# UV Photoelectron and ab Initio Quantum Mechanical Characterization of Nucleotides: The Valence Electronic Structures of 2'-Deoxycytidine-5'-phosphate

**Kenzabu Tasaki, Xu Yang, Shigeyuki Urano, Sharon Fetzer, and Pierre R. LeBreton***

*Contribution from the Department of Chemistry, University of Illinois at Chicago, Chicago, Illinois 60680.* *Received May 25, 1989*

**Abstract:** He I UV photoelectron spectroscopy and ab initio molecular orbital calculations with the 4-31G basis set have been employed to characterize the valence electronic structures of neutral and anionic 2'-deoxycytidine-5'-phosphate (5'-dCMP). Theoretical ionization potentials (IP's) of neutral molecules and of anions have been obtained by applying Koopmans' theorem to ab initio SCF results. The ionization potentials predicted from the SCF calculations have been compared to He I photoelectron spectra of the neutral model compounds, 1-methylcytosine, tetrahydrofuran, cyclopentanol, and the trialkylphosphate esters, trimethyl, triethyl, and tributylphosphate. The SCF calculations predict a value (8.94 eV), for the highest occupied $\pi$ orbital in 1-methylcytosine, which agrees well with the experimental vertical IP (8.65 eV). However, IP's for the highest occupied lone-pair orbitals in tetrahydrofuran, cyclopentanol, and the trialkylphosphates are predicted to be higher than the experimental IP's by 1.51, 0.84, and 1.18–1.41 eV, respectively. Ionization potentials obtained from SCF calculations on the model anions $CH_3O^-$, $PO_2^-$, and $PO_3^-$ have been compared with experimental adiabatic IP's. The results of the SCF calculations for these model anions and for $H_2PO_4^-$ have also been compared with results from third-order Møller–Plesset perturbation calculations. The comparison of 4-31G SCF results with experimental data and with results from post-SCF calculations on the model compounds and anions permits an individual orbital evaluation of the accuracy of IP's calculated for neutral and anionic 5'-dCMP. In both the neutral and the anion, the electron distributions for all of the upper occupied orbitals are localized and similar to those appearing in 1-methylcytosine, 2'-deoxyribose, and methyl phosphate or $H_2PO_4^-$. In neutral 5'-dCMP, the first and second to fourth highest occupied molecular orbitals, with IP's of 8.7 and 9.4–9.7 eV, respectively, are orbitals located on the base. The fifth to eighth highest occupied orbitals, with IP's of 9.7, 10.2, 10.2–11.1, and 11.1 eV, respectively, are located on the sugar group. The ninth and tenth highest occupied orbitals, both of which have IP's in the range 11.2–11.3 eV, are associated with the phosphate group. In anionic 5'-dCMP, the first ionization potentials associated with the phosphate, base, and sugar groups occur at 4.6, 6.1, and 6.8 eV, respectively. The present description of orbital structure in anionic 5'-dCMP suggests that in electrophilic attack of nucleotides, product distributions of electrostatically controlled reactions will be different from product distributions of orbitally controlled reactions. A comparison of DNA and RNA alkylation patterns in methylation and ethylation reactions supports this conclusion.

Gas-phase UV photoelectron spectroscopy, when used in conjunction with results from molecular orbital (MO) calculations, has provided detailed, experimentally based pictures of valence orbital structures in DNA components such as nucleotide bases,[1-13] nucleosides,[14] and phosphate esters.[15,16] The primary goal of this investigation is to obtain valence electronic descriptions of neutral and anionic forms of the nucleotide 2'-deoxycytidine-5'-phosphate. The anionic form of this molecule, which contains 160 electrons, is the simplest among all of the common nucleotides occurring in DNA and RNA. Parts of the nucleotide have already been examined. Photoelectron spectra of eight methyl-substituted cytosines[7] and the cytidine analogue, 2',3',5'-tri-*O*-methylcytidine,[14] have been previously reported and assigned. The good agreement found in past studies,[1,3,7b,10,12] between photoelectron spectra and the energetic ordering and spacing of valence orbitals of nucleotide components predicted by SCF calculations, lends confidence to descriptions of electronic structures in neutral nucleotide components which this combined use of theory and experiment provides.

However, when this approach, which relies upon the direct comparison of SCF results with photoelectron spectra, is applied to intact nucleotides, it meets with difficulties. One problem is that nucleotides decompose at the pressures and temperatures required for gas-phase He I photoelectron measurements. Another problem is that in nucleotides there are a large number of orbitals with low ionization potentials (IP's). This abundance of low-energy orbitals gives rise to a loss of the spectroscopic resolution which makes photoelectron studies of smaller molecules so useful. To date, no valence orbital photoelectron spectra of nucleotides have been reported.

The present investigation employs a less direct approach which relies upon experimental and theoretical results obtained for small model compounds and anions. By using this approach, SCF results on neutral and anionic 2'-deoxycytidine-5'-phosphate (5'-dCMP and 5'-dCMP$^-$) have been corrected by examining model compounds and anions with orbital structures similar to those in the nucleotide. The SCF calculations on both forms of the nucleotide have been carried out with the 4-31G basis set. In most of the model studies, IP's predicted by 4-31G SCF calculations were compared with vertical and adiabatic IP's obtained from experiment. In some cases, results from SCF calculations were compared to results from more accurate theoretical treatments employing third-order Møller–Plesset perturbation theory (MP3).

(1) Padva, A.; LeBreton, P. R.; Dinerstein, R. J.; Ridyard, J. N. A. *Biochem. Biophys. Res. Commun.* **1974**, *60*, 1262.
(2) Lauer, G.; Schäfer, W.; Schweig, A. *Tetrahedron Lett.* **1975**, 3939.
(3) Padva, A.; O'Donnell, T. J.; LeBreton, P. R. *Chem. Phys. Lett.* **1976**, *41*, 278.
(4) Dougherty, D.; Wittel, K.; Meeks, J.; McGlynn, S. P. *J. Am. Chem. Soc.* **1976**, *98*, 3815.
(5) Peng, S.; Padva, A.; LeBreton, P. R. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2966.
(6) Dougherty, D.; McGlynn, S. P. *Chem. Phys.* **1977**, *67*, 1289.
(7) (a) Yu, C.; Peng, S.; Akiyama, I.; Lin, J.-H.; LeBreton, P. R. *J. Am. Chem. Soc.* **1978**, *100*, 2303. (b) Urano, S.; Yang, X.; LeBreton, P. R. *J. Mol. Struct.* In press.
(8) Padva, A.; Peng, S.; Lin, J.-H.; Shahbaz, M.; LeBreton, P. R. *Biopolymers* **1978**, *17*, 1523.
(9) Peng, S.; Lin, J.-H.; Shahbaz, M.; LeBreton, P. R. *Int. J. Quantum. Chem. Quantum. Biol. Symp.* **1978**, *5*, 301.
(10) O'Donnell, T. J.; LeBreton, P. R.; Shipman, L. L. *J. Phys. Chem.* **1978**, *82*, 343.
(11) Lin, J.-H.; Yu, C.; Peng, S.; Akiyama, I.; Li, K.; Lee, L.-K.; LeBreton, P. R. *J. Am. Chem. Soc.* **1980**, *102*, 4627.
(12) O'Donnell, T. J.; LeBreton, P. R.; Petke, J. D.; Shipman, L. L. *J. Phys. Chem.* **1980**, *84*, 1975.
(13) Lin, J.; Yu, C.; Peng, S.; Akiyama, I.; Li, K.; Lee, L.-K.; LeBreton, P. R. *J. Phys. Chem.* **1980**, *84*, 1006.
(14) Yu, C.; O'Donnell, T. J.; LeBreton, P. R. *J. Phys. Chem.* **1981**, *85*, 3851.
(15) (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.
(16) LeBreton, P. R.; Fetzer, S.; Tasaki, K.; Yang, X.; Yu, M.; Slutskaya, Z.; Urano, S. *J. Biomol. Str. Dyn.* **1988**, *6*, 199.

0002-7863/90/1512-538$02.50/0 

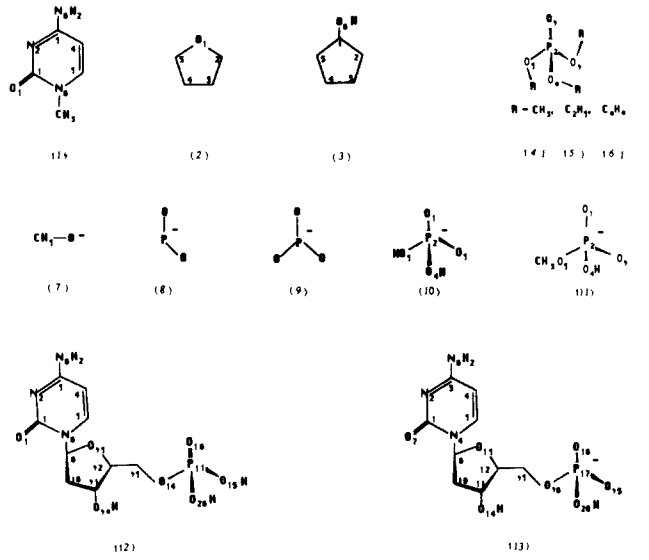

**Figure 1.** Structures of 1-methylcytosine (**1**), tetrahydrofuran (**2**), cyclopentanol (**3**), trimethyl (**4**), triethyl (**5**), and tributyl phosphate (**6**), $CH_3O^-$ (**7**), $PO_2^-$ (**8**), $PO_3^-$ (**9**), $H_2PO_4^-$ (**10**), $CH_3HPO_4^-$ (**11**), 5′-dCMP (**12**), and 5′-dCMP$^-$ (**13**).

1-Methylcytosine (**1**) was chosen as a model for the base group in 5′-dCMP and 5′-dCMP$^-$, because it occurs exclusively in the amino-oxo tautomeric form.[17] Tetrahydrofuran (**2**) and cyclopentanol (**3**) were used to model the oxygen atom lone-pair electrons of the deoxyribose group in 5′-dCMP and 5′-dCMP$^-$. The trialkylphosphate esters, trimethyl (**4**), triethyl (**5**), and tributyl phosphate (**6**), were employed as model compounds for the phosphate group in 5′-dCMP.

In this work, an evaluation of 4-31G SCF ionization potentials predicted for the negative phosphate group of 5′-dCMP$^-$ was carried out by comparison of 4-31G SCF results with experimental results for the small phosphorus- and oxygen-containing model anions $CH_3O^-$ (**7**), $PO_2^-$ (**8**), and $PO_3^-$ (**9**). The IP's obtained from SCF calculations for anions **7** to **9** and for $H_2PO_4^-$ (**10**) were also compared to theoretical IP's from MP3 calculations. Finally, calculations at the 4-31G SCF level were carried out on $CH_3HPO_4^-$ (**11**), which is a model anion for the phosphate group in 5′-dCMP$^-$. Figure 1 shows the structures of model compounds **1**–**6**, model anions **7**–**11**, and neutral (**12**) and anionic (**13**) 5′-dCMP.

