Charge Localization in Stacked Radical Cation DNA Base Pairs and the Benzene Dimer Studied by Self-Interaction Corrected Density-Functional Theory

Yves A. Mantz,*,† Francesco Luigi Gervasio,† Teodoro Laino,‡,§ and Michele Parrinello†

Computational Science, Department of Chemistry and Applied Biosciences, ETH Zurich, USI Campus, Via Giuseppe Buffi 13, CH-6900 Lugano, Switzerland, and Scuola Normale Superiore di Pisa, Piazza dei Cavalieri 7, I-56125 Pisa, Italy

Received: May 19, 2006; In Final Form: October 10, 2006

The incomplete cancellation of the electron self-interaction can be a serious shortcoming of density-functional theory especially when treating odd-electron systems. In this work, several popular and potentially viable correction schemes are applied in order to characterize the electronic structure of stacked molecular pairs, consisting of a neutral molecule and adjacent radical cation, as a function of separation distance. The unphysical sharing of the positive charge between adjacent molecules separated by 6−7 Å is corrected for by applying a new empirical scheme proposed by VandeVondele and Sprik [*Phys. Chem. Chem. Phys.* **2005**, *7*, 1363] with a unique choice of parameters. This method is subsequently applied to characterize the electronic structure of two neighboring guanines excised from a canonical Arnott B-DNA structure and will be used in future investigations of certain model DNA fibers.

1. Introduction

Systems with an odd number of electrons are of great interest in the chemical sciences and merit intense theoretical examination. Yet, density-functional theory (DFT), one of the most powerful techniques for modeling electronic structure changes in systems containing ∼100 atoms, must be applied with great care. The approximate description of the electron exchange (X) and correlation (C) in a DFT calculation is unable to correct for the spurious self-interaction of the electron included in the Hartree (electron−electron) repulsion energy, resulting in the well-known self-interaction error (SIE) of DFT.[1] Importantly, the SIE is not necessarily bad: It can successfully emulate unspecified short- and long-range electron correlation effects that are missing from the approximate XC functional,[2,3] thereby improving the accuracy of the DFT description. As a consequence, uncorrected DFT is very successful in describing certain molecular properties including standard covalent bonds and interacting closed shells.[2,4] However, the SIE is also responsible for some dramatic failures of DFT,[5] including (the modeling of) certain atomic properties,[1] the dissociation of odd-electron systems,[2,3] the underestimation of reaction barriers, and charge transfer processes.[6,7] Prototypical problems that are difficult to solve are the geometry of the quartz Al center[8,9] and the solvent shell structure of the OH radical in liquid water.[10,11] Due to the ambivalent role played by the SIE, a great deal of effort[3,12,13] is being devoted to design new XC functionals including both exact or Hartree−Fock exchange, thereby eliminating the SIE, as well as electron correlation effects mimicked by the SIE, potentially leading to a general way to deal with all types of systems. However, in certain applications, such as the dissociation of radical cation dimers including H_2^+ or polyatomic radical cations where the SIE is detrimental,[2,3] it can be more practical either to employ hybrid functionals[5] or simply to correct standard XC functionals by applying a self-interaction correction (SIC).[1,14,15]

In this work, the DFT description of the electronic structure of overlapping molecular pairs that have an unpaired electron—resembling a (kind of) polyatomic radical cation—is shown to be remarkably improved by the application of a new empirical SIC scheme proposed by VandeVondele and Sprik.[15] In particular, this method is applied to examine a biologically relevant system, a positively charged guanine dimer that possesses the same geometry as two adjacent guanine nucleotides in a DNA helix. This sequence is very important due to its relatively low ionization potential and, thus, pivotal role played in DNA charge transfer, which is being studied in order to assess the feasibility of DNA-based electronic devices as well as to understand repair mechanisms within damaged DNA.[16,17] Upon application of the SIC, a physically reasonable localization of the charge is observed that is consistent with other SIE-free calculations. A similar result is obtained for a pair of stacked adenine bases, which are more easily ionized than adjacent thymine or cytosine nucleotides, thus providing a key step in improving the description of the hole wavefunction in DNA. Accordingly, the stage is set for future investigations of certain oxidized DNA fibers to complement previous studies by our laboratory of the guanine:cytosine dodecamer.[18,19]

