Theoretical Investigation of the Hydrogen Bond Strengths in Guanine–Cytosine and Adenine–Thymine Base Pairs

Ian R. Gould[†] **and Peter A. Kollman***

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446

Received May 1, 1992. Revised Manuscript Received November 2, 1993[⊗]

Abstract: In this paper we present the results of our theoretical investigations on the Watson–Crick base pairs of 9-methylguanine – 1-methylcytosine (GC) and 9-methyladenine–1-methylthymine (AT) and the Hoogsteen base pair of 9-methyladenine–1-methylthymine (AT). We have performed full geometry optimization of the bases and the base pairs within the *ab initio* molecular orbtial (MO) framework at the Hartree–Fock (HF) self-consistent field (SCF) level with the 6-31G* basis set. We have extended these calculations by performing SCF and MP2 calculations, utilizing the Dunning double-ζ plus polarization basis set (DZP), upon the 6-31G* SCF optimized bases and base pairs. Vibrational analyses were performed at the 6-31G* level to enable the calculated interaction energies to be corrected to "predict" a ΔH_{298} for the base pairs. We have further validated this theoretical model by showing that it reproduces the ΔH_{298} for $(H_2O)_2$, HCN···HF, and CH_3CN···HF. We find that for the AT base pairs, the Hoogsteen orientation is ~1 kcal/mol more stable than the Watson–Crick orientation, with a ΔH_{298} of –12.8 kcal/mol in excellent agreement with the experimental value of –13.0 kcal/mol. We do not, however, reproduce the experimental value of ΔH_{298} for Watson–Crick GC of –21.0 kcal/mol; we calculate a ΔH_{298} of –25.3 kcal/mol. These *ab initio* energies are also compared with those found with a recently derived molecular mechanics model. The molecular mechanics calculations do a good job of reproducing the GC base pair energies found quantum mechanically and also lead to AT base pair energies in quite close agreement with experiment.

Introduction

The hydrogen bonds between base pairs in DNA are arguably the most important ones in all of biological chemistry.[1] Their role in providing structure and directionality as well as plasticity in base pair opening allows genetic replication to occur with high fidelity. Computer simulation methods such as molecular mechanics, molecular dynamics (MD), and Monte Carlo (MC) have begun to allow the study of DNA structure and dynamics[2,3] to complement high-resolution X-ray single-crystal studies and NMR studies in solution.[4,5] A critical element in such theoretical studies has been the empirical parameters required to carry them out. Before applying a molecular mechanics model to double helical DNA, one must analyze the properties of fragments that make up the DNA, specifically, the hydrogen bond strength of isolated base pairs. There are two methods available to determine the isolated hydrogen bonding energies, experiments and *ab initio* calculations. Mass spectrometric experiments on base pairing energies of 1-methylthymine–9-methyladenine and 1-methylcytosine–9-methylguanine have led to enthalpies for base pair formation of –13.0 and –21.0 kcal/mol, respectively.[6]

A number of *ab initio* calculations have been applied to nucleic acid base pairing over the years, but most have been at the SCF level, which does not include dispersion attraction. There are two notable exceptions, one in which the dispersion term was estimated empirically[7] and the other in which MP2 energies were calculated with the MIDI-1 and MINI-1 basis sets.[8] Given the important role of the dispersion correction to intermolecular attraction of molecules as large as the nucleic acid bases, the inclusion of correlation effects is critical for a quantitative comparison of the calculations with experiment. It is also essential that the calculations be carried out with a sufficiently flexible basis set that such a comparison with experiment can be done. Only recently, with the advent of direct SCF and MP2 methodologies,[9–12] have such calculations become generally feasible.

We have performed geometry optimizations on the bases and the base pairs at the SCF HF/3-21G and HF/6-31G* levels and vibrational frequency analyses at both levels. In addition we have performed single-point MP2/6-31G*//HF/3-21G and MP2/DZP//HF/6-31G* calculations on the base and the base pairs. This last model MP2/DZP//HF/6-31G* combined with the HF/6-31G*//HF/6-31G* vibrational frequencies and the SCF basis set superposition error, HF/DZP//HF/6-31G*, leads to excellent agreement with experiment for the AT base pair. This model also leads to good agreement for the hydrogen bond energy in the water dimer. However, the calculated ΔH_{298} for GC base pair formation using this method is significantly too negative. This suggests that the calculations have not converged for this system or that the experiment may need to be reevaluated.

Methodology

All four bases, 1-methylcytosine, 1-methylthymine, 9-methyladenine, and 9-methylguanine, were fully optimized at the HF/3-21G[13] and HF/6-31G*[14] levels. The analytical second derivatives were then obtained

---

[†] Present address: Department of Chemistry, University of Manchester, Oxford Road, Manchester, M13 9PL, U.K.

[⊗] Abstract published in *Advance ACS Abstracts,* February 15, 1994.

(1) Watson, J.; Crick, H. C. *Nature* **1953,** *171,* 753.

(2) Pearlman, D. A.; Kollman, P. A. *J. Mol. Biol.* **1991,** *220,* 457.

(3) Withka, J. M.; Swaminathan, S.; Srinivasam, J.; Beveridge, D. L.; Bolton, P. H. *Science* **1992,** *255,* 597.

(4) Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. V.; Kopka, M. L. *Science* **1982,** *216,* 475–485.

(5) Nerdal, W.; Hare, D.; Reid, B. *Nucl. Acids. Res.* **1989,** *28,* 10008.

(6) Yanson, I.; Teplitsky, A.; Sukhodub, L. *Biopolymers* **1979,** *18,* 1149.

(7) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987,** *109,* 1302.

(8) Hrouda, V.; Florian, J.; Hobza, P. *J. Phys. Chem.* **1993,** *97,* 1542.

(9) Almlof, J.; Faegri, K., Jr.; Korsell, K. *J. Comput. Chem.* **1982,** *3,* 385.

(10) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. *Chem. Phys. Lett.* **1989,** *152,* 165.

(11) Head-Gordon, M.; Frisch, M. J.; Pople, J. A. *Chem. Phys. Lett.* **1988,** *153,* 503.

(12) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1990,** *166,* 275.

