

# Fast Approximate Methods for Calculating Nucleic Acid Base Pair Interaction Energies

EDWARD C. SHERER, DARRIN M. YORK, CHRISTOPHER J. CRAMER

Department of Chemistry and Supercomputer Institute, University of Minnesota, 207 Pleasant St.  
SE, Minneapolis, Minnesota 55455-0431

Received 4 April 2002; Accepted 19 June 2002

**Abstract:** Interaction enthalpies for six base pairs have been computed at a variety of efficient levels of electronic structure theory and compared to experiment. In addition to previously defined levels of theory, modified Hamiltonians with adjusted parameters in hybrid Hartree–Fock/density functionals and semiempirical neglect-of-diatomic-differential-overlap models were examined. Of the pure and hybrid density functional levels, *mPWPW91/MIDI!* performed most satisfactorily, as judged by comparison not only to the available experimental data, but also to data from more robust electronic structure methods for 22 additional base pairs. The low computational cost of the *mPWPW91/MIDI!* model was further exploited in an investigation of various base trimers, tetramers, and one base pentamer. A carefully reparameterized semiempirical model, *PM3<sub>BP</sub>*, was able to achieve similar levels of accuracy at a still greater savings in terms of computational effort.

© 2002 Wiley Periodicals, Inc. J Comput Chem 24: 57–67, 2003

**Key words:** nucleic acid; base pair; hydrogen bond; hybrid density functional; semiempirical theory; base trimer; parameterization

## Introduction

Nucleic acids canonically form Watson–Crick base pairs, where adenine (A) pairs with thymine (T) or uracil (U), and guanine (G) pairs with cytosine (C).<sup>1</sup> Insofar as they play an integral role in chemistry and biology, enormous effort continues to be expended on improving our understanding of the structure, dynamics, and energetics of nucleic acid base-pair interactions.<sup>2–10</sup> Current methods of calculating the structures and interaction energies of isolated nucleic acid base pairs typically rely on the exploitation of medium-sized *ab initio* geometry optimizations followed by single-point energy calculations that take some account of electron correlation.<sup>2,3,9</sup> While such methodology has proven successful in the modeling of base-pair dimers, it remains largely impractical, owing to computational cost, for the study of interactions between more than two nucleic acid bases. With interest growing in the use of fully quantum mechanical simulations to model biomolecules,<sup>11</sup> it is interesting to ask whether less demanding levels of electronic structure theory may be able to predict base-pair interaction energies with a similar degree of accuracy, either through favorable cancellation of errors, or by careful parameterization over the limited chemical space of nucleic acid base functionality.

As a first step in that direction, we report here studies with a variety of density functional theory (DFT) methods, including hybrid Hartree–Fock/DFT methods, designed to determine the optimum functional for predicting base-pair interaction energies in

hydrogen-bonded (as opposed to stacked) complexes. The qualities of the methods are judged by comparison both to experimental interaction enthalpies and to interaction energies computed from high-level single-point energy calculations in cases where experimental data are not available. Using these and additional data for isolated nucleic acid bases, we then adjust the parameters of the semiempirical neglect-of-diatomic-differential-overlap<sup>12</sup> (NDDO) model *PM3*<sup>13</sup> to reproduce structures, dipole moments, and interaction energies. The resulting methods should be particularly useful in situations where high efficiency in the modeling of such systems is a requirement.

## Theoretical Methods

### *Ab initio and Density Functional Levels of Theory*

The geometries of all nucleic acid base pairs for which interaction energies are available from mass-spectrometry<sup>14,15</sup> were fully op-

**Correspondence to:** C. J. Cramer; e-mail: [cramer@chem.umn.edu](mailto:cramer@chem.umn.edu)

This article includes supplementary material available from the authors upon request or via the Internet at <ftp://ftp.wiley.com/public/journals/jcc/suppmat/23/57> or <http://www.interscience.wiley.com/jpages/0192-8651/suppmat/23.57.html>

timized in the gas phase (as were the corresponding base monomers) at the Hartree–Fock (HF) level using the MIDI!,<sup>16</sup> 6-31G(d), and 6-31+G(d,p) basis sets.<sup>17</sup> Geometry optimizations using these basis sets were also carried out at the DFT level using a variety of different functionals. Two “pure” DFT functionals were surveyed. The first combined the gradient-corrected exchange functional of Becke<sup>18</sup> with the gradient-corrected correlation functional of Lee, Yang, and Parr<sup>19</sup> (BLYP). The second employed the gradient corrected exchange<sup>20</sup> and correlation<sup>21</sup> functionals of Perdew and coworkers, as modified by Adamo and Barone<sup>22</sup> (*mPWPW91*). In addition, two hybrid HF/DFT functionals<sup>23</sup> were examined. The so-called B3LYP functional defines the energy as<sup>24</sup>

$$E_{xc}^{\text{B3LYP}} = (1 - a)E_x^{\text{LSDA}} + aE_x^{\text{HF}} + b\Delta E_x^{\text{B}} + (1 - c)E_c^{\text{LSDA}} + cE_c^{\text{LYP}} \quad (1)$$

where LSDA is the local spin-density approximation<sup>25,26</sup> (i.e., the DFT energy without any gradient correction) and  $a$ ,  $b$ , and  $c$  take the values 0.20, 0.72, and 0.81, respectively. The *mPWPW91* functional is a hybrid modification of the pure *mPWPW91* functional and is defined by<sup>22</sup>

$$E_{xc}^{\text{mPWPW91}} = aE_x^{\text{HF}} + (1 - a)(E_x^{\text{LSDA}} + \Delta E_x^{\text{PW}}) + E_c^{\text{LSDA}} + \Delta E_c^{\text{PW91}} \quad (2)$$

where the single parameter  $a$  is 0.25. To improve the performance of the hybrid levels in some instances,  $a$ ,  $b$ , and  $c$  were here treated as adjustable parameters, as described in more detail in the text.

The geometries of all base pairs other than those noted above were optimized only at the *mPWPW91*/MIDI! level of theory, following identification of that level as being optimal in comparison to the others.

All stationary points were verified as minima by computation of analytic vibrational frequencies. These frequencies were used to compute zero-point vibrational energies and thermal contributions to 298 K enthalpies using standard ideal-gas statistical mechanics and the rigid-rotor harmonic-oscillator approximation.<sup>27</sup> Spirko et al. have found harmonic frequencies for base pairs to agree well with independently calculated anharmonic frequencies.<sup>28</sup> The frequencies were not scaled,<sup>29</sup> because for pure DFT methods most scale factors differ insignificantly from 1.00. Scaling the HF frequencies, which represent the limit in overestimation, by a factor of 0.9, has a fairly constant and small effect, favoring every base pair by 0.4 to 0.5 kcal/mol over the corresponding separated monomers.

### Interaction Energies and Enthalpies

The “raw” interaction energy  $\Delta E$  at a given level of theory is computed as

$$\Delta E[A \cdot B] = E_{A \cdot B}^{\alpha\beta} - E_A^\alpha - E_B^\beta \quad (3)$$

where  $A$  and  $B$  are nucleic acid base monomers,  $A \cdot B$  is the base pair, the subscripts on the r.h.s. refer to the structure for which the optimized geometry was employed in computing the energy, and

the superscripts refer to the basis set employed, where  $\alpha$  is the basis set residing on the atoms of monomer  $A$  and  $\beta$  is the basis set residing on the atoms of monomer  $B$ .

