JCTC Journal of Chemical Theory and Computation

A Semiempirical Quantum Model for Hydrogen-Bonded Nucleic Acid Base Pairs

Timothy J. Giese, Edward C. Sherer, Christopher J. Cramer, and Darrin M. York*

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

Received April 15, 2005

Abstract: An exploratory semiempirical Hamiltonian ($PM3_{BP}$) is developed to model hydrogen bonding in nucleic acid base pairs. The $PM3_{BP}$ Hamiltonian is a novel reparametrization of the PM3 Hamiltonian designed to reproduce experimental base pair dimer enthalpies and high-level density-functional results. The parametrization utilized a suite of integrated nonlinear optimization algorithms interfaced with a d-orbital semiempirical program. Results are compared with experimental values and with benchmark density-functional (*m*PWPW91/MIDI!) calculations for hydrogen-bonded nucleic acid dimers and trimers. The $PM3_{BP}$ Hamiltonian is demonstrated to outperform the AM1, PM3, MNDO, and MNDO/H Hamiltonians for dimer and trimer structures and interaction enthalpies and is shown to reproduce experimental dimer interaction enthalpies that rival density-functional results for an over 3 orders of magnitude reduction in computational cost. The tradeoff between a high accuracy gain for hydrogen bonding at the expense of sacrificing some generality is discussed. These results provide insight into the limits of conventional semiempirical forms for accurate modeling of biological interactions.

1. Introduction

The accurate calculation of the electronic structure and associated properties of biomolecules remains an important challenge in computational biochemistry.[1] Biological processes are often mediated by a delicate balance of subtle and highly specific molecular interactions that allow the myriad of cellular events to proceed under physiological conditions. It is a goal of applied quantum chemistry to provide accurate, robust methods to model these interactions that include specific binding and recognition events as well as complex catalytic reaction mechanisms.[2−10] Unfortunately, for many biological applications, accurate ab initio methods are thwarted by the computational cost associated with the inherently large system size, broad temporal domain, or high degree of phase-space sampling required by the problem. A pragmatic alternative is to take recourse into empirical or semiempirical quantum methods that are able to provide accuracy that often surpasses low-level ab initio methods[11] for a fraction of the computational cost.

Semiempirical quantum methods have traditionally not been considered to be of sufficient accuracy for biological chemistry, largely because their development has focused on more general ground-state thermochemical applications.[12,13] Because of their immense computational advantage, there has been a recent resurgence in interest to develop new semiempirical quantum models[14,15,17,18] specifically designed to provide high accuracy for biological reactions[19] and that can be used with linear-scaling electronic structure[20,21] and implicit solvent methods[22,23] as well as hybrid quantum mechanical/molecular mechanical (QM/MM) simulations.[24,25]

The interaction of nucleic acid bases in DNA and RNA structures plays an integral role in macromolecular structure and function.[26,27] Nucleic acid bases can interact via specific hydrogen-bonding arrangements and aromatic base stacking.[28] These interactions have been an area of intense investigation both experimentally and with electronic structure methods.[29] Hydrogen-bonding interactions between nucleic acid base pairs is vital to the integrity of duplex DNA and responsible for the transfer of genetic information. An accurate description of nucleic acid base pairs requires a proper description of the dipole moments and delocalization

* Author to whom correspondence should be addressed. E-mail: york@chem.umn.edu

10.1021/ct050102l CCC: $30.25
Published on Web 09/13/2005

of π bonds of the individual bases, and of intermolecular hydrogen bonding.[30] These features are not adequately reproduced by any of the standard semiempirical models.[31−33]

In this paper, an exploratory PM3$_{BP}$ Hamiltonian is developed specifically for hydrogen bonding in nucleic acid base pairs. The purpose of this paper is to explore the parametrizational limits of existing Hamiltonian forms in adequately modeling biologically relevant interactions. This is a key step toward the development of simple quantum Hamiltonian models that provide accuracy comparable to the highest feasible ab initio methods for biomolecules and, therefore, can be readily extended to linear-scaling quantum calculations[34−36] or hybrid QM/MM simulations.[37,38] Achievement of this goal would represent a major advance in the modeling of important biological reactions. With careful parametrization of the semiempirical PM3$_{BP}$ Hamiltonian, accuracy comparable to density-functional theory results are obtained with over 3 orders of magnitude less computational cost. The results presented here demonstrate promise for the future development of extremely fast quantum models especially designed for biological systems.

2. Background

The formalism for the electronic part of the MNDO,[14,39,40] AM1,[41] PM3,[12,42] and MNDO/H[43] Hamiltonians is based on the neglect of the diatomic differential overlap (NDDO) approximation and is identical for all the methods (see ref 44 for an overview). The four Hamiltonians differ only in the way core−core repulsions are treated. In the MNDO method, the repulsion between two nuclear cores (A and B) is calculated as

$$E_N^{MNDO}(A,B) = Z'_A Z'_B \langle s_A s_A | s_B s_B \rangle \, (1 + e^{-\alpha_A R_{AB}} + e^{-\alpha_B R_{AB}}) \quad (1)$$

where Z'_A and Z'_B are the effective nuclear charges (nuclear charge minus number of core electrons), $\langle s_A s_A | s_B s_B \rangle$ is a Coulomb repulsion integral between an *s*-symmetry orbital centered on A and an *s*-symmetry orbital centered on B, and α_A and α_B are parameters in the exponential term that account for decreased screening of the nucleus by the electrons at small interatomic distances. For O−H and N−H bonds, a modified form of the screening term is used

$$E_N^{MNDO}(A,H) = Z'_A Z'_H \langle s_A s_A | s_H s_H \rangle \, (1 + R_{AH} e^{-\alpha_A R_{AH}} + e^{-\alpha_H R_{AH}}) \quad (2)$$

For many intermolecular interactions, particularly hydrogen bonds, the MNDO model is problematic and often incorrectly predicts essentially unbound hydrogen-bonded complexes. The PM3 and AM1 models include a set of Gaussian core−core terms that alleviate excessive repulsion at close range and offer significant improvement for intermolecular interactions. The modified core−core term takes the form

$$E_N^{AM1/PM3}(A,B) = E_N^{MNDO}(A,B) + \frac{Z'_A Z'_B}{R_{AB}} \left(\sum_k a_{kA} e^{-b_{kA}(R_{AB} - c_{kA})^2} + \sum_k a_{kB} e^{-b_{kB}(R_{AB} - c_{kB})^2} \right) \quad (3)$$

These terms considerably improve the description of hydrogen bonds; although, they are, in general, still considerably underbound. Alternatively, one could substitute the Gaussian core−core terms by other functions[45,46] or introduce new functional forms to the Hamiltonians.[14,18] A promising approach is to design new semiempirical methods based on density-functional theory, such as the SCC-DFTB method.[47]

The MNDO/H Hamiltonian is a modification of the MNDO Hamiltonian, where nuclear repulsion in bonds of the type A···H taking part in hydrogen bonds A···H−D (A, D = N, O, F) takes the form

$$E_N^{MNDO/H}(A,H) = Z'_A Z'_H \langle s_A s_A | s_H s_H \rangle \, (1 + e^{-\alpha R_{AH}^2}) \quad (4)$$

where α was proposed[43] to equal 2.0 Å^{-2}. As part of the MNDO/H modification to MNDO, the user must choose which pairs A···H take part in the formation of hydrogen bonds. We have chosen to use the default settings as implemented in the MNDO97 program;[48] that is, the minimum and maximum A···H distances were chosen as 1.1 and 5.0 Å, respectively, and a minimum A···H−D angle of 90 degrees was selected.