**Photoemission Data for Model Compounds.** Of the 11 model compounds and anions considered here photoelectron spectra have previously been reported and interpreted for **1**,[7] **2**,[18-20] **3**,[21] **4**,[15,16] **5**,[15b] **7**,[22] **8**,[23a] and **9**.[23b-d] Assignments[7] for bands arising from the four highest occupied $\pi$ orbitals and the two highest occupied lone-pair orbitals of **1** were based on results from semiempirical HAM/3 calculations and IP perturbation patterns occurring in the spectra of eight different methyl-substituted cytosines.[7] For **2**, previous reports have yielded three values (9.57,[18] 9.71,[19] and 9.74[20] eV) for the vertical IP of the highest occupied oxygen atom lone-pair orbital. For **3**, only the adiabatic IP (9.72 eV) has been reported.[21]

**Table 1.** Assignments of Photoelectron Spectra of Trimethyl (**4**), Triethyl (**5**), and Tributyl Phosphate (**6**)

| taken from ref 16 and this work[a] | taken from ref 15b[b] | taken from ref 15a[a] |
|---|---|---|
| | Trimethyl Phosphate | |
| 10.81 (8e) | 10.90 (8e) | 10.82 (8e) |
| 11.40 (9a) | 11.45 ($2a_2$) | 11.36 (9a) |
| 11.93 (7e) | 12.05 (7e) | 11.9–12.1 (9e) |
| 12.8 (8a, 6e) | 12.7–13.0 ($7a_1$, 6e) | 12.4–12.9 (8a, 6e, 7a) |
| 14.40 (7a) | 14.49 ($6a_1$) | |
| | Triethyl phosphate | |
| 10.53 (11e) | 10.69 (11e) | |
| 10.98 (12a) | 11.10 ($3a_2$) | |
| 11.52 (10e) | 11.66 (10e) | |
| 12.1 (11a, 9e) | 12.1–12.4 ($9a_1$, 9e) | |
| 13.2 (10a) | 13.26 ($8a_1$) | |
| | Tributyl Phosphate | |
| 10.3 (17e) | | |
| 10.7–... (10a, 8e, 9a...) | | |

[a] Assignments based on $C_3$ symmetry. [b] Assignments based on $C_{3v}$ symmetry.

In early investigations[15] of the spectrum of **4**, there were discrepancies concerning the assignments. In the first study,[15a] a $C_3$ symmetry was assumed, and the spectrum was compared with that of trimethyl phosphite. In a later study,[15b] spectra of **4** and **5** were measured and interpreted with the aid of semiempirical CNDO/2 calculations. In the latter study, a $C_{3v}$ geometry was assumed for both molecules. Previous assignments of the spectra of **4** and **5** are given in Table I. According to the first assignment,[15a] the energy region between 12.4 and 12.9 eV contains the band arising from the 7a orbital ($6a_1$ in $C_{3v}$). However, in the assignment based on the CNDO/2 calculations this band lies between 14.4 and 14.5 eV.[15b]

Subsequent to the first two studies, the photoelectron spectrum of **4** was measured and interpreted by carrying out ab initio SCF calculations with the STO-3G, STO-3G*, and 4-31G basis sets, assuming $C_3$ symmetry.[16] All three ab initio calculations supported the assignment based on CNDO/2 calculations.[15b]

**Molecular Orbital Investigations of Nucleotides.** The approach employed in this examination of valence electrons in 5′-dCMP and 5′-dCMP$^-$ relies on a localized description of valence orbitals in nucleotides. Prior to this investigation, valence and core electron photoionization spectra for several single base containing fragments of B-DNA and Z-DNA were predicted from results of ab initio SCF calculations by using a single $\zeta$ basis set.[24] The results for the sequence S–P–S–C, consisting of two deoxyribose groups (S), a phosphate anion (P), and cytosine (C), also yielded a description containing localized valence orbitals. In the earlier study,[24] however, no experimental ionization potentials were provided in order to evaluate the theoretical results.

Previous theoretical descriptions of nucleotides,[25] bases,[26] and phosphate esters[27] and model compounds of nucleotide bases, such

(17) Szczesniak, M.; Kwiatkowski, J. S.; KuBulat, K.; Szczepaniak, K.; Person, W. B. *J. Am. Chem. Soc.* **1988**, *110*, 8319. The assignment of the photoelectron spectrum of cytosine in ref 7a assumed that in the gas phase only the amino-oxo tautomer occurs. Results of matrix isolation experiments indicate that in inert environments a significant fraction of cytosine is in the amino-hydroxy tautomeric form. However, for 1-methylcytosine (**1**) only the amino-oxo form occurs.

(18) Bain, A. D.; Bünzli, J. C.; Frost, D. C.; Weiler, L. *J. Am. Chem. Soc.* **1973**, *95*, 291.

(19) Gerson, S. H.; Worley, S. D.; Bodor, N.; Kaminski, J. J.; Flechtner, T. W. *J. Electron Spectrosc. Relat. Phenom.* **1978**, *13*, 421.

(20) Kimura, K.; Katsumata, S.; Achiba, Y.; Yamazaki, T.; Iwata, S. In *Handbook of Hel Photoelectron Spectra of Fundamental Organic Molecules*; Halsted Press: New York, 1981; pp 20, 46, 47, 49, 57, 105, 106, 188, 191, and 209.

(21) Traeger, J. C. *Org. Mass Spectrom.* **1985**, *20*, 223.

(22) Engelking, P. C.; Ellison, G. B.; Lineberger, W. C. *Chem. Phys.* **1978**, *69*, 1826.

(23) (a) Wu, R. L. C.; Tiernan, T. O. *Bull. Am. Phys. Soc.* **1982**, *27*, 109. (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*; Gole, J. L., Stwalley, W. C., Eds.; ACS Symposium Series 179; American Chemical Society: Washington, DC, 1982; p 609.

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

(25) (a) Pullman, A. In *Molecular Basis of Cancer Part A: Macromolecular Structure, Carcinogens, and Oncogen*; Rein, R., Ed.; Alan R. Liss, Inc.: New York, 1985; pp 55–69. (b) 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. (c) Pullman, A.; Pullman, B. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1980**, *7*, 245.

(26) (a) Pullman, A.; Armbruster, A. B. *Theor. Chim. Acta* **1977**, *45*, 249. (b) Bonnaccorsi, R.; Scrocco, E.; Tomasi, J.; Pullman, A. *Ibid.* **1975**, *36*, 339.

(27) (a) Pullman, A.; Berthod, H. *Chem. Phys. Lett.* **1976**, *41*, 205. (b) Perahia, D.; Pullman, A.; Berthod, H. *Theor. Chim. Acta* **1975**, *40*, 47.

as pyrimidine,[28] have been employed to determine how electronic structure influences interactions of DNA with electrophiles. Much of this work has been directed toward understanding reactions with alkylating agents, many of which are mutagens and carcinogens. Bulky alkylating agents,[25c,26b] derived from large aromatic molecules such as benzo[*a*]pyrene, benz[*a*]anthracene, and acetylaminofluorene as well as small alkylating agents[25c,26b,28] which lead to DNA methylation and ethylation, have been examined.

Two different models[25a,28] have evolved describing the manner in which nucleotide electronic structure influences alkylation reactions. One model[25a] has emphasized the role which the electrostatic potential plays in influencing reactivity. The other model[28] yields a more complicated picture of how electronic structure influences nucleotide reaction properties. According to the latter model, DNA alkylation patterns are determined by a complex interplay of electrostatic and orbital interactions. An important parameter in the second model is the distance between reactants in the transition state. For cases where the transition state occurs when the reactants are close together, short-range orbital interactions govern reaction properties; for cases where the transition state occurs while the reactants are widely separated, long-range electrostatic interactions play a more important role in influencing reaction properties. According to this model, the relative importance of orbital versus electrostatic influences on DNA alkylation reactions varies depending upon the alkylating agent. A secondary goal of the present investigation has been to obtain evidence concerning the manner in which electronic structure influences reactions of nucleotides with small alkylating agents that lead to DNA methylation and ethylation.

## Experimental Section

He I photoelectron spectra of molecules **1**–**6** were measured with a Perkin-Elmer PS-18 spectrometer. The spectrometer probe temperatures used in the experiments are given in Figures 2–4. All of the spectra were calibrated by using the $^2P_{3/2}$ and $^2P_{1/2}$ bands of argon and xenon. Samples of **1** and **2** were obtained from Vega-Fox Biochemicals and Fisher Scientific, respectively. Compounds **3**–**6** were purchased from Aldrich Chemical Co. Samples of **1**–**5** were used without further purification. Compound **6** was vacuum distilled (5 Torr at 73–75 °C). This removed small traces of water. For all compounds, spectra obtained from a single sample over a period of 3 h were identical, indicating that no decomposition occurred.

**Molecular Orbital Calculations and Geometries.** For 5′-dCMP (**12**) and 5′-dCMP$^-$ (**13**) and for the model compounds and anions studied here, IP's have been obtained from results of SCF calculations through the application of Koopmans' theorem.[29] The SCF calculations employed the 4-31G basis set.[30] Because of disk space constraints, this basis set is one of the largest which can be applied to **12** and **13** by using readily accessible supercomputers. Computations for **12** and **13** were carried out on Cray X-MP/48 computers. For the model compounds and anions, Cray X-MP/48 and IBM 3090-600E computers were employed. The GAUSSIAN86 program[31] was used for all calculations.

Molecular orbital diagrams were drawn from valence molecular orbital coefficients obtained from results of the 4-31G SCF calculations. The sizes of atomic orbitals appearing in the orbital diagrams are proportional to the coefficients of the 4-31G SCF orbital wave functions. In drawing orbital diagrams, the inner Gaussian terms of the 4-31G 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 of these atoms.[20] Only atomic orbitals with coefficients larger than 0.2 are shown in the diagrams. The positive and negative regions of the orbital wave functions are represented by shaded and open atomic orbitals.

**1-Methylcytosine (1).** The geometry of **1** employed in the calculations was obtained by combining crystallographic coordinates for cytosine[32] with those for the methyl group in 1-methylthymine.[33]

**Tetrahydrofuran (2) and Cyclopentanol (3).** The geometry of **2** was taken from gas-phase electron diffraction data,[34] which are consistent with conformations described by a freely pseudorotating molecule.[34,35] Calculations were carried out on the $C_s$ conformation.

The geometry of **3** was based on gas-phase electron diffraction data for cyclopentane and methanol. The geometries of the five-membered ring and the OH group of **3** were taken to be the same as those in cyclopentane[36] and methanol.[37] Calculations were carried out on the $C_s$ conformation. In the calculations, the hydrogen of the OH group is cis with regard to the methene hydrogen.

**Trimethyl (4), Triethyl (5), and Tributyl Phosphate (6).** In the geometry of **4**, P=O and P—O bond lengths were obtained from crystallographic data for tri-*p*-nitrophenyl phosphate.[38] Other bond lengths and bond angles were the same as those employed in earlier classical potential calculations on **4**.[39] The geometries of **5** and **6** are similar to that of **4**. In **4**–**6**, the dihedral angles ($\phi$) describing rotation about the P—O bonds in the O=P—O—C sequences were 60°.[40] The geometries of the alkyl groups of **5** and **6** were based on crystal data for magnesium diethyl phosphate.[41] The conformations around the O-$CH_2$ and $CH_2$-$CH_2$ bonds were assumed to be trans. The C-C-C bond angles of **6** were taken to be 112°, based on electron diffraction data for *n*-alkanes.[42] In all of the calculations a $C_3$ symmetry was assumed.

**$CH_3O^-$ (7), $PO_2^-$ (8), $PO_3^-$ (9), $H_2PO_4^-$ (10), and $CH_3HPO_4^-$ (11).** Geometries of the closed-shell anions, **7**–**9**, and of their ground-state neutral radicals were fully optimized by using energy gradients at the SCF level with the 6-31G* basis set.[43] The symmetries and the ground states of anions **7**–**9** and of the corresponding ground-state radicals are as follows: $CH_3O^-$ ($C_{3v}$, $^1A_1$), $CH_3O^\bullet$ ($C_{3v}$, $^2A_1$), $PO_2^-$ ($C_{2v}$, $^1A_1$), $PO_2^\bullet$ ($C_{2v}$, $^2A_1$), $PO_3^-$ ($D_{3h}$, $^1A_1'$), and $PO_3^\bullet$ ($D_{3h}$, $^2A_1'$).[44] Because **10** was employed as a model for the phosphate group in 5′-dCMP$^-$, for which the geometry was based on crystallographic data, the geometry of **10** ($C_1$, $^1A_1$) used in the calculations was also taken from crystallographic data. Here, data for (−)-ephedrine dihydrogen phosphate has been employed.[45] For **10** the bond lengths P—O and P=O were averaged over the reported values.[45] The O–H bond lengths and the P–O–H bond angles were 1.06 Å and 107.6°. The torsional angles $\phi$ (O=P—O—H) associated with rotation around the P–OH bonds were 60°. These bond lengths and bond angles resemble those occurring in similar compounds such as cyclohexylammonium phosphoenolpyruvate.[46]

The geometry of $CH_3HPO_4^-$ was based on the structure of the phosphate group in 5′-dCMP$^-$ (**13**).[47] All HCO bond angles were 109.5°.

