Aqueous Ionization and Electron-Donating Properties of Dinucleotides: Sequence-Specific Electronic Effects on DNA Alkylation

Nancy S. Kim, Qiqing Zhu, and Pierre R. LeBreton*

Contribution from the Department of Chemistry, The University of Illinois at Chicago, Chicago, Illinois 60607-7061

Received June 17, 1999

Abstract: He I, UV photoelectron data and results from self-consistent-field (SCF) and post-SCF molecular orbital calculations were used to evaluate multiple gas-phase ionization potentials (IPs) of the phosphorylated dinucleotide $^{5'}$pGpAp both isolated and in a cluster with Na^+ and H_2O. SCF calculations with a split-valence basis set indicate that the ground-state valence orbital structure of $^{5'}$pGpAp is generally localized on the base, sugar, or phosphate groups and that orbitals in the dinucleotide are similar to orbitals in the model compounds and anion, 1,9-dimethylguanine, 9-methyladenine, 3-hydroxytetrahydrofuran, and $H_2PO_4^-$. This correlation parallels that in mononucleotides (Fernando, H.; Papadantonakis, G. A.; Kim, N. S.; LeBreton, P. R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5550−5555; Kim, N. S.; LeBreton, P. R. *J. Am. Chem. Soc.* **1996**, *118*, 3694−3707) and permits the correction of SCF ionization potentials using experimental IPs for the model compounds and theoretical IPs from post-SCF calculations on $H_2PO_4^-$. A comparison of IPs of $^{5'}$pGpAp, $^{5'}$pGpA, 2'-deoxyguanosine 5'-phosphate (5'-dGMP$^-$), and 2'-deoxyadenosine 5'-phosphate (5'-dAMP$^-$), both isolated and in gas-phase clusters with Na^+ and H_2O, indicates that the electrostatic influence of Na^+ and the anionic phosphate groups is great. IPs decrease significantly as the number of phosphate groups increases. For 5'-dGMP$^-$, $^{5'}$pGpA, and $^{5'}$pGpAp, the lowest adiabatic guanine ionization potentials are 5.2, 3.1, and 1.6 eV, respectively; the lowest phosphate vertical IPs are 5.1, 3.1, and 1.9 eV. Gas-phase IPs were combined with hydration energies obtained with a Langevin dipole relaxation model to evaluate the 22 lowest aqueous ionization energies in $^{5'}$pGpAp as well as ionization energies in the other nucleotides with and without counterions. The electrostatic dependence of the gas-phase IPs on the number of phosphate groups is modulated in aqueous solution. For 5'-dGMP$^-$, $^{5'}$pGpA, and $^{5'}$pGpAp, the aqueous guanine and phosphate ionization energies lie in the ranges 4.2−4.9 eV and 5.5−6.1 eV, respectively. Solvent relaxation similarly modulates Na^+ electrostatic effects. It also alters the relative energies of ionization events. In the gas phase, the first adiabatic base IP is equal to or larger than the lowest phosphate vertical IP. In water, the base ionization energies are 1.2−1.3 eV smaller than the phosphate ionization energies. Guanine gas-phase π ionization energies in five different B-DNA oligomer models, each containing six stacked base pairs, were compared with site-specific reactivity data for methylation at the guanine N7 and O^6 atoms by the carcinogen, *N*-methyl-*N*-nitrosourea (MNU). O^6 methylation is associated with MNU carcinogenesis; N7 methylation is the principal DNA reaction pathway. At both guanine atoms, reactivity increases as the lowest guanine π ionization potential decreases. This is consistent with a description of transition states in which activation barriers are lowered as nucleotide base π polarizability increases.

Introduction

The evaluation of nucleotide ionization threshold energies provides information about nucleotides that is important for understanding DNA biochemistry and biophysics, especially for understanding mechanisms of carcinogenesis.[1−3] For example, the energetics of nucleic acid ionization in biologically relevant environments is central to an understanding of mechanisms of DNA damage caused by high-energy photons. In an aqueous environment, ionizing radiation causes both indirect DNA damage through reaction with OH radicals formed via H_2O radiolysis[4] and direct DNA damage involving DNA photochemistry.[4−6] Current evidence links direct DNA photoionization to mechanisms leading to strand scission and base oxidation.[6,7] Hydroxyl radical scavenging experiments with eukaryotic cells

(1) Kim, H. S.; LeBreton, P. R. *J. Am. Chem. Soc.* **1996**, *118*, 3694.

(2) Fernando, H.; Papadantonakis, G. A.; Kim, N. S.; LeBreton, P. R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5550.

(3) (a) Averbeck, D. *Ann. Phys.* **1997**, *22*, C1−157. (b) Williams, D. *Nature* **1994**, *371*, 556. (c) Jin, Y.; Burns, F. J.; Garte, S. J.; Hosselet, S. *Carcinogenesis* **1996**, *17*, 873. (d) Garte, S. J.; Burns, F. J. *Environ. Health Perspect.* **1991**, *93*, 45. (e) Mizuno, T.; Kyoizumi, S.; Suzuki, T.; Iwamoto, K. S.; Seyama, T. *Oncogene* **1997**, *15*, 1455. (f) Preston, D. L.; Kusumi, S.; Tomonaga, M.; Izumi, S.; Ron, E.; Kuramoto, A.; Kamada, N.; Dohy, H.; Matsui, T.; Nonaka, H.; Thompson, D. E.; Soda, M.; Mabuchi, K. *Radiat. Res.* **1994**, *137*, S68. (g) Hancock, S. L.; McDougall, I. R.; Constine, L. S. *Int. J. Radiat. Oncology Biol. Phys.* **1995**, *31*, 1165. (h) Bedford, J. S. *Int. J. Radiat. Oncology Biol. Phys.* **1991**, *21*, 1457.

(4) Becker, D.; Sevilla, M. D. In *Advances in Radiation Biology, Vol. 17*; Lett, J. T., Adler, H., Eds.; Academic Press: New York, 1993; pp 121−180.

(5) (a) Melvin, T.; Botchway, S. W.; Parker, A. W.; O'Neill, P. *J. Am. Chem. Soc.* **1996**, *118*, 10031. (b) deLara, C. M.; Jenner, T. J.; Townsend, K. M. S.; Marsden, S. J.; O'Neill, P. *Radiat. Res.* **1995**, *144*, 43.

(6) (a) Malone, M. E.; Cullis, P. M.; Symons, M. C. R.; Parker, A. W. *J. Phys. Chem.* **1995**, *99*, 9299. (b) Shulte-Frohlinde, D.; Opitz, J.; Görner, H.; Bothe, E. *Int. J. Radiat. Biol.* **1985**, *48*, 397. (c) Opitz, J.; Shulte-Frohlinde, D. *J. Photochem.* **1987**, *39*, 145.

(7) Steenken, S.; Goldbergerova, L. *J. Am. Chem. Soc.* **1998**, *120*, 3928.

indicate that, under some conditions, the contribution from direct ionization to total radiation-induced DNA damage is as large as 35%.[5]

Because the preparation of intact gas-phase nucleotides is difficult and because the large number of valence orbitals in nucleotides with similar energies will give rise to poorly resolved ionization spectra, the experimental gas-phase evaluation of ionization potentials (IPs) for intact nucleotides has not yet been carried out. Evaluations of the IPs of DNA components that have relied exclusively on electronic structure calculations have focused only on the bases and sugars.[8,9] These theoretical investigations have not examined intact nucleotides or DNA components containing anionic phosphate groups.

The evaluation of gas-phase IPs of intact nucleotides has been carried out using a combined theoretical and experimental approach. Here, IPs were provided for $11-14$ of the lowest energy ionization events in 2′-deoxyguanosine 5′-phosphate (5′-dGMP$^-$) and in the corresponding adenine, cytosine, and thymine mononucleotides, 5′-dAMP$^-$, 5′-dCMP$^-$, and 5′-dTMP$^-$.[1,10−13] This approach relies on application of Koopmans' theorem[14] to results from ab initio, self-consistent-field (SCF) molecular orbital calculations on intact nucleotides and on corrections obtained from results of post-SCF calculations and from experimental gas-phase IP data for nucleotide components. The post-SCF calculations were carried out on a model of anionic phosphate groups in nucleotides. The experimental IPs were measured for nucleotide base and sugar model compounds. Earlier investigations of mononucleotide ionization potentials[1,2,10−13,15] indicated that at the SCF level the upper occupied orbitals are largely localized on either the base, sugar, or phosphate groups. The combined theoretical and experimental method employed in evaluating nucleotide gas-phase IPs exploited this finding and the observation that SCF calculations describe differences in IPs caused by changes in structure more accurately than they provide absolute values of ionization energies.[1,11,16,17] To examine the energetics of DNA ionization in the most useful way, it is important to recognize that in physiological surroundings DNA interacts with water, counterions, and protein.[18,19] Interactions with water and small counterions, like Na$^+$, generally occur dynamically[18] on a time scale of pico- to nanoseconds.[18a,b,20] Results from investigations of isolated nucleotide−water−counterion clusters, in fixed geometries based on crystallographic data and in geometries that molecular dynamics calculations indicate are statistically significant, demonstrate that counterion electrostatic interactions significantly increase nucleotide IPs.[1,10,11,15] For example, the lowest energy base IP in the isolated gas-phase nucleotide, 5′-dGMP$^-$, is approximately 1.8 eV smaller than that of 5′-dGMP$^-$ in a gas-phase cluster in which the phosphate group interacts with an Na$^+$ ion through a single H$_2$O solvation shell.[1]

In addition to the influence of electrostatic interactions with counterions, the influence of bulk solvent relaxation on nucleotide ionization energies is also significant. One major influence is that bulk solvent relaxation significantly modulates the increase in nucleotide IPs due to electrostatic interaction with counterions. When water relaxation is introduced, by including free energies of hydration, the difference between the Gibbs free energy of base ionization for 5′-dGMP$^-$ without Na$^+$ and for 5′-dGMP$^-$ in which the phosphate group interacts with a solvated Na$^+$ ion decreases from 1.8 eV in the gas phase to approximately 0.1 eV.[1] Water relaxation also alters the relative energies of ionization events. In the gas phase, the lowest energy vertical ionization event in all mononucleotides occurs at the negatively charged phosphate group.[1,10,11] If water relaxation is included, base ionization of 5′-dGMP$^-$ requires approximately 0.8 eV less energy than phosphate ionization.[1] While the energetics of nucleotide ionization in aqueous solution is different from that in the gas phase, guanine is the most easily ionized base in both environments.[2,11,21] This observation is consistent with the important role that guanine plays in photoinduced DNA strand scission[22,23] and in DNA electron transfer.[24,25] The first goal of this investigation is to evaluate multiple ionization energies in the intact phosphorylated dinucleotide $^{5'}$pGpAp with water and Na$^+$ ions. This system exhibits some of the most important interactions that occur in native DNA in a biological environment.

Detailed information about DNA electronic properties, including ionization energies, will ultimately provide a fundamental understanding of the electronic factors that influence site-specific and sequence-specific electrophilic attack of DNA by small alkylating agents that are antitumor agents and carcinogens.[1,12,13,26,27] Double- and single-stranded DNA alkylation patterns arising from electrophiles with significant S$_N$2 character indicate that most reaction occurs at the bases and that N7 of guanine is the most reactive site.[26,28,29] Methyl methanesulfonate,

(8) (a) Saito, I.; Takayama, M.; Sugiyama, H.; Nakamura, T. *J. Photochem. Photobio. A: Chem* **1997**, *106*, 141. (b) Sugiyama, H.; Saito, I. *J. Am. Chem. Soc.* **1996**, *118*, 7063. (c) Saito, I.; Takayama, M.; Sugiyama, H.; Nakatani, K.; Tsuchida, A.; Yamamoto, M. *J. Am. Chem. Soc.* **1995**, *117*, 6406.

(9) (a) Colson, A.-O.; Sevilla, M. D. *J. Phys. Chem.* **1995**, *99*, 3867. (b) Sevilla, M. D.; Besler, B.; Colson, A.-O. *J. Phys. Chem.* **1995**, *99*, 1060. (c) Colson, A.-O.; Besler, B.; Sevilla, M. D. *J. Phys. Chem.* **1993**, *97*, 8092. (d) Colson, A.-O.; Besler, B.; Close, D. M.; Sevilla, M. D. *J. Phys. Chem.* **1992**, *96*, 661. (e) Hutter, M.; Clark, T. *J. Am. Chem. Soc.* **1996**, *118*, 7554.

(10) Kim, N. S.; LeBreton, P. R. *Biospectroscopy* **1997**, *3*, 1.

(11) Fernando, H.; Papadantonakis, G. A.; Kim, N. S.; LeBreton, P. R. In *Molecular Modeling of Nucleic Acids*; ACS Symposium Series 682; Leontis, N., SantaLucia, J., Jr., Eds.; American Chemical Society: Washington, DC, 1997; pp 18−40.

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

(13) Kim, H. S.; Yu, M.; Jiang, Q.; LeBreton, P. R. *J. Am. Chem. Soc.* **1993**, *115*, 6169.

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

(15) Kim, H. S.; LeBreton, P. R. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3725.

(16) Lin, J.; Yu, C.; Peng, S.; Akiyama, I.; Li, K.; Lee, L.-K.; LeBreton, P. R. *J. Am. Chem. Soc.* **1980**, *102*, 4627.

(17) (a) Rossman M. A.; Leonard, N. J.; Urano, S.; LeBreton, P. R. *J. Am. Chem. Soc.* **1985**, *107*, 3884. (b) Padva, A.; O'Donnell, T. J.; LeBreton, P. R. *Chem. Phys. Lett.* **1976**, *41*, 278.

(18) (a) Young, M. A.; Jayaram, B.; Beveridge, D. L. *J. Am. Chem. Soc.* **1997**, *119*, 59. (b) Cheatham, T. E., III; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193. (c) Pack G. R.; Wong. L.; Lamm G. *Biopolymers* **1999**, *49*, 575.

(19) Luger, K.; Maäer, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. *Nature* **1997**, *389*, 251.

