DNA Photoionization and Alkylation Patterns in the Interior of Guanine Runs

Qiqing Zhu and Pierre R. LeBreton*

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

Received July 11, 2000

Abstract: Of the DNA bases, guanine has the smallest ionization potential (IP). The recent combined use of experimental gas-phase photoelectron data and self-consistent field (SCF) and post-SCF descriptions has provided new information about the energetics of guanine ionization at the nucleotide and dinucleotide levels (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.; Zhu, Q.; LeBreton, P. R. *J. Am. Chem. Soc.* **1999**, *121*, 11516−11530). The energetic matching of the highest occupied π orbitals and the π interaction which occurs in regions of stacked guanines (G runs) give rise to sequence-specific regions of low IP. This is described by a Koopmans analysis of results from ab initio SCF calculations with 6-31G*, 6-31G, 3-21G, and STO-3G basis sets and with ZINDO semiempirical calculations on oligonucleotides and oligonucleotide models. At the ab initio 3-21G SCF level, model calculations indicate that the lowest IP of guanine in a double-stranded G run containing three guanine−cytosine base pairs is approximately 1 eV smaller than that of free guanine. The ab initio and semiempririal results indicate that, in double-stranded oligonucleotides containing G runs with three or four guanines, an interior guanine has the lowest IP. This occurs for oligonucleotides in both the A- and B-DNA conformations, and for oligonucleotides with strong phosphate−counterion interactions. The finding that the lowest energy base ionization occurs from the interior of G runs differs from the result reported for a model employing a double-stranded G run without sugar and phosphate groups (Saito, I.; Nakamura, T.; Nakatani, K. *J. Am. Chem. Soc.* **2000**, *122*, 3001−3006) where the guanine at the 5′-end has the lowest IP. However, it is consistent with earlier results indicating that, in G runs, the molecular electrostatic potential is more negative around guanines in the interior than around guanines at the ends. The low interior IPs correlate with guanine two-photon ionization patterns. They also correlate with the high reactivities at interior sites exhibited by the electrophilic antitumor agent bis-2-cholroethylmethylamine (nitrogen mustard) and by the potent carcinogen *N*-methyl-*N*-nitrosourea.

Introduction

Much of the biochemistry and biophysics of DNA relies on the electron-donating properties of nucleotides, which, in the simplest sense, are reflected in ionization energies. For example, electron donation, expressed as the susceptibility of nucleotides to electrophilic attack, plays a ubiquitous role in mechanisms of chemical carcinogenesis and cancer chemotherapy.[1,2] Similarly, nucleotide ionization is an initiating step associated with radiation-induced DNA strand scission.[3,4] In addition to influencing mechanisms of DNA damage induced by electrophiles and ionizing radiation, nucleotide electron donation and ionization also play a central role in mechanisms responsible for electron transport in oligonucleotides.[5]

Gas-phase ionization energies for nucleotide bases were measured in early mass spectrometry experiments.[6] Photoelectron measurements of nucleotide components have yielded multiple gas-phase ionization potentials (IPs) of valence π and lone-pair orbitals in all of the DNA and RNA bases.[7,8] Photoelectron experiments were also carried out on sugar model compounds,[9] phosphate esters,[9a,10] and nucleoside analogues.[11]

Ionization potentials of nucleotide components have been evaluated in several recent theoretical investigations,[12−15] and the influence of water hydrogen bonding, base-pair hydrogen bonding,[14c,15] and base stacking[12,13] on base IPs has been examined. While direct experimental measurement of IPs in intact nucleotides has not yet been carried out, multiple valence

(1) (a) Swenson, D. H.; Lawley, P. D. *Biochem. J.* 1978, *171*, 575. (b) Krytopoulos, S. A.; Anderson, L. M.; Chhabra, S. K.; Souliotis, V. L.; Plesta, V.; Valavanis, C.; Georgiadis, P. *Cancer Detect. Prev.* **1997**, *21*, 391. (c) Lind, M. J.; Ardiet, C. *Cancer Surveys* **1993**, *17*, 157.

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

(3) (a) Melvin, T.; Botchway, S. W.; Parker, A. W.; O'Neill, P. *J. Am. Chem. Soc.* **1996**, *118*, 10031. (b) Gorner, H.; Gurzadyan, G. G. *J. Photochem. Photobiol., A: Chem.* **1993**, *71*, 155. (c) Schulte-Fronhlinde, D.; Simic, M. G.; Gorner, H. *Photochem. Photobiol.* **1990**, *52*, 1137.

(4) Malone, M. E.; Cullis, P. M.; Symons, M. C. R.; Parker, A. W. *J. Phys. Chem.* **1995**, *99*, 9299.

(5) (a) Dandliker, P. J.; Holmlin, R. E.; Barton, J. K. *Science* **1997**, *275*, 1465. (b) Holmlin, R. E.; Dandliker, P. J.; Barton, J. K. *Angew. Chem., Int. Ed.* **1997**, *36*, 2714.

(6) Lifschitz, C.; Bergmann, E. D.; Pullman, B. *Tetrahedron Lett.* **1967**, *46*, 4583.

(7) (a) Padva, A.; LeBreton, P. R.; Dinerstein, R. J.; Ridyard, J. N. A. *Biochem. Biophys. Res. Commun.* **1974**, *60*, 1262. (b) Hush, N. S.; Cheung, A. S. *Chem. Phys.* **1975**, *34*, 11. (c) Orlov, V. M.; Smirnov, A. N.; Varshavsky, Ya. M. *Tetrahedron Lett.* **1976**, *48*, 4377.

(8) Urano, S.; Yang, X.; LeBreton, P. *J. Mol. Struct.* 1989, *214*, 315 and references therein.

(9) (a) Tasaki, K.; Yang, X.; Urano, S.; Fetzer, S.; LeBreton, P. R. *J. Am. Chem. Soc.* **1990**, *112*, 538. (b) Kim, H. S.; Yu, M.; Jiang, Q.; LeBreton, P. R. *J. Am. Chem. Soc.* **1993**, *115*, 6169.

(10) (a) Cowley, A. H.; Lattman, M.; Montag, R. A.; Verade, J. E. *Inorg. Chim. Acta* **1977**, *25*, L151. (b) Chattopadhyay, S.; Findley, G. L.; McGlynn, S. P. *J. Electron. Spectrosc. Relat. Phenom.* **1981**, *6*, 27.

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

10.1021/ja002523c CCC: $19.00
Published on Web 12/06/2000

nucleotide ionization potentials have been evaluated using a combination of results from theory and experiment.[9,16−19] Early work focused on isolated nucleotides.[9] Recently, these investigations were expanded to nucleotides in more physiological environments containing water and counterions.[16−18]

The application of Koopmans' theorem[20] to results from ab initio self-consistent field (SCF) electronic structure calculations with 3-21G[21] or 6-31G[22] basis sets yields IPs for the highest occupied base orbital in 1,9-dimethylguanine that differ from the experimental gas-phase vertical IP by less than 0.3 eV.[16] Among the DNA and RNA bases, guanine has the lowest ionization potential, and for all of the bases the highest occupied molecular orbitals (HOMOs) are π orbitals.[8,23] Hydrogen bonding and base stacking, which is associated with $\pi\angle\pi$ interactions, influence ionization energies. For example, Becke3LYP/D95*//UHF/6-31G* density functional calculations[15,24] predict that the adiabatic ionization potential of guanine in a guanine−cytosine Watson−Crick base pair is 0.7−0.8 eV smaller than that of isolated guanine. Similarly, SCF calculations with a 6-31G* basis set[22b,25] and density functional calculations indicate that the vertical IP of two stacked 9-methylguanines (9-MeG's) at the distance and orientation occurring in B-DNA is 0.3−0.7 eV smaller than that of isolated 9-MeG.[12a,13] Interestingly, stacking influences on base π ionization energies exhibit sequence dependence. The energy resonance of occupied π orbitals that occurs in sequences of multiple stacked guanines (G runs) give rise to larger guanine π perturbations than those in stacks containing guanines and different DNA bases.[12a] The base IPs associated with G runs are the smallest known ionization potentials in DNA.

The influence of stacking interactions on guanine IPs has been examined by employing models of single- and double-stranded dinucleotides consisting of *N*-methylated bases in orientations that occur in B-DNA.[12,13,18] Here vertical ionization potentials of 9-methylguanine were evaluated in SCF 6-31G* calculations on single-stranded stacks containing as many as four 9-MeG's.[12b] Ionization potentials were also evaluated at the 6-31G* level for models containing two and three stacked 9-MeG/1-methylcytosine (1-MeC) base pairs in duplex 5-mers[12c] and at the 3-21G level[18] for model 6-mers containing one or two 9-MeG/1-MeC base pairs. Results obtained for the single- and double-stranded models of G runs indicate that the largest contribution to the highest occupied base orbital (HOBO) is from the 9-MeG at the 5′-end.[12a−c,18]

Earlier investigations have linked DNA ionization properties to DNA photolysis, photocleavage, oxidation, and alkylation chemistry.[12,13,16,18,26−29] The low IP of guanine has been related to the finding that, after relaxation times on the order of 5 μs, calf thymus DNA photoionization results in hole migration to guanine accompanied by deprotonation,[28] and that DNA double-strand scission by direct radiation and by photocleaving molecules occurs primarily at guanine sites.[12a,b,d,e,26,27,29] Considerable attention has been focused on the observation that the HOBO in models of G runs is concentrated on guanine at the 5′-end, and that photochemical DNA-cleaving agents preferentially induce strand breaks at the 5′-end.[12a,b,27]