(13) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980,** *102,* 939.

(14) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973,** *28,* 213.

0002-7863/94/1516-2493$04.50/0

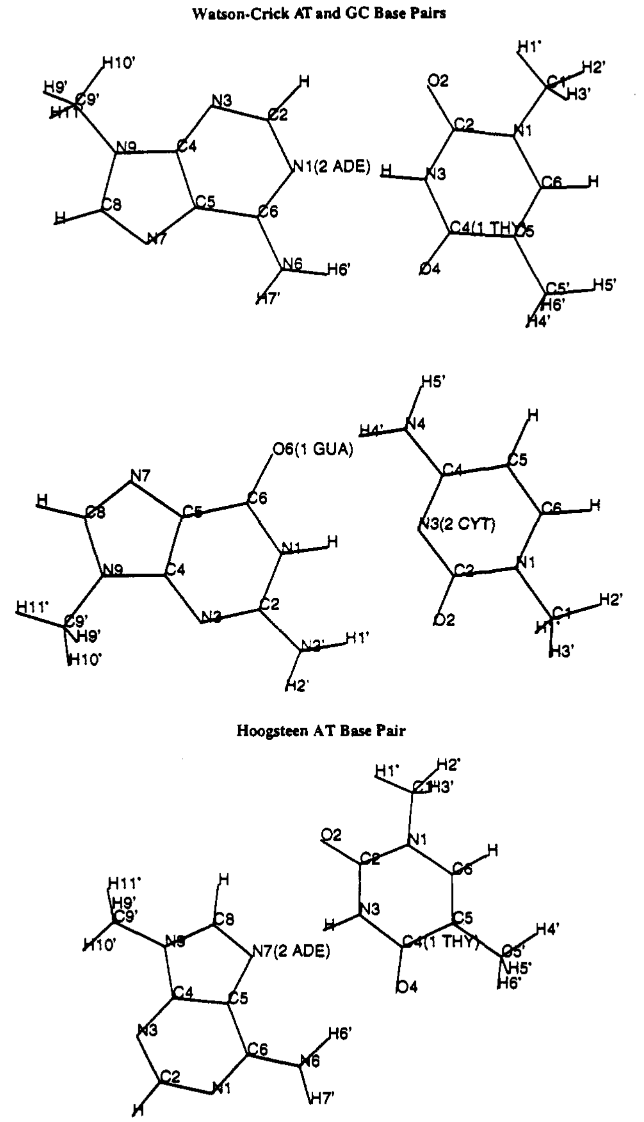

Figure 1. Watson–Crick AT and GC base pairs and Hoogsteen AT base pair.

for these monomers. In addition, single-point MP2/6-31G*//HF/3-21G, HF/DZP[15]//HF/6-31G*, and MP2/DZP//HF/6-31G* energy calculations were performed. All MP2 calculations in this paper employed the "frozen" core approximation.

In order to set up the Watson–Crick GC and AT and Hoogsteen AT base pairs, we used Saenger's principles of nucleic acid structure[16] to obtain the distances between the monomers in the base pairs. We used the optimized monomer parameters to define the individual nucleic acids in the base pairs and defined the distance between the two monomers through the N–H···N distance; this is best illustrated by referring to Figure 1, in which we illustrate the orientations of the Watson–Crick AT and GC and Hoogsteen AT base pairs and also the atom numbering. Full geometry optimization was performed on all degrees of freedom for the base pairs at the HF/3-21G and HF/6-31G* levels of theory. We also obtained the analytical second derivatives and frequencies for these optimized structures.

For each of the base pairs we then performed single-point MP2/6-31G* energy calculations on the HF/3-21G structures together with two subsequent calculations in which the distance between the N–H···N distance was increased by 0.1 and 0.2 Å, respectively, while the remaining geometrical parameters were held constant. For the geometry corresponding to the lowest energy at this MP2/6-31G*//HF/3-21G level we then performed counterpoise correction calculations, as described by Boys and Bernardi,[17] to obtain estimates of the basis set superposition error (BSSE) at the SCF and MP2 levels.

For the base pairs optimized at the HF/6-31G* level we performed a series of single-point energy calculations at the HF level with the 6-31G** [14] and DZP[15] basis sets and obtained estimates of the BSSE. We have also performed single-point MP2/DZP//HF/6-31G* energy calculations upon the base pairs. Combination of the MP2/DZP//HF/6-31G* interaction energies with the HF/6-31G* vibrational frequencies and the HF/DZP//HF/6-31G* BSSE gives our best estimates of the ΔH_{298} for the base pairs. We have also calculated the hydrogen bond strength of the linear water dimer using this same procedure, in order to evaluate the procedure on a complex where there is both experimental and very high level *ab initio* data.

Molecular mechanics optimization of the bases and base pairs was performed using the double precision minimization module of AMBER 3A.[18] The zero point and thermal energies, at 298 K, for the molecular mechanically optimized bases and base pairs were calculated using the NMODE program of AMBER 3A.[18]

All quantum mechanical calculations were performed with the GAUSSIAN 90 program[19] on a range of machines including an FPS 522 and Cray YMP/864.

Results

Geometry and Vibrational Spectra. In Tables 1–4 (supplementary material), we report the HF/3-21G, HF/MINI-1,[8] HF/6-31G* optimized, and experimental[20] geometries for the four bases. Direct comparison of the theoretical structures and the experimentally determined structures should be carried out with some caution, since the latter are for the bases attached to a furanose ring. In addition, the geometries for the HF/MINI-1[8] are for adenine and thymine, not for the 9- and 1-methyl derivatives, respectively. We find that the bond lengths found by the HF/3-21G and HF/6-31G* procedures are in general agreement with experiment, whereas for HF/MINI-1[8] the bond lengths are consistently too long. The bond angles calculated using these three procedures are in good agreement with experiment. In Table 5 we report the total energies for the individual bases. The bond distances and bond angles of the individual bases in the optimized dimers are given in Tables 6–11 (supplementary material), and where possible, these have been compared with experiment.[16] However, such comparisons should be viewed with caution since the GC experimental values refer to duplexes of GpC and Watson–Crick AT for ApU. Only for the Hoogsteen AT is a direct comparison with experimental values possible since there has been a neutron diffraction study of 9-methyladenine–1-methylthymine. Again, optimizations at HF/3-21G and 6-31G* perform much better than the HF/MINI-1[8] in reproducing the experimental bond legnths, while all three methods give bond angles in reasonable agreement with experiment.