(20) Seibel, G. L.; Singh, U. C.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 6537.

(21) (a) Urano, S.; Yang, X.; LeBreton, P. R. *J. Mol. Struct.* **1989**, *214*, 315. (b) Orlov, V. M.; Smirnov, A. N.; Varshavsky, Ya., M. *Tetrahedron Lett.* **1976**, 4377. (c) Hush, N. S.; Cheung, A. S. *Chem. Phys. Lett.* **1975**, *34*, 11. (d) Steenken, S. *Chem. Rev.* **1989**, *89*, 503.

(22) (a) Gasper, S. M.; Schuster, G. B. *J. Am. Chem. Soc.* **1997**, *119*, 12762. (b) Nakatani, K.; Shirai, J.; Sando, S.; Saito, I. *J. Am. Chem. Soc.* **1997**, *119*, 7626. (c) Melvin, T.; Plumb, M. A.; Botchway, S. W.; O'Neill, P.; Parker, A. W. *Photochem. Photobiol.* **1995**, *61*, 584.

(23) Meggers, E.; Kusch, D.; Spichty, M.; Wille, U.; Giese, B. *Angew. Chem., Int. Ed.* **1998**, *37*, 460.

(24) (a) Kelley, S. O.; Barton, J. K. *Science*, **1999**, *283*, 375. (b) Hall, D. B.; Holmlin, R. E.; Barton, J. K. *Nature* **1996**, *382*, 731.

(25) Holmlin, R. E.; Dandliker, P. J.; Barton, J. K. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 2715.

(26) Singer, B.; Grunberger, D. *Molecular Biology of Mutagens and Carcinogens*; Plenum Press: New York, 1983; Chapter 4.

(27) (a) Krytopoulos, S. A.; Anderson, L. M.; Chhabra, S. K.; Souliotis, V. L.; Plesta, V.; Valavanis, C.; Georgiadis, P. *Cancer Detect. Prev.* **1997**, *21*, 391. (b) Lind, M. J.; Ardiet, C. *Cancer Surveys* **1993**, *17*, 157.

dimethyl and diethyl sulfate, and the reactive methane diazonium ($CH_3N_2^+$) and methyl chloroethylaziridinium ions, derived from *N*-methyl-*N*-nitrosourea (MNU) and bis-2-chloroethylmethyl-amine (nitrogen mustard), exhibit this selectivity. For MNU, the most reactive DNA site (N7 of guanine) is different from the minor methylation site (O^6 of guanine) most closely linked with MNU carcinogenic mechanisms.[30]

One description of electronic factors that strongly influence the observed reaction patterns of methylating and ethylating reagents, and of other alkylating agents, has focused on the electrostatic potentials associated with nucleotides and nucleotide components.[31−33] A different description places greater emphasis on base polarizabilities as reflected in base π ionization potentials.[1,12,13] In the first description, activation barriers for electrophilic attack of DNA by methyl chloroethylaziridinium ions and similar alkylating agents are determined primarily by the nucleotide electrostatic potential. In this description, base reaction sites with large negative potentials give rise to transition states with low activation barriers.[32,33] In the second description,[1,13] nucleotide base polarizability is more important than the electrostatic potential in determining selectivity. Here S_N2 activation barrier heights for reactions with DNA decrease as base π polarizability increases, and π polarizabilities increase as π ionization energies decrease. The second goal of this investigation is to further assess these different descriptions of DNA susceptibility to electrophilic attack and to explore prospects for employing sequence-specific influences on ionization energies as a tool for understanding sequence-specific reactivity toward small methylating agents with significant S_N2 character.

Methods

Evaluation of Nucleotide of Gas-Phase Ionization Potentials. Gas-phase ionization potentials of isolated 5′-dGMP$^-$ and 5′-dAMP$^-$ and of the phosphorylated dinucleotides 5′pGpA and 5′pGpAp were determined by correcting values obtained via application of Koopmans' theorem to results from SCF calculations with the 3-21G basis set[34] and the Gaussian 94 program.[35] The same procedure was employed in evaluating IPs from clusters of mononucleotides and phosphorylated dinucleotides containing H_2O molecules and Na^+ counterions. Calculations were carried out with a SGI Cray Origin2000 cluster containing 768 R10000 processors. For a system containing 5′pGpAp, three Na^+ ions, and five H_2O molecules, the 3-21G SCF calculation required 225 MB of disk space and 53 min of processing time.

The correction procedure employed eqs 1 and 2, which were developed in earlier investigations.[1,2,10,11] It relies on the observation that the upper occupied orbitals in the isolated nucleotides and in nucleotide clusters are generally localized on either a base, sugar, or phosphate group and that these orbitals are similar to orbitals occurring in model base and sugar compounds and phosphate anions. As in this investigation, the approach can be employed at the ab initio level, or it can be applied to much larger systems for which semiempirical methods have recently been developed.[36] Here, the model base and sugar compounds employed are 1,9-dimethylguanine (1,9-Me_2G), 9-methyladenine (9-MeA), and 3-hydroxytetrahydrofuran (3-OH-THF). The model anion is $H_2PO_4^-$. According to eqs 1 and 2, the difference between the corrected IP associated with an occupied orbital in an isolated nucleotide or cluster and the IP obtained from 3-21G SCF calculations equals the difference between the corrected IP and the calculated IP associated with the corresponding orbital in 1,9-Me_2G, 9-MeA, 3-OH-THF, or $H_2PO_4^-$.

$$IP_{corr}(i) = IP_{calc}(i) + \Delta IP(i) \tag{1}$$

$$\Delta IP(i) = IP(i) - IP'_{calc}(i) \tag{2}$$

In eq 1, $IP_{corr}(i)$ is the corrected IP associated with the ith orbital in a nucleotide or dinucleotide that is isolated or in a cluster, and $IP_{calc}(i)$ is the IP obtained from 3-21G SCF calculations. When eq 2 is used to correct base or sugar ionization potentials, $IP(i)$ is the experimental adiabatic or vertical ionization potential associated with the ith orbital in 1,9-Me_2G, 9-MeA, or 3-OH-THF that most closely correlates with the ith orbital in the isolated nucleotide or in the cluster. $IP'_{calc}(i)$ is the ionization potential of the ith orbital in 1,9-Me_2G, 9-MeA, or 3-OH-THF obtained from 3-21G SCF results.

For 1,9-Me_2G, 9-MeA, or 3-OH-THF, the first adiabatic IPs have been measured. For higher energy ionization events only the vertical experimental IPs are available. Where possible, the adiabatic IPs were employed in eq 2 to correct the calculated ionization potentials. This permits a more direct comparison between results from the present investigation and results from future experimental measurements of aqueous ionization thresholds, such as the comparison reported earlier for tryptophan and indole.[2]

Theoretical Vertical Ionization Potentials of $H_2PO_4^-$. When eq 2 is used to correct IPs of the anionic phosphate group, $IP(i)$ is one of the four lowest energy vertical ionization potentials of $H_2PO_4^-$, obtained from a combination of previously reported[37] second-order Möller−Plesset (MP2) perturbation calculations[38] and second-order perturbation calculations (CASPT2)[37,39] with a complete active space self-consistent-field (CASSCF)[40] reference wave function. The CASSCF calculations employed a double-ζ quality basis set containing diffuse functions with 9 electrons in 10 active orbitals.[37]

The lowest energy vertical IP of $H_2PO_4^-$ was obtained as the difference between MP2 ground-state energies of the closed-shell anion, $H_2PO_4^-$, and the open-shell radical, $H_2PO_4^•$, calculated with a 6-31+G* basis set.[38] Because of limitations in computational resources, vertical ionization potentials of $H_2PO_4^-$ were evaluated at the post-SCF level. The geometry of the anion was optimized at the SCF level with a 6-31+G* basis set, and the same geometry was used for the radical.

The second through fourth IPs of $H_2PO_4^-$ were evaluated by employing the energies of the first three excited states of $H_2PO_4^•$,[37] obtained from CASPT2 calculations. The second through fourth IPs of $H_2PO_4^-$ were calculated by adding the CASPT2 excitation energies for $H_2PO_4^•$ to the value for the first IP. In two supplementary calculations[41] employing the equations of motion coupled cluster approximation with singles and doubles (EOM-CCSD CD) and the

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

(29) Brookes, P.; Lawley, P. D. *Biochem. J.* **1961**, *80*, 496.

(30) (a) Sukumar, S.; Babaracid, M. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 718. (b) Sukumar, S.; Notario, V.; Martin-Zanca, D.; Babaracid, M. *Nature* **1983**, *306*, 658. (c) Goth, R.; Rajewsky, M. F. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 639.

(31) (a) Ford, G. P.; Wang B. *Carcinogenesis* **1993**, *14*, 1465. (b) Kohn, K. W.; Hartley, J. A.; Mattes, W. B. *Nucleic Acids Res.* **1987**, *15*, 10531.

(32) Pullman, A.; Pullman, B. *Q. Rev. Biophys.* **1981**, *14*, 289.

(33) (a) Broch, H.; Hamza, A.; Vasilescu, D. *Int. J. Quantum. Chem.: Quantum. Biol. Symp.* **1996**, *23*, 21. (b) Mattes, W. B.; Hartley, J. A.; Kohn, K. W. *Nucleic Acids Res.* **1986**, *14*, 2971.

(34) (a) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. von R. *J. Comput. Chem.* **1983**, *4*, 294. (b) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939.

(35) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, Head-Gordon, C.; Gonzalez, C.; Pople, J. A. *Gaussian 94*; Gaussian, Inc.: Pittsburgh, PA, 1995.

(36) Jug, K.; Zerner, M. C. *Int. J. Quantum Chem.* **1996**, *23*, 1735.

(37) Fetzer, S. M.; LeBreton, P. R.; Rohmer, M.-M.; Veillard, A. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1997**, *65*, 1095.

(38) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiley: New York, 1986; pp 38−40, 86.

(39) (a) Andersson, K.; Malmqvist, P.-A.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218. (b) Andersson, K.; Malmqvist, P.-A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483.

(40) Siegbahn, P. E. M.; Almlöf, J.; Heiberg, A.; Roos, B. O. *J. Chem. Phys.* **1981**, *74*, 2384.

(41) Personal communication. Prof. Marcel Nooijen, Department of Chemistry, Princeton University.

Table 1. Theoretical and Experimental AdiabaticIonization Potentials of Phosphorus- and Oxygen-Containing Anions

	adiabatic ionization potential[a]		
anions	MP2[b]	MP2[c]	experimental
OH^-	1.59	1.60	1.829 ± 0.010[d]
CH_3O^-	1.37	1.38	1.570 ± 0.022[e]
PO_2^-	3.19	3.16	3.3 ± 0.2[f]
PO_3^-	5.02	4.98	4.9 ± 1.3[g]
HCO^-	−0.16	−0.18	0.313 ± 0.005[d]

[a] All ionization potentials in eV. [b] Theoretical IPs from MP2 calculations with 6-31+G* SCF geometry optimization of the closed-shell anion and the neutral radical. [c] From MP2 calculations with MP2/6-31+G* geometry optimization. [d] Taken from ref 44. [e] Taken from ref 45. [f] Taken from ref 46. [g] Estimated from heats of formation of PO_3 and $PO_3^{•-}$. See refs 47.

equations of motion many-body perturbation theory (EOM-MBPT[2])[42] with the Aug-CC−PVDZ basis set for heavy atoms and the CC−PVDZ basis set for hydrogen,[43] none of the four lowest excitation energies of $H_2PO_4^•$ differed from the CASPT2 values reported here by more than 0.1 eV.

While there are no experimental data available with which to test the values of the calculated vertical IPs of $H_2PO_4^-$, experimental values of the lowest energy adiabatic IPs of several phosphorus- and oxygen-containing anions have been measured. Table 1 shows a comparison of experimental adiabatic IPs with adiabatic MP2/6-31+G* ionization potentials obtained by optimizing the geometries of both the ground-state anions and the ground-state radicals at the 6-31+G* SCF and at the MP2 level. The results in Table 1 indicate that the calculated IPs obtained using the two geometry optimization methods differ by less than 0.1 eV. In cases where the experimental IPs are known to high accuracy, a comparison demonstrates that MP2 results with SCF geometry optimization underestimate the anion stabilities relative to the radical stabilities by 0.11−0.47 eV.

To further test the MP2/6-31+G* value for the lowest energy IP of $H_2PO_4^-$, a G2 calculation[48] with the 6−311G(dp) and several extended basis sets was carried out for $H_2PO_4^-$ with C_{2v} symmetry and $H_2PO_4^•$ with C_s symmetry. According to the results from the G2 calculations, the adiabatic IP of $H_2PO_4^-$ is 4.40 eV. This is consistent with the MP2/6-31+G* value of 4.89 eV for the first vertical IP.

Molecular Orbital Diagrams and Molecular Electrostatic Potentials. Orbital diagrams were drawn from molecular orbital coefficients obtained from results of the 3-21G SCF calculations. In the molecular orbital diagrams, 2s, 2p, and 3p orbitals of C, N, and P are shown for which the inner or outer coefficients are greater than 0.20, 0.17, and 0.15 for sugar, phosphate, and base orbitals, respectively. The variation in cutoff criteria used for the different groups reflects the differences between the number of group atoms over which the molecular orbitals are delocalized. This delocalization increases in the order sugar < phosphate < base. The sizes of the atomic orbitals in the molecular orbital diagrams are proportional to the coefficients. Molecular electrostatic potential (MEP) contour maps were obtained from a 3-21G SCF wave function using the MOLDEN program.[49]

Geometries. The geometries of 1,9-Me_2G, 9-MeA, 3-OH-THF, and $H_2PO_4^-$ employed in calculations of gas-phase IPs were the same as those described earlier.[1,10,11,13,37] Heavy atom bond lengths and bond angles for ^5pGpAp were obtained from crystallographic data for a B-DNA dodecamer.[50] The ^5pGpAp geometry corresponds to that occurring at positions 4 and 5 of strand A of the dodecamer. The heavy atom geometries of ^5pGpA, 5′-dGMP$^-$, 5′-dAMP$^-$, and 3′-dAMP$^-$ were the same as the geometries of corresponding groups in ^5pGpAp.

