# Ab Initio Study of the Ionization of the DNA Bases: Ionization Potentials and Excited States of the Cations

**Emilie Cauët,[†] Dominique Dehareng,[‡] and Jacques Liévin*[,†]**

*Service de Chimie Quantique et Photophysique, Université Libre de Bruxelles, CP 160/09, 50 Avenue F. D. Roosevelt, B-1050 Bruxelles, Belgium, and Centre d'ingénierie des protéines, Université de Liège, BAT. B6, 3 allée de la Chimie, B-4000 Liège 1, Belgium*

*Received: March 21, 2006*

The ionization of the four DNA bases is investigated by means of ab initio calculations. Accurate values of the gas-phase vertical and adiabatic ionization potentials (IP) are obtained at the MP2/6-31G(2d(0.8,$\alpha_d$),p) level of theory. The need of introducing extra polarization to the standard 6-31G(d,p) basis set is demonstrated by test calculations and an optimal value of $\alpha_d = 0.1$ is obtained. Ionization to electronically excited radical cations is also considered. The low-lying excited states of the cations are characterized for the first time. The topology of the corresponding potential energy surfaces is qualitatively described in terms of the stationary points (minima and saddle points) located on these surfaces. A conical intersection is characterized for the first time on the ground-state potential energy surface of all cations. It arises from the crossing of the adiabatic surfaces of the ground and first excited state at planar geometries. A nonplanar minimum is observed for the cytosine cation only. The geometry and electronic changes occurring along these surfaces are analyzed, leading to a comparison between the different nucleobase cations. The study of larger ionized systems related to DNA is rendered possible thanks to the optimized medium size basis set proposed in this work, as exemplified by the calculation of the IP of a stacked dimer of guanines.

## 1. Introduction

Radiation damage to DNA, induced by ionizing radiation, oxidizing agents, and photoirradiation, has recently attracted significant attention. In the last years, both experimental and theoretical works have been devoted to the investigation of the main oxidation reactions of DNA. Information about the created radicals was obtained, and a better knowledge of the mechanisms of the oxidation reactions in DNA was acquired.[1−3] Several studies of DNA-mediated charge transport are intended to probe oxidative damage to DNA at a distance from the initial oxidation site.[4−8] Long-range oxidative damage in DNA occurs indeed as a result of electron migration through the $\pi$-stacked DNA base pairs. Theoreticians have interpreted the dynamics of this electron-transfer process in terms of tunneling and hopping mechanisms.[9−13]

One of the results of the radiation damage to DNA is the formation of radical cations of DNA bases. Only few theoretical studies are available in the literature on these systems. Comprehensive ab initio studies of the electronic structure of the neutral DNA bases have already been published,[14−18] but no attention has been paid at our knowledge to the excited states of the corresponding cationic radicals. An accurate ab initio investigation of their electronic structure would thus be worthwhile. It would indeed provide a better understanding of the reactive processes involving cations of DNA bases in the framework of radiation damage.

Several ab initio and DFT studies aimed at estimating as accurately as possible the threshold energies needed for ionizing the DNA and RNA bases. They used different approaches to deal with the open shell structure of the radical cations. Sevilla et al.[19] have employed restricted HF and MP2 calculations with small double-$\zeta$ basis sets to investigate the gas-phase ionization potentials (IP) and electron affinities of the DNA bases. Crespo-Hernàndes et al.[20] adopted the PMP2 projection technique to correct the spin contamination induced by the unrestricted sheme. Their calculations performed with the 6-31++G(d,p) basis set are in good agreement with the corresponding experimental gas-phase values of the IPs of the four isolated DNA bases. Some authors[21−23] also calculated these IPs with B1LYP or B3LYP DFT, but obtained systematically underestimated values resulting from the use of the unrestricted formalism.

Recent experiments of ion impact with DNA study the formation of ionized DNA bases.[24,25] Less information is available in the literature about ionized clusters of DNA bases also produced in these experiments. Theoretical support on ionized clusters of DNA bases would thus also be welcome. The possibility of base stacking within these clusters should be explored. Unfortunately, the blowup of the computer costs with the molecular size renders such calculations difficult or even impossible. Only a few calculations performed at a low level of theory provide IP values of stacked DNA bases. Sugiyama et al.[26] and Prat et al.[27] used Koopmans' theorem HF/6-31G(d) calculations to evaluate the vertical IP of stacked dimers. The latter authors also used density functional theory (B3LYP/6-31G(d)). Some other Koopmans' theorem HF/6-31G(d) calculations were performed by Schumm et al.[28] on 6-mer model sequences centered on guanines. One of the major aims of the present work will be to determine on the isolated bases a level of calculation beyond Koopmans' approximation that could be applied to the study of small clusters.

* Address correspondence to this author. E-mail: jlievin@ulb.ac.be.
[†] Université Libre de Bruxelles.
[‡] Université de Liège.

10.1021/jp0617625 CCC: $33.50 
Published on Web 05/13/2006

In this work, we investigate the electronic structure of the radical cations of the DNA bases by ab initio calculations. In section 3, we present accurate values of the vertical and adiabatic IPs, calculated in the gas-phase using the restricted MP2 approach. The influence of the level of calculation (basis set, electron correlation, geometry optimization, restricted vs unrestricted solutions) on the calculated IP values is discussed. The need for extra polarization of the basis set is demonstrated, and an optimized version of the medium size 6-31G(2d(0.8,$\alpha_d$),p) basis set is proposed. The interest of such a medium size basis set for studying DNA base clusters is demonstrated on a stacked cluster of two guanines. The first excited states of the radical cations are characterized for the first time in section 4, and the topology of the low-lying potential energy surfaces is analyzed. The geometries of the stationary points on these surfaces and the corresponding electronic structures are presented and discussed in section 5.

## 2. Methods of Calculation

Hartree−Fock (HF), second-order Møller−Plesset pertubation theory (MP2)[29,30] and B3LYP[31,32] density functional theory (DFT) calculations have been used in a complementary way to investigate the ionization of the four DNA bases in the gas phase. All calculations were performed with the Gaussian 98[33] and Gaussian 03[34] program suite running on the Compaq alpha servers of the ULB/VUB computer center.

Energy differences ($\Delta E$), like ionization potentials and electronic excitation energies, were calculated at a hybrid level of theory in which geometry optimizations are performed at a lower level than the $\Delta E$ calculations. This is justified by the fact that geometries are less sensitive to the correlation and basis set effects than energies. The hybrid approach thus leads to a drastic savings in computer time in the geometry optimization step, without significant loss of accuracy.[35−37] The standard notation M2/B2//M1/B1 will be used hereafter to qualify such a level of theory, method M1 and basis set B1 being used for optimizing the geometries, and method M2 and basis set B2 for the single point calculations providing the $\Delta E$ values.

Geometry optimizations were carried out at the HF level of theory with use of the 6-31G(d,p) medium size polarized basis set.[38−40] When symmetry constraints were imposed to the geometry optmization, we always verified by a frequency calculation if the stationary point corresponded to a true minimum or a saddle point. Stability calculations[40,41] were also performed to diagnose eventual HF instabilities. For some chosen stationary points, starting from the corresponding HF geometry, we further optimized the geometry at a higher level of theory (MP2 and/or B3LYP) to investigate the effect of electron correlation on the calculated geometries and energies.

Correlated methods (M2) and a more extended basis set (B2) were used in all $\Delta E$ calculations. Concerning B2, the need for extra polarization of the basis set will be demonstrated in section 3.2, in which we propose an optimized version of the 6-31G-(2d(0.8,$\alpha_d$),p) basis set[36] for the calculation of IPs. This basis set of moderate size giving results of comparable accuracy than more extended basis sets from the literature has been adopted in all further calculations.

As for the method (M2), MP2 has been used more systematically than DFT, this preference being mainly guided by our work in progress on the ionization of stacked clusters of DNA bases. It is well-known indeed that the stability of such stacked $\pi$-systems is determined by the dispersion energy contributions, which are taken into account by MP2[36,42,43] but not by DFT.[44] Hybrid functionals such as X3LYP, yet dedicated to the description of van der Waals interactions,[45] seem indeed to fail at reproducing stacking stability.[46] Most of the calculations presented in this paper were thus performed with MP2, but calculations of IPs with B3LYP are also presented for comparison purpose.

The use of restricted versus unrestricted methods merits some comments, when as for IP calculations one calculates energy differences between closed- and open-shell systems. Crespo-Hernàndes et al.[20] showed that unrestricted MP2 (UMP2) overestimates by up to 0.75 eV the IPs of the DNA bases, as a result of the contamination of the cation doublet state by higher spin states. They obtain a good agreement with experiment (within 0.15 eV or less) by applying spin corrections by means of the PMP2 projection technique. In the present work, we rather used restricted open-shell methods (ROHF and ROMP2) in all calculations on the cations. However, with no analytic gradients being available in Gaussian at the ROMP2 level, we thus performed the MP2 geometry optimizations with UMP2, but calculate the $\Delta E$ values within the hybrid RMP2/B2//UMP2/B1 approach. Crespo-Hernàndes et al. also showed, in agreement with Bertran et al.,[47] that unrestricted B3LYP calculations do not overestimate the spin polarization, related to the spin contamination, as do MP2. They rather found that B3LYP calculations underestimate by a reasonable extent (within 0.2 eV) the IP values. We thus performed all our B3LYP calculations within the unrestricted formalism. To simplify the notation, MP2 will be used hereafter in place of RMP2 (ROMP2 for cation and RMP2 for neutral molecule) and B3LYP in place of UB3LYP.

The influence of the level of calculation (basis set, electron correlation, geometry optimization, restricted vs unrestricted solutions) on vertical and adiabatic IPs will be further discussed in sections 3.1 to 3.5, on the basis of systematic comparisons. In section 4, we will investigate in the same way the first excited state of the DNA bases cations. The use of single configuration approaches like HF and MP2 is, however, conditioned by the fact that the considered states are the lowest of a given symmetry in the considered molecular point group, and that no configuration mixing occurs in the studied regions of the PES. It will be shown below that these prescriptions are actually verified around the stationary point geometries. However, the characterization of conical intersections on the lowest PES has required the use of a multiconfigurational treatment. State-average CASSCF calculations[48−51] were used for that purpose, together with the unconstrained algorithm[52−54] implemented in Gaussian 03 for finding the minimum energy point on the conical intersection seam.