When we performed analytical second derivative calculations on the HF/3-21G optimized bases, we found that 9-methylguanine and 1-methylthymine both had a single negative frequency associated with rotation of a methyl group; for the latter it was the methyl at C5. We found that the 1-methylcytosine and 9-methyladenine were at true minima, as characterized by the lack of any negative frequencies. Upon inspection of the harmonic frequencies for the base pairs calculated at the HF/3-21G level, we found that the Watson–Crick GC and AT and the Hoogsteen AT all contained a single negative frequency. Since the single negative frequencies in the base pairs were associated with the same mode as the individual bases, we choose not to reevaluate

---

(15) Dunning, T. H.; Hay, P. J. *Modern Theoretical Chemistry*; Plenum: New York, 1976; Chapter 1, pp 1–28.

(16) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, Berlin, Heidelberg, Tokyo, 1984; pp 123 and 124 and references within.

(17) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(18) Seibel, G.; Singh, U. C.; Weiner, P. K.; Caldwell, J. W.; Kollman, P. A. AMBER 3.0, Revision A, Regents of the University of California, 1989.

(19) Frisch, M. J.; Head-Gordon, M.; Trucks, G. W.; Foresman, J. B.; Schlegel, H. B.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Defrees, D. J.; Fox, D. J.; Whiteside, R. A.; Seeger, R.; Melius, C. F.; Baker, J.; Kahn, L. R.; Stewart, J. J. P.; Fluder, E. M.; Topiol, S.; Pople, J. A. Gaussian 90; Gaussian Inc.: Pittsburgh, PA 15213.

(20) (a) Taylor, R.; Kennard, O. J. *J. Mol. Struct.* **1982**, *78*, 1. (b) Voet, D.; Rich, A. *Prog. Nucl. Acid. Res. Mol. Biol.* **1970**, *10*, 183.

Table 5. Total Energies for Individual Bases[a]

level of theory	energy, au
1-Methylcytosine	
HF/3-21G//HF/3-21G	–429.232 194 1
HF/6-31G*//HF/3-21G	–431.643 050 2
MP2/6-31G*//HF/3-21G	–432.924 846 7
HF/6-31G*//HF/6-31G*	–431.645 734 5
HF/DZP//HF/6-31G*	–431.736 963 1
MP2/DZP//HF/6-31G*	–433.039 296 1
1-Methylthymine	
HF/3-21G//HF/3-21G	–487.801 916 3
HF/6-31G*//HF/3-21G	–490.535 122 5
MP2/6-31G*//HF/3-21G	–491.963 951 6
HF/6-31G*//HF/6-31G*	–490.554 651 5
HF/DZP//HF/6-31G*	–490.642 581 7
MP2/DZP//HF/6-31G*	–492.095 139 3
9-Methyladenine	
HF/3-21G//HF/3-21G	–500.718 816 2
HF/6-31G*//HF/3-21G	–503.550 865 0
MP2/6-31G*//HF/3-21G	–505.108 126 7
HF/6-31G*//HF/6-31G*	–503.550 703 0
HF/DZP//HF/6-31G*	–503.648 231 0
MP2/DZP//HF/6-31G*	–505.222 845 6
9-Methylguanine	
HF/3-21G//HF/3-21G	–575.181 663 2
HF/6-31G*//HF/3-21G	–578.424 179 5
MP2/6-31G*//HF/3-21G	–580.153 965 4
HF/6-31G*//HF/6-31G*	–578.427 485 5
HF/DZP//HF/6-31G*	–578.546 015 3
MP2/DZP//HF/6-31G*	–580.288 786 0

[a] The MP2 energies are for the "frozen" core approximation.

the geometries at this level, and we have assumed that any error in the calculated interaction will be matched by a corresponding error in the energy of the individual base.

For the individual bases optimized at the HF/6-31G* level there are several interesting observations to be made about the calculated harmonic frequencies. The HF/3-21G optimized geometries of the bases were used as the starting structures for the HF/6-31G* optimizations and, as a consequence of this, the amine groups of the 9-methylguanine, 1-methylcytosine, and 9-methyladenine were planar at the start and finish of the optimizations. Inspection of the 1-methylthymine frequencies revealed the structure to be a true minima. For 9-methylguanine and 1-methylcytosine there was a single negative frequency in both spectra, identified as the low frequency inversion mode of the amine group. For 9-methyladenine we found two negative frequencies, one associated with the amine group inversion and the other associated with the rotation of the methyl group.

Gould et al.[21] have shown that it is necessary at the HF/6-31G** level of optimization of cytosine and guanine to allow the amine group to be nonplanar in order to obtain true minima. Calculations at this level have been able to reproduce the infrared spectra of guanine and cytosine in better agreement with experiment than found in previous calculations. Therefore, we decided to re-optimize 9-methylguanine, 1-methylcytosine, and 9-methyladenine with the amine groups nonplanar. Calculation of the harmonic frequencies of the optimized bases with nonplanar amine groups revealed that 9-methylguanine and 1-methylcytosine were true minima, while for 9-methyladenine there was a single negative frequency identified as the rotation of the methyl group. It is these final reoptimized HF/6-31G* geometries that we report and have used as our basis for further studies.

Inspection of the harmonic frequencies of the base pairs revealed that the Watson–Crick GC was a true minima at the HF/6-31G* level, whereas the Watson–Crick and Hoogsteen AT base pairs each had a single negative frequency associated with the rotation of the methyl group of the adenine. As the methyl group of the adenine is far from the interface between the adenine and thymine in both the Watson–Crick and Hoogsteen base pairs, (see Figure 1), we do not believe that its orientation will have a significant effect on the interaction energy. We therefore chose not to re-optimize the 9-methyladenine or the AT base pairs. It is interesting to note that in the base pairs, the amine groups of the bases become planar, which is not surprising given the greatest overlap between orbitals involved in hydrogen bonding.

