JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

Quantum-Mechanical Analysis of the Energetic Contributions to π Stacking in Nucleic Acids versus Rise, Twist, and Slide

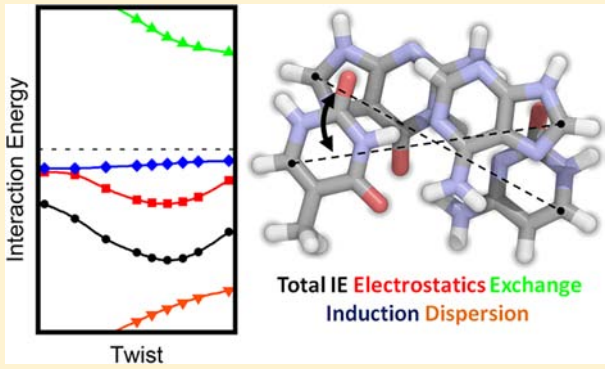
Trent M. Parker,[†] Edward G. Hohenstein,[†] Robert M. Parrish,[†] Nicholas V. Hud,[‡] and C. David Sherrill*[,†]

[†]Center for Computational Molecular Science and Technology, School of Chemistry and Biochemistry, and School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0400, United States

[‡]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332-0400, United States

Supporting Information

ABSTRACT: Symmetry-adapted perturbation theory (SAPT) is applied to pairs of hydrogen-bonded nucleobases to obtain the energetic components of base stacking (electrostatic, exchange-repulsion, induction/polarization, and London dispersion interactions) and how they vary as a function of the helical parameters Rise, Twist, and Slide. Computed average values of Rise and Twist agree well with experimental data for B-form DNA from the Nucleic Acids Database, even though the model computations omitted the backbone atoms (suggesting that the backbone in B-form DNA is compatible with having the bases adopt their ideal stacking geometries). London dispersion forces are the most important attractive component in base stacking, followed by electrostatic interactions. At values of Rise typical of those in DNA (3.36 Å), the electrostatic contribution is nearly always attractive, providing further evidence for the importance of charge-penetration effects in π−π interactions (a term neglected in classical force fields). Comparison of the computed stacking energies with those from model complexes made of the "parent" nucleobases purine and 2-pyrimidone indicates that chemical substituents in DNA and RNA account for 20−40% of the base-stacking energy. A lack of correspondence between the SAPT results and experiment for Slide in RNA base-pair steps suggests that the backbone plays a larger role in determining stacking geometries in RNA than in B-form DNA. In comparisons of base-pair steps with thymine versus uracil, the thymine methyl group tends to enhance the strength of the stacking interaction through a combination of dispersion and electrosatic interactions.



INTRODUCTION

Non-covalent forces govern self-assembly, molecular recognition, protein folding, nucleic acid structure, drug binding, intercalation, and a wide variety of other biochemical phenomena.[1−5] A thorough understanding of non-covalent forces and their effects on biomolecular structure and function would aid the progress of drug design, biotechnology, and nanotechnology. Experimental studies of the intrinsic strengths and geometrical preferences of non-covalent forces are hindered by complications arising from environmental effects, steric constraints, and competing non-covalent interactions. Even the non-covalent forces that govern the structure of a DNA duplex are not understood in detail. Limited gas-phase experimental studies have measured interaction energies between nucleic acid bases, but such studies have not succeeded in the simultaneous determination of both geometry and energetics.[6,7] Quantum-mechanical studies can avoid such difficulties and are capable of providing insight into the strength, geometrical dependence, and physical origin of molecular interactions.[8,9]

DNA and RNA duplexes are primarily stabilized by the forces of Watson−Crick hydrogen bonding and base stacking (both interstrand and intrastrand).[4,10,11] These interactions are similar in magnitude, and both are essential for the stability of a nucleic acid duplex in aqueous solution.[11−13] Hydrogen bonding is primarily electrostatic in nature (although it also has a non-negligible dispersion component),[14] and it can be well-described by second-order Møller−Plesset perturbation theory (MP2).[15] Accordingly, Šponer, Jurečka, and Hobza have reported very accurate benchmarks for hydrogen-bonding energies in a wide variety of nucleic acid base pairs, including both nucleobases found in natural nucleic acids and ones with nonstandard substituents.[16] Base stacking is harder to describe theoretically,[15,17] and very reliable results appear to require the use of highly correlated methods such as coupled-cluster theory through perturbative triples [CCSD(T)].[18] Benchmark interaction energies for the 10 DNA base-pair steps at their gas-phase minima with respect to the helical parameters Rise (see Figure 1), Twist (see Figure 1), and Propeller Twist have been computed via estimates of the CCSD(T) complete basis set (CBS) limit.[19] Also, Cooper et al. used van der Waals density functional theory (vdW-DFT) to generate full potential energy surfaces with respect to Twist for all DNA and RNA base-pair

Received: June 28, 2012
Published: December 24, 2012

   dx.doi.org/10.1021/ja3063309 | J. Am. Chem. Soc. 2013, 135, 1306−1316

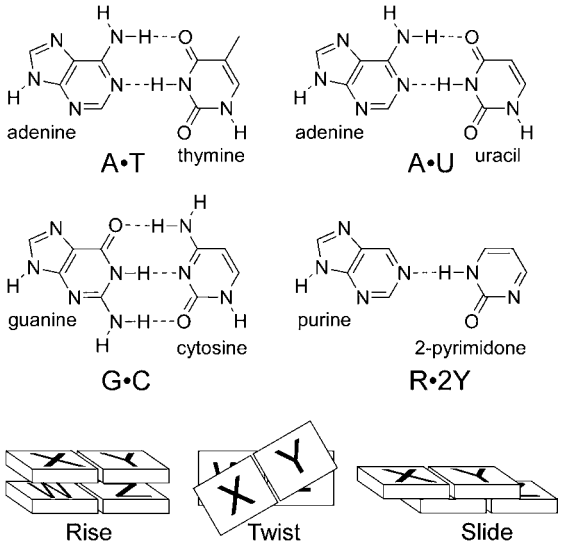


Figure 1. Natural and parent nucleobases considered in the present study and illustrations of the helical parameters Rise, Twist, and Slide.

steps.[20] Their finding of an average Twist angle of 34 ± 10° at the energetic minima is in accord with that of B-DNA as determined from multiple analyses of X-ray crystal structures deposited in the Nucleic Acids Database[21] (36 ± 7°).[22−26]

Considerably more detailed insight into the nature of π-stacking interactions in DNA and RNA can be provided by symmetry-adapted perturbation theory (SAPT),[27] which (unlike the methods mentioned above) can dissect base-stacking interactions into contributions from electrostatic, exchange-repulsion, induction/polarization, and London dispersion forces (attractive van der Waals terms) in a rigorous way. Unfortunately, the computational expense of SAPT has until recently prevented its widespread application beyond small complexes, motivating significant research on increasing its computational efficiency. One popular approach is to utilize DFT to describe monomer properties.[28−32] Alternatively, our group has recently implemented density fitting to allow efficient wave-function-based SAPT computations (DF-SAPT).[33−35] SAPT has previously been applied to hydrogen-bonded and stacked nucleobases,[36−38] but the computations have been limited to fixed gas-phase or average crystal geometries.

In this work, using our new DF-SAPT code, we performed the first energy-component analysis of stacking interactions for all DNA and RNA base-pair steps and for two related nucleobases as functions of the helical parameters Rise, Twist, and Slide (see Figure 1). Having these energy decompositions as functions of base-step geometry is crucial because previous computational studies[39] have revealed that dynamic fluctuations in base-stacking geometry can change computed base-stacking energies by as much as 3 kcal mol^{-1}. Information regarding the magnitudes of base-pair stacking interactions as functions of geometry should provide deeper insight into the fundamental forces that govern DNA and RNA equilibrium and dynamical structures and the origin of energetic penalties associated with helix unwinding and nucleobase destacking transitions that are associated with protein and drug binding events (e.g., intercalation).