## 3. Ionization Potentials of the Isolated DNA Bases

**3.1. Ab Initio Calculation of Ionization Potentials.** The IP of an *n*-electron system, calculated at a given level of theory X, implies the following energy difference:

$$\mathrm{IP} = E_X(n-1;G_{n-1}) - E_X(n;G_n) \qquad (1)$$

where $E_X(n;G_n)$ is the energy of the *n*-electron neutral system calculated at a geometry $G_n$ and $E_X(n-1;G_{n-1})$ is the energy of the $(n-1)$-electron ionic species calculated at a geometry $G_{n-1}$. To calculate IPs, two separate energy calculations, performed on the neutral and cationic species, are thus in principle needed, except within the framework of the Koopmans theorem's approximation. Recall indeed that in this particular case, the vertical IP is evaluated as minus the HF energy of the molecular orbital of the neutral system from which an electron is extracted

**TABLE 1: Gas-Phase Ionization Potentials of Guanine (in eV) and Comparison of Various Levels of Calculation with the Experimental Value**

| ionization | M2/B2[a,b] | IP (eV) |
|---|---|---|
| vertical | Koopmans/6-31G(d,p) | 8.10 |
| | HF/6-31G(d,p) | 7.22 |
| | HF/cc-PVTZ | 7.24 |
| | HF/aug-cc-PVDZ | 7.30 |
| | HF/6-311++G(d,p) | 7.39 |
| | B3LYP/6-31G(d,p)[c] | 7.65 |
| | B3LYP/6-311++G(d,p)[c] | 8.07 |
| | UMP2/6-31G(d,p) | 8.35 |
| | PMP2/6-31G(d,p) | 7.91 |
| | MP2/6-31G(d,p) | 7.88 |
| | MP2/cc-PVTZ | 8.25 |
| | MP2/aug-cc-PVDZ | 8.30 |
| | MP2/6-311++G(d,p) | 8.27 |
| | exptl[d] | 8.24 |
| adiabatic | MP2/6-31G(d,p) | 7.42 |
| | MP2/aug-cc-PVDZ | 7.84 |
| | MP2/6-311++G(d,p) | 7.82 |
| | exptl[e] | 7.77 |

[a] Calculated in this work by using the hybrid M2/B2//M1/B1 approach, with HF/6-31G(d,p) for M1/B1, unless otherwise indicated. [b] HF and MP2 refer to restricted and B3LYP to unrestricted calculations (see text section 2). [c] Calculated at the B3LYP/6-31G(2d(0.8,0.1),p) geometry. [d] Reference 61. [e] Reference 62.

to form the cation. This very simple way of estimating IP from a single HF calculation on the neutral system has been widely used in the literature. It gives reasonably accurate results for many systems, as a consequence of fortuitous compensations of errors (see next section for examples). Fundamentally Koopmans' values do not take electron relaxation accompanying ionization into account. Moreover, being based on HF theory, they suffer from the lack of electron correlation.

Two different energy quantities are usually related to ionization, corresponding to the vertical and adiabatic IPs, respectively. The vertical IP, to be referred to as $IP_{vert}$, corresponds to a vertical excitation occurring at the equilibrium geometry $G_n$ of the neutral molecule. The ab initio determination of $IP_{vert}$ thus implies the calculation of the neutral and ionic species at the same geometry ($G_{n-1} = G_n$ in (1)). The adiabatic IP, to be referred to as $IP_{adia}$, corresponds to an ionizing excitation from the equilibrium geometry of the neutral species to the equilibrium geometry of the cation. The latter process thus implies a separate geometry optimization on both species, taking the geometry relaxation accompanying ionization into account.

Many effects are in competition in an IP calculation, which makes such calculations challenging. Indeed, in addition to the geometry relaxation, one also observes a relaxation of the electron density. The latter tends to be more contracted on the atomic nuclei in the cation than in the neutral system. This means that accurate predictions of IPs are only reached when the level of calculation is able to take this electron relaxation properly into account. The method of calculation is thus of major importance, by the way it quantitatively accounts for the correlation energy difference between the $n$ and $(n - 1)$ electrons systems. The flexibility of the basis set is also very important, in particular in the molecular frontier region where the electron probability more significantly changes upon ionization.

Before addressing in the next subsection the problem of basis set optimization, let us quantify on a chosen example the relative importance of the different effects just mentioned. Table 1 reports the vertical and adiabatic IPs of guanine calculated at different levels of theory. These results are showing up: (i) the importance of the electronic relaxation, which can be estimated (−0.88 eV) from the difference between Koopmans' and HF values (recall that Koopmans' approximation consists of calculating the IP from the neutral HF solution); (ii) the order of magnitude of the basis set effect at HF (up to +0.17 eV) and MP2 (up to +0.42 eV) levels (the effect of additional polarization and/or diffuse functions with respect to the reference 6-31G(d,p) basis appears clearly, when using cc-PVTZ,[55,56] aug-cc-PVDZ,[57] or 6-311++G(d,p)[58]); (iii) the importance of the correlation effects (+0.66 to +1 eV depending on the basis set); (iv) the interplay between basis set extension and correlation effects, as shown by the differences observed in (ii) and (iii); (v) the fortuitous agreement between Koopmans' and experimental values, explained by a compensation between electron relaxation and electron correlation effects; (vi) the effect of the geometry relaxation (−0.46 eV), which can be estimated from the ($IP_{vert} - IP_{adia}$) difference; (vii) the importance of the spin contamination introduced by unrestricted MP2 (0.47 eV) (note the good agreement between restricted and projected PMP2); and (viii) the underestimate of the IP by B3LYP (see section 2), with a basis set effect of similar amplitude as that with MP2 (0.4 eV).

This example shows how the competition between different effects changes the IP values. It also demonstrates that a correlated approach like MP2 used with a basis set of reasonable size is nevertheless able to provide accurate predictions.

**3.2. Optimization of a Basis Set for the Calculation of Ionization Potentials.** The conclusion of the previous section is quite pessimistic for the study of larger ionized systems, for instance clusters of the DNA bases. The computer cost of an MP2 calculation indeed formally scales as $O(m^4)$ to $O(m^5)$, with $m$ the basis set size. Taking again the monomer of guanine as an example, the CPU time of an MP2/6-31G(d,p) ($m = 179$) calculation scaling in $m^4$ ($m^5$) is multiplied by a factor of 6 (9), 8 (13), and 25 (56) when using 6-311++G(d,p) ($m = 277$), aug-cc-PVDZ ($m = 298$), or cc-PVTZ ($m = 400$), respectively. An additional factor of $n^4$ ($n^5$) is of course to be considered, whatever the basis set, when going from the monomer to an $n$-cluster. This simple evaluation shows how the computer time explodes with basis set and cluster sizes. It would thus be worthwhile to develop a basis set of intermediate size adapted to the description of ionization. That is what we did, following the same ideas as in our previous work on the stabilities of biomolecular complexes.[36,37] In that previous work, we kept the medium size basis set 6-31G(d,p), for describing the bonding interactions within the monomers, but we simply augmented it by diffuse polarization functions intended to improve the representation of nonbonding intermolecular interaction. For that purpose, a single polarization d Gaussian function was added on the second row atoms (C, N, and O), leading to the so-called 6-31G(2d(0.8,$\alpha_d$),p) basis set. In such a basis set, each heavy atom is thus polarized by two d Gaussian functions. The first one with an exponent of 0.8 corresponds to the polarization function of the standard 6-31G(d,p) basis set, optimized for describing the short-range interactions occurring in bonded systems. The second has a Gaussian exponent $\alpha_d$ to be optimized for the properties of interest. We found[36] that the value $\alpha_d = 0.2$ significantly improved the description of three types of intermolecular interactions occurring in biomolecular complexes: stacking, H-bonding, and cation−$\pi$ interactions. Similar basis sets are used in the literature[59] for dealing with biomolecules, following the work of Šponer and co-workers,[60] who proposed an $\alpha_d$ value of 0.25.

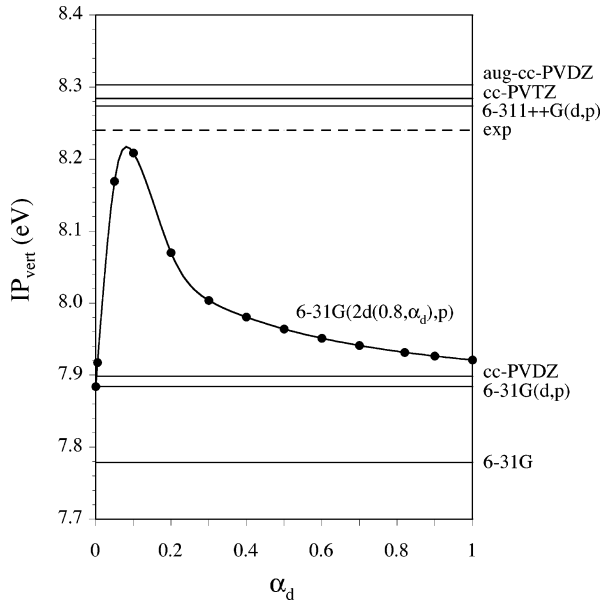


**Figure 1.** MP2/6-31G(2d(0.8,$\alpha_d$),p) vertical ionization potentials (eV) of guanine as a function of the $\alpha_d$ exponent of the d-polarization functions on C, N, and O atoms. The vertical IPs obtained with extended basis sets of the literature (6-311++G(d,p), aug-cc-PVDZ, and cc-PVTZ) are shown with horizontal lines. The experimental value corresponds to the dotted line.