Because we are working with relatively small basis sets, the raw interaction energies are significantly overestimated. We employ the procedure of Xantheas<sup>30</sup> to correct for basis-set superposition error (BSSE) according to

$$\Delta E^*[A \cdot B] = \Delta E[A \cdot B] + E_{\text{BSSE}}[A \cdot B] \quad (4)$$

where the correction  $E_{\text{BSSE}}$  is defined by

$$E_{\text{BSSE}}[A \cdot B] = \{E_{A \cdot B}^{\alpha\beta} - E_{A \cdot B}^{\alpha\beta}[A]\} + \{E_{A \cdot B}^{\alpha\beta}[B] - E_{A \cdot B}^{\alpha\beta}[B]\} \quad (5)$$

where all subscripts and superscripts on the r.h.s. have the same meaning as in eq. (3).

We define 298 K gas-phase enthalpies as

$$H_{298}^0 = E + \text{ZPVE} + (\Delta H_{298}^0 - \Delta H_0^0) \quad (6)$$

where  $E$  is the electronic energy, ZPVE is the zero-point vibrational energy, and the remaining term on the r.h.s. accounts for 298 K thermal contributions. The standard state pressure is 1 atm. In the case of a base pair,  $E$  is the electronic energy plus  $E_{\text{BSSE}}$ . A dimerization enthalpy is then the difference between the enthalpy of the dimer and the enthalpies of the separated monomers.

In the case of trimers, tetramers, etc., eqs. (3) and (5) generalize, respectively, to

$$\Delta E[A_1 A_2 \cdots A_n] = E_{A_1 A_2 \cdots A_n}^{\alpha_1 \alpha_2 \alpha_3 \cdots \alpha_n} - \sum_{i=1}^n E_{A_i}^{\alpha_i} \quad (7)$$

and

$$E_{\text{BSSE}}[A_1 A_2 \cdots A_n] = \sum_{i=1}^n (E_{A_1 A_2 \cdots A_n}^{\alpha_i} - E_{A_1 A_2 \cdots A_n}^{\alpha_1 \alpha_2 \alpha_3 \cdots \alpha_n}[A_i]) \quad (8)$$

### Semiempirical Calculations

Semiempirical quantum calculations were performed using the MNDO,<sup>31</sup> AM1,<sup>32</sup> and PM3,<sup>13</sup> Hamiltonian models, in addition to a newly parameterized Hamiltonian designed specifically for nucleic acid base pairing. The details of the parameterization procedure for the last model are presented elsewhere.<sup>33</sup>

### Software

*Ab initio* and density functional calculations were carried out with the Gaussian 98 suite of electronic structure programs.<sup>34</sup> Semiempirical calculations were performed using a modified version of MNDO97.<sup>35</sup>

## Results

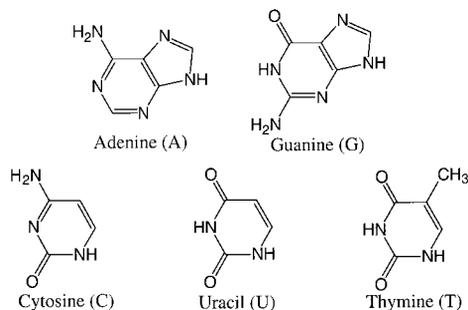
### Nomenclature

We here adopt the base-pair labeling scheme utilized by Hobza and Sponer.<sup>2,3</sup> The monomers that make up each base pair are illustrated in Figure 1. The interaction geometry of a given base pair is normally referred to as Watson-Crick (WC), Hoogsteen (H), wobble (WB), or the “reverse” form of these pairs, in which case the label is preceded by an R, for example, RWC for reverse Watson-Crick. Each of the six possibilities is illustrated in Figure 2. Although the structure of every base pair examined here has not been explicitly drawn, their geometries may be qualitatively inferred by reference to Figure 2. Complete geometrical details are available in the supporting information. We also add a seventh geometrical category corresponding to dimers labeled by Hobza and Sponer<sup>2</sup> as GC1, GA4, GA2, and GG4. As these base pairs are consistently formed with hydrogen bonds to the N3 of guanine, we choose to label these pairs as PN3N#, where the # corresponds to the nitrogen atom that forms the hydrogen bond to N3 of guanine, and “P” represents purine. For instance, the base pair GC1 is formed when the N4 and N3 of cytosine hydrogen bond to the N3 and N2 of guanine, respectively. This structure would then be labeled PN3N4.

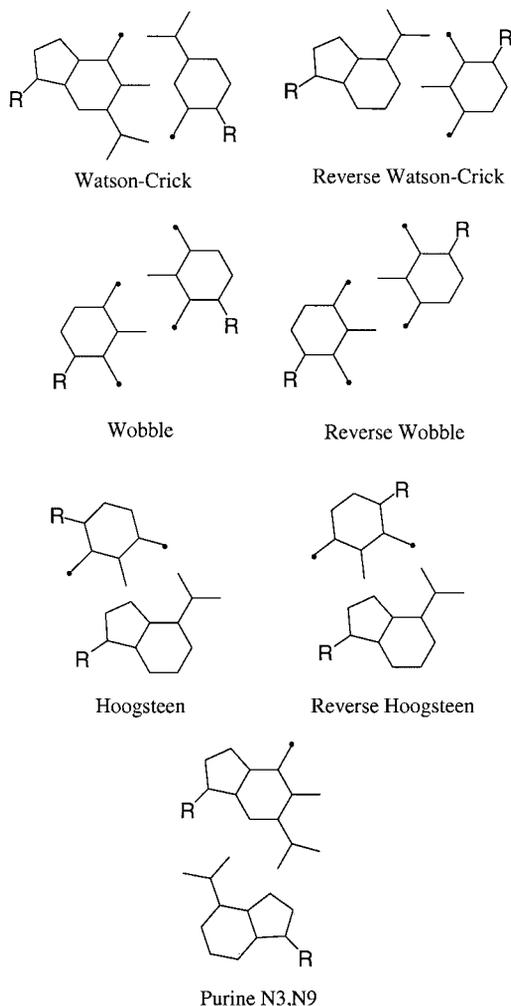
From this point forward we will always list the two bases in a base pair in alphabetical order (e.g., CG, not GC). Although this deviates slightly from the notation found in some earlier works, it greatly facilitates the readability of tables containing many different base pairs. We will not enforce alphabetical order in trimers and higher order structures.

### DFT Interaction Energies

In the late 1970s, Yanson et al.<sup>14,15</sup> carried out mass-spectrometric measurements of temperature-dependent, gas-phase, base-pairing equilibrium constants to determine interaction enthalpies for various base pairs. In particular, they measured values (kcal mol<sup>-1</sup>) of: CG, -21.0; CC, -16.0; AU, -14.5; AT, -13.0; UU, -9.5; and TT, -9.0 (all measurements were assigned to have an error bar of  $\pm 1.0$ ). They also showed that these binding enthalpies were independent of methylation at the N1 positions of the pyrimidines and the N9 positions of the purines. Moreover, changes induced by methylation of specific base pair functional groups (or a lack thereof) were interpreted by Yanson et al. to indicate that the bases



**Figure 1.** Structures of nucleic acid base monomers.



**Figure 2.** Common base-pair complexation geometries; oxygen is depicted as a dot and the positions at which ribosyl substitution would be found in nucleosides are labeled with a substituent R.

were hydrogen bonding and not stacking. Based on our own calculations reported below as well as the conclusions of Yanson et al., we assign the lowest energy base-pair geometries for the measured species to be CG, WC; CC, RWB; AU, WC; AT, WC; UU, RWB; and TT, RWB. In several cases, however, other geometries have very similar interaction energies (vide infra), so this distinction is not of paramount importance.

