A CCSD(T) Study of the Relative Stabilities of Cytosine Tautomers

Rika Kobayashi

University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England

Received: July 10, 1998; In Final Form: October 19, 1998

Coupled-cluster (CC) calculations at the single, double (SD), and perturbative triple (T) excitation level have been carried out to establish the relative stabilities of the tautomers of cytosine. The basis set used was of triple-ζ quality, making this study the highest quality approach to date. The sensitivity of the results is discussed in terms of geometry, basis set, and method and in light of previous studies. In particular, the order of stability of the tautomers at the CCSD level found in a recent study by Fogarasi is found not to be the case at the CCSD(T) level. Comparison is also made with B3LYP density-functional theory results.

1. Introduction

Nucleic acid bases make up the structure of DNA and play a major part in the transmission of the genetic code. They predominantly occur in the one isomeric form whose interactions ensure the faithful replication of the code. However, they can exist in other tautomeric forms (structures differing in the position of a proton), determined by their environments, which could lead to base-pair mismatching and thus mutations.[1] Consequently, there have been many studies on the structures and energetics of the tautomers. This paper is concerned with cytosine (see Figure 1 for structures and atom numbering), which is proving extremely difficult to study theoretically.

Experimentally, both the amino-hydroxy (**2**) and amino-oxo (**1**) forms have been identified in matrix isolation infrared studies,[2] with **2** found in higher concentration. There is also some evidence for the presence of a small amount of the imino-oxo form (**3**). Microwave spectroscopy has yielded rotational constants for three tautomers which correspond to these structures.[3]

On the theoretical side, many calculations have been carried out on the relative stabilities of the cytosine tautomers (refs 4−9 and references therein). The energy differences between them are all quite small and extremely sensitive to the approach used, giving rise to various results. The highest quality calculation to date is that of Fogarasi,[4] who calculated single-point coupled-cluster single and double (CCSD) energies at MP2-optimized geometries with a DZP basis. Calculations were presented on the five lowest isomeric forms of cytosine, and differences were found in the energy orderings depending on the method used. In particular, a difference in ordering was found between MP2 and CCSD, which generally are qualitatively comparable methods. The final order of stability Fogarasi arrived at was

$$\text{amino-hydroxy} < \text{imino-oxo} < \text{amino-oxo}$$

a qualitatively new result ascribed to the added accuracy of the coupled-cluster method.

Because of computational limitations, Fogarasi was not able to investigate the effects of including the triple excitations (T), which can be significant.[8,9] The contributions of the triple excitations have been considered in earlier studies by Leś et al.[9] through CCSD+T calculations and by Gould et al.[8] through QCISD(T) and full MP4, but these were with what would now be considered small basis sets. There is also a study by Colominas et al.,[5] who grafted the MP4 contribution using a 6-31G(d) basis to the MP2/6-311++G(d,p) result.

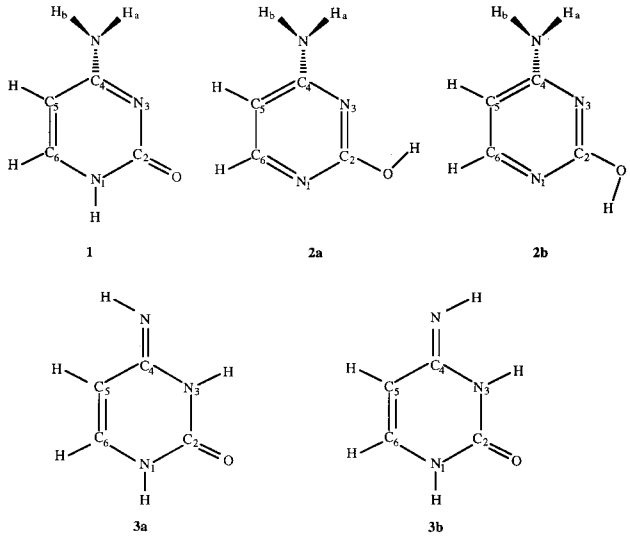

Figure 1. Structures of the cytosine tautomers investigated in this study.