In this paper we investigate how the same kind of basis set could also improve the description of the electronic changes accompanying ionization. For that purpose we have optimized the $\alpha_d$ exponent in the following way. The vertical ionization potentials of guanine and adenine have been calculated at the MP2/6-31G(2d(0.8,$\alpha_d$),p)//HF/6-31G(d,p) level, as a function of the $\alpha_d$-exponent. The corresponding curve is plotted in Figure 1 for guanine. One observes a spectacular change of the IP value between $\alpha_d = 0$ and 0.2, with a maximum (8.21 eV) at $\alpha_d = 0.1$. The latter value is close to that obtained with more extended basis sets of the literature already used above (8.27, 8.30, and 8.25 eV for 6-311++G(d,p), aug-cc-PVDZ, and cc-PVTZ, respectively). It agrees also pretty well with the corresponding experimental value ($8.24 \pm 0.03$ eV).[61] The exponent $\alpha_d = 0.1$ is thus optimal for calculating the $IP_{vert}$ of guanine. Compared to the value of 7.88 eV obtained at $\alpha_d = 0$, which corresponds to the standard 6-31G(d,p) basis set, the gain introduced by the additional polarization is significant (0.33 eV).

Figure 1 thus clearly demonstrates the sensitivity of the IP with the basis set flexibility at long range. Three classes of basis sets are observed: (i) the unpolarized valence-DZ 6-31G (IP $\sim$ 7.8 eV), (ii) the singly polarized valence-DZ 6-31G(d,p) and cc-PVDZ (IP $\sim$ 7.9 eV), and (iii) the multiply polarized and/or augmented basis sets (IP between 8.2 and 8.3 eV). Our optimized 6-31G(2d(0.8, 0.1),p) basis set belongs to the third class and thus seems to provide a diffuse polarization of equal efficiency to that encountered in the larger basis sets of the literature. The computer costs are, however, spectacularly smaller, given the low value of *m* (234 for guanine) as compared to the other basis sets (see above). The scaling factor with respect to 6-31G(d,p) is consequently only of 3 (4).

The $\alpha_d$ exponent has also been optimized in the same way for the $IP_{vert}$ of adenine. The $IP_{vert}(\alpha_d)$ curve presents a very similar shape to that observed in Figure 1 for guanine. The maximum (8.63 eV) also located at $\alpha_d = 0.1$ is within 0.1 eV from the larger aug-cc-PVDZ and 6-311++G(d,p) basis sets. The gain of 0.29 eV with respect to 6-31G(d,p) is of the same order of magnitude as for guanine, which confirms the efficiency of the diffuse d polarization with the same exponent value. A good agreement is also observed with the experimental value.

**3.3. MP2 Calculation of the Vertical and Adiabatic IPs of the Four DNA Bases.** Without further basis set optimization we have calculated the vertical and adiabatic IPs of the four DNA bases at the MP2/6-31G(2d(0.8,0.1),p)//HF/6-31G(d,p) level. No constraint of planarity has been imposed in the geometry optimization carried out on the neutral and cationic species. Geometries of the cations will be discussed in more detail in sections 4.1 and 5, but let us simply say here that for neutral systems, in agreement with previous work,[20,35] we find that thymine is planar (except the methyl group hydrogens), whereas the amino groups in cytosine, guanine, and adenine are not. We also confirm the planarity of all radical cations, previouly reported by Improta et al.[21]

The calculated IP are collected in Table 2, where they are compared to the corresponding gas-phase experimental results.[61,62] Other levels of calculation are also given for comparison purpose: M2/B2//HF/6-31G(d,p) level, with M2 = HF and MP2 and B2 = aug-cc-PVDZ and 6-311++G(d,p). Koopmans' values are also given for information for the optimized basis set only. Comparison is also made with theoretical results from the literature: the HF and MP2/6-31+G(d) calculations of Sevilla et al.,[19] the PHF and PMP2 results of Crespo-Hernàndes et al.,[20] and the OVGF/6-311G(d,p) results of Close.[63] Our MP2/6-31G(2d(0.8,0.1),p) results are found to be in good agreement (within 0.2 eV) with the gas-phase experimental data. They lie in most cases within the experimental uncertainties, the largest deviations being observed for adenine and cytosine for which all high-level calculations overestimate the IPs. Note that Crespo-Hernàndes et al.[20] consider with criticism the experimental values obtained for cytosine. The gain introduced by the $\alpha_d = 0.1$ polarization with respect to 6-31G-(d,p) is around 0.2−0.3 eV for all bases and for both $IP_{vert}$ and $IP_{adia}$. The trends pointed out in the previous section are thus confirmed. The results also compare well with the larger basis sets results (within 0.1 eV) and with those of the literature. Let us note the large value of the vertical IP for thymine obtained by Sevilla et al.[19] This suggests us that the energy of the ionic species could correspond to the first excited state of the cation. This point will be confirmed latter in section 4.4 on the basis of the calculations of the excited states of the radical cations.

**3.4. B3LYP Calculations of the Vertical and Adiabatic IPs of the Four DNA Bases.** The use of the 6-31G(2d-(0.8,0.1),p) basis set has been tested on B3LYP DFT calculations. Geometries were optimized at this level. The results are reported in Table 3, where a comparison with other basis sets is provided. Improta et al.[21] use B1LYP instead of B3LYP. All results of Table 3, involving augmented basis sets, agree within less than 0.2 eV. These results demonstrate again the efficiency of the 6-31G(2d(0.8,0.1),p) basis set. The gain brought by the additional $\alpha_d = 0.1$ polarization is of 0.2−0.3 eV, as observed at the MP2 level. The systematic underestimation by 0.1−0.3 eV of the IPs by unrestricted DFT is also confirmed.

**3.5. Qualitative Interpretation of the $\alpha_d = 0.1$ Polarization.** The results presented in the previous section clearly demonstrate that the addition of diffuse functions to the basis set significantly improves the calculated IP values. The kind and the number of functions added to a reference VDZ singly polarized basis set varies, however, among the different tested basis sets, but the corresponding gain in the IP values is, however, of the same order of magnitude. In the case of our 6-31G(2d(0.8,0.1),p) basis set one only adds a single d function on all heavy atoms. Other basis sets of the literature involve more diffuse functions, even

**TABLE 2: Gas-Phase Vertical and Adiabatic Ionization Potentials of Guanine, Adenine, Cytosine, and Thymine (in eV) and Comparison of various levels of calculation with the corresponding experimental values.**

| | Gua | | Ade | | Cyt | | Thy | |
|---|---|---|---|---|---|---|---|---|
| M2/B2[a] | $IP_{vert}$ | $IP_{adia}$ | $IP_{vert}$ | $IP_{adia}$ | $IP_{vert}$ | $IP_{adia}$ | $IP_{vert}$ | $IP_{adia}$ |
| HF/6-31G(d,p) | 7.22 | 6.70 | 7.59 | 7.19 | 9.14 | 7.82 | 8.41 | 7.92 |
| MP2/6-31G(d,p) | 7.88 | 7.42 | 8.34 | 7.94 | 8.80 | 8.49 | 8.87 | 8.60 |
| MP2/6-31G(d,p)//UMP2/6-31G(d,p) | - | 7.53 | - | - | - | 8.53 | - | - |
| Koopmans/6-31G(2d(0.8,0.1),p) | 8.29 | - | 8.53 | - | 9.34 | - | 9.65 | - |
| HF/6-31G(2d(0.8,0.1),p) | 7.34 | 6.83 | 7.67 | 7.28 | 9.25 | 7.92 | 8.50 | 8.01 |
| MP2/6-31G(2d(0.8,0.1),p) | 8.21 | 7.75 | 8.63 | 8.23 | 9.07 | 8.78 | 9.13 | 8.87 |
| HF/aug-ccpVDZ | 7.30 | 6.81 | 7.63 | 7.26 | 9.26 | 7.91 | 8.47 | 7.98 |
| MP2/aug-ccpVDZ | 8.30 | 7.84 | 8.75 | 8.33 | 9.19 | 8.88 | 9.23 | 8.94 |
| HF/6-311++G(d,p) | 7.39 | 6.88 | 7.74 | 7.35 | 9.32 | 7.99 | 8.56 | 8.05 |
| MP2/6-311++G(d,p) | 8.27 | 7.81 | 8.72 | 8.31 | 9.14 | 8.84 | 9.20 | 8.88 |
| HF/6-31+G(d)[b] | 7.29 | 6.87 | 7.73 | 7.36 | 8.45 | 7.99 | 8.99 | 8.10 |
| MP2/6-31+G(d)[b] | 8.04 | 7.66 | 8.58 | 8.18 | 8.82 | 8.74 | 10.33 | 8.85 |
| PHF/6-31++G(d,p)[c] | 6.97 | 6.48 | 7.36 | 7.09 | 7.69 | 7.35 | 8.21 | 7.79 |
| PMP2/6-31++G(d,p)[c] | 8.33 | 7.9 | 8.62 | 8.23 | 8.69 | 8.78 | 9.07 | 8.74 |
| OVGF-MP2/6-311G(d,p)[d] | 8.13 | - | 8.49 | - | 8.79 | - | 9.13 | - |
| exptl[e] | 8.24 | 7.77 | 8.44 | 8.26 | 8.94 | 8.68 | 9.14 | 8.87 |

[a] Calculated in this work at the M2/B2//HF/6-31G(d,p) level, unless otherwise indicated. [b] From ref 19, geometries optimized at the ROHF/6-31G(d) level. [c] From ref 20, geometries and energies calculated at the same level. [d] From ref 63, geometries optimized at the MP2/6-311G(d,p) level. [e] From refs 61 and 62.

