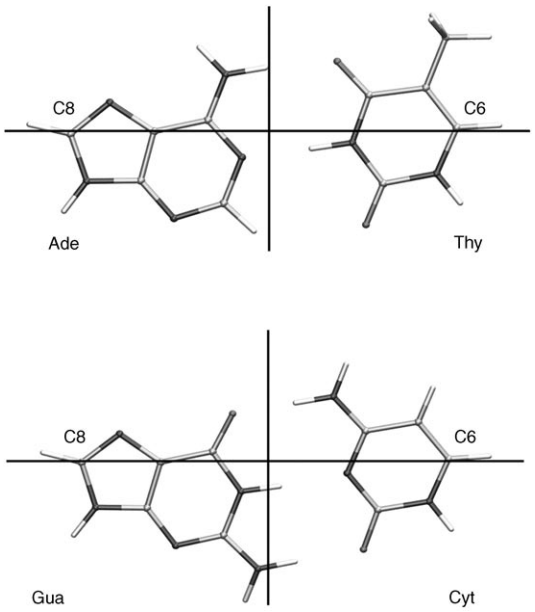

Figure 2. Watson–Crick AT (top) and GC base pairs, with their long (C8–C6) and short axes indicated.

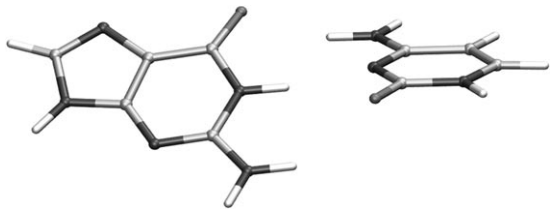

Figure 3. Base-pair propeller twisting (exaggerated).

twist, which is negative in B-DNA.[9,15,48] Then helical twist is introduced as an in-plane rotation of the upper base pair in the step along its long-axis center in the right-handed sense (Figure 1). Rotational matrices for propeller and helical twist are independent, and the outcome does not depend on the order of the rotations. Finally, vertical separation is introduced by shifting the upper base pair in the vertical (z) direction. It is numerically described by rise, here defined as the vertical distance between the original centers of the C6–C8 axes around which all rotations are specified and which are fixed when rotations are applied.

We consider helical twist, propeller twist, and rise as the most important parameters for our study. Description of irregular B-DNA helices would require a wide range of additional parameters (buckle, shear, shift, slide, roll, etc.).[9–15,48,49] These parameters are often interdependent, and the outcome of structure calculation depends on the order in which rotational matrices are applied, choice of reference frames, and so on. Further explanation is beyond the scope of this paper and the reader is referred to specialized literature.[48] Here, ambiguity is reduced as we vary only a limited number of parameters; all geometries are listed in the Supporting Information.

Optimization of vertical separation and rise: The vertical separation between consecutive base pairs in DNA molecules is always optimized.[13] Any vertical compression or extension of the largely flat base pairs is associated with substantial energy penalty. Thus, the vertical distance must be reoptimized after all other parameters are set up. The effective physical distance between base pairs is always around 3.3–3.4 Å, irrespective of the DNA form. (The base pair distance in A-DNA along the global helical axis is smaller due to inclination of the base pairs, but the inter-base-pair separation remains 3.3–3.4 Å in the local coordinate frame.[13,48]) Note that vertical separation and rise are not the same. By vertical separation we mean the actual physical contact between the base pairs due to the balance of dispersion attraction and short-range repulsion.[13] Rise is an associated geometrical parameter that can be defined, for example, as vertical intrastrand distance between the C1′ atoms in either local or global coordinate frame, by the vertical distance of centers of the C6–C8 axes, and so on. Depending on definition, computer programs may provide different numerical values of rise for identical geometries. For example, the base pairs can be buckled (counterrotation of bases around short base-pair axis that is perpendicular to the C6–C8 axis) in the opposite sense (nonzero cup; Figure 4).[14,48,49] Then, for example, the base-pair centers may be buckled away from each other while the base pair ends are clamped together (positive cup), or vice versa (negative cup; Figure 4). The numerical value of rise then depends on the exact location

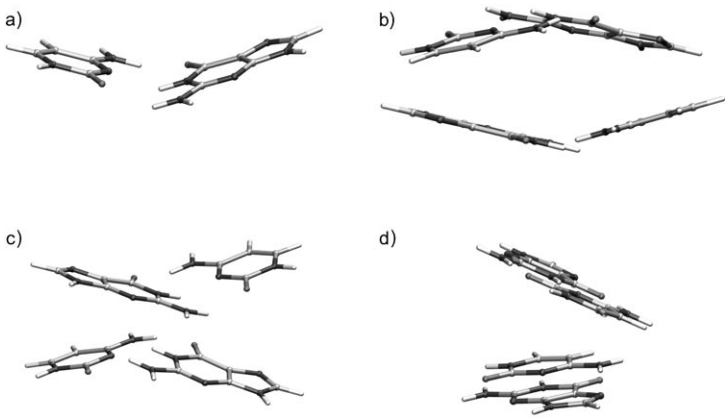

Figure 4. a) Base-pair buckle, i.e., counterrotation of the base pairs along the short base-pair axis. b) Positive cup (difference in buckling) separates the base pair centers (where steric clashes due to propeller twist may occur) and clamps together the base pair ends. c) Negative cup. d) Base-pair roll opens the minor groove and eliminates minor-groove clashes. All maneuvers are shown for CG step and are exaggerated to better illustrate the movements.[9–15,48,49]

of the points defining rise and how they move when rotations of bases are applied. In this paper, rise is defined in terms of the vertical distance of the original C6–C8 axis centers around which all rotations are defined. Thus, optimized rise would systematically increase and decrease for positive and negative cup, respectively, and is not systematically affected by helical twist and propeller twist.

Interaction (stacking) energy: The interaction energy $\Delta E^{A//C}$ of a stacked dimer A//C (// denotes stacking) is defined as the electronic energy difference between the dimer ($E^{A//C}$) and the isolated monomers (E^{A}, E^{C}), corrected for basis set superposition error [Eq. (1)].[50] As we use rigid monomers, deformation energy is not applicable.

$$\Delta E^{A//C} = E^{A//C} - (E^{A} + E^{C}) \quad (1)$$

When the studied systems consist of two stacked base pairs AB and CD, the total stacking energy $\Delta E^{AB//CD}$ can be expressed as a difference in electronic energy of the complex and the base pairs [Eq. (2)]

$$\Delta E^{AB//CD} = E^{AB//CD} - (E^{AB} + E^{CD}) \quad (2)$$

or by means of four pairwise dimer interaction energies and the four-body term ΔE^{4} [Eq. (3)].

$$\Delta E^{AB//CD} = \Delta E^{A//C} + \Delta E^{A//D} + \Delta E^{B//C} + \Delta E^{B//D} + \Delta E^{4} \quad (3)$$

We do not include the two horizontal H-bonding energies (A···B and C···D contributions). Stability of H-bonded base pairs was studied elsewhere.[51,52] Analysis of planar base pairs would require additional in-plane geometry adjustments and thus complicate the stacking analysis.