Table 12. Hydrogen Bond Distances (Å)

base pair	3-21G[a]	6-31G*[b]	MP2/ 6-31G*[b]	6-31G*[c]	MM[d]	Saenger[e]
GC Watson–Crick						
O6···N4	2.77	2.87	2.87	2.93	2.86	2.91
N1···N3	2.91	3.01	3.01	3.05	2.92	2.95
N2···O1	2.86	2.96	2.96	3.01	2.84	2.86
AT Watson–Crick						
N6···O4	2.96	3.16	3.06	3.08	2.90	2.95
N1···N3	2.78	2.98	2.88	3.01	2.91	2.82
AT Hoogsteen						
N6···O4	2.99	3.19	3.09	3.12	2.90	2.86
N7···N3	2.75	2.95	2.85	2.97	2.90	2.93

[a] Fully optimized geometries at HF/3-21G. [b] Lowest energies using HF/3-21G optimized geometry plus 0.1 or 0.2 Å, depending on lowest total energy. [c] Fully optimized geometries at HF/6-31G*. [d] Optimized distance at the molecular mechanics level, see ref 37. [e] Hydrogen bond distances from ref 16.

In Table 12, we report the hydrogen bond distances in the base pairs calculated at the various levels of theory. The level of agreement between the HF/3-21G and HF/6-31G* structures and the experimental values reported by Saenger[16] is quite good. One should realize that the experimental data are based on X-ray crystallographic studies and crystal packing effects could lead to differences with the gas phase for these bases of the order of 0.01 Å in bond lengths and 1° in bond angles. In addition, as noted above, only for the Hoogsteen AT base pairs do the experimental and theoretical systems correspond to the same complexes. Thus, it is interesting to see that the important N7···N3 distance is well reproduced at the HF/6-31G* level, whereas this distance is much smaller at the HF/3-21G level.

Ab Initio Base Pair Energies and Enthalpies. In Table 13 we report the calculated base pair energies, based upon the HF/3-21G optimized bases and base pairs. As can be seen, the 3-21G basis greatly exaggerates the base pairing energies and is clearly inadequate for studying molecular interaction energies. The values calculated at the HF/6-31G* and MP2/6-31G* levels of theory are more reasonable. However, to compare directly to experiment, one must correct for basis set superposition error and translational, rotational, and vibrational energies. Applying these corrections at the SCF level yields a reasonably satisfactory, if somewhat too small, base pair energy for the Watson–Crick GC, whereas for the AT base pairs the energies are far too small in comparison to experiment. The MP2 values, when adjusted similarly, give too great a base pair energy for GC but too small an interaction energy for both AT base pairs.

At this point it is important to consider the basis set superposition error (BSSE): its origin, a method for its evaluation, and the appropriateness of its application at varying theoretical levels. In the supermolecule method, the interaction energy is calculated as the difference between the total energy of the supermolecule XY and the sum of the energies of the subsystems X and Y, $\Delta E = E_{XY} - (E_X + E_Y)$. With a truncated basis set, ΔE may be significantly affected by the BSSE, which originates from the fact that in the calculation of the supermolecule the subsystem A tends to improve its energy using the basis set of the subsystem B and vice versa. This results in an artificial energy lowering of the energy of the supermolecule. The correction of this defect, introduced by Boys and Bernardi,[17] the "counterpoise correction" (CPC), is generally accepted to be valid at the SCF

---

(21) s(a) Gould, I. R.; Vincent, M. A.; Hillier, I. H.; Lapinski, L.; Nowak, M. J. *Spectrochim. Acta* **1992**, *48a*, 811. (b) Gould, I. R.; Vincent, M. A.; Hillier, I. H. *Spectrochim. Acta* **1993**, *49a*, 1727.

Table 13. Hydrogen Bond Energies[h] of Base Pairs

model		GC Watson–Crick	AT Hoogsteen	AT Watson–Crick
HF/3-21G//HF/3-21G	$-\Delta E$	39.8	23.1	22.1
HF/6-31G*//HF/3-21G	$-\Delta E$[a]	25.8	11.8	11.1
HF/6-31G*//HF/3-21G(BSSE)	$-\Delta E$[b]	22.8	9.09	8.37
HF/6-31G*//HF/3-21G(BSSE)	$-\Delta E_0$[c]	20.2	7.39	6.87
HF/6-31G*//HF/3-21G(BSSE)	$-\Delta H_{298}$[d]	19.7	6.49	5.77
MP2/6-31G*//HF/3-21G	$-\Delta E$	32.8	19.4	18.5
MP2/6-31G*//HF/3-21G(BSSE)	$-\Delta E$[e]	26.0	13.5	12.5
MP2/6-31G*//HF/3-21G(BSSE)	$-\Delta E_0$[f]	23.4	11.8	11.0
MP2/6-31G*//HF/3-21G(BSSE)	$-\Delta H_{298}$[g]	22.9	10.9	9.9

[a] HF/6-31G* energies at HF/3-21G optimized geometries plus variation in the N–H···N distance—as described in the text, see Figure 1. [b] BSSE calculated at the SCF level. [c] Includes correction for zero-point energies calculated at the HF/3-21G level. [d] Corrected for translational, rotational, and vibrational energies as determined at the HF/3-21G level. [e] BSSE calculated at the MP2 level. [f] Includes correction for zero-point energies calculated at the HF/3-21G level. [g] Corrected for translational, rotational, and vibrational energies as determined at the HF/3-21G level. [h] All energies are in kcal/mol.

level. The BSSE calculated using the full counterpoise procedure is defined:

$$\mathrm{BSSE(X{-}Y)} = E(\mathrm{X})_{\mathrm{X}} - E(\mathrm{X})_{\mathrm{XY}} + E(\mathrm{Y})_{\mathrm{Y}} - E(\mathrm{Y})_{\mathrm{XY}}$$

where $E(\mathrm{X})_{\mathrm{XY}}$ and $E(\mathrm{X})_{\mathrm{X}}$ represent the energy of X calculated using its geometry within the dimer and the basis functions of X plus Y in the former, and those of X alone in the latter. The CPC can be taken as an upper bound estimate of the BSSE, and it is known that it works generally well at the SCF level, except when minimal basis sets[22,23] are used.