Table 1 reports the mean signed error (MSE, theory minus experiment), mean unsigned error (MUE), and root-mean-square error (RMS) in the calculation of the six base-pairing enthalpies for several different levels of electronic structure theory. For comparison, results obtained from Hobza and Sponer's<sup>2</sup> recommended optimal level of theory for computation of base-pairing energies are included (the PM3<sub>BP</sub> level, which is also included, is discussed further below).

The excellent price-performance ratio of the *m*PWPW91/MIDI! level of theory, together with other factors discussed further below (including molecular geometries), prompted us to examine

**Table 1.** Errors (kcal mol<sup>-1</sup>) in Computed 298 K Interaction Enthalpies for Six Experimentally Characterized Base Pairs.

Level of theory	MSE	MUE	RMS
HF/MIDI!	4.8	4.8	5.2
HF/6-31G(d)	3.6	3.6	4.1
HF/6-31+G(d,p) <sup>a</sup>	2.9	2.9	3.8
BLYP/MIDI!	2.4	2.4	2.9
BLYP/6-31G(d)	1.5	1.8	2.3
BLYP/6-31+G(d,p) <sup>a</sup>	1.2	2.0	2.3
<i>m</i> PWPW91/MIDI!	0.4	1.3	1.6
<i>m</i> PWPW91/6-31G(d)	0.0	1.5	1.9
<i>m</i> PWPW91/6-31+G(d,p) <sup>a</sup>	-0.7	2.3	2.4
B3LYP/MIDI!	1.4	1.6	2.2
B3LYP/6-31G(d)	0.9	1.5	2.0
B3LYP/6-31+G(d,p) <sup>a</sup>	0.0	2.4	2.4
<i>m</i> PW1PW91/MIDI!	0.2	1.4	1.8
<i>m</i> PW1PW91/6-31G(d)	-0.2	1.7	2.1
<i>m</i> PW1PW91/6-31+G(d,p) <sup>a</sup>	-1.0	2.6	2.8
<i>m</i> B3LYP/MIDI! <sup>b</sup>	0.2	1.4	1.7
PM3 <sub>BP</sub>	0.1	1.2	1.5
Ref. 71 <sup>c</sup>	1.1	1.4	1.9

<sup>a</sup>Computed over only AT and GC.

<sup>b</sup>Using optimized values  $a = 13$ ,  $b = 69$ ,  $c = 73$ .

<sup>c</sup>Geometry optimization, ZPVE, and thermal contributions at HF/6-31G(d) level followed by BSSE correction and MP2/6-31G(d) calculation with all d polarization functions using an exponent of 0.25. Values for UU and AU at this level were computed as part of this work, because they do not appear in the original reference.

this level of theory in more detail for a number of base pairs where experimental data are not available, but for which in many instances there *are* data from high level calculations using the recommended level of Hobza and Sponer.<sup>2,3</sup> Interaction enthalpies for 35 different base pair geometries are listed in Table 2. For the 28 cases computed by Hobza and Sponer, the mean unsigned difference between their MP2 enthalpies and those from the *m*PWPW91/MIDI! level of theory is 1.1 kcal mol<sup>-1</sup>.

The *m*PWPW91/MIDI! level was also used to compute various trimerization and tetramerization enthalpies. Yanson et al.<sup>15</sup> have reported enthalpies of trimerization for CCC, UUA, UUU, and UUT based on mass spectral studies. Because Yanson et al. employed cytosine methylated at the exocyclic amino group, they assigned a trimer geometry labeled in Figure 3 as CCC3. They further assumed the UUA system to have the geometry labeled in Figure 3 as UUA1, and offered no proposals for the structures of UUU and UUT. Additional structures in Figure 3 were explored computationally, and the data are found in Table 3. Computed trimerization enthalpies are considerably smaller than those reported by Yanson et al., even given the rather broad range of the experimental values. This point will be discussed in more detail below.

Although they have not been examined by mass spectrometry, TAT and CGC(H)<sup>+</sup> trimers (the latter involves one protonated base and is thus cationic) are of special interest because they occur in DNA triple helices (Fig. 3). Prior computational studies of these bases have arrived at similar interaction energies to those com-

puted at the *m*PWPW91/MIDI! level (Table 3).<sup>36,37</sup> A separate classical electrostatic approach to calculate intermolecular complexation energies using charge distributions derived from quantum mechanical calculations arrived at an interaction energy for TAT of -20.9 kcal/mol, which is in good agreement with the fully quantum mechanical values.<sup>38</sup>

Table 3 also contains an enthalpy of tetramerization computed at the *m*PWPW91/MIDI! level for the GGGG tetrad (Fig. 4). In relatively large concentrations, guanine forms a gel in water, and the structure formed in solution has been identified as a guanine quadruplex, or tetrad.<sup>39-41</sup> At moderate ionic strength, this tetrad forms a square coplanar conformation of four guanines that are bound together in a Hoogsteen fashion so that each guanine is involved in four intermolecular hydrogen bonds (GGGG1 in Fig. 4). The structure and properties of the guanine tetrad (as well as U, T, A, and mixed tetrads) have been modeled at the HF and B3LYP levels by Gu and Lezczynski<sup>42-47</sup> and at the B3LYP level by Meyer et al.<sup>48,49</sup> At the B3LYP/6-311G(d,p)/HF/6-311G(d,p) level,<sup>42</sup> the GGGG1 interaction energy is predicted to be -66.2 kcal mol<sup>-1</sup>, while at the B3LYP/TZVP level it is -64.6 kcal mol<sup>-1</sup>. The *m*PWPW91/MIDI! model is in almost exact agreement with the latter value.

In part to demonstrate the efficiency of the methodology, we have also computed the interaction energy for a guanine pentad (Fig. 5). This structure, with a computed interaction enthalpy of -68.5 kcal mol<sup>-1</sup>, requires a substantial rearrangement of the GGGG1 tetrad in order to form. The interaction enthalpy of the tetrad portion of the pentad, -50.5 kcal mol<sup>-1</sup>, is listed in Table 2 for structure GGGG2 (Fig. 4). The GGGGG structure of Figure 5 may or may not represent the global minimum in the gas phase—we have not searched the potential energy surface sufficiently to assess this point.

### Semiempirical Interaction Energies

The continued design of accurate quantum methods that can be applied to increasingly large systems is clearly an endeavor of great importance. The DFT work presented here prompted us to consider whether a still more efficient quantum model could be developed specifically to model hydrogen bonding in nucleic acid base pairs.

Semiempirical models typically require orders of magnitude less computational effort than *ab initio* ones, and hence, can be extended to very large systems, especially when used in conjunction with linear-scaling electronic structure methods.<sup>11,50,51</sup> The trade-off is that these methods are less robust; i.e., they have limited reliability outside the scope of the chemistry for which the parameters were designed. However, within this scope, semiempirical methods can be extremely powerful models for applications that involve large molecular systems and/or long time domains, and in many instances they serve as the present method of choice for hybrid quantum mechanical/molecular mechanical applications to biological macromolecules.

