Ab Initio Calculations with Electronic Correlation (MP2) on the Nucleic Acid Bases and Their Methyl Derivatives

Eugene L. Stewart,† Charles K. Foley,‡ Norman L. Allinger,† and J. Phillip Bowen*,†

Contribution from the Computational Center for Molecular Structure and Design, Department of Chemistry, The University of Georgia, Athens, Georgia 30602, and Cray Research, Inc., P.O. Box 12746, Research Triangle Park, North Carolina 27709

Received February 15, 1994. Revised Manuscript Received May 12, 1994®

Abstract: High-level *ab initio* optimizations have been carried out on the nucleic acid bases thymine, cytosine, guanine, and adenine in an attempt to calculate gas-phase structures of these bases and their methyl derivatives. Calculations were carried out at both the Hartree–Fock (HF) and Møller–Plesset (MP2) levels of theory, utilizing the standard 6-31G** basis set. Full optimizations on these structures were undertaken, resulting in planar or near-planar (C_s symmetry) conformers. The results show a large deviation between the HF and MP2 geometries, implying that full optimization is necessary with the inclusion of electron correlation. The calculated rotational constants of the bases are in excellent agreement with those derived from gas-phase microwave results and are in accord with the existence of the planar forms in the gas phase. The MP2 structures of guanine and the methyl derivatives of the nucleic acid bases compare well with structures obtained by room temperature X-ray and neutron diffraction.

Introduction

With the growing importance of computational chemistry in understanding structure–activity relationships in biological systems, accurately characterizing certain classes of compounds with computational methods has become imperative. Because most biological systems involve macromolecules, calculations involving these systems are limited to molecular mechanics for the most part. Possibly the most biologically important macromolecule is the DNA helical strand. In understanding the biological activity of this molecule, a complete understanding of its structure and conformation is of great importance. For the purposes of computational investigations, such as docking and binding studies, in a molecular mechanics formalism, accurate characterization of the three-dimensional structure of this molecule in the gas phase is necessary. Ultimately, the addition of solvation effects will be made to truly model the DNA system, but, initially, characterization of its structure in the gas phase is desirable.

Among the principal structural features of the DNA molecule are the nucleic acid base pairs. The literature contains much structural data, primarily room temperature X-ray diffraction data, on derivatives of these bases, nucleotides, and nucleosides.[1] Initially, when first formulating a force field to model these bases, one wishes to characterize the simplest nucleic acid bases, i.e., those without the deoxyribose sugar moiety or the phosphate backbone, in the gas phase. Unfortunately, though, structures of even the nucleic acid bases and their methyl derivatives have usually been studied at room temperature by X-ray or neutron diffraction only.[2] The most significant problem with these structural results is that they contain errors due to thermal motion and uncertain distortions resulting from the intermolecular hydrogen bonding of the base pairs. Additionally, since X-ray diffraction spectroscopy maps the electron density, other apparent distortions may result from lone pairs present on oxygen and nitrogen. One would optimally prefer data which were acquired in the gas phase or at low temperatures in which many of these problems are alleviated. Unfortunately, the only gas-phase data currently available are the rotational constants derived from microwave studies on cytosine,[3] adenine,[4] and thymine.[5] Because of this limited amount of gas-phase data, semiempirical and *ab initio* formalisms have become the methods of choice for gas phase structural studies of these compounds.

To date, the best *ab initio* calculations on the nucleic acid bases, their derivatives, and their tautomers have usually involved full geometry optimizations at the Hartree–Fock (HF) LCAO-MO level of theory utilizing relatively large basis sets with polarization functions (6-31G* and 6-31G**).[6] Some post-Hartree–Fock methods have also been used, mainly Møller–Plesset (MP2), on the HF optimized geometry in an attempt to improve the energetics of these structures, their tautomers, and some important DNA lesions such as thymine glycol and 5,6-dihydrothymine.[6] Other computational studies at this level of theory have involved the modeling of the low-energy tautomers of cytosine in the aqueous phase using the reaction field and continuum solvation models.[7] Santhosh and Mishra recently mapped the potential energy surfaces of gaunine, guanine-O_2, and guanine-O_2–water with various semiempirical methods.[8] Only recently, Leszczyński performed full geometry optimizations at the MP2 level of theory using the 6-31G* basis set on the C_s structure of cytosine, its tautomers, and 1-methylcytosine.[9]

