UV Photoelectron and Ab Initio Quantum Mechanical Characterization of 2′-Deoxyguanosine 5′-Phosphate: Electronic Influences on DNA Alkylation Patterns

Ho Soon Kim, Min Yu, Qing Jiang, and Pierre R. LeBreton*

Contribution from the Department of Chemistry, The University of Illinois at Chicago, Chicago, Illinois 60680

Received October 13, 1992

Abstract: Data from He(I) UV photoelectron measurements have been used in conjunction with results from ab initio SCF molecular orbital calculations, and ab initio and semiempirical post-SCF calculations to describe the valence electronic structures of neutral and anionic 2′-deoxyguanosine 5′-phosphate (5′-dGMP). Analogous to 2′-deoxycytidine 5′-phosphate (Tasaki, K.; Yang, X.; Urano, S.; Fetzer, S.; LeBreton, P. R. *J. Am. Chem. Soc.* **1990**, *112*, 538–548), results from calculations on 5′-dGMP and 5′-dGMP$^-$, carried out at the SCF level with a split-valence basis set, indicate that valence orbital electron distributions are localized on the base, sugar, or phosphate groups. This provides evidence that valence electron ionization in 5′-dGMP and 5′-dGMP$^-$ is similar to ionization in the model compounds 1,9-dimethylguanine (**1**), 3-hydroxytetrahydrofuran (**2**), trimethyl phosphate (**3**), methyl phosphate, and H_3PO_4 and in the model anions $CH_3HPO_4^-$ and $H_2PO_4^-$. According to SCF calculations with the 6-31G basis set, the five upper occupied base orbitals and the two upper occupied sugar orbitals in 5′-dGMP and 5′-dGMP$^-$ are similar to orbitals appearing in **1** and **2**. Supplementary semiempirical HAM/3 configuration interaction (CI) calculations on 1,9-dimethylguanine (**1**) and 3-hydroxytetrahydrofuran (**2**) indicate that three of the first five ionization potentials (IPs) in **1**, and the first two IPs in **2** are associated with cation states which are qualitatively well described by Koopmans' theorem. The results of the HAM/3 CI calculations on **1** indicate that only two low-energy ionization events are influenced by CI effects. These cause hole-mixing associated with configurations arising from removal of electrons from the second and third highest occupied π orbitals. In a parallel investigation of neutral and anionic phosphate, results from second-order Møller–Plesset perturbation (MP2) calculations and from ab initio configuration interaction calculations using the CI singles (CIS) method indicate that CI effects strongly influence the second lowest energy ionization event arising from removal of electrons from H_3PO_4 and the five lowest energy ionization events arising from removal of electrons from the closed shell anion $H_2PO_4^-$. The similarity between valence orbital structure in the nucleotides and in the model compounds and anions makes it possible, employing experimental photoelectron data and computational results obtained at the post-SCF level for the model compounds and anions, to individually correct IPs calculated for 5′-dGMP and 5′-dGMP$^-$ at the 6-31G SCF level. For 5′-dGMP, this approach yields IPs of 8.4, 9.8, and 11.4–11.6 eV for the base, sugar, and phosphate groups, respectively. For 5′-dGMP$^-$ these groups have IPs of 5.9, 6.6, and 4.6 eV. A comparison of nucleotide IPs with alkylation patterns occurring in polynucleotide reactions of *N*-methyl-*N*-nitrosourea, dimethyl and diethyl sulfate, and methyl and ethyl methanesulfonate demonstrates that the percent alkylation which occurs at the different bases increases as the nucleotide base π ionization potentials decrease. These results are consistent with the observation that these alkylating agents are soft electrophiles which exhibit increasing S_N2 reactivity toward nucleotide bases with increasing π polarizabilities.

Introduction

Gas-phase UV photoelectron (PE) spectroscopy has provided an opportunity to experimentally test and scale theoretical descriptions of valence electrons in DNA and RNA components. This method has already been applied to naturally occurring DNA and RNA bases,[1] to synthetic bases which act as antimetabolites,[2] to nucleosides,[3] and to phosphate esters.[4] Before investigations employing photoelectron data, descriptions of the valence electronic properties of nucleotides relied solely on theoretical results.[5] More recently, gas-phase UV photoelectron data has been used in conjunction with results from ab initio SCF molecular orbital calculations with the 4-31G basis set[6] to describe the valence electronic structure in the neutral and anionic forms of the smallest nucleotide in DNA, 2′-deoxycytidine 5′-phosphate (5′-dCMP and 5′-dCMP$^-$).[7] In a preliminary report, this approach was also employed to describe changes in ionization potentials (IPs) of anionic 2′-deoxyguanosine 5′-phosphate (5′-dGMP$^-$) which occur when DNA undergoes the B to Z conformational transition.[8] One goal of the present investigation is to provide a complete description of the valence electronic structures of neutral and anionic forms of 2′-deoxyguanosine 5′-phosphate, the largest DNA

(1) (a) Padva, A.; LeBreton, P. R.; Dinerstein, R. J.; Ridyard, J. N. A. *Biochem. Biophys. Res. Commun.* **1974**, *60*, 1262. (b) Lauer, G.; Shafer, W.; Schweig, A. *Tetrahedron Lett.* **1975**, *45*, 3939. (c) Hush, N. S.; Cheung, A. S. *Chem. Phys. Lett.* **1975**, *34*, 11. (d) Padva, A.; O'Donnell, T. J.; LeBreton, P. R. *Chem. Phys. Lett.* **1976**, *41*, 278. (e) Dougherty, D.; Wittel, K.; Meeks, J.; McGlynn, S. P. *J. Am. Chem. Soc.* **1976**, *98*, 3815. (f) Peng, S.; Padva, A.; LeBreton, P. R. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2966. (g) Dougherty, D.; McGlynn, S. P. *J. Chem. Phys.* **1977**, *67*, 1289. (h) Yu, C.; Peng, S.; Akiyama, I.; Lin, J.; LeBreton, P. R. *J. Am. Chem. Soc.* **1978**, *100*, 2303. (i) Lin, J.; Yu, C.; Peng, S.; Akiyama, I.; Li, K.; Lee, L.-K.; LeBreton, P. R. *J. Phys. Chem.* **1980**, *84*, 1006. (j) Lin, J.; Yu, C.; Peng, S.; Akiyama, I.; Li, K.; Lee, L. K.; LeBreton, P. R. *J. Am. Chem. Soc.* **1980**, *102*, 4627. (k) LeBreton, P. R.; Yang, X.; Urano, S.; Fetzer, S.; Yu, M.; Leonard, N. J.; Kumar, S. *J. Am. Chem. Soc.* **1990**, *112*, 2138. (l) Urano, S.; Yang, X.; LeBreton, P. R. *J. Mol. Struct.* **1989**, *214*, 315.

(2) (a) Peng, S.; Lin, J.; Shahbaz, M.; LeBreton, P. R. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1978**, *5*, 301. (b) Padva, A.; Peng, S.; Lin, J.; Shahbaz, M.; LeBreton, P. R. *Biopolymers* **1978**, *17*, 1523.

(3) Yu, C.; O'Donnell, T. J.; LeBreton, P. R. *J. Phys. Chem.* **1981**, *85*, 3851.

(4) (a) Cowley, A. H.; Lattman, M.; Montag, R. A.; Verkade, J. G. *Inorg. Chim. Acta* **1977**, *25*, L151. (b) Chattopadhyay, S.; Findley, G. L.; McGlynn, S. P. *J. Electron Spectrosc. Relat. Phenom.* **1981**, *24*, 27. (c) LeBreton, P. R.; Fetzer, S.; Tasaki, K.; Yang, X.; Yu, M.; Slutskaya, Z.; Urano, S. *J. Biomol. Struct. Dynamics* **1988**, *6*, 199.

(5) Vercauteren, D. P.; Clementi, E. *Int. J. Quantum. Chem.: Quantum. Biol. Symp.* **1983**, *10*, 11.

(6) (a) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724; (b) Hehre, W. J.; Lathan, W. A. *J. Chem. Phys.* **1972**, *56*, 5255.

0002-7863/93/1515-6169$04.00/0

nucleotide, using an approach which is improved over that previously reported.[7,8]

The low volatility of nucleotides and the complexity of these molecules, which contain large numbers of π and lone-pair electrons, requires an indirect application of gas-phase photoelectron data to the investigation of valence electronic properties. For 5′-dCMP and 5′-dCMP$^-$, this approach depended on the finding that, in SCF descriptions of the neutral and anionic forms of the nucleotide, the upper occupied orbitals are localized and can be clearly assigned to either the base, sugar, or phosphate groups. In this indirect approach, results from SCF calculations on the nucleotide and on model compounds and anions, with electronic structures similar to those appearing in the nucleotide, have been compared to PE data on the model compounds and to results from post-SCF calculations on the model anions. In the investigation of 5′-dCMP and 5′-dCMP$^-$, post-SCF calculations were carried out using Møller–Plesset perturbation theory[9] to obtain reliable IPs of model anions for which PE data are not available.

In the earlier descriptions of valence electrons in nucleotides,[7,8] no investigation was carried out to determine whether configuration interaction (CI) influences ionization events associated with the base and sugar groups. In order to begin describing ionization events involving the phosphate group in 5′-dCMP$^-$, the IPs of the model anion, $H_2PO_4^-$, were examined.[7] The first IP was calculated by employing third-order Møller–Plesset perturbation (MP3) theory with a 6-31+G* basis set.[10,11] The first IPs of the related phosphorus and oxygen containing anions PO_2^-, PO_3^-, and CH_3O^- were also examined using the same approach. In these investigations, MP3 calculations were carried out on the ground-state closed-shell anions and on the ground-state neutral radicals formed by removal of an electron. In cases where comparison was possible, the first IPs obtained from the differences in the energies of the neutral radicals and the closed-shell anions differed by less than 0.5 eV from experimental IPs. However, this approach only provides reliable information about the lowest energy anion ionization potential.

In the present examination of 5′-dGMP and 5′-dGMP$^-$, photoelectron data for model compounds has been used to test computed IPs. The present investigation has employed photoelectron data for 1,9-dimethylguanine (**1**), 3-hydroxytetrahydrofuran (**2**), and trimethyl phosphate (**3**). Here, 1,9-dimethylguanine (**1**) was employed for spectroscopic comparison with results from ab initio SCF calculations on the base group because gas-phase PE measurements[1i,k] and IR measurements in N_2 and Ar matrices[12] demonstrate that guanine occurs in more than one tautomeric form and that the populations of different forms are strongly dependent on environment. The tautomeric structure of 1,9-dimethylguanine (**1**) is restricted to the amino keto form which occurs in DNA and RNA.[13]

The present investigation includes the following improvements. Ab initio SCF calculations were carried out using the 6-31G basis set,[14] which is larger than the 4-31G basis set previously employed.[7,8] The Koopmans' description of base ionization potentials in 1,9-dimethylguanine (**1**) has been modified to include recently described ionization events associated with hole-mixing.[15] In addition, semiempirical and ab initio post-SCF calculations have been employed to search for evidence of CI effects in oxygen atom lone-pair ionization events in 3-hydroxytetrahydrofuran (**2**), in the neutral phosphate model compound, H_3PO_4, and in the model anion, $H_2PO_4^-$. Finally, a more reliable description of nucleotide phosphate group ionization is provided by improved post-SCF methods, which combine results from second-order Møller–Plesset perturbation (MP2) theory and CI singles (CIS) calculations and which have been used to characterize the second through fifth lowest energy ionization events occurring in the model anion, $H_2PO_4^-$.

Photoelectron Spectra of 1,9-Dimethylguanine (1), 3-Hydroxytetrahydrofuran (2), and Trimethyl Phosphate (3). He(I) photoelectron spectra have previously been reported and interpreted for molecules **1**,[1k] **2**,[4] and **3**[8]. These are shown in Figure 1 along with spectrometer probe temperatures used in the experiments, assignments, and orbital diagrams obtained from an analysis of the spectra which employed Koopmans' theorem.[16] The photoelectron spectrum of 1,9-dimethylguanine (**1**) is shown at the top of Figure 1. The assignment in Figure 1 has been obtained[1k] from results from ab initio SCF calculations using 3-21G[17] and 4-31G basis sets. This assignment is in agreement with semiempirical HAM/3 calculations[18] and is consistent with one based on a comparison of the spectrum of **1** with spectra of guanine and hypoxanthine, with three methyl substituted hypoxanthines, and with five other methyl substituted guanines.[1i,k] Finally, the assignment of the spectrum of **1**, given in Figure 1, is generally consistent with a recent assignment based on results from HAM/3 CI calculations.[15] The results obtained from the CI calculations indicate that, of the seven lowest energy PE bands in the spectrum of **1**, five are well described within the framework of Koopmans' theorem. However, the results also indicate that CI effects influence the interpretation of two ionization events in the low-energy region of the PE spectrum of **1**. These effects are due to hole-mixing associated with the mixing of configurations arising from removal of electrons from the second and third highest occupied π orbitals. In Figure 1, these orbitals are labeled B_2 and B_4. According to the results from the HAM/3 CI calculations, the cation states which arise from hole-mixing lie in the unresolved energy region of the spectrum, between 9.5 and 10.0 eV. In Figure 1, this region contains the B_2 and B_4 orbitals.

The PE spectrum of 3-hydroxytetrahydrofuran (**2**) is shown in the middle of Figure 1 along with IPs and assignments[8] for the two lowest energy bands originating from the upper occupied oxygen atom lone-pair orbitals. The oxygen atom lone-pair interactions in **2** mimic sugar lone-pair interactions in deoxyribonucleotides.

The photoelectron spectrum of trimethyl phosphate (**3**) is shown at the bottom of Figure 1 along with vertical ionization potentials and assignments of bands arising from the upper occupied orbitals. In early investigations[4] of the spectrum of trimethyl phosphate (**3**), there were discrepancies concerning the assignments. In the first assignment,[4a] a C_3 symmetry was assumed for the molecule, and the spectrum was compared with that of trimethyl phosphite. In the second assignment,[4b] spectra of **3** and triethyl phosphate were measured and interpreted with the aid of semiempirical

(7) Tasaki, K.; Yang, X; Urano, S.; Fetzer, S.; LeBreton, P. R. *J. Am. Chem. Soc.* **1990**, *112*, 538.

(8) Yu, M.; Kim, H. S.; LeBreton, P. R. *Biochem. Biophys. Res. Commun.* **1992**, *184*, 16.

(9) Pople, J. A.; Seeger, R.; Krishnan, R. *Int. J. Quantum Chem.* **1977**, *11*, 149.