Thus, we adopted a conventional semiempirical framework (NDDO) and reparameterized the Parameterized Model 3 (PM3)<sup>13</sup> Hamiltonian now taking the somewhat larger set of DFT results as the parameterization standard—a hierarchical parameterization, one might say. DFT structures, dipole moments, and relative

**Table 2.** Interaction Enthalpies (298 K, kcal mol<sup>-1</sup>) for Various Base Pairs.

Base pair <sup>a</sup>	Geometry <sup>b</sup>	$\Delta H_{298}^c$			Experiment <sup>d</sup>
		PM3 <sub>BP</sub>	mPWPW91/MIDI!	MP2 <sup>c</sup>	
AA1	RWB	-12.5	-11.8	-9.3	
AA2	RH	-12.0	-10.5	-8.8	
AA3	RH	-10.2	-9.0	-7.8	
AC1	RWB	-15.0	-13.8	-11.7	
AC2	RH	-14.6	-12.4	-11.4	
AG1	WC	-12.6	-12.6	-13.3	
AG2	PN3N9	-8.9	-8.7	-8.9	
AG3	H	-13.3	-11.7	-12.3	
AG4	PN3N1	-9.7	-10.1	-9.9	
ATH	H	-13.6	-11.2	-11.4	
ATRH	RH	-13.7	-10.7	-11.3	
ATRWC	RWC	-12.5	-10.6	-10.4	
ATWC	WC	-12.4	-11.3	-10.5	-13.0 ± 1.0
AU	WC	-12.6	-11.4	-10.7	-14.5 ± 1.0
CC	RWB	-18.8	-17.0	-15.5	-16.0 ± 1.0
CG1	PN3N3	-12.2	-12.3	-12.3	
CG2	RWB	-8.6	-8.4		
CGWC	WC	-21.4	-22.4	-21.9	-21.0 ± 1.0
CT1	RWC		-10.5	-9.5	
CT2	WC		-9.4	-9.5	
CU1	RWC	-10.2	-10.6		
CU2	WC	-10.6	-9.2		
GG1	RWB	-16.9	-17.6	-21.5	
GG2	H		-9.8		
GG3	RH	-12.3	-12.3	-16.6	
GG4	PN3N3	-6.3	-8.0	-9.4	
GT1	RWB		-12.5	-13.3	
GT2	WB		-11.1	-13.0	
GU1	RWB	-11.5	-12.6		
GU2	WB	-11.0	-10.8		
TT1	WB	-8.7	-8.6	-9.2	
TT2	RWB	-8.4	-9.4	-9.1	-9.0 ± 1.0
TT3	RWB	-8.9	-7.9	-9.1	
UU1	WB	-8.7	-8.3		
UU2	RWB	-8.7	-9.4	-9.0	-9.5 ± 1.0
UU3	RWB	-8.8	-7.5		

<sup>a</sup>Using nomenclature convention of reference 2; see also text.<sup>b</sup>See Figure 2.<sup>c</sup>Geometry optimization, ZPVE, and thermal contributions at HF/6-31G(d) level followed by BSSE correction and MP2/6-31G(d) calculation with all d polarization functions using an exponent of 0.25. Reference 2. Values for UU and AU at this level were computed as part of this work, because they do not appear in the original reference.<sup>d</sup>From mass spectrometry; refs. 14 and 15.

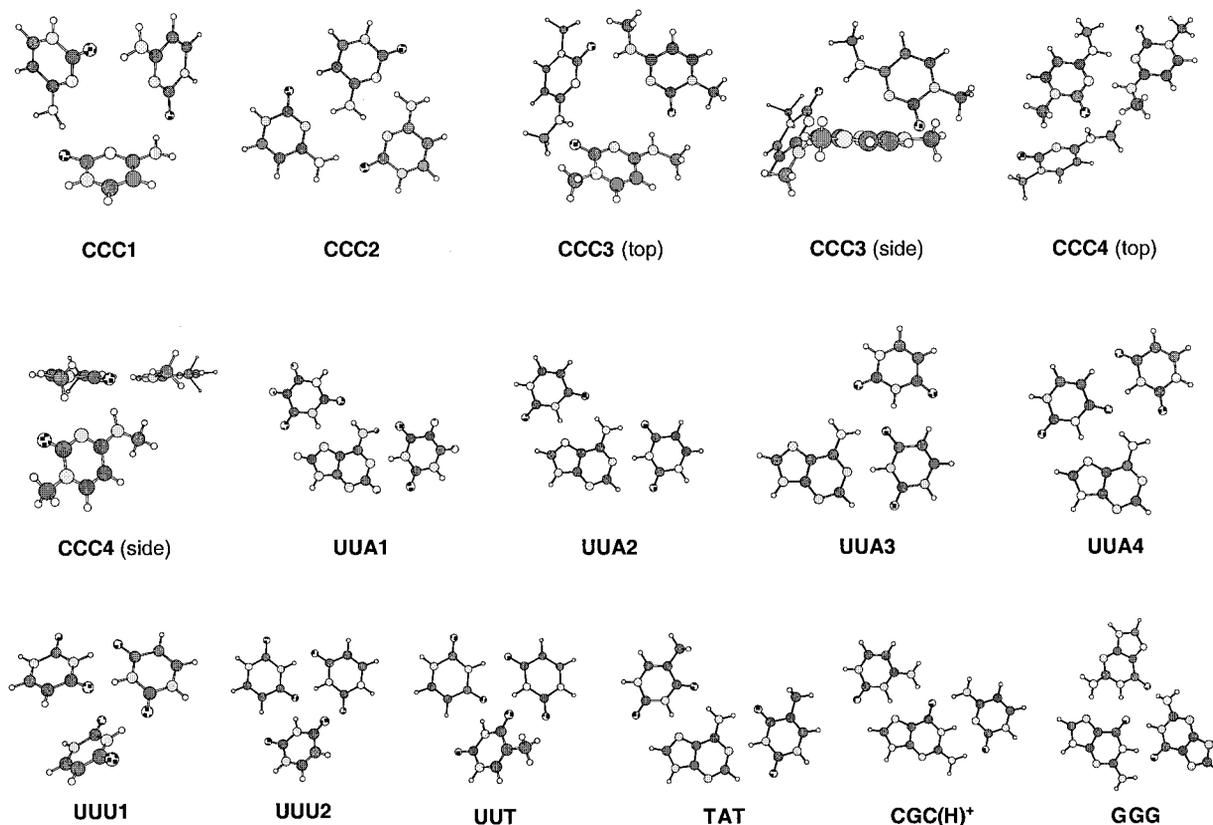
pairing energies served as the “training set” for optimization of the semiempirical parameters. Full details of the reparameterization process (including all parameters) are included in another article focusing on strategies for carrying out such reparameterizations in general.<sup>33</sup> Table 4 lists for a variety of properties the differences between the newly parameterized and developmental PM3<sub>BP</sub> (BP for base pair) model and either experiment, where available, or other levels of theory. For comparison, results from the original PM3 parameterization, as well as from Austin Model 1 (AM1)<sup>32</sup> and Modified Neglect of Differential Overlap (MNDO)<sup>31</sup> are also

included. In addition, the PM3<sub>BP</sub> base-pair interaction energies are listed in Table 2.

## Discussion

### Base-Pair Structures

Density functional theory is well established to provide accurate molecular structures in general when modern functionals are



**Figure 3.** Structures for various base trimers. Atomic legend: white, hydrogen; dark gray, carbon; light gray, nitrogen; stippled, oxygen.

used.<sup>27</sup> However, the performance of DFT with respect to the prediction of hydrogen bonded intermolecular complexes is less well established. High-quality experimental crystal structures are available for three base pairs in particular, namely two-base-step helical DNA sections of AU and CG pairs, and a Hoogsteen structure for 9-methylA:1-methylT (ATH) has been determined from neutron scattering.<sup>52–54</sup> All of the DFT levels surveyed in Table 1 had consistent average errors in hydrogen bond lengths of about 0.1 Å. With smaller basis sets, the error was typically in the direction of making the bonds too short (presumably in part from greater BSSE) while with larger basis sets, the error tended to be in the opposite direction. It should be noted that explicit inclusion of waters of hydration and counterions into gas phase optimizations has been found to bring computed geometries into better agreement with crystallography,<sup>5,55</sup> so one probably should not overly interpret the theoretical structures. It suffices to note that the *mPWPW91/MIDI!* geometries suffer from no apparent pathologies in the course of giving good energetic agreement with experiment and higher level calculations.