In the mononucleotides and the phosphorylated dinucleotides, the C−H bond lengths were 1.08 and 1.09 Å for sp^2 and sp^3 C atoms, respectively.[51] The O−H and N−H bond lengths were 0.96 and 1.01 Å.[51] For ^5pGpAp and ^5pGpA, the H atom bond angles and the torsional angles describing O−H rotation were obtained from 3-21G SCF optimization calculations on 5′-dGMP$^-$, 5′-dAMP$^-$, and 3′-dAMP$^-$.

In calculations of the Koopmans' ionization potentials of ^5pGpAp, ^5pGpA, 5′-dGMP$^-$, and 5′-dAMP$^-$, the heavy atom geometries obtained from the oligomer crystal structure were adjusted by making small changes (<5°) in the dihedral angles describing rotation about the glycosidic and phosphate−ester bonds and small changes (<2.2°) in the angles describing the sugar pucker. These adjustments were introduced in order to enhance the localization of the valence orbitals and to improve correlation with orbitals in the model compounds. In all cases, the adjustments resulted in valence orbital energy changes of less than 0.1 eV, as calculated at the 3-21G SCF level.

Figure 1 contains two views of cluster **A**, which contains ^5pGpAp, three Na^+ ions, and five water molecules. The figure also shows clusters containing 5′-dGMP$^-$ or 5′-dAMP$^-$, an Na^+ ion, and four or five water molecules. In all of the clusters, the geometries of the nucleotides are the same as the geometries of the isolated nucleotides. The positions and orientations of the water molecules and the Na^+ ions in the clusters, 5′-dGMP$^-$·4H_2O·Na^+ and 5′-dAMP$^-$·5H_2O·Na^+, are the same as those of corresponding water molecules and Na^+ ions in earlier investigations.[10,15] The geometries of 5′-dGMP$^-$·4H_2O·Na^+ and 5′-dAMP$^-$·5H_2O·Na^+ are also similar to the local geometries occurring in cluster **A**. 5′-dGMP$^-$·4H_2O·Na^+ contains Na^+(1) and water molecules W(1) and W(4) of cluster **A**; 5′-dAMP$^-$·5H_2O·Na^+ contains Na^+(2) and water molecules W(2) and W(5). Besides water molecules with geometries similar to those occurring in cluster **A**, 5′-dGMP$^-$·4H_2O·Na^+ and 5′-dAMP$^-$·5H_2O·Na^+ each contain two extra water molecules belonging to the inner solvation shell of Na^+. The 5′-dAMP$^-$·5H_2O·Na^+ cluster also has a water molecule hydrogen bonded to N3 of adenine.

The geometry of cluster **A** was obtained by combining X-ray data with results from geometry optimization calculations on components of the cluster. The O atoms of water molecules W(4) and W(5) are located 3.0 and 2.8 Å from O^6 of guanine and N7 of adenine, respectively. Each of these water molecules is coplanar with the base with which it interacts most strongly. The positions of W(4) and W(5) were chosen on the basis of X-ray data that describes hydration sites on B-DNA.[50,52]

In cluster **A**, each of the Na^+ ions interacts strongly with a different phosphate group through an intermediary H_2O molecule. The binding of Na^+ at phosphate groups[20,53] of nucleotides is favorable. Furthermore, the strong interaction of one Na^+ per nucleotide is consistent with the average counterion binding level that occurs in the counterion condensation model.[54] In DNA solutions that are 9 mM in phosphate and Na^+, more than 70% of all of the phosphate charge is neutralized.[55] In the cluster, the distance (2.4 Å) between each Na^+ ion and the O atom of the nearest H_2O molecule is the same as that obtained from a

(42) (a) Nooijen, M.; Bartlett, R. J. *J. Chem. Phys.* **1995**, *102*, 3629. (b) Stanton, J. F.; Bartlett, R. J. *J. Chem. Phys.* **1993**, *98*, 7029.

(43) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1993**, *98*, 1358, and references therein.

(44) Lias, S. G.; Bartmess, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. *J. Phys. Chem. Ref. Data* **1988**, *17*, Suppl. 1.

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

(46) Wu, R. L. C.; Tiernan, T. O. *Bull. Am. Phys. Soc.* **1982**, *27*, 109.

(47) (a) Henchman, M.; Vigianno, A. A.; Paulson, J. F.; Freedman, A.; Wormhoudt, J. *J. Am. Chem. Soc.* **1985**, *107*, 1453. (b) Unkel, W.; Freedman, A. *AIAA* **1983**, *21*, 1648. (c) 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, DC, 1982; p 609.

(48) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 42.

(49) Schaftenaar, G. *QCPE Bull.* **1992**, *12*, 3.

(50) Westhof, E. *J. Biomol. Struct. Dyn.* **1987**, *5*, 581.

(51) Bowen, H. J. M.; Donohue, J.; Jenkins, D. G.; Kennard, O.; Wheatley, P. J.; Wiffen, D. H. In *Tables of Interatomic Distances and Configuration in Molecules and Ions*; Sutton, L. E., Jenkins, D. G., Mitchell, A. D., Cross, L. C., Eds.; The Chemical Society, 1958; pp M67, S7, S8, S15−S17.

(52) Schneider, B.; Cohen, D.; Berman, M. *Biopolymers* **1992**, *32*, 725.

(53) (a) York, D. M.; Darden, T.; Deerfield, D.; Pedersen, L. G. *Int. J. Quantum Chem.: Quantum. Biol. Symp.* **1992**, *19*, 145. (b) Guldbrand, L. E.; Forester, T. R.; Lynden-Bell, R. M. *Mol. Phys.* **1989**, *67*, 473. (c) Seeman, N. C.; Rosenberg J. M.; Suddath, F. L.; Kim, J. J. P.; Rich, A. *J. Mol. Biol.* **1976**, *104*, 109.

(54) Manning, G. S. *Acc. Chem. Res.* **1979**, *12*, 443.

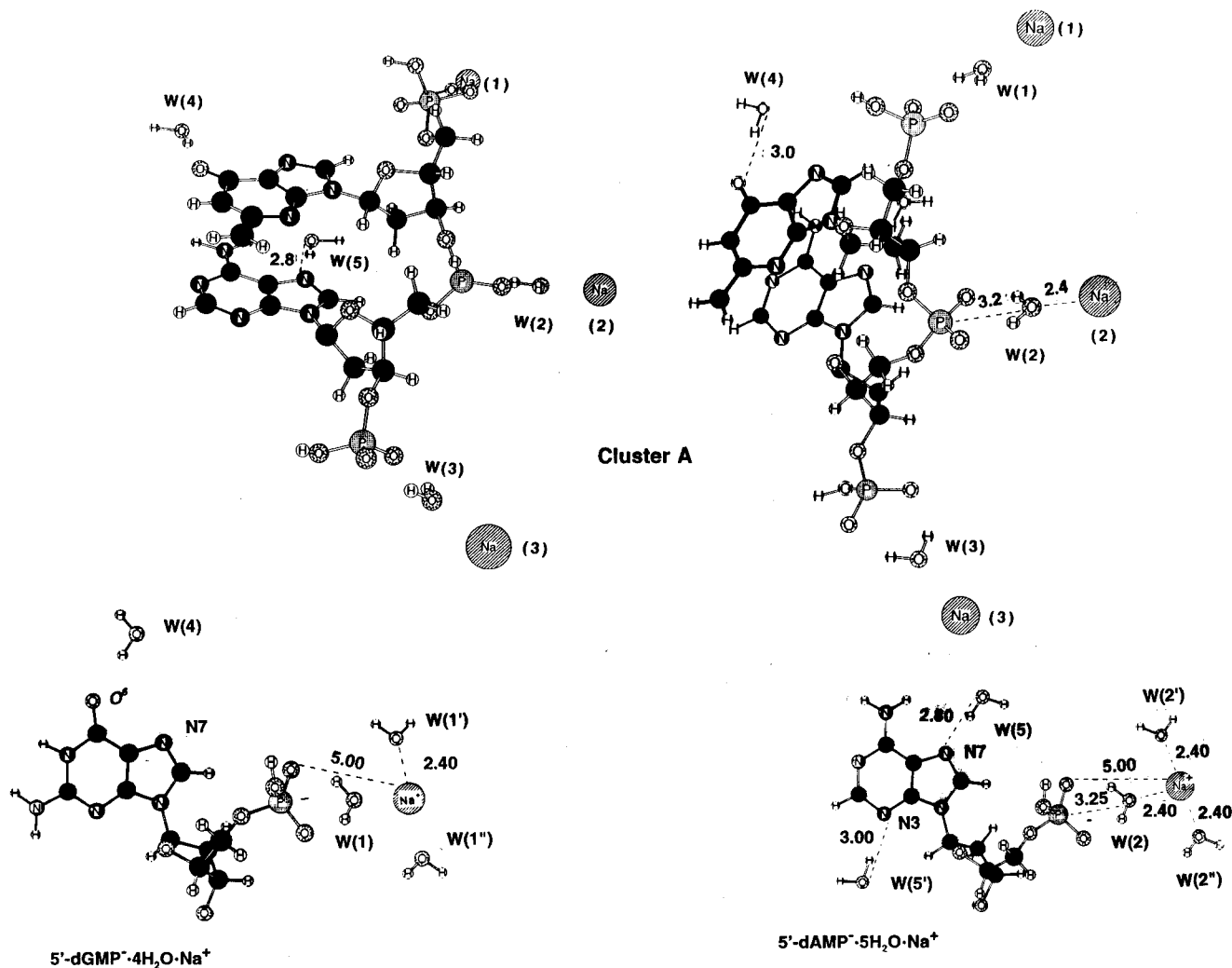(55) Flock, S.; Labarbe, R.; Houssier, C. *Biophys. J.* **1996**, *71*, 1519.

Figure 1. Structure of cluster **A** (above) containing the phosphorylated 2′-deoxydinucleotide ^{5}pGpAp with three Na$^+$ counterions and five water molecules. All distances are in Å. Two views are given. The left-hand view shows the orientation of the stacked bases with guanine above adenine, the orientation of the sugar phosphate backbone, and the positions of water molecules W(2)−W(5). The right-hand view shows the positions of water molecule W(1) and of the three Na$^+$ ions. Structures of the clusters 5′-dGMP$^-$·4H$_2$O·Na$^+$, and 5′-dAMP$^-$·5H$_2$O·Na$^+$ are shown at the lower left and right, respectively.

separate 6-31G SCF optimization calculation on a cluster that models the interaction of Na$^+$ with an inner solvation shell containing six water molecules (6H$_2$O·Na$^+$).[15] In the optimization calculation on the 6H$_2$O·Na$^+$ cluster, Na$^+$ and the O atoms of the H$_2$O molecules were restricted to octahedral symmetry and the overall symmetry with H atoms was restricted to D_{2h}.

In cluster **A**, the local geometry of water molecule W(1) relative to the Na$^+$ ion and the P atom of the phosphate group with which it interacts was obtained from an earlier, partial geometry optimization calculation at the 6-31G SCF level.[15] Here, a fixed geometry cluster containing three water molecules and Na$^+$ (3H$_2$O·Na$^+$) was docked to a fixed geometry component of cluster **A** containing 5′-dGMP$^-$ and W(4) (5′-dGMP$^-$·H$_2$O). In the docking calculation, the geometry of the 3H$_2$O·Na$^+$ cluster was the same as that shown in Figure 1 for Na$^+$, (W1), W(1′), and W(1″) in 5′-dGMP$^-$·4H$_2$O·Na$^+$, as well as that associated with Na$^+$ and three of the water molecules in the 6H$_2$O·Na$^+$ cluster.[10,15] In docking the 3H$_2$O·Na$^+$ cluster to the 5′-dGMP$^-$·H$_2$O cluster, the local symmetry of the negatively charged O atoms of the phosphate group, Na$^+$, W(1), W(1′), and W(1″) was restricted to C_{2v}. Finally, in cluster **A**, the local phosphate−water orientations of W(2) and W(3) are the same as that of W(1).

The geometry of the cluster containing ^{5}pGpA, two Na$^+$ ions, and four water molecules (^{5}pGpA·4H$_2$O·2Na$^+$) was obtained from that for cluster **A** by replacing the terminal 3′-phosphate group with an OH group and removing W(3) and Na$^+$(3).