(10) The 6-31+G* basis set is the 6-31G* basis set with diffuse functions for non-hydrogen atoms.

(11) (a) Kuchitsu, K. *J. Chem. Soc. Jpn.* **1959**, *32*, 748. (b) Bartell, L. S.; Kohl, D. A. *Chem. Phys.* **1963**, *39*, 3097. (b) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294. (c) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *Chem. Phys.* **1984**, *80*, 3265. (d) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiley & Sons: New York, 1986; p 86.

(12) (a) Szczepaniak, K.; Szczesniak, M.; Person, W. B. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1986**, *13*, 287. (b) Szczepaniak, K.; Szczesniak, M. *J. Mol. Struct.* **1987**, *156*, 29. (c) Szczepaniak, K.; Szczesniak, M.; Person, W. B. *Chem. Phys. Lett.* **1988**, *153*, 39. (d) Szczepaniak, K.; Szczesniak, M.; Szajda, W.; Person, W. B.; Leszczynski, *Can. J. Chem.* **1991**, *69*, 1705.

(13) Sanger, W. *Principles of Nucleic Acid Structure*; Springer: Berlin, 1984; pp 110 ff, and references therein.

(14) (a) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257. (b) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213. (c) Gordon, M. S. *Chem. Phys. Lett.* **1980**, *76*, 163.

(15) Yu, M.; Jiang, Q.; LeBreton, P. R. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1992**, *19*, 27.

(16) Koopmans, T. *Physica* **1934**, *1*, 104.

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

(18) (a) Asbrink, L. C.; Fridh, C.; Lindholm, E. *Chem. Phys. Lett.* **1977**, *52*, 63, 69, 72. (b) Asbrink, L.; Fridh, C.; Lindholm, E. *J. Electron Spect.* **1979**, *16*, 65. (c) Chong, D. P. *Theor. Chim. Acta* **1979**, *51*, 55. (d) Urch, D. S.; Bergknut, T. K.; Young, L. M.; Kim, R. S.; Chong, D. P.; Andermann, G. *Spectr. Lett.* **1980**, *13*, 487.

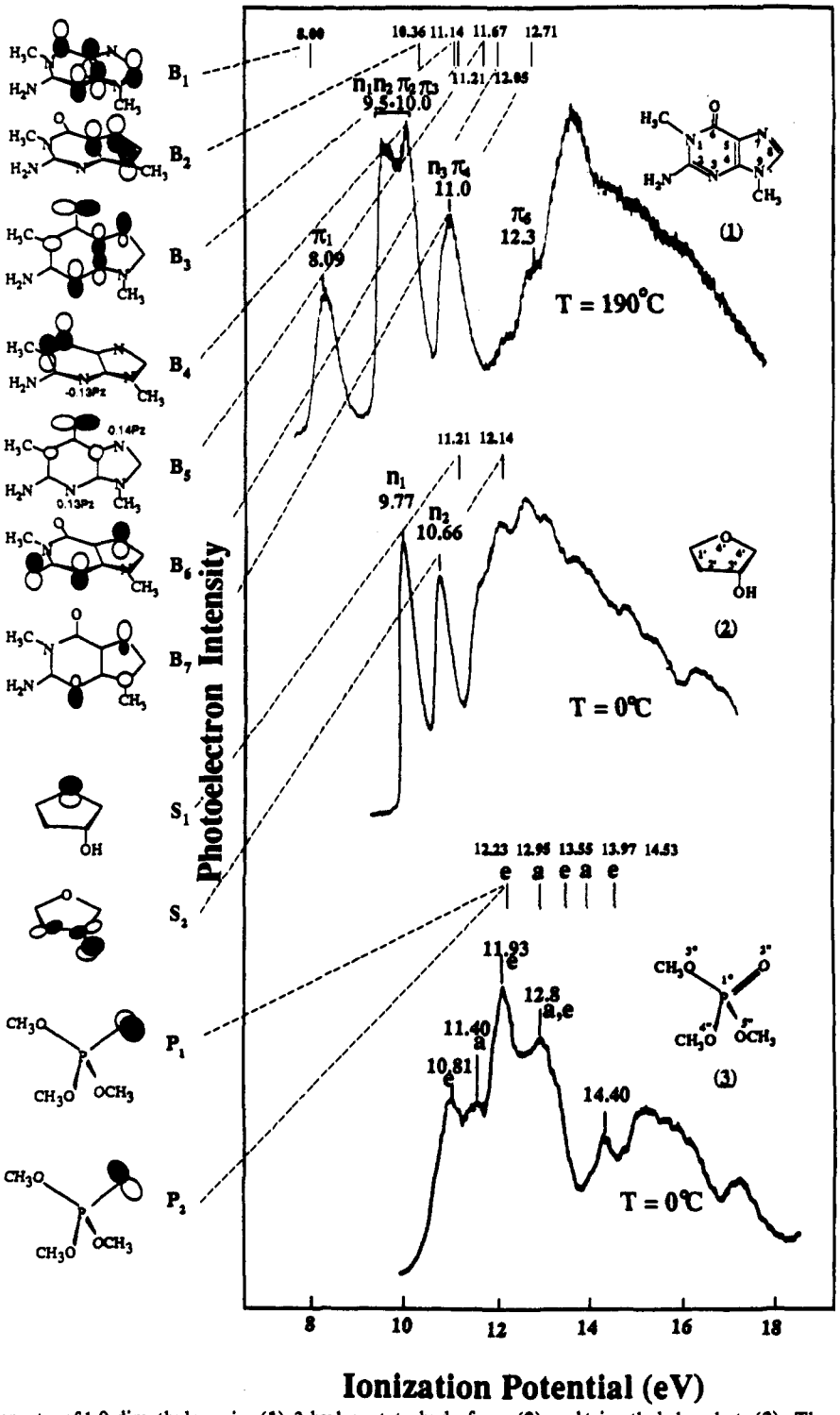

Figure 1. Photoelectron spectra of 1,9-dimethylguanine (**1**), 3-hydroxytetrahydrofuran (**2**), and trimethyl phosphate (**3**). Theoretical ionization potentials and orbital diagrams obtained from 6-31G SCF calculations. For **1** orbital diagrams illustrate major atomic orbital contributions to the seven highest occupied molecular orbitals (B_1–B_7). For π orbitals (B_1, B_2, B_4, and B_6) the viewing angle is different than for lone-pair orbitals (B_3, B_5, and B_7). For **2** and **3**, orbital diagrams for the two highest occupied oxygen atom lone-pair orbitals, S_1 and S_2, and P_1 and P_2 are also shown. Criteria used to construct the orbital diagrams are described in the text.

CNDO/2 calculations. In this second assignment, a C_{3v} geometry was assumed for both molecules. Later the PE spectrum of **3** was interpreted by carrying out ab initio SCF calculations with the STO-3G,[19] STO-3G*,[20] and 4-31G basis sets, assuming a C_3 symmetry.[4c] The results from the ab initio calculations, employing C_3 symmetry, support the assignments based on CNDO/2 calculations.[7]

(19) Hehre, W. J.; Ditchfield, R.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1970**, *52*, 2769.
(20) The STO-3G* basis set consists of a STO-3G basis set plus six phosphorus d-type orbitals. See ref 4c.

DNA and RNA Methylation and Ethylation. Much detailed experimental data is available describing patterns of DNA and RNA alkylation by mutagenic and carcinogenic methyl and ethyl methanesulfonates (MeMs and EtMs), dimethyl and diethyl sulfates (Me_2SO_4 and Et_2SO_4), and *N*-methyl- and *N*-ethyl-*N*-nitrosoureas (MeNU and EtNU).[21-25] In general, yields from DNA methylation reactions are 5–20 times larger than yields from ethylation reactions.[21,22] For EtNU, compared to MeNU, the smaller percentage of nucleotide ethylation which occurs is due to the facile hydrolysis of the more reactive ethylation reaction

intermediate.[21] For MeMs, EtMs, Me_2SO_4, Et_2SO_4, and MeNU, reaction with the bases is greatly favored over reaction with the sugar and phosphate groups.[21-23] Base ring nitrogen atoms and exocyclic oxygen atoms are attack sites, while reaction at the exocyclic amino groups is negligible.

Of biochemical interest is alkylation of the *O6* atom of guanine, which is a minor alkylation site. In reactions of DNA and RNA with MeMS, EtMS, Me_2SO_4, Et_2SO_4 and MeNU, the percent alkylation at *O6* of guanine is in all cases less than 10%,[21-23] Nevertheless, the slow depurination which accompanies guanine *O6* alkylation (compared to N7 or N3 alkylation[33,26]), the facile saturation of DNA repair mechanisms for guanine *O6* alkylation,[22,27] and the formation of a miscoding base caused by guanine *O6* alkylation[21,24,25,28,29] provide evidence that this lesion is important to mechanisms of mutagenesis and carcinogenesis involving methylating and ethylating reagents.

The DNA and RNA alkylation patterns exhibited by MeMs, EtMs, Me_2SO_4, Et_2SO_4, and MeNU are in some ways similar. In *in vitro* reactions of these five reagents with single-stranded DNA and RNA, the percentage of reaction which occurs at the different bases decreases in the order guanine (62–79%) > adenine (12–23%) > cytosine (2–10%) > thymine/uracil (0–3%).[22] In all cases, guanine (N7) is the most reactive site, accounting for between 62 and 77% of all nucleotide alkylation.

Previous quantum mechanical descriptions of nucleotide base model compounds, nucleotide bases, and nucleosides have been employed to determine how electronic structure influences reactions of DNA with methylating and ethylating agents.[30-37] In one set of studies, the methyl and ethyl cation affinities of nucleotide bases and base model compounds were examined.[31,33,34] However, experimental alkylation patterns do not correlate with cation affinities. Other investigations[30,36,37,38] have focussed on reactions of nucleotide bases and base model compounds with methane and ethanediazonium ions ($CH_3N_2^+$ and $C_2H_5N_2^+$), which are reactive intermediates formed in the base catalyzed activation of MeNU and EtNU.[21,28]

Two approaches have been employed to investigate nucleotide methylation and ethylation. The first focusses on properties of the reactants for clues concerning electronic factors that influence alkylation patterns. This approach has led to descriptions of the electrostatic potential surrounding DNA.[32,35] It has also led to an examination of the structure of the lowest unoccupied orbitals in methane and ethanediazonium ions.[38] The second approach, which is the more rigorous, has focussed on descriptions of transition states occurring in reactions of nucleotide base model compounds and nucleosides with methane and ethanediazonium ions.[30,36,37] Using this approach, a description of reactions has emerged which is consistent with a model proposed by Klopman.[39] According to this model, the activation energies of alkylation reactions are determined by a combination of short-range orbital interactions and long-range electrostatic interactions, and the relative balance of these interactions is strongly dependent upon the distance between reactants in the transition state.

In addition to obtaining a photoelectron based description of the valence electronic structures of 5′-dGMP (**4**) and 5′-dGMP$^-$ (**5**), a second goal of the present investigation is to obtain a better understanding of structural factors which influence DNA and RNA methylation and ethylation patterns.

Methods

At the SCF level, the ionization potentials of the model compounds **1–3**, and H_3PO_4, of the model anion $H_2PO_4^-$, and of 5′-dGMP (**4**) and 5′-dGMP$^-$ (**5**) have been obtained by applying Koopmans' theorem to results from calculations employing the 6-31G basis set. The same approach has been used to calculate IPs of $CH_3H_2PO_4$ and $CH_3HPO_4^-$. For anionic 5′-dGMP$^-$ (**5**), with 36 atoms and 180 electrons, the 6-31G SCF wave function contains 582 primitive and 237 contracted basis functions.

In order to obtain accurate ionization potentials for $H_2PO_4^-$, the ab initio SCF calculations were supplemented by post-SCF calculations employing second-order Møller–Plesset perturbation (MP2) theory on the ground states of $H_2PO_4^-$ and neutral $H_2PO_4^{\bullet}$, and the CI singles (CIS) method[40] on excited states of $H_2PO_4^{\bullet}$. The influence of CI on low-energy ionization events in H_3PO_4 was examined by carrying out MP2 calculations on the ground states of H_3PO_4 and $H_3PO_4^+$, and CIS calculations on the first excited state of $H_3PO_4^+$. The Gaussian 90 program[41] was used for all ab initio calculations.

Results from recent semiempirical HAM/3 CI calculations[42,43] indicate that hole mixing contributes to two of the seven lowest energy ionization events in 1,9-dimethylguanine (**1**).[15] In the present investigation, the effects of CI on the two lowest IPs of 3-hydroxytetrahydrofuran (**2**) were examined using the same approach. All computations were carried out on Cray 2, Cray Y/MP, and IBM 3090/600J/VF and 3090/120E computers.

Molecular orbital diagrams were drawn from molecular orbital coefficients obtained from results of the 6-31G SCF calculations. The inner Gaussian terms of the 6-31G expansions were used for the 2p and 3p orbitals of C, N, O, and P, while the outer Gaussian terms were used for the 2s and 3s orbitals.[14a,44] The sizes of the atomic orbitals in the molecular orbital diagrams are proportional to the coefficients. Only atomic orbitals with coefficients greater than 0.20 are shown.