Complete basis set limit of the MP2 stabilization (interaction) energies: The RI-MP2 calculations were performed with aug-cc-pVDZ and aug-cc-pVTZ basis sets of atomic orbitals. Several robust extrapolation schemes have been suggested in the literature.[41,42] In the present paper we used two methods. First, we applied schemes of Helgaker et al. [Eq. (4)][41]

$$E_{X}^{HF} = E_{CBS}^{HF} + A\mathrm{e}^{-\alpha X} \text{ and } E_{X}^{corr} = E_{CBS}^{corr} + BX^{-3} \quad (4)$$

where E_X and E_{CBS} are energies for the basis set with the largest angular momentum X and for the complete basis set, respectively, and α is a parameter fitted by the authors.[41] More details can be found in our previous papers.[39,40,51] Key calculations were also repeated with the Truhlar extrapolation scheme.[42] The method of Helgaker et al. is likely more conservative and is available also for basis sets larger than aug-cc-pVTZ, while Truhlar's scheme for VTZ-quality basis set is closer to the CBS limit.[51]

Correction for higher order correlation effects: The difference between CCSD(T) and MP2 interaction energies ($\Delta E^{CCSD(T)} - \Delta E^{MP2}$) exhibits only a small basis-set dependence. Therefore, we utilized the 6-31G*(0.25) basis set to approximate the CBS CCSD(T) interaction energy [Eq. (5)].

$$\Delta E_{CBS}^{CCSD(T)} = \Delta E_{CBS}^{MP2} + (\Delta E^{CCSD(T)} - \Delta E^{MP2})|_{6\text{-}31G*(0.25)} \quad (5)$$

All pairwise base–base terms are characterized at the $\Delta E_{CBS}^{CCSD(T)}$ level, while the $\Delta E^{AB//CD}$ calculations to determine the four-body corrections ΔE^{4} are done at the RI-MP2/aug-cc-pVDZ level.

Empirical potential calculations: The force field consisted of a van der Waals term taken from the force field of Cornell et al. (AMBER)[53] and the Coulombic term with atom-centered point charges. To improve the agreement between QM and force-field calculations, the radius of the methyl hydrogen atoms of thymine was reduced to 1.087 Å, that is, by 0.4 Å.[54] The standard value caused premature clashes in steps in which the methyl group contacts another base on propeller twisting, that is, the original hydrogen atoms appeared oversized. This adjustment is needed solely for our analysis and we do not suggest it for DNA simulations, since it could cause some other imbalances of the force field. The impact of oversized methyl hydrogen atoms is assumed to be rather small in simulations. DNA is very flexible and can avoid these clashes by subtle adjustments of many other parameters, in contrast to our conformational scans. The charges were derived by using the electrostatic potential (ESP) fitting for the monomers at the MP2/aug-cc-pVDZ level. The potential is thus very similar to AMBER, except that AMBER for condensed-phase simulations has charges derived at the HF level to implicitly account for polarization. The MP2 charges allow direct comparison with QM data.[26]

Solvation calculations: Solvent effects were determined as the loss of solvation free energy in an imaginary process in which two fully solvated bases or base pairs assemble into a given conformation (that used in the gas-phase studies) in a fully solvated stacked base dimer (two bases, see below) or stacked base-pair step (four bases). Thus, the solvent effect on nucleobase stacking is computed as the difference in solvation free energy between the stacked dimer and the isolated bases. Similarly, the solvent effect on base-pair stacking is determined as the difference in solvation free energy between the stacked tetramer and the isolated H-bonded pairs. Solvation free energies were determined at the MST/HF/6-31G(d) and MST/B3LYP/6-31G(d) levels.[55–57] Electrostatic and nonelectrostatic terms were considered to evaluate the solvation free energy. The electrostatic contribution was determined by using the integral equation formalism.[58–60] In the MST model, the nonelectrostatic components include two terms, the cavitation and van der Waals contributions. Cavitation was determined by using the Claverie–Pierotti formalism, and the van der Waals term was computed by using a linear relationship for the solvent-exposed surface with the atomic tensions obtained by fitting

the experimental solvation free energies (see refs. [55–57] for details). In particular, calculations were carried out in four different solvents (water, octanol, carbon tetrachloride, and chloroform) for which MST parameters are available.[61–63] Calculations were carried out with consideration of gas-phase optimized geometries. All computations were performed with a locally modified version of Gaussian 03.[64]

Abbreviations: SV, aDZ, and aTZ stand for the RI-MP2 (or MP2) interaction energies evaluated with the 6-31G*(0.25), aug-cc-pVDZ, and aug-cc-pVTZ basis sets, respectively. CBS is the MP2 basis set limit, Δ(T) is difference between CCSD(T) and MP2 data evaluated with the 6-31G*(0.25) basis set, SV(T) the CCSD(T)/6-31G*(0.25) interaction energy, and CBS(T) the final CBS estimate after the CCSD(T) correction [Eq. (5)]. The stacking geometries are abbreviated in the following way: XY means base pair step (stacking of two base pairs) 5′-XY-3′ while X//Y means stacking of two bases X and Y. In some cases, the geometry of the base pair step is specified by the abbreviation XY*a*/*b* which means 5′-XY-3′ base pair step with propeller twist *a* and rise *b*. Thus, abbreviation CG0/3.19 means 5′-CG-3′ step with propeller twist 0 and rise 3.19 Å. Finally, "s" and "is" indicate intra- and interstrand stacking.

Table 1. Dependence of intrinsic stacking energy (force field estimate [kcal mol^{-1}]) for ten unique B-DNA base-pair steps for assumed propeller twist of 0° and rise of 3.36 Å.

Base-pair step	Helical twist 30°	Helical twist 36°	Helical twist 42°
AA	−14.74	−15.01	−14.91
AT	−15.15	−14.78	−14.63
TA	−14.56	−14.67	−14.41
GG	−13.09	−13.57	−13.73
GC	−16.56	−16.12	−15.53
CG	−16.39	−16.41	−16.05
GA	−14.19	−14.29	−14.09
AG	−14.54	−14.65	−14.60
TG	−15.70	−15.81	−15.55
GT	−15.16	−14.98	−14.86

Table 2. Intrinsic stacking energy (force field estimate [kcal mol^{-1}]) for ten unique base pair steps, assuming optimized propeller twist [°] and rise [Å].

Base-pair step	Stacking energy	vdW energy	Electrostatic energy	Propeller twist	Rise[a]	Energy gain due to propeller twist[b]
AA	−16.54	−16.64	+0.10	20	3.15	−1.52
AT	−15.85	−17.82	+2.23	14	3.22	−0.68
TA	−14.80	−15.73	+0.93	6	3.25	−0.08
GG	−13.98	−17.37	+3.39	6	3.30	−0.30
GC	−16.12	−17.53	+1.41	0	3.35	–
CG	−16.58	−16.46	−0.12	0	3.29	–
GA	−14.29	−16.29	+2.00	0	3.36	–
AG	−15.25	−17.15	+1.90	8	3.28	−0.68
TG	−16.02	−15.92	−0.10	2	3.28	−0.02
GT	−15.09	−17.65	+2.56	6	3.29	−0.10

[a] Defined as vertical distance between the original C6–C8 axes centers around which all rotations are defined (see Methods Section). [b] Energy difference between structures with optimal and 0° propeller twist, both with optimized vertical separations.

Results and Discussion

Helical twist: Table 1 shows that the variation in stacking energy is only modest in the helical twist range 30–42°.[65] The weak energy dependence reflects mutual compensation of inter- and intrastrand stacking on changing the helical twist. Thus, we restricted the QM analysis to a helical twist of 36°.

Propeller twist and vertical separation: Propeller twist improves intrastrand stacking but can lead to steric clashes. Table S1 (Supporting Information) lists empirical potential stacking energies for propeller twists in the range of 0–30°. Vertical separation was optimized (see Methods Section for definition) for each value of propeller twist. Propeller twist introduces notable changes in the stacking energy in several steps. Table 2 summarizes optimal classical (force-field) energy (decomposed into van der Waals and electrostatic components), optimal rise, propeller twist, and the gain in stability due to optimization of propeller twist. Propeller twist improves intrastrand stacking and also leads to steric interstrand minor-groove clashes, mainly in pyrimidine–purine steps.[9–16] The largest energy gain due to propeller twist is found in the AA(TT) step. The significant intrinsic tendency of the AA step to be propeller-twisted is one of the main contributions to the unique properties of A-tracts (B-DNA stretches with several consecutive adenines in one strand).[66–73] Note (see Table S1, Supporting Information) that when propeller twist goes from 0 to optimum 20°, the vertical distance (rise) between the base pairs is reduced by 0.2 Å, reflecting the actual compaction of the AA step (see Methods Section). Calculations with fixed rise would miss the energy gain due to propeller twist.