With the recent implementation for massively parallel computers of the coupled-cluster energy with perturbative triples correction,[10] such calculations are no longer prohibitively costly, and so, an improvement on these results through calculations at the highest correlated level to date, CCSD(T) with triple-ζ quality basis, is presented.

2. Computational Details

Large-scale coupled-cluster geometry optimizations are still beyond the scope of current technology, and so, optimized MP2 geometries were chosen for the energy calculations. Starting from the planar structures given by Fogarasi, the geometries of the isomers were optimized with the CADPAC program package[11] on an SGI Origin 2000. The basis set used throughout was the Dunning cc-pvtz basis set,[13] excluding the *f* functions, thus giving a total of 245 basis functions. Single-point CCSD(T) energies (with frozen core) were then calculated at these

10.1021/jp9829546 CCC: $15.00
Published on Web 12/03/1998

TABLE 1. Energies (au) and Relative Stabilities (kcal mol^{-1}) of the Planar Cytosine Tautomers Using the cc-pvtz(-f) Basis at Various Levels of Theory

conformer	SCF	MP2	CCSD//MP2	CCSD(T)//MP2	DFT//MP2	DFT//DFT
			Energy			
cytosine **1**	−392.743 555	−394.080 649	−394.095 496	−394.162 796	−394.855 778	−394.855 970
cytosine **2a**	−392.745 189	−394.082 028	−394.096 116	−394.163 619	−394.853 869	−394.854 057
cytosine **2b**	−392.746 340	−394.083 239	−394.097 302	−394.164 792	−394.855 087	−394.855 286
cytosine **3a**	−392.741 532	−394.079 574	−394.096 465	−394.163 207	−394.853 422	−394.853 674
cytosine **3b**	−392.738 650	−394.077 060	−394.094 042	−394.160 829	−394.850 697	−394.850 951
			Relative Stability			
cytosine **1**	+1.748	+1.625	+1.133	+1.253	−0.434	−0.429
cytosine **2a**	+0.722	+0.760	+0.744	+0.736	+0.764	+0.771
cytosine **2b**						
cytosine **3a**	+3.017	+2.300	+0.525	+0.995	+1.045	+1.012
cytosine **3b**	+4.826	+3.877	+2.046	+2.487	+2.755	+2.720

TABLE 2. Fully Optimized MP2 and DFT Geometries for the Cytosine Tautomers with a cc-pvtz(-f)[a]