Calculations of anion IP's were carried out by three different methods: (1) SCF calculations on the anions at the 4-31G level using Koopmans' theorem[29] (method I); (2) MP3 calculations[50] on both anions and radicals

(28) Ford, G. P.; Scribner, J. D. *J. Am. Chem. Soc.* **1983**, *105*, 349.
(29) Koopmans, T. *Physica* **1933**, *1*, 104.
(30) (a) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *Chem. Phys.* **1971**, *54*, 724. (b) Hehre, W. J.; Lathan, W. A. *Ibid.* **1972**, *56*, 5255.
(31) The GAUSSIAN 86 program was obtained from the following: Frisch, M. J.; Binkley, J. S.; Schlegel, H. B.; Raghavachari, K.; Melius, C. F.; Martin, R. L.; Stewart, J. J. P.; Bobrowicz, F. W.; Rohlfing, C. M.; Kahn, L. R.; Defrees, D. J.; Seeger, R.; Whiteside, R. A.; Fox, D. J.; Fleuder, E. M.; Pople, J. A. Carnegie-Mellon Quantum Chemistry Publishing Unit: Pittsburgh PA, 1984.
(32) Barker, D. L.; Marsh, R. E. *Acta Crystallogr.* **1964**, *17*, 1581.
(33) Hoogsteen, K. *Acta Crystallogr.* **1963**, *16*, 28.
(34) Geise, H. J.; Adams, W. J.; Bartell, L. S. *Tetrahedron* **1969**, *25*, 3045.
(35) Almenningen, A.; Seip, H. M.; Willadsen, T. *Acta Chem. Scand.* **1969**, *23*, 2748.
(36) Adams, W. J.; Geise, H. J.; Bartell, L. S. *J. Am. Chem. Soc.* **1970**, *17*, 5013.
(37) *Tables of Interatomic Distances and Configurations in Molecules and Ions*; The Chemical Society, Burlington House: London, 1965.
(38) Ul-Haque, M.; Caughlan, C. N. *Chem. Commun.* **1967**, 202.
(39) Khetrapal, C. L.; Govil, G.; Yeh, H. J. C. *J. Mol. Struc.* **1984**, *116*, 303.
(40) The sign of $\phi$ is positive if in the O—C—P=O sequence the movement of the O—C bond toward the P=O bond involves a right-handed screw motion.
(41) Ezra, F. S.; Collin, R. L. *Acta Crystallogr.* **1973**, *B29*, 1398.
(42) (a) Kuchitsu, K. *J. Chem. Soc. Jpn.* **1959**, *32*, 748. (b) Bartell, L. S.; Kohl, D. A. *Chem. Phys.* **1963**, *39*, 3097.
(43) (a) Hariharan, P. C.; Pople, J. A. *Chem. Phys. Lett.* **1972**, *66*, 217. (b) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. *Chem. Phys.* **1982**, *77*, 3654.
(44) Lohr, L. L.; Boehm, R. C. *J. Phys. Chem.* **1987**, *91*, 3203.
(45) Hearn, R. A.; Bugg, C. E. *Acta Crystallogr.* **1972**, *B28*, 3662.
(46) Watson, D. G.; Kennard, O. *Acta Crystallogr.* **1973**, *B29*, 2358.
(47) Arnott, S.; Hukins, D. W. L. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1505.
(48) (a) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294. (b) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *Chem. Phys.* **1984**, *80*, 3265. Also, see: 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.
(49) This is the 6-31G* basis set augmented by diffuse functions for non-hydrogen atoms. See ref 42 and 48.
(50) Pople, J. A.; Seeger, R.; Krishnan, R. *Int. J. Quantum. Chem.* **1977**, *11*, 149.

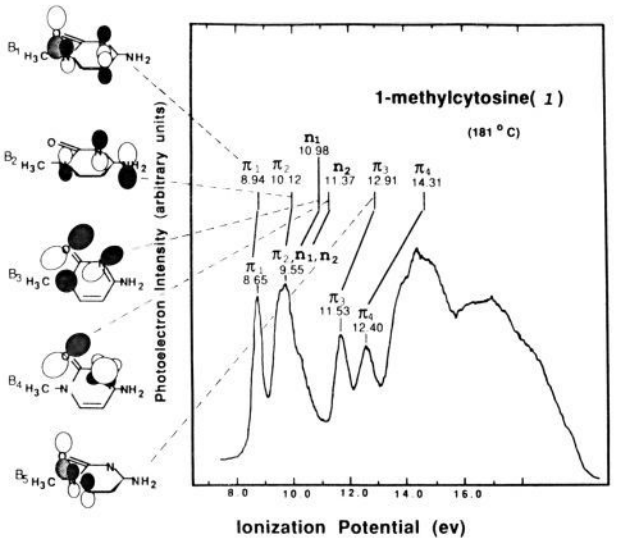


**Figure 2.** Photoelectron spectrum of 1-methylcytosine (**1**). Theoretical ionization potentials and orbital diagrams obtained from 4-31G SCF calculations. Orbital diagrams illustrate major atomic orbital contributions to the five highest occupied molecular orbitals ($B_1$–$B_5$). In $\pi$ and lone-pair orbital diagrams the viewing angle is different. For criteria used to construct the orbital diagrams, see text.

at the 6-31+G*[48,49] level, with use of the optimized geometries for anions and the same geometries for the corresponding radicals (method II); (3) the same level of calculation as method II, however, optimized geometries of both the anion and the radicals were used (method III). In the MP3 calculations, the core electrons were frozen, and the nonsinglet states were described by using the unrestricted Hartree–Fock formalism.[51]

**5′-dCMP (12) and 5′-dCMP⁻ (13).** The geometry of the phosphate group of 5′-dCMP (**12**) was based on that for **4**.[16] The OH bond lengths and H–O–P bond angles are 1.06 Å and 107.6°. The rest of the geometry was based on optimized crystallographic parameters for 5′-dCMP⁻ (**13**) in B-DNA.[47]

The geometry of **13** was obtained by combining the geometry of cytosine, based on crystallographic data for cytidine,[52] with geometries of the sugar and phosphate groups taken from the optimized geometry of **13** in B-DNA.[47] Bond lengths and bond angles involving the hydrogen atom were as follows: C–H = 1.11 Å, O–H = 0.95 Å, HCH = 109.5°, and HOC = 105.0°.[37]

## Results

**1-Methylcytosine (1).** Figure 2 shows the photoelectron spectrum of **1** along with assignments obtained from 4-31G SCF calculations. In Figure 2, theoretical IP's obtained from the 4-31G calculations are given above the experimental vertical IP's. The figure also contains orbital diagrams for the six highest occupied orbitals as predicted from the 4-31G results. Table II contains atomic charges for **1** and for the model compounds, tetrahydrofuran (**2**), cyclopentanol (**3**), and trimethyl phosphate (**4**). The results in Table II were obtained from a Mulliken population analysis[53] of the 4-31G SCF results. The assignment of the spectrum of **1**, given in Figure 2, agrees with that most recently reported.[7b] A comparison of the theoretical and experimental results indicates that the first IP of **1** is accurately calculated to within 0.29 eV and that the theoretical IP's for the $\pi_2$, $\pi_3$, and $\pi_4$ orbitals differ from the experimental values by 0.57, 1.38, and 1.91 eV, respectively. For the two highest occupied lone-pair orbitals ($n_1$ and $n_2$), which have large contributions from the oxygen atom 2p atomic orbitals, the theoretical IP's (10.98 and 11.37 eV) are 1.4–1.8 eV higher than the experimental values. The results in Figure 2 are consistent with the observation that, in general, 4-31G SCF calculations predict vertical IP's for the highest occupied $\pi$ orbitals which agree with experiment to within 0.2–0.5 eV but yield IP's for uppermost occupied oxygen atom lone-pair orbitals which are more than 0.8 eV larger than the experimental values.[20]

**Table II.** Atomic Charges for 1-Methylcytosine, Tetrahydrofuran, Cyclopentanol, Trimethyl Phosphate, and $H_2PO_4^-$ [a]

| 1-methylcytosine (**1**) | | tetrahydrofuran (**2**) | | cyclopentanol (**3**) | |
|---|---|---|---|---|---|
| $C_1$ | 1.02 | $O_1$ | −0.69 | $C_1$ | 0.14 |
| $N_2$ | −0.71 | $C_2$ | 0.02 | $C_2$ | −0.30 |
| $C_3$ | 0.63 | $C_3$ | −0.34 | $C_3$ | −0.32 |
| $C_4$ | −0.22 | $C_4$ | −0.34 | $C_4$ | −0.32 |
| $C_5$ | 0.31 | $C_5$ | 0.02 | $C_5$ | −0.30 |
| $N_6$ | −0.98 | | | $O_6$ | −0.74 |
| $O_7$ | −0.67 | | | | |
| $N_8$ | −0.83 | | | | |
| $C_9$ | −0.16 | | | | |
| trimethyl phosphate (**4**) | | $H_2PO_4^-$ (**10**) | | $CH_3HPO_4^-$ (**11**) | |
| $O_1$ | −1.01 | $O_1$ | −1.08 | $O_1$ | −1.08 |
| $P_2$ | 2.44 | $P_2$ | 2.23 | $P_2$ | 2.26 |
| $O_3$ | −0.96 | $O_3$ | −1.08 | $O_3$ | −1.08 |
| $O_4$ | −0.96 | $O_4$ | −0.98 | $O_4$ | −0.97 |
| $O_5$ | −0.96 | $O_5$ | −0.94 | $O_5$ | −0.95 |
| C | −0.11 | | | C | −0.07 |

[a] Obtained from 4-31G SCF calculations. The numbering of atoms is shown in Figure 1.

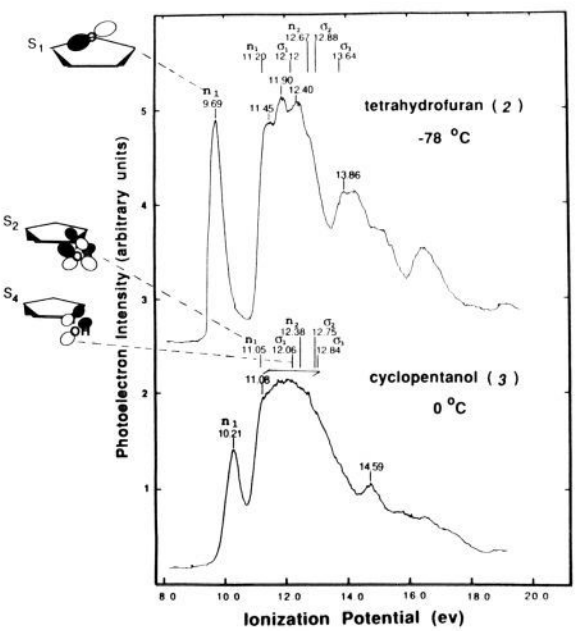


**Figure 3.** Photoelectron spectrum of tetrahydrofuran (**2**) and cyclopentanol (**3**) and theoretical ionization potentials and orbital diagrams obtained from 4-31G SCF calculations. Orbital diagrams are given for the HOMO of tetrahydrofuran ($S_1$) and for the HOMO and NHOMO of cyclopentanol ($S_2$ and $S_4$).