The organization of this article is as follows: In section 2, the models and methods employed are introduced. In sections 3and 4, the electronic structures of a stacked benzene radical cation and benzene molecule as well as a guanine ion and molecule in eclipsed geometries are examined by performing reference calculations using computational methods that are not suffering from the SIE (such as Hartree−Fock and post-Hartree−Fock techniques) and comparing the results obtained to those from uncorrected DFT as well as from DFT calculations employing different SIC schemes. Subsequently, in section 5, the most appropriate SIC is identified and applied to investigate a guanine dimer excised from an Arnott B-DNA. In section 6,

---

* Corresponding author. E-mail: ymantz@phys.chem.ethz.ch.

† ETH Zurich, USI Campus.

‡ Scuola Normale Superiore di Pisa.

§ Present address: Physical Chemistry Institute, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.

10.1021/jp063080n CCC: $37.00

this method is tested further by calculating ionization potentials of isolated benzene and the DNA bases as well as probing the electronic structure of the other nucleotide stacks. In addition, the physical nature of the SIE in all of the systems studied is briefly discussed. Conclusions are drawn in section 7.

2. Computational Methods

Following the geometry optimization of the neutral molecule and radical cation, selected molecular pairs were placed in eclipsed (sandwich) conformations in order to maximize the intermolecular interactions. Using the isolated (neutral and radical cation) molecular geometries calculated at the appropriate levels of theory, the electronic structure as a function of the separation between molecular planes was calculated at the following levels of theory using the Gaussian 03 software package:[20] restricted open-shell Hartree−Fock (ROHF), DFT-ROBLYP and DFT-ROB3LYP, unrestricted MP2 (UMP2), and complete active space multiconfiguration SCF, termed CASSCF-(X,Y), involving X electrons within a subset, Y, of the ROHF orbitals, as explained in refs 21 and 22. By restricting the orbitals, mixing with higher spin states (spin contamination) is avoided.[23]

Additionally, DFT-ROBLYP calculations with the SIE and with several different SICs were carried out as implemented in the Quickstep module[24] of the CP2K package. In Quickstep, a SIC can be applied by adding to the (restricted open-shell) Kohn−Sham energy,

$$E_{\mathrm{KS}} = \frac{1}{2}\sum_{i,\sigma}\langle\psi_{i\sigma}|\nabla^2|\psi_{i\sigma}\rangle + \int \mathrm{d}\mathbf{r}\,\rho(\mathbf{r})\,\upsilon(\mathbf{r}) + \frac{1}{2}\int \mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}'\frac{\rho(\mathbf{r})\,\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{\mathrm{XC}}[\rho_\alpha(\mathbf{r}),\rho_\beta(\mathbf{r})]$$

$$E_{\mathrm{KS}} = T[\{\psi_{i\sigma}\}] + E_{\mathrm{EXT}}[\rho(\mathbf{r})] + E_{\mathrm{H}}[\rho(\mathbf{r})] + E_{\mathrm{XC}}[\rho_\alpha(\mathbf{r}),\rho_\beta(\mathbf{r})] \quad (1)$$

a correction term, E_{SIC}, according to three schemes,

$$E_{\mathrm{SIC}}^{\mathrm{PZ}} = -E_{\mathrm{H}}[m(\mathbf{r})] - E_{\mathrm{XC}}[m(\mathbf{r}),0] \quad (2)$$

$$E_{\mathrm{SIC}}^{\mathrm{M}} = -E_{\mathrm{H}}[m(\mathbf{r})] - (E_{\mathrm{XC}}[\rho_\alpha(\mathbf{r}),\rho_\beta(\mathbf{r})] - E_{\mathrm{XC}}[\rho_\alpha(\mathbf{r}) - m(\mathbf{r}),\rho_\beta(\mathbf{r})]) \quad (3)$$