Other than Leszczyński's work, no full geometry optimizations at the MP2 level of theory have been reported on the nucleic acid bases and their methyl derivatives.[10] As will be demonstrated in this study, calculations of this type are important, since large

† The University of Georgia.
‡ Cray Research, Inc.
® Abstract published in *Advance ACS Abstracts,* June 15, 1994.

(1) (a) Kraut, J.; Jensen, L. H. *Acta Crystallogr.* **1963**, *16*, 79. (b) Rao, S. T.; Sundaralingam, M. *J. Am. Chem. Soc.* **1970**, *92*, 4963. (c) Lai, T. F.; Marsh, R. E. *Acta Crystallogr.* **1972**, *B28*, 1982. (d) Takimoto, M.; Takenata, A.; Sasada, Y. *Bull. Chem. Soc. Jpn.* **1981**, *54*, 1635.

(2) (a) Kvick, A.; Koetzle, T. F.; Thomas, R. *J. Chem. Phys.* **1974**, *61*, 2711. (b) Rossi, M.; Kistenmacher, T. J. *Acta Crystallogr.* **1977**, *B33*, 3962. (c) Stewart, R. F.; Jensen, L. H. *J. Chem. Phys.* **1964**, *40*, 2071. (d) Thewalt, U.; Bugg, C. E.; Marsh, R. E. *Acta Crystallogr.* **1971**, *B27*, 2358.

(3) Brown, R. D.; Godfrey, P. D.; McNaughton, D.; Pierlot, A. P. *J. Am. Chem. Soc.* **1989**, *111*, 2308.

(4) Brown, R. D.; Godfrey, P. D.; McNaughton, D.; Pierlot, A. P. *Chem. Phys. Lett.* **1989**, *156*, 61.

(5) Brown, R. D.; Godfrey, P. D.; McNaughton, D.; Pierlot, A. P. *J. Chem. Soc., Chem. Commun.* **1989**, 37.

(6) (a) Gould, I. R.; Hillier, I. H. *Chem. Phys. Lett.* **1989**, *161*, 185. (b) Leszczyński, J. *Chem. Phys. Lett.* **1990**, *174*, 347. (c) Leś, A.; Adamowicz, L.; Nowak, M. J.; Lapinski, L. *Spectrochim. Acta* **1992**, *48A*, 1385. (d) Miaskiewicz, K.; Miller, J.; Osman, R. *Int. J. Radiat. Biol.* **1993**, *63*, 677.

(7) Gould, I. R.; Green, D. V. S.; Young, P.; Hillier, I. H. *J. Org. Chem.* **1992**, *57*, 4434.

(8) Santhosh, C.; Mishra, P. C. *Int. J. Quantum Chem.* **1992**, *42*, 1659.

(9) Leszczyński, J. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1992**, *19*, 43.