**Tetrahydrofuran (2) and Cyclopentanol (3).** Figure 3 shows spectra of **2** and **3** together with IP's and assignments for bands originating from the upper occupied orbitals. The value of the first ionization potential reported here for molecule **2** (9.69 eV) is closer to the most recent values (9.71[19] and 9.74 eV[20]) than to that originally reported (9.57 eV).[18] Other vertical IP's given in Figure 3 agree with previous results[20] to within 0.24 eV. For molecule **3**, the first vertical ionization potential is 0.49 eV higher than the reported adiabatic ionization potential.[21] In Figure 3 theoretical IP's obtained from the 4-31G SCF calculations are given above the experimental values. The figure also contains orbital diagrams showing the highest occupied molecular orbitals (HOMOs) of **2** and **3**. In both molecules the first IP's arise from lone-pair orbitals ($n_1$) associated with O atoms. For **2** and **3**, the

(51) Pople, J. A.; Nesbet, R. K. *Chem. Phys.* **1954**, *22*, 571.
(52) Furberg, S. *Acta Crystallogr.* **1965**, *18*, 313.
(53) Mulliken, R. S. *Chem. Phys.* **1955**, *23*, 1833.

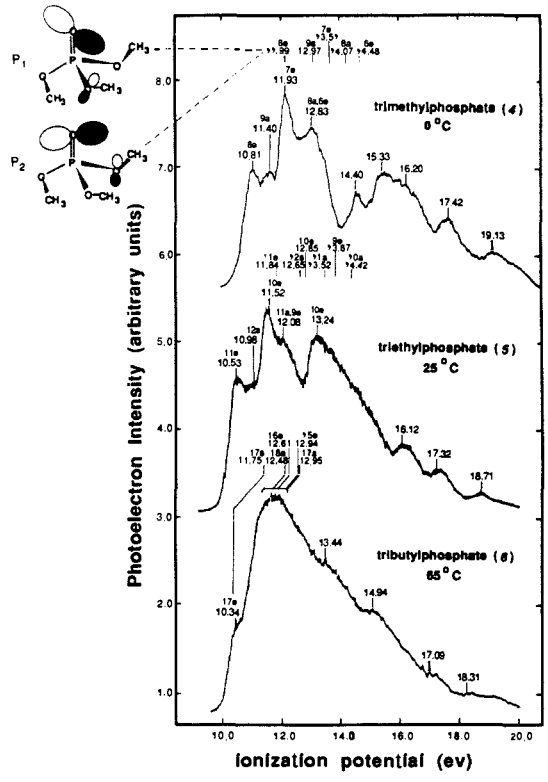


**Figure 4.** Photoelectron spectra of trimethyl (**4**), triethyl (**5**), and tributyl phosphate (**6**) and theoretical ionization potentials and orbital diagrams obtained from 4-31G SCF calculations. Orbital diagrams are given for the highest occupied molecular orbitals ($P_1$ and $P_2$) in **4**.

current assignments agree with previous assignments.[18-20]

The spectra of molecules **2** and **3**, which indicate that in each case the first band is well resolved, whereas the rest of the spectrum is poorly resolved, agree with the calculations. According to the 4-31G SCF results, the differences between the IP's of the HOMO and the next highest occupied orbital (NHOMO) in molecules **2** and **3** are 0.92 and 1.01 eV, respectively. Experimentally, these differences are 1.76 and approximately 0.9 eV. The 4-31G SCF calculations predict that the differences between the IP's of the NHOMO and the third highest occupied orbitals are 0.55 and 0.32 eV for **2** and **3**. Experimentally, these differences are 0.45 eV and less than 0.5 eV.

As in the case of 1-methylcytosine, the 4-31G SCF calculations yield ionization potentials for the O atom lone-pair orbitals ($n_1$) which are too large. The theoretical IP's for the $n_1$ orbitals in **2** and **3** are 1.51 and 0.84 eV higher than the experimental IP's.

**Trimethyl (4), Triethyl (5), and Tributyl Phosphate (6).** Figure 4 shows the spectra of molecules **4**, **5**, and **6** along with vertical ionization potentials and assignments of bands arising from the upper occupied orbitals. In Figure 4, theoretical IP's are given above each spectrum. The figure also contains orbital diagrams of the HOMOs of molecule **4**. According to the orbital diagrams, the HOMOs in **4** are O atom lone-pair orbitals. Results of the 4-31G calculations indicate that the orbital diagrams of the HOMOs are similar in molecules **4**, **5**, and **6**.

The experimental first IP's of **4**, **5**, and **6** (10.81, 10.53, and 10.34 eV, respectively) decrease as the size of the alkyl group increases. The results of 4-31G SCF calculations, which yield IP's of 11.99, 11.84, and 11.75 eV, agree with this trend.

The resolution of the spectra decreases as the size of the alkyl groups increases. The spectra of **4** and **5** exhibit five maxima in the energy region from 10.0 to 15.0 eV. In the same energy region, the spectrum of **6** exhibits only a broad maximum with poorly resolved shoulders. The 4-31G SCF results are consistent with this decrease in resolution. For example, the differences between the theoretical IP's of the HOMO and the NHOMO of **4**, **5**, and **6** decrease in the order 0.98, 0.81, and 0.73 eV, respectively. Similarly, the theoretical differences between the HOMO and the third highest occupied orbital in molecules **4**, **5**, and **6** decrease in the order 1.52, 1.01, and 0.86 eV.

**Table III.** Optimized Geometries of Anions and Neutral Radicals[a,b]

| | $CH_3O^-$ | $CH_3O^\bullet$ |
|---|---|---|
| C-O | 1.393 | 1.652 |
| C-H | 1.124 | 1.078 |
| ∠HCO | 114.00 | 102.00 |

| | $PO_2^-$ | $PO_2^\bullet$ | $PO_3^-$ | $PO_3^\bullet$ |
|---|---|---|---|---|
| P-O | 1.483 | 1.445 | 1.467 | 1.455 |
| ∠OPO | 119.02 | 134.59 | 120.00 | 120.00 |

[a] Numbering of atoms is given in Figure 1. For anions and radicals, geometries were optimized at the 6-31G* SCF level. [b] Bond lengths in Å. Angles in deg.

**Table IV.** Theoretical and Experimental Ionization Potentials of Model Anions[a]

| anions | method I[b] | method II[c] | method III[d] | exptl[e] |
|---|---|---|---|---|
| $CH_3O^-$ (**7**) | 1.68 | 1.40 | 1.45 | 1.57[f] |
| $PO_2^-$ (**8**) | 3.20 | 3.20 | 3.40 | 3.3 ± 0.2[g] |
| $PO_3^-$ (**9**) | 6.60 | 5.03 | 5.13 | 4.9 ± 1.3[h] |
| $H_2PO_4^-$ (**10**) | 5.31 | 4.87 | | |
| $CH_3HPO_4^-$ (**11**) | 5.59 | | | |

[a] All ionization potentials in eV. [b] Based on 4-31G SCF calculations on the anion. [c] Based on MP3/6-31+G* calculations on the anion and on the radical, with use of the optimized anion geometry for both anion and radical. [d] Based on MP3/6-31+G* calculations on the anion and radical, with use of the optimized anion and the optimized radical geometries. Method III has not been applied to $H_2PO_4^-$ because of the excessive amount of time required to carry out the calculation. The estimated CPU time is more than 10 h on an IBM 3090-600 VF computer. [e] Adiabatic ionization potentials. [f] Reference 22. [g] Reference 23a. [h] Estimated from heats of formation of $PO_3$ and $PO_3^-$. See ref 23b-d and 44.

In the spectra of molecules **4** and **5**, the energy region from 10.5 to 13.0 eV contains five bands, arising from the 8e, 9a, 7e, 8a, and 6e orbitals of molecule **4** and the 11e, 12a, 10e, 11a, and 9e orbitals of molecule **5**. The assignments for the two spectra are consistent with their similar appearance. The assignments of the spectra of **4** and **5**, given in the figure, are the same as those reported in other recent investigations.[15,16]

As in the case of molecules **1-3**, the 4-31G SCF calculations on **4**, **5**, and **6** predict IP's for the O atom lone-pair orbitals, which are too large. The theoretical values of the IP's for the HOMOs in **4**, **5**, and **6** are 1.18, 1.31, and 1.41 eV higher than the experimental values.

**$CH_3O^-$ (7), $PO_2^-$ (8), $PO_3^-$ (9), $H_2PO_4^-$ (10), and $CH_3HPO_4^-$ (11).** Table II contains atomic charges obtained for **10** and **11** from Mulliken population analyses of results from calculations employing method I. Table III lists geometric parameters for $CH_3O^-$, $CH_3O^\bullet$, $PO_2^-$, $PO_2^\bullet$, $PO_3^-$, and $PO_3^\bullet$ obtained from SCF optimization calculations. For $PO_2^\bullet$, experimental values for the O-P-O bond angle (135.3 ± 0.8°) and the P-O bond length (1.4665 ± 0.0041 Å) obtained from microwave spectra[54] agree well with the calculated bond angle (134.59°) and bond length (1.445 Å).

Table IV gives theoretical IP's for the model anions (**7-11**). For anions **7-9**, Table IV contains IP's calculated via methods I, II, and III. For **10** the table contains results from methods I and II. For **11** the table only includes a result from the 4-31G SCF calculations (method I). Table IV also shows experimental adiabatic IP's for anions **7-9**.[22,23] For **10** and **11** we found no reported experimental data.

A comparison of the results in Table IV for $CH_3O^-$ (**7**), $PO_2^-$ (**8**), and $PO_3^-$ (**9**) indicates that ionization potentials obtained from MP3 calculations, where only the anion geometry was optimized (method II), are similar to ionization potentials obtained when both the anion and the neutral radical geometries were optimized (method III). For anions **7** and **8**, both methods II and III yield

(54) Kawaguchi, K.; Saito, S.; Hirota, E. *Chem. Phys.* **1985**, *82*, 4893.

**Table V.** Atomic Charges in 5′-dCMP and 5′-dCMP$^{-}$ [a,b]

| | cytosine | | 2′-deoxyribose | | phosphate | |
|---|---|---|---|---|---|---|
| | | | 5′-dCMP (**12**) | | | |
| | $C_1$ | 1.03 | $C_9$ | 0.43 | $O_{16}$ | −1.01 |
| | $N_2$ | −0.69 | $C_{10}$ | −0.33 | $P_{17}$ | 2.34 |
| | $C_3$ | 0.61 | $C_{11}$ | 0.06 | $O_{18}$ | −0.99 |
| | $C_4$ | −0.28 | $C_{12}$ | 0.12 | $O_{19}$ | −0.93 |
| | $C_5$ | 0.31 | $O_{13}$ | −0.67 | $O_{20}$ | −0.93 |
| | $N_6$ | −1.01 | $O_{14}$ | −0.71 | | |
| | $O_7$ | −0.67 | $C_{15}$ | 0.05 | | |
| | $N_8$ | −0.82 | | | | |
| total | | −0.37 | | 0.90 | | −0.53 |
| | | | 5′-dCMP$^{-}$ (**13**) | | | |
| | $C_1$ | 1.03 | $C_9$ | 0.44 | $O_{16}$ | −1.01 |
| | $N_2$ | −0.72 | $C_{10}$ | −0.34 | $P_{17}$ | 2.27 |
| | $C_3$ | 0.61 | $C_{11}$ | 0.06 | $O_{18}$ | −1.07 |
| | $C_4$ | −0.30 | $C_{12}$ | 0.13 | $O_{19}$ | −1.08 |
| | $C_5$ | 0.32 | $O_{13}$ | −0.68 | $O_{20}$ | −0.96 |
| | $N_6$ | −1.01 | $O_{14}$ | −0.73 | | |
| | $O_7$ | −0.70 | $C_{15}$ | 0.09 | | |
| | $N_8$ | −0.83 | | | | |
| total | | −0.39 | | 0.79 | | −1.40 |

[a] Obtained from 4-31G SCF calculations. [b] The numbering of atoms is described in Figure 1.