The geometry of **1** employed in the calculations was obtained by optimizing heavy atom bond lengths and bond angles, and H

(21) Lawley, P. D. In *Chemical Carcinogenesis*; *ACS Monograph 182*; Searle, C. E., Ed.; American Chemical Society: Washington, 1984; pp 325–484.
(22) Singer, B.; Grunberger, D. *Molecular Biology of Mutagens and Carcinogens*; Plenum Press: New York, 1983; pp 65–68 and 221–244.
(23) Beranek, D. T.; Weis, C. C.; Swenson, D. H. *Carcinogenesis* **1980**, *1*, 595.
(24) Singer, B. *J. Natl. Cancer Inst.* **1979**, *62*, 1329.
(25) Saffhill, R.; Margison, G. P.; O'Connor, P. J. *Biochim. Biophys. Acta* **1985**, *823*, 111.
(26) (a) Lawley, P. D.; Orr, D. J.; Jarman, M. *Biochem. J.* **1975**, *145*, 73. (b) Lawley, P. D.; Shah, S. A. *Chem. Biol. Interactions* **1973**, *7*, 115. (c) Lawley, P. D.; Thatcher, C. J. *Biochem. J.* **1970**, *116*, 693. (d) Lawley, P. D. In *Topics in Chemical Carcinogenesis*; Nakahara, W., Takayama, S., Sugimura, T., Odashima, S., Eds.; University of Tokyo Press: Tokyo, 1972; pp 237–258. (e) Singer, B. *Nature* **1976**, *264*, 333. (f) Singer, B.; Fraenkel-Conrat, H. *Biochemistry* **1975**, *14*, 772. (g) Singer, B.; Bodell, W. J.; Cleaver, J. E.; Thomas, G. H.; Rajewsky, M. F.; Thon, W. *Nature* **1978**, *276*, 85. (h) Singer, B. *J. Toxicol. Environ. Health* **1977**, *2*, 1279. (i) Singer, B. In *Molecular and Cellular Mechanisms of Mutagenesis*; Lemontt, J. F., Generoso, W. M., Eds.; Plenum Press: New York, 1982; pp 1–42. (j) Pegg, A. E.; Singer, B. *Cancer Invest.* **1984**, *2*, 221.
(27) Sklar, R.; Brady, K.; Strauss, B. *Carcinogenesis* **1981**, *2*, 1293.
(28) Hathway, D. E.; Kolar, G. F. *Chem. Soc. Rev.* **1980**, *9*, 241.
(29) Loveless, A. *Nature* **1969**, *223*, 206.
(30) Ford, G. P.; Scribner, J. D. *J. Am. Chem. Soc.* **1983**, *105*, 349.
(31) Pullman, A.; Armbruster, A. M. *Theor. Chim. Acta* **1977**, *45*, 249.
(32) Pullman, A.; Pullman, B. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1980**, *7*, 245.
(33) Miertus, S.; Trebaticka, M. *J. Theor. Biol.* **1984**, *108*, 509.
(34) Miertus, S.; Trebaticka, M. *Coll. Czech. Chem. Commun.* **1983**, *48*, 3517.
(35) Pullman, A.; Pullman, B. In *Carcinogenesis: Fundamental Mechanisms and Environmental Effects*; Pullman, B., Ts'o, P. O. P., Gelboin, H., Eds.; Reidel: Dordrecht, Holland, 1980; pp 55–66.
(36) Mohammad, S. N.; Hopfinger, A. J. *J. Theor. Biol.* **1980**, *87*, 401.
(37) Ford, G. P.; Scribner, J. D. *Chem. Res. Toxicol.* **1990**, *3*, 219.
(38) Sapse, A. M.; Allen, E. B.; Lown, J. W. *J. Am. Chem. Soc.* **1988**, *110*, 5671.
(39) Klopman, G. In *Chemical Reactivity and Reaction Paths*; Klopman, G., Ed.; Wiley-Interscience: New York, 1974; pp 55–165.
(40) Foresman, J.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *J. Phys. Chem.* **1992**, *96*, 135.
(41) Frisch, M. J.; Head-Gordon, M.; Trucks, G. W.; Foresman, J. B.; Schlegel, H. B.; Raghavachari, K.; Robb, M.; Binkley, J. S.; Gonzalez, C.; Defrees, D. J.; Fox, D. J.; Whiteside, R. A.; Seeger, R.; Melius, C. F.; Baker, J.; Martin, R. L.; Kahn, L. R.; Stewart, J. J. P.; Topiol, S.; Pople, J. A. Gaussian, Inc.: Pittsburgh PA, 1990.
(42) (a) Chong, D. P. *J. Mol. Sci.* **1982**, 2, 55. (b) Lindholm, E.; Asbrink, L. *Molecular Orbitals and Their Energies, Studied by the Semiempirical HAM Method*; Springer-Verlag: New York, 1985; pp 235ff.
(43) Koenig, T.; Southworth, S. J. *J. Am. Chem. Soc.* **1977**, *99*, 2807.
(44) Clark, T. *A Handbook of Computational Chemistry*; John Wiley: New York, 1985; p 267.

atom bond angles at the 6-31G SCF level. The C–H bond lengths were 1.10 and 1.08 Å and the N–H bond lengths were 1.01 Å.[46] The geometry of **2** was obtained by combining electron diffraction data for tetrahydrofuran[45] with standard OH bond lengths and bond angles. The C–O bond length in the hydroxyl group was 1.43 Å;[46] the CCO bond angle was 111.4°.

For trimethyl phosphate (**3**), the P=O and P—O bond lengths were taken from crystallographic data on tri-*p*-nitrophenyl phosphate.[47] Other bond lengths and bond angles were based on previously reported classical potential calculations on **3**,[48] which indicated that the most stable geometry has C_3 symmetry. In the present investigation, the dihedral angles (φ) describing rotation about the P—O bonds in the O=P—O—C sequences[49] was optimized at the 6-31G level. The optimized value of φ (48.1°) obtained from the 6-31G calculation is similar to an earlier value (53.1°) obtained from STO-3G* calculations and used in recent assignments of the PE spectrum of **3**.[4c,7]

The geometries employed in the SCF and MP2 calculations on H_3PO_4 and $H_2PO_4^-$ were obtained by optimization at the 6-31+G* level.[10,11] For MP2 and CIS calculations on $H_3PO_4^+$ and $H_2PO_4^•$, geometries were taken to be the same as those of H_3PO_4 and $H_2PO_4^-$, respectively.[50]

Heavy atom bond angles and bond lengths in 5'-dGMP$^-$ (**5**) and in the base and sugar groups of 5'-dGMP (**4**) were taken from crystallographic data for deoxycytidylyl-(3'-5')-deoxyguanosine.[51] C–H and N–H bond lengths in the base and sugar groups were the same as those used in **1**. Other bond lengths and bond angles involving the hydrogen atoms in the base and sugar groups were O–H = 0.95 Å, HCH = 109.5°, HOC = 105.0°, HNH = 120°, and HNC = 120°.[46] The geometry of the phosphate group of 5'-dGMP (**4**) was based on that for trimethyl phosphate. The O–H bond lengths and HOP bond angles were 0.95 Å and 107.6°. Torsional angles used to define the conformations of **4** and **5** were taken from optimized crystallographic data for 5'-dGMP$^-$ in B-DNA.[52] The bond lengths, bond angles, and conformations employed in the 6-31G SCF calculations on $CH_3H_2PO_4$ and $CH_3HPO_4^-$ were the same as those used for the phosphate groups in calculations on 5'-dGMP (**4**) and 5'-dGMP$^-$ (**5**), respectively.

Results

1,9-Dimethylguanine (1), 3-Hydroxytetrahydrofuran (2), and Trimethyl Phosphate (3). Figure 1 shows a comparison of experimental vertical IPs and theoretical IPs, obtained from 6-31G SCF calculations on molecules **1–3**. The figure also contains orbital diagrams for the highest occupied orbitals obtained from the 6-31G results. For **1**, orbital diagrams were drawn for the seven highest occupied orbitals. For **2** and **3**, diagrams are drawn for the two highest occupied orbitals. Table I lists atomic charges, obtained from a Mulliken population analysis[53] of the 6-31G wave functions for molecules **1–3**.

Table I. Atomic Charges for 1,9-Dimethylguanine (**1**), 3-Hydroxytetrahydrofuran (**2**), Trimethyl Phosphate (**3**), $CH_3H_2PO_4$, $CH_3HPO_4^-$ [a] and $H_2PO_4^-$ [a]

1[b]		2[b]		3[b]		$CH_3H_2PO_4$		$CH_3HPO_4^-$		$H_2PO_4^-$	
N_1	−1.04	C_1'	0.06	P_1''	2.03	P_1''	1.87	P_1''	1.76	P_1''	1.72
C_2	1.05	C_2'	−0.41	O_2''	−0.87	O_2''	−0.81	O_2''	−0.88	O_2''	−0.90
N_3	−0.64	C_3'	0.08	O_3''	−0.85	O_3''	−0.82	O_3''	−0.93	O_3''	−0.90
C_4	0.71	C_4'	0.05	O_4''	−0.85	O_4''	−0.77	O_4''	−0.86	O_4''	−0.87
C_5	−0.08	O_3'	−0.72	O_5''	−0.85	O_5''	−0.85	O_5''	−0.84	O_5''	−0.87
C_6	0.86	O_4'	−0.69	C_3''	−0.13	C_5''	−0.11	C_5''	−0.09		
N_7	−0.48			C_4''	−0.13						
C_8	0.35			C_5''	−0.13						
N_9	−0.97										
C_1	−0.21										
N_2	−0.99										
O_6	−0.59										
C_9	−0.18										

[a] Obtained from 6-31G SCF calculations. [b] The numbering of atoms is shown in Figure 1.

Table II. Atomic Charges in 5'-dGMP and 5'-dGMP$^-$ [a]

guanine		2'-deoxyribose		phosphate	
		5'-dGMP (4)			
N_1	−1.04	C_1'	0.46	P_1''	1.90
C_2	1.03	C_2'	−0.40	O_2''	−0.82
N_3	−0.60	C_3'	0.10	O_3''	−0.78
C_4	0.71	C_4'	0.12	O_4''	−0.82
C_5	−0.09	O_3'	−0.72	O_5''	−0.91
C_6	0.85	O_4'	−0.67	C_5''	0.08
N_7	−0.48				
C_8	0.37				
N_9	−1.01				
N_2	−0.98				
O_6	−0.59				
	−0.38[b]		0.40[b]		−0.02[b]
		5'-dGMP$^-$ (5)			
N_1	−1.04	C_1'	0.45	P_1''	1.81
C_2	1.02	C_2'	−0.40	O_2''	−0.91
N_3	−0.60	C_3'	0.09	O_3''	−0.92
C_4	0.69	C_4'	0.11	O_4''	−0.90
C_5	−0.09	O_3'	−0.74	O_5''	−0.92
C_6	0.84	O_4'	−0.68	C_5''	0.08
N_7	−0.50				
C_8	0.35				
N_9	−1.00				
N_2	−0.99				
O_6	−0.61				
	−0.39[b]		−0.31[b]		−0.92[b]

[a] Obtained from 6-31G SCF calculations. The numbering of atoms is shown in Figures 2 and 3. [b] Total charge including heavy atoms and hydrogen atoms.

The results at the top of Figure 1 indicate that the first IP of **1**, which arises from a π orbital, is accurately calculated to within 0.1 eV and that the theoretical IPs of the second, third, and fourth highest occupied π orbitals are larger than the experimental values by 0.4–0.9, 1.2–1.7, and 1.1 eV, respectively. For the three highest occupied lone-pair orbitals in **1**, the theoretical IPs are 1.1–1.6, 1.7–2.2, and 1.7 eV larger than the experimental IPs. The difference between the experimental and theoretical IPs obtained from 6-31G calculations on **1** are similar to differences previously obtained between experimental IPs and theoretical IPs obtained from 4-31G calculations on 1-methylcytosine and methyl substituted guanines.[1k,7] Both 4-31G and 6-31G basis sets predict ionization potentials for the highest occupied π orbitals which agree with experiment to within 0.5 eV but yield IPs for the uppermost occupied lone-pair orbitals which are more than 1.0 eV larger than the experimental values.[1k,l,7]

The results in the middle of Figure 1 indicate that the two highest occupied molecular orbitals in 3-hydroxytetrahydrofuran (**2**) arise primarily from oxygen lone-pair atomic orbitals. The

(45) Geise, H. J.; Adams, W. J.; Bartell, L. S. *Tetrahedron* **1969**, *25*, 3045.
(46) Bowen, H. J. M.; Donohue, J.; Jenkin, D. G.; Kennard, O.; Wheatley, P. J.; Whiffen, D. H. In *Tables of Interatomic Distances and Configuration in Molecules and Ions*; Sutton, L. E., Jenkin, D. G., Mitchell, A. D., Cross, L. C., Eds.; The Chemical Society: London, 1958; pp S7, S8, S15–S17.
(47) Ul-Haque, M.; Caughlan, C. N. *Chem. Commun.* **1967**, 202.
(48) Khetrapal, C. L.; Govil, G.; Yeh, H. J. C. *J. Mol. Struct.* **1984**, *116*, 303.
(49) The sign of (φ) is positive if, in the O=P—O—C sequence, the movement of the O—C bond toward the P=O bond involves a right-handed screw motion.
(50) In previous MP3/6-31+G* test calculations in which the first IPs of CH_3O^-, PO_2^-, and PO_3^- were evaluated by calculating ground-state energies for the closed shell anions and for the open shell radicals two methods were examined. In the first method, the geometry of the closed shell anion was optimized at the 6-31G* level and the same geometry was used for the radical. In the second method, optimized geometries for both the closed shell anions and the open shell neutral radicals were used. When compared with experimental IPs, results obtained from the second method were not significantly more accurate than results obtained from the first method. See Table IV of ref 7.
(51) Cruse, W. B. T.; Egert, E.; Kennard, O.; Sala, G. B.; Salisbury, S. A.; Visamitra, M. A. *Biochemistry* **1983**, *22*, 1833.
(52) Arnott, S.; Hukins, D. W. L. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504.