0002-7863/94/1516-7282$04.50/0 © 1994 American Chemical Society

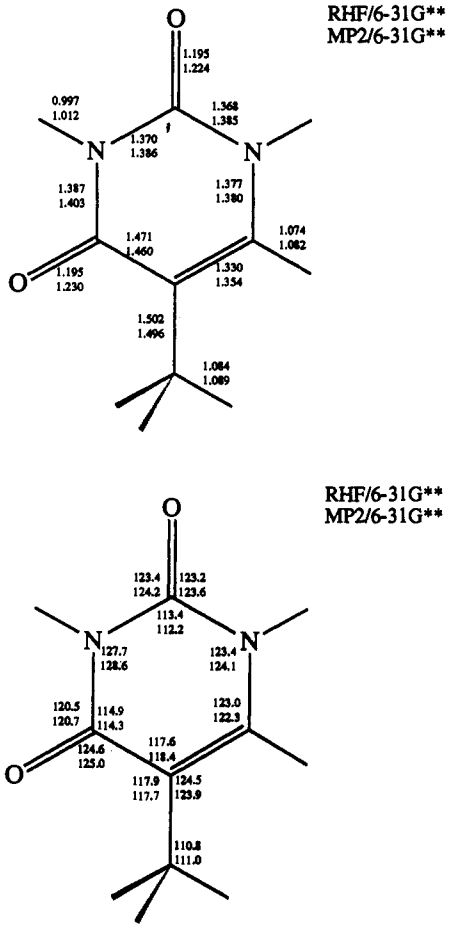


Figure 1. RHF/6-31G** and MP2/6-31G** bond lengths (Å) and bond angles (deg) for thymine.

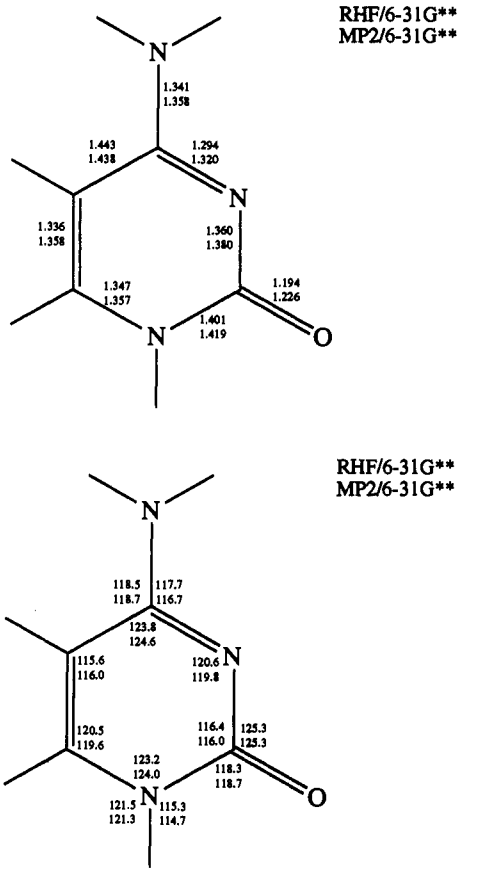


Figure 2. RHF/6-31G** and MP2/6-31G** bond lengths (Å) and bond angles (deg) for cytosine.

differences in geometries occur upon the inclusion of electron correlation. In this investigation, full geometry optimizations of the four nucleic acid bases (adenine, guanine, cytosine, and thymine) and their methyl derivatives have been undertaken at the MP2/6-31G** level of theory to obtain the structures of the C_s conformers of these compounds.

Methodology

Ab initio LCAO-MO methods[11] were used in the calculation of the C_s structures of adenine, guanine, thymine, and cytosine. Additional calculations were performed on the methyl derivatives of these compounds (9-methyladenine, 9-methylguanine, 1-methylthymine, and 1-methylcytosine) for comparison with available X-ray and neutron diffraction data. All calculations utilized the Gaussian 90 and 92[12] suites of programs. For the nucleic acid bases, full geometry optimizations (without symmetry restrictions) with standard gradient methods were used at both the HF and MP2 levels with the standard 6-31G** basis set. For the methyl derivatives, the C_s structures were calculated only at the MP2 level of theory employing the 6-31G** basis set.

At the HF level of theory, stationary points were verified through vibrational frequency calculations. MP2 vibrational frequency calculations were also carried out analytically on the nucleic acid bases to verify that the optimized structures were stationary points. All calculations were undertaken on an IBM RISC6000 cluster of machines, a Cray Y-MP/464, or a Cray C98/8512. For the larger nucleic acid bases, specifically the purines, an enormous amount of CPU time and other resources was necessary for optimization. For example, starting at the HF geometry, guanine required 7.6 h of user CPU time and 7 min of system CPU time on a Cray Y-MP/464. Data are not available for other resources used such as disk space, memory, etc. One must realize, though, that the time required for a geometry optimization is extremely dependent on the starting geometry.