The AT base-pair step also exhibits substantial propeller twisting. On the other hand the CG step (Supporting Information) shows steep deterioration of stacking with increasing propeller twist due to mutual steric clashes between the N2 amino groups and N3 atoms of the two symmetrically oriented guanines (Figure 1 bottom).[9,74] This clash can be eliminated by a number of other conformational parameters. Among them, positive cup (difference of buckles separating the base pair centers, Figure 4b)[14,74] and roll (local bending to the major groove alleviates minor groove clash, Figure 4d)[9,14,74] are the most efficient, together with base-pair stagger (effectively shifting the propeller axis to the minor groove).[14] The rather weak stacking in the GG(CC) step is

caused by unfavorable electrostatic energy between the guanines and cytosines in the strands.[19,70]

Reference ab initio calculations: Table 3 lists the main part of the QM data. The ab initio calculations were carried out either for the optimal force-field geometry or for a geometry that is close to the force-field minimum (two-dimensional search; helical twist 36°, rise and propeller twist varied). Since the force field and QM potential energy surfaces are not identical, we sometimes slightly further adjusted the geometry, partly based on preliminary MP2/6-31G*(0.25) data (not shown). All geometries are nevertheless sufficiently close to the minimum, since the potential energy surface is flat around the minima. When the force field predicted a marked propeller twist, the calculations were carried out for propeller-twisted structures (Supporting Information). The remaining steps with small propeller twist were considered with flat base pairs, while flat and propeller twisted geometries were evaluated for the AA step. Rise was taken either as the optimum force-field value (Table 2) or was reduced by up to 0.1 Å. The reason is that the ab initio data usually predict slightly reduced vertical distance of stacked bases compared with the force field (see below). Both the optimal force-field rise and its reduced value are entirely suitable for the reference calculation. Our set of geometries partly reflects the interplay between stacking and local variations and is thus superior to using uniform geometries.[19]

Table 3. Individual interaction energy terms in the ten unique base-pair steps [kcal mol^{-1}].[a]

Base-pair step and individual base-base terms	aDZ	aTZ	CBS[b]	Δ(T)	CBS(T)	E_{pot}	E_{vdW}
GC0/3.25[a]							
G//Cs	−11.21	−11.93	−12.22 (−12.39)	1.42	−10.80	−10.12	−6.92
C//Cis	2.89	2.88	2.87 (2.87)	0.22	3.09	2.93	−0.61
G//Gis	1.48	1.35	1.29 (1.24)	0.63	1.93	1.71	−2.82
CG0/3.19							
G//Cs	−8.05	−8.43	−8.57 (−8.62)	0.69	−7.88	−6.60	−5.43
G//Gis	−4.03	−4.56	−4.79 (−4.94)	0.88	−3.91	−4.11	−4.50
C//Cis	1.16	1.12	1.10 (1.08)	0.14	1.24	1.04	−0.82
GG0/3.36							
G//Gs	−4.90	−5.42	−5.64 (−5.77)	2.10	−3.54	−4.31	−8.49
C//Cs	−2.13	−2.52	−2.67 (−2.76)	1.05	−1.62	−1.73	−5.30
C//Gis	−3.41	−3.54	−3.59 (−3.61)	−0.10	−3.68	−3.39	−1.91
G//Cis	−4.64	−4.68	−4.70 (−4.70)	−0.12	−4.82	−4.4	−1.21
GA10/3.15							
A//Gs	−10.51	−11.29	−11.60 (−11.79)	2.47	−9.14	−8.94	−7.87
T//Cs	−5.06	−5.55	−5.75 (−5.87)	1.05	−4.69	−5.15	−5.65
A//Cis	−0.18	−0.27	−0.31 (−0.33)	0.00	−0.31	−0.09	−1.81
T//Gis	0.49	0.41	0.38 (0.36)	0.19	0.58	0.46	−1.37
AG08/3.19							
A//Gs	−8.87	−9.30	−9.46 (−9.55)	1.89	−7.58	−7.44	−7.55
T//Cs	−6.22	−6.64	−6.81 (−6.90)	0.73	−6.07	−6.23	−5.95
T//Gis	−0.19	−0.45	−0.56 (−0.63)	0.09	−0.47	−0.84	−2.01
A//Cis	−0.11	−0.21	−0.26 (−0.29)	0.08	−0.18	−0.41	−1.39
TG0/3.19							
T//Gs	−5.92	−6.34	−6.50 (−6.59)	0.83	−5.67	−5.68	−5.58
A//Cs	−5.46	−5.94	−6.12 (−6.21)	1.16	−4.96	−4.37	−5.31
A//Gis	−4.32	−4.72	−4.88 (−4.99)	0.66	−4.22	−4.45	−3.93
T//Cis	−1.12	−1.14	−1.15 (−1.16)	0.00	−1.15	−1.17	−0.66
GT10/3.15							
T//Gs	−5.62	−6.42	−6.74 (−6.94)	1.78	−4.96	−5.80	−7.82
A//Cs	−6.26	−6.84	−7.07 (−7.21)	1.63	−5.44	−5.05	−6.56
T//Cis	0.22	0.21	0.20 (0.20)	0.10	0.30	0.21	−0.62
A//Gis	−3.99	−4.20	−4.29 (−4.34)	0.23	−4.06	−3.97	−2.50
AT10/3.26							
A//Ts	−7.39	−7.99	−8.24 (−8.39)	1.60	−6.64	−7.42	−7.43
T//Tis	0.8	0.75	0.73 (0.72)	0.16	0.88	0.40	−1.04
A//Ais	−0.87	−1.01	−1.07 (−1.12)	0.15	−0.92	−1.18	−1.81
TA08/3.16							
A//Ts	−6.47	−6.91	−7.09 (−7.19)	1.02	−6.07	−6.72	−6.04
A//Ais	−1.87	−2.24	−2.40 (−2.50)	0.85	−1.55	−1.24	−2.50
T//Tis	0.60	0.58	0.57 (0.56)	0.14	0.70	0.48	−0.68
AA0/3.24							
A//As	−7.74	−8.35	−8.59 (−8.74)	2.34	−6.25	−6.0	−6.53
T//Ts	−4.13	−4.85	−5.15 (−5.32)	1.29	−3.86	−5.04	−5.58
A//Tis	−1.71	−1.75	−1.77 (−1.79)	0.07	−1.71	−2.11	−1.83
T//Ais	−1.31	−1.37	−1.39 (−1.40)	0.09	−1.30	−1.62	−1.43
AA20/3.05							
A//As	−7.66	−8.33	−8.60 (−8.76)	2.54	−6.06	−6.0	−6.8
T//Ts	−4.61	−5.22	−5.47 (−5.63)	1.29	−4.18	−5.6	−6.4
A//Tis	−2.05	−2.23	−2.31 (−2.37)	−0.03	−2.34	−2.1	−1.5
T//Ais	−2.01	−2.18	−2.25 (−2.30)	0.09	−2.16	−2.1	−1.2

[a] See Methods Section for abbreviations; helical twist 36°, all geometries are provided in Supporting Information. [b] Values of Helgaker et al.[41] (in parentheses the value extrapolated according to Truhlar[42]).

The first five columns of Table 3 contain aug-cc-pVDZ, aug-cc-pVTZ, and extrapolated values of the MP2 stacking energies, followed by the higher order electron correlation correction and the final stacking energy CBS(T) (sum of columns 3 and 4). Note that the extrapolations of Helgaker et al. and Truhlar provide essentially identical results. The last two columns present the empirical potential values and their van der Waals (vdW) component. In the abbreviation of the step, the first number is the propeller twist and the second number the rise (see Methods Section for all abbreviations). Note that steps with twofold symmetry have two equivalent intrastrand stacking contributions.