**TABLE 3: Gas-Phase Ionization Potentials of the DNA Bases (in eV) from DFT Calculations**

| | Gua | | Ade | | Cyt | | Thy | |
|---|---|---|---|---|---|---|---|---|
| method/basis set | $IP_{vert}$ | $IP_{adia}$ | $IP_{vert}$ | $IP_{adia}$ | $IP_{vert}$ | $IP_{adia}$ | $IP_{vert}$ | $IP_{adia}$ |
| B3LYP/6-31G(d,p)[a] | 7.65 | 7.32 | 7.99 | 7.77 | 8.35 | 8.23 | 8.72 | 8.49 |
| B3LYP/6-31G(2d(0.8,0.1),p)[a] | 7.94 | 7.61 | 8.23 | 8.02 | 8.64 | 8.52 | 8.97 | 8.74 |
| B3LYP/6-311++G(d,p)[a] | 8.02 | 7.69 | 8.33 | 8.11 | 8.74 | 8.63 | 9.05 | 8.81 |
| B1LYP/6-31+G(d,p)[b] | 7.89 | 7.52 | 8.16 | 7.95 | 8.62 | 8.47 | 8.90 | 8.66 |
| B3LYP/6-31++G(d,p)[c] | 7.98 | - | 8.26 | - | 8.69 | - | 9.01 | - |
| exptl[d] | 8.24 | 7.77 | 8.44 | 8.26 | 8.94 | 8.68 | 9.14 | 8.87 |

[a] Calculated in this work at the B3LYP/B2//B3LYP/6-31G(2d(0.8,0.1),p) level. [b] From ref 21; geometries optimized at the B1LYP/6-311G(d,p) level. [c] From ref 22; geometries and energies calculated at the same level. [d] From refs 61 and 62.

**TABLE 4: Spatial Extent $\langle R^2 \rangle$ Values (in au) Calculated from the MP2 Densities of the Neutral and Cationic Species of the Four DNA Bases[a]**

| | Ade | | Thy | | Cyt | | Gua | |
|---|---|---|---|---|---|---|---|---|
| method/basis set | neutral | cation | neutral | cation | neutral | cation | neutral | cation |
| MP2/6-31G(d,p) | 1146.39 | 1122.33 | 1116.74 | 1092.62 | 834.33 | 808.54 | 1487.46 | 1458.41 |
| MP2/6-31G(2d(0.8,0.1),p) | 1150.12 | 1125.46 | 1120.14 | 1093.96 | 837.62 | 809.62 | 1491.91 | 1460.31 |

[a] All values are calculated at the optimized geometry of the neutral system.

on hydrogens for some basis sets: (1s+1p) and TZ valence flexibility in 6-311++G(d,p), (1s+1p+1d) in aug-cc-PVDZ, and finally 1d in 6-31G(2d(0.8,0.1),p). To better understand the effect of the polarization on the vertical ionization we report in Table 4 the expectation values $\langle R^2 \rangle$, calculated by Gaussian from the MP2 densities. This property allows us to quantify the spatial extent of the electronic wave function. It has already been used with success by Eisfeld et al.[64] for investigating the valence, Rydberg, and mixed valence/Rydberg character of excited electronic states. We use it here in a similar way to measure the electronic relaxation accompanying ionization. One sees that the spatial extent drops by 24 up to 32 au when the system ionizes. This corresponds to an average reduction of this extent by about 2.5%, which gives a measure of the expected density contraction in the cation. Comparing $\langle R^2 \rangle$ values obtained from the 6-31G(d,p) and 6-31G(2d(0.8,0.1),p) allows us to follow the changes occurring in the electronic structure of the neutral and ionic systems when extra polarization is provided: the spatial extent increases on average by 3.7 and 1.8 for the neutral and ionic species. This means that the neutral system takes more advantage of the $\alpha_d$ polarization than the cation, which explains the corresponding increase of the IP value. Looking again to Figure 1, one sees that the peak in the IP($\alpha_d$) curve is acute, suggesting that the spatial region in which electronic rearrangement accompanying ionization is quite localized and corresponds to the long tail part of the wave function. We also compared the $\langle R^2 \rangle$ values for the set of extended basis sets considered in the previous sections. We do not detail here these results, but only report the conclusion, which is that the spatial extents calculated with all extended basis sets are close to those obtained with our 6-31G(2d(0.8,0.1),p) basis set. This means that in all these basis sets the region of space important for ionization is properly covered, whatever the number and nature of the diffuse functions. Let us point out the sensitivity of the spatial extent, which is an interesting indicator for investigating ionization.

**3.6. IP of Clusters of DNA Bases.** Our results show that the addition of a diffuse d function to the 6-31G(d,p) basis set

improves as well the interaction energies of stacked bases[36,37] as the IPs of isolated DNA bases. Optimal values of the Gaussian $\alpha_d$-exponent are found to be of the same order of magnitude, 0.2 and 0.1 for interaction energies and IPs, respectively. This result suggests that this kind of basis set is well suited for studying the ionization of stacked clusters of DNA bases. As an example we calculated the ionization potential of a stacked dimer of guanine. The geometrical structure, derived from the X-ray structure of tc3 transposase (protein code 1TC3, residues A7 and A8), is taken from our previous work on stair motifs at the protein−DNA interfaces.[36,65] The MP2 vertical IP is calculated to be of 7.67, 7.86, and 8.01 eV, when using the 6-31G(d,p) and the 6-31G(2d(0.8,$\alpha_d$),p) basis sets with $\alpha_d = 0.2$ and 0.1, respectively. These values are to be compared with the corrresponding IPs of isolated guanine of 7.88, 8.07, and 8.21 eV, respectively. The basis set effect on the IP values follows the trends pointed out in section 3.5, but it is striking to observe that it is the same on the isolated guanine as on the dimer, the IP being systematically smaller by 0.2 eV in the latter case. A very simple explanation of this finding is that, as confirmed by a Mulliken population analysis of the dimer, 98% of the positive charge is born by one of the guanines. The electronic relaxation of the ionization, which we demonstrated to be sensitive to the diffuse basis set extension, will thus essentially occur on this single guanine. Let us note that the use of an exponent of 0.1 or 0.2 changes the IP by 0.15 eV. The former value is certainly better for predicting the IP value, but the latter is more adapted for describing the stack stability. We would thus recommend the use of $\alpha_d = 0.2$ in all studies concerned by the stack stability or by the reactivity in ionized stacked systems. Let us note moreover that the vertical IP of isolated guanine, calculated with $\alpha_d = 0.2$, is not too far from the experimental value (see Figure 1).

Other authors also calculated the IP of a stacked dimer of guanine, but for *N*-methylated bases. A direct comparison is unfortunately not possible with our MP2 results on the dimer, because we use normal guanines and also a different stack geometry. Sugiyama et al.[26] obtained an IP value of 7.75 eV using Koopmans' HF/6-31G(d). The latter value for the dimer is 0.5 eV smaller than that for an isolated guanine. Prat et al.[27] obtained for the vertical IP of a similar dimer a value of 7.34 eV with Koopmans' HF/6-31G(d) and 6.64 eV with B3LYP/6-31G(d). The corresponding IPs decrease from the dimer to the monomer by 0.4 and 0.7 eV, respectively. It is difficult to estimate the accuracy of the Koopmans' values, subject to the error compensation already discussed above. Concerning the B3LYP calculations, we have shown that the monomer's IP is systematically underestimated. We must recall besides that this approach is unable to take the dispersion attraction in a stack into account. Our MP2 results can be considered of higher accuracy given the use of the MP2 method and of our optimized basis set. The method of calculation opens the door to the study of ionization in larger systems.

## 4. Excited States of the Radical Cations of DNA Bases

**4.1. Characterization of the First Excited States.** We reported in section 3.3 that our geometry optimizations performed on the ground states of the radical cations demonstrate their planarity, in agreement with the DFT calculations of Improta et al.[21] Such planar systems belong to the $C_s$ point group symmetry, in which two irreductible representations A′ and A″ coexist. The ground state of all four DNA bases is found to be of $^2$A″ symmetry. Following the comments made in section 2, we exploited the possibility of characterizing the lowest electronic state of the second symmetry $^2$A′, by means of the HF and MP2 monoconfigurational approaches used in the previous sections. Following the same computational strategy, we first optimized the geometry at the HF/6-31G(d,p) level and found a stationary point for all radical cations. To ascertain the nature of these stationary points we calculated the second derivative Hessian matrix and determined the vibrational frequencies. The A′ states of $Ade^+$, $Thy^+$, and $Gua^+$ do not have imaginary frequencies, while $Cyt^+$ has one imaginary frequency. This means that the $^2$A′ stationary points of the former three cationic bases correspond to real minima on the corresponding Born−Oppenheimer PES. For $Cyt^+$, the stationary point is a saddle point, suggesting the possible existence of nonplanar equilibrium structures. Further geometry optimizations were then carried out, starting with a symmetry broken geometry ($C_1$ symmetry) close to the $^2$A′ saddle point of $Cyt^+$, and a new minimum, corresponding to a nonplanar $^2$A structure, has been found. The possible existence of nonplanar minima for the three other bases was also considered, and searchs were undertaken but without success.

All the stationary points we just characterized correspond to the lowest electronic states of their symmetry ($^2$A′, $^2$A″, or $^2$A), but we cannot assert at this point that the lowest $^2$A′ states actually correspond to the first excited state of the planar cations. To answer to this question we need to evaluate for each cation the relative energy position of the second state of $^2$A″ symmetry (2 $^2$A″) with respect to the first $^2$A′ (1 $^2$A′). SDCI/6-31G(2d-(0.8,0.1),p) calculations were performed for that purpose at the previously optimized geometries of the A″ and A′ stationary points. A limited active space involving the two lowest occupied molecular orbitals of a′ and a″ symmetries has been used. These calculations demonstrate that, for all cations, the 1 $^2$A′ state is significantly more stable than the 2 $^2$A″ state (by 2.59, 1.50, 2.96, and 3.87 eV for $Gua^+$, $Ade^+$, $Cyt^+$, and $Thy^+$, respectively). This confirms thus clearly that the 1 $^2$A′ is the first excited state for all the cations. The lack of configuration mixing (absolute value of the main configuration coefficient larger than 0.97) also confirms the validity of the monoconfiguration zeroth order approaches in the vicinity of the considered stationary points.