IP's which agree with the experimental adiabatic IP's to within 0.2 eV. For **9**, where the experimental uncertainty is 1.3 eV, methods II and III yield values which differ by only 0.1 eV. These observations provide evidence that the IP (4.87 eV) predicted by method II for $H_2PO_4^-$ (**10**) has accuracy in the range ±0.3 eV.[55]

The results in Table IV also demonstrate that 4-31G SCF calculations (method I) yield ionization potentials for anions **7** and **8** which agree with the experimental adiabatic ionization potentials to within 0.11 eV and with the values obtained by using methods II and III to within 0.3 eV. However, method I predicts an IP for **9** which is 1.7 eV higher than the experimental value and more than 1.4 eV higher than the values predicted by methods II and III. Method I also predicts an IP for **10** which is 0.44 eV higher than that obtained via method II. The results in Table IV generally indicate that as the ionization potentials of oxygen- and phosphorus-containing anions increase, the accuracy of the 4-31G SCF results decreases.

**5′-dCMP (12).** Figure 5 gives IP's and orbital diagrams describing the 11 highest occupied orbitals in 5′-dCMP (**12**) as predicted by 4-31G SCF calculations. Table V contains atomic charges in **12** obtained from Mulliken population analyses of the 4-31G SCF results.

In Figure 5, all of the 11 highest occupied orbitals in **12** are localized on either the base, sugar, or phosphate group. For example, in the HOMO, designated $B_1$, only base atoms have orbital coefficients larger than 0.20. The next three highest occupied orbitals ($B_2$ to $B_4$) and the seventh orbital ($B_5$) are also on the base. The fifth ($S_1$), sixth ($S_2$), tenth ($S_3$), and eleventh ($S_4$) highest occupied orbitals are on the sugar. The eighth ($P_1$) and ninth ($P_2$) orbitals are associated with the phosphate group. A localized orbital description breaks down in the eighteenth highest occupied orbital in **12**. This orbital has substantial contributions from both the sugar and phosphate groups.

Evidence for the localized valence electronic structure in **12**, indicated by the orbital diagrams in Figure 5, is also provided by Mulliken population analyses of the 4-31G SCF results for 1-methylcytosine (**1**), trimethyl phosphate (**4**), and **12**. According to the Mulliken population analyses, summarized in Tables II and V, the atomic charges associated with **1** and **4** are similar to those occurring when the base and the phosphate groups are incorporated into **12**.

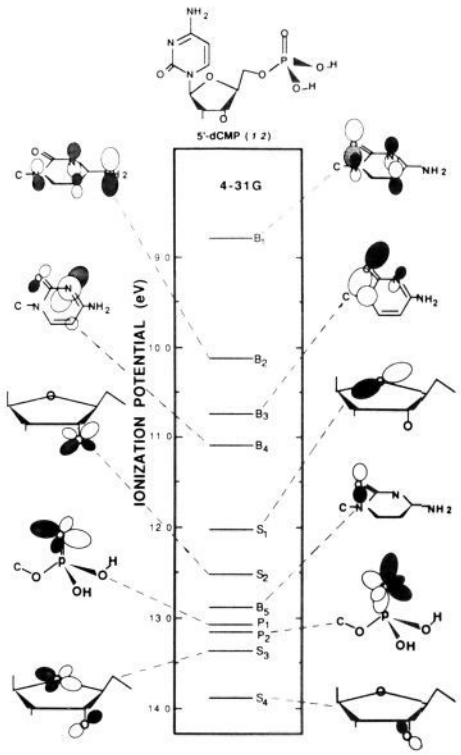


**Figure 5.** Valence orbital IP's and orbital diagrams for 5′-dCMP (**12**) obtained from 4-31G SCF calculations. Orbitals located primarily on the base, sugar, and phosphate groups are denoted B, S, and P, respectively.

The 4-31G SCF results also demonstrate that the electron distributions in almost all of the upper occupied orbitals are similar to orbitals appearing in the model compounds. For example, a comparison of the diagrams in Figures 1 and 5 indicates that the five highest occupied base orbitals in the nucleotide ($B_1$–$B_5$) correspond to orbitals occurring in **1**. While small differences occur, they are due largely to the cutoff criteria used in drawing the diagrams.[54]

In addition, base IP's for **12** obtained from 4-31G SCF calculations are similar to those for **1**. In **12**, the calculated IP's of the $B_1$–$B_5$ orbitals are 8.80, 10.15, 10.73, 11.09, and 12.88 eV, respectively. In **1**, the corresponding calculated IP's are 8.94, 10.12, 10.98, 11.37, and 12.91 eV.

While the correlation between the orbital diagrams for **12** and for the model compounds also extends to the sugar and phosphate groups, the correspondence for the sugar and phosphate orbitals is less strict than for the base orbitals. For example, while the $S_1$, $S_2$, and $S_4$ orbitals in **12** are also found in tetrahydrofuran (**2**) and cyclopentanol (**3**), the $S_3$ orbital, which is formed from a mixing of 2p orbitals on the $O_{13}$ and $O_{14}$ atoms in **12**, does not occur in the model compounds. Differences also occur between the $S_2$ and $S_4$ orbital diagrams in **3** and **12**. However, when effects associated with the cutoff criteria are accounted for, the orbital diagrams of the $S_2$ and $S_4$ orbitals in **12** are similar to those in **3**. Another apparent difference between the lone-pair orbitals in **2** or **3** and those in **12** involves the calculated ionization potentials. In **12**, the IP's of the $S_1$, $S_2$, and $S_4$ orbitals are predicted to be 0.83, 1.46, and 1.81 eV larger than the IP's predicted for the corresponding orbitals in **2** and **3**.

Apparent differences also occur between the upper occupied phosphate orbitals in 5′-dCMP (**12**) and those in the model compound trimethyl phosphate (**4**). The $P_1$ and $P_2$ orbitals in **12** have different diagrams than the $P_1$ and $P_2$ orbitals in **4**. Here, the observed discrepancies between the orbitals occurring in the nucleotide and in the model compound are due to differences in conformation and in the number of ester bonds. This is demon-

(55) The results of the post SCF calculations qualitatively support a description of electron detachment from negative phosphorus- and oxygen-containing anions which is based on Koopmans' theorem. For example, the MP3 calculations indicate that for $PO_2^-$, $PO_3^-$, and $H_2PO_4^-$ and for their corresponding ground-state open shell radicals the reference wave functions make up more than 90% of the total wave functions. Other investigations support the qualitative description obtained by applying Koopmans' theorem to the low-energy IP's of nucleotide bases. See ref 12.

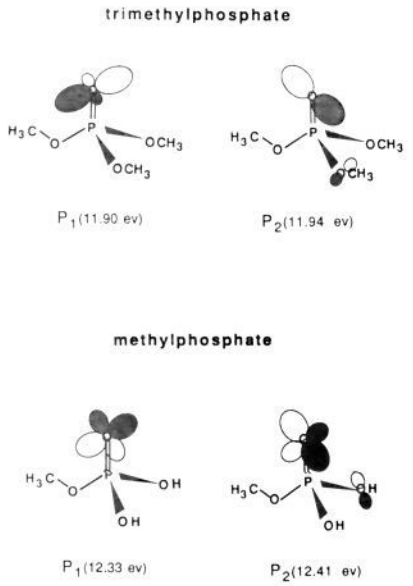


**Figure 6**. Orbital diagrams and ionization potentials for methyl phosphate and trimethyl phosphate (**4**) obtained from 4-31G SCF calculations. For **4**, a $C_s$ geometry similar to that of the phosphate group in 5′-dCMP (**12**) was employed. The dihedral angle $\phi_1$ [O—C—P═O] around one of the P—O bonds was 60°. The other two dihedral angles, $\phi_2$ and $\phi_3$, were 53.1°. See note 40. In all other respects, the geometry was the same as that used in Figure 4. For methyl phosphate, the geometry of all heavy atoms and most H atoms was the same as that of the phosphate group in **12**. The C–H bond lengths of the methyl group were taken to be 1.1 Å.

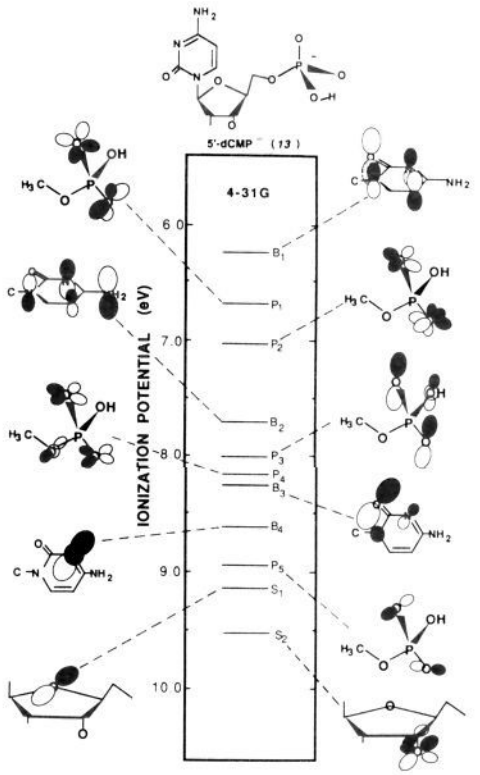


**Figure 7**. Valence orbital ionization potentials and orbital diagrams for 5′-dCMP⁻ (**13**) obtained from 4-31G SCF calculations.

strated in Figure 6 which contains orbital diagrams and IP's obtained from 4-31G SCF calculations for the $P_1$ and $P_2$ orbitals in **4** when the conformation is modified to become more similar to that of the phosphate group in **12**. Figure 6 also contains orbital diagrams for methyl phosphate in a conformation similar to that occurring in **12**. The results in Figure 6 indicate that diagrams for the $P_1$ and $P_2$ orbitals in methyl phosphate are very similar to those in **12**. The orbital diagrams for **4** in Figure 6 have characteristics which are between the orbital diagrams for methyl phosphate in the same figure and for **4** in Figure 3.

The results in Figures 4 and 6 also point out that the changes in the ionization potentials of the $P_1$ and $P_2$ orbitals associated with changes in conformation and with changes in the number of ester bonds are less than 0.5 eV. In **4**, the IP's of the $P_1$ and $P_2$ orbitals in Figure 4 are predicted to be 0.34 and 0.42 eV lower than the IP's of the corresponding orbitals in methyl phosphate.

While the results in Figures 5 and 6 point out that diagrams for the $P_1$ and $P_2$ orbitals in **12** and in methyl phosphate are similar, the 4-31G calculations predict that the ionization potentials of these orbitals in **12** are significantly higher than in methyl phosphate. The predicted IP's for the $P_1$ and $P_2$ orbitals (13.07 and 13.16 eV) in **12** are 0.74 and 0.75 eV higher than the values for methyl phosphate.

**5′-dCMP⁻ (13).** Figure 7 shows ionization potentials and orbital diagrams for the eleven highest occupied orbitals in 5′-dCMP⁻ (**13**) as predicted by 4-31G SCF calculations. Table V contains atomic charges obtained from Mulliken population analyses of the 4-31G SCF results.