Results and Discussion

Nucleic Acid Bases. The results of the HF and MP2 calculations of thymine, cytosine, guanine, and adenine are illustrated in Figures 1–4. For guanine, the calculated structures were compared to room temperature X-ray diffraction results of the monohydrate.[2d] The conformations of these rings are not displayed in these figures but were found to deviate only slightly from the C_s structures. The HF and MP2 vibrational frequency results, however, show that cytosine is a transition-state structure in the planar C_s conformation. The imaginary frequency is the result of the planar geometry of the amine (NH_2) side chain. Similar results were found in the cases of guanine and adenine. It is interesting to note that *ab initio* calculations tend to find the planar amino group to be a transition state, while the pyramidal geometry is a minimum lying approximately 0.4 kcal/mol lower in energy than the planar form.[13] Due to this small energy difference and for computational ease, the calculations reported herein involved only the planar (C_s) conformers of the nucleic acid bases and their methyl derivatives.

The resulting bond lengths and bond angles are essentially as were predicted: there are significant differences between bond

(10) After this manuscript was submitted and during the review process, Šponer and Hobza reported optimizations of the nucleic acid bases with the 6-31G*, 6-31G**, DZP+, and DZ(2d) basis sets at the MP2 level of theory (Šponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161). No optimizations, though, were reported for the methyl derivatives of the bases. The results obtained in this study are comparable to the results obtained by Šponer and Hobza where they overlap.

(11) Hehre, W. H.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(12) Gaussian 90 and Gaussian 92, Revision B. M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople, Gaussian, Inc.: Pittsburgh, PA, 1992.

(13) Riggs, N. V. *Chem. Phys. Lett.* **1991**, *177*, 447.

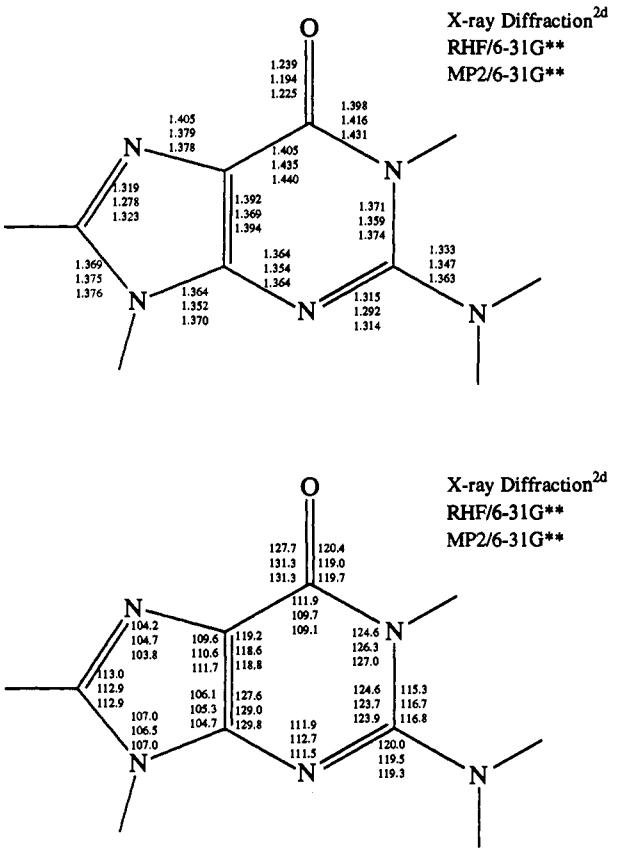


Figure 3. X-ray diffraction, RHF/6-31G**, and MP2/6-31G** bond lengths (Å) and bond angles (deg) for guanine.