	MP2					DFT				
parameter	**1**	**2a**	**2b**	**3a**	**3b**	**1**	**2a**	**2b**	**3a**	**3b**
N1−C2	1.413	1.325	1.330	1.387	1.380	1.425	1.325	1.332	1.392	1.385
C2−N3	1.376	1.333	1.328	1.373	1.380	1.369	1.333	1.326	1.374	1.382
N3−C4	1.312	1.335	1.334	1.398	1.404	1.314	1.337	1.337	1.406	1.412
C4−C5	1.429	1.396	1.399	1.449	1.447	1.437	1.405	1.408	1.457	1.455
C5−C6	1.352	1.377	1.375	1.343	1.341	1.353	1.379	1.377	1.342	1.340
C6−N1	1.351	1.341	1.342	1.373	1.378	1.351	1.339	1.340	1.375	1.380
C2−O	1.219	1.349	1.349	1.217	1.218	1.215	1.347	1.347	1.214	1.214
C4−N	1.367	1.373	1.372	1.283	1.283	1.362	1.368	1.366	1.278	1.277
N1−H	1.006			1.003	1.003	1.007			1.005	1.005
N3−H				1.008	1.007				1.009	1.008
C5−H	1.071	1.073	1.073	1.072	1.071	1.078	1.080	1.079	1.078	1.076
C6−H	1.075	1.078	1.077	1.074	1.074	1.081	1.084	1.084	1.080	1.080
N−Ha	1.005	1.005	1.005			1.006	1.006	1.007		
N−Hb	1.002	1.003	1.003			1.004	1.005	1.004		
N−H				1.014	1.018				1.016	1.019
O−H		0.965	0.965				0.966	0.967		
C6−N1−C2	123.8	114.1	114.6	123.3	123.3	123.3	114.3	114.7	123.2	123.2
N1−C2−N3	116.0	128.3	128.3	113.2	113.3	116.1	128.1	128.2	113.5	113.6
C2−N3−C4	119.9	116.3	115.8	128.1	128.0	120.4	116.4	116.0	127.8	127.7
N3−C4−C5	124.4	121.1	121.7	113.7	113.5	124.0	121.0	121.5	113.7	113.4
C4−C5−C6	116.2	116.6	116.6	119.9	120.2	116.1	116.2	116.2	120.0	120.3
C5−C6−N1	119.6	123.7	123.1	121.7	121.7	120.0	123.9	123.4	121.8	121.8
N1−C2−O	118.9	115.4	116.9	122.7	123.3	118.3	115.6	116.6	122.3	123.0
N3−C4−N	117.0	116.6	116.3	117.2	125.0	117.1	116.8	116.4	117.3	124.7
C2−N1−H1	115.0			115.1	115.2	115.2			115.3	115.3
C2−O−H		105.4	105.0				106.4	105.9		
C2−N3−H				115.6	113.5				115.8	113.8
C4−N−Ha	114.3	113.8	113.7			116.1	115.8	115.7		
C4−N−Hb	117.7	116.2	116.6			119.7	118.2	118.6		
C4−N−H				109.2	110.9				110.6	112.2
C4−C5−H	122.3	121.8	121.9	119.5	118.1	122.3	122.0	122.0	119.1	117.8
C5−C6−H	123.7	120.7	121.0	122.9	123.0	123.0	120.3	120.6	122.6	122.7
C2−N3−C4−N	177.0	177.0	177.1			177.8	177.6	177.8		
N3−C4−N−Ha	15.4	19.6	18.5			11.8	16.1	14.9		
C5−C4−N−Hb	−25.5	−26.8	−26.8			−18.7	−21.3	−20.7		
A (MHz)	3897	3911	3972	3877	3896	3879	3901	3968	3863	3877
B (MHz)	2027	2029	2012	2028	2011	2025	2027	2008	2024	2009
C (MHz)	1335	1337	1337	1332	1327	1331	1335	1334	1328	1323
μ (D)	6.12	4.35	3.15	4.54	2.30	6.36	4.49	3.15	4.61	2.40
A[a] (MHz)	3872		3952	3848		3872		3952	3848	
B[a] (MHz)	2025		2009	2026		2025		2009	2026	
C[a] (MHz)	1330		1332	1328		1330		1332	1328	

[a] Bond lengths are in angstroms. Bond angles are in degrees. [b] Experimental rotational constants from Brown et al.[3]

geometries using the parallel program package NWChem[12] on a 128 processor CRAY T3E at CINECA Computer Centre. The energies obtained are presented in Table 1. Table 1 also contains the results of density-functional theory (DFT) calculations using the B3LYP functional[14] at the DFT-optimized geometries and the MP2-optimized geometries. Second-derivative DFT calculations were carried out and confirmed that the planar structures for tautomers **1**, **2a**, and **2b** were transition states with the imaginary frequency corresponding to an out-of-plane motion of the NH_2 group.

The geometry optimizations were repeated at the DFT level, distorting the NH_2 group sufficiently to prevent it from optimizing back to the planar structure, and their nature again was confirmed by second-derivative calculations, from which the zero-point energies were later calculated. The DFT geometries were then used as the starting points for MP2 reoptimi-