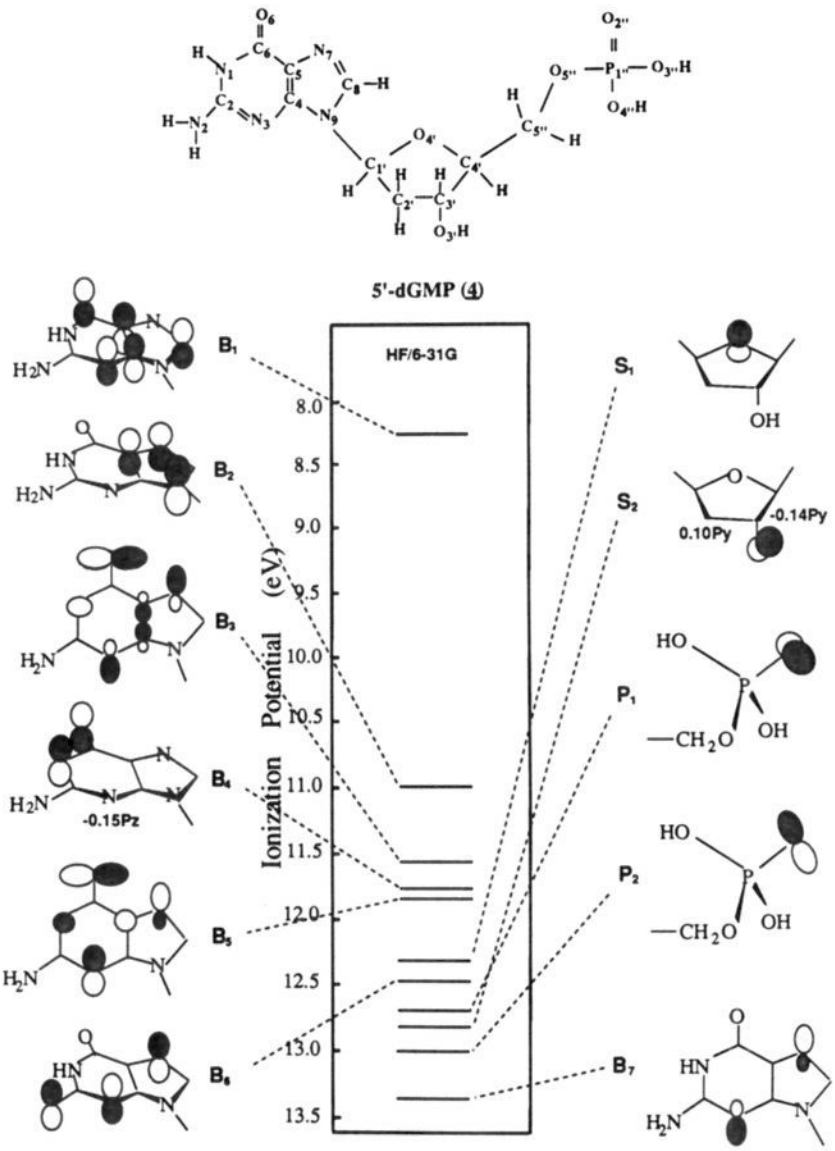

Figure 2. Ionization potentials and orbital diagrams of 5′-dGMP (**4**) obtained from 6-31G SCF calculations. Orbitals localized primarily on the base, sugar, and phosphate groups are denoted B, S, and P, respectively.

highest occupied molecular orbital (HOMO), with an IP of 9.77 eV, is associated with the cyclic O atom, while the second orbital, with an IP of 10.66 eV, is associated with the hydroxyl O atom. The assignments of these two bands are in agreement with the previously reported spectra of tetrahydrofuran and cyclopentanol[7,54,55] which have first IPs of 9.69 and 10.21 eV, respectively. A comparison of the theoretical and experimental results again indicates that the 6-31G SCF calculations yield ionization potentials for lone-pair orbitals which are too large. The theoretical IPs for the first and second highest occupied orbitals in **2** (S_1 and S_2) are 1.44 and 1.48 eV larger than the experimental IPs.

The results at the bottom of Figure 1 indicate that in trimethyl phosphate (**3**) the two highest occupied molecular orbitals, labeled P_1 and P_2, arise primarily from O atom lone-pair orbitals. The assignments in Figure 1 are based on results from the 6-31G SCF calculations and agree with previously reported assignments based on results from 4-31G calculations.[4c,7] As for 1,9-dimethylguanine (**1**) and 3-hydroxytetrahydrofuran (**2**), the 6-31G SCF calculations on trimethyl phosphate (**3**) predict IPs for the O atom lone-pair orbitals, which are too large. The theoretical IPs of the P_1 and P_2 orbitals in **3** are 1.42 eV larger than the experimental values.

5′-dGMP (4). Figure 2 shows IPs and orbital diagrams obtained from 6-31G SCF calculations describing the 11 highest occupied orbitals in 5′-dGMP (**4**). Table II lists atomic charges in **4** and in 5′-dGMP$^-$ (**5**) obtained from Mulliken population analyses of 6-31G SCF wave functions. Results from the 6-31G calculations indicate that, for the eleven upper occupied orbitals in **4**, the mixing of molecular orbital coefficients between atomic orbitals on the base, sugar, and phosphate groups is small.

According to the 6-31G results, the HOMO in 5′-dGMP (**4**), designated B_1 in Figure 2, has major contributions only from atoms on the base. In B_1, only atomic orbitals on the base have molecular orbital coefficients greater than 0.20. The next four highest occupied orbitals (B_2 to B_5), the seventh orbital (B_6), and the eleventh highest occupied orbital (B_7) are also on the base. The sixth (S_1) and ninth (S_2) orbitals are on the sugar, and the eighth (P_1) and tenth (P_2) orbitals are on the phosphate group. According to results from the 6-31G calculations, the localized orbital description of 5′-dGMP (**4**) breaks down for the 12th highest occupied orbital which resides on both the base and sugar groups.

The localized valence electronic structure in 5′-dGMP (**4**), indicated by the orbital diagrams in Figure 2, is also reflected in the Mulliken population analyses of the 6-31G SCF wave functions for 1,9-dimethylguanine (**1**), 3-hydroxytetrahydrofuran (**2**), trimethyl phosphate (**3**), and **4**. According to the Mulliken population analyses, the atomic charges associated with the C, N, O, and P atoms in **1**, **2**, and **3** are similar to those occurring

on the corresponding atoms of the base, sugar, and the phosphate groups in 5'-dGMP (4).

In addition to predicting that the valence orbitals in 5'-dGMP (4) are localized, the 6-31G SCF orbital diagrams for 4 also indicate that the electron distributions in the upper occupied orbitals are similar to those occurring in corresponding orbitals of the model compounds. For example, a comparison of the diagrams in Figures 1 and 2 demonstrates that the seven highest occupied base orbitals in the nucleotide (B_1–B_7) correspond to orbitals occurring in 1. While small differences occur, they are due largely to the cutoff criteria used. In Figures 1 and 2, molecular orbital coefficients less than 0.20 are given in cases where there are discrepancies between diagrams for specific molecular orbitals which appear in the nucleotide and in the model compounds. Not only are the B_1–B_7 orbital diagrams in 5'-dGMP (4) similar to those in 1,9-dimethylguanine (1) but also the base IPs for 4, obtained from 6-31G SCF calculations, are similar to those for 1. In 4, the calculated IPs of the B_1–B_7 orbitals are 8.24, 10.97, 11.53, 11.75, 11.82, 12.45, and 13.36 eV, respectively. In 1 the IPs of the corresponding orbitals are 8.00, 10.36, 11.14, 11.21, 11.67, and 12.05 eV.

There is also a correspondence between the diagrams for the S_1 and S_2 orbitals in 3-hydroxytetrahydrofuran (2), shown in Figure 1, and the S_1 and S_2 orbitals in 5'-dGMP (4), shown in Figure 2. However, the average calculated difference between the IPs of the S_1 and S_2 orbitals in 5'-dGMP (4) and the IPs of the corresponding orbitals in 3-hydroxytetrahydrofuran (2) is greater than the average calculated difference between IPs of corresponding orbitals in 4 and in 1,9-dimethylguanine (1). In 5'-dGMP (4), the calculated IPs of the S_1 and S_2 orbitals are 12.29 and 12.74 eV. In 2, the calculated S_1 and S_2 ionization potentials are 11.21 and 12.14 eV.

Like the upper occupied base and sugar orbitals in 5'-dGMP (4), the upper occupied phosphate orbitals, P_1 and P_2, have orbital diagrams which are similar to those in the model compound, in this case trimethyl phosphate (3). Results from the 6-31G SCF calculations agree with the earlier 4-31G results[7] in predicting that the P_1 and P_2 orbital diagrams in 3 are also similar to those in methyl phosphate. In 3, the IPs of the P_1 and P_2 orbitals, calculated at the 6-31G SCF level, are 12.23 eV. In methyl phosphate, they are 12.34 and 12.55 eV. In 4, the calculated P_1 and P_2 ionization potentials are 12.80 and 12.97 eV.

5'-dGMP$^-$ (5). Figure 3 shows ionization potentials and orbital diagrams for the 12 highest occupied orbitals in 5'-dGMP$^-$ (5) obtained from the 6-31G SCF calculations. The results of Mulliken population analysis, given in Table II, indicate that in 5, the negative charge of the anion is located on the phosphate group (–0.92 eu). A comparison of atomic charges in 5'-dGMP (4) and in 5 demonstrates that in both the neutral molecule and the anion, the charges on the sugar and base groups are similar. In 5, the charges on the sugar and base groups differ from those in 4 by only –0.09 and –0.01 eu, respectively. The similar charges on corresponding atoms of the base and sugar groups suggest that there are similarities between base and sugar orbitals in 5'-dGMP (4) and 5'-dGMP$^-$ (5).

This similarity between the base and sugar groups in 4 and 5 is reflected in the orbital diagrams in Figures 2 and 3. The 6-31G SCF calculations predict that for 5'-dGMP$^-$ (5) all 12 of the highest occupied orbitals are localized on either the base, sugar, or phosphate groups. The 13th highest occupied orbital is distribution between the sugar and phosphate groups. According to 6-31G SCF calculations, the first (B_1), sixth (B_2), seventh (B_3), tenth (B_5), and eleventh (B_4) highest occupied orbitals are on the base; the second (P_1), third (P_2), fourth (P_3), fifth (P_4), and eighth (P_5) highest occupied orbitals are on the phosphate group, and the ninth (S_1) and twelfth (S_2) orbitals are on the sugar group. The results in Figures 2 and 3 indicate that the upper occupied base (B_1–B_5) and sugar (S_1 and S_2) orbitals in 5'-dGMP$^-$ (5) have electron distributions like those of corresponding orbitals in 5'-dGMP (4) and in 1,9-dimethylguanine (1) and 3-hydroxytetrahydrofuran (2). For the B_3, B_4, and B_5 orbitals, small apparent differences occurring in the diagrams for 4 and 5 in Figures 2 and 3 are exaggerated by the cutoff criterion. This is indicated by the molecular orbital coefficients given in the figures.

The 6-31G calculations also indicate that there is correspondence between valence electronic properties of the phosphate group in 5 and of the model anions, $CH_3HPO_4^-$ and $H_2PO_4^-$. Tables I and II provide evidence of this correspondence. A comparison of the results of the Mulliken population analyses for 5'-dGMP$^-$ (5), $CH_3HPO_4^-$, and $H_2PO_4^-$ indicates that charges occurring on corresponding atoms of the three anions are similar. The similar charge distributions in $H_2PO_4^-$, $CH_3HPO_4^-$, and 5 are also reflected in the orbital diagrams for the three anions. This is demonstrated by comparing the phosphate group orbitals (P_1–P_5) shown for 5'-dGMP$^-$ (5) in Figure 3, with the P_1–P_5 orbitals in $H_2PO_4^-$ and $CH_3HPO_4^-$, shown in Figure 4. It is also interesting to note that not only are the diagrams for the P_1–P_5 orbitals in $H_2PO_4^-$ similar to those in $CH_3HPO_4^-$ but also the ionization potentials predicted by 6-31G SCF calculations for corresponding orbitals in $H_2PO_4^-$ and $CH_3HPO_4^-$ are almost equal. The calculated difference between IPs of corresponding orbitals is less than 0.45 eV. On the other hand, differences between calculated IPs for the P_1–P_5 orbitals in $CH_3HPO_4^-$ and 5'-dGMP$^-$ (5) are larger. In 5, each of the phosphate orbitals is predicted to have an IP which is 0.60–0.82 eV larger than that of the corresponding orbital in $CH_3HPO_4^-$.

Photoelectron Corrected Valence Ionization Potentials of 5'-dGMP (4). Base Orbitals. According to the results in Figures 1 and 2 and Tables I and II, electron distributions in the base orbitals (B_1–B_5) of 1,9-dimethylguanine (1) are similar to those of the upper occupied base orbitals in 5'-dGMP (4). This similarity provides evidence that the accuracy of IPs of the upper occupied base orbitals in 4, obtained from 6-31G SCF calculations, is approximately the same as the accuracy of the calculated IPs of these orbitals in 1.

For 5'-dGMP (4) corrected IPs for the base orbitals have been obtained from eqs 1 and 2.

$$IP_{corr,4}(B_i) = IP_{calc,4}(B_i) - \Delta IP_i \quad (1)$$

$$\Delta IP_i = IP_{calc,1}(B_i) - IP_{exp,1}(B_i) \quad (2)$$

In eqs 1 and 2, $IP_{calc,4}(B_i)$ and $IP_{calc,1}(B_i)$ are the ionization potentials of the B_i orbitals (i = 1–5) in 4 and 1, as predicted by 6-31G SCF calculations. $IP_{exp,1}(B_i)$ is the experimental ionization potential of the B_i orbital in 1,9-dimethylguanine (1). The corrected ionization potentials, $IP_{corr,4}(B_i)$, of the base orbitals in 5'-dGMP (4) are given in Figure 5. The hatched regions in the figure denote uncertainty in the corrected IPs. This arises because the PE spectrum of 1,9-dimethylguanine (1) is not well resolved in the energy range between 9.0 and 10.5 eV. In this range, the spectrum of 1 contains four overlapping bands.

Sugar Orbitals. While the orbital diagrams and the results of population analyses indicate that the S_1 and S_2 orbitals in 5'-dGMP (4) are similar to the S_1 and S_2 orbitals in 3-hydroxytetrahydrofuran (2), results from the 6-31G SCF calculations predict that the S_1 and S_2 ionization potentials in 4 are 1.08 and 0.60 eV larger than in 2. However, the large calculated differences between IPs of corresponding S_1 and S_2 orbitals in 4 and 2 are an artifact of the 6-31G SCF calculations. This aberration also occurs in results from 4-31G SCF calculations, where it is greater.[7] The differences in calculated IPs are related to differences in the energetic ordering of the S_1 and S_2 orbitals relative to other occupied orbitals in 5'-dGMP (4) and in 3-hydroxytetrahydrofuran (2). In 4, the S_1 and S_2 orbitals are the sixth and ninth highest occupied orbitals. In 2, they are the first and second highest occupied orbitals. Evidence that large differences in calculated IPs of the corresponding S_1 and S_2 orbitals in 2 and 4 are due to computational error is provided by results from a supplementary 6-31G SCF calculation on 2'-deoxyribophosphate,