The population analysis indicates that in **13**, the negative charge of the anion is located on the phosphate group (−1.4 e). A comparison of atomic charges in **12** and **13** indicates that in both molecules the charge on the base is nearly the same. In **13**, the base is more negative than in **12** but by only 0.02 e. The total atomic charge on the sugar group in **13** differs from that in **12** by 0.11 e. The similarities between **12** and **13** with regard to the atomic charges on the base and sugar groups suggest that the orbital structures of these groups are alike in 5′-dCMP and 5′-dCMP⁻.

This correspondence is demonstrated by the orbital diagrams in Figure 7. The 4-31G SCF calculations predict that for 5′-dCMP⁻ all 12 of the highest occupied orbitals are localized on either the base, sugar, or phosphate groups. According to 4-31G SCF calculations, the first ($B_1$), fourth ($B_2$), seventh ($B_3$), and eighth ($B_4$) highest occupied orbitals are on the base; the second ($P_1$), third ($P_2$), fifth ($P_3$), sixth ($P_4$), and ninth ($P_5$) highest occupied orbitals arise from the phosphate group; and the tenth ($S_1$) and eleventh ($S_2$) orbitals are on the sugar group.

The results in Figure 7 demonstrate that the upper occupied base ($B_1$ to $B_4$) and sugar ($S_1$ and $S_2$) orbitals in 5′-dCMP⁻ (**13**) have electron distributions like those of corresponding orbitals in 5′-dCMP (**12**). A comparison of Figures 5 and 7 indicates that, except for the $B_4$ orbital, the diagrams for each pair of corresponding orbitals in **12** and **13** are nearly identical. For the $B_4$ orbitals in **12** and **13**, small apparent differences occurring in the orbital diagrams are exaggerated by the cutoff criterion.

A comparison of the results of Mulliken population analyses for the model anions, $H_2PO_4^-$ (**10**) and $CH_3HPO_4^-$ (**11**), and for **13** in Tables II and V indicates that the electronic structures of the model anions are like that of the phosphate group in **13**.

The nearly equal atomic charges occurring in **10**, **11**, and **13** are reflected in the orbital diagrams. This is demonstrated by Figure 8 which contains diagrams of the $P_1$ to $P_5$ orbitals in **10** and **11**. The diagrams in Figure 8 were obtained by employing the same geometries used in the calculations for Table IV. Like the upper occupied phosphate orbitals in **13**, the $P_1$–$P_5$ orbitals in **10** and **11** arise primarily from O atom 2p orbitals. The results in Figures 7 and 8 point out that the mixing of the $P_1$–$P_5$ orbitals is very similar in **11** and **13**. In **10**, the mixing of the $P_1$–$P_5$ orbitals is slightly different. This is due to differences in phosphate conformation and in ester linkage.

Finally, it is interesting to observe that, although orbital diagrams for the $P_1$–$P_5$ orbitals in **10** are different than those in **11**, the ionization potentials predicted by 4-31G SCF calculations for corresponding orbitals are almost equal. In **11**, each of the $P_1$–$P_5$ orbitals is predicted to have an IP which is only 0.10–0.27 eV lower than that of the corresponding orbital in **10**. On the other hand, differences between the calculated IP's for the $P_1$–$P_5$ orbitals in **11** and **13** are much larger. In **11**, each of the phosphate orbitals

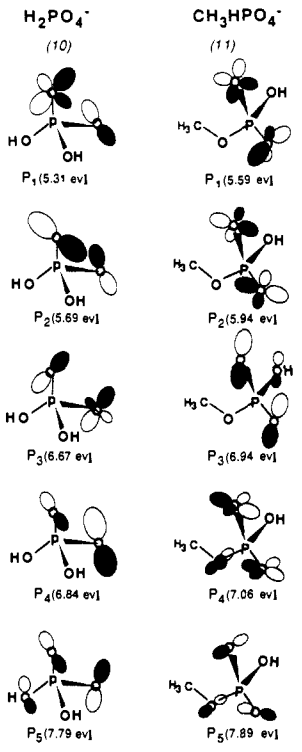


**Figure 8.** Orbital diagrams and ionization potentials for $H_2PO_4^-$ (**10**) and $CH_3HPO_4^-$ (**11**) obtained from 4-31G SCF calculations.

is predicted to have an ionization potential which is more than 1.05 eV smaller than that of the corresponding orbital in **13**.

## Discussion

**Valence Orbital Ionization Potentials in 5'-dCMP (12). Base Orbitals.** According to the results in Figures 2 and 5 and Tables II and V, electron distributions in the base orbitals ($B_1$–$B_5$) of 1-methylcytosine (**1**) correspond closely to those of the upper occupied base orbitals in 5'-dCMP (**12**). This observation points out that the correct IP's of the base orbitals in **12** are nearly equal to the IP's of **1**.

The combined use of photoelectron and computational results for the model compounds and for **12** has provided a set of estimated ionization potentials for **12**. These are listed in Table VI along with IP's obtained from 4-31G SCF calculations for **12**, **1**, tetrahydrofuran (**2**), cyclopentanol (**3**), and trimethyl phosphate (**4**). Table VI also lists experimental vertical IP's of valence orbitals in molecules **1**–**4**.

**Sugar Orbitals.** The orbital diagrams and the results of population analyses indicate that the S1, S2, and S4 orbitals in **12** are similar to the corresponding orbitals in **2** and **3**. However, according to 4-31G SCF calculations, the IP's for $S_1$, $S_2$, and $S_4$ orbitals in **12** (12.03, 12.51, and 13.87 eV) are significantly higher than the IP's (11.20, 11.05, and 12.06 eV) predicted for the corresponding orbitals in **2** and **3**. The high ionization potentials calculated for the $S_1$, $S_2$, and $S_4$ orbitals in **12** compared to the values in **2** and **3** are an artifact of the 4-31G SCF calculations. This artifact is related to the observation that the $S_1$ and $S_2$ orbitals in **2** and **3** are the HOMOs, and the $S_4$ orbital in **3** is the NHOMO, while the $S_1$, $S_2$, and $S_4$ orbitals in **12** are more deeply buried in the valence manifold. In **12**, these orbitals are the fifth, sixth, and eleventh highest occupied orbitals. This dependence of predicted IP's obtained from 4-31G calculations, on the energetic ordering of orbitals within the valence manifold, is commonly observed. For example, the experimental ionization potentials for the O atom lone-pair orbitals in ethanol and benzyl alcohol differ by only 0.03 eV.[20,56] However, 4-31G calculations predict that the IP for the lone-pair orbital in ethanol, which is the HOMO, is 0.61 eV smaller than that of the lone-pair orbital in benzyl alcohol, which is the fourth highest occupied orbital. A similar trend is observed when the experimental lone-pair IP's for acetic acid[57] and benzene acetic acid[58] are compared with the results from 4-31G calculations.

**Table VI.** Theoretical and Estimated Ionization Potentials of Neutral 5'-dCMP[a]

| Base Orbitals | | | | | |
|---|---|---|---|---|---|
| | | 1-methylcytosine (**1**) | | 5'-dCMP (**12**) | |
| orbital | orbital type | $IP_{calc}$[b] | $IP_{exp}$[c] | $IP_{calc}$[b] | $IP_{est}$[d] |
| $B_1$ | $\pi$ | 8.94 | 8.65 | 8.80 | 8.7 |
| $B_2$ | $\pi$ | 10.12 | 9.40–9.70 | 10.15 | 9.4–9.7 |
| $B_3$ | n | 10.98 | 9.40–9.70 | 10.73 | 9.4–9.7 |
| $B_4$ | n | 11.37 | 9.40–9.70 | 11.09 | 9.4–9.7 |
| $B_5$ | $\pi$ | 12.91 | 11.53 | 12.88 | 11.5 |

| Sugar Orbitals | | | | | |
|---|---|---|---|---|---|
| | | tetrahydrofuran (**2**), cyclopentanol (**3**) | | 5'-dCMP (**12**) | |
| orbital | orbital type | $IP_{calc}$[b] | $IP_{exp}$[c] | $IP_{calc}$[b] | $IP_{est}$[d] |
| $S_1$ | n | 11.20 | 9.69 | 12.03 | 9.7 |
| $S_2$ | n | 11.05 | 10.21 | 12.51 | 10.2 |
| $S_3$ | n | | | 13.44 | 10.8 |
| $S_4$ | n | 12.06 | 11.1 | 13.87 | 11.1 |

| Phosphate Orbitals | | | | | |
|---|---|---|---|---|---|
| | | trimethyl phosphate (**4**) | | 5'-dCMP (**12**) | |
| orbital | orbital type | $IP_{calc}$[b] | $IP_{exp}$[c] | $IP_{calc}$[b] | $IP_{est}$[d] |
| $P_1$ | n | 11.90 | 10.81 | 13.07 | 11.2 |
| $P_2$ | n | 11.94 | 10.81 | 13.16 | 11.3 |

[a] All ionization potentials in eV. [b] Theoretical ionization potentials obtained from 4-31G SCF calculations. [c] Experimental vertical ionization potentials. [d] Estimated ionization potentials. See text.

The observation that orbitals with similar experimental IP's and electron distributions are predicted by 4-31G SCF calculations to have significantly different IP's, when the energetic ordering of these orbitals within the valence manifold is different, provides evidence that the IP's for the $S_1$, $S_2$, and $S_4$ orbitals in **12** are nearly equal to the experimental values for the corresponding orbitals in **2** and **3**. In Table VI, the estimated ionization potentials of the $S_1$, $S_2$, and $S_4$ orbitals in **12** have been obtained by employing this approximation.[59]

In Table VI, the estimated ionization potentials for the $S_3$ orbital in **12** has been obtained by combining 4-31G SCF results for **12** with photoelectron data for **3**. According to the calculations, the difference between the IP's of the $S_2$ and $S_3$ orbitals is 0.68 times as large as the difference between the IP's of the $S_2$ and $S_4$ orbitals. Employing this scaling factor and the experimental value (0.89 eV) for the difference between the IP's for the $S_2$ and $S_4$ orbitals in **3** yields an estimated value of 10.8 eV for the IP for the $S_3$ orbital in **12**.

**Phosphate Orbitals.** The correspondence between methyl phosphate and the phosphate group in 5'-dCMP (**12**) which is demonstrated by results of the population analyses and by the orbital diagrams indicates that the ionization potentials of the $P_1$ and $P_2$ orbitals in **12** are almost equal to the values in methyl phosphate.

However, the calculated results predict that the IP's for the $P_1$ and $P_2$ orbitals in **12** are 0.74 and 0.75 eV larger than the IP's

(56) The influence of the cutoff criteria is illustrated by the second highest occupied $\pi$ orbitals ($B_2$) in **1** and **12**. Figure 2 indicates that in **1** there is no contribution to the $B_2$ orbital from the 2p $\pi$ atomic orbital of the $C_4$ atom, while Figure 5 indicates that in **12** the $B_2$ orbital has a significant 2p $\pi$ contribution from the $C_4$ atom. Actually, for **1** and **12**, the inner orbital 2p $\pi$ coefficients on the $C_4$ atoms are 0.18 and 0.20, respectively. Similar observations explain small differences in the diagrams of the $B_4$ and $B_5$ orbitals in **1** and **12**.

(57) Ballard, R. E.; Gunnell, G. G.; Hagan, W. P. *J. Electron Spectrosc. Relat. Phenom.* **1979**, *16*, 435.

(58) Klasinc, L.; Kovac, B.; Gusten, H. *Pure Appl. Chem.* **1983**, *55*, 289.