TABLE 3. Energies (au) and Relative Stabilities (kcal mol^{-1}) of the Fully Optimized Cytosine Isomers Using the cc-pvtz(-f) Basis at Various Levels of Theory

conformer	SCF	MP2	CCSD//MP2	CCSD(T)//MP2	DFT//MP2	DFT//DFT
			Energy			
cytosine **1**	−392.743 565	−394.081 399	−394.096 236	−394.163 731	−394.855 778	−394.856 136
cytosine **2a**	−392.745 479	−394.083 275	−394.097 268	−394.164 996	−394.854 126	−394.854 444
cytosine **2b**	−392.746 568	−394.084 367	−394.098 340	−394.166 047	−394.855 275	−394.855 603
cytosine **3a**	−392.741 532	−394.079 574	−394.096 465	−394.163 207	−394.853 674	−394.853 674
cytosine **3b**	−392.738 650	−394.077 060	−394.094 042	−394.160 829	−394.850 697	−394.850 951
			Relative Stability			
cytosine **1**	+1.884	+1.862	+1.320	+1.453	−0.316	−0.335
cytosine **2a**	+0.683	+0.685	+0.673	+0.660	+0.721	+0.727
cytosine **2b**						
cytosine **3a**	+3.160	+3.008	+1.177	+1.782	+1.005	+1.210
cytosine **3b**	+4.969	+4.585	+2.697	+3.274	+2.873	+2.919

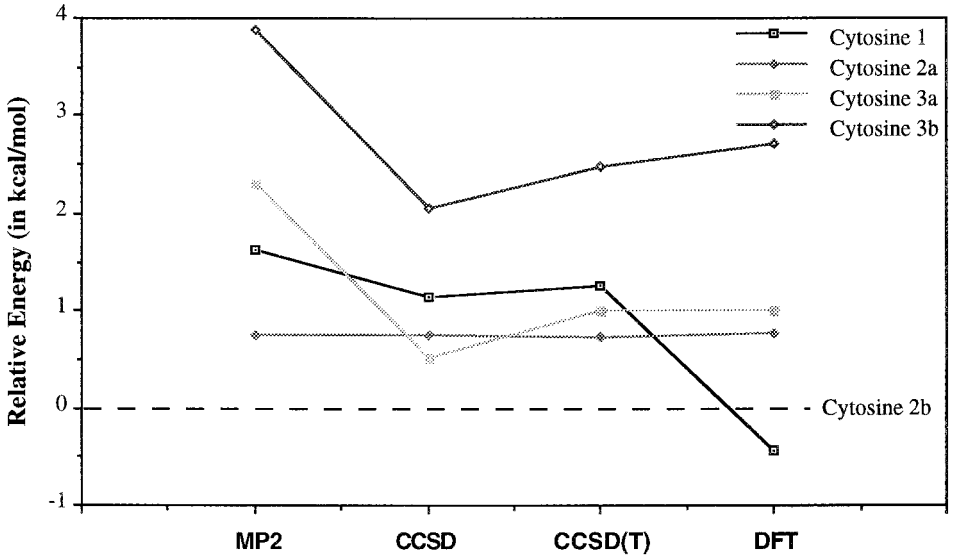

Figure 2. Energies (kcal/mol) of the planar cytosine tautomers relative to **2b**.

zations. The fully optimized structures and corresponding energies are presented in Tables 2 and 3, respectively.

3. Results and Discussion

3.1. Geometries. The structural features of the cytosine tautomers have already been discussed extensively[4,7] and will not be dealt with in depth here. Any differences in geometry from previous works were found to be quantitative rather than qualitative, and so, only the parameters for the fully optimized geometries are presented (Table 2). The changes in going from planarity to nonplanarity only appeared to affect the parameters associated with the NH_2 group, the rest of the ring framework remaining almost identical. However, as will be discussed later, this is enough to make a significant difference.

The MP2-optimized geometries of all the isomers mirror previously reported MP2 results.[4,5] However, the bond lengths are consistently shorter than those of Fogarasi[4] (using a DZP basis) by, on average, 0.01 Å. They correlate well with those of Colominas et al.,[5] who used a 6-31g(d) basis, with the bond lengths differing by only about 0.007 Å.