The individual intrastrand stacking contributions are in the range of −10.8 (GC step) to −1.6 kcal mol^{-1} (C//C stack

in the GG step). The individual interstrand terms are within the range of −4.8 (G//C term in the GG step) and +3.1 kcal mol^{-1} (C//C term in the GC step). The CCSD(T) corrections are in the range of −0.1 to +2.5 kcal mol^{-1}. The differences between the aDZ and CBS(T) data vary in the range of +0.2 to −1.6 kcal mol^{-1}, that is, the aDZ stacking stabilization is in most cases weakly exaggerated. The empirical potential values are within +1.3 to −1.1 kcal mol^{-1} of the CBS(T) data. The average root mean square deviation between classical and CBS(T) data is only 0.5 kcal mol^{-1}, the Pearson correlation coefficient is 0.99, and the scaling coefficient is 1.00. The largest individual differences between QM and classical values are found for T// T intrastrand stacking and may be partly due to methyl-group parametrization (see Table 3 and Methods Section). In summary, despite the complexity of the stacking interaction, it can be well reproduced by empirical force fields. The van der Waals terms vary from −5.3 to −7.9 and from −0.6 to −3.9 kcal mol^{-1} for the individual intra- and interstrand terms, respectively. Thus, they are considerably more uniform than the full stacking energies. This reflects the role of the electrostatic terms in determining sequence-dependent changes in gas-phase stacking.

Table 4. Stacking energies in B-DNA base pair steps [kcal mol^{-1}].

	CBS(T), intrastr.[a]	CBS(T), interstr.[a]	CBS(T) total[a]	SV,[b] total	Potential, total	$\Delta E^{AB//CD}$ (ΔE^4)[c]
GC0/3.25	−21.6	+5.1	−16.6	−14.1	−15.6	−15.8 (+1.2)
CG0/3.19	−15.8	−2.7	−18.4	−13.8	−16.3	−17.3 (+1.1)
GG0/3.36	−5.2	−8.5	−13.7	−11.5	−13.8	−11.2 (+2.2)
GA10/3.15	−13.8	+0.3	−13.6	−12.1	−13.7	−12.9 (+0.7)
AG08/3.19	−13.6	−0.7	−14.3	−12.2	−14.9	−12.5 (+0.8)
TG0/3.19	−10.6	−5.4	−16.0	−12.5	−15.7	−15.1 (+0.9)
GT10/3.15	−10.4	−3.8	−14.2	−12.3	−14.6	−13.4 (+0.8)
AT10/3.26	−13.3	0.0	−13.3	−11.6	−15.6	−13.3 (0.0)
TA08/3.16	−12.1	−0.9	−13.0	−11.2	−14.2	−12.8 (+0.2)
AA0/3.24	−16.1	3.0	−13.1	−12.0	−14.7	−13.1 (0.0)
AA20/3.05	−10.2	−4.5	−14.7	–	−15.8	−14.7 (0.0)

[a] Intrastr. and interstr. stand for sum of the two intra- and interstrand stacking terms. Total = intrast. + intrestr., i.e., base-pair-step stacking energy within the pair-additive approximation, without the many-body term. [b] MP2/6-31G*(0.25) reference[19] without the many-body term, with helical twist 36°, propeller twist 0°, and rise 3.36 Å. [c] Total stacking energies after adding the four-body term listed in parentheses.

Table 4 summarizes, in a compact way, the data from Table 3. The first three columns list the intrastrand, interstrand, and total CBS(T) stacking energies, while the fourth column gives the decade-old MP2/6-31G*(0.25) (SV) reference data. The last column shows the total stacking energies after addition of the four-body correction ΔE^4 (in parentheses, aDZ level). There is a substantial degree of mutual compensation of intra- and interstrand stacking terms. The weakest intrastrand stacking of −5.2 kcal mol^{-1} occurs in the GG step, due to the unfavorable intrastrand electrostatic interaction between the cytosines and guanines, which is compensated for by the most attractive interstrand interaction of −8.5 kcal mol^{-1}. Entirely opposite distribution is seen for the GC step, with the most attractive intrastrand interaction of −21.6 kcal mol^{-1} and the most repulsive interstrand term of +5.1 kcal mol^{-1}.

The decade-old SV reference data suggest that the individual base-pair steps have, without the many-body term, a very narrow range of stacking energies varying from −11.2 to −14.1 kcal mol^{-1}. This range is expanded from −13.0 to −18.4 kcal mol^{-1} at the CBS(T) level. Thus, the CBS(T) method reveals a nonnegligible enhancement of stacking stabilization compared with the SV level and, more importantly, the individual base-pair steps are no longer isoenergetic. We consider this to be a significant qualitative difference between the SV and CBS(T) reference data. It stems from the cumulation of quantitative differences for the individual base–base terms.

The average force-field stacking energy (−15.8 kcal mol^{-1}) is close to the average CBS(T) value (−14.7 kcal mol^{-1}), but the spread of values is larger for the CBS(T) method (range and standard deviation 5.4 and 1.7 kcal mol^{-1} for CBS(T) and 2.6 and 1.1 kcal mol^{-1} for classical evaluation; not shown). It is then clear that, though the force field is able to capture reasonably well individual intra- and interstrand interactions, it does not reproduce equally well the sum of these interactions. Analysis of the individual terms (see Table 3) shows that many discrepancies between CBS(T) and classical values arise from T//T stacking. This may reflect the difficulties in parameterization of the van der Waals parameters for the methyl group (see Methods Section) and possibly use of ESP charges when the methyl group comes into close contact with the other monomer. However, nonnegligible differences are also seen in the three GC-containing steps. These could, for example, reflect lack of polarization in the calculation of the MP2 atomic charges. Without additional extended calculations, however, it is not possible to pinpoint the exact origin of such moderate discrepancies between the force-field and QM data. For example, anisotropic short-range repulsion may also lead to differences between simple force fields assuming spherical atoms and QM calculations.[75] The absolute force-field values are in a better agreement with the new reference QM data (compared with older SV data[19]). The relative discrepancies appear to be modestly larger, in line with the differences between the old and new QM reference values. Comparison with several older studies (published before 1995) on base stacking calculations can be found in Figure 3 of ref. [19]. As the results are very different from those of modern ab initio calculations, these studies are not discussed here. Neither semiempirical approaches nor the sandwich model of stacking with explicit out-of-plane π charges provides a satisfactory estimate of intrinsic base stacking.

The four-body term evaluated at the aDZ level matches almost exactly (within 0.2 kcal mol^{-1}) earlier SV estimates.[19] The four-body interaction is negligible, except for the GG step, for which it amounts to +2.2 kcal mol^{-1} repulsion, which is 20% of the total stacking energy. After considering the four-body term, the final range of stacking energies is widened to −11.2 to −17.3 kcal mol^{-1}. Note that the MP2 method still neglects the nonadditivity of dispersion.[19]

To complete the evaluation, Table 5 lists stacking energies for optimized gas-phase geometries of ten stacked nucleobase dimers, assuming a vertical separation of bases of 3.3 or 3.4 Å and coplanarity of the bases. These reference structures have no relation to nucleic acids and were obtained by force field search with rigid MP2-optimized monomers (see Figure 3 in ref. [26] and Supporting Information). While in some cases the CBS(T) values are better approximated by aDZ data, for dimers with large CCSD(T) correction the SV interaction energies match the CBS(T) data better. The accuracy is thus not systematically improved when the 6-31G*(0.25) basis set is replaced by the aug-cc-pVDZ basis set at the MP2 level.