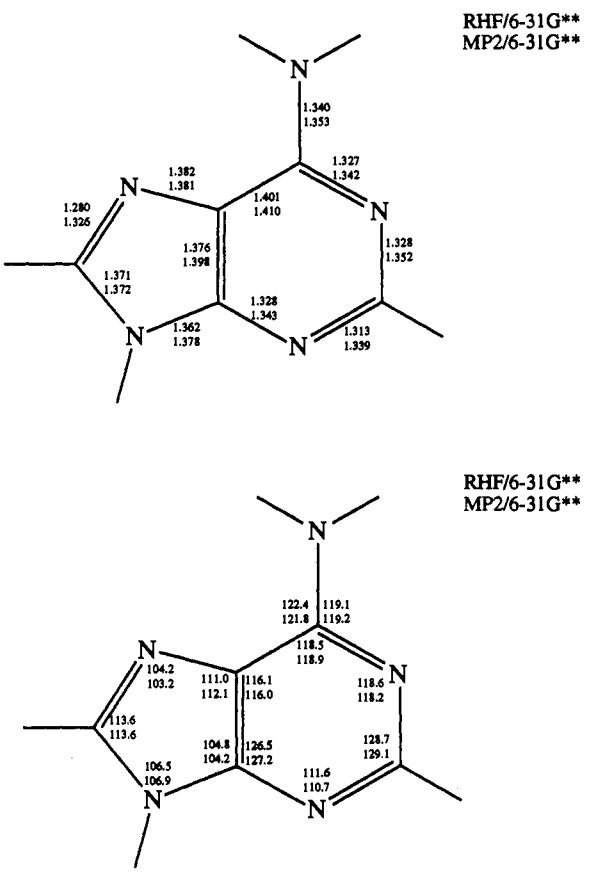
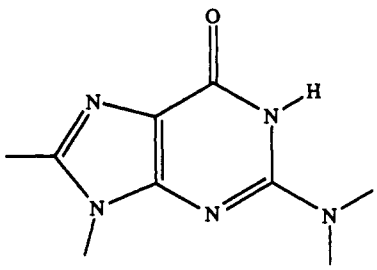


Figure 4. RHF/6-31G** and MP2/6-31G** bond lengths (Å) and bond angles (deg) for adenine.

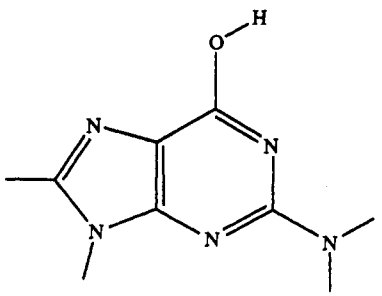
lengths and angles calculated at the HF level of theory and those calculated with electron correlation. The variations in the bond lengths differ somewhat and are as small as 0.001 Å to as large as 0.045 Å. The bond angles are somewhat erratic, varying from 0.1° to approximately 3.0°. This clearly illustrates the importance of full geometry optimization for molecules such as these after the inclusion of electron correlation in the wave function: significant changes in the geometry may occur. The most significant changes in geometry when correlation effects are incorporated occurs in those portions of the nucleic acid bases in which conjugation is significant. For example, in guanine, many of the bond lengths increase by up to 0.04 Å after correlation effects are included.

The most interesting structural feature is the measure of the C—C=O bond angle for guanine observed at both the HF and MP2 levels of theory. In both instances, the bond angle measures 131.3° (Figure 3), which agrees with experiment and is a significant deviation from the trigonal-planar geometry of an sp^2-hybridized carbon (approximately 120°). It is believed that this angle strain may be related to a preference of a second tautomer for guanine (Figure 5). For this reason, further *ab initio* optimizations at the MP2 level of theory were undertaken to determine if this second tautomer (**1B**) is, in fact, more stable than the one calculated above (**1A**). Results of these MP2 optimizations showed, in fact, that these two tautomers are essentially equal in energy, with tautomer **1B** being only 0.278 kcal/mol higher in energy than **1A**.

Table 1 illustrates a comparison of the calculated and experimental rotational constants of adenine, cytosine, and thymine. Brown and co-workers[3,4,5] carried out microwave studies on these three bases in order to (1) determine the dominant tautomer of these compounds in the gas phase, (2) compare the microwave and calculated rotational constants to obtain structure, and (3) determine if these molecules were planar in the gas phase. Table 1 shows an excellent correspondence between the experimentally derived rotational constants and those calculated at the MP2 level of theory. This good correlation between theory and experiment clearly supports the conclusion drawn by Brown and co-workers: the gas-phase conformations of these three structures are essentially C_s planar. Based on this excellent correlation, it can also be concluded that the MP2 structures of adenine, cytosine, and thymine are essentially those found in the gas phase. The



1A



1B

Figure 5. Two major tautomeric forms of guanine.