In addition, through energy-component analysis we sought to quantify and categorize the contributions of the purine and pyrimidine aromatic rings and nucleobase substituents to the total stacking energy in DNA/RNA base-pair steps. Presently, even the origin of duplex stabilization by the thymine methyl group is not well understood. Theory and experiment agree that the presence of the methyl group at the C5 position of thymine contributes to the stability of a nucleic acid duplex.[20,40−45] However, there are conflicting results regarding the origin of the increased stability observed for base pairs containing thymine versus uracil.[46−49] A number of factors have been proposed as the origin of the increased thermal stability of the RNA duplex with respect to a DNA duplex of analogous sequence, including more interstrand base-pair stacking (associated with nonzero Slide), stronger base stacking, stronger base-pair hydrogen bonding, and greater preorganization of the RNA backbone (due to more restricted sugar pucker and 2-OH hydrogen bonding).[39,46,50,51] Here we examined differences between the DNA and RNA nucleobase stacking energy components. To understand more fully how substituents tune base-stacking energetics, we also considered base-pair steps formed from base pairs in which purine is hydrogen-bonded in the Watson−Crick geometry to 2-pyrimidone, a model base pair with one hydrogen bond, which we will denote as R·2Y (see Figure 1).

In this study, we focused on understanding the energetic consequences of varying the base-pair step parameters Twist and Rise (see Figure 1) using base pairs with idealized pairing geometries (planar base pairs with optimal hydrogen-bond lengths). Twist and Rise are the most significant base-pair step parameters that define the B-form helix of duplex DNA, the helical structure assumed by DNA under physiological conditions. X-ray crystal structures have revealed that the base-pair steps within a B-form helix have Rise values near 3.3 Å, and it is anticipated that the energetic components of base stacking (particularly dispersion) should vary strongly with Rise. The average Twist angle for B-form DNA, also derived from crystal structure analysis, is around 36° per base-pair step.[25,26] However, it is less obvious how the energetic components of stacking should vary with Twist and how sequence-specific differences in overall stacking energy as a function of Twist contribute to local variations in helix structure. The A-form helix, which is assumed by duplex RNA, by RNA−DNA hybrid duplexes, and by DNA under dehydrating conditions, has average values of Rise and Twist similar to those for B-form DNA. However, the A-form helix, unlike the B-form helix, has an appreciable nonzero value for the helical parameter Slide.[25] Thus, Slide is arguably the most distinguishing helical feature between B-form and A-form base-pair steps, and it is one of the only base-pair step parameters (besides Rise and Twist) with an average value that deviates significantly from zero (0.26 Å for the B-form helix and −1.5 Å for the A-form helix).[25,26] To determine the effect of Slide on base-pair stacking energetics, our analysis included variation of Slide at the most favored values of Twist and Rise for each base-pair step.

The conformational space explored in the current study includes, to a good approximation, the base-pair geometries of canonical duplex nucleic acids with nearly ideal Watson−Crick geometries. As in a previous study[20] we observed a good correlation between the energetically favored average geometries and the experimentally derived average helical parameters for DNA, although there are varying degrees of agreement between the theoretical and experimental values for individual base steps. By calculating the energy of base steps as a function of Slide, we were able to examine, to a first-order approximation, the differential energetic contributions of base-

dx.doi.org/10.1021/ja3063309 | J. Am. Chem. Soc. 2013, 135, 1306−1316

pair stacking interactions to the relative stabilities of RNA and DNA helices. A more comprehensive exploration of conformational space will be required in order to confirm the conclusions drawn from the current study regarding the energetics of RNA base-pair stacking, as RNA base-pair steps have appreciable Roll, which is beyond the scope of the current study (which already includes over 900 dinucleotide step computations). Nevertheless, our inclusion of uracil as well as noncanonical nucleobases provides some important insights regarding the stacking energetics of RNA nucleobases and the relative contributions of base-pair stacking energies in defining the helical structure of RNA. In particular, our computations indicate that base-pair stacking energies play a more significant role in defining the local B-form DNA helix than the A-form RNA helix. DF-SAPT has allowed us to provide a systematic energy-component analysis for this most fundamental interaction of biology.

■ THEORETICAL METHODS

Here we applied DF-SAPT to stacked base pairs to obtain interaction energies in terms of their electrostatic, exchange, induction/polarization, and London dispersion components.[27] Each hydrogen-bonded base pair is treated as a monomer in SAPT, as our goal is to provide a better understanding of base stacking between these hydrogen-bonded base pairs. This means that our computations do not explicitly determine energy components for the attraction between one base and its hydrogen-bonded partner, although the interaction is included in the Hartree−Fock computations on each base pair as part of the SAPT procedure. Likewise, this approach is unable to separate out three-body induction or other many-body terms explicitly (from the point of view in which each base is a separate monomer). However, we are unaware of a method capable of performing explicit four-body energy-decomposition analysis, and even a study of three-body effects would require careful selection of a model that would be computationally tractable while remaining reasonably accurate. Thus, while a study including at least three-body effects would certainly be more complete, we nevertheless believe that the two-body SAPT description of interactions between base pairs remains very informative.

Because of the size of these systems and the number of computations required, the present SAPT computations did not include the effect of intramonomer electron correlation (i.e., we employed the SAPT0 approximation).[27,34] When paired with a truncated aug-cc-pVDZ basis set[52,53] that neglects diffuse functions on H atoms and diffuse d functions on other atoms, this level of theory (SAPT0/jun-cc-pVDZ, later more simply called SAPT0/jaDZ) is known to obtain reasonably accurate stacking energies.[33,34] A complete description of the SAPT0 computations and the grouping of energy components has been provided previously.[33] The SAPT computations were performed with our DF-SAPT program developed within the framework of PSI.[35,54,55]

The geometry of each base pair was optimized at the B3LYP-D/aug-cc-pVDZ level of theory.[56,57] These geometry optimizations were performed with Q-Chem 3.2.[58] The Twist and Rise base-pair step geometrical parameters were scanned from 0 to 60° in 5° or 10° increments and from 3.1 to 3.7 Å in 0.2 Å increments, respectively, with fixed planar geometries. Rise and Twist are defined according to ref 25 and the Nucleic Acids Database.[21] Rise is the separation between the planes of the two base pairs, and Twist is the in-plane counterclockwise rotation of the top base pair about the midpoint of the base-pair long axis (pyrimidine C6 to purine C8) (see Figure 1).[25,59]

After the minimum-energy Rise and Twist of each base-pair step were found, the Slide parameter was scanned from −3.0 to +3.0 Å in 0.25 Å increments at the fixed values of Rise and Twist specific to each step. Slide is the horizontal displacement of the top base pair along the base-pair long axis. We do not discount the possibility that other base-pair step geometrical parameters may possibly have a significant role in the energetics of base stacking, but these three (Rise, Twist, and Slide) vary the most in known crystal structures of nucleic acids. A limited scan of what is most likely the next most important geometry parameter, Roll, did not provide much additional energetic stabilization (see Figure S22 in the Supporting Information).

Distributed multipole analysis (DMA)[60] computations were performed with Molpro[61] at the Hartree−Fock/6-311G** level of theory to generate atom-centered multipoles, including charges, dipoles, quadrupoles, octopoles, hexadecapoles, and 32-poles. The interactions between these multipoles were calculated with the inclusion of all terms depending on R^{-n} with $n \leq 6$ (up to 32-pole−charge, hexadecapole−dipole, and octopole−quadrupole) using an in-house program. This method of acquiring the long-range electrostatic interaction is both converged with respect to multipole order (as shown in Figures S12−S15 in the Supporting Information) and stable and converged with respect to basis-set size (as shown in Figures S16−S19 in the Supporting Information). The difference between DMA electrostatics and SAPT electrostatics (which includes the diffuse nature of electron density) was used here as an estimate of "charge penetration," an electrostatic contribution due to orbital overlap whose importance to π stacking has recently been revealed as a favorable term in all π-stacking systems studied to date.[62]

The notation X·Y indicates an X base hydrogen-bonded to a Y base. WX:YZ indicates an X·Y base pair stacked on top of a W·Z base pair, with the former rotated counterclockwise relative to the latter. For the four possible nucleobases in DNA or RNA, there are 16 possible dinucleotide steps in this notation, but because of symmetry (i.e., WX:YZ = YZ:WX), there are only 10 unique pairs. Three of the base-pair steps are identical for DNA and RNA (CG:GC, GC:GC, and GG:CC), giving 17 unique base-pair steps when both DNA and RNA are considered (Figure 1).