With respect to the planarity of base-pair structures, canonical pairs such as ATWC, CGWC, and noncanonical pairs like CUWC optimized to essentially perfectly planar structures. Other less standard pairing motifs tended to be predicted to adopt more highly propeller twisted or bent geometries; these structures typi-

cally showed considerably reduced interaction energies. Such results are entirely in keeping with prior analysis.<sup>2,48</sup>

Insofar as the  $PM3_{BP}$  Hamiltonian was parameterized including monomer and base pair geometries in the penalty function, it necessarily predicts molecular geometries in reasonably close agreement with the *mPWPW91/MIDI!* level. This includes hydrogen-bond lengths, where agreement tends to be within 0.05 Å for the majority of cases. The seemingly larger RMS error in Table 4 (0.151 Å) is heavily skewed by one anomalously large error, namely for GG1, where a single hydrogen bond is predicted to be too long by 0.740 Å.

#### Base-Pair Interaction Energies

As noted in Table 1, the performances of the small basis-set models *mPWPW91/MIDI!* and *mB3LYP/MIDI!* are considerably better than those of other DFT models making use of larger basis sets. Indeed, the two smaller models have accuracies equivalent to the much more computationally demanding level of Spomer and Hobza, which has been documented at some length to be useful in the study of base-pairing interactions.<sup>2,3</sup>

It must be noted that there is potentially some ambiguity in deciding exactly how to compare the computed interaction enthalpies with those measured<sup>14,15</sup> in the mass spectral experiments. As

**Table 3.** Interaction Enthalpies (298 K, kcal mol<sup>-1</sup>) for Various Base Complexes of Higher Molecularity.

Base pair <sup>a</sup>	$\Delta H_{298}^{\circ}$		Experiment <sup>c</sup>
	<i>m</i> PWPW91/MIDI!	MP2 <sup>b</sup>	
CCC1	-14.0		-33 (-38) ± 4
CCC2	-28.8		-33 (-38) ± 4
CCC4	-22.0		-33 (-38) ± 4
UUA1	-21.0		-27 (-29) ± 4
UUA2	-21.4		-27 (-29) ± 4
UUA3	-17.0		-27 (-29) ± 4
UUA4	-17.4		-27 (-29) ± 4
UUU1	-8.5		-20 (-22) ± 4
UUU2	-11.3		-20 (-22) ± 4
UUT	-7.1		-23 (-25) ± 4
TAT	-21.3	-23.8	
CGC <sup>+</sup>	-68.3	-65.2	
GGG	-36.8		
GGGG1	-64.5	-64.6 <sup>d</sup>	
GGGG2	-50.5		
GGGGG	-68.5		

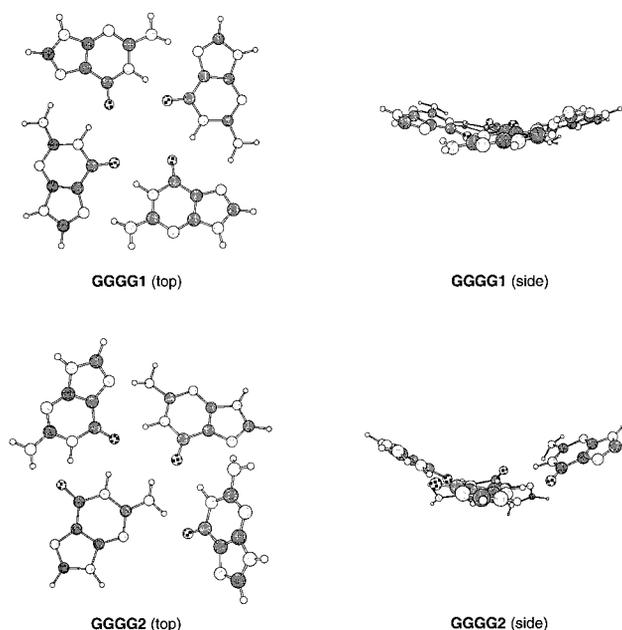
<sup>a</sup>See Figure 3.<sup>b</sup>Geometry optimization, ZPVE, and thermal contributions at HF/6-31G(d) level followed by BSSE correction and MP2/6-31G(d) calculation with all d polarization functions using an exponent of 0.25. Ref. 36.<sup>c</sup>From mass spectrometry; ref. 15.<sup>d</sup>Complexation energy at the B3LYP level with a triple- $\zeta$  basis set. Ref. 49.

Table 2 indicates, for some of the base pairs there are multiple geometries predicted to have interaction enthalpies sufficiently similar to one another that they might be expected to contribute in a nontrivial way to an equilibrium population. In the case of a mixture of different isomers, it is not obvious how to decompose into individual contributions the single experimental “enthalpy” derived from a van’t Hoff plot of dimer/monomer mass-peak intensity ratios vs. temperature. A case in point is the AT pair, where the comparable energies of the Hoogsteen, reverse Hoogsteen, Watson–Crick, and reverse Watson–Crick motifs have been noted previously.<sup>56,57</sup>

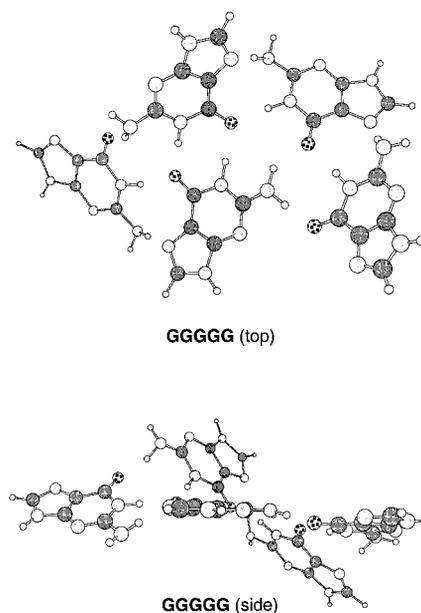
Considering this base pair, Brameld et al.<sup>56</sup> suggested that the experimental number should be reinterpreted based on equilibrium averaging. Their rationale appears to be that with *free energy*, one may write<sup>58</sup>

$$G(X) = -RT \ln \sum_i^{\text{isomers}} e^{-G(X_i)/RT} \quad (9)$$

where  $G(X)$  refers to the free energy of the full population of molecule  $X$ , which is constituted of several isomers  $X_i$ , each of which has a component free energy associated with it. If the entropies of all of the isomers are identical, then eq. (9) also holds with enthalpy  $H$  in place of free energy  $G$  at every occurrence. The results of assuming such a prescription for the enthalpy of binding at 298 K (computed as the difference between the enthalpy of the base-pair population and the sum of the individual monomer

**Figure 4.** Optimized structures of guanine tetramers. Legend as in Figure 3.

enthalpies, where population averaging is not an issue) are presented in Table 5 for AT, TT, and UU. The CC and CG base pairs have a single hydrogen-bonding geometry that is sufficiently favored over all others to completely dominate the equilibrium at 298 K. With respect to AU, numerous studies have shown that the extra methyl group on T has only a very small effect on base-

**Figure 5.** Optimized structure of guanine pentamer. Legend as in Figure 3.

**Table 4.** Root-Mean-Square Differences between Semiempirical Models and Experiment or Other Theory for Various Nucleic Acid Base and Base-Pair Properties.