$$E_{\mathrm{SIC}}^{\mathrm{SS}} = -aE_{\mathrm{H}}[m(\mathbf{r})] - bE_{\mathrm{XC}}[m(\mathbf{r}),0] \quad (4)$$

where the electronic spin density $m(\mathbf{r}) = \rho_\alpha(\mathbf{r}) - \rho_\beta(\mathbf{r})$, while $\rho_\alpha(\mathbf{r}) = \sum_i^{N+1}|\psi_{i\alpha}(\mathbf{r})|^2$ is the α electron density and $\rho_\beta(\mathbf{r}) = \sum_i^{N}|\psi_{i\beta}(\mathbf{r})|^2$ is the β electron density of the $2N+1$ electron system. Also, $\rho(\mathbf{r}) = \rho_\alpha(\mathbf{r}) + \rho_\beta(\mathbf{r})$. Note that the radical or singly occupied molecular orbital (SOMO) is closely related to $m(\mathbf{r})$ in these (restricted) calculations, because $\psi_{i\alpha}(\mathbf{r}) = \psi_{i\beta}(\mathbf{r})$.

All three SIC schemes in eqs 2−4 are a simplification of that proposed by Perdew and Zunger, who started with the fact that the functional, $E_{\mathrm{H}}[\rho(\mathbf{r})]$, contains a self-interaction term for every electron and proved[1] that the exact exchange-correlation functional, $E_{\mathrm{XC}}[\rho_\alpha(\mathbf{r}),\rho_\beta(\mathbf{r})]$ would cancel these terms exactly but is unknown. Accordingly, to correct for the practical use of any approximate $E_{\mathrm{XC}}[\rho_\alpha(\mathbf{r}),\rho_\beta(\mathbf{r})]$, the electron self-repulsion, and its incomplete cancellation, can be subtracted out of E_{KS} on an orbital-by-orbital basis,[1]

$$E_{\mathrm{SIC}} = \sum_{i,\sigma} - E_{\mathrm{H}}[\rho_{i\sigma}(\mathbf{r})] - E_{\mathrm{XC}}[\rho_{i\sigma}(\mathbf{r}),0] \quad (5)$$

where $\rho_{i\sigma} = |\psi_{i\sigma}(\mathbf{r})|^2$, and $\sigma = \alpha$ or β. However, this correction is both expensive and dependent on the nature of the $\psi_{i\alpha}(\mathbf{r})$'s, i.e., its magnitude is affected by a unitary transformation of the $\psi_{i\alpha}(\mathbf{r})$'s, which is problematic[14] when attempting to minimize the resultant energy, $E_{\mathrm{KS}} + E_{\mathrm{SIC}}$. Motivated by this fact and the observation that the most relevant self-interaction is typically that of the unpaired electron, the refined correction, $E_{\mathrm{SIC}}^{\mathrm{PZ}}$ in eq 2, inspired by Perdew and Zunger,[1] was proposed by Mauri and co-workers.[14] The $E_{\mathrm{SIC}}^{\mathrm{PZ}}$ is a functional *only* of $m(\mathbf{r})$, a physical observable, and subtracts the self-interaction term associated with the unpaired electron from both $E_{\mathrm{H}}[\rho(\mathbf{r})]$ and $E_{\mathrm{XC}}[\rho_\alpha(\mathbf{r}),\rho_\beta(\mathbf{r})]$. A different SIC, $E_{\mathrm{SIC}}^{\mathrm{M}}$ in eq 3, also proposed by Mauri and co-workers,[14] subtracts the unpaired electron self-interaction from $E_{\mathrm{H}}[\rho(\mathbf{r})]$ and replaces $E_{\mathrm{XC}}[\rho_\alpha(\mathbf{r}),\rho_\beta(\mathbf{r})]$ for the $2N+1$ electron system with that for the N electron system without the unpaired electron. These corrections are implemented using restricted orbitals in order to avoid an unphysical solution to $m(\mathbf{r})$.[14] Last, $E_{\mathrm{SIC}}^{\mathrm{SS}}$ in eq 4 is an empirical "scaled SIC" including the adjustable parameters a and b introduced by VandeVondele and Sprik,[15] allowing great flexibility in treating systems that are not well described by the other methods.