Determination of Free Energies for Aqueous Nucleotide Ionization. Gibbs free energies for aqueous ionization ($\Delta G_{\mathrm{ION_{aq}}}$) of ^{5}pGpAp, ^{5}pGpA, 5′-dGMP$^-$, and 5′-dAMP$^-$ of cluster **A** and of ^{5}pGpA·4H$_2$O·2Na$^+$, 5′-dGMP$^-$·4H$_2$O·Na$^+$, and 5′-dAMP$^-$·5H$_2$O·Na$^+$ were evaluated using eq 3.[1,2,10]

$$\Delta G_{\mathrm{ION_{aq}}} \approx \mathrm{IP_{corr}} + \Delta\Delta G_{\mathrm{HYD}} + V_0 \qquad (3)$$

$$\Delta\Delta G_{\mathrm{HYD}} = \Delta G_{\mathrm{HYD\,2}} - \Delta G_{\mathrm{HYD\,1}} \qquad (4)$$

In eq 3, IP$_{corr}$ is the gas-phase ionization potential associated with removal of an electron from a specific nucleotide or nucleotide cluster orbital. Here, the entropy contribution to gas-phase ionization is taken to be negligible.[1,56] $\Delta\Delta G_{\mathrm{HYD}}$ is given by eq 4, where $\Delta G_{\mathrm{HYD\,1}}$ is the Gibbs free energy of hydration of the nucleotide before ionization and $\Delta G_{\mathrm{HYD\,2}}$ is the hydration energy after ionization. V_0 is the energy difference between an electron in aqueous solution at the conduction-band-edge and an electron in the gas phase.[57−59] Different values of V_0 ranging between −1.3 and −0.12 eV have been reported, the most recent of which are greater than or equal to −0.75 eV.[59] In the present investigation a value of −1.3 eV has been employed. It is possible

(56) Pearson, R. G. *J. Am. Chem. Soc.* **1986**, *108*, 6109.
(57) Grand, D.; Bernas, A.; Amouyal, E. *Chem. Phys.* **1979**, *44*, 73.
(58) Goulet, T.; Bernas, A.; Ferradini, C.; Jay-Gerin, J.-P. *Chem. Phys. Lett.* **1990**, *170*, 492.

that the more negative value of V_0 employed here corresponds to an extension into the energy range of trapped states lying below the conduction-band-edge.[58] However, in test studies, where this choice of V_0 and eq 3 were used to evaluate the conduction-band-edge ionization thresholds of indole and tryptophan from gas-phase IPs, the aqueous thresholds obtained from gas-phase data differ from the experimental aqueous ionization thresholds by less than 0.2 eV.[2,57]

Evaluation of Gibbs Free Energies of Hydration. Values of $\Delta G_{HYD\,1}$ and $\Delta G_{HYD\,2}$ were calculated for isolated nucleotides and nucleotide clusters by employing the noninterative Langevin dipole solvation model and the ChemSol 1.0 program.[60,61] In cases where tested, the version of the Langevin dipole model used here provides Gibbs free energies of hydration which agree well with experimental values for models of nucleotide components. When employed with charge distributions based on results from 6-31G* SCF and 6-31+G** MP2 calculations, the method provides hydration energies for neutral heterocycles which differ from experiment by no more the 3.0 kcal/ mol. For $H_2PO_4^-$, the method provides a value that is within experimental uncertainty.[61] This feature made the method well suited for the present investigation. The atomic charges of the nucleotides and nucleotide clusters before ionization, which are needed to calculate $\Delta G_{HYD\,1}$ were determined from potential-derived charges[62] for the ground states of closed-shell nucleotides or nucleotide clusters obtained from 3-21G SCF results. Hydration energies after ionization, $\Delta G_{HYD\,2}$, were obtained by modifying the atomic charges of the closed-shell nucleotide anions.[1,2,10]

To treat ionization at phosphate groups, the modified atomic charges used to calculate $\Delta G_{HYD\,2}$ were obtained by subtracting the charge of a single electron distributed over the nucleotide phosphate atoms from the charge distribution of the closed-shell nucleotide or nucleotide cluster. For 5'-dGMP$^-$ and $^{5'}$pGpAp, the distribution of the electron removed from the phosphate with 5' linkage to 2'-deoxyguanosine (dGuo) was obtained by subtracting the atomic charge distribution for the ground state of the isolated gas-phase radical, 5'-dGMP$^\bullet$, from the charge distribution of 5'-dGMP$^-$. The charge distribution of 5'-dGMP$^\bullet$ was obtained from an open-shell 3-21G SCF calculation on the ground state of 5'-dGMP$^\bullet$. For 5'-dAMP$^-$, the charge distribution of the electron removed via phosphate ionization was similarly obtained. The charge distribution for electrons removed via ionization of the phosphate group with 3' linkage to 2'-deoxyadenosine (dAdo) in $^{5'}$pGpAp was obtained by subtracting the charge distribution for the ground state of 3'-dAMP$^\bullet$ from the charge distribution of 3'-dAMP$^-$. In 5'-dGMP$^-$, 5'-dAMP$^-$, and 3'-dAMP$^-$, the highest occupied molecular orbitals (HOMOs) are O atom lone-pair orbitals on the anionic phosphate groups.

For $^{5'}$pGpAp and $^{5'}$pGpA, the charge distribution for electrons removed from the phosphate group with diester linkage was obtained from calculations on a phosphate diester model anion containing two sugar groups but no bases. Here, one sugar group, 2',5'-dideoxyribose, has 3' phosphate linkage; the second, 2'-deoxyribose, has 5' phosphate linkage. The differences in charge distributions before and after ionization were obtained by subtracting the charges of the ground-state radical from the charges of the closed-shell anion. The same difference was subtracted from the atomic charges of the closed-shell dinucleotides $^{5'}$pGpAp and $^{5'}$pGpA.[63]

To treat base and sugar ionization of the nucleotides and nucleotide clusters, the charge differences used to calculate $\Delta G_{HYD\,2}$ were obtained using a method similar to that used in calculating charge differences in the phosphate groups.[64]

Results

Figure 2 shows gas-phase He I photoelectron (PE) spectra of the model compounds 1,9-dimethylguanine (1,9-Me$_2$G), 9-methyladenine (9-MeA), and 3-hydroxytetrahydrofuran (3-OH-THF). The figure also shows vertical IPs associated with the seven highest occupied orbitals in 1,9-Me$_2$G (G$_1$ to G$_7$) and 9-MeA (A$_1$ to A$_7$), and with the two highest occupied orbitals in 3-OH-THF (S$_1$ and S$_2$). Assignment of the PE spectra of the three model compounds was carried out earlier.[13,16,65] For 1,9-Me$_2$G, the second through fifth and sixth and seventh bands are unresolved and lie in broad energy regions of high photoelectron intensity between 9.5 and 10.0 eV and around 11.0 eV. For 9-MeA, the second and third bands are unresolved and lie between 9.3 and 9.5 eV, and the fourth and fifth bands are unresolved and lie in the energy region 10.1−10.3 eV. In addition to vertical IPs, Figure 2 gives adiabatic IPs associated with the lowest energy bands of 1,9-Me$_2$G, 9-MeA, and 3-OH-THF. The first adiabatic IPs of 1,9-Me$_2$G and 9-MeA, the first and second vertical IPs of 3-OH-THF, and the second through seventh vertical IPs of 1,9-Me$_2$G and 9-MeA are shown together on an energy level diagram in the center of the figure. The shaded areas and the areas containing diagonal lines represent unresolved energy regions in the spectra of 1,9-Me$_2$G and 9-MeA, respectively.

In Figure 2, the assignment of the spectrum of 1,9-Me$_2$G is consistent with an earlier assignment based on results from semiempirical HAM/3 configuration interaction (CI) calculations.[66] The CI calculations indicate that of the first seven photoelectron bands in the spectrum of 1,9-Me$_2$G Koopmans' theorem gives a qualitatively accurate description of five. However, a single-orbital, Koopmans' description of ionization is qualitatively inadequate for two ionization events. Here, an accurate description of ionization must include hole-mixing associated with the mixing of configurations arising from removal of electrons from the second and third highest occupied π orbitals, G$_2$ and G$_4$ in Figure 2. Results from the HAM/3 CI calculations indicate that cation states arising from hole-mixing lie in a broad unresolved region of the spectrum between 9.5

(59) (a) Bernas, A.; Ferradini, C.; Jay-Gerin, J.-P. *Photochem. Photobio. A: Chem.* **1998**, *117*, 171. (b) Coe, J. V.; Earhart, A. D.; Cohen, M. H.; Hoffman, G. J.; Sarkas, H. W.; Bowen, K. H. *J. Chem. Phys.* **1997**, *107*, 6023. (c) Bernas, A.; Ferradini, C.; Jay-Gerin, J.-P. *Chem. Phys.* **1997**, *222*, 151.

(60) (a) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161. (b) Warshel, A.; Åqvist, J. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 267.

(61) Florián. J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583.

(62) Mishra, P. C.; Kumar, A. In *Molecular Electrostatic Potentials, Concepts and Applications*; Murray, J. S., Sen, K., Eds.; Elsevier: New York, 1996; pp 262−267, and references therein.

(63) In calculating $\Delta G_{HYD\,2}$ for ionization of the phosphate with diester linkage in $^{5'}$pGpAp and $^{5'}$pGpA, the charge differences occurring at guanine N9 (0.005 e) and at the P atom of the phosphate at the 5' end (0.003 e) were taken to be the same as that of the H atoms at the 1' and 5' positions, respectively, of the 2',5'-dideoxyribose group in the model anion. The charge difference at adenine N9 (0.01 e) was taken to be the same as that of one of the H atoms at the 1' position of the 2'-deoxyribose group. The charge difference at the P atom of the phosphate at the 3' end of $^{5'}$pGpAp (0.02 e) was taken to be the same as that of the hydroxy H atom at the 3' position of the 2'-deoxyribose group.

(64) For base ionization, the charge differences were obtained by subtracting the charge of a single electron distributed over the base atoms. For guanine ionization, the distribution of the removed electron was obtained by subtracting the atomic charge distribution for the ground state of the radical cluster 5'-dGMP$^\bullet$·4H$_2$O·Na$^+$ from that of the corresponding ground-state, closed-shell cluster 5'-dGMP$^-$·4H$_2$O·Na$^+$. The geometry of the clusters is that given at the lower left of Figure 1. In 5'-dGMP$^-$·4H$_2$O· Na$^+$, the HOMO is a base orbital. For adenine ionization, the distribution of the removed electron was obtained in an analogous fashion using results from 3-21G SCF calculations on 5'-dAMP$^-$·5H$_2$O·Na$^+$ and 5'-dAMP$^-$·5H$_2$O· Na$^+$. The differences in charge distributions obtained for base ionization of 5'-dGMP$^-$·4H$_2$O·Na$^+$ and 5'-dAMP$^-$·5H$_2$O·Na$^+$ were used to describe differences in charge distributions associated with guanine and adenine ionization in $^{5'}$pGpA, $^{5'}$pGpAp, $^{5'}$pGpA·4H$_2$O·2Na$^+$, and cluster **A**. For sugar ionization, the modified charges used to calculate $\Delta G_{HYD\,2}$ were obtained by subtracting the charge of an electron distributed over the sugar atoms. The distribution of the removed electron was evaluated by subtracting the charge distribution for the ground state of the radical cation 3-OH-THF$^+$ from that of the ground state of 3-OH-THF.

(65) LeBreton, P. R.; Yang, X.; Urano, S.; Fetzer, S.; Yu, M.; Leonard, N. J.; Kumar, S. *J. Am. Chem. Soc.* **1990**, *112*, 2138.

(66) Yu, M.; Jiang, Q.; LeBreton, P. R. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1992**, *19*, 27.
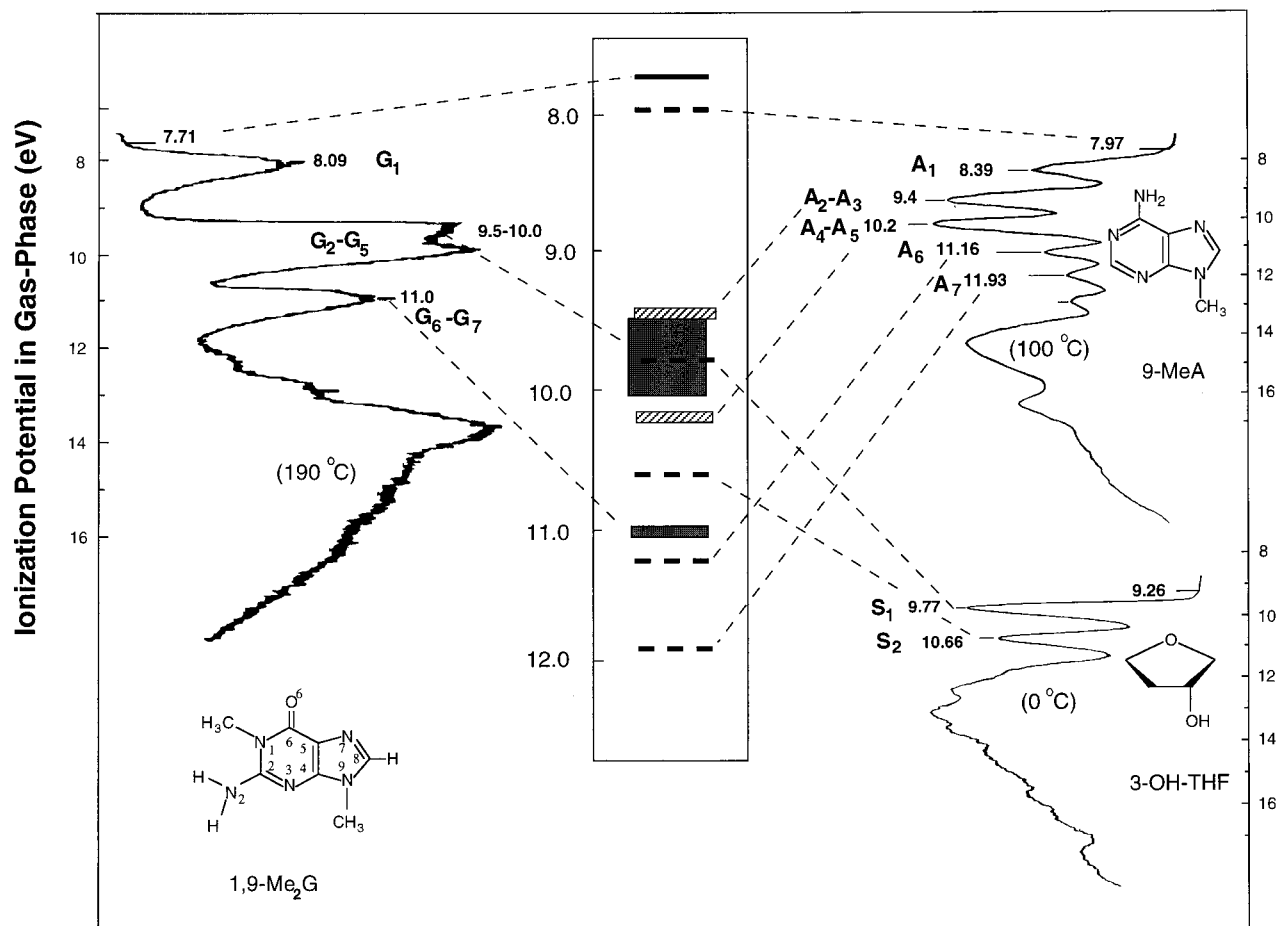
Figure 2. Gas-phase He I UV photoelectron spectra of the model compounds 1,9-dimethylguanine (1,9-Me$_2$G), 9-methyladenine (9-MeA), and 3-hydroxytetrahydrofuran (3-OH-THF). The figure shows vertical ionization potentials based on earlier assignments (refs 13, 65, and 16) associated with the seven highest occupied orbitals in 1,9-Me$_2$G and 9-MeA and the two highest occupied orbitals in 3-OH-THF. These are labeled G$_1$–G$_7$, A$_1$–A$_7$, and S$_1$ and S$_2$, respectively. In addition to vertical IPs, adiabatic IPs associated with the lowest energy photoelectron band are also given for each molecule. The figure contains an energy level diagram (center) in which all of the model compound ionization potentials are given together. The diagram gives adiabatic IPs associated with the G$_1$ and A$_1$ bands and vertical IPs associated with all of the other bands. Shaded regions in the energy level diagram correspond to energy regions in the spectrum of 1,9-Me$_2$G which contain unresolved bands and for which the vertical IPs are uncertain. Areas with diagonal lines correspond to similar energy regions in the spectrum of 9-MeA.

and 10.0 eV, which also contains bands arising from the G$_3$ and G$_5$ orbitals.