Table 5. Stacking energies for optimal arrangements of ten stacked nucleobase dimers, gas-phase minima.[a]

	aDZ	aTZ	CBS	Δ(T)	CBS(T)	SV[b]
A//A	−10.5	−11.1	−11.4	2.82	−8.5	−8.8
G//G	−13.8	−14.5	−14.8	2.13	−12.7	−11.3
A//C	−11.3	−11.9	−12.2	1.98	−10.2	−9.5
G//A	−13.2	−13.8	−14.0	2.62	−11.4	−11.2
C//C	−10.1	−10.7	−11.0	0.98	−10.0	−8.3
A//U	−10.7	−11.3	−11.6	1.81	−9.8	−9.1
G//C	−11.1	−11.7	−12.0	1.40	−10.6	−9.3
C//U	−10.2	−10.8	−11.1	0.68	−10.4	−8.5
U//U	−7.7	−8.3	−8.5	1.04	−7.5	−6.5
G//U	−12.4	−13.1	−13.4	1.31	−12.1	−10.6

[a] geometries taken from ref. [26] and listed in Supporting Infomation. [b] MP2/6-31G* (0.25) energies.[26]

Vertical separation: Base stacking energy is very sensitive to any vertical compression or extension of the stacked base-pair steps.[13] While a pure vertical compression of extended stacked systems is entirely prevented by the large short-range repulsion, pure vertical extension would lead to a major loss of dispersion attraction. Both compression and extension are associated with large energy gradients, and therefore vertical distance between stacked base pairs is always optimized independently of the other structural parameters.[13] Earlier SV calculations slightly underestimated the optimal vertical distance between bases (3.25 rather than 3.35 Å) compared to high-resolution crystal structures and well-calibrated force fields.[26] Therefore, for three base-pair steps we evaluated a few points with varied vertical distance. The AA20/3.25 geometry gives −15.45 kcal mol^{-1} stabilization at the aDZ level, which is changed by −0.65 and −0.84 kcal mol^{-1} on vertical compression by 0.1 and 0.2 Å, respectively. The corresponding CBS(T) values are −14.07, −0.66, and −0.67 kcal mol^{-1} (for full set of data, see Supporting Information). In the case of GG0/3.46, GG0/3.36, and GG0/3.26 geometries, the aDZ values are −14.26, −0.82, and −1.23 kcal mol^{-1}, and the CBS(T) values −12.87, −0.79, and −1.23 kcal mol^{-1}. Thus, for the AA step the CBS(T) method leads to slightly larger intermonomer separation compared to the aDZ level, while for the GG step there appears to be no change at all. The respective empirical potential values are −16.21, −0.33, and −0.06 kcal mol^{-1} for the AA, and −13.40, −0.29, and +0.07 kcal mol^{-1} for the GG step. For AA0/3.44, we obtained −14.19 kcal mol^{-1} at the aDZ level, which is improved by −0.58 and −0.70 kcal mol^{-1} on vertical compression by 0.1 and 0.2 Å, respectively. The corresponding CBS(T) values are −12.57, −0.53, and −0.55 kcal mol^{-1}, while the force-field values are −14.74, −0.28, and +0.14 kcal mol^{-1}. The data clearly demonstrate that the force field gives systematically a higher rise than QM. As indirectly follows from analysis of stacking energies and vertical separations of base pairs in DNA crystal structures,[13] the currently used AMBER van der Waals parameters[53] reproduce the vertical base-pair separations in high-resolution oligonucleotide X-ray structures and the rise of about 3.36 Å for idealized geometries. Combining both pieces of data, the QM method underestimates the X-ray vertical separation of base pairs. The effect is perhaps a little less apparent for the later AA0/3.44–3.24 set of geometries, but it is primarily due to the T//T base–base term. The A//A term tends to further compact the step (see Supporting Information). Three points for each step are not sufficient to determine the exact optimal rise, but it is clear that the CBS(T) level still subtly reduces the vertical distance between base pairs compared to modern force fields and values derived from oligonucleotide X-ray structures. The origin of this systematic difference is not clear, and further studies will be needed to clarify this point. Due to the extraordinary sensitivity of stacking to the vertical distance between the extended aromatic systems, the difference is not negligible.[13]

Local conformational variations: Local conformational variations in B-DNA are likely driven by base-stacking forces. There have been years of effort to derive the rules of sequence-dependent local conformational variations in B-DNA[9–15,20,71–74] but they remain rather elusive, largely because we are dealing with very subtle and delicate effects. Although the medium-level QM method and force field provide excellent qualitative description of base stacking, the changes seen above when using the CBS(T) method instead could easily obscure subtle quantitative interplay between base stacking and the local conformational variations. It is therefore tempting to apply the CBS(T) method for eventual reassessment of selected stacking geometries related to local B-DNA conformational variations. Thus, we compared three geometries (with force field optimal rise) of the CG step, which is known to be sensitive to minor-groove amino-group clashes caused by propeller twisting.[9–15,74] These were structures with a) a propeller twist of 0°, b) a propeller twist of −12° and a base-pair roll of 12°, and c) a propeller twist

of −12° and a cup of 28° (see Methods Section and refs. [14] and [74]). Both roll and cup belong to common maneuvers that eliminate amino-group clashing in the minor groove. Structure a) has an aDZ energy of −18.97, which changes by +1.37 and +2.25 kcal mol^{-1} for structures b) and c), respectively. These values are −18.43, −0.53, and +2.38 kcal mol^{-1} at the CBS(T) level and −16.19, −0.32 and +0.1 kcal mol^{-1} for the force field. Thus, the relative stabilities of the three distinct configurations of the CG step are quite sensitive to the level of calculation. These preliminary values indicate that an accurate analysis of the relation between local conformational variations of B-DNA and the stacking energies may require calculations at a high QM level, while some subtle variations of stacking energies associated with local conformational variations may be not well captured by empirical force fields. We plan to address this issue in future, since considerably more points on the potential energy surface (better sampling) will be needed for further insights, especially with regard to adjustment of vertical separation.

Solvation effects: The gas-phase calculations reflect stacking stabilization in the absence of any other contributions. In the following, we present continuum solvent estimates of solvation effects for the same stacking arrangements. Although such calculations on single geometries cannot be, for a variety of reasons, directly compared with thermodynamics of stacking in nucleic acids or in solution, they give insights into the interplay between intrinsic stacking and solvation effects for DNA-like arrangement of nucleobases. Note that theoretical assessment of thermodynamic parameters for stacking in water was attempted, for example, by Florian et al., based on MP2/6-31G*(0.25) calculations of gas-phase interaction energy and the Langevin dipole solvent model.[21a] Such calculations, however, require more thorough searches over the potential energy surface and are based on a number of approximations, which include calibration with experimental data.[21a] Such analysis is beyond the scope of our study.

Let us first consider the stacked base–base arrangements analyzed in Table 5. When two fully solvated nucleobases stack, a major loss of solvation leads to positive values of the differential solvation free energy. In water the loss of solvation ranges from 3.0 to 7.0 kcal mol^{-1} (MST/B3LYP/6-31G(d) level). It is reduced to 1–3 kcal mol^{-1} in CCl_4 (see Table 6). Not surprisingly, stacking of guanines and cytosines, which are the most polar bases (free energies of dehydration for the single nucleobases of 22–27 kcal mol^{-1}), implies a larger desolvation cost than for stacking of adenine and uracil (free energies of dehydration for the free nucleobases of 16–17 kcal mol^{-1}). Due to the neglect of entropic effects, values in Tables 5 and 6 cannot be directly summed to determine the stability of stacked bases in condensed phase. However, the results show that the absolute values of the intrinsic stacking energies are only slightly larger than the dehydration terms.[21] In general, dimers having the best intrinsic stacking energies are more disfavored by solvation

Table 6. Solvation contributions to nucleobase stacking [kcal mol^{-1}]. Geometries are optimal for gas-phase dimers (see Table 5 and Supporting Infomation). A positive value means that solvent disfavors the stacked complex. Roman values: HF/6-31G(d) data, italic values: B3LYP/6-31G(d) data.