Table 1. RHF/6-31G**, MP2/6-31G**, and Microwave Rotational Constants for Adenine, Cytosine, and Thymine

structure	rotational constant[a]	RHF/ 6-31G**	MP2/ 6-31G**	microwave
adenine	A	2426.745	2360.114	2371.873 (4)
	B	1596.1980	1574.0749	1573.3565 (8)
	C	962.8687	944.2852	946.2576 (4)
cytosine	A	3959.885	3863.816	3871.547 (9)
	B	2068.046	2013.579	2024.930 (3)
	C	1358.546	1323.733	1330.306 (1)
thymine	A	3256.3030	3197.6939	3201.2088 (20)
	B	1425.7965	1394.6240	1404.8110 (27)
	C	99.7621	97.6903	98.6095 (24)

[a] The calculated and experimental rotational constants are expressed in MHz. The gas-phase microwave results for cytosine, adenine, and thymine are taken from refs 3, 4, and 5, respectively.

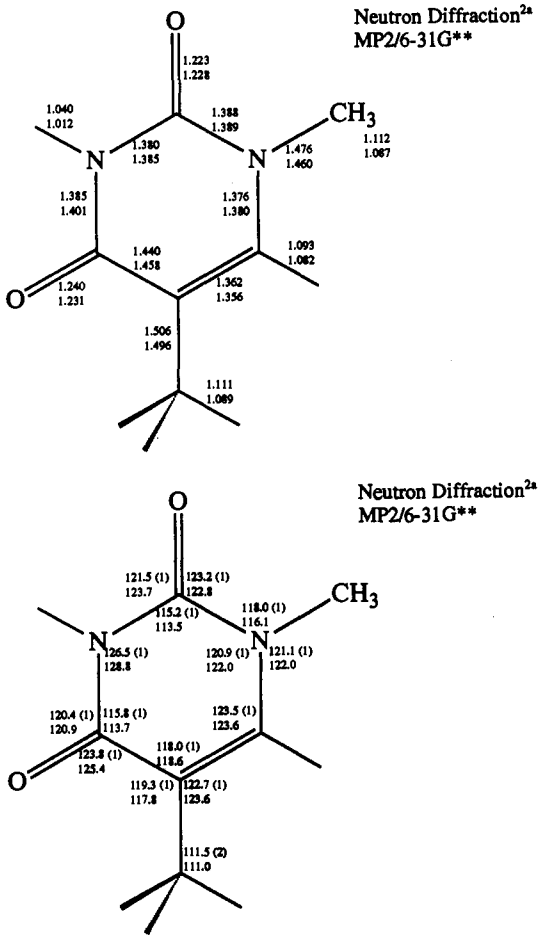


Figure 6. Neutron diffraction and MP2/6-31G** bond lengths (Å) and bond angles (deg) for 1-methylthymine.

HF geometries, however, do not correlate as well; thus, again, verifying that geometry optimizations at the MP2 level of theory are important for these compounds, not only for calculating energies, but also for calculating structures.

Methyl Derivatives. The structural results of the MP2 optimizations of the methyl derivatives (1-methylthymine, 1-methylcytosine, 9-methylguanine, and 9-methyladenine) are illustrated in Figures 6–9 and are compared with available room temperature X-ray diffraction and neutron diffraction data.[2a–c] For the majority of the structural results, there is good agreement between the MP2/6-31G** and experimental data. Most of the differences in bond lengths are within 0.010 Å, with the N–H or C–H bond lengths deviating the most, usually around 0.020 Å. This deviation is expected, though, since the bond lengths calculated by *ab initio* methods are r_e and should be somewhat shorter than those obtained by neutron diffraction (r_α). This phenomenon is most pronounced