Other successful Hamiltonian forms of note, although not directly compared against here, include the PDDG/PM3 and PDDG/MNDO Hamiltonians, which employ pairwise distance-dependent Gaussian core−core terms,[49,50] the AM1/d model for molybdenum with bond-specific (i.e., pairwise) core−core exponential repulsion terms,[51] a redefinition of core−core terms for hydrogen-bonded systems,[45,46] the use of bond-based corrections for improving heats of formation,[52] and potential energy scaling procedures.[53]

3. Methods

This section describes the methods used to develop the semiempirical PM3$_{BP}$ model that is subsequently analyzed and tested. The first subsection describes the quantum data set used as the reference data to fit the semiempirical PM3$_{BP}$ parameters for nucleic acid base pairs. The second subsection describes the details of the parametrization procedure itself.

3.1. Quantum Dataset for Nucleic Acid Base Pairs. The quantum reference data set employed here to parametrize the new semiempirical method has been described in detail elsewhere[54] and is briefly summarized here. Geometries were optimized with the Kohn−Sham density-functional theory (DFT) method using the *m*PWPW91 exchange-correlation functional[55,56] with the MIDI! basis set.[57] Stationary points were verified to be minima through standard frequency calculations (positive Hessian eigenvalues for all vibrational modes) that were also used (unscaled) to calculate zero-point and thermal contributions to the gas-phase enthalpy at 298.15 K and 1 atm. Basis set superposition errors were corrected using the procedure of Xantheas.[58] All ab initio calculations were performed with the Gaussian 98 suite of programs.[59] The quantum reference data[54] calculated at this level will henceforth be designated as "*m*PWPW" in the tables and text. The interaction enthalpies obtained from the quantum data set were previously demonstrated[54] to compare favorably with available experimental[60,61] values for AT, GC, UU, CC, TT, and AU pairs and also with computations[62] for a large

number of other base pairs carried out with larger basis sets and more complete levels of electronic structure theory and were, thus, deemed to be an appropriate and convenient test set against which to parametrize the semiempirical model. Formally, the contributions to the experimental interaction enthalpies require sampling of all relevant conformations. In the present work, calculated enthalpies are based on the single, lowest-energy ab initio configuration, as in other work,[54] or on a Boltzmann-weighted average.

3.2. Semiempirical Parametrization Procedure. This section describes the $PM3_{BP}$ parametrization procedure for nucleic acid base pairs based on the density-functional quantum data set described in the previous section. The first step is to construct an appropriate $\chi^2(\boldsymbol{\lambda})$ merit function that measures the goodness of fit of a set of molecular properties, calculated with a set (vector) of semiempirical parameters $\boldsymbol{\lambda}$, with the corresponding reference values. The second step is to use nonlinear optimization methods to find a suitable set of parameters by minimization of the $\chi^2(\boldsymbol{\lambda})$ merit function.

3.2.1. Construction of the $\chi^2(\boldsymbol{\lambda})$ Merit Function. The form of the $\chi^2(\boldsymbol{\lambda})$ merit function used in this work is given by

$$\chi^2(\boldsymbol{\lambda}) = \sum_{i}^{\text{mol}} \sum_{\alpha}^{\text{prop}} w_{i\alpha}[Y_{i\alpha}^{\text{PM3}_{\text{BP}}}(\boldsymbol{\lambda}) - Y_{i\alpha}^{\text{REF}}]^2 \quad (5)$$

$$w_{i\alpha} = (\sigma_{i\alpha}{}^2)^{-1} \quad (6)$$

where the first sum with index i in eq 5 runs over molecules (or complexes) and the second sum with index α runs over properties of the molecule (or complex). The argument λ represents a trial set of $PM3_{BP}$ parameters that are the variational degrees of freedom, $Y_{i\alpha}^{\text{PM3}_{\text{BP}}}(\boldsymbol{\lambda})$ is the value of the property α for molecule (complex) i calculated with the trial parameter set λ, $Y_{i\alpha}^{\text{REF}}$ is the corresponding reference value (taken either from the experiment or calculated with DFT), and $w_{i\alpha}$ is the associated least-squares weight in the fitting. The weights $w_{i\alpha}$ are proportional to the inverse square of the $\sigma_{i\alpha}$ values in eq 6. The $\sigma_{i\alpha}$ values have the same units as the molecular property to which they are associated and control the sensitivity of the merit function to deviations of that property from the reference value. The properties contained in the χ^2 merit function, the number of reference data, and the σ weight values for each property are summarized in Table 1.

For the semiempirical calculations, a modified version of the MNDO97[48] program was used. The properties considered include relative energies, optimized bond lengths, angles, torsions, and dipole moments for neutral species. Each structure of the data set is fully optimized at the semiempirical level for a given set of parameters before calculating these properties and constructing the $\chi^2(\boldsymbol{\lambda})$ function. Previous work in the development of specific reaction parameter Hamiltonians did not perform geometry optimization but, instead, performed single-point calculations at a stationary point and penalized the norm of the gradient in the $\chi^2(\boldsymbol{\lambda})$ function. This procedure works well for very simple molecules and reactions with only small allowable variations in a few semiempirical parameters.[63] This is not a productive strategy in the present case. The large number of degrees of freedom make it extremely difficult to lock down the norm of the gradient to a sufficient degree so that they accurately reflect the energies and geometries associated with the fully optimized geometries, especially when kcal/mol accuracy is the primary goal.

Table 1. Reference Data Contained in the χ^2 Merit Function[a]

description	N	σ	unit
bond lengths	999	0.005	Å
bond angles	1518	2.000	deg
bond torsion angles	1075	5.000	deg
dipole moments (*m*PWPW)	36	0.040	D
dipole moments (exptl)	5	0.040	D
H-bond distances	65	0.020	Å
H-bond angles	65	0.040	deg
intermolecular heavy atom distances	65	0.020	Å
dimerization enthalpies (*m*PWPW)	18	0.500	kcal/mol
dimerization enthalpies[b] (exptl)	11	0.500	kcal/mol
conformationally relative dimerization enthalpies[c] (*m*PWPW)	30	0.250	kcal/mol

[a] The terms exptl and *m*PWPW refer to available experimental and DFT data, respectively. If not explicitly specified, the reference data refers to the use of DFT reference data. N is the number of reference data points for the given property, and σ is the associated weight within the χ^2 definition. [b] A dimerization enthalpy is defined as the difference in enthalpy between a dimer and isolated monomers. Only six unique experimental values were used in the parametrization but appear more than once in the χ^2 definition since the experimental numbers are not orientationally distinguishable. [c] The conformationally relative dimerization enthalpies are defined here as the difference in dimerization enthalpies between base-pair arrangements, for example, the difference in dimerization enthalpy between AT_{WC} and AT_{RWC}.

3.2.2. Nonlinear Optimization of the $\chi^2(\boldsymbol{\lambda})$ Merit Function. Semiempirical parameters were obtained by optimization of the $\chi^2(\boldsymbol{\lambda})$ merit function of eq 5 with respect to the set of semiempirical parameters $\boldsymbol{\lambda}$ for H, N, and O atoms (parameters for C atoms were held fixed to the PM3 values). For this purpose, a suite of integrated nonlinear optimization methods for semiempirical parameter development has been used. The details of the integrated suite are forthcoming;[64] a brief overview of the algorithms is provided here.