(59) The $S_4$ band in the photoelectron spectrum for **3** is not resolved (see Figure 3); however, the 4-31G SCF calculations indicate that this band arises from the NHOMO. Consistent with the 4-31G results, the $S_4$ band in **3** has been assigned to the shoulder occurring at 11.08 eV in the spectrum.

of the $P_1$ and $P_2$ orbitals in methyl phosphate. As in the case of the $S_1$, $S_2$, and $S_4$ ionization potentials in **12** compared to the values in tetrahydrofuran (**2**) and cyclopentanol (**3**), these differences arise from differences in the energetic positions of the $P_1$ and $P_2$ orbitals in the valence manifolds of **12** versus methyl phosphate. The 4-31G SCF calculations predict that the $P_1$ and $P_2$ orbitals in **12** are the eighth and the ninth highest occupied orbitals. In methyl phosphate, the $P_1$ and $P_2$ orbitals are the HOMO and NHOMO.

In Table VI, the estimated IP's for the $P_1$ and $P_2$ orbitals in **12** were taken to be the same as in methyl phosphate. Experimental IP's for the $P_1$ and $P_2$ orbitals in methyl phosphate have not been measured because the molecule decomposes in the gas phase. Instead, estimated ionization potentials ($IP_{est}(P_i)$) for the $P_1$ and $P_2$ orbitals in methyl phosphate were obtained by using eq 1 and 2. In eq 1 and 2, $IP_{exp,4}(P_i)$ represents the experimental

$$IP_{est}(P_i) = IP_{calc}(P_i) - \Delta IP \quad (1)$$

$$IP = IP_{calc,4}(P_i) - IP_{exp,4}(P_i) \quad (2)$$

ionization potentials for the $P_i$ orbital in trimethyl phosphate (**4**), while $IP_{calc,4}(P_i)$ and $IP_{calc}(P_i)$ are the ionization potentials of the $P_i$ orbitals ($i$ = 1, 2) in **4** and methyl phosphate as predicted by 4-31G SCF calculations. Values of $IP_{est}(P_i)$ for the $P_1$ and $P_2$ orbitals are 11.2 and 11.3 eV, respectively. The reliability of eq 1 and 2 depends on the observation that the perturbation of IP's due to methyl substitution is accurately predicted by 4-31G SCF calculations.[7b,20,60]

**Valence Orbital Ionization Potentials in 5′-dCMP⁻ (13). Base and Sugar Orbitals.** As expected, the results from the 4-31G SCF calculations given in Figure 7 indicate that the upper occupied base ($B_1$–$B_4$) and sugar ($S_1$ and $S_2$) orbitals in 5′-dCMP⁻ (**13**) have ionization potentials which are smaller than those of corresponding orbitals in 5′-dCMP (**12**). For the $B_1$–$B_4$ orbitals, the calculated IP's in **13** are 2.41–2.55 eV smaller than the values in **12**; for the $S_1$ and $S_2$ orbitals the calculated IP's in **13** are 2.91 and 3.05 eV smaller. Nevertheless, the 4-31G SCF results indicate that there is a strong correspondence between the base and sugar orbitals in **12** and **13**. For 5′-dCMP and 5′-dCMP⁻, this is demonstrated by similarities between the results of population analyses for the base and sugar groups, between the relative energetic ordering of the base and sugar orbitals, and between diagrams of the base and sugar orbitals.

This correspondence provides evidence that the accuracy of ionization potentials of the upper occupied base and sugar orbitals in **13**, obtained from 4-31G SCF calculations, is approximately the same as the accuracy of the calculated ionization potentials of these orbitals in **12**.

Employing this approximation, estimated values of the ionization potentials ($IP_{est,13}(i)$) of the base and sugar orbitals in **13** have been obtained from eq 3 and 4. In eq 3, $IP_{calc,13}(i)$ represents

$$IP_{est,13}(i) = IP_{calc,13}(i) - \Delta IP \quad (3)$$

$$\Delta IP = IP_{calc,12}(i) - IP_{exp}(i) \quad (4)$$

the calculated ionization potential for the $i$th base or sugar orbital in **13**. In eq 4, $IP_{calc,12}(i)$ and $IP_{exp}(i)$ are values of the ionization potentials of the $i$th orbital obtained from calculations on **12** and from photoelectron spectra of the model compounds 1-methylcytosine (**1**), tetrahydrofuran (**2**), and cyclopentanol (**3**). Table VII lists estimated IP's of the base and sugar orbitals in **13**, obtained via eq 3 and 4.

Evidence supporting the validity of eq 3 and 4 is provided by a consideration of the ionization potentials for the highest occupied $\pi$ orbital ($B_1$) in **13**. A supplementary supermolecule calculation has indicated that the low IP of the $B_1$ orbital in **13**, compared to that in **12**, is due primarily to a through-space interaction of the negatively charged phosphate group with the base orbitals.

(60) Urano, S.; Fetzer, S.; Harvey, R. G.; Tasaki, K.; LeBreton, P. R. *Biochem. Biophys. Res. Commun.* **1988**, *154*, 789.

**Table VII.** Theoretical and Estimated Ionization Potentials of 5′-dCMP⁻[a]

| orbital | orbital type | $IP_{calc}$[b] | $IP_{est}$[c] |
|---|---|---|---|
| | Base Orbitals | | |
| $B_1$ | $\pi$ | 6.25 | 6.1 |
| $B_2$ | $\pi$ | 7.74 | 7.0–7.3 |
| $B_3$ | n | 8.24 | 6.9–7.2 |
| $B_4$ | n | 8.65 | 7.0–7.3 |
| | Sugar Orbitals | | |
| $S_1$ | n | 9.12 | 6.8 |
| $S_2$ | n | 9.51 | 7.2 |
| | Phosphate Orbitals | | |
| $P_1$ | n | 6.69 | 4.6 |
| $P_2$ | n | 7.05 | 4.8 |
| $P_3$ | n | 8.02 | 5.4 |
| $P_4$ | n | 8.16 | 5.4 |
| $P_5$ | n | 8.94 | 5.9 |

[a] All ionization potentials in eV. [b] Theoretical ionization potentials obtained from 4-31G SCF calculations. [c] Estimated ionization potentials. See text.

The calculation was carried out at the 4-31G SCF level on 1-methylcytosine (**1**) in the presence of $H_2PO_4^-$ (**10**). In the calculation, the geometries of **1** and **10** were the same as those described above, and the relative position of the molecule and anion was chosen so that the orientation of **10** relative to **1** was the same as that of the base and the phosphate groups in **13**. Evidence indicating the importance of through-space effects on the base IP's of **13** is provided by the observation that the calculated value (6.48 eV) for the IP of the $B_1$ orbital in the supermolecule is almost equal to the calculated value (6.25 eV) for the $B_1$ orbital in **13**.

In order to assess whether the accuracy of the 4-31G SCF values for the IP's of a neutral molecule are of the same order as the accuracy of values for the IP's of the molecule in the presence of an anion, two more calculations were performed. These were carried out in order to compare calculated $\pi$ ionization potentials of a supermolecule system containing ethylene and $PO_3^-$. One calculation was performed at the 4-31G SCF level. The other was performed by using more reliable second-order Møller–Plesset perturbation theory (MP2).[61,62] In both calculations, the geometry of ethylene was taken from electron diffraction data,[37] and the geometry of $PO_3^-$ was the same as that used above.[63]

The 4-31G SCF calculations yielded a value of 6.15 eV for the highest occupied orbital in the supermolecule, which is almost exclusively an ethylene $\pi$ orbital. The magnitude of the difference between the $\pi$ ionization potential obtained from the 4-31G calculations and that obtained from MP2 calculations is of the same order as the difference between the ionization potential of ethylene obtained from 4-31G calculations and the experimental vertical ionization potential. In both cases, these differences are less than 0.5 eV. These differences are also similar to the difference (0.3 eV) between the IP of 1-methylcytosine obtained from 4-31G calculations and the experimental value. The results of the supermolecule calculations on ethylene plus $PO_3^-$ support the conclusion that 4-31G calculations provide base and sugar IP's in 5′-dCMP⁻ with the same accuracy as they provide base and sugar IP's in 5′-dCMP.

**Phosphate Orbitals.** Similarities between results of the 4-31G population analyses for the phosphate group of 5′-dCMP⁻ (**13**) and $CH_3HPO_4^-$ (**11**) and the correlation between the orbital diagrams for **13** and **11** indicate that ionization potentials for the $P_1$–$P_5$ orbitals in **13** are nearly equal to those in **11**. In Table VII estimated IP's for the $P_1$–$P_5$ orbitals in **13** were taken to be equal

(61) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(62) Results obtained from MP2 calculations on the anions listed in Table IV are comparable to those obtained from MP3 calculations. Values of IP's obtained from MP2 calculations with method II are 1.45, 3.34, and 4.90 eV for $CH_3O^-$, $PO_2^-$, and $PO_3^-$, respectively.

(63) The MP2 calculations employed the 6-31+G* basis set with five d functions. The calculations were carried out with a procedure similar to method II in which only the geometry of the negatively charged supermolecule was employed.

to those in **11**. It is likely that this approximation is reliable, even though the calculated values of the IP's for the $P_1$–$P_5$ orbitals in **13** are more than 1.1 eV larger than the corresponding values in **11**. These differences are due to differences in the energetic ordering of the $P_1$–$P_5$ orbitals in the valence manifolds of **11** and **13**. In **11** the $P_1$–$P_5$ orbitals are the five highest occupied orbitals, while in **13** they are predicted to be the second, third, fifth, sixth, and ninth highest occupied orbitals.[64]

Unlike the base and sugar model compounds (**1**–**3**), no photoemission data are available for **11**. Furthermore, because of the large amount of computational time needed, it is not currently feasible to carry out post SCF calculations on **11**. In the present investigation, estimated values ($IP_{est,11}(i)$) of the ionization potentials for the five highest occupied orbitals in **11** were obtained from eq 5. In eq 5, $IP_{est,11}(i)$ is the estimated ionization potential

$$IP_{est,11}(i) = 0.55IP_{calc,11}(i) + 1.54 \quad (5)$$

for the $i$th phosphate orbital ($i$ = 1–5) in **11**, and $IP_{calc,11}(i)$ is the ionization potential obtained from 4-31G SCF calculations.

Equation 5 was derived by comparing values of the first five IP's in $PO_2^-$ predicted by MP2 calculations[61] with values predicted by 4-31G SCF calculations. The values for the first to fifth IP's of $PO_2^-$, obtained from the MP2 calculations, are 3.34, 3.80, 5.33, 6.16, and 6.16 eV, respectively.[66] The values obtained from the 4-31G SCF calculations are 3.20, 4.91, 5.44, 8.16, and 8.77 eV. Equation 5 is the result of a linear regression analysis of these two sets of values.[67]

A test of the accuracy of this scaling of phosphate anion IP's is provided by considering the estimated value (5.05 eV) obtained from eq 5 for the first IP of $PO_3^-$. This estimated value agrees well with the values of 5.03 and 5.13 eV listed in Table IV and obtained from MP3 calculations.

The estimated ionization potentials for the $B_1$–$B_4$, $S_1$ and $S_2$, and $P_1$–$P_5$ orbitals in **13** are significantly different than the ionization potentials obtained by applying Koopmans' theorem directly to the 4-31G SCF results. According to the estimated IP's in Table VII, the first to fifth highest occupied orbitals in **13** are on the phosphate group, the sixth to eighth and eleventh highest occupied orbitals are on the base group, and the ninth and tenth orbitals are on the sugar group.