Property	RMS Error			
	PM3 <sub>BP</sub>	PM3	AM1	MNDO
Monomer bond lengths vs. <i>mPWPW91/MIDI!</i> (Å) <sup>a</sup>	0.015	0.021	0.022	0.015
Monomer valence bond angles vs. <i>mPWPW91/MIDI!</i> (deg) <sup>a</sup>	2.6	2.9	2.6	2.0
Monomer torsion angles vs. <i>mPWPW91/MIDI!</i> (deg) <sup>a</sup>	0.7	2.7	0.7	1.1
Monomer dipole moments vs. B3LYP/cc-pVTZ <sup>a,b</sup>	0.1	0.7	0.4	0.7
Base-pair hydrogen-bond lengths vs. <i>mPWPW91/MIDI!</i> (Å) <sup>c</sup>	0.151	0.165	0.665	2.479
Base-pair hydrogen-bond angles vs. <i>mPWPW91/MIDI!</i> (deg) <sup>c</sup>	5.5	5.5	23.3	42.5
Base-pair interaction energies vs. experiment (kcal mol <sup>-1</sup> ) <sup>c</sup>	1.5	7.1	7.1	12.7
Base-pair interaction energies vs. <i>mPWPW91/MIDI!</i> <sup>c</sup>	1.3	5.8	5.7	10.3

<sup>a</sup>For A, C, G, T, and U.<sup>b</sup>Ref. 72.<sup>c</sup>See Table 2.

pairing with A, so while we have not done a full survey of all AU geometries, we would expect the net effect of averaging for AU to be very similar to that for AT.

Table 6 shows the results of error analysis performed for the PM3<sub>BP</sub> and *mPWPW91/MIDI!* levels, and for the multistep MP2 level of theory of Hobza and Sponer, using population averaged enthalpies for AT, AU (assuming an effect identical to that found in AT, i.e., lowering the binding enthalpy relative to the single most negative value by 0.5 kcal mol<sup>-1</sup>), TT, and UU. Comparing to Table 1, the effect of averaging improves the performance of the *mPWPW91/MIDI!* and multistep MP2 models to a significant extent, and has a small mixed effect on PM3<sub>BP</sub>, improving the mean unsigned error at the expense of the mean signed error. This suggests that the averaging procedure may have merit. However, to the extent that the predicted changes in the base-pairing enthalpies are at most half as large in magnitude as the reported experimental error in the measurement, some skepticism may be warranted.

On the technical front, the rather poor performance, documented in Table 1, of standard density functionals with moderate to large basis sets when used to compute hydrogen-bonding inter-

action enthalpies has been previously noted for select cases.<sup>2,3,9</sup> It is thus a given that the remarkably accurate predictions made by the *mPWPW91/MIDI!* and *mB3LYP/MIDI!* levels of theory must arise in part from a favorable cancellation of errors associated with the smaller basis set. Because our goal in this instance, however, is to develop efficient, approximate methods, we are happy to take advantage of this cancellation. We note as an aside that in this article we have carried out a far more complete study of the *mPWPW91/MIDI!* level than of the *mB3LYP/MIDI!* level simply because the former, as a pure DFT method, avoids the bottleneck of computing 4-index exact exchange integrals.<sup>27</sup> However, this difference in efficiency will probably be small in most cases, and may ultimately be eliminated with the development of linear-scaling algorithms.

It is a somewhat more subtle point how one chooses to view PM3<sub>BP</sub>. The poor performance of semiempirical levels of theory (as well as ab initio Hartree-Fock theory with minimal basis sets) represents another phenomenon previously noted for base pairing.<sup>2,5,9</sup> However, insofar as semiempirical models are intrinsically parametric, it is not obvious that improving their performance by parameter adjustment over a more restricted data set should be regarded as introduction of canceling error. Rather, it is simply more focused modeling, a philosophy that prevails in analogous areas like force-field development. In any case, the good performance of the *mPWPW91/MIDI!* and PM3<sub>BP</sub> models suggests that

**Table 5.** Population-Averaged Interaction Enthalpies (298 K, kcal mol<sup>-1</sup>) for Three Base Pairs.<sup>a</sup>

Base pair	$\Delta H_{298}^{\circ}$			
	PM3 <sub>BP</sub>	<i>mPWPW91/MIDI!</i>	MP2 <sup>b</sup>	Experiment <sup>c</sup>
AT	-14.1	-11.8	-11.9	-13.0 ± 1.0
TT	-9.3	-9.6	-9.8	-9.0 ± 1.0
UU	-9.4	-9.5	n/a	-9.5 ± 1.0

<sup>a</sup>Equation 9 and accompanying discussion.<sup>b</sup>Geometry optimization, ZPVE, and thermal contributions at HF/6-31G(d) level followed by BSSE correction and MP2/6-31G(d) calculation with all d polarization functions using an exponent of 0.25. Ref. 2.<sup>c</sup>From mass spectrometry; refs. 14 and 15.**Table 6.** Errors (kcal mol<sup>-1</sup>) in Computed 298 K Interaction Enthalpies for Six Experimentally Characterized Base Pairs Including Equilibrium Averaging.

Level of theory	MSE	MUE	RMS
<i>mPWPW91/MIDI!</i>	0.1	1.1	1.4
PM3 <sub>BP</sub>	-0.4	1.1	1.5
Ref. 71 <sup>a</sup>	0.6	1.2	1.5

<sup>a</sup>See footnote c of Table 1.

they will prove useful in future modeling efforts where efficient surveys of hydrogen-bonding interactions in large molecules or large sets of small molecules like nucleic acid bases are of interest prior to resorting (if possible) to more complete levels of electronic structure theory.

A critical caveat that must be mentioned is that we have not addressed the utility of the models developed herein for the quantitative estimation of base *stacking* interactions, compared to hydrogen bonding. Modern DFT functionals and semiempirical methods tend to do quite poorly in the evaluation of dispersion interactions, and this point has been noted particularly for the case of base-stacking interactions.<sup>2,9,59–61</sup> Interestingly, Elstner et al. recently demonstrated that the explicit augmentation of standard functionals with a dispersion model, thereby creating a method they refer to as SCC-DFTB, led to significant improvement in the estimation of stacking energies.<sup>62</sup> These and other studies<sup>63,64</sup> suggest that extensions of the model to systems having both stacking and hydrogen-bonding interactions merit exploration.

### Trimers and Larger Structures

Yanson et al.<sup>15</sup> have reported interaction enthalpies for four nucleic acid base trimers in addition to the dimers discussed above (Table 3). The trimerization enthalpies are determined from analysis of mass spectral peak intensities in multicomponent mixtures, under the assumption that ionization efficiencies remain constant on a per base basis and that entropy changes for monomer/dimer equilibria are the same as those for dimer/trimer equilibria; the appearance of two values in Table 3 prior to the estimated error of  $\pm 4$  kcal mol<sup>-1</sup> reflects different choices of dimer equilibrium constants to be used in analyzing the peak intensity ratios.

Agreement between the *mPWPW91/MIDI!* level and experiment is quite poor, with the theoretical values inevitably being considerably below the experimental ones. However, even without taking account of the computed trimerization energies, the experimental values seem difficult to rationalize. Thus, for instance, Yanson et al. report the CCC trimerization enthalpy to be in the range of  $-35$  kcal mol<sup>-1</sup>. If one ignores any special three-body term, this would imply an interaction enthalpy of  $-12$  kcal mol<sup>-1</sup> between each pair of cytosine molecules. While at first this might seem plausible, given that the CC dimerization enthalpy is  $-16$  kcal mol<sup>-1</sup>, it must be recalled that the CC dimer has no remaining open sites with which to hydrogen bond to another monomer.