**4.2. Characterization of Conical Intersections on the Lowest Potential Energy Surfaces.** The crossing of the lowest $^2$A′ and $^2$A″ adiabatic PES suggests the occurrence of conical intersections (CI) on the lowest potential energy surfaces for all cations. Indeed the coupling matrix element $H_{12} = \langle \Psi_1^d | \hat{H} | \Psi_2^d \rangle$ within the diabatic representation $\{\Psi_1^d, \Psi_2^d\}$ vanishes for symmetry reasons at planar geometries, but not necessarily when nonplanar displacements are considered. The two directions $x_1$ and $x_2$ describing the typical double cone form of the potential energy function have been determined by means of the algorithm of Robb and co-workers.[66,67] This means that the point of minimum energy has been searched in the $(n-2)$-dimensional hyperline (with $n$ the number of nuclear coordinates) in which the ground and excited states are degenerated. The two orthogonal directions defining the conical intersection in the adiabatic basis $\{\Psi_1^a, \Psi_2^a\}$ are given by:

$$x_1 = \frac{\partial(E_1 - E_2)}{\partial q(a')} \text{ and } x_2 = \left\langle \Psi_1^a \left| \frac{\partial}{\partial q(a'')} \right| \Psi_2^a \right\rangle \qquad (2)$$

where $x_1$ and $x_2$ are the gradient difference and the nonadiabatic coupling vectors respectively, and $q$(a′) and $q$(a″) define nuclear displacements preserving or not the planar geometry, respectively.

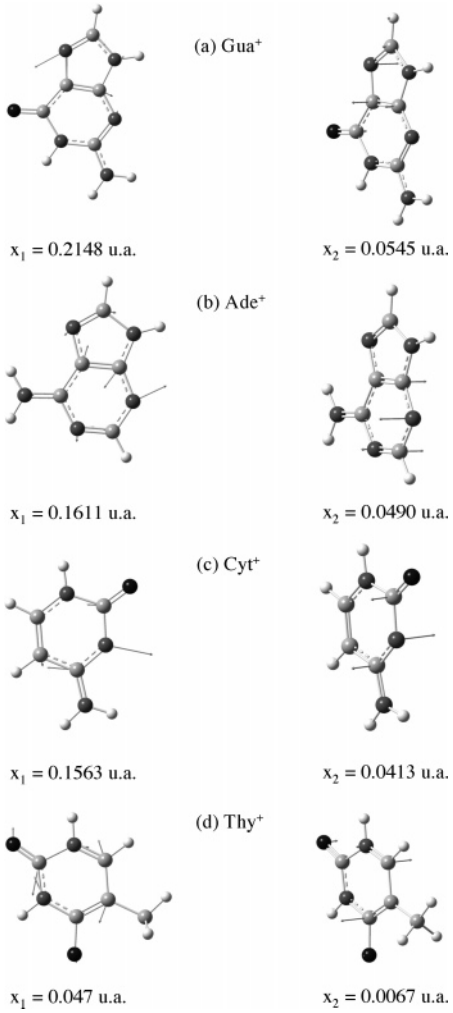


**Figure 2.** Normalized modes of displacements (in au), calculated at CASSCF(4,7)/6-31G(d,p) level, corresponding to vectors $x_1$ and $x_2$ at the conical intersection points for the different cations. Arrows have been multiplied by a factor of 1.5. The values of the norm of $x_1$ and $x_2$ (in au) are given for each intersection point. Carbon, nitrogen, and oxygen atoms are drawn in light gray, dark gray, and black, respectively.

Such calculations were performed at the state-average CASSCF(4,7)/6-31G(d,p) level (7 active electrons in 4 active MOs). The conical intersection point has been located for all cations and the displacements vectors $x_1$ and $x_2$ were determined. Figure 2 shows the normalized nuclear displacements corresponding to these vectors. The values of the norm of $x_1$ and $x_2$ at the intersection points are also indicated on the figure. Their values ($0.047 \leq x_1 \leq 0.215$ au and $0.007 \leq x_2 \leq 0.054$ au) give qualitative information on the diabatic vs adiabatic behavior of the corresponding systems in the vicinity of the intersections. Figures3 and 4 allow a visualization of the conical intersection topologies. Figure 3 shows for each cation the energy variation of the two adiabatic interacting states as a function of $(x_1 - x_{1e})$ and $(x_2 - x_{2e})$, $x_{1e}$ and $x_{2e}$ referring to the geometry at the intersection point. The double cone structure clearly appears, with its common apex taken as the origin of the energy scale. Note that for reasons of readability of the graphs this scale is different for the $x_1$ and $x_2$ variations. The conical structures are better visualized in the 3D pictures proposed in Figure 4 for $Cyt^+$ and $Thy^+$. These two examples has been chosen to illustrate two typical cases of conical topologies. The cone structure is in all cases symmetrical with respect to $x_2$, for symmetry reasons, but the same is not true for the $x_1$ variation, as can be seen in Figure 3. One observes, however, a quasisymmetry in the case of $Cyt^+$ only. In this particular case only the slopes at the origin of the $x_1$ variation have opposite signs, and one can expect that the two planar stationary points ($^2A''$ and $^2A'$) are located on both sides of the conical intersection. The other cations exhibit a different situation, with slopes in $x_1$ of equal signs, suggesting then the existence of two planar stationary points on the same side of the conical intersection. $Thy^+$ differs moreover from the other cations by the smaller values of the coupling vectors, with as a result the very flat conical intersection topology seen in Figure 4.

**4.3. Global Topology of the Lowest Potential Energy Surfaces.** The results presented in previous sections provide a global picture of the topology of the lowest potential energy surfaces of the nucleobase cations. Conical intersections on these surfaces are characterized for the first time. Such features are of particular interest, they can indeed govern dynamic processes following an energy deposit (collisons or photons) of the order of magnitude of the energy difference between the stationary points, as in the case of the neutral DNA bases, for which ultrafast nonradiative decay following photon excitation has been observed[68] and interpreted by means of theoretical calculations.[14,16−18]

The four cations are found to obey to three distinct cases of topology. Let us label minima and saddle points by M and S, respectively, and number them accordingly.

A first case is given by $Cyt^+$, which is the only cation to have a nonplanar minimum (M2) in addition to the ground state $^2A''$ planar minimum (M1). It is also the only one to have a saddle point (S1) and not a minimum as a planar $^2A'$ stationary point. The topology of the conical intersection energy curve along $x_1$ points out the intermediate position of the CI between the two critical points M1 and S1.

A second case regroups $Ade^+$ and $Gua^+$, exhibiting two planar minima only, corresponding to the $^2A''$ (M1) and $^2A'$ (M2) states, respectively. Despite coupling vectors of the same magnitude as $Cyt^+$, they differ, however, by the topology of the conical intersection energy curve along $x_1$: apparently, the two minima could lie on the same side by reference to the position of the CI.

$Thy^+$ forms the last case, with a conical intersection topology along $x_1$ like that of $Ade^+$ and $Gua^+$, but with significantly weaker couplings than all other cations. This means that one could qualitatively expect a rather adiabatic behavior for this cation while all the others should behave more diabatically. This means that the change of spin or charge excess position on the skeleton without electronic excitation (i.e., by remaining on the lowest energy surface) is much easier in $Thy^+$ than in the other base cations, in the region of the CI.

The nonplanarity of a critical point in the vicinity of a CI is obviously related to the topology of the energy curve along the out-of-plane coordinate, i.e., $x_2$. At first sight, $Cyt^+$ does not differ from $Ade^+$ and $Gua^+$ on this feature, nor on the value of the coupling elements. The effect explaining this difference is linked with the proximity of the CI relative to either of the planar critical points. As a matter of fact, the topology of the energy curve along $x_2$ inducing a nonplanarity would influence their geometries only if they are close to the apex.

To determine this proximity, root-mean-square differences RMS($x$,$y$) of geometrical parameters between couples of critical points $x$ and $y$ have been calculated as:

$$\mathrm{RMS}(x,y) = [\sum_{i=1}^{N}(d_{i,x} - d_{i,y})^2/N]^{1/2} \qquad (3)$$

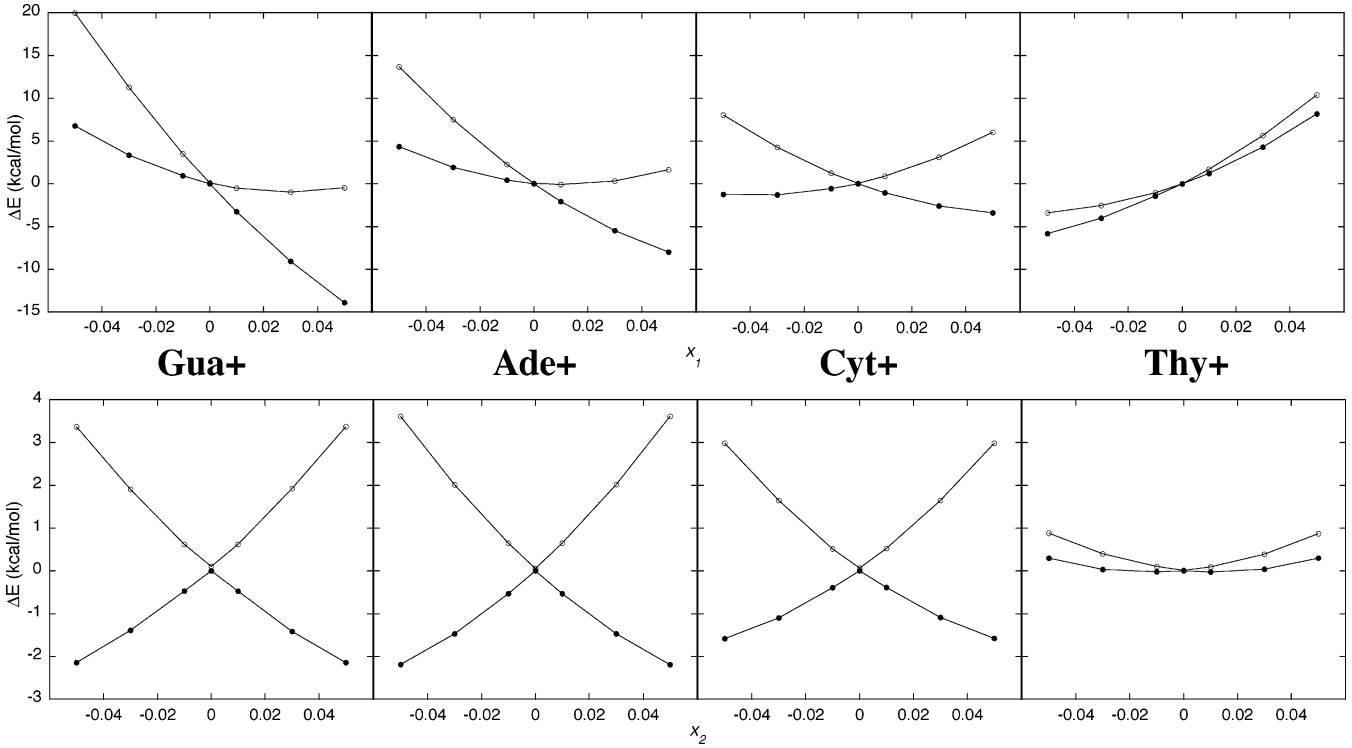


**Figure 3.** CASSCF(4,7)/6-31G(d,p) cuts in the two interacting adiabatic potential energy surfaces of the different cations as a function of ($x_1 - x_{1e}$) and ($x_2 - x_{2e}$), $x_{1e}$ and $x_{2e}$ referring to the geometry at the intersection point. The common apex of the double cone structure is the origin of the energy scale. This scale is different for the $x_1$ and $x_2$ variations.