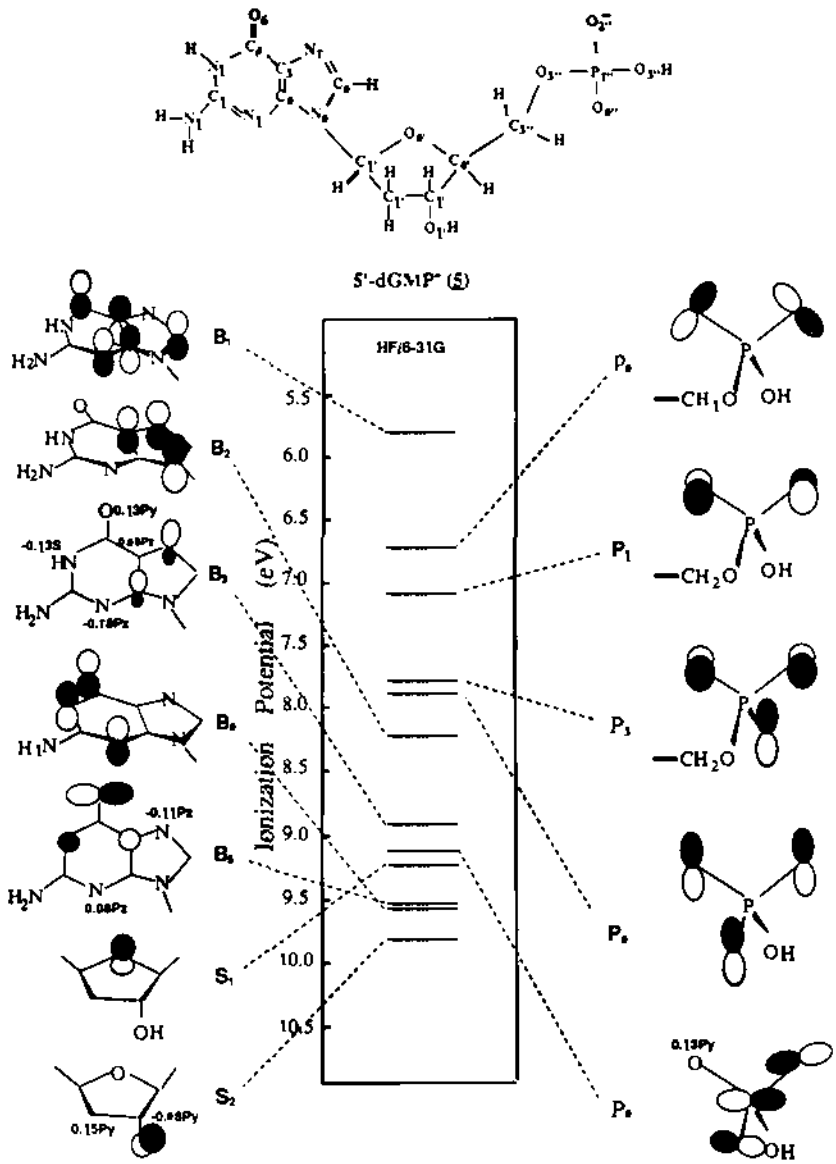

Figure 3. Ionization potentials and orbital diagrams of 5′-dGMP²⁻ (5) obtained from 6-31G SCF calculations.

which has a structure and geometry identical to **4**, but which has an H atom substituted for the base group. In this molecule, as in **2**, the S_1 orbital is the HOMO, and the calculated S_1 and S_1 ionization potentials (11.51 and 12.24 eV) differ from those in **2** by only 0.30 and 0.10 eV, respectively.

The observation that the S_1 and S_1 orbitals in 5′-dGMP (**4**) and 3-hydroxytetrahydrofuran (**2**) have similar electron distributions and that the orbitals are associated with atoms having similar total charges provides evidence that the true IPs for the S_1 and S_1 orbitals in **4** are nearly equal to the experimental IPs of the S_1 and S_1 orbitals in **2**. Figure 5 gives corrected IPs of the S_1 and S_1 orbitals in **4**, which have been obtained by employing this approximation.

Phosphate Orbitals. A comparison of the population analyses given in Tables I and II and of orbital diagrams for the P_1 and P_1 orbitals in 5′-dGMP (**4**) and in $CH_3H_1PO_4$ provide evidence that IPs of the P_1 and P_2 orbitals in **4** and in $CH_3H_1PO_4$ are nearly equal. Corrected IPs for the P_1 and P_1 orbitals in 5′-dGMP (**4**) have been estimated to be equal to the ionization potentials ($IP_{corr,CH_3H_1PO_4}(P_i)$) of the P_1 and P_1 orbitals in methyl phosphate which have been calculated using eqs 3 and 4.[56]

$$IP_{corr,CH_3H_1PO_4}(P_i) = IP_{calc,CH_3H_1PO_4}(P_i) - \Delta IP_i \qquad (3)$$

$$\Delta IP_i = IP_{calc,3}(P_1) - IP_{exp,3}(P_i) \qquad (4)$$

In eqs 3 and 4, $IP_{calc,CH_3H_1PO_4}(P_i)$ and $IP_{calc,3}(P_i)$ are the ionization potentials of the P_i orbitals (i = 1, 2) in methyl phosphate and trimethyl phosphate (3) as predicted by 6-31G SCF calculations. $IP_{exp,3}(P_i)$ is the experimental ionization potential for the P_i orbital in trimethyl phosphate (3). Values of $IP_{corr,CH_3H_1PO_4}(P_i)$ for the P_1 and P_2 orbitals in methyl phosphate and 5′-dGMP (**4**) are in the range of 11.4–11.6 eV. The reliability of eqs 3 and 4 depends on the observation that the difference between valence electron IPs, which are due to the addition of heteroatoms or of methyl groups, are more accurately predicted by both ab initio and semiempirical SCF calculations than are the absolute values of IPs.[1d,1k,57,58] The corrected IPs of the P_1 and P_1 orbitals in 5′-dGMP (**4**) are given in Figure 5.

Corrected Valence Ionization Potentials of 5′-dGMP²⁻ (5). Base and Sugar Orbitals. The results from the 6-31G SCF calculations on 5′-dGMP (**4**) and 5′-dGMP²⁻ (**5**), given in Figures 2 and 3, indicate that the upper occupied base (B_1–B_5) and sugar (S_1 and S_1) orbitals in 5′-dGMP²⁻ (**5**) have smaller IPs than those of corresponding orbitals in 5′-dGMP (**4**). For the B_1–B_5 orbitals, the calculated IPs in **5** are 2.18–2.72 eV smaller than the IPs in **4**. For the S_1 and S_1 orbitals, the calculated IPs in **5** are 3.05 and 2.93 eV smaller. Nevertheless, the 6-31G SCF results indicate that there are similarities between base and sugar orbitals in 5′-dGMP (**4**) and 5′-dGMP²⁻ (**5**). This is demonstrated by the correspondence in **4** and **5** between the results of Mulliken population analyses for the base and sugar groups, by the similar

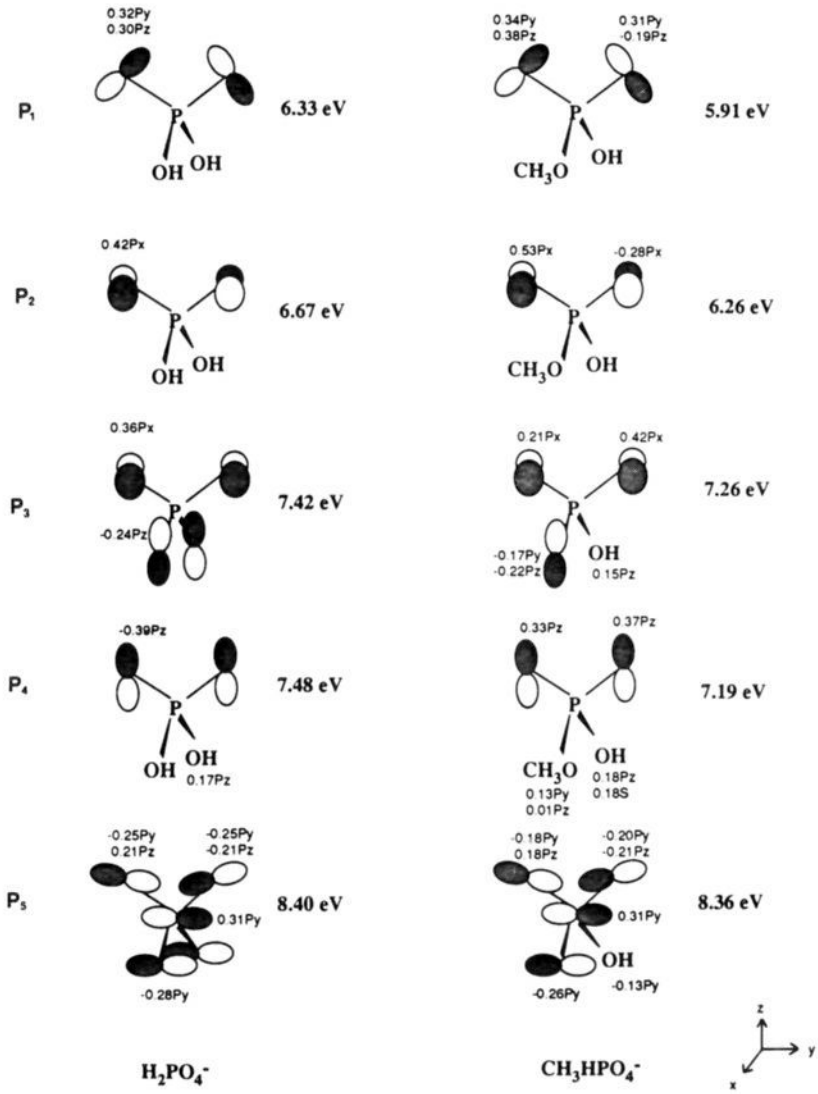

Figure 4. Orbital diagrams and ionization potentials of the five highest occupied orbitals in $H_2PO_4^-$ and $CH_3HPO_4^-$ obtained from 6-31G SCF calculations.

relative energetic ordering of the base and sugar orbitals, and by similarities between base and sugar orbital diagrams.

The correspondence between the 6-31G SCF descriptions of valence electronic structures of the base and sugar groups in neutral and anionic forms of the nucleotide provide evidence that the accuracy of IPs for the base and sugar orbitals obtained from 6-31G SCF calculations on 5′-dGMP (**4**) is approximately the same as that obtained from calculations on 5′-dGMP$^-$ (**5**). This observation has led to eqs 5 and 6 which have been employed to obtain corrected base and sugar IPs in 5′-dGMP$^-$ (**5**).

$$IP_{corr,5}(i) = IP_{calc,5}(i) - \Delta IP_i \quad (5)$$

$$\Delta IP_i = IP_{calc,4}(i) - IP_{corr,4}(i) \quad (6)$$

In eq 5, $IP_{calc,5}(i)$ is the ionization potential of the ith base or sugar orbital in **5** obtained from 6-31G SCF calculations. In eq 6, $IP_{calc,4}(i)$ is the ionization potential of the ith orbital obtained from 6-31G SCF calculations on 5′-dGMP (**4**), and $IP_{corr,4}(i)$ is the corrected ionization potential of the ith base or sugar orbital in **4**. Here, $IP_{corr,4}(i)$ was obtained from eq 1 for the base orbitals, and from the PE spectrum of 3-hydroxytetrahydrofuran (**2**) for the sugar orbitals. Figure 6 shows corrected IPs of the base and sugar orbitals in **5**, obtained from eqs 5 and 6.

The validity of eqs 5 and 6 relies on the observation that the smaller IPs of the corresponding upper occupied base and sugar orbitals in 5′-dGMP$^-$ (**5**), compared to 5′-dGMP (**4**), are due primarily to through-space interactions in **5** associated with the negatively charged phosphate group and that the perturbation of base and sugar orbitals arising from through-space interactions is accurately represented by results from 6-31G SCF calculations. Support of this observation is provided, in part, by results of a supplementary supermolecule calculation carried out at the 6-31G SCF level on 9-methylguanine in the presence of $H_2PO_4^-$. In this calculation, the geometry of 9-methylguanine was based on that of **1**, and the geometry of $H_2PO_4^-$ was the same as that described above. The relative position of the molecule and anion was chosen so that the orientation of $H_2PO_4^-$, relative to 9-methylguanine, was the same as that of the base and the phosphate groups in **5**. The importance of through-space effects on the base IPs of **5** is demonstrated by the finding that the calculated IP of the B_1 orbital in the supermolecule (5.70 eV) is almost equal to the calculated IP of the B_1 orbital in **5** (5.78 eV).

In order to assess the consistency of 6-31G SCF predictions concerning the influence of through-space interactions on IPs, another supplementary 6-31G SCF calculation was carried out on a second supermolecule. This supermolecule contained 9-methylguanine and PO_2^-. In this calculation, the distances between the charged O atoms of PO_2^- and 9-methylguanine were similar to distances between guanine and the negatively charged O atoms of the phosphate group in 5′-dGMP$^-$ (**5**). In the supermolecule, the distances between one O atom and the C_8 and N_7 atoms of 9-methylguanine were 4.3 and 5.1 Å, respectively.

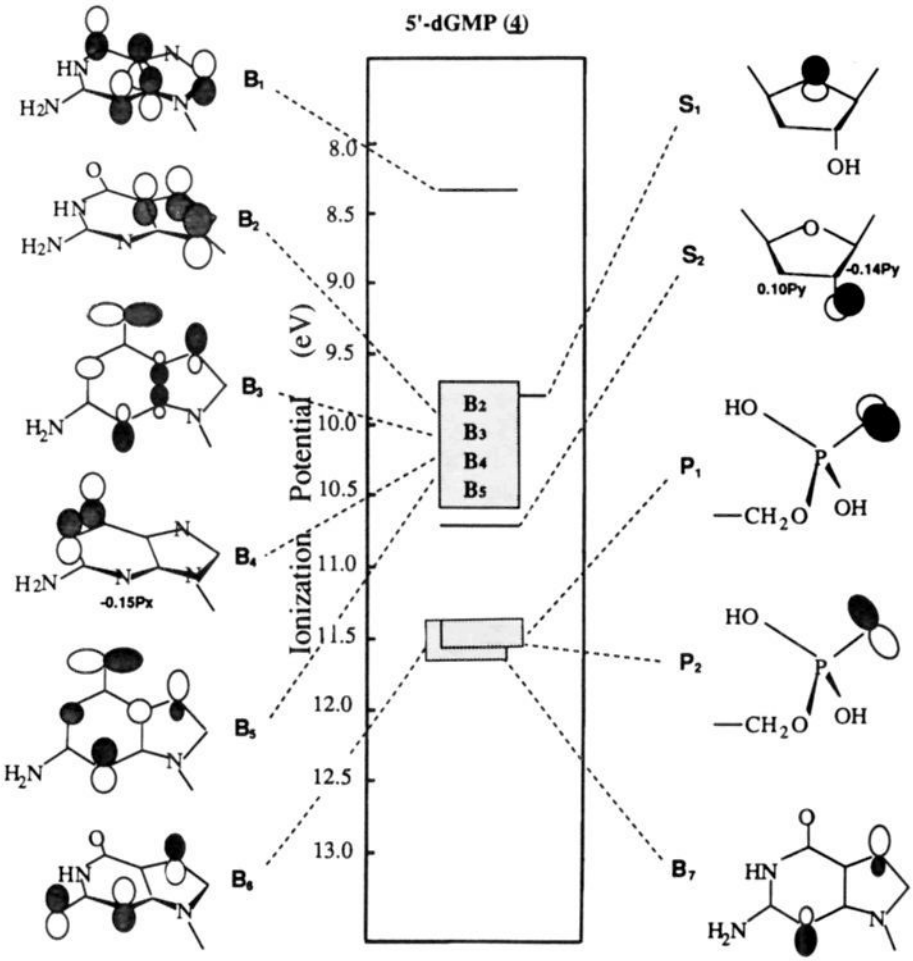

Figure 5. Photoelectron corrected valence electron ionization potentials of 5′-dGMP (**4**).