Three nonlinear optimization methods working in concert were applied in the present work: (1) genetic algorithm, (2) Monte Carlo simulated annealing, and (3) direction set minimization methods. Genetic algorithms[65,66] have been demonstrated elsewhere to be useful in semiempirical parameter optimization.[51,63,67,68] The implementation of the genetic algorithm was loosely based on the description by Goldberg[65] and tailored for the several issues encountered in the semiempirical optimization application. A new method of partitioning subsets of the population (referred to as "tribes") into different local minima on the $\chi^2(\boldsymbol{\lambda})$ surface called *fitness-weighted eigenvector niching* was employed. The fitness of members was determined from a Gaussian distribution of the χ^2 values of population members within a "tribe" with the Gaussian width proportional to the χ^2 variance. The full details of the method are described elsewhere.[64]

Table 2. Parameters in the PM3 and $PM3_{BP}$ Hamiltonians[a]

parameter	H	C	N	O
U_{ss} (eV)	−12.755 118 80	−47.270 320 00	−48.794 933 85	−87.387 093 24
	−13.073 321 00	*−47.270 320 00*	*−49.335 672 00*	*−86.993 002 00*
U_{pp} (eV)		−36.266 918 00	−46.579 459 03	−71.702 685 70
		−36.266 918 00	*−47.509 736 00*	*−71.879 580 00*
β_s (eV)	−4.878 234 60	−11.910 015 00	−14.338 846 65	−46.877 410 23
	−5.626 512 00	*−11.910 015 00*	*−14.062 521 00*	*−45.202 651 00*
β_p (eV)		−9.802 755 00	−19.308 628 53	−24.742 325 18
		−9.802 755 00	*−20.043 848 00*	*−24.752 515 00*
α (eV)	3.356 386 00	2.707 807 00	2.830 545 00	3.217 102 00
	3.356 386 00	*2.707 807 00*	*2.830 545 00*	*3.217 102 00*
H_{sp} (eV)		2.290 980 00	1.136 713 00	0.593 883 00
		2.290 980 00	*1.136 713 00*	*0.593 883 00*
G_{ss} (eV)	15.023 337 45	11.200 708 00	12.415 221 41	15.261 643 45
	14.794 208 00	*11.200 708 00*	*11.904 787 00*	*15.755 760 00*
G_{pp} (eV)		10.796 292 00	13.966 114 12	13.659 300 75
		10.796 292 00	*11.754 672 00*	*13.654 016 00*
G_{sp} (eV)		10.265 027 00	7.345 402 26	10.413 326 25
		10.265 027 00	*7.348 565 00*	*10.621 160 00*
G_{p2} (eV)		9.042 566 00	10.410 219 25	12.420 510 17
		9.042 566 00	*10.807 277 00*	*12.406 095 00*
ζ_s ($Å^{-1}$)	0.967 807 00	1.565 085 00	2.028 094 00	3.796 544 00
	0.967 807 00	*1.565 085 00*	*2.028 094 00*	*3.796 544 00*
ζ_p ($Å^{-1}$)		1.842 345 00	2.313 728 00	2.389 402 00
		1.842 345 00	*2.313 728 00*	*2.389 402 00*
a_1 (unitless)	1.121 725 94	0.050 107 00	1.501 671 53	−1.131 176 77
	1.128 750 00	*0.050 107 00*	*1.501 674 00*	*−1.131 128 00*
b_1 ($Å^{-2}$)	5.095 167 07	6.003 165 00	5.903 991 75	6.009 998 15
	5.096 282 00	*6.003 165 00*	*5.901 148 00*	*6.002 477 00*
c_1 (Å)	1.536 937 00	1.642 214 00	1.710 426 69	1.607 311 00
	1.537 465 00	*1.642 214 00*	*1.710 740 00*	*1.607 311 00*
a_2 (unitless)	−1.064 925 25	0.050 733 00	−1.515 716 18	1.130 989 09
	−1.060 329 00	*0.050 733 00*	*−1.505 772 00*	*1.137 891 00*
b_2 ($Å^{-2}$)	6.023 153 66	6.002 979 00	5.975 794 98	5.872 165 45
	6.003 788 00	*6.002 979 00*	*6.004 658 00*	*5.950 512 00*
c_2 (Å)	1.571 307 32	0.892 488 00	1.710 935 13	1.603 474 21
	1.570 189 00	*0.892 488 00*	*1.716 149 00*	*1.598 395 00*

[a] Standard notation for parameters taken from refs 42 and 70. The original PM3 parameters are shown in italics immediately below the $PM3_{BP}$ values. Note: the parameters for C were held fixed to the standard PM3 values.

Several genetic algorithm runs were performed, with the number of generations ranging from 50 to 200 using a population of 64−128 members. The final population from the genetic algorithm optimization was then passed to a Monte Carlo simulated annealing procedure. The Monte Carlo procedure[69] used multidimensional simplex moves and variable exponentially decaying annealing schedules to explore the local region of parameter space around the final population provided by the genetic algorithm. The resulting parameters were then passed to a quadratically convergent direction set optimization method[69] to arrive at the final optimized $PM3_{BP}$ parameter set (Table 2). Recently, these methods have been extended and improved to make the parametrization more (although not completely) automated and robust, a detailed description of which is forthcoming.[64]

4. Results and Discussion

This section presents results and compares the performance of semiempirical Hamiltonian models with respect to experimental and density-functional calculations for nucleic acid base dimers and trimers. A detailed comparison and extended discussion of nucleic acid base monomer geometries and dipole moments are provided in the Supporting Information. The semiempirical methods include the new $PM3_{BP}$ method of the present work and the conventional semiempirical AM1,[41] PM3,[12,42] MNDO,[39,40] and MNDO/H[43] Hamiltonian models. The error metrics (error = calculated − experimental/reference value) shown in the tables are the maximum (signed) error (MAXE), root-mean-square error (RMSE), mean unsigned error (MUE), and mean signed error (MSE).

4.1. Relationship between the PM3 and $PM3_{BP}$ Parameters. Overall, the $PM3_{BP}$ parameters do not change dramatically from the PM3 parameters that were the starting point for optimization (Table 2). Note that the parameters for carbon in the $PM3_{BP}$ method were held fixed to the PM3 values, as were the H_{sp} parameters for each atom. For hydrogen, the greatest change occurs for the β_s and G_{ss} parameters that were shifted from the PM3 values in the positive direction by 0.75 and 0.21 eV, respectively. Similarly

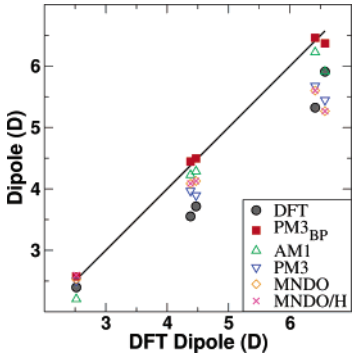

Figure 1. Regression of semiempirical and DFT *m*PWPW[54] dipole moments for nucleic acid bases with B3LYP/cc-pVTZ[71] (*x*-axis reference) values. A linear fit for each method produces intercept (*b*), slope (*m*), and correlation coefficient (*c*) values of DFT: $b = 0.253$ D, $m = 0.809$, $c = 0.982$; $PM3_{BP}$: $b = 0.081$ D, $m = 0.981$, $c = 1.000$; AM1: $b = 0.116$ D, $m = 0.918$, $c = 0.996$; PM3: $b = 0.289$ D, $m = 0.818$, $c = 0.988$; MNDO: $b = 0.349$ D, $m = 0.807$, $c = 0.985$; and MNDO/H: $b = 0.367$, $m = 0.805$, $c = 0.985$.