Indeed, based on studies using cytosine monomer monomethylated at the exocyclic amino group, Yanson et al. propose a cyclic structure like that labeled CCC3 in Figure 3. This cyclic structure has only three hydrogen bonds in it, and they are not particularly ideal in a geometric sense. If we assume the interaction enthalpy for every base pair in Table 2 to be entirely associated with hydrogen bonding, we can assign an average hydrogen bond strength of about  $-6$  kcal mol<sup>-1</sup> (an intuitively reasonable value). On that basis, we would expect the interaction enthalpy in the CCC trimer to be no better than  $-18$  kcal mol<sup>-1</sup>, and possibly a bit less, because the hydrogen bonds in the tricycle cannot adopt ideal linear geometries; and, indeed, our computed interaction enthalpy for CCC1 (the trimer lacking exocyclic *N*-methylation) is only  $-14.0$  kcal mol<sup>-1</sup>; when the methyl groups are included, this value drops to  $-12.8$  kcal mol<sup>-1</sup>. In the absence of methyl groups,

the more linear trimer CCC2 (Fig. 3) shows a stronger interaction enthalpy that is near the lower error limit of the experimental range, but this structure is *not* accessible to an *N*-methylated system because it uses both amino protons of the central cytosine in hydrogen bonds. There *is* an *N*-methylated structure, CCC4 (Fig. 4), which involves the interaction of one out-of-plane cytosine with the reverse wobble cytosine dimer, which does show an interaction enthalpy more favorable than that of the cyclic trimer, namely  $-22.0$  kcal mol<sup>-1</sup>. This value is consistent with adding to the existing CC dimer a single new hydrogen bonding interaction worth  $-6$  kcal mol<sup>-1</sup>, assuming the pairing interaction enthalpy of the *N*-methylated species is the same as that of the nonmethylated base pair that has already been discussed above. There does not seem to be any way to achieve as large an interaction enthalpy as that reported experimentally.

The same situation holds true for the UUU and UUT cases, where again Yanson et al.<sup>15</sup> proposed cyclic trimer structures that would necessarily have to enjoy much stronger hydrogen bonding interactions than those found in the dimers. Calculations do not support such a supposition.

In the case of UUA, which Yanson et al.<sup>15</sup> propose to have a linear structure like those of UUA1 and UUA2 in Figure 3, the *mPWPW91/MIDI!* estimate is only just below the lower error bar [and averaging of UUA1 and UUA2 using eq. (9) very slightly improves the agreement], suggesting that experiment for this trimer may possibly be more reliable than for the homotrimers. We did locate two additional trimer geometries, UUA3 and UUA4 (Fig. 3), that are more cluster-like than linear, but their interaction enthalpies are sufficiently less favorable than UUA1 and UUA2 that they are not expected to contribute significantly to a 298 K equilibrium population.

Beyond the analysis offered above, some support for the accuracy of the theoretical predictions over the reported experimental values may be derived from comparing to other base trimers where, although experimental data are not available, higher level calculations *have* been done.<sup>36–38</sup> In particular, for TAT and CGC(H)+ (structures in Fig. 3), the agreement in Table 3 between *mPWPW91/MIDI!* and multistep MP2 calculations is reasonably good, with no apparent tendency to grossly underestimate the interaction enthalpy.

The theory/theory comparison continues to hold at the level of the GGGG tetrad, where agreement between *mPWPW91/MIDI!* and B3LYP/TZP is quantitative (Table 4). In addition, *mPWPW91/MIDI!* suggests that there is a drastic drop in the additional interaction enthalpy associated with adding a fifth guanine to the complex, consistent with the experimental observation of an equilibrium heavily dominated by the tetrad at high concentrations. It must be noted, however, that we cannot compare to any other quantitative estimate of the pentad interaction enthalpy, in part because only a level of theory as efficient as *mPWPW91/MIDI!* makes the calculation tractable.

### Comparison to Molecular Mechanics

Quantum mechanical models for base-pairing will be of primary utility when it proves desirable to have a wave function (or probability density) in hand for analysis. However, for purposes of biomolecular simulation, it is clear that standard technologies

relying on molecular mechanics force fields will not be fully supplanted by QM alternatives for a long time to come (and indeed, in the absence of accounting for dispersion in some fashion, the models presented here would be unsuitable for that purpose in any case). With this point in mind, and moreover noting the dedicatory nature of the special issue in which this article appears, it is appropriate to compare the present results with those obtained using the most popular force fields developed within Peter Kollman's group in the 1980s and 1990s.

Using the Weiner et al. force field,<sup>65</sup> Cieplak and Kollman<sup>66</sup> reported 298 K "statistical" free energies of dimerization that, lacking any thermal contributions from translation, rotation, or vibration, may be regarded as essentially potential energies of interaction. If we correct for the ideal gas translational and rotational thermal contributions associated with dimerization, and ignore changes in the vibrational partition functions (these data were not reported in the original publication<sup>66</sup>), the ATWC and CGWC dimerization enthalpies would be  $-10.9$  and  $-17.4$  kcal mol<sup>-1</sup>. These values are smaller in magnitude than those determined experimentally by 2–4 kcal mol<sup>-1</sup>, although this error might be slightly reduced if vibrational changes were to be considered, because the dimer has six more vibrational degrees of freedom than do the separated monomers.

With the second-generation Cornell et al. force field,<sup>67</sup> Cornell et al.<sup>68</sup> reported explicitly computed 298 K dimerization enthalpies of  $-10.7$ ,  $-11.2$ , and  $-24.8$  kcal mol<sup>-1</sup> for ATH, ATWC, and CGWC, respectively (these results also appear in a later article by Gould and Kollman<sup>69</sup> where the performances of Hartree–Fock theory and second-order perturbation theory for predicting these dimerization enthalpies are examined using the 6-31G(d) basis set). The Cornell et al.<sup>67</sup> force field overshoots the CGWC interaction enthalpy by about as much as it is underestimated by the Weiner et al. force field,<sup>65</sup> while the AT results are about the same. Despite these quantitative differences in specific interaction enthalpies, the Cornell et al.<sup>67</sup> force field has proven to be a particularly useful tool for biomolecular simulations,<sup>6,70</sup> and its continued utility in future applications seems assured.

## Significance

The *mPWPW91/MIDI!* level of theory predicts interaction enthalpies with a mean unsigned error of 1.3 kcal/mol compared to experiment for the AT, AU, CC, CG, TT, and UU base pairs. This level of theory also agrees well with interaction enthalpies computed from more complete levels of electronic structure theory in cases where experimental data are not available. Extension of the *mPWPW91/MIDI!* model to the prediction of trimer interaction enthalpies suggests that previously reported experimental values may be overestimated. Treating the *mPWPW91/MIDI!* structures and energies as a training set for a focused reparameterization of PM3 provides the PM3<sub>BP</sub> model, which reproduces the DFT predictions with a high degree of fidelity at a still greater savings in computational resources. As research focusing on the development of more diverse nucleotide analogs accelerates, the use of the *mPWPW91/MIDI!* and PM3<sub>BP</sub> models to economically screen optimized structures and interaction energies should prove to be especially efficient.

## Acknowledgments

Computational resources were provided by the Minnesota Supercomputing Institute.

## Supporting Information

Computed 298 K interaction enthalpies for experimentally characterized base pairs at various levels of theory, errors in computed 298 K interaction enthalpies for these base pairs as a function of B3LYP parameters, and stereostructures for all *mPWPW91/MIDI!* dimers (five pages).