The DFT geometries were much the same as the MP2 geometries, except for the degree of pyramidization of the NH_2 group. However, as Tables 1 and 3 show, the differences in geometry between the MP2 and DFT structures amount to at most 0.035 kcal mol^{-1} in the relative energy and, thus, are not believed to be significant. In general, the geometries do not appear to be so dependent on the basis set, and in addition, the changes in geometry going from method to method are unlikely to be great.

Agreement with the experimental rotational constants[3] is closer than previously found,[4,6,8] with the DFT values being almost exact in some cases.

3.2. Relative Stabilities. The relative stabilities of the planar cytosine structures (Table 1) are plotted in Figure 2. They reproduce those found by Fogarasi, including the unexpectedly different orderings between MP2 (**2a** < **1** < **3a**) and CCSD (**3a** < **2a** < **1**). In addition, there was another surprising change in order going from CCSD to CCSD(T) (**2a** < **3a** < **1**).

For the fully optimized structures (Table 3 and Figure 3), the additional stabilization obtained from relaxing the NH_2 group is enough to change the ordering of the CCSD energies of tautomers **2a** and **3a** yet again. This shows that it is probably not reasonable to use planarity as an approximation, as it can lead to qualitative as well as quantitative errors. The difference between CCSD and CCSD(T) found in the planar case also carries over to the fully optimized results, indicating that the triple excitation contributions should not be neglected at the coupled-cluster level.

The final order at the CCSD(T) level is thus

amino-hydroxy	<	amino-oxo	<	imino-oxo
2b < **2a**	<	**1**	<	**3a** < **3b**

in agreement with the gas-phase experimental studies.[2] The *qualitatively new picture* found by Fogarasi at the CCSD level is not seen at the CCSD(T) level.

The relative stabilities obtained at the DFT level reproduced the "incorrect" results of previous DFT studies,[5,8] which found

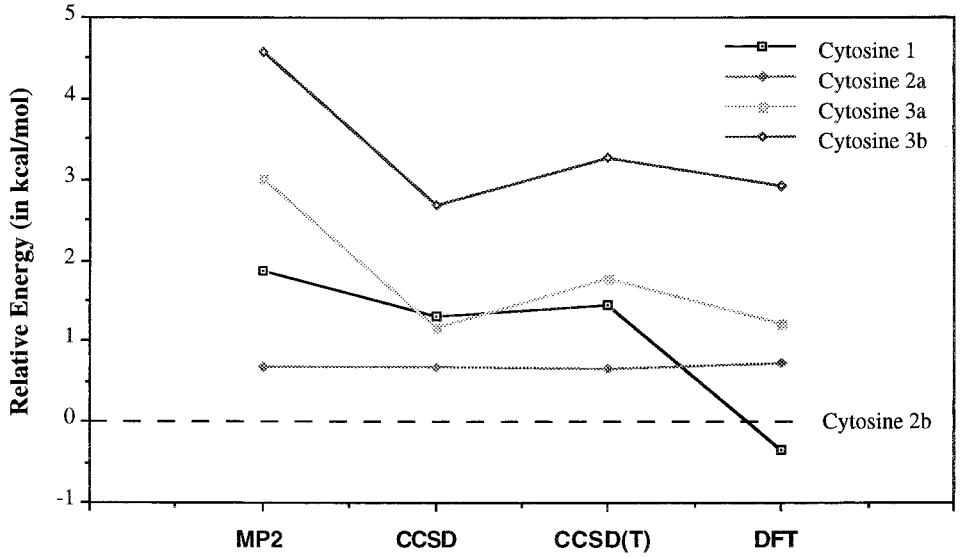

Figure 3. Energies (kcal/mol) of the fully optimized cytosine tautomers relative to **2b**.