Figure 3 shows orbital diagrams and IPs associated with the upper occupied π and lone-pair orbitals of 1,9-Me$_2$G, 9-MeA, and 3-OH-THF which were obtained from 3-21G SCF calculations. The figure also gives vertical IPs obtained from He I photoelectron spectra. Finally, the figure gives experimental adiabatic IPs associated with the HOMOs. Figure 4 contains orbital diagrams and IPs associated with the four upper occupied lone-pair orbitals of H$_2$PO$_4^-$ which were obtained from 3-21G SCF calculations. The figure also gives the four lowest energy vertical IPs obtained from a combination of MP2/6-31+G* and CASPT2 calculations.[37]

Nucleotide Gas-Phase Ionization Potentials and Aqueous Ionization Thresholds. Results from SCF calculations on $^{5'}$pGpA and $^{5'}$pGpAp with the 3-21G basis set, like results from earlier split-valence basis set calculations on 5'-dGMP$^-$ and 5'-dAMP$^-$, indicate that many of the upper occupied orbitals are localized on individual base, sugar, and phosphate groups.[1,10,13] The electron distributions in most of the upper occupied orbitals are also similar to those associated with the model compounds 1,9-Me$_2$G, 9-MeA, and 3-OH-THF and with H$_2$PO$_4^-$. The upper part of Figure 5 shows gas-phase ionization potentials of 5'-dGMP$^-$, 5'-dAMP$^-$, $^{5'}$pGpA, and $^{5'}$pGpAp, which were obtained

by employing eqs 1 and 2 to correct 3-21G SCF results. The gas-phase IPs are given for both the isolated mononucleotides and phosphorylated dinucleotides and for the mononucleotides and dinucleotides in clusters with Na$^+$ ions and H$_2$O molecules.

For all of the nucleotides without Na$^+$ and H$_2$O, Figure 5 shows gas-phase adiabatic IPs associated with the highest occupied base orbitals and vertical IPs associated with the highest occupied phosphate orbitals. For $^{5'}$pGpA and $^{5'}$pGpAp, the highest occupied phosphate orbitals are on the phosphate at the 5' end and on the phosphate with diester linkage, respectively. In both cases, G$_1$ is the highest occupied base orbital. The values of the lowest base versus phosphate IPs of all of the isolated nucleotides, except 5'-dAMP$^-$, are nearly equal. For 5'-dAMP$^-$, the gas-phase vertical IP of the anionic phosphate group is 0.8 eV smaller than the adiabatic IP of the base. The results at the top of Figure 5 for the nucleotides without Na$^+$ and H$_2$O indicate that the IPs decrease as the number of phosphate groups increases. For 5'-dGMP$^-$, 5'-dAMP$^-$, $^{5'}$pGpA, and $^{5'}$pGpAp, the lowest energy phosphate IPs are 5.1, 5.2, 3.1, and 1.9 eV, respectively. In the clusters, each phosphate group has a strong electrostatic interaction with an Na$^+$ ion through an intermediary H$_2$O molecule. This electrostatic interaction causes a significant increase in all of the nucleotide gas-phase IPs. Because interaction is strongest
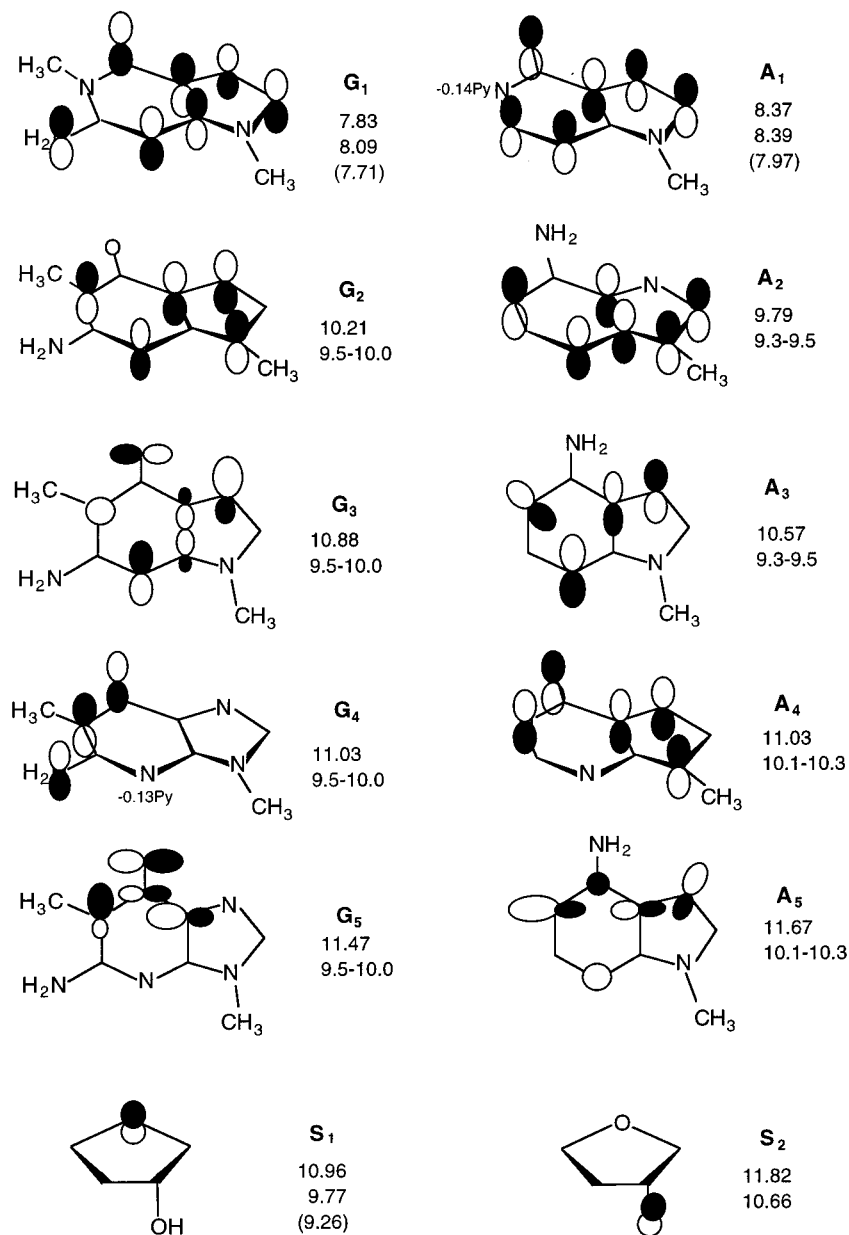
Figure 3. Orbital diagrams and ionization potentials associated with the upper occupied orbitals in 1,9-Me$_2$G (G$_1$−G$_5$), 9-MeA (A$_1$−A$_5$), and 3-OH-THF (S$_1$ and S$_2$). Ionization potentials obtained from SCF calculations with the 3-21G basis set are given without parentheses. For base π orbitals (G$_1$, G$_2$, G$_4$, A$_1$, A$_2$, and A$_4$) the viewing angle is different from that of base lone-pair orbitals (G$_3$, G$_5$, A$_3$, and A$_5$). Vertical IPs obtained from He I UV photoelectron spectra are given below the calculated IPs. Experimental adiabatic G$_1$, A$_1$, and S$_1$ ionization potentials are given in parentheses. Criteria used to draw the orbital diagrams are described in the text.
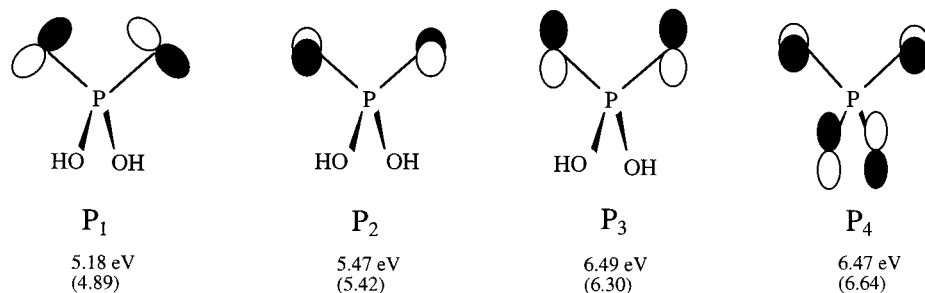
Figure 4. Diagrams of the four upper occupied orbitals in H$_2$PO$_4^-$ and vertical ionization potentials obtained from results of 3-21G SCF calculations. Values of vertical ionization potentials of H$_2$PO$_4^-$ obtained from MP2/6-31+G* and CASPT2 calculations on H$_2$PO$_4^-$ and H$_2$PO$_4^\bullet$ are given in parentheses.

between the Na$^+$ ions and the phosphate groups, the increase of the phosphate IPs in the clusters is larger than that of the base IPs.

The lower part of Figure 5 contains aqueous ionization energies associated with the lowest energy base and phosphate ionization events in the nucleotides not incorporated in clusters,
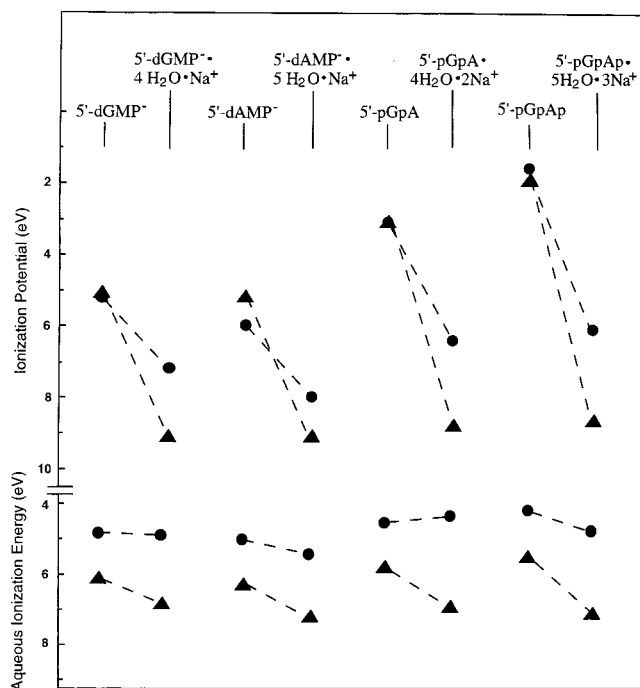
Figure 5. (Above) Corrected gas-phase ionization potentials associated with the highest occupied guanine and adenine orbitals (circles) and phosphate orbitals (triangles) in 5′-dGMP⁻ and 5′-dAMP⁻ and in the phosphorylated 2′-deoxydinucleotides, $^{5′}$pGpA and $^{5′}$pGpAp. Results, obtained from eqs 1 and 2, are shown for the isolated nucleotides and dinucleotides and for the clusters 5′-dGMP⁻·4H$_2$O·Na⁺, 5′-dAMP⁻·5H$_2$O·Na⁺, $^{5′}$pGpA·4H$_2$O·2Na⁺, and $^{5′}$pGpAp·5H$_2$O·3Na⁺ (cluster **A**). The geometries are described in Figure 1 and in the text. Dashed lines connect corresponding ionization events of nucleotides not incorporated in clusters with ionization events of nucleotides in clusters. (Below) Aqueous ionization energies, $\Delta G_{\text{ION}_{aq}}$, obtained from eqs 3 and 4 for the gas-phase ionization events given above.

in cluster **A** and in $^{5′}$pGpA·4H$_2$O·2Na⁺, 5′-dGMP⁻·4H$_2$O·Na⁺, and 5′-dAMP⁻·5H$_2$O·Na⁺. The ionization energies in the lower part of Figure 5 are significantly different from the gas-phase IPs in the upper part. These differences arise from a sum of contributions from $\Delta\Delta G_{\text{HYD}}$ (the difference between hydration energies before and after ionization) and V_0 (the difference between the energy of a vacuum electron and an electron at the conduction-band-edge or in trapped states slightly below the band edge). For the nucleotides without Na⁺, the absolute magnitude of the differences between the gas-phase and aqueous ionization energies shown in Figure 5 lies in the range 0.4−3.6 eV. For the clusters, the differences lie in the range 1.3−2.6 eV.