Nucleotide base ionization potentials are related to the reactivities of bases toward small alkylating electrophiles with significant S_N2 character.[9,16,18,30] This includes antitumor drugs and carcinogenic alkylating agents such as bis-2-chloroethylmethylamine (nitrogen mustard), dimethyl sulfate, and *N*-methyl-*N*-nitrosourea (MNU)[2,31] that are both site and sequence specific. For these electrophiles, the most reactive site on DNA is the N7 atom of guanine.[2,32] For MNU, reaction at guanine O^6 is important to carcinogenic mechanisms.[2,33] The reactivity at N7 and O^6 of guanine in G runs is greater than that at guanine in sequences in which there is no nearest-neighbor interaction with other guanines.[33b,34−36] These characteristics of small alkylating electrophiles with significant S_N2 character are consistent with a reaction model in which site- and sequence-specific reactivities increase as nucleotide base polarizability increases.[16,18] Reactivity is related to ionization potentials because the relative polarizabilities of a series of structurally similar molecules generally increase as the ionization potentials decrease.[37]

Earlier examinations of sequence-specific DNA ionization potentials that focused on DNA model systems[12,13,18] provided

(12) (a) Sugiyama, H.; Saito, I. *J. Am. Chem. Soc.* **1996**, *118*, 7063. (b) Saito, I.; Takayama, M.; Sugiyama, H.; Nakamura, T. *J. Photochem. Photobiol., A: Chem* **1997**, *106*, 141. (c) Saito, I.; Nakamura, T.; Nakatani, K. *J. Am. Chem. Soc.* **2000**, *122*, 3001. (d) Saito, I.; Nakamura, T.; Nakatani, K.; Yoshioka Y.; Yamaguchi, K.; Sugiyama, H. *J. Am. Chem. Soc.* **1998**, *120*, 12686. (e) Yoshioka, Y.; Kitagawa, Y.; Takano, Y.; Yamaguchi, K.; Nakamura, T.; Saito, I. *J. Am. Chem. Soc.* **1999**, *121*, 8712.

(13) Prat, F.; Houk, K. N.; Foote, C. S. *J. Am. Chem. Soc.* **1998**, *120*, 845.

(14) (a) Colson, A.-O.; Besler, B.; Close, D. M.; Sevilla, M. D. *J. Phys. Chem.* **1992**, *96*, 661. (b) Colson, A.-O.; Besler, B.; Sevilla, M. D. *J. Phys. Chem.* **1993**, *97*, 8092. (c) Colson, A.-O.; Besler, B.; Close, D. M.; Sevilla, M. D. *J. Phys. Chem.* **1993**, *97*, 13852. (d) Sevilla, M. D.; Besler, B.; Colson, A.-O. *J. Phys. Chem.* **1995**, *99*, 1060. (e) Dolgounitcheva, O.; Zakrzewski, V. G.; Ortiz, J. V. *J. Am. Chem. Soc.*, in press.

(15) Hutter, M.; Clark, T. *J. Am. Chem. Soc.* **1996**, *118*, 7554.

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

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

(18) Kim, N. S.; Zhu, Q.; LeBreton, P. R. *J. Am. Chem. Soc.* **1999**, *121*, 11516.

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

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

(21) (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.

(22) (a) Hehre, W. J.; Ditchfield R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257. (b) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654.

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

(24) (a) Becke, A. D. *Phys. Rev.* **1988**, *A38*, 3098. (b) Parr, R. G.; Yang W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.

(25) Hariharan, P. C.; Pople, J. A. *Chem. Phys. Lett.* **1972**, *66*, 217.

(26) Melvin, T.; Plumb, M. A.; Botchway, S. W.; O'Neill, P.; Parker, A. W. *Photochem. Photobiol.* 1995, *61*, 584.

(27) Saito, I.; Takayama, M.; Sugiyama, H.; Nakatani, K.; Tsuchida, A.; Yamamoto, M. *J. Am. Chem. Soc.* **1995**, *117*, 6406.

(28) Candeias, L. P.; O'Neill, P.; Jones, G. D. D.; Steenken, S. *Int. J. Radiat. Biol.* **1992**, *61*, 15.

(29) Croke D. T.; Blau, W.; OhUigin, C.; Kelly, J. M.; McConnell, D. J. *Photochem. Photobiol.* **1988**, *47*, 527.

(30) (a) Lawley P. D. In *Chemical Carcinogenesis*; Searle, C. E., Ed.; ACS Monograph 182; American Chemical Society: Washington, DC, 1984; pp 325−484. (b) Hathway, D. E.; Kolar, G. F. *Chem. Soc. Rev.* **1980**, *9*, 241.

(31) Willman, D. E. V.; Connors, T. A. In *Topics in Molecular and Structural Biology*; Neidle, S. N., Waring, M. J., Eds.; Verlag Chemie: Weinheim, 1983; Vol. 3, pp 233−282.

(32) (a) Brookes, P.; Lawley, P. D. *Biochem. J.* **1961**, *80*, 496. (b) Lawley, P. D.; Thatcher, C. J. *Biochem. J.* **1970**, *116*, 693. (c) Margison, G. P.; O'Connor, P. J. *Biochim. Biophys. Acta* **1973**, *331*, 349. (d) Beranek, D. T.; Weis, C. C.; Swenson, D. H. *Carcinogenesis* **1980**, *1*, 595.

(33) (a)Goth, R.; Rajewsky, M. F. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 639. (b) Richardson, F. C.; Boucheron, J. A.; Skopek, T. R.; Swenberg, J. A. *J. Biol. Chem.* **1989**, *264*, 838.

(34) Mattes, W. B.; Hartley, J. A.; Kohn, K. W. *Nucleic Acids Res.* **1986**, *14*, 2971.

(35) Dolan, M. E.; Oplinger M.; Pegg, A. E. *Carcinogenesis* **1988**, *9*, 2139.

(36) Wurdeman, R. L.; Church, K. M.; Gold, B. *J. Am. Chem. Soc.* **1989**, *111*, 6408.

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

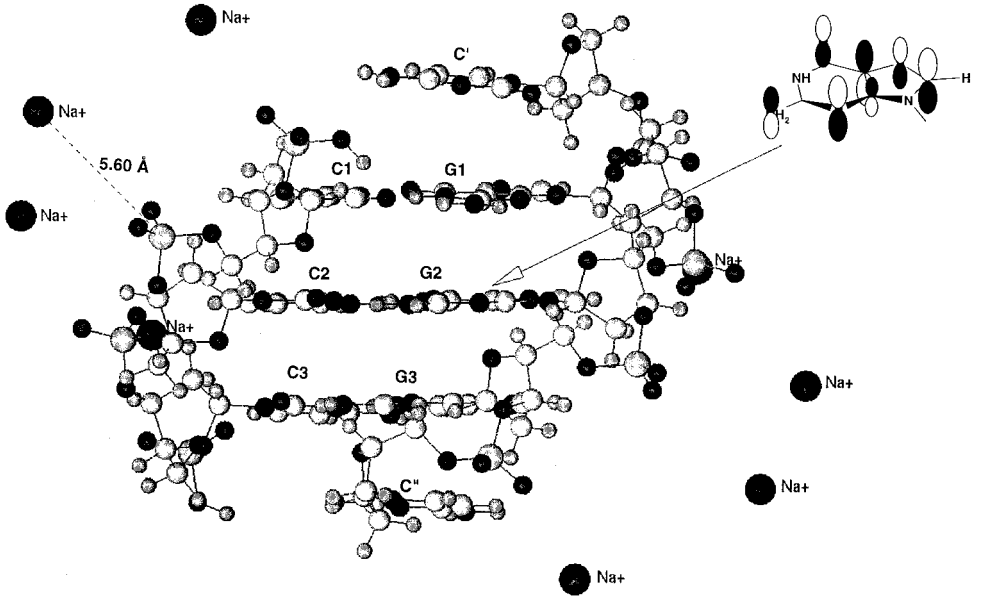


Figure 1. Description of the highest occupied molecular orbital in the oligonucleotide−counterion complex

$$\begin{matrix} CpG_1pG_2pG_3pC + 8\ Na^+ \\ 3'\text{-}\ pC_1pC_2pC_3p \end{matrix}$$

obtained from results of ab initio calculations with the 3-21G basis set indicating HOMO charge localization on the central guanine, G_2. The oligonucleotide is in the standard B-DNA conformation.

evidence that guanine ionization properties are related to sequence-specific DNA chemistry which depends on electron donation. However, in the examination of DNA models the influence of the sugar−phosphate backbone that contains anionic charge centers which can strongly perturb the ionization properties of neighboring groups has not been included. In the earlier investigations of DNA models, the effects of components that occur in a physiological environment on DNA ionization were also omitted. Under physiological conditions, DNA interacts with protein, water, and counterions.[38,39] Among DNA interactions with small monovalent counterions, those occurring at the anionic phosphate groups are highly favorable,[39a−c,e,40] and occur dynamically on a time scale of 10^{-12} to 10^{-9} s.[39a,d,e] The main goal of the present investigation is to examine base ionization properties of G runs in DNA structures that include the sugar−phosphate backbone, and to determine whether phosphate−counterion interactions influence the sites in G runs at which the lowest energy base ionization occurs.

Methods

Descriptions of Oligonucleotide Ionization Properties. The energetic ordering of gas-phase base ionization potentials of oligonucleotides, oligonucleotide−Na^+ complexes, and oligonucleotide models was determined by applying Koopmans' theorem to results from SCF calculations. Ab initio calculations with the STO-3G,[41] 3-21G, 6-31G, and 6-31G* basis sets and semiempirical ZINDO calculations[42] were carried out with an SGI Cray Origin2000 cluster containing 1528 R10000 processors. Ab initio calculations were performed using the Gaussian 98 program.[43]

For all oligonucleotides, complexes, and models, both semiempirical and ab initio calculations were performed to determine how the results varied depending on the computational method employed. Whenever possible, ab initio calculations were carried out with multiple basis sets to examine the basis set dependence of the results. The maximum size of the basis set employed for a specific sequence depended on the sequence size. Because this investigation focuses on a relatively simple electronic property, the electron distribution in the highest occupied base orbital, different computational methods gave similar results. For the oligonucleotide complex in Figure 1, which contains 258 atoms or ions and 1336 electrons, an ab initio 3-21G SCF calculation required 1041 MB of disk space and 466 min of processing time.