## References

1. Watson, J.; Crick, H. C. *Nature* 1953, 171, 737.
2. Hobza, P.; Sponer, J. *Chem Rev* 1999, 99, 3247.
3. Sponer, J.; Hobza, P.; Leszczynski, J. In *Computational Molecular Biology*; Leszczynski, J., Ed.; Elsevier: Amsterdam, 1999; p 85. Vol. 8.
4. Hernández, B.; Luque, F. J.; Orozco, M. J. *Comp Chem* 1999, 20, 937.
5. Fonseca Guerra, C.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J Am Chem Soc* 2000, 122, 4117.
6. Cheatham, T. E., III; Kollman, P. A. *Annu Rev Phys Chem* 2000, 51, 435.
7. Beveridge, D. L.; McConnell, K. J. *Curr Opin Struct Biol* 2000, 10, 182.
8. Orozco, M.; Luque, F. J. *Chem Rev* 2000, 100, 4187.
9. Sponer, J.; Leszczynski, J.; Hobza, P. *J Mol Struct (Theochem)* 2001, 573, 43.
10. Sherer, E. C.; Cramer, C. J. *J Comp Chem* 2001, 22, 1167.
11. Khandogin, J.; Hu, A. G.; York, D. M. *J Comp Chem* 2000, 21, 1562.
12. Segal, G. A., Ed. *Semiempirical Methods of Electronic Structure Calculation, Parts A and B*; Plenum: New York, 1977.
13. Stewart, J. J. P. *J Comp Chem* 1989, 10, 209.
14. Sukhodub, L. F.; Yanson, I. K. *Nature* 1976, 264, 245.
15. Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* 1979, 18, 1149.
16. Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor Chim Acta* 1996, 93, 281.
17. Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
18. Becke, A. D. *Phys Rev A* 1988, 38, 3098.
19. Lee, C.; Yang, W.; Parr, R. G. *Phys Rev B* 1988, 37, 785.
20. Burke, K.; Perdew, J. P.; Wang, Y. In *Electronic Density Functional Theory. Recent Progress and New Directions*; Dobson, J. F.; Vignale, G.; Das, M. P., Eds.; Plenum Press: New York, 1998, p 81.
21. Perdew, J.; Wang, Y. *Phys Rev B* 1992, 45, 13244.
22. Adamo, C.; Barone, V. *J Chem Phys* 1998, 108, 664.
23. Becke, A. D. *J Chem Phys* 1993, 98, 5648.
24. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J Phys Chem* 1994, 98, 11623.
25. Slater, J. C. *Quantum Theory of Molecules and Solids, Vol. 4: The Self-Consistent Field for Molecules and Solids*; McGraw-Hill: New York, 1974.
26. Vosko, S. H.; Wilks, L.; Nussair, M. *Can J Phys* 1980, 58, 1200.
27. Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; Wiley: Chichester, 2002.
28. Spirko, V.; Sponer, J.; Hobza, P. *J Chem Phys* 1997, 106, 1472.
29. Scott, A. P.; Radom, L. *J Phys Chem* 1996, 100, 16502.

30. Xantheas, S. S. *J Chem Phys* 1996, 104, 8821.
31. Dewar, M. J. S.; Thiel, W. *J Am Chem Soc* 1977, 99, 4899.
32. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J Am Chem Soc* 1985, 107, 3902.
33. York, D. M.; Sherer, E. C.; Cramer, C. J., to be submitted.
34. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98 (Revision A.10)*; Gaussian, Inc.: Pittsburgh, PA, 2001.
35. Thiel, W. *MNDO97*; Universität Zürich: Zürich, Switzerland, 1998.
36. Sponer, J.; Burda, J. V.; Mejzlik, P.; Leszczynski, J.; Hobza, P. *J Biomol Struct Dynam* 1997, 14, 613.
37. Venkateswarlu, D.; Leszczynski, J. *J Phys Chem A* 1999, 103, 3489.
38. Pundlik, S. S.; Gadre, S. R. *J Phys Chem B* 1997, 101, 9657.
39. Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc Natl. Acad Sci USA* 1962, 48, 2013.
40. Phillips, K.; Dauter, Z.; Murchie, A. I. H.; Lilley, D. M. J.; Luisi, B. *J Mol Biol* 1997, 273, 171.
41. Simonsson, T. *J Biol Chem* 2001, 382, 621.
42. Gu, J.; Leszczynski, J.; Bansal, M. *Chem Phys Lett* 1999, 311, 209.
43. Gu, J.; Leszczynski, J. *J Phys Chem A* 2000, 104, 1898.
44. Gu, J.; Leszczynski, J. *J Phys Chem A* 2000, 104, 6308.
45. Gu, J.; Leszczynski, J. *J Phys Chem A* 2000, 104, 7353.
46. Gu, J.; Leszczynski, J. *J Phys Chem A* 2001, 105, 10366.
47. Gu, J.; Leszczynski, J. *Chem Phys Lett* 2001, 335, 465.
48. Meyer, M.; Steinke, T.; Brandl, M.; Suhnel, J. *J Comp Chem* 2001, 22, 109.
49. Meyer, M.; Schneider, C.; Brandl, M.; Suhnel, J. *J Phys Chem A*, to appear.
50. Gogonea, V.; Merz, K. M. *J Phys Chem A* 1999, 103, 5171.
51. van der Vaart, A.; Gogonea, V.; Dixon, S. L.; Merz, K. M. *J Comp Chem* 2000, 21, 1494.
52. Frey, M. N.; Koetzle, T. F.; Lehmann, M. S.; Hamilton, W. C. *J Chem Phys* 1973, 59, 915.
53. Seeman, N. C.; Rosenberg, J. M.; Suddath, F. L.; Kim, J. J. P.; Rich, A. *J Mol Biol* 1976, 104, 109.
54. Rosenberg, J. M.; Seeman, N. C.; Day, R. O.; Rich, A. *J Mol Biol* 1976, 104, 145.
55. Fonseca Guerra, C.; Bickelhaupt, F. M. *Angew Chem Int Ed Engl* 1999, 38, 2942.
56. Brameld, K.; Dasgupta, S.; Goddard, W. A., III. *J Phys Chem B* 1997, 101, 4851.
57. Kratochvil, M.; Sponer, J.; Hobza, P. *J Am Chem Soc* 2000, 122, 3495.
58. Ben-Naim, A. *Statistical Thermodynamics for Chemists and Biochemists*; Plenum: New York, 1992.
59. Hobza, P.; Kabelác, M.; Sponer, J.; Mejzlik, P.; Vondrásek, J. *J Comp Chem* 1997, 18, 1136.
60. Sokalski, W. A.; Roszak, S. M. *J Mol Struct (Theochem)* 1991, 234, 387.
61. Sponer, J.; Leszczynski, J.; Hobza, P. *J Comp Chem* 1996, 17, 841.
62. Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J Chem Phys* 2001, 114, 5149.
63. Kurita, N.; Araki, M.; Nakao, K.; Kobayashi, K. *Int J Quantum Chem* 2000, 76, 677.
64. Kurita, N.; Kobayashi, K. *Comput Chem* 2000, 24, 351.
65. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J Comp Chem* 1986, 7, 230.
66. Cieplak, P.; Kollman, P. A. *J Am Chem Soc* 1988, 110, 3734.
67. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J Am Chem Soc* 1995, 117, 5179.
68. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J Am Chem Soc* 1993, 115, 9620.
69. Gould, I. R.; Kollman, P. A. *J Am Chem Soc* 1994, 116, 2493.
70. Cheatham, T. E., III; Kollman, P. A. In *Structure, Motion, Interaction, and Expression of Biological Macromolecules*; Sarma, R. H.; Sarma, M. H., Eds.; Adenine Press: New York, 1998; p 99.
71. Sponer, J.; Leszczynski, J.; Hobza, P. *J Phys Chem* 1996, 100, 1965.
72. Li, J.; Xing, J.; Cramer, C. J.; Truhlar, D. G. *J Chem Phys* 1999, 111, 885.