■ RESULTS AND DISCUSSION

Reliability of the Theoretical Method. Before analysis of the data, it was useful to verify the reliability of our methods by comparison to high-quality literature data where available. Šponer et al.[19] attempted to obtain CCSD(T)/CBS estimates for base-pair steps by summing the four individual base−base interactions and then applying three- and four-body corrections computed at the MP2/aug-cc-pVDZ level of theory. The base−base interactions were evaluated by MP2/aug-cc-pVDZ/aug-cc-pVTZ extrapolations plus four two-body ΔCCSD(T) corrections [CCSD(T)-MP2] obtained with the small 6-31G*(0.25) basis set. The 6-31G*(0.25) basis set may be too small for a high-quality ΔCCSD(T) correction.[63−66] More recently, Hill and Platts[67] dispensed with the four-body correction by computing the interaction between the two base pairs directly, and they used larger basis sets (aug-cc-pVTZ and aug-cc-pVQZ) in their MP2 extrapolations. However, they were unable to improve upon the ΔCCSD(T) component of the energies, and they had to employ local approximations to make their larger MP2 computations possible. The local approximation appears to be reliable for non-covalent interactions as long as basis sets larger than aug-cc-pVDZ are used,[68] but Hill and Platts were concerned[67] that it could lead to errors as large as 1 kcal mol^{-1} for these systems because of the limitations of having only two-body (and not three- or four-body) ΔCCSD(T) corrections. Thus, the best available CCSD(T)/CBS estimates for base-pair steps are useful but not definitive.

Table 1 presents interaction energies at the equilibrium geometries (with respect to Rise and Twist) of the base-pair steps considered. Here we systematically varied only the base-pair Rise and Twist parameters, whereas the geometries of Šponer et al. also optimized Propeller Twist (providing additional stabilization of up to 1.5 kcal mol^{-1}).[19] In view of this difference and the approximations made in the CCSD(T)/CBS estimates, the present SAPT0/jaDZ interaction energies

Table 1. DNA, RNA, and Purine·2-Pyrimidone (R·2Y) Base-Pair Step Interaction Energies (kcal mol^{-1})

step	CCSD(T)/CBS[a]	CCSD(T)/CBS[b]	SAPT0/jaDZ
CG:CG	−17.3	−16.16	−15.69
GC:GC	−15.4	−14.70	−14.48
CA:TG	−15.1	−13.96	−13.63
AA:TT	−13.1	−11.98	−12.10
TA:TA	−12.8		−11.92
AC:GT	−13.4	−12.01	−11.29
AG:CT	−13.5	−12.39	−11.20
AT:AT	−13.3	−11.99	−10.87
GA:TC	−12.9	−11.26	−10.22
GG:CC	−11.5	−10.29	−9.32
CA:UG			−13.56
AA:UU			−11.92
UA:UA			−11.78
AG:CU			−11.61
AC:GU			−10.43
GA:UC			−9.82
AU:AU			−9.33
2YR:2YR			−9.56
R2Y:R2Y			−8.75
RR:2Y2Y			−7.28

[a]Data from ref 19, as reported in ref 67, which claims to correct some summation errors. Individual base-pair interactions were estimated by MP2/aug-cc-pVDZ/aug-cc-pVTZ extrapolations plus ΔCCSD(T)/6-31G*(0.25) corrections; the sum of the pairwise interactions was corrected with a four-body MP2/aug-cc-pVDZ correction. [b]Data from ref 67; values were estimated from an aug-cc-pVTZ/aug-cc-pVQZ extrapolation of DF-LMP2 energies plus the ΔCCSD(T)/6-31G*(0.25) corrections from ref 19.

compare favorably to the best literature results: they are within 7−21% of the values of Šponer et al.[19] and within 1−10% of the more recent values of Hill and Platts.[67] The SAPT0/jaDZ values nearly always show lower binding energy than the CCSD(T)/CBS estimates, which might be anticipated given the extra stabilization resulting from varying Propeller Twist in previous studies. Hence, the present SAPT0/jaDZ data appear to be reliable for these systems. This level of theory also performed well[34] for the S22 test set of non-covalent interactions,[69] for which higher-quality benchmark data are available.[70]

Substituent Effects. Table 1 also shows the minima for the three base-pair steps of our purine·2-pyrimidone model systems (which we denote as R·2Y, R = purine, Y = pyrimidine, 2Y = 2-pyrimidone). 2-Pyrimidone contains the same C2 carbonyl substituent that is present in the three purine bases cytosine, thymine, and uracil (see Figure 1). However, 2-pyrimidone lacks the C4 carbonyl group found in thymine and uracil, the C5 methyl group in thymine, and the amino group in cytosine. Similarly, purine lacks the C6 amino substituent of adenine and the carbonyl and amino substituents of guanine (Figure 1). We observed a stabilization of 2.0−6.1 kcal mol^{-1} for the natural base-pair steps relative to our model "parent" R·2Y base-pair steps, with the natural bases' $-NH_2$, $=O$, and $-CH_3$ substituents (relative to purine and 2-pyrimidone) accounting for 20−40% of their stacking energy at their respetive Rise/Twist global minima. This result is again in accord with the prediction that all substituents exhibit a stabilizing effect on $\pi-\pi$ interactions.[62,71,72] Such significant contributions to the stacking energy from these exocyclic substituents seems rather surprising. However, experiments involving artificial nucleosides have suggested that substituents might provide such considerable contributions to the stacking energies of the natural bases. For example, Sun and McLaughlin[45] compared the thermodynamic stabilities of A·T and A·m^{3}2P base pairs, where m^{3}2P is a base-pairing analogue of thymine that lacks the C4 carbonyl group. Their study indicated that the removal of this *non-hydrogen-bonding* carbonyl group of thymine by substitution of m^{3}2P reduced the enthalpy contribution to the free energy of duplex stability ($\Delta\Delta H°$) by 12 kcal mol^{-1} per substitution relative to A·T base pairs. As noted by Sun and McLaughlin, the removal of the non-hydrogen-bonding carbonyl group of T may also result in the removal of favorable solvent interactions with the minor groove. Moreover, m^{3}2P is actually a *pyridine* nucleobase that is connected to the backbone through a carbon atom (a C-nucleoside), an additional difference from T that could alter the energetics of base pairing and base stacking beyond the loss of carbonyl stacking interactions.

Energy Components. Figure 2 presents plots showing the variation of the SAPT0 energy components as functions of Twist for all of the base-pair steps analyzed in the present study. In our scans of the total SAPT0 energy versus Twist and Rise, the minimum-energy configurations were found at a Rise of 3.3 Å for every base-pair step considered, so energy versus Twist plots for this fixed value of Rise are presented in Figure 2. Plots of energy versus Twist for other values of Rise are provided in the Supporting Information. The London dispersion and exchange-repulsion terms are much larger in magnitude than the electrostatic and induction terms, so plotting all four terms in the figures would make it hard to distinguish the variation of electrostatics and induction with respect to Twist. Fortunately, the dispersion and exchange terms are of opposite sign and are usually similar in magnitude for stacked $\pi-\pi$ interactions, and thus, we have plotted their sum ("net dispersion") instead of the individual components in the figures. Separate plots of the exchange and dispersion terms are given in Figures S4−S7 in the Supporting Information. The dispersion term is always attractive and decays smoothly with increasing Twist and increasing Rise, as expected from its R^{-6} dependence. The exchange-repulsion term increases rapidly when atoms come into close contact through local steric clashes at low values of Rise or low values of Twist.

The SAPT0 components show clear general trends across the base-pair steps. The induction component between the hydrogen-bonded base pairs, reflecting dipole/induced-dipole and similar polarization contributions, is significantly attractive and varies in magnitude between ∼1 and 4 kcal mol^{-1} across all base-pair steps. It changes the least of all terms with increasing Twist and decays slowly and smoothly toward zero with respect to Rise. Even at 3.1 Å, the lowest value of Rise considered in the present study, induction changes by 2 kcal mol^{-1} or less with respect to Twist. At larger values of Rise, induction changes by 1 kcal mol^{-1} or less with respect to Twist for all steps except AA:TT. However, even in that case, the changes in induction are minor compared with the changes in electrostatics and net dispersion.