A systematic procedure was used to identify the most appropriate SIC scheme for the systems investigated in this work. Initially, $m(\mathbf{r})$ was calculated for the uncorrected DFT-BLYP geometries (with the SIE) of the neutral molecule and radical cation employing either the $E_{\mathrm{SIC}}^{\mathrm{PZ}}$ or $E_{\mathrm{SIC}}^{\mathrm{M}}$ corrections as well as the $E_{\mathrm{SIC}}^{\mathrm{SS}}$ scheme. For the latter, the parameter space was initially scanned by performing 64 single-point calculations[25] and adjusting the parameters a and b from 0 to 1.4 in increments of 0.2. For promising values of a and b [i.e., those yielding $m(\mathbf{r})$'s consistent with benchmark calculations as shown in the figures] as well as other selected values, the geometries of the neutral molecule and radical cation were optimized using the empirical SIC, and binding energies were calculated. Results were tested for convergence with respect to basis set size; the calculated $m(\mathbf{r})$'s shown are essentially converged, as are the binding energies obtained for the guanine system excised from AB-DNA, which are presented for three basis sets of increasing quality.

Additional technical details of this work are also important. The Gaussian calculations were performed with a 6-31G* basis set and checked for convergence with respect to both basis set size and the density matrix convergence criterion. Notably, converged Gaussian DFT results were obtained only after increasing this criterion from 10^{-3} ("normal") to 10^{-6} ("tight"), requiring a different SCF procedure—CDIIS with dynamic damping, instead of the default method—to facilitate convergence. In the Quickstep calculations, the valence electrons were described both with a double-ζ valence basis set with polarization functions, DZVP,[24] and an auxiliary plane-wave basis set with a density cutoff of 350 Ry, while the core electrons and nuclei were treated with Goedecker−Teter−Hutter pseudopotentials.[26,27] Energies were tested for convergence with respect to the wavefunction gradient (standard 10^{-6} criterion) and cell size, which was required to be 20 Å in a periodic image decoupling scheme.[28] As a consistency check, results obtained using Quickstep with the choice of parameters $a = 0.0$ and $b = 0.0$ were observed to be very similar to the Gaussian ROBLYP results; small discrepancies are due to the slightly different basis sets and use of pseudopotentials.

3. Results: Neutral Benzene and Benzene Radical Cation

In this section, several different SICs are applied to the DFT description of a polyatomic radical cation dimer consisting of

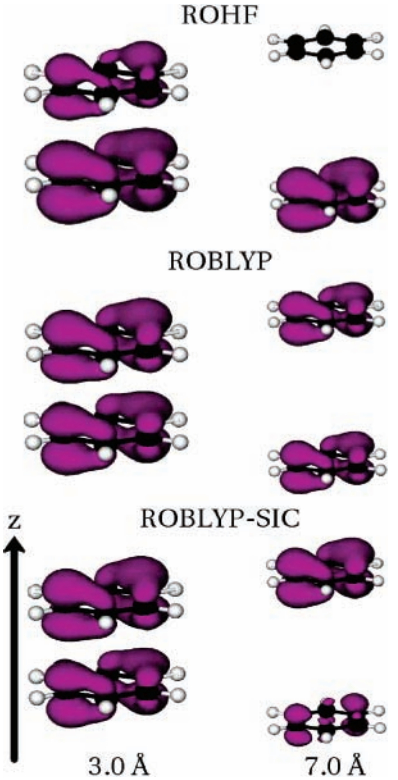

Figure 1. Isosurface contours of $m(\mathbf{r})$ (=0.002) for separations of either 3.0 or 7.0 Å between a benzene radical cation and molecule calculated at different levels of theory. The "ROBLYP-SIC" results were obtained by setting $a = 0.8$ and $b = 0.5$. In all figures, the radical cation possesses the smaller z-coordinate.

a stacked benzene molecule, the prototypical conjugated molecule, and its radical cation. The electronic structure as a function of separation distance is compared to that predicted at the HF level of theory, which is sufficient to yield a qualitatively accurate picture of the SOMO, although binding energies are not quantitatively correct due to the absence of electron correlation. For this system, the most appropriate SIC is slightly different than that for stacked radical cation DNA base pairs (sections 4 and 5), underscoring the sensitivity of the results to the magnitude of the correction and the limitations of the empirical SIC approach, i.e., the fact that a single correction is not appropriate for all stacked π systems.