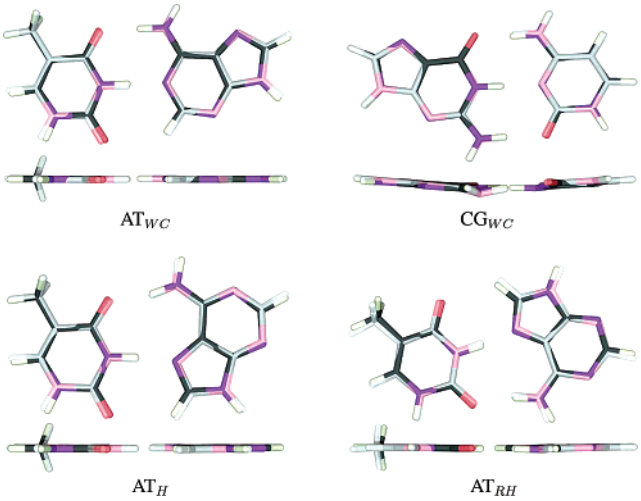

Figure 2. Superimposed root-mean-squared fit of $PM3_{BP}$ (lighter colors) geometry optimized structures to DFT *m*PWPW[54] structures (darker colors) for AT_{WC} (upper left), CG_{WC} (upper right), AT_{H} (lower right), and AT_{RH} (lower left). Each pane shows the face-on view (upper) and side view (lower).

for nitrogen and oxygen, the β_s and G_{ss} parameters also exhibited significant change; however, the β_s parameter was shifted toward slightly more negative values, whereas the G_{ss} parameter was shifted by +0.51 eV for oxygen and −0.49 eV for nitrogen. For nitrogen, the β_p and G_{pp} parameters exhibited even more pronounced change from the PM3 values than the β_s and G_{ss} parameters (+0.74 and 2.21 eV, respectively). Aside from these parameters, the $PM3_{BP}$ parameters deviated from the PM3 values by typically only a few percent or less.

4.2. Nucleic Acid Base Monomers. In this section, a comparison of the internal geometry and dipole moments for cytosine, guanine, adenine, thymine, and uracil nucleotide bases is briefly summarized. An extended discussion can be found in the Supporting Information. All of the semiempirical methods perform reasonably well for the internal geometries of the base monomers with respect to the *m*PWPW results of Sherer et al.,[54] with the $PM3_{BP}$ method performing best overall.

The semiempirical dipole moments for the nucleic acid bases are compared with the density-functional calculations of Sherer et al.,[54] and the high basis-set level (B3LYP/cc-pVTZ) calculations of Li et al.[71] are illustrated in Figure 1. The $PM3_{BP}$ method performs the best with respect to the high basis DFT results, with a RMSE of 0.101 D. The MNDO/H method has the largest RMSE (0.717 D) of the semiempirical methods. It is of interest to note that the DFT dipole moments calculated with the smaller basis set (*m*PWPW) have a RMSE (0.761 D), with respect to the higher basis-set (B3LYP/cc-pVTZ) values, that is larger than any of the semiempirical RMSE values.

4.3. Nucleic Acid Base Dimers. The main focus of this paper is on the structure and binding enthalpy of hydrogen-bonded nucleic acid base pairs. A host of standard and nonstandard base-pairing interactions were considered, including Watson−Crick (WC), reverse Watson−Crick (RWC), Hoogstein (H), reverse Hoogstein (RH), and mismatched base pairs. Figure 2 illustrates various base pair geometries (*m*PWPW geometries) for a representative subset of hydrogen-bonded dimers ($PM3_{BP}$ geometries are RMS overlaid onto the DFT structures of Sherer et al.;[54] figures of the DFT structures and their relation to the nomenclature are found within the appendix of ref 54). Subsection 4.3.1 provides a brief summary comparison of the semiempirical results for gas-phase intermolecular hydrogen-bonding and binding enthalpies with recent density-functional calculations,[54] a full discussion of which is provided in the Supporting Information. Subsection 4.3.2 compares density-functional and semiempirical hydrogen-bond lengths and dimerization enthalpies to experimental values. Subsection 4.3.3 examines an adiabatic binding potential energy curve for a representative hydrogen-bonded base pair and addresses potential problems associated with the use of Gaussian core−core functions.

4.3.1. Comparison with Density-Functional Calculations. This subsection provides a brief summary of the comparison results for the set of 31 hydrogen-bonded nucleic acid base dimers calculated with the semiempirical and density-functional (*m*PWPW) quantum models. An extended discussion and presentation of data is provided in the Supporting Information and has also been discussed, in part, elsewhere.[54]

The MNDO Hamiltonian does not predict stable hydrogen bonds, MNDO/H forms hydrogen-bond lengths that are too short, AM1 predicts hydrogen-bond lengths that are too long, and PM3 and $PM3_{BP}$ predict hydrogen-bond lengths that agree the most closely with the *m*PWPW values of any of the other semiempirical models considered. Results for the hydrogen-bond angles are qualitatively similar in that errors are most significant for the MNDO and AM1 methods. A comparison of semiempirical and *m*PWPW dimerization enthalpies shows that MNDO is critically underbound; AM1 and PM3 are significantly underbound by over 5 kcal/mol on average, whereas MNDO/H predicts dimers that are

Table 3. Comparison of Semiempirical and DFT Binding Enthalpies with Experimental Values for Nucleic Acid Base Dimers[a]

molecule	exptl	*m*PWPW	$PM3_{BP}$	AM1	PM3	MNDO	MNDO/H
CG_{WC}	−21.0	−22.4	−21.4	−13.8	−11.8	−3.9	−29.2
AT_{WC}		−11.3	−12.4	−4.9	−5.8	−0.6	−17.7
AT_{RWC}		−10.6	−12.5	−4.7	−5.9	−0.4	−17.1
AT_{H}		−11.2	−13.6	−4.9	−6.8	−0.9	−17.4
AT_{RH}		−10.7	−13.7	−5.0	−6.9	−1.0	−17.2
AT*	−13.0	−11.1	−13.5	−4.9	−6.7	−0.8	−17.4
CC	−16.0	−17.0	−18.8	−7.5	−9.4	−3.4	−26.6
TT_1		−8.6	−8.7	−5.9	−4.6	−0.5	−13.4
TT_2		−9.4	−8.4	−6.0	−4.4	−0.4	−13.8
TT_3		−7.9	−8.9	−5.9	−4.7	0.2	−13.0
TT*	−9.0	−9.2	−8.7	−5.9	−4.6	−0.4	−13.6
UU_1		−8.3	−8.7	−6.0	−4.5	−1.8	−13.4
UU_2		−9.4	−8.7	−6.0	−4.3	−0.3	−13.8
UU_3		−7.5	−8.8	−5.9	−4.6	−0.2	−13.0
UU*	−9.5	−9.2	−8.8	−6.0	−4.5	−1.6	−13.6
AU_{WC}	−14.5	−11.4	−12.6	−4.9	−5.8	−0.4	−17.8
MSE		**0.5**	**0.1**	**6.7**	**6.7**	**12.1**	**−5.9**
MUE		**1.3**	**1.1**	**6.7**	**6.7**	**12.1**	**5.9**
RMSE		**1.6**	**1.4**	**7.1**	**6.9**	**12.5**	**6.4**
MAXE		**3.1**	**−2.8**	**9.6**	**9.2**	**17.1**	**−10.6**