The electrostatic term is strongly attractive at low Rise and highly variable with Twist, showing a single well-defined minimum versus Twist at low Rise that is retained at large Rise for YR:YR steps, decays significantly for RR:YY steps, and nearly disappears for RY:RY steps. At large Rise, the electrostatic term generally dominates the shape of the plot of interaction energy versus Twist. Electrostatics account for

dx.doi.org/10.1021/ja3063309 | J. Am. Chem. Soc. 2013, 135, 1306−1316

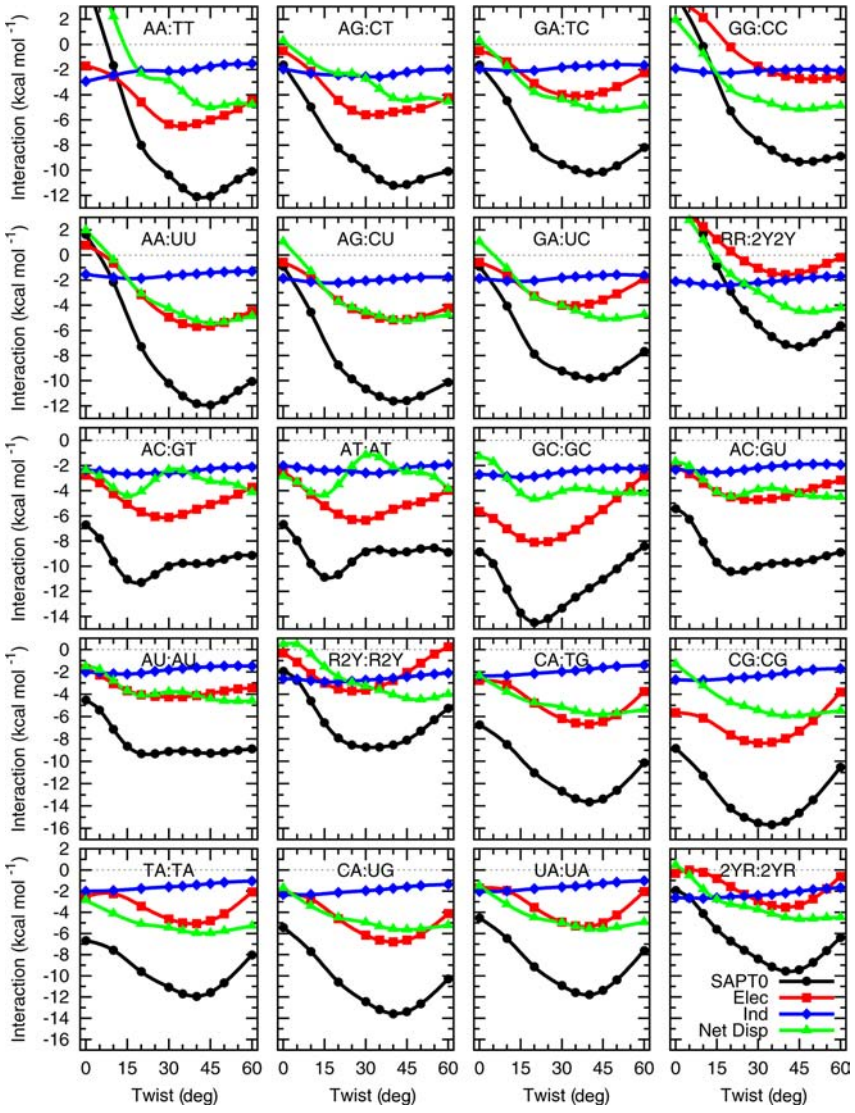


Figure 2. SAPT interaction energy components as functions of Twist at Rise = 3.3 Å for all DNA, RNA, and purine·2-pyrimidone (R·2Y) base-pair steps. Global minima with respect to Rise and Twist were found at this value of Rise for all base-pair steps.

~50% of the magnitude of the total energy change at Rise = 3.3 Å, ~75% at 3.5 Å, and >90% at 3.7 Å. Electrostatics are attractive for all values of Twist at Rise = 3.1 Å. The electrostatic contribution remains significant (−3.4 to +7.2 kcal mol^{-1}) even for the largest value of Rise considered here (3.7 Å). It becomes significantly repulsive at low Twist and large Rise for most base-pair steps and is repulsive at all values of Twist in some cases, such as GG:CC at large Rise.

The uniformly attractive electrostatic interaction between base-pair steps at low Rise can be explained by charge penetration.[62] Due to the compact, positively charged nuclei and the diffuse, negatively charged electron density, weakly interacting dimers of neutral chemical species tend to have attractive electrostatics at short range regardless of the nature of their dipoles, quadrupoles, and so on. Using stacked substituted benzene dimers as simple model systems, we have shown[62] that these charge-penetration effects can be large at intermolecular distances less than 4 Å; the stacking distances in DNA and RNA are certainly within this limit. We can also demonstrate explicitly the importance of charge-penetration effects on the electrostatic interactions between stacked nucleobases. Figure 3 plots the electrostatic energy of atom-centered multipoles

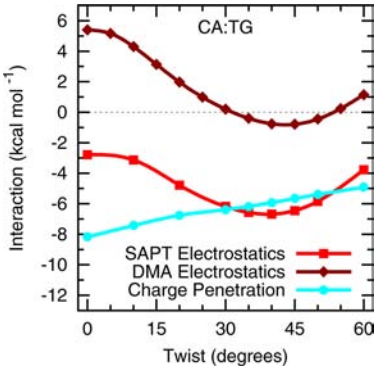


Figure 3. SAPT, DMA, and charge-penetration electrostatic energies as functions of Twist for the CA:TG base pair at Rise = 3.3 Å (plots for all of the base-pair steps are provided in the Supporting Information).

(charges, dipoles, quadrupoles) derived from DMA[60] on each isolated base pair as well as the more rigorous SAPT electrostatic energy (which includes charge penetration) versus Twist for the CA:TG base-pair step at Rise = 3.3 Å. While the

dx.doi.org/10.1021/ja3063309 | J. Am. Chem. Soc. 2013, 135, 1306−1316

multipole−multipole interactions from DMA give a qualitatively correct shape for the curve, the magnitude is completely incorrect because multipole models break down at short range. Electrostatics are attractive by 3−7 kcal mol^{-1} across the range of Twist, whereas DMA predicts the electrostatics of CA:TG to be repulsive by 1−6 kcal mol^{-1}. The remainder of the electrostatic energy is the result of charge penetration from the diffuse π electron clouds in the bases, which accounts for 5−8 kcal mol^{-1} of attraction between the base pairs. This result is typical for all base-pair steps. Plots of DMA and charge-penetration electrostatics for other base-pair steps, along with the demonstrated convergence of the DMA results with respect to multipole order and basis set, are presented in Figures S8−S19 in the Supporting Information. In view of the large magnitude of the charge-penetration energy present in these systems at near-equilibrium geometries, this effect cannot be ignored when modeling stacking interactions. The most popular empirical force fields currently used in macromolecular modeling do not include this term explicitly,[73] even though it is highly important in the interactions of nucleic acids.

Net dispersion becomes repulsive at low Twist, with 0° as a local maximum for all steps. For Rise = 3.1 Å, the exchange term dominates, and net dispersion is highly repulsive. For larger values of Rise, the exchange term quickly decays, and the dispersion term begins to dominate. Net dispersion is the dominant term contributing to the shape of interaction energy versus Twist for Rise < 3.3 Å (see Figure S1 in the Supporting Information). This term contributes ∼85% of the total energy change versus Twist at Rise = 3.1 Å, ∼45% at 3.3 Å, and significantly less for greater values of Rise. The net dispersion versus Twist curves become very flat at 3.5 Å and beyond, changing by less than 2 kcal mol^{-1} on average with respect to Twist.

Twist Dependence of Net Interaction Energies. At Twist = 0° for RR:YY steps, a purine base is stacked directly on top of another purine base, and their paired pyrimidine bases are likewise stacked directly on top of one another, resulting in a highly repulsive exchange term relative to the value at higher Twist. These RR:YY steps feature energetic minima for Twist = 40−50° at all values of Rise. These minima are fairly broad, with the energy increasing slowly for larger values of Twist up to 60°, the largest Twist considered here. All YR:YR steps have deep, well-defined minima around Twist = 30−40° for all values of Rise studied. Compared with the YR:YR steps, the RY:RY steps tend to be somewhat flatter with respect to Twist, even at low Rise. The minima for RY:RY steps occur at smaller Twist angles (15−20°) than for the YR:YR steps.