	$\Delta\Delta G_{sol}$ (water)	$\Delta\Delta G_{sol}$ ($CHCl_3$)	$\Delta\Delta G_{sol}$ (octanol)	$\Delta\Delta G_{sol}$ (CCl_4)
A//A	3.5/*3.0*	1.8/*1.7*	2.6/*2.3*	0.9/*0.8*
G//G	8.1/*6.7*	4.3/*3.7*	6.0/*5.1*	2.4/*1.9*
A//C	5.9/*4.7*	3.4/*3.1*	4.6/*3.7*	1.8/*1.4*
G//A	6.0/*4.9*	3.2/*3.1*	4.5/*3.8*	1.6/*1.2*
C//C	8.0/*6.0*	4.7/*3.5*	6.1/*4.6*	2.5/*1.9*
A//U	5.4/*4.4*	2.9/*2.2*	4.0/*3.3*	1.5/*1.2*
G//C	8.2/*6.8*	5.0/*4.1*	6.4/*5.4*	2.8/*2.1*
C//U	7.9/*5.9*	4.5/*3.1*	5.9/*4.6*	2.4/*1.9*
U//U	4.9/*3.7*	2.7/*1.3*	3.7/*2.9*	1.5/*1.2*
G//U	8.8/*7.0*	5.0/*3.7*	6.5/*5.3*	2.7/*2.0*

(stacking energies in Table 5 and changes in solvation free energies in Table 6 are anticorrelated with Pearson's correlation coefficients around 0.8). Nevertheless, there are dimers with good intrinsic stacking and rather small desolvation term such as G//A, and others like C//C with moderate intrinsic stacking and quite large desolvation penalty. The G//A stack obviously has a large van der Waals overlap of bases, while the C//C stack has better electrostatic attraction. The smallest desolvation penalty is predicted for the adenine dimer, which reflects the known ability of adenines to self-associate in water. While the range of gas-phase stacking energies in Table 5 is −7.5 to −12.7 kcal mol^{-1}, adding the solvation terms (water) results in a reduced energy range of −3.8 to −6.0 kcal mol^{-1} for the set of stacked geometries studied.

When two base pairs are H-bonded, the dimer is much less solvated than the individual bases. The desolvation cost associated with stacking of two H-bonded base pairs is thus reduced. This feature, in conjunction with the different arrangement of bases in optimum dimers (Table 5) and in B-DNA-like stacked base-pair steps (Tables 3 and 4), explains why desolvation values in Table 7 are much smaller than the sum of the corresponding base-stacking desolvation energies. Solvent effects associated with the stacking of base pairs range from 1.4–9.8 (water) to 0.5–3.3 kcal mol^{-1} (CCl_4), without significant differences between HF/6-31G(d) and B3LYP/6-31G(d) MST estimates. For comparison, intrinsic stacking energies in Table 4 are in the range −13 to −17 kcal mol^{-1}.

There is no clear relationship between the type of stacked base pairs and the free energy of desolvation. There is a rough relationship between the intrinsic interaction energies (Table 4) and the free energies of desolvation (Table 7). Larger values of intrinsic stacking energies correlate (r = 0.86 for water) with larger desolvation costs. This finding can be explained by considering that dipole annihilation and burial of polar patches, which is one of the mechanisms for stabilization of stacking in the gas phase, reduces the ability of the solvent to interact with the stacked base pairs. For the GG, GC, and CG steps, the water solvation term almost

Table 7. Solvation contributions to base pair stacking [kcal mol^{-1}]. Geometries are those in Tables 3 and 4. Positive sign means that solvent disfavors stacking. Values in roman type: HF/6-31G(d) data, data in italics: B3LYP/6-31G(d) data.

	$\Delta\Delta G_{sol}$ (water)	$\Delta\Delta G_{sol}$ (octanol)	$\Delta\Delta G_{sol}$ ($CHCl_3$)	$\Delta\Delta G_{sol}$ (CCl_4)
GC 0/3.25	5.3/*5.5*	4.1/*3.9*	3.3/*3.1*	1.7/*1.7*
CG 0/3.19	9.8/*8.1*	7.5/*6.4*	5.6/*4.7*	3.3/*2.5*
GG 0/3.36	2.0/*1.8*	1.8/*1.8*	1.1/*1.1*	0.6/*0.5*
GA 10/3.15	3.3/*2.7*	2.5/*2.4*	1.7/*1.6*	0.9/*0.9*
AG 08/3.19	3.5/*2.7*	2.5/*2.8*	1.8/*1.7*	1.0/*0.7*
TG 0/3.19	7.1/*5.9*	5.4/*4.6*	3.7/*2.9*	1.9/*1.4*
GT 10/3.15	3.9/*3.2*	3.0/*2.8*	2.2/*2.1*	1.2/*1.1*
AT 10/3.26	2.2/*1.4*	1.3/*1.0*	0.9/*0.7*	0.5/*0.5*
TA 08/3.16	4.7/*3.6*	3.5/*3.0*	2.2/*1.8*	0.9/*0.5*
AA 0/3.24	4.2/*3.2*	2.6/*2.5*	1.7/*1.5*	0.8/*0.7*
AA 20/3.05	4.2/*3.2*	2.3/*2.1*	1.5/*1.3*	0.7/*0.6*

exactly cancels the variations in gas-phase stacking energy and makes these steps isoenergetic (within 1 kcal mol^{-1}). In B-DNA, however, the stacked base pairs are only partially hydrated. Thus, it is fair to assume that solvent compensation of the intrinsic stacking terms in B-DNA is smaller than in our calculations. In agreement with this qualitative consideration, experiments reveal the following nearest-neighbor stacking enthalpies in B-DNA: CG −10.6, GC −9.8, and GG −8.0 kcal mol^{-1}.[3] This stability order matches the gas-phase data but with reduced slope, as expected. When considering nearest-neighbor free energies, the stability order of GC and CG steps is interchanged.[3] This may reflect the large internal flexibility of the pyrimidine–purine CG step, which was not included in our calculations). Nearest-neighbor enthalpies also suggest that three AT-containing steps are mutually isoenergetic, similar to the four mixed CG+AT steps. This is also in a qualitative agreement with the gas-phase trends, and suggests that the intrinsic gas-phase stacking energies are at least partially reflected in nearest-neighbor stacking enthalpies. We wish to underline, however, that the comparison must be made with caution. The outcome of the experiments is affected by a number of contributions (besides base–base intrinsic energies), so the agreement could still be coincidental. It is not possible to directly compare steps with different composition of bases, since the experimental data do not separate stacking and base-pairing thermodynamic contributions.[3]

Conclusion

Base-stacking energies in ten unique B-DNA base-pair steps and some other stacking arrangements were evaluated by the second-order Møller–Plesset (MP2) method and complete basis set (CBS) extrapolation with aug-cc-pVDZ and aug-cc-pVTZ basis sets of atomic orbitals. Corrections for triple electron correlation contributions by means of the CCSD(T) method were done with the 6-31G*(0.25) basis set to give final stacking energy values denoted CBS(T). The B-DNA stacking geometries were considered with helical twist of 36°, optimized propeller twist, and optimized rise, based on preliminary empirical potential search. Thus, all calculations in this study are based on single-point calculations on selected geometries without applying the QM gradient optimization. Note that QM gradient optimization is not suitable for calculations of base stacking for a variety of reasons. These include the large basis set superposition error, which is not corrected by standard gradient techniques and which would vertically compress the stacked system. Gradient optimization would also lead to undesirable out-of-plane deformations of the rings and exocyclic groups of bases (this may be relevant to the gas phase but certainly not to DNA), and most likely the optimized geometries would slide far away from the B-DNA stacking arrangements.

The individual intrastrand stacking contributions are in the range of −10.8 to −1.6 kcal mol^{-1}, while the individual interstrand terms lie between −4.8 and +3.1 kcal mol^{-1}. The CCSD(T) corrections are in the range of −0.1 to +2.5 kcal mol^{-1}. There is a substantial degree of mutual compensation of intra- and interstrand stacking terms in the individual base-pair steps, which is most significant in the three steps consisting of two GC base pairs. In most steps the four-body interaction is negligible. An exception is the GG step, in which the four-body term is +2.2 kcal mol^{-1} and thus amounts to 20% of the total stacking energy. After considering the four-body term, the final range of stacking energies in the ten B-DNA steps is −11.2 to −17.3 kcal mol^{-1} (Table 5).