Descriptions of the HOBOs were obtained from a Mulliken population analysis.[44] In the tables, arrows denote guanines on which the total population of the highest occupied base orbital is greater than 10%. The length of each arrow is proportional to the population on the guanine to which the arrow points.

Geometries. Sequences which are rich in guanine, such as G runs, favor an A-DNA conformation.[45−47] In this investigation, G runs have

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

(39) (a) Seibel, G. L.; Singh, U. C.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 6537. (b) Pack G. R.; Lamm G.; Wong. L.; Clifton, D. In *Theoretical Biochemistry and Molecular Biophysics*; Beveridge, D. L., Lavery, R., Eds.; Adenine: Guilderland, NY, 1990; pp 237−246. (c) York, D. M.; Darden, T.; Deerfield, D.; Pedersen, L. G. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1992**, *19*, 145. (d) Young, M. A.; Jayaram, B.; Beveridge, D. L. *J. Am. Chem. Soc.* **1997,** *119*, 59. (e) Young, M. A.; Ravishanker, G.; Beveridge, D. L. *Biophys. J.* **1997**, *73*, 2313.

(40) (a) Seeman, N. C.; Rosenberg, J. M.; Suddath, F. L.; Kim, J. J. P.; Rich, A. *J. Mol. Biol.* **1976**, *104*, 109. (b) Rosenberg, J. M., Seeman, N. C.; Day, R. A.; Rich, A. *J. Mol. Biol.* **1976**, *104*, 145.

(41) (a) Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657. (b) Hehre, W. J.; Ditchfield R.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1970**, *52*, 2769.

(42) Zerner, M. C.; Ridley, J. E.; Bacon, A. D.; Edwards, W. D.; Head, J. D.; McKelvey, J.; Culberson, J. C.; Knappe, P.; Cory, M. G.; Weiner, B.; Baker, J. D.; Parkinson, W. A.; Kannis, D.; Yu, J.; Roesch, N.; Kotzian, M.; Tamm, T.; Karelson, M. M.; Zheng, X.; Pearl, G.; Broo, A.; Albert, K.; Cullen, J. M.; Cramer, C. J.; Truhlar, D. G.; Li, J.; Hawkins G. D.; Liotard, D. A. *ZINDO 99.1*; Molecular Simulations Inc.: San Diego, CA, 1999.

(43) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle E. S.; Pople, J. A. *Gaussian 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.

(44) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(45) Malinina, L.; Fernandez, L. G.; Huynh-Dinh, T.; Subirana, J. A. *J. Mol. Biol.* **1999**, *285*, 1679.

been examined in the standard A- and B-DNA conformations,[48] and in geometries taken from oligonucleotide crystal data.[45−47] The standard geometries were generated using the InsightII program.[49] Both in the standard geometries and in geometries from crystal data, the C−H, N−H, and O−H bond lengths were 1.09, 1.03, and 1.03 Å, respectively. Calculations of single-stranded oligonucleotides were carried out on geometries which correspond to that of a single strand in a standard double-stranded B-DNA conformation.

In model oligonucleotide calculations, stacking and hydrogen-bonding interactions were examined for 9-MeG and 1-MeC using base geometries and relative orientations obtained from the InsightII description of standard B-DNA. Here, the C−N9 and C−N1 bond lengths for the methyl groups in 9-MeG and 1-MeC are 1.49 Å. In these calculations, methyl groups on guanine and cytosine mimic the electronic influence that glycosidic bond formation has on base IPs.

Oligonucleotide−Na^+ complexes were examined in geometries in which there are favorable interactions of the Na^+ ion with the anionic phosphate groups. In the calculations of the complexes, each phosphate group interacts strongly with a single Na^+ ion. This is consistent with the finding that more than 70% of the phosphate charge is neutralized in a DNA solution that is 9 mM in phosphate and Na^+.[50] The local Na^+−phosphate geometry in the nucleotide complexes is based on that obtained from earlier ab initio 6-31G SCF calculations in which a Na^+ ion in an octahedral aqueous solvation shell was docked to the two negatively charged O atoms of phosphate.[16,18,19] Here, the Na^+ ion is in the plane containing the P atom and the two O atoms. The local symmetry of the Na^+ ion, the P atom, and the two O atoms is C_{2v}. For A- and B-DNA the distances between each Na^+ ion and the nearest P atom are 5.0 and 5.6 Å, respectively.

Results

A test 6-31G* calculation was carried out to examine the highest occupied molecular orbital in a small system containing two guanine−cytosine base pairs with the sugar phosphate backbones and six Na^+ ions.

$$\begin{matrix} 5'\text{-}pG_1pG_2p + 6\ Na^+ \\ 3'\text{-}pC_1pC_2p \end{matrix}$$

A standard B-DNA conformation was employed, and the Na^+ ions were located in positions corresponding to those employed in the oligonucleotide calculations. Here, the 6-31G* results indicating that the majority of the HOMO electron density (0.85) is on G_1 at the 5′-end agree with earlier calculations[12a] on the small model system

$$\begin{matrix} 5'\text{-}G_1G_2 \\ 3'\text{-}C_1C_2 \end{matrix}$$

The calculations were then extended to double-stranded oligonucleotides. Figure 1 shows the structure of an oligonucleotide−counterion complex which was examined in this investigation. The complex contains a G run with three guanines. The structure contains three stacked guanine−cytosine base pairs in a B-DNA conformation and additional stacked cytosines at each end of the G run. Each strand has a sugar−phosphate backbone with four phosphate groups. The complex also contains eight Na^+ counterions. Each of the Na^+ ions interacts strongly with one of the phosphate groups. The figure shows the highest occupied molecular orbital in the complex as described by ab initio SCF calculations with the 3-21G basis set. The results indicate that the HOMO contribution from the guainine in the middle of the G run is 72%.

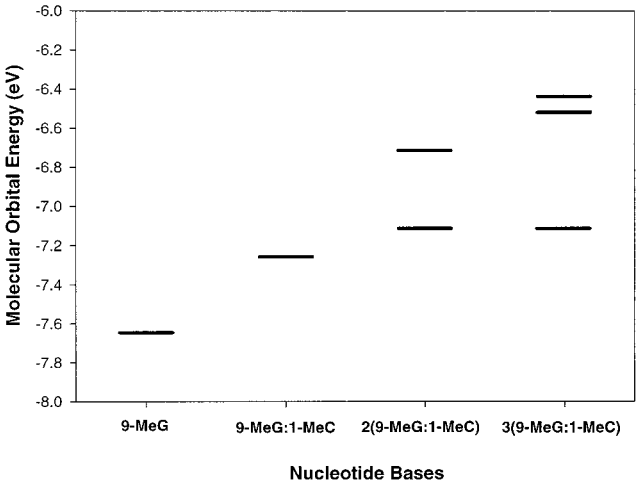


Figure 2. Energy levels associated with the highest occupied orbitals in systems containing three 9-MeG's and three 1-MeC's obtained from 3-21G SCF calculations. All of the energies are associated with the highest occupied π orbitals on one or more of the 9-MeG's. Results on the left, labeled 9-MeG, show the HOMO energy of 9-MeG in a system in which the shortest distance between molecules is 15 Å. Results second from the left, labeled 9-MeG:1-MeC, correspond to a system which contains a Watson−Crick base pair between a 9-MeG and a 1-MeC. The shortest distance between all other molecules is 15 Å. The results second from the right, labeled 2(9-MeG:1-MeC), correspond to a system which contains two stacked 9-MeG:1-MeC base pairs in a relative orientation which models that of A-DNA. The shortest distance between all other molecules is 15 Å. The results on the right, labeled 3(9-MeG:1-MeC), correspond to a system which contains three stacked 9-MeG:1-MeC base pairs in the A-DNA conformation. For all of the HOMOs, the relative magnitudes of the molecular orbital coefficients on the most strongly contributing 9-MeG are similar to those in the HOMO of Figure 1.

Figure 2 contains an energy level diagram showing ab initio 3-21G SCF energies of the highest occupied orbitals in systems containing three 9-MeG's and three 1-MeC's. All of the energies correspond to π orbitals localized on 9-MeG. The results farthest to the left, labeled 9-MeG, show the energy of the highest occupied π orbital of 9-MeG in a system where there is little intermolecular interaction. The results second from the left, 9-MeG:1-MeC, give the energy of the highest occupied π orbital of 9-MeG in a system which contains a Watson−Crick base pair between 9-MeG and 1-MeC. The results second from the right, 2(9-MeG:1-MeC), give the energies of the two highest occupied π orbitals on 9-MeG's in a system which contains two stacked 9-MeG:1-MeC base pairs in an orientation that models A-DNA. The results on the far right, 3(9-MeG:1-MeC), give the energies of the three highest occupied π orbitals on 9-MeG's in a system that contains three stacked 9-MeG:1-MeC base pairs in a model A-DNA geometry. Within the context of Koopmans' theorem, the results indicate that the formation of the 9-MeG:1-MeC Watson−Crick base pair decreases the IP of 9-MeG by 0.4 eV. The results also indicate that base stacking further reduces the ionization potential. For 2(9-MeG:1-MeC) and 3(9-MeG:1-MeC) the calculated IPs are 0.9 and 1.2 eV smaller than that for 9-MeG. These results are consistent with results from earlier 6-31G* and density functional calculations of adiabatic and vertical IPs of stacked and base-paired 9-MeG's.[12a,13,15]

(46) Champion, D.; Kumar, C. S.; Ramakrishnan, B.; Gautham, N.; Viswamitra, M. A. Nucleic Acid Data Base ADJ069. Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.-H.; Srinivasan, A. R.; Schneider, B. *Biophys. J.* **1992**, 63, 751.

(47) Gao, Y.-G.; Robinson, H. H.; Wang, A. H.-J. *Eur. J. Biochem.* **1999**, *261*, 413.