Table 2 summarizes the average values and standard deviations of the minimum-energy Twist angle for each class of base-pair steps at each value of Rise considered. All of the global minima for Rise and Twist were found at Rise = 3.3 Å, where the DNA base-pair minima have an average Twist of 34 ± 11° (where 11° is the standard deviation). This is in accord with averages from the Nucleic Acids Database (36 ± 7°).[21] Additionally, for every base-pair step, our values of the minimum-energy Twist angle are in excellent agreement (within our resolution of Twist) with the results of Cooper et al.,[20] which were obtained using vdW-DFT, providing additional support for the reliability of vdW-DFT.

The minimum-energy Twist values for six of the DNA steps (AA:TT, AG:CT, CA:TG, CG:CG, GA:TC, and TA:TA) are in good agreement with the experimental ranges[22−24] (see Table S1 in the Supporting Information). The deviations of the calculated minimum-energy Twist values for the AC:GT, AT:AT, and GC:GC base steps are notable, as these base steps have well-defined minima with respect to Twist (the potential energy wells are flatter for the other RY:RY steps). The computed minima for these cases occur for Twist angles of 15−20°, whereas the average experimental Twist angles for these cases are 29−38°. The minimum for GC:GC is the most well-defined, with an interaction energy of −14.5 kcal mol^{-1} at Twist = 20°. This minimum is strongly favored by the electrostatic term, and in this geometry, the amino substituent of one GC base pair is directly stacked above the carbonyl substituent of the other base pair. Because of the C_2 symmetry of the GC:GC base step, two of these interactions are present, forming a very favorable cyclic, four-membered electrostatic contact. It is unclear why the well-defined theoretical minima for this case and the AC:GT and AT:AT cases occur at noticeably smaller Twist angles than the experimental averages, whereas the agreement is better for base-pair steps with flatter potential energy curves. It is conceivable that backbone contributions, solvent interactions, or steric effects are more important for these steps.

Table 2. Optimal Twist Angles (Average ± Standard Deviation) at Various Rise Values for DNA, RNA, and R·2Y Base-Pair Steps

nucleic acid	base-pair steps	3.1 Å	3.3 Å	3.5 Å	3.7 Å
DNA	RR steps	46 ± 3°	41 ± 3°	41 ± 3°	43 ± 5°
	RY steps	18 ± 3°	18 ± 3°	20 ± 0°	20 ± 0°
	YR steps	38 ± 3°	38 ± 3°	38 ± 3°	37 ± 6°
	average	36 ± 13°	34 ± 11°	34 ± 10°	34 ± 11°
RNA	RR steps	46 ± 3°	43 ± 3°	43 ± 3°	44 ± 5°
	RY steps	30 ± 17°	22 ± 3°	23 ± 3°	30 ± 12°
	YR steps	38 ± 3°	38 ± 3°	38 ± 3°	37 ± 6°
	average	39 ± 11°	35 ± 10°	36 ± 9°	38 ± 10°
R·2Y	RR:2Y2Y	45°	45°	45°	45°
	R2Y:R2Y	40°	30°	30°	30°
	2YR:2YR	45°	40°	40°	40°
	average	44 ± 2°	39 ± 7°	39 ± 7°	39 ± 7°

The average Twist angles for RNA base-pair steps are also listed in Table 2. The average of 35 ± 10° at Rise = 3.3 Å is nearly identical to that found for DNA (34 ± 11°). This similarity remains for all values of Rise, with differences between the averages over RNA and DNA base-pair steps of no more than 4°. This suggests that the thymine methyl group does not greatly affect the optimal Twist angle if only Rise and Twist are varied. When particular base steps are compared (see Table S2 in the Supporting Information), the optimal Twist angle for each RNA step is within 5° of the corresponding DNA value except for a 10° variation between AU:AU (25°) and AT:AT (15°).

Table 2 also contains corresponding values for the R·2Y unsubstituted model system (i.e., RR:2Y2Y, R2Y:R2Y, and 2YR:2YR base-pair steps). The overall R·2Y averages are weighted by the number of each type of step in DNA (four RR:YY, three RY:RY, and three YR:YR) to make them more comparable to the DNA averages. The R·2Y average optimal Twist angles vary even less with respect to Rise than those for the DNA and RNA steps. The average Twist of 39 ± 7° at Rise = 3.3 Å is only slightly larger than that of either DNA or RNA steps, and the ranges within one standard deviation show significant overlap. Among three classes of steps (RR:YY,

dx.doi.org/10.1021/ja3063309 | J. Am. Chem. Soc. 2013, 135, 1306−1316

RY:RY, and YR:YR) in DNA and RNA, only the RY:RY steps differ significantly in their optimal Twist angles (15−25° at Rise = 3.3 Å) relative to the parent R2Y:R2Y system (30°).

Variations in Slide. In addition to Rise and Twist, Slide was also considered at the optimum Rise and Twist for each base-pair step. SAPT potential energy curves for Slide parameters ranging from −3.0 to 3.0 Å are presented in the Supporting Information. The effect of varying Slide in this way ranged from no energy change to an extra stabilization of up to 2 kcal mol^{-1}, with most cases stabilized by about 1 kcal mol^{-1}. Such changes are modest relative to the average interaction energy of 12 kcal mol^{-1} for the nucleobase steps, indicating that the general qualitative conclusions drawn from the limited potential scans in this work are relevant for base steps in typical B-DNA geometries.

B-form DNA has an average Slide of +0.26 Å, as determined from crystallographic data.[25,26] Four base-pair steps have an average crystallographic Slide that deviates from this value by more than 0.25 Å, namely, CA:TG (+1.88 Å),[74] AT:AT (−0.57 Å), CG:CG (+0.68 Å), and AA:TT (−0.16 Å) (see Table 3). For these four steps, our computed values for Slide are in good agreement with the experimental values (i.e., a minimum on the SAPT potential energy curve exists within one standard deviation of the corresponding experimental value). For the remaining steps with average Slide values near 0.26 Å, the SAPT potential energy minima and crystallographic averages do not correlate well.

Table 3. Experimental Slide Values (Average ± Standard Deviation)[25,26] from the Nucleic Acids Database (NDB), SAPT0/jaDZ-Computed Minimum-Energy Slide Values, and Energy Penalties for Base Stacking If the NDB Slide Instead of the SAPT0 Slide Is Adopted (ΔE_{slide})

step	Slide (Å)		ΔE_{slide} (kcal mol^{-1})
	NDB	SAPT0	
	B-Form DNA		
AA:TT	−0.16 ± 0.4	+0.00	0.03
AC:GT	+0.06 ± 0.3	−0.50	0.80
AG:CT	+0.34 ± 0.4	+1.50	1.08
AT:AT	−0.57 ± 0.2	−0.50	0.00
CA:TG	+1.88 ± 1.0	+1.25	0.35
CA:TG+[a]	+0.81 ± 0.6	+1.25	0.31
CG:CG	+0.68 ± 0.3	+1.00	0.04
GA:TC	−0.01 ± 0.4	−0.75	0.84
GC:GC	+0.31 ± 0.8	−0.50	2.60
GG:CC	+0.28 ± 0.3	+1.50	1.19
TA:TA	+0.38 ± 0.7	−1.00	1.08
total	+0.26 ± 0.8	+0.20 ± 0.9	0.80 ± 0.75
	A-Form RNA		
AA:UU	−1.27 ± 0.4	+1.25	1.72
AC:GU	−1.43 ± 0.3	−0.50	3.03
AG:CU	−1.50 ± 0.3	+1.75	1.72
AU:AU	−1.36 ± 0.3	−0.50	2.05
CA:UG	−1.46 ± 0.2	+1.25	2.51
CG:CG	−1.89 ± 0.4	+1.00	4.72
GA:UC	−1.70 ± 0.5	−0.75	0.61
GC:GC	−1.39 ± 0.2	−0.50	3.37
GG:CC	−1.78 ± 0.3	+1.50	1.84
UA:UA	−1.45 ± 0.2	−1.00	0.36
total	−1.56 ± 0.3	+0.44 ± 0.9	2.19 ± 1.23

[a]CA:TG base steps from the NDB are bimodal; a few steps have large positive values of slide, resulting in a very over-twisted state (see ref 23). These states are removed from the row labeled CA:TG+.