The distances between the other O atom and the C_8 and N_7 atoms were 5.5 and 6.6 Å. In **5**, the distances between the $O_{2''}$ atom and the C_8 and N_7 atoms are 4.3 and 5.0 Å. The distances between the $O_{4''}$ atom and the C_8 and N_7 atoms are 5.1 and 6.6 Å. For the supermolecule, the results from the 6-31G SCF calculations indicate that the B_1 ionization potential is 2.4 eV smaller than the B_1 ionization potential of isolated 9-methylguanine. This difference is nearly equal to the difference (2.5 eV), obtained from 6-31G SCF calculations, between the ionization potentials of the B_1 orbitals in 5′-dGMP$^-$ (**5**) and 5′-dGMP (**4**).

Phosphate Orbitals. Similarities between results of the 6-31G SCF population analyses for the phosphate group of 5′-dGMP$^-$ (**5**) and the model anion, $CH_3HPO_4^-$, and the correlation between the orbital diagrams for **5** and $CH_3HPO_4^-$ provide evidence that phosphate group ionization potentials in **5** are nearly equal to those in $CH_3HPO_4^-$. For 5′-dGMP$^-$ (**5**) corrected phosphate IPs of the P_1–P_5 orbitals were taken to be equal to those in $CH_3HPO_4^-$. It is likely that this approximation is reliable, even though the IPs of the P_1–P_5 orbitals in **5**, obtained from 6-31G SCF calculations, are 0.60–0.82 eV larger than the corresponding calculated IPs in $CH_3HPO_4^-$. These differences are again expected to be due largely to differences in the energetic ordering of the P_1–P_5 orbitals in the valence manifolds of **5** and $CH_3HPO_4^-$. In **5** the P_1–P_5 orbitals are predicted to be the second, third, fourth, fifth, and eighth highest occupied orbitals, while in $CH_3HPO_4^-$ they are the five highest occupied orbitals.

Unlike the base and sugar model compounds, 1,9-dimethylguanine (**1**) and 3-hydroxytetrahydrofuran (**2**), no photoemission data is available for $CH_3HPO_4^-$. Furthermore, because of the large amount of computation time required, it is not currently feasible to carry out reliable post-SCF calculations on $CH_3HPO_4^-$. Corrected values ($IP_{corr,CH_3HPO_4^-}(i)$) of the five lowest energy ionization potentials of $CH_3HPO_4^-$ were obtained from eq 7.

$$IP_{corr,CH_3HPO_4^-}(P_i) = IP_{calc,CH_3HPO_4^-}(P_i) - \Delta IP_i \qquad (7)$$

Here $IP_{corr,CH_3HPO_4^-}(P_i)$ is the corrected value of the ith lowest energy ionization potential of $CH_3HPO_4^-$, and $IP_{calc,CH_3HPO_4^-}(P_i)$ is the value of the ith ionization potential obtained from 6-31G SCF calculations. In eq 7, $\Delta IP(i)$ is equal to the difference between the value of the ith ionization potential of $H_2PO_4^-$, obtained from post-SCF calculations, and the value obtained from the 6-31G SCF calculations. At the post-SCF level, the first ionization potential of $H_2PO_4^-$ was obtained from results of MP2/6-31+G* calculations on the closed-shell ground-state anion and on the ground-state neutral radical. In test calculations on oxygen and phosphorus containing anions, this method yielded values of 1.51, 3.34, and 4.90 eV for the first IPs of CH_3O^-, PO_2^-, and PO_3^-, respectively. The MP2/6-31+G* values agree well with the experimental values of 1.57, 3.3 ± 0.2 and 4.9 ± 1.3 eV.[59,60]

Post-SCF values for the second, third, fourth, and fifth IPs of $H_2PO_4^-$ were obtained from calculations of excited-state energies of the neutral radical, employing the CI singles approach. The values for the first to fifth IPs of $H_2PO_4^-$, obtained from the combined use of MP2 and CIS calculations, are 4.89, 6.21, 6.36, 7.24, and 8.80 eV, respectively. The values obtained from the 6-31G SCF calculations are 6.33, 6.67, 7.42, 7.48, and 8.42 eV.

In a test of the accuracy of the method employed here to calculate ionization potentials of $H_2PO_4^-$, IPs of NO_2 were obtained using a similar method. The first IP was obtained from MP3 calculations.[61] The energies of the first through fourth excited states of NO_2^+ were obtained from CI singles calculations. The theoretical IPs (11.15, 12.09, 13.11, 13.95, and 14.83 eV), obtained in this manner, differed from the experimental IPs (11.23, 13.02, 13.61, 14.08, and 14.53 eV) by an average of 0.39 eV.[54]

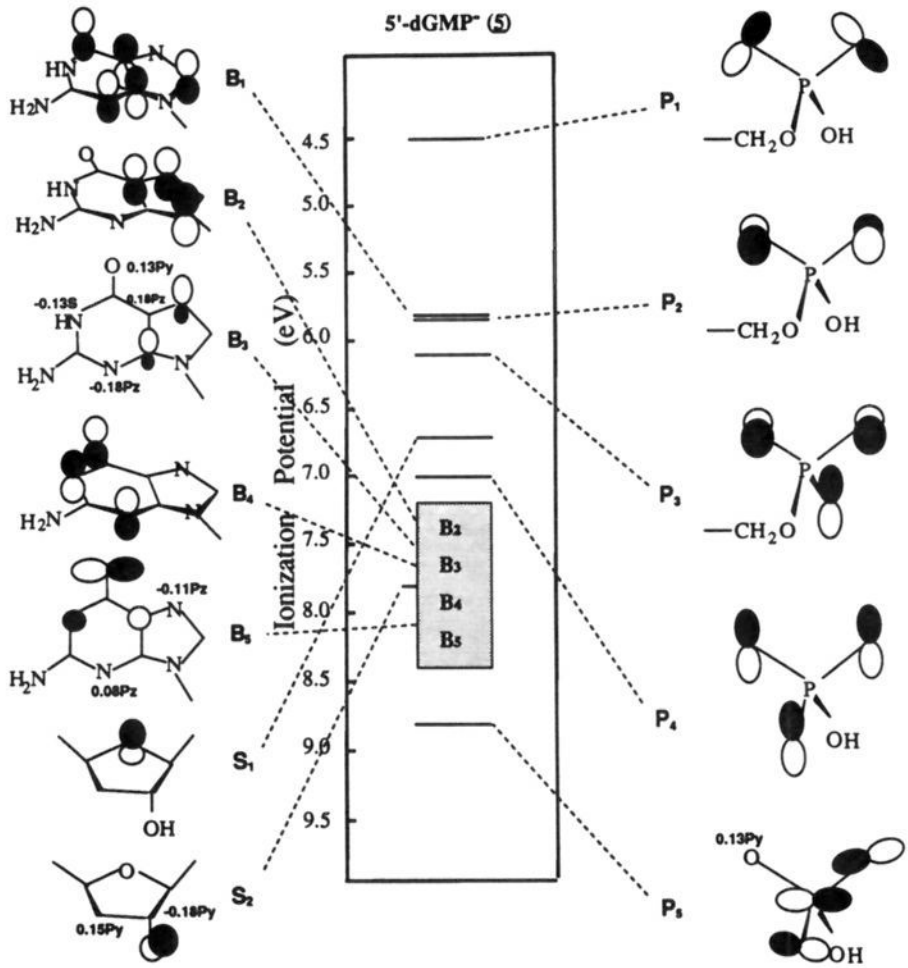

Figure 6. Corrected valence electron ionization potentials of 5'-dGMP⁻ (**5**).

Configuration Interaction Descriptions of Electron Holes. In many cases Koopmans' theorem does not yield accurate ionization potentials; however, even in these cases, the description of ionization which the theorem provides is sometimes qualitatively valid. When Koopmans' theorem describes ionization events in a qualitatively accurate manner, the diagrams in Figures 1–6, which describe occupied orbitals in molecules and anions before removal of an electron, also represent the electron hole which is created after ionization occurs. In order to test the description of electron holes obtained by an application of Koopmans' theorem, a series of calculations, at the post-SCF level, were carried out to examine ionization events in the model compounds and anions 1,9-dimethylguanine (**1**), 3-hydroxytetrahydrofuran (**2**), H_3PO_4, and $H_2PO_4^-$.

Because of restrictions on computational resources, it is not possible at this time to carry out rigorous ab initio CI calculations on **1** and **2** or on the radical cations derived from **1** and **2**. However, recent calculations on **1**, carried out employing the semiempirical HAM/3 CI method, have indicated that, of the seven lowest energy radical cation states associated with **1**, five are qualitatively well described within the context of Koopmans' theorem.[15,62] The only significant CI effects in the low-energy region of the PE spectrum of 1,9-dimethylguanine (**1**) give rise to hole mixing between the radical cation states arising from removal of electrons from the B_2 and B_4 orbitals. This description, provided by the HAM/3 CI calculation, is consistent with the results in Figure 1. The figure indicates that in the PE spectrum of **1**, bands arising from cation states associated with removal of electrons from the B_2 and B_4 orbitals appear in a region of the spectrum which is poorly resolved, and which has contributions from several cation states of nearly equal energy. The top of Figure 7 shows composite electron holes, predicted by the HAM/3 CI calculation, for the cation states B_2' and B_4' which arise from hole mixing associated with the B_2 and B_4 orbitals. It should be noted that, while the HAM/3 CI calculation indicates that the description of the states arising from removal of electrons in the B_2 and B_4 orbitals in Figures 1–3, 5, and 6 requires modification, the hole-

(53) Mulliken, R. S. *Chem. Phys.* **1955**, *23*, 1833.
(54) Kimura, K.; Katsumata, S.; Achiba, Y.; Yamazaki, T.; Iwata, S. *Handbook, of HeI Photoelectron Spectra of Fundamental Organic Molecules*; Halsted Press: New York, 1981; pp 36, 57, 209.
(55) (a) Bain, A. D.; Bünzli, J. C.; Frost, D. C.; Weiler, L. *J. Am. Chem. Soc.* **1973**, *95*, 291. (b) Gerson, S. H.; Worley, S. D.; Bodor, N.; Kaminski, J. J.; Flechtner, T. W. *J. Electron Spectrosc. Relat. Phenom.* **1978**, *13*, 421.
(56) For methyl phosphate experimental IPs have not been measured because of the ease with which samples of methyl phosphate are hydrolyzed by trace quantities of water.
(57) Urano, S.; Fetzer, S.; Harvey, R. G.; Tasaki, K.; LeBreton, P. R. *Biochem. Biophys. Res. Commun.* **1988**, *154*, 789.
(58) Rossman, M. A.; Leonard, N. J.; Urano, S.; LeBreton, P. R. *J. Am. Chem. Soc.* **1985**, *107*, 3884.
(59) (a) Engelking, P. C.; Ellison, G. B.; Lineberger, W. C. *Chem. Phys.* **1978**, *69*, 1826. (b) Wu, R. L. C.; Tiernan, T. O. *Bull. Am. Phys. Soc.* **1982**, *27*, 109.
(60) (a) Lohr, L. L.; Boehm, R. C. *J. Phys. Chem.* **1987**, *91*, 3203. (b) Henchman, M.; Vigianno, A. A.; Paulson, J. F.; Freedman, A.; Wormhoudt, J. *J. Am. Chem. Soc.* **1985**, *107*, 1453. (c) Unkel, W.; Freedman, A. *AIAA* **1983**, *21*, 1648. (d) Freedman, A.; Warmhoudt, J. C.; Kolb, C. E. In *Metal Bonding and Interactions in High Temperature Systems, ACS Symposium Series 179*; Gole, J. L., Stwalley, W. C., Eds.; American Chemical Society: Washington, 1982; p 609.
(61) For neutral NO_2, the value of the first ionization potential obtained from MP2 calculations (10.03 eV) was less accurate than values obtained from MP2 calculations on the anions CH_3O^-, PO_2^-, and PO_3^-. However, the value obtained from MP3 calculations agreed well with the experimental value (11.23 eV).
(62) In ref 15, results from test calculations on pyrimidine employing the HAM/3 CI method has provided a description of CI effects in the low-energy region of the PE spectrum which is in good agreement with that obtained using more rigorous ab initio Green function methods.

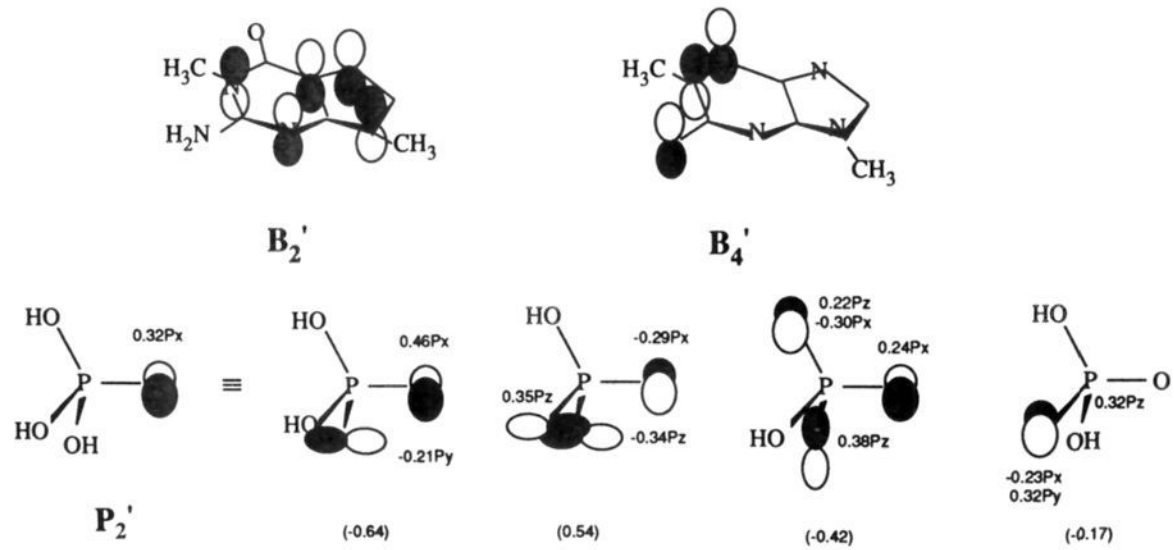

Figure 7. Composite electron holes arising from low-energy ionization events in 1,9-dimethylguanine (**1**) and H_3PO_4 obtained from CI calculations. Electron holes, B_2' and B_4' were constructed from results of HAM/3 CI calculations indicating the occurrence of hole mixing associated with the removal of electrons from the second (B_2) and third (B_4) highest occupied π orbitals in 1,9-dimethylguanine (**1**). The electron hole, P_2', shown below, was constructed by comparing results from ab initio MP2 and CIS calculations describing the ground and first excited state of $H_3PO_4^+$ with SCF calculations on H_3PO_4. Diagrams of B_2', B_4', and P_2' were drawn by using atomic orbital coefficients associated with orbitals occupied in 1,9-dimethylguanine (**1**) and H_3PO_4 but unoccupied in configurations which significantly contribute to the CI wave functions describing the radical cation states examined. Atomic orbital coefficients employed in the diagrams of electron deficiency distributions were obtained by multiplying the molecular orbital coefficients by the configuration coefficients. For P_2', contributions to the resulting hole are shown in diagrams containing molecular orbital coefficients and configuration coefficients resulting from the CIS calculation on $H_3PO_4^+$. These are given at the bottom right. It should be noted that the B_2', B_4', and P_2' composite holes provide an orbital representation of ionization events described by post-SCF wave functions. In this regard, they are not rigorous.

mixing states, B_2' and B_4', lie in the same unresolved energy region of the PE spectrum of 1,9-dimethylguanine (**1**) as that assigned to the Koopmans' states associated with the B_2 and B_4 orbitals.