The electronic structure of a stacked benzene molecule and ion is calculated at the ROHF level of theory in order to have results without the SIE. Calculations were performed by placing a neutral benzene, D_{6h} symmetry, above the benzene radical cation, D_{2h} symmetry,[29] and varying the separation distance. At a separation of 3 Å, the molecule with the isolated cation geometry is slightly favored by the unpaired electron, based on the size of the isosurface lobes of m (Figure 1, top row). This is consistent with physical intuition. As the aromatic rings are pulled apart to 7 Å, the unpaired electron is completely localized on this species, which is the expected outcome (Figure 1, top row). Furthermore, the radical "orbital" possesses a single nodal plane and resembles one of the degenerate HOMOs of an isolated benzene molecule. Thus, a correct *qualitative* description of the unpaired electron is obtained at this level of theory.

In marked contrast, at the ROBLYP level of theory, this α-electron is unrealistically shared between both aromatic rings at a separation of 7 Å and, in fact, somewhat favors the molecule with the isolated *neutral* geometry (Figure 1, middle row). A graphical depiction is obtained by comparing the projections of m along the z-axis for the radical ion and neutral molecule (Figure 2). Note that the projection obtained at a separation of 3 Å at the ROHF level of theory is generally consistent with that at the UMP2 level, although the preference of the electron for either the ion or neutral molecule is unclear, at least partly due to spin contamination affecting the UMP2 results. Also, the m profiles obtained using the BLYP functional are not improved significantly by employing the B3LYP functional (not shown), with a refined exchange term including some exact exchange.

On the other hand, the DFT electronic structure description is significantly improved by the application of the "scaled SIC" method with the choice of empirical parameters $a = 0.8$ and $b = 0.5$ (section 2), labeled "ROBLYP-SIC" in Figure 1, bottom row. In fact, the unpaired electron is now much more localized on one of the molecular species at a separation of 7 Å, while the already satisfactory "ROBLYP" results at 3 Å are, more or less, unaffected. (Apparently, the correction is unable to overcome the tendency of DFT-ROBLYP to favor the isolated neutral geometry.) By integrating the z-projections of m, the fractional electron populations on the preferred species (with the benzene cation geometry for ROHF but the benzene neutral geometry for ROBLYP-SIC) are obtained: for ROHF, they are either 0.72 (3 Å) or 1.00 (7 Å), while for $a = 0.8$ and $b = 0.5$, 0.58 (3 Å) and 0.77 (7 Å). However, this desirable outcome is sensitive to the values of a and b: A complete localization of the unpaired electron at a separation of 7 Å can be achieved with the values $a = 0.8$ and $b = 0.4$ (Figure 2). Also, neither the $E_{\mathrm{SIC}}^{\mathrm{M}}$ method nor the parameter choice, $a = 0.2$ and $b = 0$, of ref 15 is able to capture the correct behavior, yielding either a highly localized electron (at small gaps) or delocalized electron (at large gaps), respectively. The $E_{\mathrm{SIC}}^{\mathrm{PZ}}$ method (equivalent to $a = 1.0$ and $b = 1.0$) is, also, predicting a more delocalized electron relative to other (a,b) values. For certain other choices of a and b, the isosurfaces obtained are unphysical, e.g., they are not symmetric and confined to three atoms (or one half) of the neutral molecule. In these cases, converged results are difficult to obtain and are also highly sensitive to the initial wavefunction guess, indicating that the method is not able to locate the global minimum. These problems, and the fact that the optimal correction is slightly different for the benzene system, might be explained by the observation that the geometries of the molecular components are quite similar. In a diatomic radical cation such as H_2^+, a continuum of ground states ranging from covalent to ionic is predicted.[5] Thus, "correcting" the approximate DFT-BLYP functional might eliminate useful nondynamic correlation effects that are needed to describe the ground state of the system. Incidentally, the projections of m are two-dimensional and a less sensitive measure than the isosurfaces of m themselves, and consequently a "window" of parameters is obtained describing reasonably well the unpaired electron at both small and large separations (Figure 2).