Figure 6 shows aqueous free energies of ionization associated with the 22 highest occupied orbitals of $^{5′}$pGpAp. The ionization energies were obtained from eqs 1−4. All except four of these ionization events are associated with orbitals that are largely located on a single base, sugar, or phosphate group. Figure 6 contains orbital diagrams obtained from 3-21G SCF calculations which describe the electron distributions for the upper occupied molecular orbitals (G$_1$ and G$_3$−G$_5$) with contributions mainly from atomic orbitals on guanine. The figure also shows molecular orbital diagrams (A$_1$ and A$_3$−A$_5$) with large contributions from atomic orbitals on adenine. The orbital diagrams in Figures 3 and 6 indicate that the G$_1$, G$_3$−G$_5$, A$_1$, and A$_3$−A$_5$ orbitals of the dinucleotide correlate closely with individual orbitals of the model compounds 1,9-Me$_2$G and 9-MeA. Small differences appearing in diagrams of correlating orbitals are due to the cutoff criteria. In cases where differences occur, coef-

ficients of atomic orbitals that are below the cutoff are given in the orbital diagrams of Figures 3 and 6. While not shown in Figure 6, the diagrams for the second highest occupied sugar orbitals (S$_2$ and S$_2$′) in dGuo and dAdo and the three highest occupied orbitals on each of the three phosphate groups in the dinucleotide are similar to diagrams in Figures 3 and 4 of corresponding orbitals in 3-OH-THF and in H$_2$PO$_4$⁻.

In Figure 6, the region containing vertical lines indicates the uncertainty in the aqueous ionization energies associated with the G$_3$−G$_5$ orbitals that arises because these bands are in an unresolved region of the PE spectrum of 1,9-Me$_2$G. Similarly, the regions containing closely spaced diagonal lines indicate the uncertainty in values of the aqueous ionization energies associated with the A$_2$−A$_5$ orbitals, which are in unresolved regions of the PE spectrum of 9-MeA.

Unlike most of the orbitals associated with the ionization events shown in Figure 6, four (G$_2$A$_2$, A$_2$G$_2$, S$_1$G$_3$, and S$_1$′P$_4$′) are significantly delocalized. Two of these, G$_2$A$_2$ and A$_2$G$_2$, are formed from a mixing of the G$_2$ and A$_2$ orbitals in 1,9-Me$_2$G and 9-MeA. The third, S$_1$G$_3$, is formed from a mixing of the S$_1$ orbital in 3-OH-THF associated with the cyclic O atom of dGuo and the G$_3$ orbital of 1,9-Me$_2$G. The fourth, S$_1$′P$_4$′, is formed from a mixing of the S$_1$ orbital in dAdo and the P$_4$ orbital in the phosphate group with diester linkage.

In Figure 6, the evaluation of the aqueous ionization energies associated with the G$_2$A$_2$ and A$_2$G$_2$ orbitals was carried out by employing eqs 1−4 in two ways. In the first, the values of the gas-phase IPs were corrected (eq 1) using ΔIP(G$_2$) obtained from 1,9-Me$_2$G, and $\Delta\Delta G_{\text{HYD}}$ was evaluated for guanine ionization in cluster **A**. In the second, the gas-phase IPs were corrected using ΔIP(A$_2$) obtained from 9-Me$_2$A, and $\Delta\Delta G_{\text{HYD}}$ was evaluated for adenine ionization in cluster **A**. The differences between the values of the G$_2$A$_2$ and A$_2$G$_2$ aqueous ionization energies obtained from these two methods (<0.4 eV) is of approximately the same magnitude as the uncertainties in the experimental gas-phase IPs associated with the G$_2$ and A$_2$ orbitals (0.5 and 0.2 eV, respectively). The uncertainties in the G$_2$A$_2$ and A$_2$G$_2$ aqueous ionization energies are indicated in Figure 6.

The evaluation of the S$_1$G$_3$ and S$_1$′P$_4$′ aqueous ionization energies was also carried out using methods that were analogous to those employed in the evaluation of the G$_2$A$_2$ and A$_2$G$_2$ orbitals. For the S$_1$G$_3$ orbital, the uncertainty in the ionization energy, indicated by the difference between the results obtained from the two alternate methods, is represented by the square hatched energy region in Figure 6. For the S$_1$′P$_4$′ orbital, the uncertainty is indicated by the region containing the widely spaced diagonal lines. Here the large uncertainty (1.2 eV) primarily reflects the difference between the corrections (ΔIP(S$_1$) = −1.19 eV and ΔIP(P$_4$) = 0.17 eV) of the gas-phase ionization potentials.

Molecular Electrostatic Potential of 5′-dGMP⁻. In an investigation of the electrostatic environment surrounding 5′-dGMP⁻, the molecular electrostatic potential obtained from the 3-21G SCF wave function was calculated and is given in Figure 7. The figure shows contour maps associated with two planes. One contains the guanine ring atoms. The second contains the P atom and the negatively charged O atoms of the phosphate group.

Figure 7 indicates that, in the plane of the base, electrostatic minima occur in the region of the N7, N3, and O^6 atoms. The magnitudes of the minima decrease in the order N7 < O^6 < N3. This ordering is the same as that found in earlier ab initio
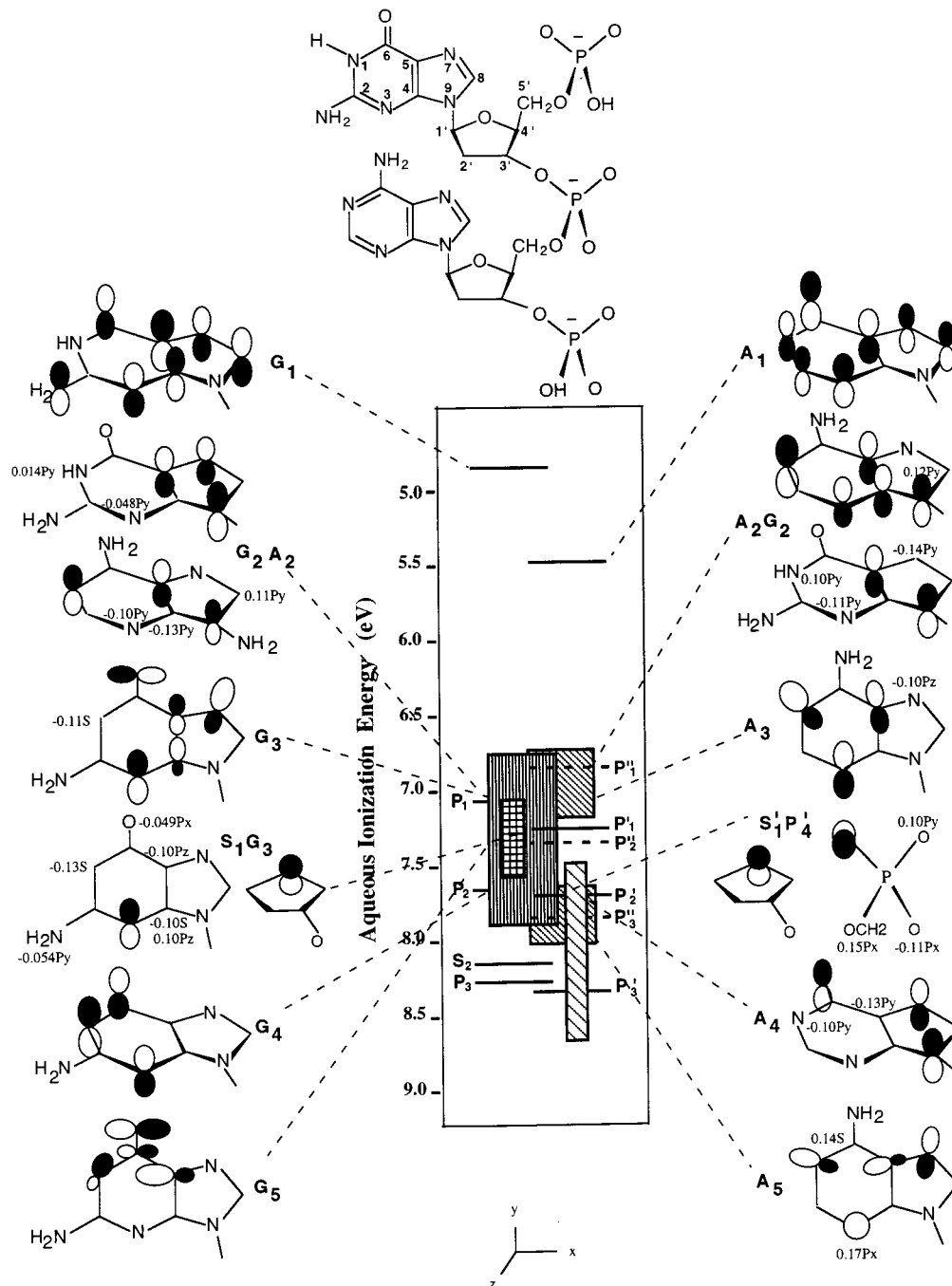
Figure 6. Aqueous ionization energies associated with the 22 highest occupied molecular orbitals of $^5{}'pGpAp$ in cluster **A**. Orbitals that are localized primarily on the guanine, adenine, sugar, and the phosphate groups are designated G, A, S, and P, respectively. Diagrams are given for molecular orbitals with large contributions from atomic orbitals on the bases or on the cyclic oxygen atom in the sugar groups. The orbital diagram associated with the second highest occupied sugar orbital, S_2, which is localized on 2′-deoxyguanosine and which correlates with the S_2 orbital of 3-OH-THF (Figure 2), is not given. Neither are diagrams associated with the phosphate orbitals P_1-P_3, $P_1''-P_3''$, and $P_1'-P_3'$ which correlate with corresponding orbitals in $H_2PO_4^-$ (Figure 3) and which are localized on the phosphates at the 5′ and 3′ ends and the phosphate with diester linkage, respectively. The hatched area and the areas containing vertical and diagonal lines indicate uncertainties in aqueous ionization energies due to unresolved bands in the gas-phase PE spectra of 1,9-Me$_2$G and 9-MeA, to hole-mixing in the spectrum of 1,9-Me$_2$G, and to delocalization of the G_2A_2, A_2G_2, S_1G_3, and $S_1'P_4'$ orbitals in the dinucleotide. See text.

SCF investigations of the electrostatic potential of isolated guanine[32] and of guanine in the guanine−cytosine base pair.[33a] Results from 6-31G SCF calculations indicate that the minimum electrostatic potential at N7 of guanine in the guanine−cytosine base pair is −95.5 kcal/mol. A 3-21G SCF calculation on isolated guanine, performed here, yielded a minimum of −96.6 kcal/mol at N7. For 5′-dGMP$^-$ in Figure 7, the minimum at N7 (−153 kcal/mol) is significantly more negative due to the

anionic phosphate group, which causes a decrease of the electrostatic potential in the entire region surrounding the nucleotide.

The results in Figure 7 demonstrate that, in addition to the minima occurring in the base region, two minima occur near the O atoms (**A** and **B**), where the electrostatic potential is −206 and −198 kcal/mol, respectively. The deeper potential minima at the phosphate group compared to the base group correlate
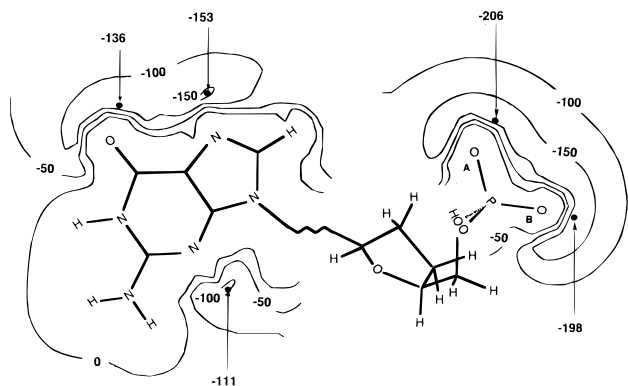
Figure 7. Molecular electrostatic potential (MEP) contour map of 5′-dGMP⁻ showing equipotential contours in two planes. One contains the heavy atoms of guanine. The second contains the P atom and the two negatively charged O atoms of the phosphate group. Contours are given in kcal mol⁻¹ e⁻¹. For the two planes examined, dark circles indicate minima in the electrostatic potential. The geometry of 5′-dGMP⁻ employed is the same as that in the 5′-dGMP⁻·4H₂O·Na⁺ cluster of Figure 1.

with atomic charge. According to the 3-21G SCF results, the most negative charge on a base atom is 0.82 e, while the charges on the anionic O atoms of the phosphate group are 0.93 and 0.97 e.

Discussion

The uncertainties in the aqueous ionization energies arising from the poor resolution of some bands in the PE spectra of 1,9-Me₂G and 9-MeA and from the delocalization of some of the nucleotide orbitals in 5′pGpAp (G₂A₂, A₂G₂, S₁G₃ and S′₁P′₄) are indicated in Figure 6. In cases where spectral resolution or orbital delocalization does not introduce uncertainty, the estimated average error in the aqueous ionization energies in Figure 6 due to the evaluation method employed is 0.5 eV. The same uncertainty is estimated for the results in Figure 5. This uncertainty includes the error introduced when the correction of gas-phase nucleotide ionization potentials is carried out using IPs from well-resolved photoelectron bands of the model compounds. In tests where the correction method (eqs 1 and 2) was applied to the five lowest energy IPs of five different methyl-substituted uracils, the average difference between the corrected IPs and the experimental IPs was 0.12 eV.[11]

For the model anion, $H_2PO_4^-$, experimental gas-phase IPs have not been measured. However, the results in Table 1 indicate that the differences between experimental first ionization potentials of P- and O-containing anions and the IPs obtained from the MP2/6-31+G* calculations employed here are less than 0.3 eV for all anions except HCO^-. A comparison[37] of the differences between the second through fourth lowest energy IPs of $H_2PO_4^-$, obtained using CASPT2 results based on CASSCF reference wave functions with different numbers of active orbitals and with results from calculations employing coupled cluster methods and many-body perturbation theory, provides evidence that the uncertainty in the gas-phase IPs associated with the P_2-P_4 orbitals is also less than 0.5 eV.

An estimate of 0.5 eV for the average error in the nucleotide aqueous ionization energies is consistent with expected errors in the values of $\Delta\Delta G_{HYD}$. For Na^+, NH_4^+, $N(CH_3)H_3^+$, Cl^-, CH_3O^-, and $C_6H_5O^-$, the differences between the values of ΔG_{HYD} obtained using the present method and the experimental values are 0.004−0.42 eV.[56,61,67] It is likely that errors in $\Delta\Delta G_{HYD}$ are of similar magnitude.