TABLE 4. Energy (au) and Relative Stabilities (kcal mol^{-1}) of the Cytosine Isomers with the cc-pvtz Basis at the MP2 and DFT Levels

conformer	SCF	MP2	DFT
		Energy	
cytosine **1**	−392.762 328	−394.365 553	−394.874 385
cytosine **2a**	−392.762 911	−394.367 513	−394.871 830
cytosine **2b**	−392.764 016	−394.368 645	−394.872 990
cytosine **3a**	−392.760 785	−394.364 041	−394.872 257
cytosine **3b**	−392.757 848	−394.361 203	−394.869 571
		Relative Stability	
cytosine **1**	+1.059	+1.940	−0.875
cytosine **2a**	+0.693	+0.710	+0.728
cytosine **2b**			
cytosine **3a**	+2.027	+2.889	+0.460
cytosine **3b**	+3.870	+4.670	+2.145

the amino-oxo form to be the lowest in energy. Kwiatkowski et al.,[6] on the other hand, found the amino-hydroxy form to be the lowest in energy but by an extremely small amount (0.09 kcal mol^{-1}). As the later DFT calculations were all carried out with the same functional (B3LYP) as that used in this study, though with differing basis sets, DFT is probably too sensitive and thus unreliable for such calculations.

Quantitatively, each method yields the same energy gap between the rotamers (structures differing in the orientation of a proton); the main difference appears between tautomers. This indicates that the methods are not able to describe uniformly the variations in bonding for otherwise similar structures. Leś et al.[9] observed upon examining the coupled-cluster wave functions that the contribution from some doubly excited determinants was more pronounced for the imino-oxo form and concluded that it was necessary to go to high orders of the perturbation series for an accurate description of the multireference character of the tautomers. Thus it would be expected that of the results presented in this paper the CCSD(T) ones will be the most accurate.

Finally, calculations were also carried out with the full cc-pvtz basis at the above geometries using MP2 and DFT (Table 4). The differences in energy were of the order of 0.1 kcal mol^{-1} and are unlikely to change the above final order.

3.3. Zero-Point Energies. Full second-derivative calculations were carried out for all five tautomers at the DFT level to determine the zero-point energies (Table 5). The DFT frequencies are regarded as being more reliable than the energies; however for cytosine, comparison with experimental data is made more difficult by the complexity of the spectra, especially with the possible presence of several tautomeric forms. There are additional complications in having to compare theoretical gas-phase harmonic frequencies with experimental matrix perturbed fundamental frequencies. Thus a more comprehensive discussion will be presented separately of the results of the frequency calculations, including those for rotamers **2a** and **3b**, which have not been considered in previous characterizations.[6,9]

The zero-point energy corrections in Table 5 were calculated from the DFT harmonic frequencies. Values obtained by scaling these frequencies by the factor of 0.97, estimated by Kwiatkowski et al.[6] to account for anharmonicity effects, are given in parentheses. The zero-point energy corrections have the effect of lowering the relative stability of **1** and raising that of **3a**. This is sufficient to raise the CCSD energy of **3a** past **1** again, thus ending with the same order of stability for all three ab initio methods.

3.4. Solvent Effects. It should be mentioned that all of the above results pertain to properties in the gas phase, and under different conditions, they can change drastically. Indeed, it is accepted that under aqueous and crystalline environments

TABLE 5. Effects of Zero-Point Energy on the Relative Stabilities of the Cytosine Isomers (kcal mol^{-1})[a]

conformer	MP2	CCSD	CCSD(T)	DFT
		Zero-Point Energy Corrected		
cytosine **1**	+1.655 (+1.661)	+1.113 (+1.119)	+1.246 (+1.252)	−0.542 (−0.536)
cytosine **2a**	+0.672 (+0.672)	+0.659 (+0.659)	+0.646 (+0.646)	+0.714 (+0.714)
cytosine **2b**				
cytosine **3a**	+3.219 (+3.212)	+1.388 (+1.381)	+1.993 (+1.987)	+1.422 (+1.415)
cytosine **3b**	+4.684 (+4.681)	+2.796 (+2.793)	+3.373 (+3.370)	+3.018 (+3.015)

[a] The values in parentheses are the zero-point energy corrections obtained from frequencies scaled by 0.97 to account for anharmonicity.

cytosine exists predominantly in the amino-oxo form, which is the form in which it interacts in DNA.