There are conflicting views[24,25] whether one should implement BSSE corrections at the correlated level. The use of the CPC method for correlated methods is rather questionable, since in the calculation of $E(\mathrm{X})_{\mathrm{XY}}$, and analogously $E(\mathrm{Y})_{\mathrm{XY}}$, excitations form occupied orbitals of X to the "occupied" orbitals of Y are allowed, which may lead to a spurious overcorrection. The "occupied" orbitals of Y are not accessible for electrons of X in the supermolecule calculation. Attempts to utilize only the virtual spaces of X and Y have led to discouraging results.[26] In addition it has been argued that the uncorrected energy converges long before the BSSE disappears.[25b,27]

Examination of the magnitudes of the BSSE's calculated at the SCF and MP2 levels in Table 13 show that at the SCF level with the 6-31G* basis set the BSSE are 3, 2.7, and 2.7 kcal/mol for Watson–Crick GC and AT and Hoogsteen AT, respectively. However, at the MP2 level the calculated BSSE's are much larger at 6.8, 6.0, and 5.9 kcal/mol for Watson–Crick GC and AT and Hoogsteen AT base pairs, respectively. The Watson–Crick AT base pair BSSE corrections compare favorably with those of Hrouda et al.,[8] who found at the HF/MINI-1//HF/MINI-1 level values of 3.9 and 8.4 for the SCF and MP2 BSSE's, respectively, and at the HF/MIDI-1//HF/MINI-1 level values of 6.9 and 11.4 for the SCF and MP2 BSSE's, respectively. It was at this stage of our investigation that we decided to reoptimize the bases and base pairs at the HF/6-31G* level, as noted previously, and to then perform a series of single-point calculations on the base pairs with increasingly larger basis sets. Furthermore, we calculated the SCF BSSE's for the Watson–Crick and Hoogsteen AT's with the hierarchy of basis sets 6-31G*, 6-31G**, and DZP to determine which basis gave the smallest BSSE. We found, not surprisingly, that the DZP basis gave the smallest SCF BSSE's for the Watson–Crick and Hoogsteen AT's of 1.13 and 1.64 kcal/mol, respectively; the values were 2.42 and 2.42 kcal/mol for the 6-31G* basis, respectively, and 2.22 and 2.20 kcal/mol for the 6-31G** basis, respectively. Since the differences in the BSSE's calculated at the 6-31G* and 6-31G** level for both AT's were similar, we only performed BSSE calculations with the 6-31G* and DZP basis sets for Watson–Crick GC. The values were 2.86 and 1.65 kcal/mol, respectively, for these two levels of calculation.

Table 14. Water Dimer Energies (kcal/mol)

level of theory		interaction energy
HF/6-31G*//HF/6-31G*	$-\Delta E$	5.63
HF/6-31G*//HF/6-31G*(HF BSSE)	$-\Delta E$	4.70
HF/6-31G*//HF/6-31G*(HF BSSE)	$-\Delta H_{298}$	2.36
MP2/6-31G*//HF/6-31G*	$-\Delta E$	7.15
MP2/6-31G*//HF/6-31G*(HF BSSE)	$-\Delta E$	6.22
MP2/6-31G*//HF/6-31G*(HF BSSE)	$-\Delta H_{298}$	3.88
MP2/6-31G*//HF/6-31G*(MP2 BSSE)	$-\Delta E$	5.31
MP2/6-31G*//HF/6-31G*(MP2 BSSE)	$-\Delta H_{298}$	2.97
HF/DZP//HF/6-31G*	$-\Delta E$	4.93
HF/DZP//HF/6-31G*(HF BSSE)	$-\Delta E$	4.47
HF/DZP//HF/6-31G*(HF BSSE)	$-\Delta H_{298}$	2.13
MP2/DZP//HF/6-31G*	$-\Delta E$	6.18
MP2/DZP//HF/6-31G*(HF BSSE)	$-\Delta E$	5.72
MP2/DZP//HF/6-31G*(HF BSSE)	$-\Delta H_{298}$	3.38
MP2/DZP//HF/6-31G*(MP2 BSSE)	$-\Delta E$	5.00
MP2/DZP//HF/6-31G*(MP2 BSSE)	$-\Delta H_{298}$	2.66
experiment[a]	$-\Delta H_{298}$	3.59 ± 0.5
MP4/6-311+G(2df,2p)[b]	$-\Delta H_{298}$	3.48

[a] Experiment, see ref 28. [b] See ref 29.

To obtain the best possible estimate of the base pair energies we decided to adopt the following strategy: the geometries of the bases and base pairs at which the energies were to be obtained were taken as the HF/6-31G* optimized structures; the zero-point energies and translational, rotational, and vibrational energies needed to correct the ΔE to a ΔH_{298} were calculated at the HF/6-31G*//HF/6-31G* level; the MP2 energies were calculated with the DZP basis set and an estimate of the BSSE was taken as the SCF BSSE calculated with the DZP basis.

To assess the validity of this model we applied it to the linear water dimer since there is both experimental[28] and high level *ab initio*[29] data available on this complex. We report our results for the water dimer in Table 14. It can be seen that using this model (MP2/DZP//HF/6-31G*(BSSE)) $-\Delta H_{298}$ is calculated to be 3.38 kcal/mol, which well reproduces the experimental and highest level theoretical $-\Delta H_{298}$ values of 3.59 and 3.48 kcal/mol, respectively.

---

(22) Carsky, P.; Urban, M. *Ab initio Calculations. Methods and Applications in Chemistry. Lecture Notes in Chemistry*; Springer: Berlin, 1980; Vol. 16.

(23) Hobza, P.; Zahradnik, R. *Weak Intermolecular Interactions in Chemistry and Biology*; Academia: Prague; Elsevier: Amsterdam, 1980.

(24) Hobza, P.; Zahradnik, R. *Chem. Rev.* **1988**, *88*, 871.