In the present investigation, HAM/3 CI calculations were carried out on 3-hydroxytetrahydrofuran (**2**). The results indicate that a description of the two lowest energy cation states based on Koopmans' theorem is qualitatively accurate and that a modification, like that required for **1**, is not needed for the description of states associated with removal of electrons from the S_1 and S_2 orbitals shown in Figure 1.

In order to test descriptions of electron holes associated with removal of electrons from the neutral and anionic phosphate groups in 5′-dGMP (**4**) and 5′-dGMP$^-$ (**5**), results from closed-shell 6-31G SCF calculations on H_3PO_4 and $H_2PO_4^-$ were compared to results from ab initio post-SCF calculations on the $H_3PO_4^+$ radical cation and on the $H_2PO_4^\bullet$ neutral radical. Here, MP2 calculations with the 6-31+G* basis set were used to describe the ground states of $H_3PO_4^+$ and $H_2PO_4^\bullet$, and CIS calculations were used to describe the excited states. For the lowest energy ionization events in H_3PO_4 and $H_2PO_4^-$, results of the MP2 calculations on the ground states of $H_3PO_4^+$ and $H_2PO_4^\bullet$ indicate that configurations, other than the principal Koopmans' configurations, which contribute to the perturbation wave functions, all have coefficients smaller than 0.1. In these calculations the perturbation wave function was normalized so that the coefficient for the principal Koopmans' configuration is 1.0. In support of the qualitative validity of a Koopmans' description of the lowest energy ionization of H_3PO_4, an examination of the lowest unoccupied molecular orbital (LUMO) in the lowest energy configuration of the ground-state perturbation wave function of $H_3PO_4^+$ indicates that the hole distribution in $H_3PO_4^+$ is similar to the distributions of the P_2 orbital of H_3PO_4 and of the P_2 orbitals of trimethyl phosphate (**3**) and 5′-dGMP (**4**), shown in Figures 1, 2, and 5. An examination of the lowest energy configuration in the perturbation wave function for the ground state of $H_2PO_4^\bullet$ yields a similar result. Here, the orbital diagram of the LUMO of $H_2PO_4^\bullet$ is nearly identical to the orbital diagrams of the P_1 orbitals of $H_2PO_4^-$, $CH_3HPO_4^-$, and 5′-dGMP$^-$ (**5**) shown in Figures 3, 4, and 6.

A comparison of the CIS wave function describing the first excited state of $H_3PO_4^+$ with the MP2 and SCF wave functions describing the ground states of $H_3PO_4^+$ and H_3PO_4, respectively, yields a description of a composite excited-state electron hole, P_2'. This is shown at the bottom of Figure 7. Holes appearing in major configurations contributing to this excited state are shown on the bottom right of the figure.[63] The composite hole, P_2', shown on the bottom left, has been constructed from the component holes. The distribution in P_2' is similar to that of the P_1 orbitals in trimethyl phosphate (**3**) and 5′-dGMP (**4**), which are shown in Figures 1, 2, and 5.

Figure 8 shows electron holes in $H_2PO_4^\bullet$ obtained from CIS calculations describing neutral radical states arising from the second through the fifth ionization events in $H_2PO_4^-$. The right side of the figure shows holes appearing in the major configurations contributing to the excited states of $H_2PO_4^\bullet$. The composite holes, labeled P_2'–P_5', are shown on the left. For two of these states, P_2' and P_3', results from the CIS calculations indicate that the composite holes are similar to holes in Koopmans' configurations, that arise from the removal of an electron from the P_2 and P_4 orbitals of $H_2PO_4^-$. According to the CIS results, the remaining two states, with composite holes, P_4' and P_5', are strongly mixed and are not similar to holes in Koopmans' configurations. These results provide evidence that the descriptions of phosphate ionization events in $H_2PO_4^-$, $CH_3HPO_4^-$ and 5′-dGMP$^-$ (**5**), which are given in Figures 3, 4, and 6, and which employ the P_2 to P_5 orbital diagrams, should be modified so as to reflect the results in Figure 8.

Discussion

Electronic Structures of 5′-dGMP (4) and 5′-dGMP$^-$ (5). Within the SCF framework, the results in Figures 2 and 5 indicate that for the base, sugar, and phosphate groups of neutral 5′-

(63) According to the CIS calculations on $H_3PO_4^+$, the difference between the first and second IPs of H_3PO_4 is 0.23 eV. This is consistent with the assignment of the PE spectrum of trimethyl phosphate (**3**) in Figure 1.

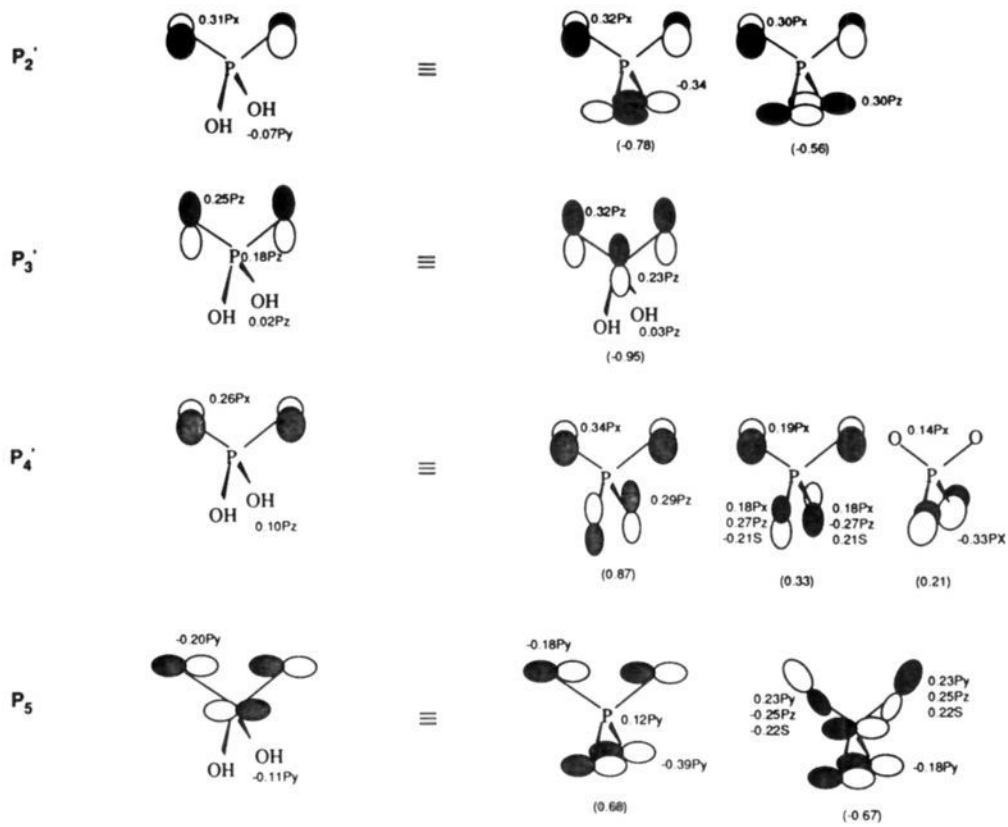

Figure 8. Electron holes (P_2'–P_5') for the second through fifth lowest energy ionization events in $H_2PO_4^-$ obtained by comparing results from ab initio MP2 and CIS calculations on $H_2PO_4^{\bullet}$ with results from 6-31G SCF calculations on $H_2PO_4^-$. Electron holes were drawn using the method employed to obtain a description of P_2' in Figure 7. Contributions to the resulting holes are shown in diagrams containing molecular orbital coefficients and configuration coefficients resulting from the CIS calculation on $H_2PO_4^{\bullet}$. These are given on the right.

dGMP (**4**) the energetic ordering of orbitals is the same before and after correction of the IPs. In both cases, the smallest IP is associated with the base. The IP of the highest occupied sugar orbital is greater than those of the upper occupied base orbitals but smaller than that of the highest occupied phosphate orbital. Within the base, the energy difference between the IPs of the B_1 and B_7 orbitals after correction (4.7 eV) is smaller than the difference before correction (5.1 eV). In 5′-dGMP (**4**), correcting the 6-31G SCF results leads to a small increase (0.09 eV) in the ionization potential of the highest occupied B_1 orbital. For other ionization events, correction leads to larger changes in the IPs. After correction, the IPs involving the B_2–B_7 orbitals in 5′-dGMP (**4**) are reduced from 0.4 to 2.2 eV; the IPs of the S_1 and S_2 orbitals are reduced 2.5 and 2.0 eV; and the IPs of the P_1 and P_2 orbitals are reduced by 1.4 eV.

For 5′-dGMP$^-$ (**5**), correcting the 6-31G SCF results similarly leads to a reduction of 11 of the 12 lowest energy ionization potentials. The IP of the B_1 orbital is again increased by 0.09 eV. However, correction of the 6-31G SCF results for 5′-dGMP$^-$ (**5**), unlike that for (**4**), gives rise to a significantly different ordering of IPs than that predicted by the uncorrected results. Before correction, the lowest IP in **5** is associated with a π orbital (B_1) on the base. After correction, the IP of the phosphate group is predicted to be 1.3 eV smaller than that of the base. This change in the ordering of the base and phosphate ionization potentials is consistent with the observation that the total charge on the phosphate group, obtained from the 6-31G SCF calculations, is –0.92 eu. The improved method employed for the scaling of phosphate group ionization events has led to values (5.8, 6.1, 7.0 and 8.8 eV) for the P_2–P_5 ionization potentials in 5′-dGMP (**5**). These differ significantly from the values (5.3, 5.8, 5.8, and 6.5 eV) which were reported earlier,[8] and which were scaled using results from less accurate excited-state energies of PO_2^- obtained from MP2 calculations.[7]

The description of the HOMO of 5′-dGMP$^-$ (**5**), which is given in Figure 6, is consistent with that previously reported for the S–P–S–G sequence.[5] The corrected ionization potential of 5′-dGMP$^-$ (**5**) which is reported here (4.5 eV) is also similar to the value (5.1 eV) which was previously reported. However, the present description of the upper occupied orbitals in 5′-dGMP$^-$, which indicates that orbital structures are largely localized on either the base sugar or phosphate groups differs from the earlier description of the S–P–S–G sequence obtained from a single ζ basis set. According to results from the previous investigation, 10 of the 12 highest occupied orbitals are significantly delocalized over two or more groups.

The results of MP2 and CIS calculations indicate that CI influences are more important in the description of the model anion, $H_2PO_4^-$, than in the neutral, base, sugar, and phosphate group model compounds, 1,9-dimethylguanine (**1**), 3-hydroxytetrahydrofuran (**2**), and H_3PO_4. Of the 11 low-energy ionization events examined in **1**, **2**, and H_3PO_4, only two do not arise primarily from a single Koopmans' configuration associated with removal of an electron from an occupied orbital in the ground-state neutral molecule. Of the five low-energy ionization events examined in $H_2PO_4^-$, it is possible to qualitatively describe two within the context of Koopmans' theorem. The calculated influence of CI on low-energy ionization events in the model compounds and anions provides evidence that similar effects influence ionization events in 5′-dGMP (**4**) and 5′-dGMP$^-$ (**5**).