[a] Comparison of binding enthalpies (kcal/mol) for nucleic acid base dimers from semiempirical ($PM3_{BP}$, AM1, PM3, MNDO, and MNDO/H) and DFT *m*PWPW[54] (*m*PWPW) calculations with experimental values[60,61] (exptl). An asterisk indicates that the value used in comparison to the experimental value is a Boltzmann-weighted average of several structures (individually listed immediately above the averaged result) at 298.15 K. Summarized at the bottom are the error metrics (bold) for the semiempirical and DFT (*m*PWPW) values with corresponding experimental results.

overbound by over 6 kcal/mol on average relative to the *m*PWPW results. The $PM3_{BP}$ dimerization enthalpies, on the other hand, are in close agreement with the *m*PWPW values with a RMSE or 1.3 kcal/mol.

4.3.2. Comparison with Experimental Values. Table 3 compares the calculated binding enthalpies with experimental values.[60,61] In some instances, a direct comparison cannot be made since the experimental values can often not distinguish between different binding orientations. In these instances, the computed value was used to compare with the experiment result from a Boltzmann-weighted average of the available minima at 298.15 K. Note that experimental measurements may contain fractions of very different binding motifs, such as stacked base interactions, which are not accounted for within the scope of the currently selected geometries. Consequently, the "MSE" and "RMSE" values reported in Table 3 must be regarded as approximate. Nonetheless, the DFT values appear to slightly underestimate the binding enthalpy (the MSE is 0.5 kcal/mol), and the overall RMSE is 1.6 kcal/mol. The MSE and RMSE values for AM1 (6.7 and 7.1 kcal/mol, respectively) and PM3 (6.7 and 7.9 kcal/mol, respectively) relative to the experimental binding enthalpies are slightly larger than the corresponding MSE and RMSE values relative to the DFT results. The MNDO/H method is still overbound with respect to experimental values, with a MSE value of −5.9 kcal/mol and RMSE of 6.4 kcal/mol. This underscores the inadequacy of these methods for biological applications where hydrogen bonding is involved. The $PM3_{BP}$ method performs best relative to the experimental binding enthalpies (Figure 3) with a MSE of only 0.1 kcal/mol and a RMSE of 1.4 kcal/mol. The largest errors occur for the CC dimer ($PM3_{BP}$ error of −2.8 kcal/mol) and the AU_{WC} dimer ($PM3_{BP}$ error of 1.9 kcal/mol). The error trends of the DFT (*m*PWPW) values with respect to those of the experiment have the same sign for these dimers (−1.0 and 3.1 kcal/mol for CC and AU_{WC}, respectively).

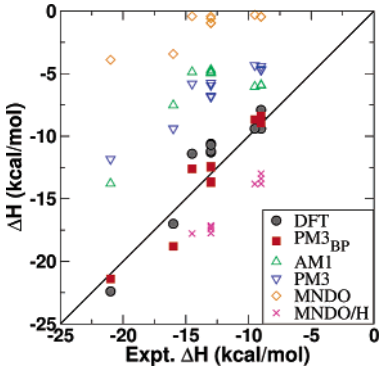

Figure 3. Regression of semiempirical and DFT *m*PWPW[54] binding enthalpies for nucleic acid base dimers with experimental[60,61] (*x*-axis reference) values. A linear fit for each method produces intercept (*b*), slope (*m*), and correlation coefficient (*c*) values of DFT: $b = 2.009$ kcal/mol, $m = 1.086$, $c = 0.940$; $PM3_{BP}$: $b = 1.510$ kcal/mol, $m = 1.117$, $c = 0.965$; AM1: $b = -0.006$ kcal/mol, $m = 0.495$, $c = 0.695$; PM3: $b = 1.211$ kcal/mol, $m = 0.598$, $c = 0.944$; MNDO: $b = 1.211$ kcal/mol, $m = 0.598$, $c = 0.945$; and MNDO/H: $b = 2.865$ kcal/mol, $m = 0.308$, $c = 0.850$.

The distance between the heavy atoms acting as hydrogen-bond acceptors and donors have been resolved for the AU_{WC} and CG_{WC} base pairs in sodium adenylyl-3′-5′-uridine (ApU) and sodium guanylyl-3′,5′-cytidine nonahydrate (GpC) crys-

Table 4. Comparison of Experimental, Semiempirical, and DFT Hydrogen-Bond Heavy-Atom Acceptor−Donor Separations for AU and CG Base Pairs[a]

	AU_{WC}		CG_{WC}		
method	$N_1\cdots H-N_3$	$N_6-H\cdots O_4$	$N_4-H\cdots O_6$	$N_3\cdots H-N_1$	$O_2\cdots H-N_2$
exptl	2.82	2.95	2.91	2.95	2.86
*m*PWPW	2.73 (−0.09)	2.84 (−0.11)	2.71 (−0.20)	2.83 (−0.12)	2.87 (0.01)
$PM3_{BP}$	2.77 (−0.05)	2.77 (−0.18)	2.76 (−0.15)	2.78 (−0.17)	2.80 (−0.06)
AM1	3.05 (0.23)	3.11 (0.16)	3.06 (0.15)	3.05 (0.10)	3.10 (0.24)
PM3	2.81 (0.01)	2.82 (−0.13)	2.82 (−0.09)	2.80 (−0.15)	2.85 (−0.01)
MNDO	4.22 (1.40)	5.10 (2.15)	4.22 (1.31)	4.12 (1.17)	4.09 (1.23)
MNDO/H	2.55 (−0.27)	2.54 (−0.41)	2.52 (−0.39)	2.55 (−0.40)	2.60 (−0.26)

[a] Comparison of semiempirical ($PM3_{BP}$, AM1, PM3, MNDO, and MNDO/H) and DFT *m*PWPW[54] (*m*PWPW) calculated hydrogen-bond heavy-atom acceptor−donor separations (Å) with experimental (exptl) X-ray crystal structure analysis of ApU and GpC.[26,72,73] Errors relative to experimental values are indicated in parentheses.

tals using X-ray diffraction.[26,72,73] Table 4 compares the experimental acceptor−donor distances with *m*PWPW and semiempirical methods. Reasonable agreement with experimental values is found between DFT and $PM3_{BP}$ with errors ranging from 0.01 to 0.2 Å. Both *m*PWPW and $PM3_{BP}$ are in reasonable agreement with crystallographic data with MUE values of 0.11 and 0.13 Å, respectively. PM3 agrees best with experimental values with a MUE error of 0.08 Å; however, an examination of the base pair geometries shows considerably artificial nonplanarity. The MNDO/H method (using the suggested α parameter[43] of 2.0 $Å^{-2}$) predicts acceptor−donor separations that are systematically too short by 0.29 Å.