(48) Arnott, S.; Hukins, D. W. L. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504.

(49) InsightII, Molecular Simulations Inc., 9685 Scranton Rd., San Diego, CA 92121-3752.

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

Table 1. Fraction of Electron Density in the Highest Occupied Base Orbital (HOBO) on Different Guanines in an Oligonucleotide Model and in Single-Stranded Oligonucleotides Containing G Runs with Three and Four Guanines[a–e]

DNA	ZINDO	ΔIP[f]	STO-3G	ΔIP[g]	3-21G	ΔIP[h]	6-31G (6-31G*)	ΔIP[i]
ss3Gm	0.69 0.29 5′-$G_1G_2G_3$	0.24	0.81 0.18 5′-$G_1G_2G_3$	0.35	0.76 0.24 5′-$G_1G_2G_3$	0.35	0.80(0.81) 0.19(0.18) 5′-$G_1G_2G_3$	0.36 (0.35)
ss3G	0.66 0.15 5′-pG_1pG_2pG_3p	0.35	0.76 0.22 5′-pG_1pG_2pG_3p	0.27	0.62 0.36 5′-pG_1pG_2pG_3p	0.34		
ss3Gc	0.66 0.26 5′-pG_1pG_2pG_3p+4Na^+	0.21	0.82 0.16 5′-pG_1pG_2pG_3p+4Na^+	0.29	0.78 0.21 5′-pG_1pG_2pG_3p+4Na^+	0.38		
ss4G	0.67 0.19 5′-pG_1pG_2pG_3pG_4p	0.35	0.82 0.14 5′-pG_1pG_2pG_3pG_4p	0.30	0.75 0.16 5′-pG_1pG_2pG_3pG_4p	0.40		
ss4Gc	0.62 0.32 5′-pG_1pG_2pG_3pG_4p+5Na^+	0.22	0.53 0.45 5′-pG_1pG_2pG_3pG_4p+5Na^+	0.20	0.46 0.49 5′-pG_1pG_2pG_3pG_4p+5Na^+	0.30		

[a] Descriptions of HOBOs were obtained from Mulliken population analyses of results from ZINDO semiempirical calculations and from ab initio SCF calculations with STO-3G and 3-21G basis sets. For ss3Gm, ab initio results from 6-31G (without parentheses) and 6-31G* calculations (with parentheses) are also given. [b] Arrows above sequences indicate guanines that contribute most to the highest occupied base molecular orbital. [c] Values above the arrows represent the fraction of electron population in the highest occupied base orbital that resides on the guanine indicated. [d] Oligonucleotides are in the standard B-DNA conformation obtained from the InsightII program. [e] For oligonucleotide complexes containing Na^+ counterions, the ions interact most strongly with phosphate. See Figure 1. [f] Energy difference between the first and second highest occupied base orbitals (eV). From ZINDO calculations. [g] From STO-3G calculations. [h] From 3-21G calculations. [i] From 6-31G and 6-31G* calculations, with and without parentheses, respectively.

A. Single-Stranded Nucleotides and Nucleotide Models with G Runs Containing Three and Four Guanines. Table 1 contains descriptions of the HOMO in the single-stranded oligonucleotide model with three stacked 9-methylguanines in the same relative orientation of the bases that occurs in B-DNA.

5′-$G_1G_2G_3$ (**ss3Gm**)

For **ss3Gm**,[51] the arrows in the table show the 9-MeG's that contribute most to the highest occupied base orbital. The number above each arrow gives the fraction of the highest occupied orbital that resides on a specific 9-MeG obtained from a Mulliken population analysis. The table also gives the energy difference between the highest and second highest occupied base orbitals. Table 1 contains results from semiempirical ZINDO calculations, and from ab initio calculations with the STO-3G, 3-21G, 6-31G, and 6-31G* basis sets.

For 5′-$G_1G_2G_3$, the descriptions of electron distributions obtained from ZINDO calculations and from all of the ab initio SCF calculations are in good agreement, indicating that the HOMO is located primarily on G_1, at the 5′-end, where 0.69−0.81 of the total electron density resides. The calculations indicate that between 0.18 and 0.29 of the HOMO electron density is on G_2, in the middle. The finding that the HOMO of **ss3Gm** is located primarily on the guanine at the 5′-end agrees with that previously reported.[12b]

Table 1 also contains descriptions of the HOBO in single-stranded oligonucleotides containing three and four stacked guanines in the relative orientations that occur in B-DNA. Here the sugar−phosphate backbone is included. Results are given both for isolated oligonucleotides and for oligonucleotides with a Na^+ ion interacting strongly with each of the phosphate groups.

5′-pG_1pG_2pG_3p (**ss3G**) 5′-pG_1pG_2pG_3p + 4 Na^+ (**ss3Gc**)

5′-pG_1pG_2pG_3pG_4p (**ss4G**) 5′-pG_1pG_2pG_3pG_4p + 5 Na^+ (**ss4Gc**)

As in earlier investigations of mononucleotides, the description of the HOMO in oligonucleotides containing anionic phosphate groups depends on the computational method employed.[16,18] In **ss3G**, the ZINDO results indicate that the HOMO is on the phosphate group between G_1 and G_2; the ab initio STO-3G and 3-21G results indicate that the HOMO is on G_1 and G_2. In **ss3Gc**, the STO-3G results indicate that the HOMO is on the phosphate group on the 5′-end; the ZINDO and 3-21G results indicate that the HOMO is on G_1 and G_2. The results in Table 1 and in subsequent tables contain descriptions of the HOBOs, which are not, in all cases, predicted to be the HOMOs.

For 5′-pG_1pG_2pG_3p with and without Na^+ (**ss3Gc** and **ss3G**), results obtained from ZINDO calculations and from ab initio STO-3G and 3-21G SCF calculations are similar, indicating that the electron distribution in the highest occupied base orbital is dependent on the specific structure of the oligonucleotide system examined and on the environment. These results suggest that, for stacked single-stranded G runs, the location of the HOBO

(51) The following notation has been employed to describe structures and sequences (xyA...Bz): xy = ds means double-stranded. xy = ss means single-stranded. A...B describes the base sequence starting from the 5′-end. 3G and 4G denote G runs with three and four guanines, respectively. For double-stranded oligonucleotides and models, the sequence of only one strand is given. z = m denotes a model oligonucleotide without a backbone. z = blank denotes an oligonucleotide with a backbone. z = c denotes an oligonucleotide complex with Na^+ counterions.

Table 2. Fraction of Electron Density in the Highest Occupied Base Orbital on Different Guanines in Double-Stranded Oligonucleotides and Oligonucleotide Models of G Runs with Three and Four Guanines in the Standard B-DNA Conformation[a]

DNA	ZINDO	ΔIP [b]	STO-3G	ΔIP [c]	3-21G	ΔIP [d]	6-31G	ΔIP [e]
ds3Gm	0.63, 0.30 / 5′-$G_1G_2G_3$ 3′-$C_1C_2C_3$	0.23	0.62, 0.35 / 5′-$G_1G_2G_3$ 3′-$C_1C_2C_3$	0.21	0.60, 0.35 / 5′-$G_1G_2G_3$ 3′-$C_1C_2C_3$	0.33	0.56, 0.39 / 5′-$G_1G_2G_3$ 3′-$C_1C_2C_3$	0.32
ds3G	0.62 / 5′-$pG_1pG_2pG_3p$ 3′-$pC_1pC_2pC_3p$	0.66	0.91 / 5′-$pG_1pG_2pG_3p$ 3′-$pC_1pC_2pC_3p$	0.69	0.89 / 5′-$pG_1pG_2pG_3p$ 3′-$pC_1pC_2pC_3p$	0.68		
ds3Gc	0.86, 0.10 / 5′-$pG_1pG_2pG_3p$ 3′-$pC_1pC_2pC_3p$+8Na^+	0.39	0.90 / 5′-$pG_1pG_2pG_3p$ 3′-$pC_1pC_2pC_3p$+8Na^+	0.42	0.81, 0.15 / 5′-$pG_1pG_2pG_3p$ 3′-$pC_1pC_2pC_3p$+8Na^+	0.46		
ds4G	0.56, 0.42 / 5′-$pG_1pG_2pG_3pG_4p$ 3′-$pC_1pC_2pC_3pC_4p$	0.23	0.52, 0.47 / 5′-$pG_1pG_2pG_3pG_4p$ 3′-$pC_1pC_2pC_3pC_4p$	0.22	0.51, 0.46 / 5′-$pG_1pG_2pG_3pG_4p$ 3′-$pC_1pC_2pC_3pC_4p$	0.33		
ds4Gc	0.55, 0.42 / 5′-$pG_1pG_2pG_3pG_4p$ 3′-$pC_1pC_2pC_3pC_4p$+10Na^+	0.22	0.51, 0.34 / 5′-$pG_1pG_2pG_3pG_4p$ 3′-$pC_1pC_2pC_3pC_4p$+10Na^+	0.22	0.37, 0.26 / 5′-$pG_1pG_2pG_3pG_4p$ 3′-$pC_1pC_2pC_3pC_4p$+10Na^+	0.56		

[a] See footnotes for Table 1. [b−d] Same as footnotes *f*, *g*, and *h*, respectively, in Table 1. [e] From 6-31G calculations.

depends on phosphate electrostatic interactions. In **ss3G**, the equal number of anionic phosphate groups in the 5′- and 3′-directions from G_2 results in location of the HOBO on G_2. When the anionic phosphate charges are shielded with Na^+ counterions in **ss3Gc**, the HOBO returns to G_1 at the 5′-end, as in the model **ss3Gm**.

Results for 5′-$pG_1pG_2pG_3pG_4p$ with and without Na^+ (**ss4Gc** and **ss4G**) are similar. Here, there are some differences between the ZINDO and the ab initio descriptions of the HOBOs; however, the trend is the same. Without Na^+, the single-stranded G run with four guanines (**ss4G**) has a HOBO that is located at interior positions, G_2 (according to ab initio results) or G_3 (according to ZINDO results). In the complex with Na^+ (**ss4Gc**), all of the calculations indicate that the HOBO electron density is more delocalized than in **ss4G**, and that, compared to **ss4G**, the HOBO electron density is shifted toward the 5′-end.