DNA and RNA Methylation and Ethylation Patterns. The Influence of Base Polarizability. Figure 9 shows plots of the percent alkylation which occurs at different bases in DNA and RNA versus the base ionization potentials of anionic nucleotides for reactions of methyl and ethyl methanesulfonates (MeMs and EtMs), dimethyl and diethyl sulfates (Me_2SO_4 and Et_2SO_4), and *N*-methyl- and *N*-ethyl-*N*-nitrosoureas (MeNU and EtNU). The upper panel of Figure 9 shows results from reactions with double-

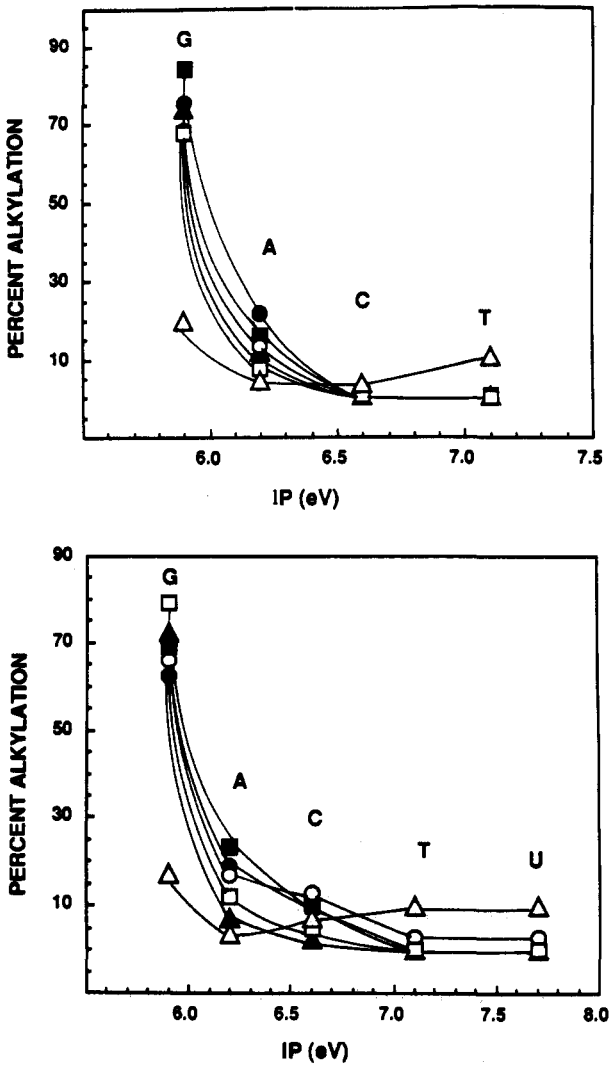

Figure 9. Alkylation patterns for reactions of methyl methanesulfonate (solid squares), ethyl methanesulfonate (open squares), dimethyl sulfate (solid circles), diethyl sulfate (open circles), *N*-methyl-*N*-nitrosourea (solid triangles), and *N*-ethyl-*N*-nitrosourea (open triangles). Percentage of total alkylation which occurs at guanine (G), adenine (A), cytosine (C), thymine (T), and uracil (U) is plotted versus base IPs in the nucleotides 5′-dGMP⁻, 5′-dCMP⁻, 2′-deoxyadenosine 5′-phosphate (5′-dAMP⁻), 2′-deoxythymidine 5′-phosphate (5′-dTMP⁻), and uridine 5′-phosphate (5′-UMP⁻). Base ionization potentials in 5′-dAMP⁻, 5′-dCMP⁻, 5′-dTMP⁻, and 5′-UMP⁻ were obtained by correcting results from 6-31G SCF calculations on the nucleotides employing geometries obtained from refs 52 and 64, and photoelectron data obtained from refs 1d,j,l and 7. The upper panel shows results for double-stranded DNA obtained from reactions with DNA from salmon sperm, calf thymus, salmon testes, rat liver and brain, human fibroblasts, and HeLa and V79 cells. For T and C, data points in the upper panel overlap for all reactions except those with *N*-ethyl-*N*-nitrosourea. The lower panel shows results for single-stranded DNA and RNA obtained from reactions with DNA from M13 phage and with RNA from TMV, yeast, HeLa cells, animal ribosomes, and m_2 phage. See ref 22.

stranded DNA. The lower panel shows results from reactions with single-stranded DNA and RNA. For all nucleotides, the lowest energy, base ionization potentials are associated with the removal of π electrons. The results in Figure 9 demonstrate the similarities in the product distributions from reactions of MeMs, EtMs, Me_2SO_4, Et_2SO_4 and MeNU, in which the base reactivities increase as the base π ionization potentials in the nucleotide anions decrease. The figure also demonstrates that the product distributions for these reactions are significantly different from that for EtNU. Specifically, EtNU shows no evidence of the base selectivity exhibited by the other methylating and ethylating reagents.

The unique alkylation pattern of EtNU is exhibited in other ways as well. For example, in *in vitro* reactions of EtNU with double-stranded DNA, and with single-stranded DNA and RNA, the percent of alkylation which occurs at the bases is less than that which occurs at the phosphate ester backbone. Approximately 38% of the reaction occurs at the bases, while from 57 to 65% occurs at the backbone. In contrast, for MeMs, EtMs, Me_2SO_4, Et_2SO_4, and MeNU the percentage of reaction which occurs at the bases is more than five times larger than that occurring at the backbone. Similarly, reactions with EtNU exhibit less base site specificity than reactions with the other methylating and ethylating reagents. For EtNU reactions with guanine, the percentage of ethylation which occurs at the N7 atom (10–11.5%) and at the *O6′* atom (7–7.8%) is nearly equal. In guanine reactions with MeMs, EtMs, Me_2SO_4, Et_2SO_4, and MeNU, reaction at N7 is more than ten times more favorable than reaction at *O6′*.

While all of the reactions considered in Figure 9 proceed via an S_N2-like mechanism,[21,28,30,37] they lie on different points along the spectrum between S_N1 and S_N2.[21,30,37] It is likely that the lower selectivity exhibited by EtNU is due to the higher reactivity and to the greater S_N1 character of the intermediate ethanediazonium ion formed in EtNU reactions.[21,28] Reactions of MeMS, EtMS, Me_2SO_4, Et_2SO_4, and of the intermediate methanediazonium ion formed from MeNU have greater S_N2 character. This is consistent with previous speculation, made within the context of Pearson's hard–soft acid–base principle,[65] that, compared to MeMs, EtMs, Me_2SO_4, Et_2SO_4 and MeNU, the higher reactivity of EtNU at the negatively charged oxygen atoms of the phosphate backbone and at the exocyclic O atoms on the bases indicates that ethanediazonium ion reactions involve a hard electrophile.[28,66]

Results from MNDO calculations indicate that, in reactions of ethanediazonium ions with nucleosides, forming bond distances in the transition states are 0.534–1.335 Å greater than those in corresponding reactions of methanediazonium ions.[37] Furthermore, in the methanediazonium ion transitions states the charge transfer is 0.015–0.125 eu greater than in the corresponding ethanediazonium ion transition states.[37] In the contracted methanediazonium ion transition states, in which significant charge transfer occurs, short-range orbital interactions are relatively more important than in the ethanediazonium ion transition states. In the ethanediazonium ion transition states, in which there is greater charge separation, long-range electrostatic forces play a more important role.[30,37] The quantitative differences which emerge from the molecular orbital calculations support the conclusion that the ethanediazonium ion transition states involve hard electrophiles and that the methanediazonium ion transition states involve soft electrophiles.

For MeMs, EtMs, Me_2SO_4, Et_2SO_4, and MeNU, the preference for reaction at cyclic N atoms of the bases over reaction at the exocyclic O atoms of the bases or at the charged O atoms of the phosphate groups is characteristic of soft electrophiles. The results in Figure 9 are consistent with a description of reactions of soft electrophiles with soft nucleophiles in which nucleophilic strength is strongly coupled to polarizability.[37] For the bases in DNA and RNA,[3,67] and for other series of structurally related unsaturated molecules,[68,69] increases in polarizability are often accompanied by decreases in π ionization potentials. According to results from

(64) Dock-Bregeon, A. C.; Chevrier, B.; Podjarny, A.; Johnson, J.; de Bear, J. S.; Gough, G. R.; Gilham, P. T.; Moras, D. *J. Mol. Biol.* **1989**, *206*, 459.
(65) (a) Pearson, R. G. *Hard and Soft Acids and Bases*; Dowden, Hutchinson and Ross: Stroudsburg, PA, 1981. (b) Ho, T.-L. *Hard and Soft Acids and Bases Principle in Organic Chemistry*; Academic Press: New York, 1977.
(66) March, J. *Advanced Organic Chemistry*; John Wiley: New York, 1985; p 324.
(67) Ts'o, P. O. P. In *Basic Principles in Nucleic Acid Chemistry*; Ts'o, P. O. P., Ed.; Academic Press: New York, 1974; pp 526–562.
(68) Zegar, I. S.; Prakash, A. S.; Harvey, R. G.; LeBreton, P. R. *J. Am. Chem. Soc.* **1985**, *107*, 7990.
(69) Fetzer, S. M. Ph.D. Dissertation, The University of Illinois at Chicago, Chicago, IL, 1992.

6-31G SCF calculations, the polarizabilities of uracil, thymine, cytosine, adenine, and guanine are 7.63, 9.30, 8.39, 10.74, and 11.29 Å^3, respectively.[70] In small soft nucleophiles, such as thiourea and iodide ions, the manner in which polarizability correlates with nucleophilic strength has long been recognized[71] and has been explicitly incorporated into classical descriptions of nucleophilic substitution such as that provided by a modification of the Swain–Scott relationship.[71a,72]

Earlier treatments focussed mainly on the polarizabilities of donor lone-pair electrons of the nucleophile and on the importance of attractive London dispersion forces.[71b,73] However, for large soft nucleophiles with labile, delocalized π electronic systems, the importance of π electron polarizabilities has also been recognized.[71b] The reactions of these nucleophiles with soft electrophiles can be strongly influenced by the ability of the π electrons to adjust, in an energetically favorable way, to changes in geometry and charge distribution which occur as the transition state is approached. This is consistent with the notion that softness is synonymous with polarizability.[66]

Steric Interactions in Nucleoside–Methanediazonium Ion Transition States. While the data in Figure 9 support descriptions of reactions of MeMs, EtMs, Me_2SO_4, Et_2SO_4, and MeNU with DNA and RNA bases in which short-range electronic interactions are important, steric influence is also expected to be strong. Both DNA secondary structure and the structure of individual nucleotides influence alkylation patterns. However, the steric influence on methylation and ethylation which is associated with DNA secondary structure, specifically the shielding which arises via base pairing and stacking, is modest. For example, the strong temperature dependence of adenine (N1) ethylation in reactions of EtNU with poly(rA)·poly(rU) is attributed to hydrogen bonding associated with the adenine–uracil base pair.[74] However, in reactions with double-stranded DNA, no such temperature dependence occurs for the ethylation of guanine (*O6*), thymine (*O4*), or cytosine (*O2*), even though these three oxygen atoms are hydrogen-bonded.[74] The conclusion that conformational effects on methylation and ethylation patterns are small is supported by the observation that ethylation and methylation patterns measured for single-stranded DNA and RNA are similar to those measured for double-stranded DNA.[22] For example, in reactions of EtMs with double-stranded DNA, the fraction of nucleotide modification which occurs at guanine, phosphate, adenine, and cytosine is 67.9, 13.0, 7.7, and 0.7%, respectively. In reactions of EtMs with single-stranded DNA and RNA, the fraction of nucleotide modification which occurs at guanine, phosphate, adenine, and cytosine is 79.0, 10.0, 12.0, and 5.0%. In these experiments, no modification of uracil or thymine was detected.[22]

It is likely that steric interactions involving nucleotide primary structure are more important. This speculation is supported by the results in Table III which lists differences in MNDO core repulsion energies and 6-31G SCF nuclear repulsion energies for different transition states occurring in reactions of methanediazonium ions with 2'-deoxyguanosine. The table also lists differences in core and nuclear repulsion energies occurring in reactions with 2'-deoxycytidine and 2'-deoxyadenosine. For the guanosine, cytidine, and adenosine reactions, the repulsion energy differences have been calculated relative to the guanine (*O6*), cytosine (*O2*), and adenine (N1) transition states, respectively. Both the core and the nuclear repulsion energies were calculated using previously reported MNDO transition-state geometries.[37] For the nucleosides, Table III also gives the average change in the two H—N—C=N dihedral angles of the exocyclic amino groups, which occurs as a reaction proceeds from the isolated reactants to the transition state.

Table III. Steric Interaction in Nucleoside–Methanediazonium Ion Transition States[a]

site	methylation $\Delta E_{\text{core repulsion}}$[b,c]	methylation $\Delta E_{\text{nuclear repulsion}}$[b,d]	$\Delta\varphi_{av}$[e,f]
guanine (O6)	0.0	0.0	3.0
guanine (N7)	491.8	986.5	9.0
guanine (N3)	1123.1	2187.5	44.0
cytosine (O2)	0.0	0.0	8.7
cytosine (N3)	39.3	87.5	17.1
adenine (N1)	0.0	0.0	30.7
adenine (N7)	435.2	829.7	41.8
adenine (N3)	823.2	1582.5	16.6

[a] Transition-state geometries taken from ref 37. [b] In eV. [c] Obtained from MNDO calculations. [d] Obtained from 6-31G SCF calculations employing MNDO transition-state geometries. [e] Average change in exocyclic amino group dihedral angles H—N—C=N which occur as reaction proceeds from the isolated reactants to the transition state. [f] In degrees.

In single-stranded DNA, the exocyclic guanine (*O6*) and the cytosine (*O2*) sites are unhindered. For the guanine (N3), adenine (N1), and adenine (N7) sites, steric interference arises through repulsion between the attacking methanediazonium ion and the exocyclic amino groups on the bases. This is demonstrated by the large distortion of the amino group dihedral angles which occurs in the transition states of these reactions, compared to the small amino group distortion which occurs in the guanine (*O6*) or cytosine (*O2*) transition states. For the adenine (N3) transition state, repulsion of the deoxyribose group is also large. This is indicated by the large core and nuclear repulsion energies associated with reaction at adenine (N3), compared to reaction at adenine (N1) or adenine (N7). For reaction at guanine (N7), the amino group distortion is small, and the core and nuclear repulsion energy differences, referenced to reaction at guanine (*O6*), are less than half the energy differences associated with reaction at guanine (N3). These results provide evidence that steric accessibility strongly influences the higher reactivity of Me_2SO_4, MeMs, MeNu, Et_2SO_4 and EtMs at guanine (N7), compared to guanine (N3), adenine (N1), adenine (N3), and adenine (N7).

Acknowledgment. Support of this work by The American Cancer Society (Grant No. CN-37), the Petroleum Research Fund (Grant No. 21314-AC), The National Cancer Institute of the National Institutes of Health (Grant No. CA41432) and Cray Research, Inc., is gratefully acknowledged. Computer access time has been provided by the Computer Center of the University of Illinois at Chicago, the Cornell National Supercomputer Facility, and the National Center for Supercomputing Applications, at the University of Illinois at Urbana-Champaign. The authors would like to thank Professor Robert Moriarty (The University of Illinois at Chicago) for helpful discussions and Professor George Ford (Southern Methodist University) for providing the nucleoside–methanediazonium ion transition-state geometries.

(70) An exception to the rule that polarizabilities increase as IPs decrease is indicated by the observation that the polarizability of thymine (9.30 Å^3) is larger than that of cytosine (8.39 Å^3), while the IP of thymine in 5'-dTMP$^-$ (7.1 eV) is larger than that of cytosine in 5'-dCMP$^-$ (6.6 eV). This exception arises because of the additive effect which the methyl group at the 5-position of thymine has on polarizability.

(71) (a) Edwards, J. O. *J. Am. Chem. Soc.* **1956**, *78*, 1819. (b) Edwards, J. O.; Pearson, R. G. *J. Am. Chem. Soc.* **1962**, *84*, 16.
(72) Swain, C. G.; Scott, C. B. *J. Am. Chem. Soc.* **1953**, *75*, 141.
(73) Bunnett, J. F. *Annu. Rev. Phys. Chem.* **1963**, *14*, 271.
(74) Bodell, W. J.; Singer, B. *Biochemistry* **1979**, *18*, 2860.