The A-form RNA double helix has an average Slide of −1.56 Å with much less deviation from the average value than is observed for the B-form DNA double helix. The average experimental value for each base-pair step is within 0.3 Å of the overall average.[25,26] Our data show that the change in Slide to transition from the B-form helix to the A-form helix is energetically uphill for all base-pair steps by 0.5−2.0 kcal mol^{-1} per base step. *None* of the base steps have an energetic minimum within one standard deviation of the Slide values of the A-RNA helix for the given step. Whereas the B-DNA average geometry for Rise, Twist, and Slide can be reasonably well described as being determined by the energetic minimum for base stacking, the canonical A-form helix does not correspond to an energetic minimum for base-pair stacking. Thus, the RNA helix, and DNA in the A-form helix, must be more influenced by other factors. In the case of RNA, greater backbone preorganization (due to the more restricted sugar conformation of the ribose versus the deoxyribose sugar) and 2′-OH hydrogen bonding are possible terms that override the tendency of base pairs to assume their theoretically optimal stacking configuration. In the case of DNA, which converts from the B-form helix to the A-form helix under dehydrating conditions,[75] it is likely that changes in solvation and cation localization provide the energy that pushes the base pairs from their (average) most favorable geometry to that associated with the A-form helix.

Effect of the Thymine Methyl Group on Base Stacking. The energetic minima of base-pair steps containing thymine bases have more stable stacking energy than the corresponding uracil-containing steps by 0.1−1.5 kcal mol^{-1} (1−14%). This result illustrates the stabilizing contribution of the thymine methyl group in base stacking, as expected on the basis of known substituent effects in $\pi-\pi$ interactions.[62,71,72] This effect is largest in RY:RY steps, where the direct interaction of the methyl group with the base pair below provides a stabilization of 0.9 (AC:GT) or 1.5 (AT:AT) kcal mol^{-1}. In RR:YY and YR:YR steps, the effect of the methyl substitutent is smaller (0.1−0.4 kcal mol^{-1}) at their respective energetic minima versus Rise and Twist, where the methyl group does not interact directly with the adjacent nucleobase. These results are in qualitative agreement with the results of experimental studies in which the stability of nucleic acid duplexes has been shown to increase in association with the replacement of a dU nucleoside with a dT nucleoside.[41,43] However, a parceling of free energy gain into enthalpy ($\Delta\Delta H°$) and entropy ($T\Delta\Delta S°$) contributions reveals a more complex and context-dependent thermodynamics for dT and dU substitutions, indicating that solvent effects, which were not considered in the present study, also contribute to the net free energy gain associated with the addition of a methyl group at the 5C position of uracil.[76] For more insight into the possible effect of solvent interactions on nucleic acid structure studied at the molecular mechanics level, see ref 77.

Figure 4 shows the differences between the energy components for DNA (i.e., T-containing) base-pair steps and those for the corresponding RNA (i.e., U-containing) base-pair steps as functions of Twist at Rise = 3.3 Å. Also shown are the differences between the average T-containing and average U-containing base-pair step energy components. The difference in induction terms changes relatively little with respect to Twist, usually by 0−0.5 kcal mol^{-1} (although by closer to 1.5 kcal

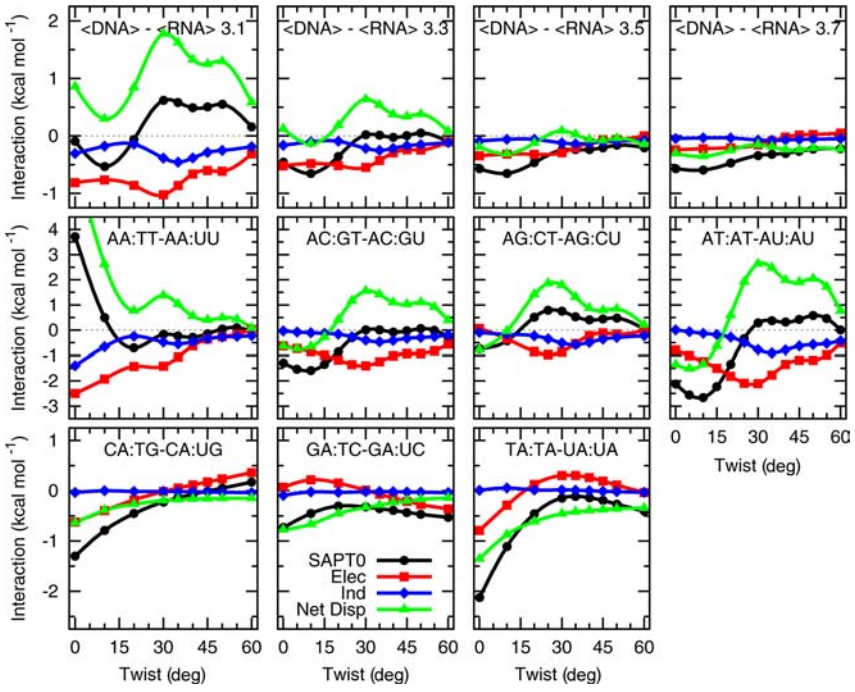


Figure 4. Effect of the thymine methyl group. First row: Differences between average DNA and average RNA base-pair step SAPT energy components versus Twist for Rise = 3.1−3.7 Å. Second and third rows: Differences between DNA and RNA base-pair step SAPT energy components versus Twist at Rise = 3.3 Å for individual base-pair steps.

mol^{-1} for AA:TT − AA:UU), with the induction term being slightly more stabilizing for thymine versus uracil. The change in electrostatics is larger and stabilizes T-containing steps by as much as 2.5 kcal mol^{-1} (AA:TT − AA:UU) or destabilizes them by nearly 0.5 kcal mol^{-1} (CA:TG − CA:UG or TA:TA − UA:UA). However, on average, the thymine leads to electrostatic stabilization. In several cases, there appears to be an anticorrelation between the differential electrostatic contribution and the differential net dispersion contribution, suggesting that these changes in electrostatics are primarily due to changes in charge-penetration terms. The dominant term in the graphs of Figure 4 is the differential net dispersion term, and the differential exchange repulsion is the dominant contributor at low Rise. As Rise increases, the strength of steric interactions with the methyl group decreases quickly, whereas the dispersion term decays much more slowly, resulting in attractive net dispersion contributions by Rise = 3.5 Å with magnitudes of up to 0.3 kcal mol^{-1}.

The AA:TT step contains a direct methyl−methyl interaction at Twist = 0°, which results in highly repulsive net dispersion and total interaction energies at low Twist. For all other steps, at Rise = 3.3 Å, the methyl group in thymine produces stabilization at low values of Twist, by up to 2.7 kcal mol^{-1} in the case of AT:AT. This stabilization is likely due to favorable direct substituent−π interactions[78,79] between the thymine methyl group and the aromatic nucleobase stacked below thymine. The partial positive charge on the methyl hydrogens can interact favorably with the negatively charged π cloud below them, and the methyl group can enhance favorable dispersion interactions. Of course, the methyl group can also lead to local steric clashes for certain base-pair steps at certain values of Twist.

The T − U differences for individual base-pair steps are qualitatively different for the base-pair steps in the second row of Figure 4 (DNA steps AA:TT, AC:GT, AG:CT, and AT:AT) and those in the third row of Figure 4 (DNA steps CA:TG, GA:TC, and TA:TA). For the first group, the thymine methyl group crosses over the edge of the nearest ring of the base below it, passing directly over another atom at Twist ≈ 30° (note the corresponding peak in the net dispersion contribution at this angle). For AT:AT, there are two of these steric clashes, leading to an even larger peak in net dispersion around 30°. Net dispersion continues to be unfavorable until Twist angles of 60° are reached, and it cancels out favorable contributions to give an overall methyl substituent effect near zero for Twist angles greater than 30°. In the second group of base-pair steps, net dispersion is favorable at low Twist angles and smoothly approaches zero at large Twist angles. In these steps, the methyl group rotates away from the base stacked below it, so it is not involved in any local steric clashes.