(25) (a) van Lenthe, J. H.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. *Adv. Chem. Phys.* **1987**, *69*, 521. (b) Schwenke, D. W.; Truhlar, D. G. *J. Chem. Phys.* **1985**, *82*, 2418. (c) Dykstra, C. E. *Ab Initio Calculations of the Structures and Properties of Molecules*; Elsevier: Amsterdam, 1988. (d) Mayer, I. *Theor. Chim. Acta* **1987**, *72*, 207. (e) Mayer, I. *Int. J. Quantum Chem.* **1983**, *23*, 341. (f) Loushin, S. K.; Liu, S.; Dykstra, C. E. *J. Chem. Phys.* **1986**, *84*, 2720.

(26) Gutowski, M.; van Lenthe, J. H.; Verbeek, J.; van Duijneveldt, F. B.; Chalasinski, G. *Chem. Phys. Lett.* **1986**, *124*, 370.

(27) Frisch, M. J.; Del Bene, J. E.; Binkley, J. S.; Schaefer, H. F. *J. Chem. Phys.* **1986**, *84*, 2279.

(28) Curtiss, L. A.; Frurip, D. J.; Blander, H. *J. Chem. Phys.* **1979**, *71*, 2703.

(29) Smith, B. J.; Swanton, D. J.; Pople, J. A.; Schaefer, H. F.; Radom, L. *J. Chem. Phys.* **1990**, *92*, 1240.

Table 15. B···HF, Dimer Energies (kcal/mol)

level of theory		interaction energy
B = HCN		
MP2/DZP//HF/6-31G*	$-\Delta E$	8.67
MP2/DZP//HF/6-31G*(HF BSSE)	$-\Delta E$	7.55
MP2/DZP//HF/6-31G*(HF BSSE)[c]	$-\Delta H_{298}$	5.00
MP2/DZP//HF/6-31G*(MP2 BSSE)	$-\Delta E$	6.37
MP2/DZP//HF/6-31G*(MP2 BSSE)[c]	$-\Delta H_{298}$	3.82
experiment[a]	$-\Delta H_{298}$	4.97
B = H_2O		
MP2/DZP//HF/6-31G*	$-\Delta E$	10.27
MP2/DZP//HF/6-31G*(HF BSSE)	$-\Delta E$	9.72
MP2/DZP//HF/6-31G*(HF BSSE)[c]	$-\Delta H_{298}$	7.45
MP2/DZP//HF/6-31G*(MP2 BSSE)	$-\Delta E$	8.67
MP2/DZP//HF/6-31G*(MP2 BSSE)[c]	$-\Delta H_{298}$	6.40
experiment[a]	$-\Delta H_{298}$	9.34 (6.2)[b]
B = CH_3CN		
MP2/DZP//HF/6-31G*	$-\Delta E$	10.16
MP2/DZP//HF/6-31G*(HF BSSE)	$-\Delta E$	8.95
MP2/DZP//HF/6-31G*(HF BSSE)[c]	$-\Delta H_{298}$	7.08
MP2/DZP//HF/6-31G*(MP2 BSSE)	$-\Delta E$	7.64
MP2/DZP//HF/6-31G*(MP2 BSSE)[c]	$-\Delta H_{298}$	5.77
experiment[a]	$-\Delta H_{298}$	6.88

[a] Experiment, see ref 30. [b] Experiment, see ref 31. [c] Correction for translational, rotational, and vibrational energies determined at the HF/6-31G*//HF/6-31G* level.

It should be stressed that it is difficult to determine accurate ΔH_{298} for weakly hydrogen bonded complexes in the gas phase, because the concentration of these complexes at equilibrium is so low. However, to further validate our theoretical model, we studied the complexes involving HF as a proton donor, which Millen and co-workers[30] had studied experimentally. We report the results of applying our model for calculating $-\Delta H_{298}$(MP2/DZP//HF/6-31G*(BSSE)) to B···HF, for B = HCN, CH_3CN, and H_2O, in Table 15. It can be seen that our model does exceptionally well in reproducing the $-\Delta H_{298}$ for the HCN···HF and CH_3CN···HF dimers with calculated values of 5.0 and 7.1 kcal/mol, respectively, which can be compared to the experimental values of 5.0 and 6.9 kcal/mol, respectively. For the H_2O···HF dimer our model gives a value of $-\Delta H_{298}$ of 7.5 kcal/mol. This is smaller than the value reported by Millen et al.[30] of 9.3 kcal/mol and larger than the experimental value of 6.2 kcal/mol reported by Sasaki and Yoshimine.[31] If the more recent value of Millen et al.[30] is correct, it suggests that, if anything, our model underestimates the magnitude of ΔH_{298}. We should like to point out that in the study of H_2O···HF by Del Bene[32b] the MP4(SDTQ)6-31+G(2d,2p)//HF/6-31G(d) $-\Delta H_{298}$ value was calculated to be 7.4 kcal/mol, the thermal corrections being calculated at the HF/6-31G(d)//HF/6-31G(d) level.

Other sources of error in our model include optimization at the SCF level only, calculation of BSSE at the SCF level only, and the evaluation of correlation only up to the MP2 level. We would like to address each of these points: First, while it is possible to perform MP2 optimizations in a "direct" fashion, the computational cost is prohibitive for a system of this size even on a supercomputer of the performance of the Cray YMP/864. Furthermore, it is doubtful whether there would be a significant change in the geometry of the complexes, compared with those optimized at the SCF level, and since the dimer surface is very "soft" near the minimum, the change in ΔE due to it is likely to be small. Second, we have given a somewhat lengthy explanation of the origins of the BSSE above, and why we have employed it at the SCF level only. The excellent agreement with experiment for $(H_2O)_2$, HCN···HF, and CH_3CN···HF would support such a choice.

In the excellent review of *ab initio* calculations on hydrogen-bonded complexes by Scheiner,[33] the work of Szalewicz et al.[34] and Del Bene[32] is cited. Both works indicate that the second-order correlation term dominates in the correlation contribution to the interaction energy and that the third-order correction is completely negligible, and in the fourth order the SDQ term cancels to a large extent the triples contribution. In addition, in the work of Szalewicz et al.[33] they found that CCSD-T interaction energies were very close to the Møller–Plesset values. Therefore, we believe that our calculations at the MP2 level are appropriate and that correlation calculations above this level would not significantly alter the interaction energy even if they were computationally tractable for these systems. However, to perform an MP4(SDTQ) or a CCSD(T) or a QCISD(T) energy calculation on the Watson–Crick GC base pair, with 343 basis functions, would require ~18 Gb of disk storage, since storage requirements scale as order N^3.