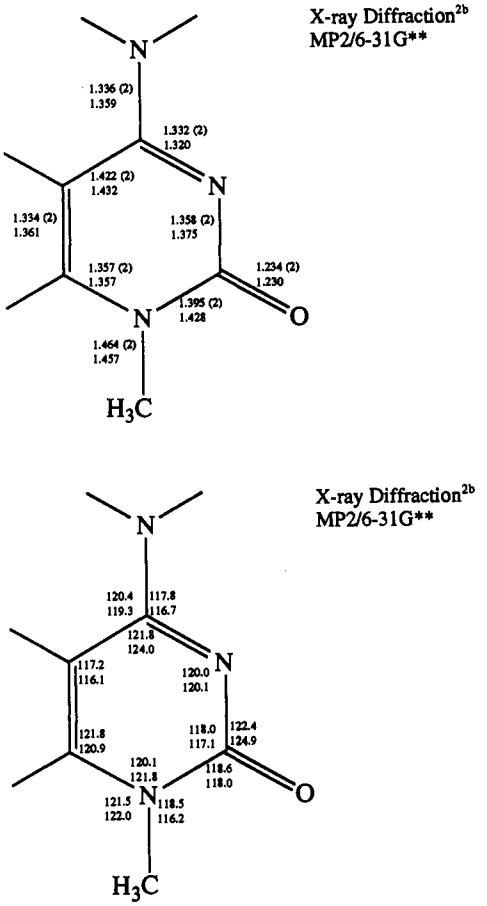


Figure 7. X-ray diffraction and MP2/6-31G** bond lengths (Å) and bond angles (deg) for 1-methylcytosine.

in the case of bonding involving hydrogen and is due to the effect of anharmonicity, which is severe because of the low mass of the hydrogen.

Excellent agreement between calculated and experimental bond lengths was found in 1-methylthymine (Figure 6). The differences in the bond lengths between the heavy atoms are quite small, less than 0.008 Å on average. All of the calculated r_e bond lengths are longer than those obtained experimentally (r_α). This systematic difference is apparently due to the thermal motions of the molecules in the crystals, which, since the structure was determined at room temperature, would tend to shorten the observed bond lengths. If the bond lengths, both calculated and experimental, of the other methyl bases are examined, the observed differences become somewhat more erratic; in a few cases, the calculated bond lengths are longer than those determined experimentally. These erratic differences could be due to (1) errors in the *ab initio* calculations, (2) thermal motion in the crystal, since many of these structural determinations were performed at room temperature, (3) crystal packing forces that, in these cases, may be attributed mainly to intermolecular hydrogen bonding of the bases in the crystal, (4) comparisons of r_e and r_α values, and (5) use of X-ray data for carboxyl oxygens and amino groups, where centers of electron densities and nuclei do not coincide.

The calculated and experimental bond angles of the methylated bases are in reasonable agreement. In most cases, the bond angles differ by approximately 1°. The largest errors appear to be approximately 2.5° and usually occur near the locations of intermolecular hydrogen bonds. Therefore, any large deviations in the bond angles are likely due primarily to crystal forces (specifically hydrogen bonding) rather than to thermal motion of the molecules in the crystal.

It is important once again to stress that X-ray and neutron diffraction results do not completely characterize the gas-phase