4.3.3. Dimer Potential Energy Curve. The use of core−core functions can lead to artificial stationary points in potential energy surfaces. In this section, the adiabatic binding energy potential energy curve for a hydrogen-bonded base pair is explored to address this issue. Figure 4 displays the adiabatic binding potential energy curve for AT_{WC} (defined here as the center of mass separation between ADE and THY) for the semiempirical methods and *m*PWPW. The monomer geometries and relative orientation with respect to one another were taken from the DFT-optimized AT_{WC} structure (Figure 2).

None of the semiempirical methods show artificial stationary points in the potential energy curve. The minimum energy separations and relative binding energies are qualitatively similar to the dimer hydrogen bond lengths and dimer binding enthalpies summarized in Section 4.3 and extensively discussed in the Supporting Information. At first glance, it appears that the AM1 method has the best qualitative *shape* of the potential energy curve when compared to DFT, although severely underbound, whereas the PM3 and $PM3_{BP}$ methods have a spuriously steep potential well near the minimum. A careful comparison of these three potential energy curves beyond 6 Å reveals that their long-range attractive tails are nearly parallel, suggesting that the steep potential well near the minimum observed in the PM3 and $PM3_{BP}$ methods are due to core−core functions. A better potential energy curve might be obtained with core−core functions with smaller Gaussian exponents or having based the parametrization off of AM1 as opposed to PM3. For the design of new-generation semiempirical methods for QM/MM methods, it is likely best to avoid completely the use of off-center Gaussian core−core terms.

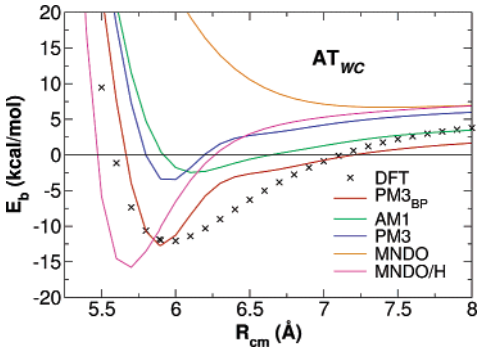

Figure 4. Binding energy of AT_{WC} as a function of rigid base-pair separation. The potential energy curve is defined as the center of mass separation between each monomer along the vector joining their center of mass. The monomer geometries and relative orientation with respect to each other were held fixed during the scan to those determined from DFT optimization of the energy minimum (see Figure 2). The DFT binding energies are counterpoise corrected, but since nonstationary points are involved, zero-point energy corrections and thermal corrections to the enthalpy were not included. The zero of energy is defined as the energy of the two isolated monomers. Since the monomers are constrained to those found in the AT_{WC} DFT optimized structure, the binding energy asymptotically reaches a positive value at large center of mass separations. The semiempirical geometries used were the same as used in the ab initio single-point calculations.

4.4. Nucleic Acid Base Trimers. The elementary next step in evaluating the limits of semiempirical methods in studying nucleic acid base interactions is to examine trimer interactions. This is an interesting test for the $PM3_{BP}$ method, since no trimer data were used in the parametrization procedure. Base pair trimers (sometimes referred to as "triplexes") have been studied in the past with Hartree−Fock[74], DFT (B3LYP[5] and *m*PWPW),[54] and MP2[75,76] methods. In addition, Yanson et al.[61] have reported trimerization enthalpies by analysis of mass spectral peak intensities in multicomponent mixtures.

Table 5 compares values for the nucleic acid base trimer binding enthalpies calculated with DFT,[54] MP2,[76] $PM3_{BP}$, AM1, PM3, MNDO, and MNDO/H. The naming of the trimers and illustrations of the DFT trimer geometries are

Table 5. Comparison of the Semiempirical and DFT Binding Enthalpies with Experimental Values for Nucleic Acid Base Trimers[a]

molecule	exptl	*m*PWPW	$PM3_{BP}$	AM1	PM3	MNDO	MNDO/H
CCC_1		−14.0	−16.8	−24.1	−8.5	−8.5	−17.4
CCC_2		−28.8	−33.4	−25.5	−17.1	−8.3	−40.2
CCC_4	−33 (−38) ± 4	−22.0	−28.9	−14.8	−13.5	−4.4	−29.1
UUA_1		−21.0	−25.2	−8.8	−12.0	−1.4	−33.9
UUA_2		−21.4	−25.1	−8.8	−11.9	−1.2	−34.1
UUA_3		−17.0	−20.5	−11.5	−10.6	−1.3	−21.9
UUA_4		−17.4	−20.6	−12.2	−11.9	−2.2	−25.0
UUA*	−27 (−29) ± 4	−21.3	−25.2	−12.0	−11.9	−1.8	−34.0
UUU_1		−8.5	−13.1	−10.0	−7.0	−2.6	−19.9
UUU_2		−11.3	−14.6	−10.3	−8.4	−0.7	−18.3
UUU*	−20 (−22) ± 4	−11.3	−14.5	−10.2	−8.3	−2.5	−19.8
UUT	−23 (−25) ± 4	−7.1	−12.7	−9.9	−6.6	−0.5	−18.6
TAT	−23.8 (MP2)	−21.3	−24.9	−8.9	−11.9	−1.0	−34.1
CGC^+	−65.2 (MP2)	−68.3	−65.2	−39.9	−46.0	−23.2	−79.8
GGG		−36.8	−33.3	−28.5	−19.6	−8.7	−40.3

[a] Comparison of binding enthalpies (kcal/mol) for nucleic acid base trimers (triplexes) from semiempirical ($PM3_{BP}$, AM1, PM3, MNDO, and MNDO/H) and DFT *m*PWPW[54] (*m*PWPW) calculations with experimental values[60,61] (exptl) or MP2/6-31G(d) calculations[76] (MP2). The MP2 results[76] involve geometry optimization, zero-point vibrational energy correction, and thermal contributions at the 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. The naming convention of the molecules follows from Sherer et al.[54] Of special note is the difference between CCC_1, CCC_2, and CCC_4: CCC_1 and CCC_2 are unmethylated cytosine triplex structures, whereas CCC_4 is a (N_1,N_4)-dimethylcytosine triplex structure. An asterisk indicates that the value is a Boltzmann-weighted average of several structures (individually listed immediately above the averaged result) at 298.15 K.

presented in ref 54. As was done with the comparisons of dimer binding enthalpies with experimental values, Boltzmann-weighted averages of the computed energies from the available geometries are used to compare to available experimental trimer enthalpy results. The different experimental values reported in Table 5 result from different choices of the dimer equilibrium constants in the analysis of the spectral intensity ratios.[61]

The AM1, PM3, and MNDO semiempirical methods considerably underestimate the binding enthalpy for all trimers when compared to experimental values, MP2, or the Boltzmann-averaged DFT data. The smallest error of all of the conventional (AM1, PM3, MNDO) semiempirical methods is with AM1, although closer inspection reveals an incorrect rank order of some of the binding enthalpies relative to that of the DFT values. The MNDO/H method, in general, is considerably overbound, but not in all cases, such as the CCC_4 trimer. The DFT values are underbound relative to those of the experiment and are typically just outside the lower bound of the experimental error. $PM3_{BP}$ is also underbound relative to experiment, but more bound than the DFT values by roughly 3 kcal/mol, which is consistent with the behavior observed with the dimers (see Supporting Information). As pointed out previously,[54] the experimentally determined binding enthalpy of (N_1,N_4)-dimethylcytosine (CCC_4) is unjustifiably overbound; a binding enthalpy of −33 kcal/mol only seems plausible if hydrogen-bonding sites are freed by demethylation of the monomers, the lowest enthalpy of which is the CCC_2 structure. In fact, the $PM3_{BP}$ binding enthalpy of the unmethylated structure exactly reproduces a binding enthalpy of −33 kcal/mol.