The results for 5′-$pG_1pG_2pG_3p$ (**ss3G**) indicating that the HOBO resides primarily on the central guanine is consistent with the observation that the molecular electrostatic potential (MEP) at the central guanine is more negative than at the guanines on the ends. In **ss3G**, 3-21G calculations indicate that the MEP in the plane of G_2 and 2.1 Å from N7 on a line bisecting the C3−N7−C8 bond angle is more negative than that at corresponding positions in the planes of G_1 and G_3 by 25.7 and 10.9 kcal mol^{-1} e^{-1}, respectively. In **ss3Gc**, the MEP at G_2 is only 6.1 kcal mol^{-1} e^{-1} more negative than that at G_1.

Results in Table 1, for the model (**ss3Gm**) and for the G runs containing sugar−phosphate backbones with (**ss3Gc** and **ss4Gc**) and without (**ss3G** and **ss4G**) Na^+ ions, indicate that the highest occupied base orbital electron distributions change when anionic phosphate groups are incorporated into the structure and when counterion interactions are included. Changes in HOBO electron distributions that depend on structure and environment are expected because of the energy matching of orbitals that occurs in the guanine bases which make up G runs. This energy resonance gives rise to labile base orbitals in which changes in spacial distributions are readily induced by structural and environmental perturbations.

B. Double-Stranded Nucleotides and Nucleotide Models with G Runs Containing Three and Four Guanines. i. B-DNA. Table 2 contains descriptions of the highest occupied base orbitals in double-stranded G runs and in a model of a double-stranded G run in the standard B-DNA conformation. The table contains results from ZINDO and ab initio STO-3G, 3-21G, and 6-31G calculations. For the model of the double-stranded G run

5′-$G_1G_2G_3$ (**ds3Gm**)
3′-$C_1C_2C_3$

the results are different from those for the model of the single-stranded G run (**ss3Gm**). The ZINDO and ab initio STO-3G, 3-21G, and 6-31G results for **ds3Gm** all indicate that the HOMO is not localized on G_1 at the 5′-end, but is delocalized over G_1 and the middle guanine, G_2, with the population on G_2 more than 1.4 times larger than that on G_1.[52]

Table 2 also gives results for oligonucleotides containing G runs with three and four guanines with and without Na^+.

5′-$pG_1pG_2pG_3p$ (**ds3G**)
3′-$pC_1pC_2pC_3p$

5′-$pG_1pG_2pG_3p$ + 8 Na^+ (**ds3Gc**)
3′-$pC_1pC_2pC_3p$

5′-$pG_1pG_2pG_3pG_4p$ (**ds4G**)
3′-$pC_1pC_2pC_3pC_4p$

5′-$pG_1pG_2pG_3pG_4p$ + 10 Na^+ (**ds4Gc**)
3′-$pC_1pC_2pC_3pG_4p$

(52) For the double-stranded model, the HOMO structure depends on the conformation. When ds3Gm is in an A-DNA conformation, 3-21G results, for which the energies are given in Figure 2, indicate that the contributions to the HOMO from the 9-MeG's at the 5′-end and in the middle are 0.78 and 0.22, respectively.

Table 3. Fraction of Electron Density in the Highest Occupied Base Orbital on Different Guanines of G Runs in Double-Stranded Oligonucleotides in the Standard A-DNA Conformation[a,b]

DNA	ZINDO	ΔIP [c]	STO-3G	ΔIP [d]
ds3G	0.93 → 5′-pG$_1$pG$_2$pG$_3$p / 3′-pC$_1$pC$_2$pC$_3$p	0.61	0.99 → 5′-pG$_1$pG$_2$pG$_3$p / 3′-pC$_1$pC$_2$pC$_3$p	0.41
ds3Gc	0.83, 0.13 → 5′-pG$_1$pG$_2$pG$_3$p / 3′-pC$_1$pC$_2$pC$_3$p +8Na$^+$	0.26	0.99 → 5′-pG$_1$pG$_2$pG$_3$p / 3′-pC$_1$pC$_2$pC$_3$p +8Na$^+$	0.15
dsT3GA	0.91 → 5′-TpG$_1$pG$_2$pG$_3$pA / 3′-ApC$_1$pC$_2$pC$_3$pT	0.52	0.99 → 5′-TpG$_1$pG$_2$pG$_3$pA / 3′-ApC$_1$pC$_2$pC$_3$pT	0.45
dsT3GAc	0.76, 0.20 → 5′-TpG$_1$pG$_2$pG$_3$pA / 3′-ApC$_1$pC$_2$pC$_3$pT +8Na$^+$	0.20	0.99 → 5′-TpG$_1$pG$_2$pG$_3$pA / 3′-ApC$_1$pC$_2$pC$_3$pT +8Na$^+$	0.12
dsAT3GAG	0.85 → 5′-ApTpG$_1$pG$_2$pG$_3$pApG / 3′-TpApC$_1$pC$_2$pC$_3$pTpC	0.45	0.92 → 5′-ApTpG$_1$pG$_2$pG$_3$pApG / 3′-TpApC$_1$pC$_2$pC$_3$pTpC	0.25
dsAT3GAGc	0.62, 0.32 → 5′-ApTpG$_1$pG$_2$pG$_3$pApG / 3′-TpApC$_1$pC$_2$pC$_3$pTpC +12Na$^+$	0.17	0.82 → 5′-ApTpG$_1$pG$_2$pG$_3$pApG / 3′-TpApC$_1$pC$_2$pC$_3$pTpC +12Na$^+$	0.04

[a] Results are for G runs with three guanines and are given for sequences with and without Na^+ counterions. See footnotes for Table 1. [b] DNA geometries obtained using the InsightII program. [c,d] Same as footnotes *f* and *g* in Table 1.

For double-stranded, like single-stranded, oligonucleotides the description of the HOMO depends on the method of calculation.[53] However, the descriptions of the highest occupied base orbitals provided by the different methods are similar. For the double-stranded oligonucleotide containing three guanines (**ds3G**), the largest population in the HOBO, like that in the double-stranded model **ds3Gm**, resides primarily on the middle guanine, G_2, not on G_1 at the 5′-end.

The results for the double-stranded oligonucleotide containing four guanines (**ds4G**) indicate that the HOBO, which is almost equally distributed on G_2 and G_3, is less localized than in **ds3G**. For **ds3G**, ZINDO and ab initio STO-3G and 3-21G calculations indicate that the HOBO population on G_2 is 0.62, 0.91, and 0.89, respectively. In **ds4G** the ratio of the G_2 to G_3 contributions is 0.7:1.1. However, like **ds3G**, the HOBO of **ds4G** is again in the interior of the G run. For the double-stranded complexes with Na^+, the results in Table 2 indicate that the HOBO in the G run with three guanines (**ds3Gc**) is also more localized than the HOBO in the G run with four guanines (**ds4Gc**). In **ds3Gc**, the G_2 contribution is more than 0.8. In **ds4Gc**, the ratio of the G_2 to G_3 contributions is 1.3:1.5. Nevertheless, in **ds4Gc**, like **ds3Gc**, the HOBO remains in the interior of the G run.

(53) For example, the 3-21G HOMO of **ds3G** is on G_2, and the HOMO of **ds4G** is on G_2 and G_3. In contrast, the STO-3G HOMOs of **ds3G** and **ds4G** are on the phosphate group between G_2 and G_3. Similarly, in **ds3Gc** and **ds4Gc**, the 3-21G HOMOs are on G_1 and G_2, and on G_2 and G_3, respectively, while the STO-3G HOMOs are on the phosphate groups between G_1 and G_2 and between G_2 and G_3.

In a further test of the basis set dependence of the description of the HOBO, a 6-31G* calculation was carried out on **ds3Gc**. Here, the 6-31G* results indicate that the fractions of the HOBO electron population on G_1 at the 5′-end and G_2 in the middle (0.15 and 0.81, respectively) agree with the 3-21G results in Table 2.

ii. A-DNA. Tables 3 and 4 give results for oligonucleotides containing G runs with three and four guanines in the standard A-DNA conformation. Both tables show results with and without Na^+ ions. In Table 3, which contains G runs with three guanines, the results for the A conformation of **ds3G** and **ds3Gc**, like the results for the B conformation in Table 2, indicate that the HOBOs reside primarily on the interior guanine, G_2. Similarly, in Table 4, which contains results for G runs with four guanines, the results for **ds4G** and **ds4Gc** in the A conformation are similar to those in Table 2 for the B conformation. The largest contributions to the HOBO come from one of the interior guanines, G_2 or G_3.

Tables 3 and 4 also list results for G runs which are incorporated in sequences containing adenine with thymine. Table 3 gives results for a G run with three guanines in sequences containing five and seven base pairs.

5′-TpG$_1$pG$_2$pG$_3$pA (**dsT3GA**)
3′-ApC$_1$pC$_2$pC$_3$pT

5′-TpG$_1$pG$_2$pG$_3$pA + 8 Na$^+$ (**dsT3GAc**)
3′-ApC$_1$pC$_2$pC$_3$pT

5′-ApTpG$_1$pG$_2$pG$_3$pApG (**dsAT3GAG**)
3′-TpApC$_1$pC$_2$pC$_3$pTpC

5′-ApTpG$_1$pG$_2$pG$_3$pApG + 12 Na$^+$ (**dsAT3GAGc**)
3′-TpApC$_1$pC$_2$pC$_3$pTpC

Table 4 shows results for G runs with four guanines in sequences containing six to eight base pairs.