The overall methyl substituent effect in thymine compared to uracil is repulsive at low Rise, attractive at low Twist (except for AA:TT), and attractive by 0.2−0.6 kcal mol^{-1} in general beyond Rise = 3.3 Å. If we compare T-containing base steps to the corresponding U-containing base steps at their respective minima with respect to Rise and Twist (Table 1), we see that the T-containing steps are generally more stabilized than their U-containing counterparts, in agreement with the experimental finding that replacing thymine with uracil in DNA destabilizes the duplex.[40−42] This comparison assumes that the optimal Twist angles computed here are representative of the Twist angles these base steps would adopt in DNA. As previously discussed, the Twist angles computed for DNA base steps do agree well with experimental average Twist angles. However, we are not aware of experimental structures and thermodynamic measurements featuring T → U substitutions for the same DNA sequence, so we cannot independently validate the appropriateness of our computed Twist angles for the uracil-containing steps for this discussion. An alternative approach

dx.doi.org/10.1021/ja3063309 | J. Am. Chem. Soc. 2013, 135, 1306−1316

would be to evaluate the stabilization due to the thymine methyl group at the average experimental Twist for each step using the experimental DNA values. Such an approach leads to both stabilizations and destabilizations (depending on the step) that average out to about zero (see Table S4 in the Supporting Information). However, this approach fails to account for the fact that the T → U substitution is likely to cause some relaxation in the Twist angle. For the AA:TT step, whether we evaluate the interaction energies at their respective minima or at the average DNA Twist, we obtain a stabilization of 0.18−0.23 kcal mol^{-1}, which compares well to the experimental values of 0.29 or 0.35 kcal mol^{-1} per AA:TT step in experiments on polyA:polyT versus polyA:polyU duplexes.[40,41]

■ CONCLUSIONS

The non-covalent interactions of base pairing and base stacking are essential in defining and maintaining the three-dimensional structure of nucleic acids. In this work, we applied quantum-mechanical energy-component analysis to decompose the base-stacking energies of the unique base-pair steps found in DNA or RNA (17 total) and the base-pair steps in the purine·2-pyrimidone (R·2Y) model system (3 total) into the physically meaningful components of electrostatics, exchange-repulsion, induction, and dispersion as functions of the Rise and Twist duplex parameters. We also performed a more limited analysis of how the energies vary with respect to Slide at the optimal theoretical values of Rise and Twist. Other coordinates, such as Propeller Twist, should be less important and were fixed at zero for computational tractability. The geometries considered are thus close to those exhibited by B-form DNA.

Our SAPT0 values compare favorably to benchmark estimated CCSD(T)/CBS values for energetic minima,[19,67] and they should be useful in testing biomolecular force fields for nucleic acids. Induction (polarization) is significant but relatively flat versus Twist. Net dispersion (exchange plus dispersion) is the dominant changing term versus Twist at low values of Rise (≤3.3 Å) whereas electrostatics is the dominant changing term versus Twist at high values of Rise (>3.3 Å). Dispersion is the dominant attractive term overall, with electrostatics becoming highly attractive at low Rise as a result of charge penetration. Exchange becomes highly repulsive at low Twist and low Rise, where steric clashes occur. Most steps show a single, well-defined energetic minimum with respect to Rise and Twist, with the optimum Rise (3.3 Å for all steps) and average Twist (34 ± 11°) in accord with the corresponding values from average B-DNA crystal structure geometry (3.36 Å, 36 ± 7°) as well as previous experimental and theoretical work. Because the backbone was not included in the present computations, the agreement observed between theory and experiment is somewhat surprising and suggests that the backbone either does not significantly influence the Rise and Twist parameters in B-form DNA or that the backbone has a local (or global) conformational minimum that is (by evolution or coincidence) an excellent match with the Twist angles that provide optimal stacking energies.

Substituents (relative to the "parents" purine and 2-pyrimidone) lead to increased dispersion and charge penetration. Total energies relative to the R·2Y reference base-pair steps are more attractive at all but a select few geometries, accounting for 20−40% of the base-stacking energy in general. This stabilization is not highly dependent on Rise or Twist, changing the depth of the potential energy surface more than its shape. This finding suggests that chemical modifications of base pairs may have substantial effects on the strength of their base-stacking interactions.

The four base-pair steps that significantly deviate from the B-DNA average Slide value of 0.26 Å have SAPT0 energetic minima that match well with their crystallographic averages. The average Slide values of the remaining base pairs (which are clustered closely around 0.26 Å) do not correlate well with their SAPT0 minima. None of the 10 base-pair steps of the A-RNA helix have crystallographic average Slide values consistent with the computed SAPT0 minima. While deviations from the average Slide in B-form DNA are well-explained by base-stacking energy, other factors must dominate the Slide preferences of A-form RNA. Thus, the RNA backbone, which differs from the DNA backbone only by the presence of the 2′-OH group, must largely dictate the helical parameters associated with base-pair stacking aside from Rise.

When the AA:TT, AC:GT, AG:CT, and AT:AT steps are compared with their uracil-containing counterparts, the most notable effect of the thymine methyl group is short-range steric clashes between the methyl hydrogens and the stacked base pair, leading to unfavorable exchange-repulsion terms at moderate to large Twist angles (or at all Twist angles for AA:TT). However, more favorable electrostatic terms can compensate for this, leading to a net stabilization at low Twist angles. For the CA:TG, GA:TC, and TA:TA steps, the net dispersion term is generally favorable because of direct methyl−π interactions, leading to stabilization (especially at low Twist). Whether we use theoretical geometries or the average experimental Twist values for DNA, we compute a stabilization of about 0.2 kcal mol^{-1} per AA:TT step due to the thymine methyl group, which compares well to the experimental values of 0.29−0.35 kcal mol^{-1}.[40,41]

■ ASSOCIATED CONTENT

Supporting Information

Comparison of theoretical Twist values to experimental averages for each base-pair step; minima of SAPT potential curves with respect to Twist; difference between DNA and RNA step interaction energies at experimental Twist angles; graphs of SAPT components vs Twist for all values of Rise considered (including comparison of SAPT electrostatics vs DMA electrostatics to deduce charge-penetration terms); graph of SAPT components vs Slide for optimal Rise and Twist; full SAPT0 data for all potential curves for the ranges of Rise, Twist, and Slide considered; and Cartesian coordinates for each base-pair step at the minimum values of Rise and Twist along with the Hartree−Fock total energies at these geometries. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

sherrill@gatech.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful to Tahir Yusufaly, Prof. Wilma Olson, and Prof. Jiri Šponer for helpful comments. This material is based upon work supported by the National Science Foundation (Grant CHE-1011360). T.M.P. was supported by a Research Experiences for Undergraduates (REU) in Chemistry Site

Award to Georgia Tech from the National Science Foundation (Grant 0851780). N.V.H. acknowledges support from the Vasser Woolley Foundation. R.M.P. is supported by the DOE Computational Science Graduate Fellowship, Grant No. DE-FG02-97ER25308. The Center for Computational Molecular Science and Technology is funded through a National Science Foundation CRIF Award (CHE-0946869) and by Georgia Tech.