The aqueous ionization energies of the mononucleotides and dinucleotides reported in the lower part of Figure 5 and in Figure 6 were evaluated under conditions in which there is bulk solvent relaxation. The average time for complete solvent relaxation (∼540 fs) is longer[68] than that for electron localization in weakly bound solvation states (∼160 fs),[69] and hydration kinetics may influence observed ionization threshold energies. However, the relaxation of polar solvents around solute molecules occurs on multiple time scales, some of which are in the range of 50 fs.[70] Rapid solvent relaxation may take place for some ionization events occurring at the fully relaxed threshold energy. This would permit the experimental observation of aqueous ionization thresholds near the energies given in Figures 5 and 6 and accounts for the agreement between experimental ionization thresholds for indole and tryptophan and thresholds evaluated using the methods employed here.[2]

The present results provide evidence that solvation significantly influences the relative energies of different nucleotide ionization events. For nucleotides without Na^+, the results in Figure 5 indicate that solvation increases phosphate ionization energies compared to base ionization energies. This is explained by the difference between $\Delta\Delta G_{HYD}$ values associated with base versus phosphate ionization. For phosphate ionization, which gives rise to a final state with an uncharged phosphate group, there is significantly less charge separation than in the initial state. Here, the hydration energy (ΔG_{HYD2}) is less negative than the hydration energy of the final state arising from base ionization, where both the base and phosphate are charged.[1,2,10] According to Figure 5, the difference between the charge distributions of the final states arising from base and phosphate ionization significantly influences the relative energies of base versus phosphate ionization events in solution. Solvation increases the energetic favorability of base ionization relative to phosphate ionization. For all of the nucleotides without Na^+, the lowest gas-phase phosphate IPs are nearly equal to or smaller than the lowest base IPs.[1,10] In water, the smallest base ionization energies are more than 1.3 eV less than the smallest phosphate ionization energies. This is consistent with the finding that, in the condensed phase, photoionization and oxidation of nucleotides results in the removal of electrons from guanine and adenine.[23,25,71]

Figure 5 indicates that water relaxation also modulates the decrease in IPs which occurs as the number of phosphate groups increases. This modulation of the electrostatic influence of the anionic phosphate groups is caused by $\Delta\Delta G_{HYD}$ values that become more positive as the number of phosphate groups increases.[72] The lowest gas-phase IPs of the base and phosphate groups of 5′pGpA and 5′pGpAp are 2.0−3.6 eV smaller than the corresponding IPs of 5′-dGMP⁻. In solution these differences are reduced to 0.3−0.6 eV.

Finally, the results in Figure 5 demonstrate how solvation modulates the electrostatic influence of Na^+ interactions. For the mononucleotides and the dinucleotides, the difference between corresponding IPs, with and without Na^+, are in the

(67) Fernando, O. U. N. H. Ph.D. Thesis, The University of Illinois at Chicago, Chicago, IL, 1998.

(68) Reuther, A.; Laubereau, A.; Nikogosyan, D. N. *J. Phys. Chem.* **1996**, *100*, 16794.

(69) Pépin, C.; Goulet, T.; Houde, D.; Jay-Gerin, J.-P. *J. Phys. Chem. A* **1997**, *101*, 4351.

(70) Jimenez, R.; Fleming, G. R.; Kumar, P. V.; Maroncelli, M. *Nature* **1994**, *369*, 471.

(71) (a) Gregoli, S.; Olast, M.; Bertinchamps, A. *Radiat. Res.* **1977**, *72*, 201. (b) Gregoli, S.; Olast, M.; Bertinchamps, A. *Radiat. Res.* **1974**, *60*, 388.

(72) This reflects the exponential dependence of the hydration energy on charge. Born, M. *Z. Phys.* **1920**, *1*, 45.

ranges 2.0−4.4 eV and 3.9−6.7 eV for base and phosphate ionization events, respectively. In aqueous solution, the differences between base and phosphate ionization energies, with and without Na^+, are only 0.03−0.6 eV and 0.7−1.5 eV. In water, the reduction of the electrostatic influence of Na^+ occurs because of differences between the relative hydration energies of the initial and final states for nucleotides with Na^+ versus nucleotides without Na^+. For clusters, where the final state (after ionization) has a net charge (−1), hydration causes a greater stabilization of the final state relative to the initial state (before ionization) than for nucleotides without Na^+, where the final state has no net charge. For clusters, this yields $\Delta\Delta G_{HYD}$ values that largely cancel the increases in gas-phase ionization energies due to the Na^+ electrostatic interaction.

The Relationship between Guanine π Ionization Energies and Sequence-Specific Reactivity with Carcinogenic Alkylating Agents. Early investigations of electronic influences on the selectivity of small electrophilic alkylating agents at different sites on DNA focused on the molecular electrostatic potential (MEP).[32,33b] The finding that, around DNA bases, the MEP is most negative in the region of the N7 atom of guanine[32,33a] is consistent with the selectivity of several small cationic alkylating agents for the N7 position of guanine.[26,28,32,33b,73] However, the difficulty associated with an electrostatic model of DNA site selectivity is that for reactions of small uncharged methylating and ethylating agents with S_N2 character, reaction at N7 of guanine is also most favored. In reactions of methyl and ethyl methanesulfonate (MeMs and EtMs) and dimethyl and diethyl sulfate (Me$_2$SO$_4$ and Et$_2$SO$_4$) yields at N7 of guanine are more than 4 times larger than at other base sites.[26,28] The electrostatic model of selectivity also fails to account for the low reactivity at the negatively charged O atoms of the phosphate groups. As Figure 7 indicates, the most negative electrostatic potential occurs in the region around phosphate. In *N*-methyl-*N*-nitrosourea (MNU) reactions with DNA and RNA, more than 80% of all alkylation occurs at the bases, while less than 20% occurs at phosphate.[26,28] The differences between experimental S_N2 reaction properties noted here and predictions based on an electrostatic model have been pointed out in a more general way in earlier work.[74] The electrostatic potential is more closely related to DNA alkylation reactions with greater S_N1 character. For example, reactions of *N*-ethyl-*N*-nitrosourea with DNA and RNA proceed through a reactive ethane diazonium ion via a mechanism with greater S_N1 character than MNU.[75] Here, 55%−65% of the nucleotide reaction occurs at phosphate.[26,28]

Earlier investigations[1,13] demonstrated that nucleotide base π ionization potentials correlate inversely with base reactivity for a number of carcinogenic methylating and ethylating reagents with significant S_N2 character. These include MeMs, EtMs, Me$_2$SO$_4$, Et$_2$SO$_4$, and MNU. As base π ionization potentials decrease, base reactivities increase. A reasonable explanation of the principle underlying this relationship is that the transition-state barrier heights are strongly influenced by the polarizability of the transition state. Here, the base π ionization energy is a marker of the ease with which electronic rearrangement occurs that is necessary to accommodate significant charge transfer in the transition state.[76] As base π ionization energies decrease, π polarizability and reactivity increase. The inverse relationship, invoked here, between π ionization energies and polarizabilities

Figure 8. Transition state for in-plane attack of a methane diazonium ion (CH$_3$N$_2^+$) at N7 of guanine and mechanism through which π polarizability in the highest occupied molecular orbital (HOMO) of guanine reduces activation barriers of transition states in which there is significant charge transfer.

Table 2. Guanine Adiabatic Ionization Potentials in Self-Complementary Model Oligonucleotide Hexamer Duplexes

sequence[a]	ionization potential[b,c]	
5′-dA$_1$G$_2$G$_3$C$_4$C$_5$T$_6$ (**I**)	6.36$_{G2}$	6.85$_{G3}$
5′-dA$_1$C$_2$C$_3$G$_4$G$_5$T$_6$ (**II**)	6.46$_{G4}$	6.82$_{G5}$
5′-dA$_1$C$_2$T$_3$A$_4$G$_5$T$_6$ (**III**)	6.78	
5′-dA$_1$C$_2$A$_3$T$_4$G$_5$T$_6$ (**IV**)	6.92	
5′-dA$_1$C$_2$G$_3$C$_4$G$_5$T$_6$ (**V**)	6.94$_{G3}$	6.98$_{G5}$

[a] Sequences of palindromic hexamer duplexes containing guanines for which reactivities at the N7 and O^6 atoms are reported. See ref 78. [b] In eV. [c] Obtained by employing eqs 1 and 2 to correct results from 3-21G SCF calculations on the hexamer combinations of the stacked pairs (9-MeG:1-MeC and 9-MeA:1-MeT) used to model the duplex oligonucleotide sequences. See text.

in a series of structurally related π systems is the same as that observed in aromatic hydrocarbons and hydrocarbon metabolites.[77] The way in which π polarizability enhances reactivity is illustrated schematically in Figure 8, which shows the S_N2 transition state for alkylation at N7 of guanine by a methane diazonium ion (CH$_3$N$_2^+$).

The current investigation of DNA components containing more than one base permits a consideration of the possibility that π polarizability is useful for understanding nucleotide electron-donating properties that influence sequence-specific DNA methylation. This possibility was examined by comparing experimental guanine methylation patterns with guanine IPs evaluated for models of different sequences of stacked base-pair hexamers. Table 2 lists five palindromic sequences of double-stranded hexamers containing guanines where reactivity at both the N7 and O^6 atoms was measured in reactions with MNU.[78]

Because the hexamer duplexes are large, guanine, gas-phase adiabatic IPs were examined in model hexamers containing stacked base pairs without the sugar phosphate backbone. The evaluation of the guanine ionization energies was carried out by employing eqs 1 and 2 to correct IPs obtained from results of 3-21G SCF calculations. In all cases, the results of the calculations indicated that the highest occupied molecular orbital (HOMO) is a π orbital that resides primarily (>90%) on a single guanine. The influence of the sugar groups was modeled by using methyl groups at positions where glycosidic bond formation occurs. The influence of energy terms associated with solvation ($\Delta\Delta G_{HYD} + V_0$) was not included. While these influences are significant, they are not expected to alter the relative ordering of the guanine IPs.

Geometries of the nucleoside model compounds, 9-methylguanine (9-MeG), 9-methyladenine (9-MeA), 1-methylcy-

(73) Hartley, J. A.; Forrow, S. M.; Souhami, R. L. *Biochemistry* **1990**, *29*, 2985.

(74) Brinck, T.; Murray, J. S.; Politzer, P. *Int. J. Quantum Chem.* **1993**, *48*, 73.

(75) Osterman-Golkar, S.; Ehrenberg, L.; Wachmeister, C. A. *Radiat. Bot.* **1970**, *10*, 303.

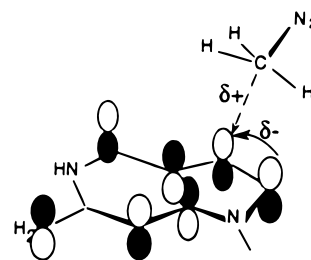(76) Ford, G. P.; Scribner, J. D. *Chem. Res. Toxicol.* **1990**, *3*, 219.

(77) Fetzer, S. M.; Huang, C.-R.; Harvey, R. G.; LeBreton, P. R. *J. Phys. Chem.* **1993**, *97*, 2385.

(78) Dolan, M. E.; Oplinger M.; Pegg, A. E. *Carcinogenesis* **1988**, *9*, 2139.
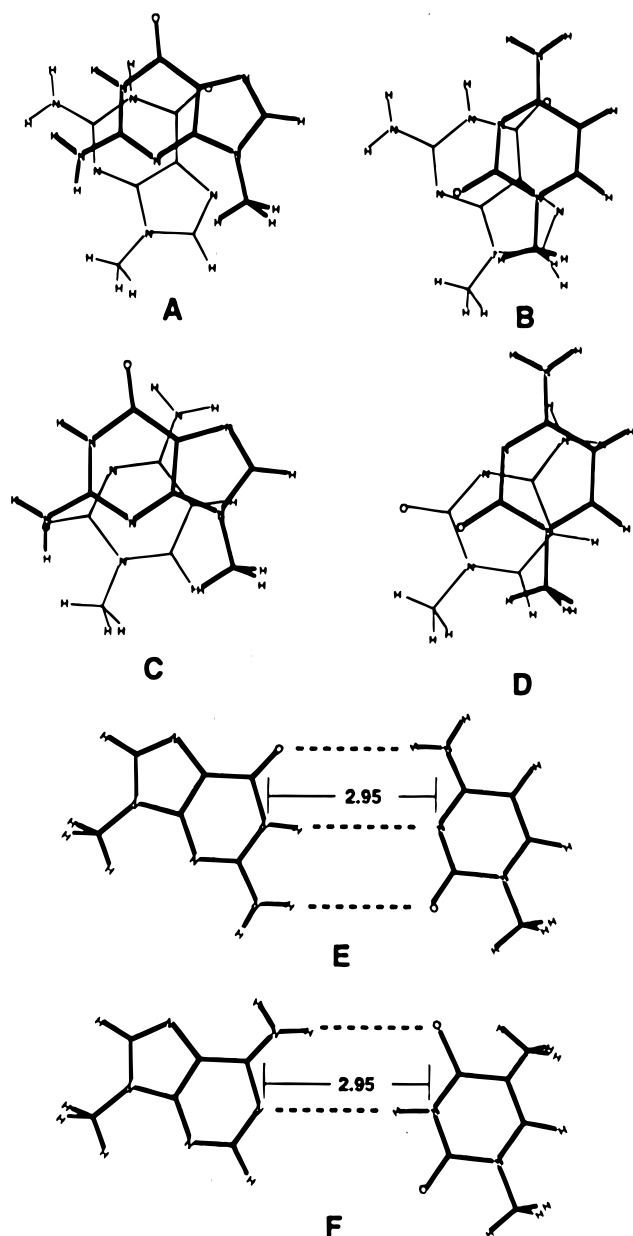
Figure 9. (Above) Projection diagrams showing orientations of purine-$5'p3'$-purine (A), pyrimidine-$5'p3'$-purine (B), purine-$5'p3'$-pyrimidine (C), and pyrimidine-$5'p3'$-pyrimidine (D) stacked bases employed in the evaluation of model oligonucleotide ionization potentials. In the diagrams, the model nucleosides, 9-methylguanine (9-MeG) and 1-methylcytosine (1-MeC), represent all stacked combinations of purines and pyrimidines. Viewed from above, the upper base (heavy lines) binds via 3′ linkage, through the phosphate, to the lower base, which has 5′ linkage. (Below) Base pair geometries (E and F).

tosine (1-MeC), and 1-methylthymine (1-MeT), were obtained from optimization calculations with the 3-21G basis set. The orientations of the model nucleosides in the base pairs, the stacking distance (3.32 Å), and the relative orientations of the stacked base pairs were obtained from average X-ray geometries for DNA reported in the Nucleic Acid Database.[79] Figure 9 shows ring orientations of the stacked purines and pyrimidines and the base pair geometries employed for the model hexamer duplexes examined. The coordinates employed in the model

hexamer calculations are given in the Supporting Information, Table S-I.