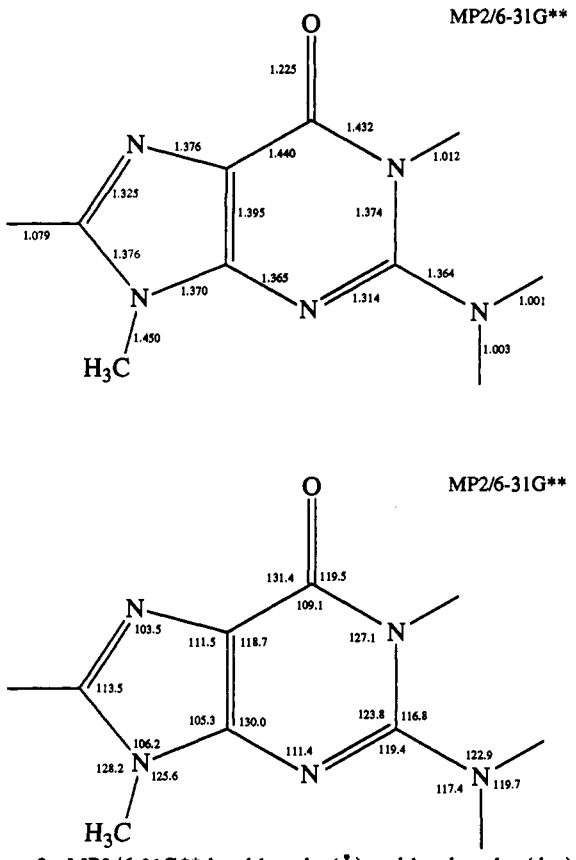


Figure 8. MP2/6-31G** bond lengths (Å) and bond angles (deg) for 9-methylguanine.

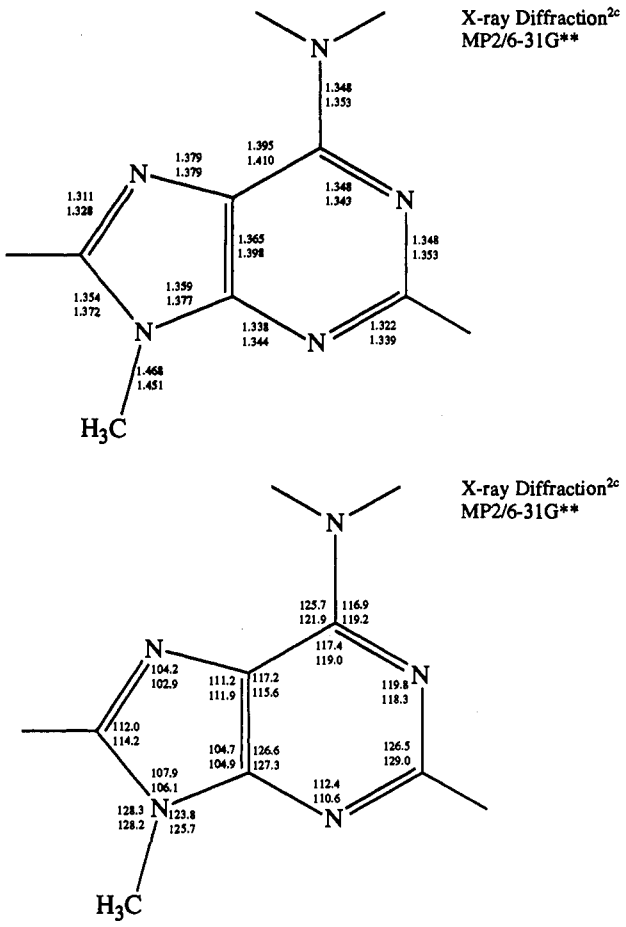


Figure 9. X-ray diffraction and MP2/6-31G** bond lengths (Å) and bond angles (deg) for 9-methyladenine.

structures due to forces within the crystal. Thus, a comparison between the *ab initio* (gas phase) and diffraction data (crystalline phase) is not entirely valid. Unfortunately, data of this type are the only data currently available for most of these molecules.

Conclusions

Geometry optimizations of the DNA nucleic acid bases have been carried out with the 6-31G** basis set, at both the HF and MP2 levels. The results show that there is a significant difference in the geometries obtained by the correlated and noncorrelated methods. Thus, in the case of these molecules, full geometry optimizations are desirable with the inclusion of electron correlation. Calculated rotational constants correlate well with those obtained by microwave spectroscopy, supporting the hypothesis that the correlated calculations are reproducing the structures of these molecules in the gas phase.

MP2 level calculations with the 6-31G** basis set were also carried out on the methyl-substituted nucleic acid bases for comparison with available X-ray and neutron diffraction data. The results of the 1-methylthymine calculation were found to be in good agreement with the neutron diffraction data, with most calculated bond lengths being about 0.008 Å longer than the experimental values. Other structural results show some deviation from the experimental values, but the agreement is reasonable.

Acknowledgment. The authors wish to gratefully acknowledge Cray Research, Inc., for a grant of computer time and the North Carolina Supercomputing Center for use of the Cray Y-MP/464.