5′-ApG$_1$pG$_2$pG$_3$pG$_4$pC (**dsA4GC**)
3′-TpC$_1$pC$_2$pC$_3$pC$_4$pG

5′-ApG$_1$pG$_2$pG$_3$pG$_4$pC + 10 Na$^+$ (**dsA4GCc**)
3′-TpC$_1$pC$_2$pC$_3$pC$_4$pG

5′-ApG$_1$pG$_2$pG$_3$pG$_4$pCpC (**dsA4GCC**)
3′-TpC$_1$pC$_2$pC$_3$pC$_4$pGpG

5′-ApG$_1$pG$_2$pG$_3$pG$_4$pCpC + 12 Na$^+$ (**dsAT3GAGc**)
3′-TpC$_1$pC$_2$pC$_3$pC$_4$pCpC

5′-TpApG$_1$pG$_2$pG$_3$pG$_4$pCpC (**dsTA4GCC**)
3′-ApTpC$_1$pC$_2$pC$_3$pC$_4$pGpG

Results for G runs containing three and four guanines in longer standard A-DNA sequences are similar to results obtained for G runs without flanking bases. In all cases, with and without Na^+, the central guanine, G_2, makes the largest contribution to the HOBO in the runs with three guanines, and an interior guanine, G_2 or G_3, makes the largest contribution to the HOBO in the runs with four guanines.

iii. Crystal Structures. Tables 5−7 show ab initio STO-3G and ZINDO descriptions of the highest occupied base orbitals in G runs containing three and four guanines in geometries obtained from crystal data.[45−47] All of these double-stranded crystal structures have individual conformational properties that differ in one way or another from standard geometries. Two[46,47] are in a general A conformation; one[45] has a kink and exhibits a hybrid A/B-DNA geometry. Results for sequences taken from a double-stranded dodecamer, d(CGCCCGCGGGCG),[45] and a double-stranded decamer, d(CCGGGCCCGG),[46] are given in Tables 5 and 6, respectively. Each of the sequences contains G runs with three guanines. Tables 5 and 6 contain results for ds3G and for

5′-CpG$_1$pG$_2$pG$_3$pC (**dsC3GC**)
3′-GpC$_1$pC$_2$pC$_3$pG

ZINDO and STO-3G results in Tables 5 and 6 obtained from both crystal structures containing geometries of **ds3G** and **dsC3GC** indicate that the contribution to the HOBO from the central guanine (G_2) is more than 0.90.

Table 4. Fraction of Electron Density in the Highest Occupied Base Orbital on Different Guanines of G Runs in Double-Stranded Oligonucleotides in the Standard A-DNA Conformation[a,b]

DNA	ZINDO	ΔIP [c]	STO-3G	ΔIP [d]
ds4G	0.49 0.47 5'-pG$_1$pG$_2$pG$_3$pG$_4$p 3'-pC$_1$pC$_2$pC$_3$pC$_4$P	0.14	0.96 5'-pG$_1$pG$_2$pG$_3$pG$_4$p 3'-pC$_1$pC$_2$pC$_3$pC$_4$P	0.09
ds4Gc	0.75 0.21 5'-pG$_1$pG$_2$pG$_3$pG$_4$p +10Na$^+$ 3'-pC$_1$pC$_2$pC$_3$pC$_4$P	0.21	0.99 5'-pG$_1$pG$_2$pG$_3$pG$_4$p +10Na$^+$ 3'-pC$_1$pC$_2$pC$_3$pC$_4$P	0.21
dsA4GC	0.66 0.30 5'-ApG$_1$pG$_2$pG$_3$pG$_4$pC 3'-TpC$_1$pC$_2$pC$_3$pC$_4$pG	0.21	0.98 5'-ApG$_1$pG$_2$pG$_3$pG$_4$pC 3'-TpC$_1$pC$_2$pC$_3$pC$_4$pG	0.17
dsA4GCc	0.71 0.17 0.10 5'-ApG$_1$pG$_2$pG$_3$pG$_4$pC +10Na$^+$ 3'-TpC$_1$pC$_2$pC$_3$pC$_4$pG	0.21	0.99 5'-ApG$_1$pG$_2$pG$_3$pG$_4$pC +10Na$^+$ 3'-TpC$_1$pC$_2$pC$_3$pC$_4$pG	0.27
dsA4GCC	0.77 0.17 5'-ApG$_1$pG$_2$pG$_3$pG$_4$pCpC 3'-TpC$_1$pC$_2$pC$_3$pC$_4$pGpG	0.25	0.96 5'-ApG$_1$pG$_2$pG$_3$pG$_4$pCpC 3'-TpC$_1$pC$_2$pC$_3$pC$_4$pGpG	0.08
dsA4GCCc	0.67 0.23 5'-ApG$_1$pG$_2$pG$_3$pG$_4$pCpC 3'-TpC$_1$pC$_2$pC$_3$pC$_4$pGpG +12Na$^+$	0.19	0.92 5'-ApG$_1$pG$_2$pG$_3$pG$_4$pCpC 3'-TpC$_1$pC$_2$pC$_3$pC$_4$pGpG +12Na$^+$	0.13
dsTA4GCC	0.78 0.13 5'-TpApG$_1$pG$_2$pG$_3$pG$_4$pCpC 3'-ApTpC$_1$pC$_2$pC$_3$pC$_4$pGpG	0.28	0.99 5'-TpApG$_1$pG$_2$pG$_3$pG$_4$pCpC 3'-ApTpC$_1$pC$_2$pC$_3$pC$_4$pGpG	0.29

[a] Results are for G runs with four guanines. See footnotes for Table 1. [b] Same as footnote *b* in Table 3. [c,d] Same as footnotes *f* and *g* in Table 1.

Tables 5 and 6 also contain results for G runs with three guanines in longer sequences with six or seven base pairs indifferent contexts.[45,46] Table 5 gives results for

5'-CpG$_1$pG$_2$pG$_3$pCpG (**dsC3GCG**)
3'-GpC$_1$pC$_2$pC$_3$pGpC

5'-GpCpG$_1$pG$_2$pG$_3$pC (**dsGC3GC**)
3'-CpGpC$_1$pC$_2$pC$_3$pG

5'-GpCpG$_1$pG$_2$pG$_3$pCpG (**dsGC3GCG**)
3'-CpGpC$_1$pC$_2$pC$_3$pGpC

Table 6 contains results for

5'-CpG$_1$pG$_2$pG$_3$pCpC (**dsC3GCC**)
3'-GpC$_1$pC$_2$pC$_3$pGpG

5'-CpCpG$_1$pG$_2$pG$_3$pC (**dsCC3GC**)
3'-GpGpC$_1$pC$_2$pC$_3$pG

5'-CpCpG$_1$pG$_2$pG$_3$pCpC (**dsCC3GCC**)
3'-GpGpC$_1$pC$_2$pC$_3$pGpG

Here, the STO-3G and ZINDO calculations used with the crystal geometries indicate, again, that the HOBO is localized primarily on G_2 in the middle of the G run. In all cases, the contribution from G_2 is more than 0.7.

Table 7 contains descriptions of the highest occupied base orbitals obtained from ZINDO and STO-3G calculations using a crystal geometry for an oligonucleotide containing a G run with four guanines. The geometry is based on that for the double-stranded decamer d(AGGGGCCCCT) in the $P2_12_12_1$ space group.[47] The table gives descriptions of the HOBOs in

Table 5. Fraction of Electron Density in the Highest Occupied Base Orbital on Different Guanines in Oligonucleotides Containing G runs with Three Guanines[a,b]

DNA	ZINDO	ΔIP [c]	STO-3G	ΔIP [d]
ds3G	0.98 5'-pG$_1$pG$_2$pG$_3$p 3'-pC$_1$pC$_2$pC$_3$p	0.82	0.99 5'-pG$_1$pG$_2$pG$_3$p 3'-pC$_1$pC$_2$pC$_3$p	0.78
dsC3GC	0.98 5'-CpG$_1$pG$_2$pG$_3$pC 3'-GpC$_1$pC$_2$pC$_3$pG	0.79	0.92 5'-CpG$_1$pG$_2$pG$_3$pC 3'-GpC$_1$pC$_2$pC$_3$pG	0.77
dsC3GCG	0.98 5'-CpG$_1$pG$_2$pG$_3$pCpG 3'-GpC$_1$pC$_2$pC$_3$pGpC	0.35	0.99 5'-CpG$_1$pG$_2$pG$_3$pCpG 3'-GpC$_1$pC$_2$pC$_3$pGpC	0.53
dsGC3GC	0.91 5'-GpCpG$_1$pG$_2$pG$_3$pC 3'-CpGpC$_1$pC$_2$pC$_3$pG	0.32	0.95 5'-GpCpG$_1$pG$_2$pG$_3$pC 3'-CpGpC$_1$pC$_2$pC$_3$pG	0.25
dsGC3GCG	0.97 5'-GpCpG$_1$pG$_2$pG$_3$pCpG 3'-CpGpC$_1$pC$_2$pC$_3$pGpC	0.66	0.99 5'-GpCpG$_1$pG$_2$pG$_3$pCpG 3'-CpGpC$_1$pC$_2$pC$_3$pGpC	0.62
dsCGC3GC	0.88 0.11 5'-pCPGpCpG$_1$pG$_2$pG$_3$pCp 3'-pGpCpGpC$_1$pC$_2$pC$_3$pGp	0.23	0.98 5'-pCPGpCpG$_1$pG$_2$pG$_3$pCp 3'-pGpCpGpC$_1$pC$_2$pC$_3$pGp	0.21

[a] Geometries were taken from X-ray data for the double-stranded dodecamer d(CGCCCGCGGGCG). See footnotes for Table 1. [b] See ref 45. [c,d] Same as footnotes *f* and *g* in Table 1.

ds4G and of G runs with four guanines in two oligonucleotides containing six and seven base pairs, **dsA4GC** and **dsA4GCC**. Like the results obtained for the standard B-DNA and A-DNA geometries in Tables 2 and 4, the results in Table 7 for **ds4G**, **dsA4GC**, and **dsA4GCC** obtained using crystal data indicate that the highest occupied base orbitals reside on the interior guanines, G_2 and G_3, of the G run. In all of the results given in Table 7, the total contributions from the interior guanines are between 0.88 and 0.98.