The CBS(T) calculations were compared with decade-old MP2/6-31G*(0.25) reference data and simple (AMBER) force field utilizing MP2-level ESP atom-centered point charges. The new reference calculations show modest increase of stacking stabilization and, surprisingly, substantially larger sequence-dependent variability of stacking energies compared to the MP2/6-31G*(0.25) data. The absolute force-field values are in better agreement with the new reference QM data, while the relative discrepancies between the force field and QM are modestly enlarged. Considering the obvious simplicity of the force field, its description of base stacking is surprisingly good. Better descriptions of base stacking should be achieved in future by polarizable force fields, while we do not see reasons to introduce more complicated electrostatic terms, such as distributed multipoles or out-of-plane charge distributions.[75]

The calculations indicate rather surprising systematic difference of ca 0.1 Å between the vertical separation of the base pairs predicted by quantum chemistry and derived from crystal structures. The force field reproduces the X-ray vertical dimensions of stacked base pair steps. This means that the optimal QM vertical base-pair distance is 3.25 Å rather than 3.35 Å. This difference was noticed in preceding MP2/6-31G*(0.25) calculations and is not mitigated when the level of theory is raised to CBS(T). The origin of this difference is not clear at this moment and deserves further studies, since the base stacking energy is very sensitive to any deviation of the vertical separation between the base pairs from its optimal value. It may be caused by a number

of factors, which include solvent effects, temperature averaging (QM data reflect interaction at 0 K), and a modest incompleteness of the QM approach, mainly regarding the higher order electron contributions.

Evaluation of three different local arrangements of the CG B-DNA step indicates sensitivity of the results to the level of calculations. Thus, finding subtle quantitative relations between local B-DNA geometrical variations and stacking may be more complicated than usually assumed, and may be beyond the accuracy of simple force fields. Therefore, extended CBS(T) calculations will be needed in future to better understand this issue.

The reference calculations were complemented by continuum solvent assessment of solvent-screening effects on stacking stabilization. Comparison with experimental nearest-neighbor DNA stacking enthalpies indicates that part of the gas-phase order of stacking stability is well reflected by the experimental enthalpies.

The present work provides the thus-far most physically complete analysis of the nature and magnitude of intrinsic base-stacking forces in B-DNA, which can be used, for example, as reference values for verification and parameterization of other computational methods and analysis of experimental data.

Acknowledgement

The Czech partners are grateful for financial support of the Wellcome Trust International Senior Research Fellowship in Biomedical Science in Central Europe GR067507 (J.S.), grants GA203/05/0009 and GA203/05/0388, Grant Agency of the Czech Republic (J.S., P.H.), LC512 research center by the Ministry of Education of the Czech Republic (J.S., P.H.) and research projects AVOZ50040507 (IBP) and AVOZ40550506 (IOCB), Ministry of Education of the Czech Republic. The Spanish team is thankful for financial support of the Spanish Ministry of Science (Grants BIO2003-06848 and SAF2002-04282), the BBVA, and La Caixa Foundations.

[1] D. H. Mathews, J. Sabina, M. Zuker, D. H. Turner, *J. Mol. Biol.* **1999**, *288*, 911–940.
[2] J. Santa-Lucia, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 1460–1465.
[3] J. SantaLucia, D. Hicks, *Annu. Rev. Biophys. Chem.* **2004**, *33*, 415–440.
[4] S. Bommarito, N. Peyret, J. SantaLucia, *Nucleic Acids Res.* **2000**, *28*, 1929–1934.
[5] E. Protozanova, P. Yakovchuk, M. D. Frank-Kamenetskii, *J. Mol. Biol.* **2004**, *342*, 775–785.
[6] a) K. M. Guckian, B. A. Schweitzer, R. X. F. Ren, C. J. Sheils, D. C. Tahmassebi, E. T. Kool, *J. Am. Chem. Soc.* **2000**, *122*, 2213–2222; b) K. M. Guckian, B. A. Schweitzer, R. X. F. Ren, C. J. Sheils, P. L. Paris, D. C. Tahmassebi, E. T. Kool, *J. Am. Chem. Soc.* **1996**, *118*, 8182–8183.
[7] J. S. Lai, J. Qu, E. T. Kool, *Angew. Chem.* **2003**, *115*, 6155–6159; *Angew. Chem. Int. Ed.* **2003**, *42*, 5973–5977.
[8] R. Luo, H. S. R Gilson, M. J. Potter, M. K. Gilson, *Biophys. J.* **2001**, *80*, 140–148.
[9] C. R. Calladine, *J. Mol. Biol.* **1982**, *161*, 343–352.
[10] K. Yanagi, G. G. Prive, R. E. Dickerson, *J. Mol. Biol.* **1991**, *217*, 201–214.
[11] A. A. Gorin, V. B. Zhurkin, W. K. Olson, *J. Mol. Biol.* **1995**, *247*, 34–48.
[12] M. Suzuki, N. Amano, J. Kakinuma, M. Tateno, *J. Mol. Biol.* **1997**, *274*, 421–435.
[13] J. Šponer, J. Kypr, *J. Biomol. Struct. Dyn.* **1993**, *11*, 277–292.
[14] J. Šponer, J. Kypr, *J. Biomol. Struct. Dyn.* **1993**, *11*, 27–41.
[15] a) C. S. Tung, S. C. Harvey, *J. Biol. Chem.* **1986**, *261*, 3700–3709; b) A. R. Srinivasan, R. Torres, W. Clark, W. K. Olson, *J. Biomol. Struct. Dyn.* **1987**, *5*, 459; c) D. Bhattacharyya, M. Bansal, *J. Biomol. Struct. Dyn.* **1990**, *8*, 539–572.
[16] V. B. Zhurkin, V. I. Poltev, V. L. Florentev, *Mol. Biol.* **1980**, *14*, 882–895.
[17] V. I. Poltev, N. V. Shulyupina, *J. Biomol. Struct. Dyn.* **1986**, *3*, 739–765.
[18] P. Hobza, J. Šponer, *Chem. Rev.* **1999**, *99*, 3247–3276.
[19] J. Šponer, H. A. Gabb, J. Leszczynski, P. Hobza, *Biophys. J.* **1997**, *73*, 76–87.
[20] J. Šponer, J. Florian, H. L. Ng, J. E. Šponer, N. Špačková, *Nucleic Acids Res.* **2000**, *28*, 4893–4902.
[21] a) J. Florian, J. Šponer, A. Warshel, *J. Phys. Chem.* **1999**, *103*, 884–892; b) R. Amutha, V. Subramanian, B. U. Nair, *Theor. Chem. Acc.* **2002**, *107*, 343–350; c) L. X. Dang, P. A. Kollman, *J. Am. Chem. Soc.* **1990**, *112*, 503–507; d) J. Norberg, L. Nilsson, *Biophys. J.* **1998**, *74*, 394–402; e) A. Pohorille, S. K. Burt, R. D. Macelroy, *J. Am. Chem. Soc.* **1984**, *106*, 402–409; f) R. A. Friedman, B. Honig, *Biophys. J.* **1995**, *69*, 1528–1535.
[22] J. Šponer, J. Leszczynski, P. Hobza, *Biopolymers* **2001**, *61*, 3–31.
[23] J. Šponer, I. Berger, N. Špačková, J. Leszczynski, P. Hobza, *J. Biomol. Struct. Dyn.* **2000**, *Special Iss. S2*, 383–407.
[24] a) E. Seibert, J. B. A. Ross, R. Osman, *J. Mol. Biol.* **2003**, *330*, 687–703; b) P. Varnai, M. Canalia, J. L. Leroy, *J. Am. Chem. Soc.* **2004**, *126*, 14659–14667; c) N. K. Banavali, A. D. MacKerell, *J. Mol. Biol.* **2002**, *319*, 141–160.
[25] Y. Zhao, D. G. Truhlar, *Phys. Chem. Chem. Phys.* **2005**, *7*, 2701–2705.
[26] J. Šponer, J. Leszczynski, P. Hobza, *J. Phys. Chem.* **1996**, *100*, 5590–5596.
[27] L. M. J. Kroon-Batenburg, F. B. VanDuijneveldt, *Theochem J. Mol. Struct.* **1985**, *22*, 185–199.
[28] P. Mignon, S. Loverix, J. Steyaert, P. Geerlings, *Nucleic Acids Res.* **2005**, *33*, 1779–1789.
[29] D. A. Bondarev, W. J. Skawinski, C. A. Venanzi, *J. Phys. Chem. B* **2000**, *104*, 815–822.
[30] D. Řeha, M. Kabeláč, F. Ryjáček, J. Šponer, J. E. Šponer, M. Elstner, S. Suhai, P. Hobza, *J. Am. Chem. Soc.* **2002**, *124*, 3366–3376.
[31] R. Wintjens, C. Biot, M. Rooman, J. Lievin, *J. Phys. Chem. A* **2003**, *107*, 6249–6258.
[32] C. Biot, E. Buisine, J. M. Kwasigroch, R. Wintjens, M. Rooman, *J. Biol. Chem.* **2002**, *277*, 40816–40822.
[33] R. Wintjens, J. Lievin, M. Rooman, E. Buisine, *J. Mol. Biol.* **2000**, *302*, 395–410.
[34] F. L. Gervasio, R. Chelli, P. Procacci, V. Schettino, *Protein J.* **2002**, *48*, 117–125.
[35] F. L. Gervasio, P. Procacci, G. Cardini, A. Guarna, A. Giolitti, V. Schettino, *J. Phys. Chem. B* **2000**, *104*, 1108–1114.
[36] M. Dracinsky, O. Castano, *Phys. Chem. Chem. Phys.* **2004**, *6*, 1799–1805.
[37] M. L. Leininger, I. M. B. Nielsen, M. E. Colvin, C. L. Janssen, *J. Phys. Chem. A* **2002**, *106*, 3850–3854.
[38] P. Hobza, J. Šponer, *J. Am. Chem. Soc.* **2002**, *124*, 11802–11808.
[39] P. Jurečka, P. Hobza, *J. Am. Chem. Soc.* **2003**, *125*, 15608–15613.
[40] P. Jurečka, J. Šponer, P. Hobza, *J. Phys. Chem. B* **2004**, *108*, 5466–5471.
[41] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A. K. Wilson, *Chem. Phys. Lett.* **1998**, *286*, 243–252.
[42] D. G. Truhlar, *Chem. Phys. Lett.* **1998**, *294*, 45–48.
[43] M. Feyereisen, G. Fitzgerald, A. Komornicki, *Chem. Phys. Lett.* **1993**, *208*, 359–363.
[44] O. Vahtras, J. Almlöf, M. Feyereisen, *Chem. Phys. Lett.* **1993**, *213*, 514–518.
[45] D. E. Bernholdt, R. J. Harrison, *Chem. Phys. Lett.* **1996**, *250*, 47–58.