For structures where experimental binding enthalpies are unavailable (TAT and CGC^+), a comparison is made to counterpoise corrected MP2 enthalpies with zero-point vibrational energy and thermal corrections to the energy.[76] In both cases, $PM3_{BP}$ agrees better with the MP2 enthalpies (−23.8 and −65.2 kcal/mol for TAT and CGC^+, respectively) than the *m*PWPW[54] enthalpies. Although the DFT values are in good agreement with the MP2 enthalpies (errors of 2.5 and 3.1 kcal/mol for TAT and CGC^+, respectively), $PM3_{BP}$ agrees exceptionally well (errors of 1.1 and 0.0 kcal/mol for TAT and CGC^+, respectively).

4.5. Transferability to Molecules not in the Parametrization Set. Although it is the purpose of the present work to focus on hydrogen bonding in nucleic acid bases, it is instructive to test and compare the $PM3_{BP}$ method with a more general set of hydrogen-bonded complexes. Toward this end, a test set of molecules was considered in order to compare the ability of the $PM3_{BP}$ method and other semiempirical methods to model intermolecular hydrogen bonding between neutral molecules[31] and some biologically relevant ions. A summary of the error metric results for the dimerization enthalpies, dipole moments, and hydrogen-bond lengths relative to the *m*PWPW values is provided in Table 6, the complete set of data being provided in the Supporting Information. Overall, the $PM3_{BP}$ method makes a considerable improvement relative to the other semiempirical methods for all of these properties. For example, the RMSE values for the dimerization enthalpy, dipole moment, and hydrogen-bond distances are 1.58 kcal/mol, 0.63 D, and 0.23 Å, respectively, for $PM3_{BP}$, whereas the next lowest RMSE values from any of the other semiempirical methods are 1.91 kcal/mol (MNDO/H), 0.83 D (PM3), and 0.26 Å (PM3), respectively. These results suggest that the strategy outlined here of careful, specific reparametrization, using some consistency constraints (such as fixing the C parameters and allowing a relatively small deviation from the more general PM3 parameter values) can assist in maintaining a significant level of robustness and transferability for the properties included in the parametrization procedure.

Table 6. Comparison of Semiempirical Dimerization Enthalpy, Dipole Moment, and Hydrogen-Bond Length Error Metrics Relative to DFT for Hydrogen-Bonded Complexes Not Contained in the Parameterization[a]

property	metric	$PM3_{BP}$	AM1	PM3	MNDO/H
ΔH (kcal/mol)	MSE	0.34	0.81	1.61	−0.93
	MUE	1.11	2.37	2.04	1.63
	RMSE	1.58	3.27	2.87	1.91
	MAXE	4.47	10.45	8.65	−5.26
μ (D)	MSE	0.17	−0.36	−0.25	0.01
	MUE	0.40	0.63	0.61	0.69
	RMSE	0.63	0.90	0.83	0.91
	MAXE	1.91	−2.78	−2.43	1.93
H···X (Å)	MSE	−0.04	0.39	0.05	−0.24
	MUE	0.19	0.39	0.19	0.32
	RMSE	0.23	0.49	0.26	0.37
	MAXE	0.76	1.27	0.72	0.90

[a] The DFT reference values were obtained at the B3LYP/6-311++G(3df,2p)//B3YLP/6-31++G(d,p) level of theory with zero-point energy and thermal corrections to the enthalpy derived from a frequency analysis at the B3YLP/6-31++G(d,p) level of theory. The dimerization enthalpy (ΔH) error metrics involved the comparison of 37 dimers. The dimerization enthalpy is defined as the difference in enthalpy between the dimer and the isolated monomers. From these 37 dimers, there were 43 hydrogen-bond lengths. The hydrogen-bond lengths were measured from the hydrogen position to the heavy atom to which it is hydrogen-bonding, denoted as H···X. The dipole moment (μ) statistics involved a total of 45 *neutral* dimers and monomers. A complete comparison of the data is tabulated in the Supporting Information.

5. Conclusion

The present paper reports an exploratory semiempirical Hamiltonian ($PM3_{BP}$) for modeling hydrogen-bonded nucleic acid bases that significantly outperforms the AM1, PM3, MNDO, and MNDO/H Hamiltonians and accurately reproduces nucleic acid base pair interaction enthalpies and optimized geometries when compared to experimental and *m*PWPW calculations. The $PM3_{BP}$ model was applied to hydrogen-bonded nucleic acid base trimers not contained in the parametrization set and found to agree much better with prior, higher-level calculations than the other tested semiempirical Hamiltonians.

Overparametrization of semiempirical methods to focused chemical problems can distort the physical nature of the model away from general applicability, which calls their very usefulness into question and limits their general predictive capability. On the other hand, the very broadly parametrized semiempirical models are not of sufficient quantitative accuracy to be useful in biological applications without additional ad hoc corrections. This suggests that the forms of current semiempirical models might be reaching their inherent limits. Consequently, further progress needs to be made in the development of new semiempirical methods[17] with treatments for those phenomena not described well with current Hamiltonian forms, such as dispersive attraction and proper polarization to electric fields while using a small basis set.

Nonetheless, the present work takes a significant step forward in testing the ability of the common semiempirical Hamiltonian forms to accommodate and reliably reproduce hydrogen-bonded nucleic acid base interactions. None of the tested Hamiltonians contain a term to properly account for the long-range dispersion effects that play an important role in stabilizing base-stacking interactions (or condensed phase simulations, in general), and further refinement of the methods to include such terms is likely to lead to more accurate and robust semiempirical models for biomolecular interactions.

Acknowledgment. D.Y. is grateful for financial support provided by the National Institutes of Health (Grant GM62248) and the Army High Performance Computing Research Center (AHPCRC) under the auspices of the Department of the Army, Army Research Laboratory (ARL) under Cooperative Agreement number DAAD19-01-2-0014. C.J.C. thanks the National Science Foundation for support (CHE02-03346). Computational resources were provided by the Minnesota Supercomputing Institute.

Supporting Information Available: The Supporting Information contains an extended discussion and comparison of nucleic acid base monomer bond lengths, angles, torsion angles, and dipole moments predicted at the semiempirical and DFT levels of theory. Nucleic acid base dimer hydrogen-bond lengths, hydrogen-bond angles, and dimerization binding enthalpies are tabulated and compared in the extended discussion. Dimerization enthalpies, hydrogen-bond lengths, and the dipole moments of the 37 hydrogen-bonded complexes not considered in the parametrization are also tabulated. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Friesner, R. A.; Beachy, M. D. *Curr. Opin. Struct. Biol.* **1998**, *8*, 257−262.

(2) Alkorta, I.; Elguero, J. *J. Phys. Chem. B* **2003**, *107* (22), 5306−5310.

(3) Brandl, M.; Meyer, M.; Sühnel, J. *J. Phys. Chem. A* **2000**, *104* (47), 11177−11187.

(4) Müller-Dethlefs, K.; Hobza, P. *Chem. Rev.* **2000**, *100* (1), 143−167.

(5) Peters, M.; Rozas, I.; Alkorta, I.; Elguero, J. *J. Phys. Chem. B* **2003**, *107* (1), 323−330.