As a quantitative assessment of the empirical SIC scheme, selected binding energies as a function of separation distance between the benzene cation and molecule are given in Table 1. At the uncorrected DFT-BLYP level, the charged pair is attractive, 11.0 kcal/mol, at a separation of 3 Å. The lack of steric repulsion combined with a slowly decaying ion−dipole electrostatic coupling yields a larger binding energy, 21.4 kcal/mol, at 7 Å. The empirical SIC scheme with the choice of parameters $a = 0.8$ and $b = 0.5$ affects these values significantly: It yields −5.4 and −5.9 kcal/mol, respectively, which is a consequence of the nature of the correction. Other values of the binding energy calculated for pertinent choices of a and

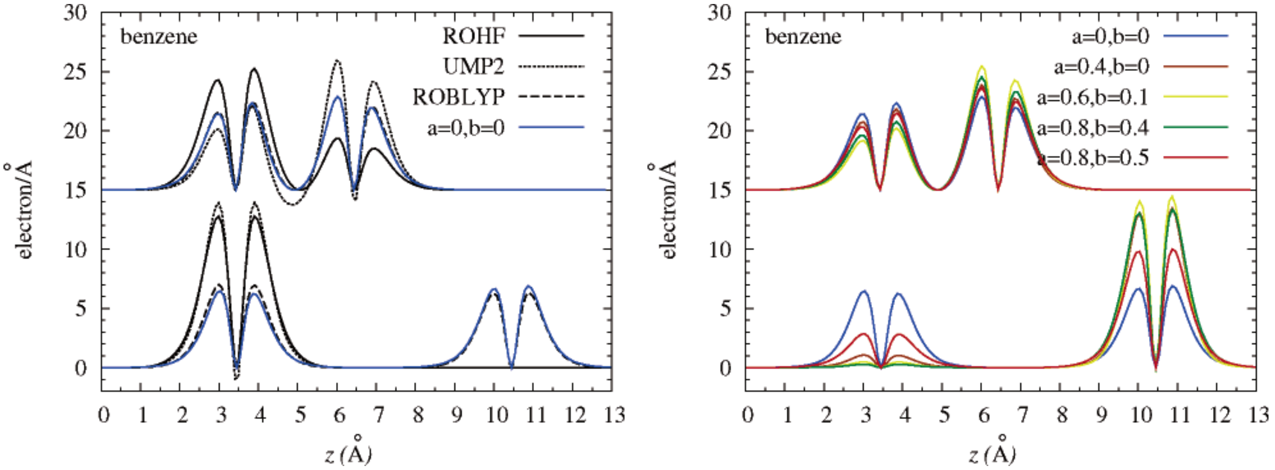

Figure 2. Projection of $m(\mathbf{r})$ for separations of 3.0 and 7.0 Å between a benzene cation (smaller z-coordinate) and molecule (0.08 Å bin).

b are also tabulated. Because correlation effects are crucial in describing dispersion interactions, a repulsion is predicted between *neutral* benzene molecules for methods such as HF and B3LYP using a 6-311++G(d,p) basis set,[30] whereas accurate potential energy profiles are given by MP2 calculations.[31]

4. Results: Neutral Guanine and Guanine Radical Cation

In order to test the empirical SIC scheme further, a very different molecule containing heteroatoms, and its radical cation, are examined. Guanine is the most easily oxidized nucleobase and is intimately involved in the mechanism of charge transport in oxidized DNA fibers. Because the conclusions drawn are particularly important in regard to future (biological) applications, first, results obtained at the ROHF level of theory are validated by comparison both to UMP2 and CASSCF results, after noting satisfactory agreement of the isolated neutral and radical cation geometries obtained to previous calculations.[32] Subsequently, both uncorrected and SIE-free DFT results are compared to these calculations.