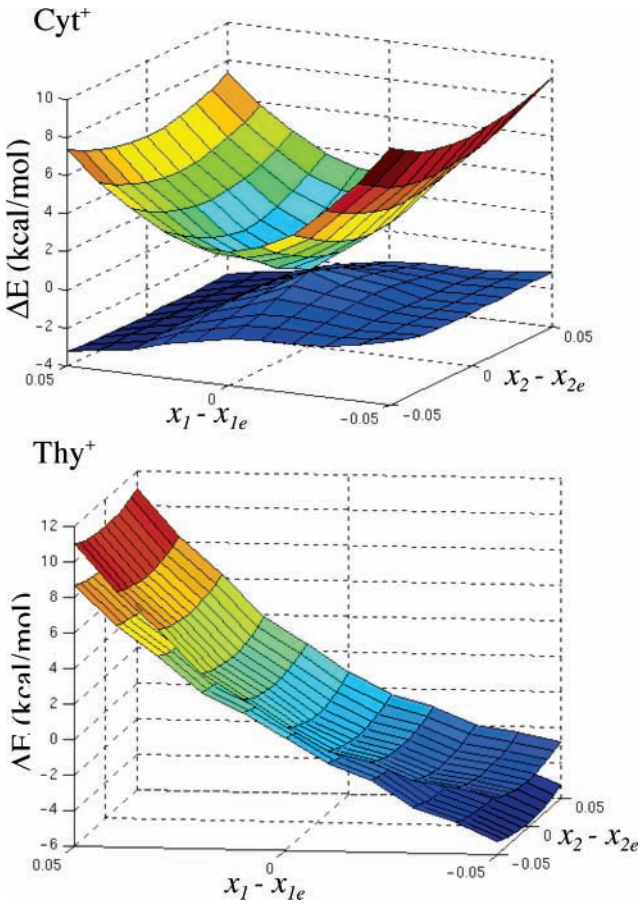


**Figure 4.** 3D representations of the conical intersections of $Cyt^+$ and $Thy^+$, calculated at CASSCF(4,7)/6-31G(d,p) level.

where $x$ and $y$ refer either to the planar $^2A''$ and $^2A'$ critical points or to the CI, and $d$ to the geometrical variables of the system, bond lengths, and angles being considered as separate variables. Also note that the sum over $N$ excludes variables involving hydrogen atoms.

**TABLE 5: Root-Mean-Square Differences of Geometrical Parameters (bond lengths (Å) and angles (deg)) between Couples of Critical Points ($^2A''$, $^2A'$, and CI) or between a Critical Point and the Mean Planar Geometry $\langle ^2A''-^2A'\rangle$**

| | RMS($x$,$\langle ^2A''-^2A'\rangle$), $x = ^2A''$ or $^2A'$ | | RMS(CI,$\langle ^2A''-^2A'\rangle$) | |
|---|---|---|---|---|
| | bond lengths | angles | bond lengths | angles |
| $Gua^+$ | 0.030 | 1.530 | 0.038 | 6.236 |
| $Ade^+$ | 0.020 | 2.488 | 0.026 | 3.678 |
| $Cyt^+$ | 0.030 | 3.636 | 0.034 | 2.654 |
| $Thy^+$ | 0.032 | 2.088 | 0.076 | 3.807 |

| | RMS(CI,$^2A''$) | | RMS(CI,$^2A'$) | |
|---|---|---|---|---|
| | bond lengths | angles | bond lengths | angles |
| $Gua^+$ | 0.0433 | 5.882 | 0.0524 | 6.919 |
| $Ade^+$ | 0.0435 | 5.271 | 0.0164 | 3.413 |
| $Cyt^+$ | 0.0630 | 5.869 | 0.0125 | 2.468 |
| $Thy^+$ | 0.0943 | 4.313 | 0.0691 | 4.372 |

RMS($x$,$\langle ^2A''-^2A'\rangle$) values were also calculated, to investigate the proximity of critical point $x$ to the mean geometry between the planar $^2A''$ and $^2A'$ points. This quantity is obtained by replacing in (3) $d_{i,y}$ by the mean value of $d_i$ between the planar points.

Table 5 presents the calculated RMS values for the bond lengths and the angles.

From the RMS($x$,$\langle ^2A''-^2A'\rangle$) values, it results that the CI lies outside the interval constituted by the planar critical points except in the case of $Cyt^+$. This confirms the conclusion drawn from the topology of the CI. From the RMS(CI,$^2A''$ or $^2A'$) values, again $Cyt^+$ shows a close proximity of the $^2A'$ state to the CI apex. Except for $Gua^+$, this state is closer to the CI than the $^2A''$. Furthermore, by comparing the values for $Ade^+$ and $Cyt^+$, one can imagine that the $Ade^+$ $^2A'$ state could be on the edge of being a saddle point.

**TABLE 6: Relative Energies (in eV) of the M2 and/or S1 Stationary Points with Respect to the Ground State Minimum M1**

| method/basis set | $Ade^+$ M2 | $Thy^+$ M2 | $Cyt^+$ S1 | $Cyt^+$ M2 | $Gua^+$ M2 |
|---|---|---|---|---|---|
| HF/6-31G(d,p) | 1.7 | 0.28 | 0.12 | −0.06 | 0.81 |
| HF/6-31G(2d(0.8,0.1),p)//HF/6-31G(d,p) | 1.73 | 0.32 | 0.21 | 0 | 0.82 |
| MP2/6-31G(d,p)//HF/6-31G(d,p) | 0.30 | 0.91 | 0.55 | 0.29 | 1.57 |
| MP2/6-31G(2d(0.8,0.1),p)//HF/6-31G(d,p) | 0.31 | 0.94 | 0.61 | 0.34 | 1.58 |
| MP2/6-31G(2d(0.8,0.1),p)//UMP2/6-31G(d,p) | - | - | 0.55 | 0.32 | - |

**4.4. Excitation Energies of the Radical Cations.** The excitation energies $E_{exc}$ are calculated as the energy differences between the S1 and M2 stationary points and the ground-state minimum M1. They have been calculated at different levels of theory to quantify both the basis set and correlation energy effects on the global energy shape of the lowest PES. The results are reported in Table 6. The MP2/6-31G(2d(0.8,0.1),p)//HF/6-31G(d,p) level, proved to be successful in the calculation of the IPs, is our better level of theory. It is compared to the corresponding results obtained with HF/6-31G(d,p), HF/6-31G-(2d(0.8,0.1),p)//HF/6-31G(d,p), and MP2/6-31G(d,p)//HF/6-31G(d,p). The first line of the table corresponds to the HF results, used in the previous section to determine the stationary point geometries. At this level, $Ade^+$ and $Gua^+$ have the largest $E_{exc}$ values (1.7 and 0.8 eV, respectively), while for $Thy^+$ and $Cyt^+$ they are smaller, even close to zero for the nonplanar minimum M2 of $Cyt^+$. In the latter case, the existence of two minima M1 and M2, close in energy, implies that a potential energy barrier separates them. We did not try to locate the corresponding transition state. Using the extended 6-31G(2d-(0.8,0.1),p) basis set does not significantly change the relative HF energies, the larger change being less than 0.1 eV. This small effect is not surprising, the 0.1 polarization being expected to play an equivalent role for the different stationary points. For instance for $Ade^+$, the MP2/6-31G(2d(0.8,0.1),p) $\langle R^2 \rangle$ values, calculated for M1 and M2 at equilibrium geometries, are 1121.94 and 1121.56 au, respectively. Let us note, however, that these values are smaller by 4 au than the one reported in Table 4. This difference corresponds to change in the electronic spatial extent accompanying the geometry relaxation within the cation. This example demonstrates again the sensitivity of $\langle R^2 \rangle$ and its interest in the analysis of ionization.

Correlation energy contributions are more important, with a different order magnitude for the four cations (from −1.4 to +0.8 eV). Electron correlation stabilizes more M1 than M2 and/or S1 in $Thy^+$, $Cyt^+$, and $Gua^+$. The $E_{exc}$ value is thus larger than that at the HF level. Oppositely, the excitation energy is severely reduced for $Ade^+$. Such differences between the four cations can be explained by the specific nature of the a′ and a″ (planar structures or nonplanar) molecular orbitals involved in the ionization of the different cations, as discussed in the next subsection from an analysis of the wave functions. The basis effect at MP2 level is very small, as already observed at the HF level, and for the same reasons.