Discussion

This investigation employs SCF descriptions of upper occupied orbitals in oligonucleotides to characterize base ionization properties. For the oligonucleotides examined, with and without Na^+, the descriptions of the highest occupied molecular orbitals depend on the computational method used. Where differences occur, they are related to whether the HOMO is a base orbital or a phosphate orbital. These differences between the HOMOs provided by different methods are not unreasonable. Tests of SCF descriptions of the gas-phase ionization of 2′-deoxyguanosine 5′-phosphate (5′-dGMP$^-$)[16] and of a phosphorylated dinucleotide containing guanine,[18] using gas-phase photoelectron data and post-SCF calculations on $H_2PO_4^-$, indicate that the guanine and phosphate ionization energies differ by less than 0.5 eV.[16,18,54]

In water, the energetics of ionization are significantly different from those in the gas phase. A theoretical description of 5′-

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

Table 6. Fraction of Electron Density in the Highest Occupied Base Orbital on Different Guanines in Oligonucleotides Containing G Runs with Three Guanines[a,b]

DNA	ZINDO	ΔIP [c]	STO-3G	ΔIP [d]
ds3G	0.96 ↓ 5′-pG$_1$pG$_2$pG$_3$p 3′-pC$_1$pC$_2$pC$_3$p	0.63	0.99 ↓ 5′-pG$_1$pG$_2$pG$_3$p 3′-pC$_1$pC$_2$pC$_3$p	0.86
dsC3GC	0.97 ↓ 5′-CpG$_1$pG$_2$pG$_3$pC 3′-GpC$_1$pC$_2$pC$_3$pG	0.68	0.99 ↓ 5′-CpG$_1$pG$_2$pG$_3$pC 3′-GpC$_1$pC$_2$pC$_3$pG	0.61
dsC3GCC	0.84 ↓, 0.16 ↓ 5′-CpG$_1$pG$_2$pG$_3$pCpC 3′-GpC$_1$pC$_2$pC$_3$pGpG	0.29	0.98 ↓ 5′-CpG$_1$pG$_2$pG$_3$pCpC 3′-GpC$_1$pC$_2$pC$_3$pGpG	0.50
dsCC3GC	0.98 ↓ 5′-CpCpG$_1$pG$_2$pG$_3$pC 3′-GpGpC$_1$pC$_2$pC$_3$pG	0.19	0.72 ↓ 5′-CpCpG$_1$pG$_2$pG$_3$pC 3′-GpGpC$_1$pC$_2$pC$_3$pG	0.20
dsCC3GCC	0.96 ↓ 5′-CpCpG$_1$pG$_2$pG$_3$pCpC 3′-GpGpC$_1$pC$_2$pC$_3$pGpG	0.38	0.99 ↓ 5′-CpCpG$_1$pG$_2$pG$_3$pCpC 3′-GpGpC$_1$pC$_2$pC$_3$pGpG	0.39

[a] Geometries were taken from X-ray data for the double-stranded decamer d(CCGGGCCCGG). See footnotes for Table 1. [b] See ref 46. [c,d] Same as footnotes *f* and *g* in Table 1.

Table 7. Fraction of Electron Density in the Highest Occupied Base Orbital on Different Guanines in Oligonucleotides Containing G Runs with Four Guanines[a,b]

DNA	ZINDO	ΔIP [c]	STO-3G	ΔIP [d]
ds4G	0.93 ↓ 5′-pG$_1$pG$_2$pG$_3$pG$_4$p 3′-pC$_1$pC$_2$pC$_3$pC$_4$p	0.20	0.88 ↓, 0.10 ↓ 5′-pG$_1$pG$_2$pG$_3$pG$_4$p 3′-pC$_1$pC$_2$pC$_3$pC$_4$p	0.21
dsA4GC	0.92 ↓ 5′-ApG$_1$pG$_2$pG$_3$pG$_4$pC 3′-TpC$_1$pC$_2$pC$_3$pC$_4$pG	0.28	0.93 ↓ 5′-ApG$_1$pG$_2$pG$_3$pG$_4$pC 3′-TpC$_1$pC$_2$pC$_3$pC$_4$pG	0.28
dsA4GCC	0.88 ↓ 5′-ApG$_1$pG$_2$pG$_3$pG$_4$pCpC 3′-TpC$_1$pC$_2$pC$_3$pC$_4$pGpG	0.12	0.78 ↓, 0.16 ↓ 5′-ApG$_1$pG$_2$pG$_3$pG$_4$pCpC 3′-TpC$_1$pC$_2$pC$_3$pC$_4$pGpG	0.15

[a] Geometries were taken from X-ray data for the double-stranded decamer d(AGGGGCCCCT). See footnotes for Table 1. [b] See ref 47. [c,d] Same as footnotes *f* and *g* in Table 1.

dGMP$^-$ solvation energies, before and after ionization, based on a Langevin dipole relaxation model[55] indicates that, when bulk water relaxation is accounted for, aqueous removal of an electron from the guanine group of 5′-dGMP$^-$ is approximately 1 eV more favorable than removal of an electron from phosphate.[16,18] This ordering of aqueous ionization events is consistent with the finding that, in 5′-dGMP$^-$,[56] in single-stranded oligonucleotides[57] and polynucleotides[28] containing guanine, and in double-stranded DNA,[26] results from photoionization and γ radiolysis experiments provide evidence that electron hole migration occurs primarily to guanine. The experimental association of the lowest aqueous IP with guanine has aided the present investigation which finds that different computational methods provide similar descriptions of the lowest energy guanine ionization event in G runs.

Results from Figure 2 and from earlier investigations[12a13,15] indicate how hydrogen bonding and base stacking reduce base ionization energies in G runs. In Table 1, the ab initio and ZINDO SCF descriptions of the lowest energy ionization event in a single-stranded model G run containing three 9-MeG's without the sugar−phosphate backbone (**ss3Gm**) are similar to previously reported results obtained from ab initio 6-31G* calculations.[12a,b] All of the calculations indicate that the first IP is 0.3−0.6 eV smaller than that of isolated 9-MeG, and that, within the context of Koopmans' theorem, the first ionization event is largely localized (69−81%) at the guanine on the 5′-end.

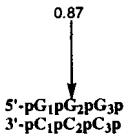
The results in Table 1 also point out that the orbital energy resonance which occurs between individual guanines in G runs makes guanine ionization properties in single-stranded oligonucleotides sensitive to details in structure and environment. For example, the ZINDO, STO-3G, and 3-21G results indicate that introduction of the sugar−phosphate backbone into single-stranded G runs (**ss3G** and **ss4G**) results in a shift of the highest occupied base orbital from the 5′-end in **ss3Gm** to interior guanines. When charges on the phosphate groups in single-stranded G runs are screened by Na^+ ions, as in **ss3Gc** and **ss4Gc**, the HOBO electron density is shifted from the interior back toward the 5′-end.

In double-stranded structures that contain the backbone and more closely resemble native DNA, the lowest energy base ionization in G runs is associated with interior guanines. This finding holds for G runs containing three and four guanines in the standard B (Table 2) and A (Tables 3 and 4) conformations. For the standard conformations, the results in Tables 2−4 also indicate that accumulation of the HOBO at interior positions of G runs occurs both with and without Na^+ counterions. The accumulation of the HOBO in the interior of G runs is also indicated when geometries are taken from X-ray structures (Tables 5−7).

In describing the HOBOs in oligonucleotides containing G runs, it is interesting to consider the number of anionic phosphate groups lying in the 5′-direction of interior guanines compared to the number of phosphate groups in the 3′-direction. In Tables 2−7, it might be expected that the HOBO residence in the interior of G runs is governed principally by charge balance. That is, the number of phosphate groups lying in the 5′-direction of the interior guanines equals the number lying in the 3′-direction. This argument suggests that the large contributions to the HOBO from interior guanines, observed in the present investigation of oligonucleotides containing eight or fewer base pairs, might not occur in longer oligonucleotides or in polynucleotides. However, evidence suggesting that this is not so is indicated by results for **dsCGC3GC** in Table 5. Here, the HOBO is localized on the central guanine (G_2), even though there are five phosphate groups lying in the 5′-direction of G_2 and only three phosphates in the 3′-direction. Charge imbalances also occur in **dsCC3GC** of Table 6 and **dsA4GCC** of Table 7.

(55) Florián. J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583.
(56) Gregoli, S.; Olast, M.; Bertinchamps, A. *Radiat. Res.* **1977**, *72*, 201.
(57) Melvin, T.; Botchway, S. W.; Parker, A. W.; O'Neill, P. *J. Chem. Soc., Chem. Commun.* **1995**, 653.

Consideration has been given to the possibility that ionization occurs with equal energy from guanines in the interior of G runs and from guanines at the 5′- or 3′-end. However, within the framework of Koopmans' theorem, this is not a frequent occurrence among the systems examined in this investigation. For the double-stranded G runs containing a sugar−phosphate backbone, with or without Na* ions, the results in Tables 2−7 indicate that the average differences between the first and second base ionization energies are 0.51, 0.34, and 0.36 eV in results obtained from 3-21G, STO-3G, and ZINDO calculations, respectively.



To examine the influence of an aqueous environment on the results for double-stranded G runs, a STO-3G calculation on ds3G in the standard A conformation was carried out in an IPCM reaction field.[58] The results shown above indicate that the HOBO is less localized than for **ds3G** in the gas phase, where STO-3G calculations indicate that the contribution from G_2 in the middle is 0.99. More importantly, like the gas-phase results, the reaction field results indicate that the largest contribution to the HOBO comes from the middle guanine.