Solvent effects for the cytosine tautomers have been investigated ab initio by Colominas et al.[5] from self-consistent reaction-field calculations using the continuum model developed by Miertus et al.,[15] which is based on a molecule-shaped cavity. They added their free energies of hydration to their MP4/6-311++G(d,p)//MP2/6-31G(d) gas-phase values which are similar to the values obtained in this study, so the solvent effects should be analogous. They found that the amino-hydroxy (**2**) and imino-oxo (**3**) forms were considerably destabilized by hydration, giving the amino-oxo (**1**) form as by far the preferred form in an aqueous environment.

4. Conclusion

The relative stabilities of the tautomers of cytosine using the CCSD(T) method with cc-pvtz(-f) basis, the highest accuracy method to date, were found to be

amino-hydroxy	<	amino-oxo	<	imino-oxo
2b < **2a**	<	**1**	<	**3a** < **3b**

in agreement with experimental observations.

The small range in energy involved makes the calculation extremely sensitive to factors such as geometry, basis set, and level of correlation, and the results change qualitatively as well as quantitatively, depending on the method. The CCSD result, which would normally be considered to be trustworthy, is found to yield an order of stability different from the CCSD(T) results, indicating that the effects of the triple excitations cannot be neglected and it is necessary to go to high order in the perturbation series. Density-functional theory proves unreliable regarding the calculation of energies, as it predicts the incorrect ordering in many cases and varies greatly with basis set.

Acknowledgment. R.K. thanks CINECA Computer Centre for their grant to use the CRAY-T3E.

References and Notes

(1) (a)Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984. (b) Blackburn, G. M., Gait, M. J. *Nucleic Acids in Chemistry and Biology*; Oxford University Press: Oxford, 1996.

(2) (a) Szczesniak, M.; Szczepaniak, K.; Kwiatkowski, J. S.; KuBulat, K.; Person, W. B. *J. Am. Chem. Soc.* **1988**, *110*, 8319. (b) Nowak, M. I.; Lapinski, L.; Fullara, J. *Spectrochim. Acta* **1989**, *A45*, 229.

(3) Brown, R. D.; Godfrey, P. D.; Mc Naughton, D.; Pierlot, A. P. *J. Am. Chem. Soc.* **1987**, *111*, 2308.

(4) Fogarasi, G. *J. Mol. Struct.* **1997**, *413*, 271.

(5) Colominas, C.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1996**, *118*, 6811.

(6) Kwiatkowski, J. S.; Leszczyński, J. *J. Phys. Chem.* **1996**, *100*, 941.

(7) Sponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161.

(8) Gould, I. R.; Burton, N. A.; Hall, R. J.; Hillier, I. J. *J. Mol. Struct.* **1995**, *331*, 147.

(9) Leś, A.; Adamowicz. L.; Bartlett, R. J. *J. Phys. Chem.* **1989**, *93*, 4001.

(10) Kobayashi, R.; Rendell, A. P. *Chem. Phys. Lett.* **1997**, *265*, 1.

(11) Amos, R. D.; Alberts, I. L.; Andrews, J. S.; Colwell, S. M.; Handy, N. C.; Jayatilaka, D.; Knowles, P. J.; Kobayashi, R.; Laidig, K. E.; Laming, G.; Lee, A. M.; Maslen, P. E.; Murray, C. W.; Rice, J. E.; Simandiras, E. D.; Stone, A. J.; Su, M.-D.; Tozer, D. J. *CADPAC: The Cambridge Analytic Derivatives Package*, Issue 6; A suite of quantum chemistry programs; Cambridge, England, 1995.

(12) High Performance Computational Chemistry Group. *NWChem: A Computational Chemistry Package for Parallel Computers*, Version 3.2; Pacific Northwest National Laboratory: Richland, WA, 1998.

(13) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(14) (a)Becke, A. *J. Chem. Phys.* **1993**, *98*, 5648. (b) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

(15) (a) Miertus, S.; Scrocco, E.; Tomasi, J. *J. Chem. Phys.* **1981**, *55*, 117. (b) Miertus, S.; Tomasi, J. *J. Chem. Phys.* **1982**, *65*, 239.