For $Cyt^+$, we reoptimized the geometries of all stationary points (M1, M2, and S1) at the UMP2/6-31G(d,p) level and recalculated the restricted MP2/6-31G(2d(0.8,0.1),p) energies at the UMP2 equilibrium geometries. The results strictly confirm (within less than 0.1 eV) the MP2 results obtained from HF optimizations. We thus consider the MP2 stationary points energy scale as reliable. It predicts a local minimum M2 of low energy (0.3 eV) for $Ade^+$ and $Cyt^+$, and of higher energy (0.94 eV) for $Thy^+$. In the case of $Cyt^+$, the planar structure M1, predicted to be isoenergetic to M2 at the HF level, is actually the global minimum of the MP2 PES, confirming the DFT results of Improta et al.[21]

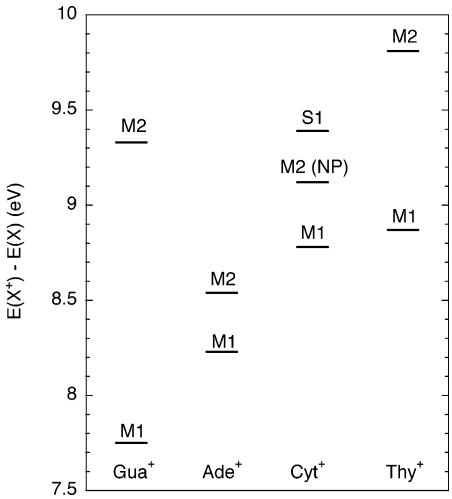


**Figure 5.** Relative energies of the stationary points of the nucleobase cations with respect to the ground-state energies of the corresponding neutral systems. All energy differences are calculated at MP2/6-31G-(2d(0.8,0.1),p)//HF/6-31G(d,p) level. The M1 values correspond to the adiabatic IPs reported in Table 2.

The case of guanine merits a further comment. In addition to its low IP, its cation is particularly stable with respect to electron excitation. Both of these features could contribute to the particular role played by this DNA base in the electron-transfer process in DNA.

Concerning the $IP_{vert}$ calculated for thymine by Sevilla et al.[19] (see Table 2), we confirm that the calculated energy of the ionic species corresponds well to the first excited state $^2A'$. To verify, we calculated the $IP_{vert}$ of thymine at the MP2/6-31G(2d-(0.8,0.1),p) level of theory leading to the cation A′. We obtained a value of 10.35 eV very close to the one of Sevilla et al.

Figure 5 summarizes the main features of the energy landscape characterizing the ionization of the DNA bases. This picture completes the qualitative overview of the topology of the lowest PES of the cations discussed in the previous section, and the quantitative determination of the ionization potentials. The bases are sorted by order of increasing IP values. All energies reported on this figure arise from our best level of calculation MP2/6-31G(2d(0.8,0.1),p)// HF/6-31G(d,p) and correspond to HF optimized geometries for all stationary points. Energies in the cationic species $X^+$ are calculated with respect to the energy at equilibrium geometry of the neutral system. The M1 values thus correspond to the adiabatic IPs reported in Table 2. We believe that the results collected in this figure provide a helpful reference for discussing any excitation process involving the DNA bases, in particular in the framework of the study of radiation damage, a hot topic in biophysical science.

## 5. Equilibrium Geometries and Electronic Structure of the Radical Cations

The equilibrium geometries of the stationary points characterized in section 4 are reported in Tables 1S−4S of the Supporting

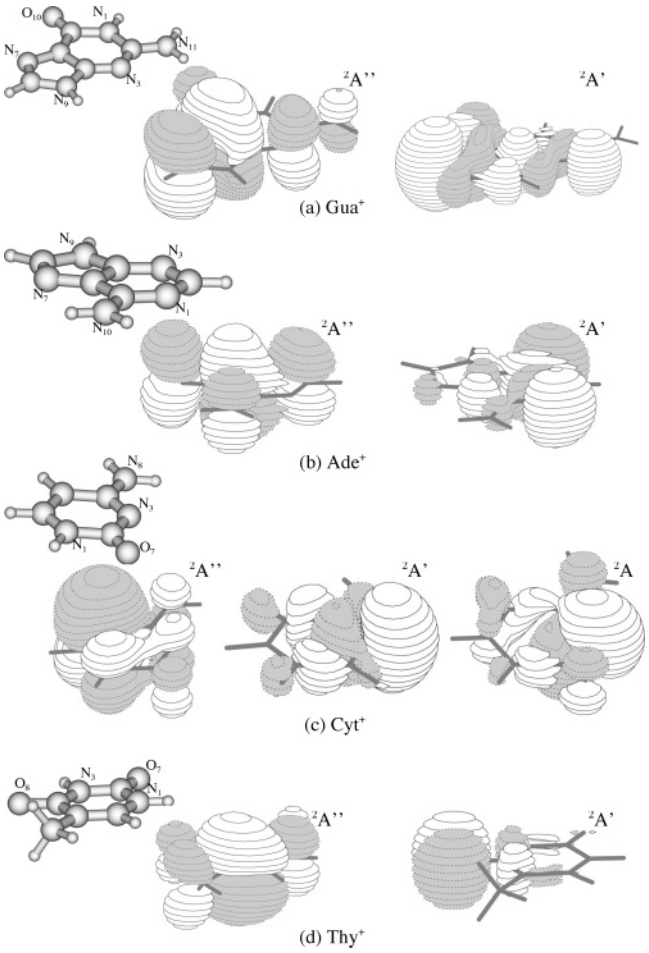


**Figure 6.** Orbital contour plot of the HOMO (HF/6-31G(d,p)) of the radical cations in their electronic states $^2A''$ (ground state), $^2A'$, and $^2A$ symmetries for (a) $Gua^+$ (7a″ and 32a′ MOs), (b) $Ade^+$ (6a″ and 29a′ MOs), (c) $Cyt^+$ (5a″and 24a′ MOs), and (d) $Thy^+$ (6a″ and 27a′ MOs).

Information for $Gua^+$, $Ade^+$, $Cyt^+$, and $Thy^+$ respectively. The corresponding Cartesian coordinates are also given in Tables 5S−8S. For sake of comparison, we have adopted for each cation the atom numbering (see the molecular frames in Figure 6) used by Improta et al.[21] The B1LYP/6-311G(d,p) equilibrium geometries calculated by the latter authors for the ground states of the cations (planar M1, $^2A''$ symmetry) are in close agreement with our HF results. The mean absolute deviation is globally of 0.018 Å for the internuclear distances and of 1° for the bond angles. A similar agreement is observed for the ground-state equilibrium geometries of the neutral bases, not reported in this work devoted to the cations. The results of Tables 1S−4S show that the geometrical changes accompanying the electronic excitation from $^2A''$ to $^2A'$ stationary points are not negligible (up to ±0.13 Å and ±13°). They essentially correspond to in-plane deformations of the aromatic cycles of the bases. The nonplanar $^2A$ structure of $Cyt^+$ is found to be more similar to the $^2A'$ than to the $^2A''$ ones. The more significant out-of-plane deformations concern the $O_7$, $N_3$, and $N_8$ atoms, as indicated by the dihedral angles given in Table 3S. Some bending angles implying these atoms also change accordingly with respect to the planar structures.

The electronic structure of all the stationary points can be first discussed in terms of the corresponding electronic configurations. Restricted for the sake of place to the $HOMO^{-3}$, $HOMO^{-2}$, $HOMO^{-1}$, and HOMO molecular orbitals (MOs), they can be written as:

$Gua^+$: 32a′ 5a″ 6a″ 7a″ ($^2A''$); 31a′ 6a″ 7a″ 32a′ ($^2A'$)

$Ade^+$: 4a″ 29a′ 5a″ 6a″ ($^2A''$); 28a′ 5a″ 6a″ 29a′ ($^2A'$)

$Cyt^+$: 23a′ 24a′ 4a″ 5a″ ($^2A''$); 23a′ 4a″ 5a″ 24a′ ($^2A'$)

$Thy^+$: 26a′ 27a′ 5a″ 6a″ ($^2A''$); 5a″ 26a′ 6a″ 27a′ ($^2A'$)

In these configurations all a′ and a″ MOs have a $\sigma$ and $\pi$ character, respectively. The two highest MOs of the ground state of all cations thus contribute to the $\pi$ structure of the aromatic planes. For all cations, except $Gua^+$, it is the $HOMO^{-2}$ which corresponds to the $\sigma$ HOMO of the $^2A'$ state. In the case of $Gua^+$, this MO is the $HOMO^{-3}$. The HOMOs of all stationary points are drawn in Figure 6.

The stability of the ground states of the cations, being of $^2A''$ symmetry, is thus related to the a″ symmetry of the HOMO. These MOs are indeed characterized by a $\pi$-bonding on the aromatic cycles: on $N_9-C_4-C_5-C_6$ for $Gua^+$, $N_9-C_4-C_5-C_6$ for $Ade^+$, $C_4-C_5-C_6$ for $Cyt^+$, and $C_4-C_5-C_6$ for $Thy^+$. Oppositely the a′ symmetry HOMOs are centered on an extra-cycle atom in $Thy^+$ ($O_8$) or mostly correspond to nitrogen lone pairs in the other cations ($N_7$ and $N_3$ in $Gua^+$, $N_3$ and $N_1$ in $Ade^+$, and $N_3$ in $Cyt^+$). One clearly sees that the HOMO of the nonplanar $^2A$ minimum of $Cyt^+$ results from a mixing of the $^2A''$ and $^2A'$ HOMOs, but with a larger amount of the latter. This explains the similarity pointed out above between the equilibrium geometries of the $^2A'$ and $^2A$ structures.

A last analysis concerns the atomic spin densities, which characterize the atoms involved in the ionization process. The spin densities (>0.1) are:

$Ade^+$ $^2A''$(M1): 0.36 on $C_5$ + 0.16 on $C_8$

$^2A'$(M2): 0.47 on $N_3$ + 0.43 on $N_1$

$Cyt^+$ $^2A''$(M1): 0.75 on $C_5$

$^2A'$(S1): 0.93 on $N_3$

$^2A$(M2): 0.46 on $C_5$ + 0.19 on $N_3$ + 0.14 on $C_6$

$Thy^+$ $^2A''$(M1): 0.65 on $C_5$ + 0.12 on $C_6$

$^2A'$(M2): 0.96 on $O_8$

$Gua^+$ $^2A''$(M1): 0.20 on $C_8$ + 0.40 on $C_5$ + 0.12 on $C_4$

$^2A'$(M2): 0.97 on $O_{10}$

These values show that the ionization occurs on lone pairs of heteroatoms for A′ species, while spin densities are more delocalized on the $\pi$ structure of the A″ species, as suggested in most cases by the MO pictures of Figure 6.