The best estimates we obtained for the Watson–Crick GC and AT base pair energies are summarized in Table 16. They are denoted as MP2/DZP//6-31G*(BSSE) and lead to values of $-\Delta H_{298}$ of 25.4 and 11.9 kcal/mol, respectively. For the Hoogsteen AT base pair, the best estimate of $-\Delta H_{298}$ is 12.8 kcal/mol. The value for the Hoogsteen AT is in excellent agreement with the experimental[6] mass spectrometric value of 13.0 kcal/mol. However, the value of 25.4 kcal/mol for the Watson–Crick GC is significantly larger than the experimental[6] value of 21.0 kcal/mol. It is instructive to compare the best estimates of $-\Delta H_{298}$ with that found at the HF/6-31G* level. In this case $-\Delta H_{298}$ values for the Watson–Crick GC and AT and Hoogsteen AT are 20.0, 8.1, and 8.7 kcal/mol, respectively.

Comparison with Other *ab Initio* and Molecular Mechanics Calculations. We have used molecular mechanics with 6-31G* electrostatic potential charges,[35] a new set of van der Waals parameters[36] without 10–12 H bonds, and vibrational corrections to calculate ΔH_{298} at the molecular mechanics level for AT and GC base pairs.[37] We present these results in Table 16; they further support the result from the quantum mechanical calculations that the GC base pair should have an enthalpy of interaction approximately twice the magnitude of an AT base pair. One can compare these hydrogen bond strengths with those presented by Weiner et al.,[38] and it is clear that those in the Weiner et al. force field are ~1 kcal/mol smaller. This may rationalize why DNA simulations using this model with explicit TIP3P water seem to have more rapid base pair openings[39] than expected from NMR experiments.

In the paper of Hobza and Sandorfy[7] they applied a lower level of quantum mechanical theory, *ab initio* with a minimal basis set and BSSE correction, and an empirical correction for dispersion energy to study 29 possible DNA base pairs. They obtained values for the $-\Delta E$ of GC of 26.5 kcal/mol, a $-\Delta E$ of 16.0 kcal/mol for Watson–Crick AT, and a $-\Delta E$ of 15.9 kcal/mol for Hoogsteen AT, close to our values in Table 16. This is somewhat surprising, given the difference in basis sets and the fact that MP2 includes other correlation effects and, without diffuse functions, not enough dispersion attraction. Nonetheless, when corrected for zero-point energies and thermal effects using our

---

(30) Adebayo, S. L. A.; Legon, A. C.; Millen, D. J. *J. Chem. Soc., Faraday Trans.* **1991**, *87(3)*, 443.

(31) Sasaki, F.; Yoshimine, M. *Phys. Rev. A* **1974**, *9*, 17.

(32) (a) Del Bene, J. E. *J. Chem. Phys.* **1987**, *86*, 2110. (b) Del Bene, J. E. *J. Phys. Chem.* **1988**, *92*, 2874.

(33) Scheiner, S. In *Reviews of Computational Chemistry*; Lipkowitz, K., Boyd, R., Eds.; VCH, 1991; Vol. 2.

(34) Szalewicz, K.; Cole, S. J.; Kolos, W.; Bartlett, R. J. *J. Chem. Phys.* **1988**, *89*, 3662.

(35) Gould, I. R.; Pearlman, D. A.; Kollman, P. A. To be submitted for publication.

(36) Spellmeyer, D.; Ferguson, D. M.; Cornell, W. D.; Cieplak, P.; Kollman, P. A. Unpublished.

(37) Cornell, W. D.; Cieplak, P.; Bayly, C.; Kollman, P. A. Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies and Free Energies of Solvation. *J. Am. Chem. Soc.*, accepted for publication.

(38) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. K. *J. Am. Chem. Soc.* **1984**, *106*, 765.

(39) Kumar, S. Ph.D. Thesis, University of Pittsburgh, 1990.

Table 16. Hydrogen Bond Energies of Base Pairs

model		GC Watson–Crick	AT Hoogsteen	AT Watson–Crick
HF/6-31G*//HF/6-31G*	$-\Delta E$	25.4	14.0	13.4
HF/6-31G*//HF/6-31G*(BSSE)	$-\Delta E$[a]	22.5	11.6	11.0
HF/6-31G*//HF/6-31G*(BSSE)	$-\Delta H_{298}$[b]	20.0	8.7	8.1
HF/DZP//HF/6-31G*	$-\Delta E$	23.9	12.4	11.7
HF/DZP//HF/6-31G*(BSSE)	$-\Delta E$[c]	22.3	11.3	10.6
HF/DZP//HF/6-31G*(BSSE)	$-\Delta H_{298}$[d]	19.7	8.4	7.8
MP2/DZP//HF/6-31G*	$-\Delta E$	29.6	16.8	15.9
MP2/DZP//HF/6-31G*(BSSE)	$-\Delta E$[e]	28.0	15.6	14.7
MP2/DZP//6-31G*(BSSE)	$-\Delta H_{298}$[f]	25.4	12.8	11.9
molecular mechanics	$-\Delta E$[g]	27.2	13.0	12.6
molecular mechanics	$-\Delta H_{298}$[h]	24.8	11.2	10.7
experiment	$-\Delta H_{298}$[i]	21.0	13.0	

[a] BSSE calculated at the SCF level. [b] After correction for translational, rotational, and vibrational energies determined at the HF/6-31G*//HF/6-31G* level. [c] BSSE calculated at the SCF level. [d] After correction for translational, rotational, and vibrational energies determined at the HF/6-31G*//HF/6-31G* level. [e] BSSE calculated at the SCF level. [f] After correction for translational, rotational, and vibrational energies determined at the HF/6-31G*//HF/6-31G* level. [g] Molecular mechanics energies from ref 32. [h] ΔH_{298} calculated using normal mode analysis for translational, rotational, and vibrational energies from ref 32. [i] See ref 6.

thermal corrections, their values for ΔH would be in line with experiment for the AT base pair and about ~3 kcal/mol too large for GC, in qualitative agreement with our results.