REFERENCES

(1) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210−1250.
(2) Salonen, L. M.; Ellermann, M.; Diederich, F. *Angew. Chem., Int. Ed.* **2011**, *50*, 4808−4842.
(3) Saenger, W. *Principles of Nucleic Acid Structure*; Springer: New York, 1984.
(4) Šponer, J.; Lankas, F. *Computational Studies of DNA and RNA*; Springer: Dordrecht, The Netherlands, 2006.
(5) Alber, F.; Dokudovskaya, S.; Veenhoff, L. M.; Zhang, W.; Kipper, J.; Devos, D.; Suprapto, A.; Karni-Schmidt, O.; Williams, R.; Chait, B. T.; Rout, M. P.; Sali, A. *Nature* **2007**, *450*, 683−694.
(6) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149−1170.
(7) Nir, E.; Kleinermanns, K.; de Vries, M. S. *Nature* **2000**, *408*, 949−951.
(8) Sherrill, C. D. *Rev. Comput. Chem.* **2009**, *26*, 1−38.
(9) Riley, K. E.; Pitoňák, M.; Jurečka, P.; Hobza, P. *Chem. Rev.* **2010**, *110*, 5023−5063.
(10) Šponer, J.; Riley, K. E.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595−2610.
(11) Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **2006**, *34*, 564.
(12) Řezáč, J.; Hobza, P. *Chem.—Eur. J.* **2007**, *13*, 2983−2989.
(13) Černý, J.; Kabeláč, M.; Hobza, P. *J. Am. Chem. Soc.* **2008**, *130*, 16055−16059.
(14) Thanthiriwatte, K. S.; Hohenstein, E. G.; Burns, L. A.; Sherrill, C. D. *J. Chem. Theory Comput.* **2011**, *7*, 88−96.
(15) Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608−15613.
(16) Šponer, J.; Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142−10151.
(17) Hobza, P.; Šponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802−11808.
(18) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479−483.
(19) Šponer, J.; Jurečka, P.; Marchan, I.; Luque, F. J.; Orozco, M.; Hobza, P. *Chem.—Eur. J.* **2006**, *12*, 2854−2865.
(20) Cooper, V. R.; Thonhauser, T.; Puzder, A.; Schroder, E.; Lundqvist, B. I.; Langreth, D. C. *J. Am. Chem. Soc.* **2008**, *130*, 1304−1308.
(21) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S. H.; Srinivasan, A. R.; Schneider, B. *Biophys. J.* **1992**, *63*, 751−759.
(22) ElHassan, M. A.; Calladine, C. R. *Philos. Trans. R. Soc. London, Ser. A* **1997**, *355*, 43−100.
(23) Gorin, A. A.; Zhurkin, V. B.; Olson, W. K. *J. Mol. Biol.* **1995**, *247*, 34−48.
(24) Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. M.; Zhurkin, V. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163−11168.
(25) Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X.-J.; Neidle, S.; Shakked, Z.; Sklenar, H.; Suzuki, M.; Tung, C.-S.; Westhof, E.; Wolberger, C.; Berman, H. M. *J. Mol. Biol.* **2001**, *313*, 229−237.
(26) Perez, A.; Noy, A.; Lankas, F.; Luque, F. J.; Orozco, M. *Nucleic Acids Res.* **2004**, *32*, 6144−6151.
(27) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887−1930.
(28) Williams, H. L.; Chabalowski, C. F. *J. Phys. Chem. A* **2001**, *105*, 646−659.
(29) Jansen, G.; Hesselmann, A. *J. Phys. Chem. A* **2001**, *105*, 11156−11157.
(30) Misquitta, A. J.; Jeziorski, B.; Szalewicz, K. *Phys. Rev. Lett.* **2003**, *91*, No. 033201.
(31) Podeszwa, R.; Szalewicz, K. *Chem. Phys. Lett.* **2005**, *412*, 488−493.
(32) Heßelmann, A.; Jansen, G.; Schütz, M. *J. Chem. Phys.* **2005**, *122*, No. 014103.
(33) Hohenstein, E. G.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *132*, No. 184111.
(34) Hohenstein, E. G.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *133*, No. 014101.
(35) Hohenstein, E. G.; Parrish, R. M.; Sherrill, C. D.; Turney, J. M.; Schaefer, H. F., III. *J. Chem. Phys.* **2011**, *135*, No. 174107.
(36) Sedlák, R.; Jurečka, P.; Hobza, P. *J. Chem. Phys.* **2007**, *127*, No. 075104.
(37) Hesselmann, A.; Jansen, G.; Schütz, M. *J. Am. Chem. Soc.* **2006**, *128*, 11730−11731.
(38) Fiethen, A.; Jansen, G.; Hesselmann, A.; Schütz, M. *J. Am. Chem. Soc.* **2008**, *130*, 1802−1803.
(39) Svozil, D.; Hobza, P.; Šponer, J. *J. Phys. Chem. B* **2010**, *114*, 1191−1203.
(40) Ross, P. D.; Howard, F. B. *Biopolymers* **2003**, *68*, 210−222.
(41) Howard, F. B. *Biopolymers* **2005**, *78*, 221−229.
(42) Soto, A. M.; Rentzeperis, D.; Shikiya, R.; Alonso, M.; Marky, L. A. *Biochemistry* **2006**, *45*, 3051−3059.
(43) Wang, S.; Kool, E. T. *Biochemistry* **1995**, *34*, 4125−4132.
(44) Chaterjee, S.; Pathmasiri, W.; Chattopadhyaya, J. *Org. Biomol. Chem.* **2005**, *3*, 3911−3915.
(45) Sun, Z.; McLaughlin, L. W. *Biopolymers* **2007**, *87*, 183−195.
(46) Acharya, P.; Cheruku, P.; Chatterjee, S.; Acharya, S.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **2004**, *126*, 2862−2869.
(47) Vakonakis, I.; LiWang, A. C. *J. Am. Chem. Soc.* **2004**, *126*, 5688−5689.
(48) Pérez, A.; Sponer, J.; Jurecka, P.; Hobza, P.; Luque, F. J.; Orozco, M. *Chem.—Eur. J.* **2005**, *11*, 5062−5066.
(49) Swart, M.; Guerra, C. F.; Bickelhaupt, F. M. *J. Am. Chem. Soc.* **2004**, *126*, 16718−16719.
(50) Šponer, J.; Zgarbová, M.; Jurečka, P.; Riley, K. E.; Šponer, J. E.; Hobza, P. *J. Chem. Theory Comput.* **2009**, *5*, 1166−1179.
(51) Isaksson, J.; Acharya, S.; Barman, J.; Chereku, P.; Chattopadhyaya, J. *Biochemistry* **2004**, *43*, 15996−16010.
(52) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(53) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796−6806.
(54) Crawford, T. D.; Sherrill, C. D.; Valeev, E. F.; Fermann, J. T.; King, R. A.; Leininger, M. L.; Brown, S. T.; Janssen, C. L.; Seidl, E. T.; Kenny, J. P.; Allen, W. D. *J. Comput. Chem.* **2007**, *28*, 1610−1616.
(55) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 556−565.
(56) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(57) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787−1799.
(58) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio, R. A., Jr.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y.

M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L., III; Zhang, W.; Bell, A. T.; Chakraborty, A. K. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172−3191.
(59) Our definition of Twist differs minimally from that in nucleic acid analysis programs in that we have no glycosal atoms and thus we must use C6 and C8 to define the base-pair long axis and Twist.
(60) Stone, A. J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047−1064.
(61) Werner, H.-J.; Knowles, P. J.; Manby, F. R.; Schütz, M.; Celani, P.; Knizia, G.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M.; Wolf, A. *MOLPRO: A Package of Ab Initio Programs*, version 2010.1; http://www.molpro.net.
(62) Hohenstein, E. G.; Duan, J.; Sherrill, C. D. *J. Am. Chem. Soc.* **2011**, *133*, 13244−13247.
(63) Janowski, T.; Pulay, P. *Chem. Phys. Lett.* **2007**, *447*, 27−32.
(64) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. *J. Phys. Chem. A* **2009**, *113*, 10146−10159.
(65) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *132*, No. 144104.
(66) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Submitted.
(67) Hill, J. G.; Platts, J. A. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2785−2791.
(68) Takatani, T.; Sherrill, C. D. *Phys. Chem. Chem. Phys.* **2007**, *9*, 6106−6114.
(69) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993.
(70) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. *J. Chem. Phys.* **2011**, *135*, No. 194102.
(71) Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 7690−7697.
(72) Ringer, A. L.; Sherrill, C. D. *J. Am. Chem. Soc.* **2009**, *131*, 4574−4575.
(73) Sherrill, C. D.; Sumpter, B. G.; Sinnokrot, M. O.; Marshall, M. S.; Hohenstein, E. G.; Walker, R. C.; Gould, I. R. *J. Comput. Chem.* **2009**, *30*, 2187−2193.
(74) This average value of Slide is higher than normal because of contributions from highly overtwisted states caused by contacts between duplexes in the crystal lattice (see ref 23).
(75) Hud, N. V.; Polak, M. *Curr. Opin. Struct. Biol.* **2001**, *11*, 292−301.
(76) Sugimoto, N.; Nakano, M.; Nakano, S. *Biochemistry* **2000**, *39*, 11270−11281.
(77) Beššeová, I.; Banáš, P.; Kührová, P.; Košinová, P.; Otyepka, M.; Šponer, J. *J. Phys. Chem. B* **2012**, *116*, 9899−9916.
(78) Wheeler, S. E.; Houk, K. N. *J. Am. Chem. Soc.* **2008**, *130*, 10854−10855.
(79) Wheeler, S. E.; Houk, K. N. *J. Chem. Theory Comput.* **2009**, *5*, 2301−2312.