[46] P. Jurečka, P. Nachtigall, P. Hobza, *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578–4582.
[47] a) J. Šponer, P. Hobza, *Chem. Phys. Lett.* **1997**, *267*, 263–270; b) M. O. Sinnokrot, E. F. Valeev, C. D. Sherrill, *J. Am. Chem. Soc.* **2002**, *124*, 10887–10893; c) M. O. Sinnokrot, C. D. Sherrill, *J. Phys. Chem. A* **2003**, *107*, 8377–8379; d) S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami, K. Tanabe, *J. Am. Chem. Soc.* **2002**, *124*, 104–112.
[48] X. J. Lu, W. K. Olson, *J. Mol. Biol.* **1999**, *285*, 1563–1575.
[49] W. K. Olson, M. Bansal, S. K. Burley, R. E. Dickerson, M. Gerstein, S. C. Harvey, U. Heinemann, X. J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, C. S. Tung, E. Westhof, C. Wolberger, H. M. Berman, *J. Mol. Biol.* **2001**, *313*, 229–237.
[50] S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553–579.
[51] J. Šponer, P. Jurečka, P. Hobza, *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151.
[52] A. Pérez, J. Šponer, P. Jurečka, P. Hobza, F. J. Luque, M. Orozco, *Chem. Eur. J.* **2005**, *11*, 5062–5066.
[53] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
[54] S. Warmlander, J. E. Šponer, J. Šponer, M. Leijon, *J. Biol. Chem.* **2002**, *277*, 28491–28497.
[55] M. Bachs, F. J. Luque, M. Orozco, *J. Comput. Chem.* **1994**, *15*, 446–454.
[56] F. J. Luque, C. Curutchet, J. Muñoz-Muriedas, A. Bidon-Chanal, I. Soteras, A. Morreale, J. L. Gelpí, M. Orozco, *Phys. Chem. Chem. Phys.* **2003**, *5*, 3827–3836.
[57] I. Soteras, C. Curutchet, A. Bidon-Chanal, M. Orozco, F. J. Luque, *J. Mol. Struct.* **2005**, *727*, 29–40.
[58] E. Cancès, B. Mennucci, *J. Math. Chem.* **1998**, *23*, 309–326.
[59] E. Cancès, B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *107*, 3032–3041.
[60] B. Mennucci, E. Cancès, J. Tomasi, *J. Phys. Chem. B* **1997**, *101*, 10506–10517.
[61] F. J. Luque, Y. Zhang, C. Aleman, M. Bachs, J. Gao, M. Orozco, *J. Phys. Chem.* **1996**, *100*, 4269–4276.
[62] F. J. Luque, C. Alemán, M. Bachs, M. Orozco, *J. Comput. Chem.* **1996**, *17*, 806–820.
[63] C. Curutchet, M. Orozco, F. J. Luque, *J. Comput. Chem.* **2001**, *22*, 1180–1193.
[64] Gaussian 03 (Revision B.04), M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. Jr., Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian, Inc., Pittsburgh PA, **2003**.
[65] A. Perez, A. Noy, F. Lankas, F. J. Luque, M. Orozco, *Nucleic Acids Res.* **2004**, *32*, 6144–6151.
[66] H. S. Koo, H. M. Wu, D. M. Crothers, *Nature* **1986**, *320*, 501–506.
[67] H. C. M. Nelson, J. T. Finch, B. F. Luisi, A. Klug, *Nature* **1987**, *330*, 221–226.
[68] A. D. DiGabriele, M. R. Sanderson, T. A. Steitz, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 1816–1820.
[69] D. M. Crothers, T. E. Haran, J. G. Nadeau, *J. Biol. Chem.* **1990**, *265*, 7093–7096.
[70] H. S. Koo, J. Drak, J. A. Rice, D. M. Crothers, *Biochemistry* **1990**, *29*, 4227–4234.
[71] W. K. Olson, V. B. Zhurkin *in Biological Structure and Dynamics, Proc. Ninth Conversation, Vol. 2* (Eds.: R. H. Sarma, M. H. Sarma), Adenine Press, Albany, NY, **1996**, pp. 341–368.
[72] M. A. Young, D. L. Beveridge, *J. Mol. Biol.* **1998**, *281*, 675–687.
[73] F. Lankaš, T. E. Cheatham, N. Špačková, P. Hobza, J. Langowski, J. Šponer, *Biophys. J.* **2002**, *82*, 2592–2609.
[74] J. Šponer, J. Kypr, *J. Biomol. Struct. Dyn.* **1990**, *7*, 1211–1220.
[75] J. Šponer, J. Leszczynski, P. Hobza, *J. Comput. Chem.* **1996**, *17*, 841–850.

Received: October 8, 2005
Published online: January 19, 2006