(6) Sivanesan, D.; Babu, K.; Gadre, S. R.; Subramanian, V.; Ramasami, T. *J. Phys. Chem. A* **2000**, *104* (46), 10887−10894.

(7) Williams, N. G.; Williams, L. D.; Shaw, B. R. *J. Am. Chem. Soc.* **1989**, *111* (18), 7205−7209.

(8) Zhanpeisov, N. U.; Šponer, J.; Leszczynski, J. *J. Phys. Chem. A* **1998**, *102* (50), 10374−10379.

(9) Kawahara, S.; Uchimaru, T.; Sekine, M. *THEOCHEM*, **2000**, *530*, 109−117.

(10) Bosch, D.; Campillo, M.; Pardo, L. *J. Comput. Chem.* **2003**, *24* (6), 682−691.

(11) For example, minimal basis set Hartree−Fock methods.

(12) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209−220.

(13) Stewart, J. J. P. *J. Mol. Model.* **2004**, *10*, 6−12.

(14) Thiel, W. In *Adv. Chem. Phys.*; Prigogine, I., Rice, S. A., Eds.; John Wiley and Sons: New York, 1996; Volume 93, pp 703−757.

(15) Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495−506.

(16) Clark, T. *THEOCHEM* **2000**, *530*, 1−10.

(17) Winget, P.; Selçuki, C.; Horn, A.; Martin, B.; Clark, T. *Theor. Chem. Acc.* **2003**, *110* (4), 254−266.

(18) Thiel, W. Semiempirical Theories. In *Handbook of Molecular Physics and Quantum Chemistry*; Wilson, S., Ed.; John Wiley and Sons: Chicester, U. K., 2003; Volume 2, pp 487−502.

(19) Lopez, X.; York, D. M. *Theor. Chem. Acc.* **2003**, *109*, 149−159.

(20) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085−1123.

(21) Yang, W.; Pérez-Jordá, J. M. In *Encyclopedia of Computational Chemistry*; von Schleyer, P. R., Ed.; John Wiley and Sons: New York, 1998; pp 1496−1513.

(22) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027−2094.

(23) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99* (8), 2161−2200.

(24) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.

(25) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186−195.

(26) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(27) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I., Jr. *Nucleic Acids: Structures, Properties, and Functions*; University Science Books: Sausalito, CA, 2000.

(28) Maki, A.; Brownewell, F. E.; Liu, D.; Kool, E. T. *Nucleic Acids Res.* **2003**, *31* (3), 1059−1066.

(29) Šponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2002**, *61*, 3−31.

(30) Šponer, J.; Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142−10151.

(31) Jurema, M. W.; Shields, G. C. *J. Comput. Chem.* **1993**, *14* (1), 89−109.

(32) Lively, T. N.; Jurema, M. W.; Shields, G. C. *Int. J. Quantum Chem.* **1994**, *52* (21), 95−107.

(33) Hobza, P.; Kabeláč, M.; Šponer, J.; Mejzkík, P.; Vondrášek, J. *J. Comput. Chem.* **1997**, *18* (9), 1136−1150.

(34) Khandogin, J.; York, D. M. *J. Phys. Chem. B* **2002**, *106*, 7693−7703.

(35) Khandogin, J.; Musier-Forsyth, K.; York, D. M. *J. Mol. Biol.* **2003**, *330*, 993−1004.

(36) Khandogin, J.; York, D. M. *Proteins* **2004**, *56*, 724−737.

(37) Gregersen, B. A.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2003**, *125*, 7178−7179.

(38) Gregersen, B. A.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2004**, *126*, 7504−7513.

(39) Dewar, M. J.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99* (15), 4899−4907.

(40) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616−626.

(41) Dewar, M. J. S.; Zoebisch, E.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(42) Stewart, J. J. P. *Rev. Comput. Chem.* **1990**, *1*, 45−81.

(43) Burstein, K. Y.; Isaev, A. N. *Theor. Chim. Acta* **1984**, *64* (5), 397−401.

(44) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; John Wiley & Sons: Chichester, England, 2004.

(45) Bernal-Uruchurtu, M.; Ruiz-López, M. *Chem. Phys. Lett.* **2000**, *330*, 118−124.

(46) Bernal-Uruchurtu, M. I.; Martins-Costa, M. T. C. C.; Millot, M. F. R.-L. *J. Comput. Chem.* **2000**, *21*, 572−581.

(47) Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. *Phys. Status Solidi B* **2000**, *217*, 357−376.

(48) Thiel, W. *MNDO97*, version 5.0; University of Zurich: Zurich, Switzerland, 1998.

(49) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601−1622.

(50) Tubert-Brohman, I.; Guimaraes, C. R. W.; Repasky, M. P.; Jorgensen, W. L. *J. Comput. Chem.* **2003**, *25*, 138−150.

(51) Voityuk, A. A.; Rösch, N. *J. Phys. Chem. A* **2000**, *104*, 4089−4094.

(52) Long, D. A.; Anderson, J. B. *Chem. Phys. Lett.* **2005**, *402*, 524−528.

(53) Brauer, B.; Chabanb, G. M.; Gerbe, R. B. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2543−2556.

(54) Sherer, E. C.; York, D. M.; Cramer, C. J. *J. Comput. Chem.* **2003**, *24*, 57−67.

(55) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108* (2), 664−675.

(56) Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54* (23), 16533−16539.

(57) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1996**, *93* (5), 281−301.

(58) Xantheas, S. S. *J. Chem. Phys.* **1996**, *104* (21), 8821−8824.

(59) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; 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.; 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.9; Gaussian, Inc.: Pittsburgh, PA, 1998.

(60) Sukhodub, L. F.; Yanson, I. K. *Nature* **1976**, *264* (5583), 245−247.

(61) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149−1170.

(62) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247−3276.

(63) Rossi, I.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *233*, 231−236.

(64) Giese, T. J.; York, D. M. Nonlinear parameter optimization algorithms. Manuscript in preparation.

(65) Goldberg, D. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley: Reading, MA, 1989.

(66) Coley, D. A. *An Introduction to Genetic Algorithms for Scientists and Engineers*; World Scientific: River Edge, NJ, 1999.

(67) Cundari, T. R.; Deng, J.; Fu, W. *Int. J. Quantum Chem.* **2000**, *77*, 421−432.

(68) Hutter, M. C.; Reimers, J. R.; Hush, N. S. *J. Phys. Chem. B* **1998**, *102*, 8080−8090.

(69) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, W. P. *Numerical Recipes in Fortran*, 2nd ed.; Cambridge University Press: Cambridge, U. K., 1992.

(70) Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1992**, *81*, 391−404.

(71) Li, J.; Xing, J.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *111* (3), 885−892.

(72) Seeman, N. C.; Rosenberg, J. M.; Suddath, F. L.; Kim, J.; Rich, A. *J. Mol. Biol.* **1976**, *104* (1), 109−144.

(73) Rosenberg, J. M.; Seeman, N. C.; Day, R. O.; Rich, A. *J. Mol. Biol.* **1976**, *104* (1), 145−167.

(74) Pundlik, S. S.; Gadre, S. R. *J. Phys. Chem. B* **1997**, *101* (46), 9657−9662.

(75) Poltev, V. I.; Shulyupina, N. V. *J. Biomol. Struct. Dyn.* **1986**, *3* (4), 739−765.

(76) Sponer, J.; Burda, J. V.; Mejzlik, P.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1997**, *14*, 613.

CT050102L