Table 2 lists adiabatic guanine IPs for the self-complementary hexamer sequences with geometries constructed using the base pair and the stacking orientations in Figure 9. For sequences **I**, **II**, and **V**, which contain two guanines, Table 2 lists two IPs. Figure 10 shows a plot of the natural log of guanine methylation yields for reactions with MNU, which proceed through a reactive $CH_3N_2^+$ ion, versus adiabatic guanine IPs in sequences **I–V**. Results are given for reactions at both the N7 and O^6 atoms of guanine. For sequences **I**, **II**, and **V**, which contain two guanines, Figure 10 gives results corresponding to the average yields per guanine and the average guanine adiabatic IPs. For reaction at O^6, the yields are, as expected,[26] smaller than at N7; however, the relationship between reactivity and guanine π ionization potentials is the same for reactions at both atoms. A linear relationship between the natural log of the yield and the π ionization potential would correspond to a reaction mechanism in which the activation barrier height increases linearly as the guanine IP increases. In fact, the fit of the results in Figure 10 is only approximately linear, demonstrating that a more refined description of the relationship between IP and reactivity is required. Nevertheless, reactivity at both N7 and O^6 decreases as the guanine π ionization potential increases.

In Table 2 the low guanine IPs in sequences **I** and **II** arise because guanine−guanine π stacking interactions cause a large splitting (0.4−0.5 eV) of the energies of the highest occupied guanine π orbitals. Stacking interactions that reduce guanine π ionization potentials are greatest in sequences containing G runs.[8] Results from measurements of sequence-specific DNA methylation by MNU and the related alkylating agent, nitrogen mustard, indicate that reaction at N7 and O^6 is more favorable for guanines in G runs than for isolated guanines.[33b,78,80] These observations are consistent with the results in Table 2 and Figure 10 and with the view that guanine π polarizabilities, which are related to π ionization potentials, strongly influence sequence-specific DNA alkylation.

A description of electrophilic attack of DNA by small charged S_N2 alkylating agents is emerging, which is represented by the free energy profile in Figure 11. While describing free energy for an aqueous reaction, the profile exhibits properties associated with gas-phase ion−molecule reactions. Specifically, gas-phase reactions have an attractive electrostatic well in the reactant channel.[81,82] In typical S_N2 ion−molecule reactions, this attractive interaction is reduced by solvent relaxation, and the well is absent in reactions occurring in aqueous solution.[82,83] However, because DNA is a polyanion, this attractive well is expected to play a significant role in reactions of cations with aqueous DNA. The occurrence of the potential well is consistent with the observed influence of cations on DNA folding[84,85] and aggregation[84,86] which depends on cation condensation at the DNA−water interface.[54] Nonreactive cations influence DNA reactivity with cationic alkylating agents via a similar mecha-

(79) Berman, H. M.; Olsen, W. K.; Beveridge, D. L.; Westbrook, J.; Belbin, A.; Demeny, T.; Hsieh, S. H.; Srinivasan, A. R.; Schneider, B. *Biophys. J.* **1992**, *63*, 751.

(80) (a) Wurdeman, R. L.; Church, K. M.; Gold, B. *J. Am. Chem. Soc.* **1989**, *111*, 6408. (b) Richardson, F. C.; Boucheron, J. A.; Skopek, T. R.; Swenberg, J. A. *J. Biol. Chem.* **1989**, *264*, 838.

(81) (a) Sharma, D. K. S.; Kebarle, P. *J. Am. Chem. Soc.* **1982**, *104*, 19. Morokuma, K. *J. Am. Chem. Soc.* **1982**, *104*, 3732. (c) Wolfe, S.; Mitchell, D. J.; Schlegel, H. B. *J. Am. Chem. Soc.* **1981**, *103*, 7694.

(82) Chabinyc, M. L.; Craig, S. L.; Regan, C. K.; Brauman, J. I. *Science* **1998**, *279*, 1882, and references therein.

(83) Bohme, D. K.; Mackay, G. I. *J. Am. Chem. Soc.* **1981**, *103*, 978.

(84) Marquet, R.; Houssier, C. *J. Biomol. Struct. Dyn.* **1991**, *9*, 159.

(85) Manning, G. S. *J. Biomol. Struct. Dyn.* **1989**, *7*, 41.

(86) (a) Bloomfield, V. A. *Biopolymers* **1991**, *31*, 1471. (b) Bloomfield, V. A.; Wilson, R. W.; Rau, D. C. *Biophys. Chem.* **1980**, *11*, 339. (c) Wilson, R. W.; Bloomfield, V. A. *Biochemistry* **1979**, *18*, 2192.
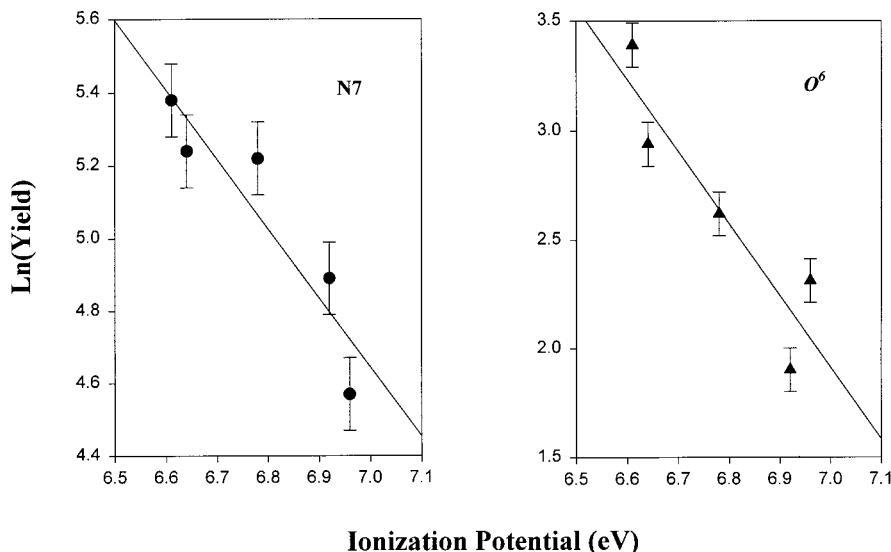
Figure 10. Plots of the natural log of guanine methylation yields (pmol adduct per μmol guanine) in reactions with *N*-methyl-*N*-nitrosourea (MNU) versus guanine adiabatic π ionization potentials in hexamer duplexes **I**−**V** of Table 2. Yields at N7 and O^6 were taken from data (ref 78) for dodecamers containing the hexamers examined. Error bars represent 10% uncertainty in the adduct yields. Solid lines represent linear least-squares fits to the results.
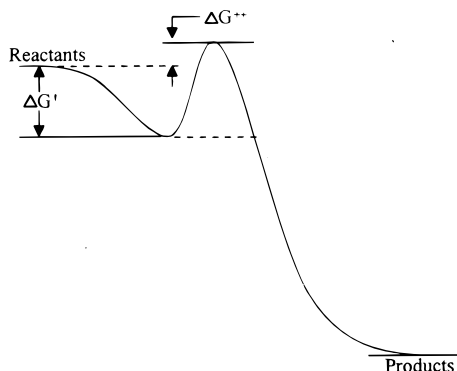

Figure 11. Schematic description of the Gibbs free energy reaction profile for the aqueous methylation of DNA bases by methane diazonium ions. The profile shows the attractive well arising from electrostatic interaction between cationic electrophiles and polyanionic DNA, for which the depth, $\Delta G'$, is diminished by increasing the ionic strength. The profile also shows the energy barrier ΔG^{++} which determines reaction selectivity and which is influenced by base π polarizability as indexed by π ionization potentials.

nism. In DNA reactions with cations, the condensation of nonreactive interfering cations reduces the well depth and inhibits reaction by inhibiting the formation of reactive complexes. Here, interfering nonreactive cations more directly influence long-range interactions on the reaction potential surface than short-range interactions that govern the detailed structure of the transition barrier. This description of the influence of nonreactive cations on S_N2 alkylation of DNA by cationic alkylating agents is consistent with the finding that increasing ionic strength,[87] adding cationic DNA affinity binding agents,[87] or incorporating tethered cations into DNA structure[88] inhibits DNA alkylation by MNU, but does not influence reactivity by neutral methylating agents such as Me_2SO_4.[88] It is also consistent with the finding that, while nonreactive cations inhibit the overall reactivity of MNU, nonreactive cations do not alter the relative reactivities at guanine N3, O^6, and N7, adenine N3, thymine O^2, and phosphate.[89]

The inhibition of reactions of cationic electrophiles with DNA by nonreactive cations does not rule out the importance of π polarizability on selectivity. The relationship between base π polarizability and the relative barrier heights at different reaction sites is consistent with the high base versus phosphate reactivity of small alkylating reagents with S_N2 character. It is also consistent with the high reactivity at guanine versus the other DNA bases and with the high reactivity occurring in sequences containing stacked guanines versus sequences containing guanine flanked by other bases.

Conclusions

The principal findings of this investigation are as follows.

1. In the phosphorylated dinucleotides, $^{5'}$pGpA and $^{5'}$pGpAp, residing in a B-DNA conformation and interacting with Na^+ counterions and H_2O molecules, most of the upper occupied orbitals are localized on base, sugar, and phosphate groups. In a cluster of $^{5'}$pGpAp containing three Na^+ ions and five water molecules, 18 of the 22 highest occupied orbitals correlate with individual orbitals associated with the model compounds and the model anion 1,9-Me_2G, 9-MeA, 3-OH-THF, and $H_2PO_4^-$.

2. In the gas phase, phosphate electrostatic interactions cause the IPs of $^{5'}$pGpA and $^{5'}$pGpAp to be significantly smaller than those of the mononucleotides, 5'-dGMP$^-$ and 5'-dAMP$^-$. For $^{5'}$pGpAp, the lowest energy guanine adiabatic and phosphate vertical IPs (1.6 and 1.9 eV, respectively) are 3.6 and 3.2 eV smaller than the lowest guanine and phosphate IPs in 5'-dGMP$^-$.

3. The decreases in gas-phase IPs associated with the buildup of nucleotide sequences is modulated by hydration. For $^{5'}$pGpAp in water without counterions, the lowest base and phosphate ionization energies (4.2 and 5.5 eV, respectively) are only 0.6 eV smaller than corresponding aqueous ionization energies in

(87) Wurdeman, R. L.; Gold, B. *Chem. Res. Toxicol.* **1988**, *1*, 146.

(88) Liang, G.; Encell, L.; Nelson, M. G.; Switzer, C.; Shuker, D. E. G.; Gold, B. *J. Am. Chem. Soc.* **1995**, *117*, 10135.

(89) (a) Weissler, M. In *N-Nitrosamines*; ACS Symposium Series 682; Anselme, J.-P., Ed.; American Chemical Society: Washington, DC, 1979; pp 57−75. (b) Jensen, D. E.; Reed, D. J. *Biochemistry* **1978**, *17*, 5098. (c) Briscoe, W. T.; Cotter, L.-E. *Chem.-Biol.* **1985**, *56*, 321. (d) Rajalakshmi, S.; Rao, P. M.; Sarma, D. S. R. *Biochemistry* **1978**, *17*, 4515. (e) Lawley, P. D.; Shah, S. A. *Biochem. J.* **1972**, *128*, 117. (f) Lawley, P. D.; Thatcher, C. J. *Biochem. J.* **1970**, *116*, 693. (g) Schultz, U.; McCalla, D. R. *Can. J. Chem.* **1969**, *47*, 2021. (h) McCalla, D. R. *Biochim. Biophys. Acta* **1968**, *155*, 114. (i) Kriek, E.; Emmelot, P. *Biochemistry* **1963**, *2*, 733.

5′-dGMP⁻. Furthermore, solvation alters the relative ionization energies of different nucleotide groups. While the lowest energy adiabatic guanine and vertical phosphate IPs are approximately equal for ⁵′pGpAp in the gas phase, the aqueous ionization energy of guanine is 1.3 eV smaller than the lowest phosphate ionization energy.

4. In the region surrounding nucleotides, the electrostatic potential is most negative around the anionic phosphate groups; however, electrophilic attack of DNA by small methylating and ethylating agents with significant S_N2 character, such as MNU, occurs primarily at the bases. This finding and the earlier observation that base reactivity increases as base π ionization potentials decrease provide evidence that the selectivity of alkylation reactions is influenced more by π polarizability, for which π ionization potentials are a marker, than by electrostatic potential.

5. Stacking interactions that occur in B-DNA sequences reduce base π ionization potentials. The smallest ionization energies occur for guanine in runs containing two or more stacked guanines. The low ionization energies in G runs correlate with increases in base reactivity toward methylating agents such as MNU. In different oligomer sequences, containing six base pairs, reaction of MNU at both N7 and O^6 occurs more favorably at guanine sites with low ionization energies.

Supporting Information Available: Base pair coordinates used in the 3-21G SCF calculations on the self-complementary hexamers listed in Table 2 are given in Table SI (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA992065B