## 6. Conclusion

The ionization of the four DNA bases in the gas phase has been studied by means of restricted HF and MP2 calculations. The first excited states of the cations have been characterized for the first time and the topology of the corresponding low-lying potential energy surfaces has been studied. Conical intersections were characterized for all cations. The shape of the potential energy surfaces in the vicinity of the intersections and the magnitude of the nonadiabatic coupling have been used to explain the differences observed in the different cations. In

particular, the existence of a nonplanar minimum in the case of $Cyt^+$ only has been emphasized. These results bring a new insight into the electronic structure of the ionized systems related to DNA, and thus provide information that could be useful in the interpretation of reactive processes involving oxidation of DNA. The level of calculation to be used for a correct description of the ionization in such systems has been established on the basis of systematic test calculations and of comparisons with results from the literature. The medium size 6-31G(2d-(0.8, $\alpha_d$),p) basis set, with an optimized value of $\alpha_d = 0.1$, has been proposed for predicting accurate energy differences related to ionization (vertical and adiabatic IPs and electronic excitation energies of the cations). It presents the advantage of describing the electronic changes accompanying ionization, as well as larger polarized and/or augmented basis sets of the literature, but at the lowest computer costs. Such basis sets are thus potentially interesting for investigating larger systems involving the DNA bases, like clusters of these bases. This opportunity has been illustrated on a single system, a stacked dimer of guanines, for which the IP has been calculated. Another feature of the proposed basis set is to also improve the accuracy of calculated interaction energies (H-bonding, cation$-\pi$ and stacking interactions) in biomolecular complexes, as demonstrated in a previous work.[36] Work on ionized DNA base clusters is in progress in our group.

**Acknowledgment.** We thank Professor D. Peeters and Drs. G. Dive, R. Wintjens, and W. Eisfeld for useful comments, and G. Destrée for his help in computational problems. The Communauté Française de Belgique (ARC contract), the COST P9 Action, and the Belgian National Fund for Scientific Research (F.R.F.C. contract) are acknowledged for support. E.C. is an ARC researcher.

**Supporting Information Available:** The equilibrium geometries of the stationary points characterized in section 4 are reported in Tables 1S to 4S for $Gua^+$, $Ade^+$, $Cyt^+$ and $Thy^+$, respectively, and the corresponding Cartesian coordinates in Tables 5S to 8S. This material is available free of charge via the Internet at http://pubs.acs.org.

## References and Notes

(1) Cadet, J.; Berger, M.; Douki, T.; Ravanat, J. L. *Rev. Physiol. Biochem. Pharmacol.* **1997**, *131*, 1.
(2) Cadet, J.; Delatour, T.; Douki, T.; Gasparutto, D.; Pouget, J. P.; Ravanat, J. L.; Sauvaigo, S. *Mutat. Res.* **1999**, *424*, 9.
(3) Cadet, J.; Douki, T.; Gasparutto, D.; Ravanat, J. L. *Mutat. Res.* **2003**, *531*, 5.
(4) Hall, D. B.; Holmlin, R. E.; Barton, J. K. *Nature* **1996**, *382*, 731.
(5) Nakatani, K.; Dohno, C.; Saito, I. *J. Am. Chem. Soc.* **1999**, *121*, 10854.
(6) Giese, B. *Acc. Chem. Res.* **2000**, *33*, 631.
(7) Schuster, G. B. *Acc. Chem. Res.* **2000**, *33*, 253.
(8) Nuñez, M. E.; Barton, J. K. *Curr. Opin. Chem. Biol.* **2000**, *4*, 199.
(9) Meggers, E.; Michel-Beyerle, M. E.; Giese, B. *J. Am. Chem. Soc.* **1998**, *120*, 12950.
(10) Ye, Y. J.; Jiang, Y. *Int. J. Quantum Chem.* **2000**, *78*, 112.
(11) Grozema, F. C.; Berlin, Y. A.; Siebbeles, L. D. A. *J. Am. Chem. Soc.* **2000**, *122*, 10903.
(12) Troisi, A.; Orlandi, G. *Chem. Phys. Lett.* **2001**, *344*, 509.
(13) Berlin, Y. A.; Burin, A. L.; Ratner, M. A. *Chem. Phys.* **2002**, *275*, 61.
(14) Perun, S.; Sobolewski, A. L.; Domcke, W. *Chem. Phys.* **2005**, *313*, 107.
(15) Nguyen, M. T.; Zhang, R.; Nam, P.-C.; Ceulemans, A. *J. Phys. Chem. A* **2004**, *108*, 6554.
(16) Schultz, T.; Samoylova, E.; Radloff, W.; Hertel, I. V.; Sobolewski, A. L.; Domcke, W. *Science* **2004**, *306*, 1765.
(17) Merchan, M.; Serrano-Andres, L. *J. Am. Chem. Soc.* **2003**, *125*, 8108.
(18) Sobolewski, A. L.; Domcke, W. *Eur. Phys. J. D* **2002**, *20*, 369.
(19) Sevilla, M. D.; Besler, B.; Colson, A. O. *J. Phys. Chem.* **1995**, *99*, 1060.
(20) Crespo-Hernàndez, C. E.; Arce, R.; Ishikawa, Y.; Gorb, L.; Leszczynski, J.; Close, D. M. *J. Phys. Chem. A* **2004**, *108*, 6373.
(21) Improta, R.; Scalmani, G.; Barone, V. *Int. J. Mass Spectrom.* **2000**, *201*, 321.
(22) Close, D. M. *J. Phys. Chem. A* **2004**, *108*, 10376.
(23) Wetmore, S. D.; Boyd, R. J.; Eriksson, L. A. *Chem. Phys. Lett.* **2000**, *322*, 129.
(24) de Vries, J.; Hoekstra, R.; Morgenstern, R.; Schlathölter, T. *Eur. Phys. J. D* **2003**, *24*, 161.
(25) de Vries, J.; Hoekstra, R.; Morgenstern, R.; Schlathölter, T. *Phys. Rev. Lett.* **2003**, *91*, 053401.
(26) Sugiyama, H.; Saito, I. *J. Am. Chem. Soc.* **1996**, *118*, 7063.
(27) Prat, F.; Houk, K. N.; Foote, C. S. *J. Am. Chem. Soc.* **1998**, *120*, 845.
(28) Schumm, S.; Prévost, M.; Garcia-Fresnadillo, D.; Lentzen, O.; Moucheron, C.; Kirsch-De Mesmaeker, A. *J. Phys. Chem. B* **2002**, *106*, 2763.
(29) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
(30) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, *153*, 503.
(31) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(32) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery J. A., Jr.; 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.; Baboul, A. G.; 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.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, Revision A. 7; Gaussian Inc.: Pittsburgh, PA, 1998.
(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
(35) Šponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161.
(36) Wintjens, R.; Biot, C.; Rooman, M.; Liévin, J. *J. Phys. Chem. A* **2003**, *107*, 6249.
(37) Cauët, E.; Rooman, M.; Wintjens, R.; Liévin, J.; Biot, C. *J. Chem. Theory Comput.* **2005**, *1*, 472.
(38) Ditchfield, R.; Hehre, W.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724.
(39) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.
(40) Bauernschmitt, R.; Ahlrichs, R. *J. Chem. Phys.* **1996**, *104*, 9047.
(41) Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1977**, *66*.
(42) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247.
(43) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 104.
(44) Mishra, B. K.; Sathyamurthy, N. *J. Phys. Chem. A* **2005**, *109*, 6.
(45) Xu, X.; Goddard, W. A., III *Proc. Natl. Acad. Sci.* **2004**, *101*, 2673.
(46) Černy, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1624.
(47) Bertran, J.; Oliva, A.; Rodriguez-Santiago, L.; Sodupe, M. *J. Am. Chem. Soc.* **1998**, *120*, 8159.
(48) Hegarty, D.; Robb, M. A. *Mol. Phys.* **1979**, *38*, 1795.
(49) Eade, R. H. E.; Robb, M. A. *Chem. Phys. Lett.* **1981**, *83*, 362.
(50) Siegbahn, E. M. *Chem. Phys. Lett.* **1984**, *109*, 417.
(51) Robb, M. A.; Niazi, U. *Rep. Mol. Theory* **1990**, *1*, 23.
(52) Bearpark, M. J.; Robb, M. A.; Schlegel, H. B. *Chem. Phys. Lett.* **1994**, *223*, 269.
(53) Ragazos, I. N.; Robb, M. A.; Bernardi, F.; Olivucci, M. *Chem. Phys. Lett.* **1992**, *197*, 217.
(54) Bernardi, F.; Robb, M. A.; Olivucci, M. *Chem. Soc. Rev.* **1996**, *25*, 321.
(55) Woon, D. E.; Dunning, T. H. J. *J. Chem. Phys.* **1995**, *98*, 4572.

(56) Dunning, T. H. J. *J. Chem. Phys.* **1989**, *90*, 1007.
(57) Kendall, R. A.; Dunning, T. H. J.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.
(58) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294.
(59) Mignon, P.; Loverix, S.; De Proft, F.; Geerlings, P. *J. Phys. Chem. A* **2004**, *108*, 6038.
(60) Hobza, P.; Šponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802.
(61) Hush, N. S.; Cheung, A. S. *Chem. Phys. Lett.* **1975**, *34*, 11.
(62) Orlov, V. M.; Smirnov, A. N.; Varshavsky, Y. M. *Tetrahedron Lett.* **1976**, *48*, 4377.
(63) Close, D. M. *J. Phys. Chem. B* **2003**, *107*, 864.
(64) Eisfeld, W.; Morokuma, K. *J. Chem. Phys.* **2001**, *114*, 9430.
(65) Rooman, M.; Liévin, J.; Buisine, E.; Wintjens, R. *J. Mol. Biol.* **2002**, *319*, 67.
(66) Ragazos, I. N.; Robb, M. A.; Bernardi, F.; Olivucci, M. *Chem. Phys. Lett.* **1992**, *197*, 217.
(67) Bearpark, M. J.; Robb, M. A.; Schlegel, H. B. *Chem. Phys. Lett.* **1994**, *223*, 269.
(68) Crespo-Hernàndez, C. E.; Cohen, B.; Hare, P. M.; Kohler, B. *Chem. Rev.* **2004**, *104*, 1977.