The estimated IP's listed in Table VII are generally similar to those reported for the S–P–S–C sequence.[24] For example, the HOMOs in both 5′-dCMP$^-$ and in the S–P–S–C sequence are phosphate orbitals, and the IP of the HOMO obtained in the present work differs from the IP previously reported for the HOMO by only 0.38 eV. However, the results obtained in the present investigation differ in some respects from those obtained in the earlier investigation. In the previous study, which used a smaller basis set, the phosphorus atomic orbitals contribute significantly to the upper occupied orbitals in the S–P–S–C sequence. Results of the 4-31G calculations indicate, however, that there are no phosphorus atomic orbitals whose coefficients are larger than 0.2 in any of the 24 highest occupied orbitals in 5′-dCMP$^-$ (**13**). Furthermore, the present investigation yields upper orbital ionization potentials for **13** which are more closely spaced than those reported for the S–P–S–C sequence. For example, the results in Table VII indicate that the difference between the IP's of the $P_1$ and $P_5$ orbitals is 1.3 eV. On the other hand, the difference between the first and fifth phosphate orbital IP's for the S–P–S–C sequence is reported to be about 4.4 eV. Moreover, the difference in the IP's for the uppermost phosphate orbital and the uppermost base orbital in **13** is 1.5 eV, while in the S–P–S–C sequence, the reported difference is approximately 2.5 eV.

While our investigation of 5′-dCMP$^-$ indicates that the ionization potential of the phosphate group is 1.5 eV lower than that of the base, this description will change when the nucleotide is no longer isolated. In a physiological environment containing water and counterions, the IP's of the phosphate orbitals in **13** increase relative to the IP's of the base orbitals. This is indicated by supplementary 3-21G[68] SCF calculations on two supermolecules. One calculation was carried out on $H_2PO_4^-$ in the presence of two water molecules and one $Na^+$ ion. The geometry of the supermolecule used in the calculation was based on an optimized geometry obtained from 3-21G SCF calculations.[69] The second calculation was carried out on 1-methylcytosine, also in the presence of two water molecules and one $Na^+$ ion. The geometry used in this calculation models the hydration structure surrounding cytosine in DNA[70] and was based on results from a Monte Carlo simulation.[71] The first calculation indicated that, in the presence of water molecules and $Na^+$, the IP of $H_2PO_4^-$ is 8.25 eV greater than that of isolated $H_2PO_4^-$. In contrast, the second calculation indicated that, in the presence of water molecules and $Na^+$, the IP of 1-methylcytosine is only 3.22 eV greater than the IP of isolated 1-methylcytosine. These calculations suggest that, in a polar medium containing counterions, the ionization potentials of the bases in nucleotides are comparable or lower than the ionization potentials of the phosphate groups.

**Table VIII.** Percent of Total Nucleic Acid in Vitro Alkylation Occurring at Phosphate

| | double-stranded[a] | double-stranded[b] | single-stranded[c] |
|---|---|---|---|
| *N*-ethyl-*N*-nitrosourea | 55.35 | 57 | 65 |
| *N*-methyl-*N*-nitrosourea | 12.1 | 17 | 10 |
| ethyl methanesulfonate | 12.0 | 13 | 10 |
| methyl methanesulfonate | 0.82 | 0.8 | 2 |
| diethyl sulfate | | 16 | 6 |
| dimethyl sulfate | | | <2 |

[a] Results obtained with DNA from salmon testes. See ref 72. [b] Average results obtained with DNA from salmon sperm, calf thymus, salmon testes, rat liver and brain, human fibroblast, and HeLa and V79 cells. See ref 73. [c] Average results obtained with DNA from M13 phage and with RNA from TMV, yeast, HeLa cells, animal ribosomes, and $\mu$2 phage. See ref 73.

(64) For the supplemental calculations the P atom of $PO_3^-$ was located 3.0 and 3.8 Å from the two C atoms of ethylene. The orientation of two of the O atoms in $PO_3^-$, relative to ethylene, was chosen to mimic the orientation of the two negatively charged O atoms of the phosphate group in 5′-dCMP$^-$, relative to the $C_4$ and $C_5$ atoms of cytosine.

(65) Evidence supporting this evaluation of the IP's predicted by 4-31G SCF calculations is provided by a calculation carried out on 2′-deoxyribose-5′-monophosphate (5′-dMP$^-$), which contains only the sugar and phosphate groups. In this calculation, the geometry used for 5′-dMP$^-$ was the same as that of the sugar and phosphate groups in **13**. The results indicate that the $P_1$–$P_5$ orbitals are the five highest occupied orbitals in 5′-dMP$^-$ and have calculated IP's which differ from those for the corresponding orbitals in **11** by less than 0.48 eV. This difference is smaller than the difference (≥1.1 eV) between the calculated IP's of corresponding orbitals in **13** and **11**.

(66) Ionization potentials from the MP2 calculations were obtained by using the 6-31+G* basis set in combination with method II. All five of the highest occupied orbitals in $PO_2^-$, like the $P_1$–$P_5$ orbitals in **11** and **13**, contain large O atom 2p contributions. The four highest occupied orbitals are O atom lone-pair orbitals. The fifth highest occupied orbital is a P–O $\sigma$ bonding orbital.

(67) (a) Lambert, J. B.; Xue, L.; Bosch, R. J.; Taba, K. M.; Marko, D. E.; Urano, S.; LeBreton, P. R. *J. Am. Chem. Soc.* **1986**, *108*, 7575. (b) Santiago, C.; Houk, K. N.; DeCicco, G. J.; Scott, L. T. *J. Am. Chem. Soc.* **1978**, *100*, 692. (c) Angeli, R. P.; Hornung, S. D.; Christoffersen, R. E. In *Proceedings of the Third International Conference on Computers in Chemical Research, Education and Technology*; Caracas, Venezuela, July, 1976. Ludeña, E. V., Sabelli, N. H., Wahl, A. C., Eds.; Plenum: New York, 1977; pp 357–379.

(68) (a) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939. (b) Gordon, M. S.; Binkley, J. S.; Pople, J. A.; Pietro, W. J.; Hehre, W. J. *Ibid.* **1982**, *104*, 2797.

(69) Prasad, C. V.; Pack, G. R. *Ibid.* **1984**, *106*, 8079.

(70) The two $H_2O$ molecules were located in the same plane as 1-methylcytosine. One was near the amino group of 1-methylcytosine and oriented so that the distance between the O atom of $H_2O$ and the $N_8$ atom was 2.2 Å. The distances between the two H atoms of $H_2O$ and the $N_8$ atom were equal. The other $H_2O$ molecule was located near the $O_7$ atom of 1-methylcytosine. It was oriented so that the distance between the O atom of $H_2O$ and the $O_7$ atom was also 2.2 Å, and the distances between the two H atoms of $H_2O$ and the $O_7$ atom were equal. The $Na^+$ ion was located above the cytosine ring at a distance of 3.7 Å from the $C_5$ atom.

(71) Clementi, E.; Corongiu, G. In *Current Aspects of Quantum Chemistry 1981, Proceeding of an International Conference and Workshop*; Barcelona, Spain, 28 September–3 October 1981, Carbo, R., Ed.; Elsevier: Amsterdam, 1982; pp 331–378.

**Nucleotide Alkylation Patterns: Ethylation versus Methylation.** In nucleotide anions the separation of sites with low ionization potentials and high electron-donating ability, which occur on the base, from sites with the largest negative charge, which occur at the phosphate group, leads to interesting speculation concerning DNA and RNA reaction properties. It is expected, for example, that nucleotides will participate differently in orbitally controlled reactions, in which the bases are more important electrophilic target sites, than in electrostatically controlled reactions, in which the phosphate groups are more important electrophilic target sites.

Evidence supporting this conclusion is provided by previous studies of alkylation patterns observed in reactions of a number of direct acting carcinogens with DNA and RNA. Table VIII lists the percentage of total nucleotide alkylation which occurs at the phosphate groups in in vitro studies of reactions of polynucleotides with *N*-alkyl-*N*-nitrosoureas, alkyl alkanesulfonates, and dialkyl sulfates. The reactions have been studied with double-stranded DNA and with single-stranded DNA and RNA. In all cases alkylation occurs primarily (>88%) at the bases and phosphate groups.[72,73]

The results demonstrate that when comparable alkylating reagents are considered, the percentage of alkylation which occurs at the phosphate group is greater for ethylation than for methylation. For example, in reactions with *N*-alkyl-*N*-nitrosoureas and alkyl methanesulfonates, the percentage of ethylation is 3.3–63 times and 5–16 times greater than the percentage of methylation. Similarly, in reactions with dialkyl sulfates, the percentage of ethylation is 3 times greater than the percentage of methylation. The results in Table VIII also suggest that the difference in the methylation versus ethylation patterns is influenced more by electronic effects than by steric effects. This conclusion is supported by the observation that the percentage of ethylation which occurs at the phosphate groups in single-stranded DNA and RNA, with extended regions of disrupted base-pairing and base-stacking, is similar to that which occurs in conformationally ordered double-stranded DNA.

The results in Table VIII are consistent with the description of the electronic structure of 5′-dCMP$^-$ provided here and with model potential surfaces for nucleotide alkylation reactions calculated by Ford and Scribner.[29] The model surfaces indicate that activation energies for the alkylation of nucleotide bases are influenced by both orbital and electrostatic interactions.

A comparison of model transition-state potential surfaces, associated with $SN_2$ ethylation and methylation reactions, points out that the greater ability of ethyl groups versus methyl groups to support positive charge has strong influence on the difference between ethylation and methylation patterns.[29] The difference between the energies required to form ethyl and methyl cations is important in determining the volume spanned by transition states that exhibit carbocationic character. According to the model of Ford and Scribner,[29] the greater energy requirements for the formation of methyl cations gives rise to transition states in DNA methylation reactions in which the reactants are closer together than in transition states of ethylation reactions. Consideration of the tighter transition states, occurring in methylation reactions, leads to the prediction that short-range orbital interactions play a significant role in determining activation energies. In ethylation reactions, which proceed via looser transition states, long-range electrostatic interactions are expected to be more important.

Within the framework of this model the present description of valence orbital structures in 5′-dCMP$^-$, which contains a negative charge located on the phosphate group and orbitals with low ionization potentials on both the base and phosphate groups, leads to the prediction that as reactions become more electrostatically controlled there will be increased reactivity at the phosphate groups and less at the base. The reactivity patterns for DNA and RNA methylation versus ethylation which are summarized in Table VIII are consistent with this prediction.

**Acknowledgment.** We thank Professors Nora Sabelli (University Illinois, Urbana-Champaign) and George Ford (Southern Methodist University) for helpful discussions. Support of this work by the Petroleum Research Fund (Grant No. 19245-AC), the National Cancer Institute of the National Institutes of Health (Grant No. CA41432), and Cray Research, Inc., is gratefully acknowledged. We also thank the Computer Center of the University of Illinois at Chicago, the Pittsburgh Supercomputer Center, the Cornell National Supercomputer Facility, and the National Center for Supercomputing Applications, University of Illinois, Urbana-Champaign.

**Registry No.** **1**, 1122-47-0; **2**, 109-99-9; **3**, 96-41-3; **4**, 512-56-1; **5**, 78-40-0; **6**, 126-73-8; **7**, 3315-60-4; **8**, 20499-58-5; **9**, 15389-19-2; **10**, 14066-20-7; **11**, 48000-95-9; **12**, 1032-65-1; **13**, 123642-69-3; $CH_3O^•$, 2143-68-2; $PO_2^•$, 12164-97-5; $PO_3^•$, 12651-01-3; *N*-ethyl-*N*-nitrosourea, 759-73-9; *N*-methyl-*N*-nitrosourea, 684-93-5; ethylmethanesulfonate, 62-50-0; methylmethanesulfonate, 66-27-3; diethyl sulfate, 64-67-5; dimethyl sulfate, 77-78-1.

(72) Beranek, D. T.; Weis, C. C.; Swenson, D. H. *Carcinogenesis* **1980**, *1*, 595.

(73) Singer, B.; Grunberger, D. *Molecular Biology of Mutagens and Carcinogens*; Plenum Press: New York, 1983; pp 65–68.