A triple-ζ plus polarization HF calculation[40] with BSSE correction yields a GC base pair energy of 24.7 kcal/mol even without dispersion and correlation effects! When these are added with thermal corrections it is likely that ref 40 would calculate a GC pair energy even larger than ours. In the recent paper of Trollope et al.,[41] they have calculated the base pair energies of Watson–Crick GC and AT and Hoogsteen AT using distributed multipoles to model the electrostatic interactions and an empirical "6–12" potential derived from the Weiner et al.[38] force field to model the dispersion–repulsion terms. Having corrected their ΔE's with the zero-point energies and thermal effects using thermal corrections calculated at the HF/3-21G level they obtain $\Delta H_{29}38$ values of 23.7 and 12.9 kcal/mol for the formation of Watson–Crick GC and AT base pairs and 13.3 kcal/mol for the formation of the Hoogsteen base pair, again in qualitative agreement with our results.

We disagree with the assertion in ref 7 that molecular mechanics methods are inherently incapable of reproducing the hydrogen bond energies as well as quantum mechanical methods. Quantum mechanical methods have not yet reached the stage of uniform reliability for macromolecules, particularly for representing the dispersion energy, and that is why in ref 7 an empirical approach was used for this contribution to the energy. In any case, we find that comparison of our best estimates for the base pair energies calculated from quantum mechanics and molecular mechanics is in reasonably good accord. We should note that the calculations in ref 7 found the AT Watson–Crick ~0.2 kcal/mol more stable than AT Hoogsteen.

We see quantum mechanical and molecular mechanical models as complementary. The former have been used in the construction of the latter, which can be applied to much larger systems. In particular, it is our opinion that suitably parametrized molecular mechanical models[37,38,42–44] can reproduce reality and accurate quantum mechanical models surprisingly well.

Discussion and Conclusions

Although the quantum mechanical model presented is not the last word, it is the most extensive applied to GC and AT base pairs. The model reproduces the experimental ΔH_{298} for the linear water dimer, HCN···HF, CH_3CN···HF, and the Hoogsteen AT base pair; however, it finds the ΔH_{298} for GC base pair formation ~4.5 kcal/mol more negative than the currently accepted experimental value. The results of the SCF calculations give base pair energies smaller than the experimental values and therefore show that correlation effects are of considerable importance.

At some point in the future, one would like to be able to perform geometry optimization of the bases and the base pairs at the MP2/6-31G* level to ascertain the importance of being at a minima at the correlated level when calculating base pair energies, though such calculations would be extremely computer intensive even on a supercomputer such as the Cray YMP/864.

Results of energy calculations using a new molecular mechanical model for DNA base pairing energies are presented, and these suggest a ΔH_{298} for AT and GC base pairs quite consistent with those derived from *ab initio* calculations. Many recent calculations suggest a GC base pair energy at least twice as strong as the AT base pair, whereas the experimental ΔH_{298}'s are –21.0 and 13.0 kcal/mol, respectively.

Because we cannot use the same level of *ab initio* calculations on the base pairs as on smaller hydrogen bonded complexes such as the water dimer, we have also compared the results of the theory which we have employed on the base pairs ((MP2/DZP/HF/6-31G*) with the Hartree–Fock basis set superposition error and vibrational corrections) with higher levels of theory on $(H_2O)_2$. The level of quantum mechanical model employed in this paper leads to a $-\Delta H_{298}$ for this dimer within 0.2 kcal/mol of the highest level of theory applied to it. In the case of the Hoogsteen AT base pair, where we can only compare our calculations with experiment, the level of agreement is excellent ($-\Delta H_{298}$ = 12.8 (calculated) and 13.0 (experimental) kcal/mol). We stress that the accuracy of the theory in these cases is fortuitously good; probably an answer within ± 1–2 kcal/mol of experiment is reasonable to expect for weak to moderate strength hydrogen-bonded systems.

Nonetheless, given the results for water dimer, HCN···HF, the CH_3CN···HF, and the AT base pair, we feel that the difference between the calculated and experimental ΔH_{298} of 4.4 kcal/mol for the GC base pair is significant and outside the range expected for errors in the theory. Thus, we suggest that this warrants an examination of the experimental value, determined nearly 15 years ago. Nonetheless, as noted by a referee of this paper, "how can we be sure that we have the correct balance of the components of the binding energies in each case, even when computed at the same level of *ab initio* theory?" In particular, the contributions from basis set superposition error and MP2 correlation corrections are significant. Of course, we *cannot* be *sure* that there is not something peculiar about GC that makes it behave differently

---

(40) Clementi, E.; Corongiu, G.; Chkrovorty, S. MOTECC; Clementi, E., Ed.; Escom Press: Leiden, 1990; pp 345–346.

(41) Trollope, K. I.; Gould, I. R.; Hillier, I. H. *Chem. Phys. Lett.* **1993**, *209*, 113.

(42) Poltev, V.; Shulyupina, N. *J. Biomol. Struct. Dyn.* **1986**, *3*, 739—singled out in ref 7 as having a molecular mechanical model that is most consistent with the quantum mechanical calculations.

(43) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comp. Chem.* **1986**, *7*, 230.

(44) Rao, S.; Kollman, P. A. *Biopolymers* **1986**, *25*, 267.

than the chemically very similar AT in these regards. Hopefully, the results presented here will be an impetus for theoreticians and experimentalists to resolve this discrepancy.

Acknowledgment. We are grateful to the NIH National Cancer Institute for research support (CA-25644 to P.A.K.) and to the Pittsburgh Supercomputer Center, supported by the NSF, for computer time. We also acknowledge the UCSF Computer Graphics Lab (supported by NIH RR01081, R. Langridge, PI). We are grateful to Dr. Neil A. Burton for reading and commenting on the manuscript.

Supplementary Material Available: Tables listing bond distances and angles (20 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, and can be ordered from the ACS; see any current masthead page for ordering information.