Sequence-Specific Photoionization and Alkylation Patterns. Earlier investigations of mechanisms of DNA photoionization have focused on the importance of the low IP of guanine and of hole migration to guanine.[26,29] Results from 266 nm two-photon laser photoionization of G runs[59] provide a test of the present description of the highest occupied base orbital in double-stranded sequences. The sequence-specific pattern arising from guanine photoionization that leads to the formation of 7,8-dihydro-8-oxoguanine (8-oxoG) via guanine radical cation hydration is shown in panel A of Figure 3. The sequence dependence of the 8-oxoG yield agrees with a description of photoionization in which guanine radical cations residing in G runs have lower energy than guanine cations flanked by other bases, and that, in G runs, the lowest energy photoionization is most likely at interior sites.

It has been postulated that the electron distribution in the highest occupied base orbital of G runs plays an important role in determining the site at which strand scission is induced by photocleaving agents, and it has been reported that photocleaving agents cause strand breaks that occur most favorably at the 5′-end of G runs.[12a,b] The reason for the discrepancy between the present results, indicating that the lowest energy ionization occurs in the interior of G runs, and the earlier reports that strand scission associated with photocleaving agents occurs at the 5′-end is not clear. One possibility is that the binding of photocleaving agents causes a shift of the HOBO from the interior to the 5′-end. Another is that the selectivity of the photocleaving mechanism depends on multiple factors including base deprotonation,[4,60] the formation of sugar radicals,[4] and the involvement of phosphate groups.[61] The present results suggest that, in considering these possibilities, the influence of the sugar−phosphate backbone on the energetics of base chemistry is not insignificant. This is likely to be true for the energies of neutral closed-shell states of the bases, and of cation and neutral radical states. In any case, in several sequences containing G runs with three guanines, riboflavin- and benzoquinone-induced photoionization occurs primarily at the interior guanine.[12e]

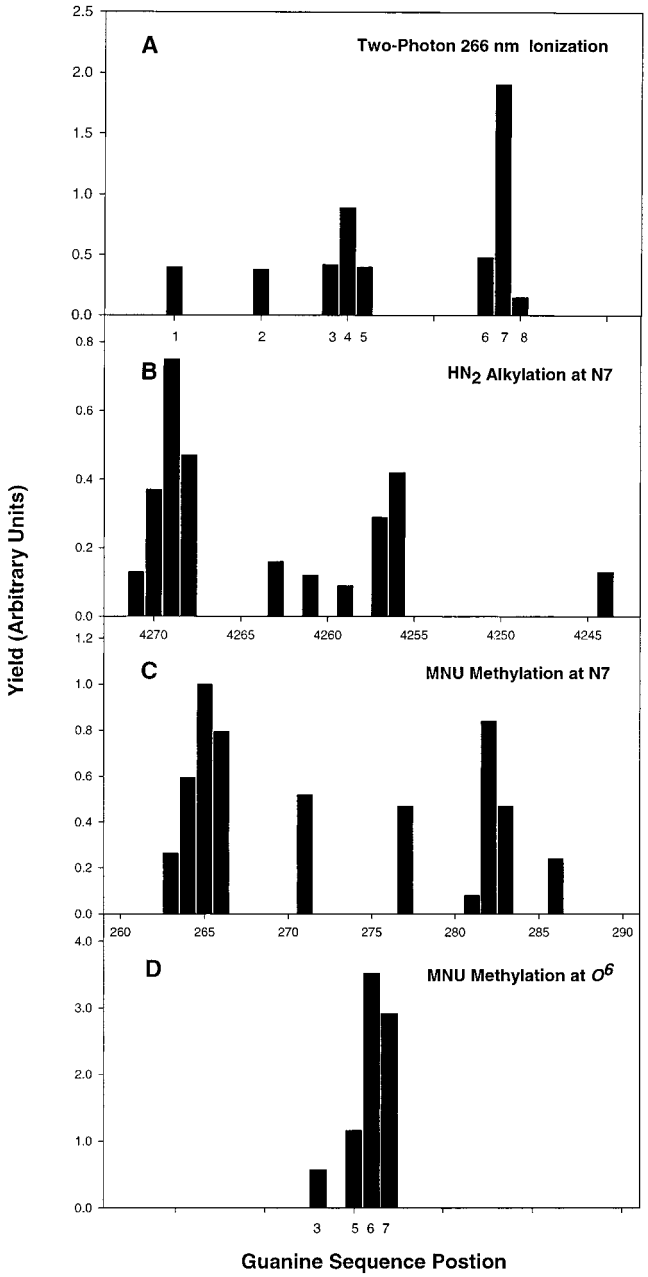


Figure 3. Photoionization and alkylation patterns for guanine in different DNA sequences. Panel A shows 7,8-dihydro-8-oxoguanine yields from hydration of guanine radical cations formed via two-photon 266 nm ionization (ref 59). Panel B contains the reaction pattern for HN2 at guanine N7 (ref 34). Panel C shows the pattern for MNU at N7 (ref 36). Panel D gives the pattern for MNU at O^6 (ref 33b). Numbering of the sequences begins from the 5′-end. See the references for details.

The present descriptions of the highest occupied base orbitals of G runs in oligonucleotides do not provide a simple explanation for the selectivity of strand scission induced by photocleaving agents. However, the finding that vertical ionization is most energetically favorable from the interior of double-stranded G runs is consistent with results from an earlier computational description of the MEP in the region near guanine N7 atoms in DNA sequences. While the ionization potential and the electrostatic potential are indicators of different tenden-

(58) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098.
(59) Spassky, A.; Angelov, D. *Biochemistry* **1997**, *36*, 6571.
(60) Steenken, S. *Chem. Rev.* **1989**, *89*, 503.
(61) Steenken, S.; Goldbergerova, L. *J. Am. Chem. Soc.* **1998**, *120*, 3928.

cies, this description of the MEP at different sites in G runs indicates that the potential is most negative at interior positions rather than at the ends.[34]

There is also correspondence between the descriptions of the highest occupied base orbitals given here, and the reactivity data of Figure 3, which shows experimental alkylation patterns[33b,34,36] for reactions of bis-2-chloroethylmethylamine (nitrogen mustard or HN2) and MNU. These electrophilic alkylating agents, which exhibit antitumor and carcinogenic activity, proceed through reactive methyl chloroethylaziridinium ions[34] and methane diazonium ions,[62] respectively. Figure 3 shows the sequence-specific reactivity of HN2 at guanine N7, and the reactivity of MNU at N7 and guanine O^6. According to the results of Figure 3, the reactivities of HN2 at N7 and of MNU at both N7 and O^6 is greatest for a guanine in a G run. Furthermore, for G runs containing three or four guanines, most reaction occurs at an interior site. In earlier investigations[9b,16,18] site-specific DNA reactivity toward electrophiles such as HN2 and MNU was considered by examining regions of high local polarizability. Here, sites of low base IP correspond to sites of high base polarizability. The greater reactivity of HN2 and MNU that occurs in the interior of G runs is consistent with the present description of IPs indicating that base electron donation and mobility is energetically most favorable at these interior sites.

In examining the reactivity data given in panels B, C, and D of Figure 3, it is important to note that, while the most reactive sites occur in the interior of the G runs, reaction also occurs at the ends. However, this is not inconsistent with a description of selectivity in which local base polarizability plays an important role. In this description, the HOBO makes a strong contribution to polarizability, but is not the sole contributor.

Conclusions

The main conclusions are the following.

(1) In almost all cases where comparisons are possible, results from SCF ab initio calculations with STO-3G, 3-21G, 6-31G, and 6-31G* basis sets and from semiempirical ZINDO calculations provide similar descriptions of the electron distribution associated with the highest occupied base orbital in oligonucleotides and model oligonucleotides containing stacked guanines. Results obtained at the ab initio SCF 3-21G level are consistent with earlier reports[12a,13,15] based on SCF 6-31G* and density functional calculations indicating that hydrogen bonding occurring in Watson−Crick guanine−cytosine base pairs and π interactions occurring in stacked guanine systems result in guanine ionization energies that are smaller than that of isolated guanine.

(2) Electron distributions of the highest occupied base orbital in G runs are sensitive to structure. In a model of a single-stranded G run containing three 9-methylguanines without a sugar−phosphate backbone, in which the 9-MeG's assume a B-DNA orientation, the present results, like earlier results,[12a−c] indicate that the lowest vertical IP is associated with the guanine at the 5′-end. The same result was obtained for a single-stranded oligonucleotide containing three guanines with the sugar−phosphate backbone, and with a Na^+ cation interacting strongly with each phosphate group. However, for double-stranded oligonucleotides that have sugar and phosphate groups and that more closely mimic native DNA, both semiempirical and ab initio SCF calculations indicate that vertical ionization is more favorable from guanines at interior positions than from guanines at the 5′- or 3′-end.

(3) The observation that guanine IPs in double-stranded G runs are smaller than those of guanines in other sequence contexts and the finding that the lowest energy IP in G runs occurs at interior guanines are consistent with earlier results indicating that the electrostatic potential is more negative in regions near interior guanines than at the ends. It is also consistent with the 266 nm two-photon ionization pattern, and with experimental DNA alkylation patterns of small electrophiles such as nitrogen mustard and *N*-methyl-*N*-nitrosourea. Both photoionization and reaction occur most favorably at interior guanine sites in G runs. Finally, the possible coupling between the electrostatic potential and HOMO electron distributions suggests that these electronic influences may sometimes act in concert. In these cases, the relative importance of orbital versus electrostatic control of sequence-specific DNA chemistry will be difficult to determine.

Acknowledgment. This paper is dedicated to the memory of Professor Michael Zerner, who provided assistance with the ZINDO calculations, and whose enthusiasm and friendship will be greatly missed. Technical assistance from Dr. Cindy Harwood and support from the American Cancer Society (Grant RPG-91-024-08-CNE) are gratefully acknowledged. Computer access time has been provided by the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign and the Computer Center of the University of Illinois at Chicago.

JA002523C

(62) Ford, G. P.; Scribner, J. D. *J. Am. Chem. Soc.* **1983**, *105*, 349.