For the separation distances tested, the agreement between the ROHF and CASSCF results is excellent (Figure 3, first and second rows). In particular, the isosurface of m obtained at the ROHF level of theory for a gap of 2.5 Å is very similar in appearance to the isosurface of the radical SOMO obtained by performing a CASSCF(7,8) calculation, initiated by interchanging molecular orbitals 82 and 83 or 82 and 84. This is not surprising, as the ground state is assigned a high weight of $0.9514^2 = 91\%$ in the CASSCF CI matrix, although significant stabilization is realized due to correlation effects, $E(\text{ROHF}) = -1078.3686601$ au versus $E(\text{CASSCF}) = -1078.4370199$ au. The active space is chosen in order to include the most relevant π orbitals of both the cation and neutral species, and the total energy is converged with respect to a unit increase of the active space [to either (7,9) or (9,8)], but a calculation including the

TABLE 1: Binding Energies (kcal/mol) for Different Charged Pairs at the ROBLYP Level of Theory

		SIC for selected values of a,b			
charged pair[a] (gap)	no SIC	0.2,0	0.8,0.4	0.8,0.5	1.0,1.0
guanine (AB-DNA)	7.6	4.3	3.7	3.6[b]	4.7
guanine (2.5 Å)	−90.7	−94.1	−109.8	−101.8	−98.0
adenine (2.5 Å)	−81.4	−84.3	−79.5	−91.5	−86.9
cytosine (2.5 Å)	−54.1	−78.5	−148.4	−141.9	−70.9
thymine (2.5 Å)	−126.2	−133.6	−145.2	−146.2	−143.8
benzene (3.0 Å)	11.0	5.1	−16.6	−5.4	−5.3
guanine (6.0 Å)	12.4	6.4	0.7	0.7	2.8
adenine (6.0 Å)	15.6	10.0	−4.2	2.7	6.1
cytosine (6.0 Å)	18.7	0.5	0.4	0.3	0.1
thymine (6.0 Å)	14.4	5.7	0.3	0.1	−0.1
benzene (7.0 Å)	21.4	12.6	−6.7	−5.9	3.4

[a] Not corrected for basis set superposition error. [b] Values using a TZV2P and a QZV3P basis set are 1.9 and 1.4 kcal/mol, respectively.

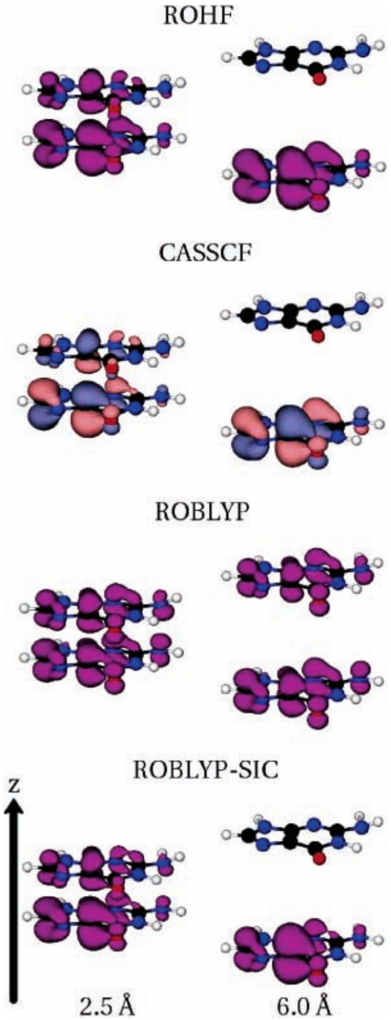

Figure 3. Isosurface contours of $m(\mathbf{r})$ (=0.002, purple lobes) or of the radical orbital ($\sqrt{0.002} = \pm 0.045$, where ice blue/pink orbitals have a positive/negative sign) for separations of either 2.5 or 6.0 Å between a guanine radical cation and molecule calculated at different levels of theory. The "ROBLYP-SIC" results were obtained by setting $a = 0.8$ and $b = 0.5$.

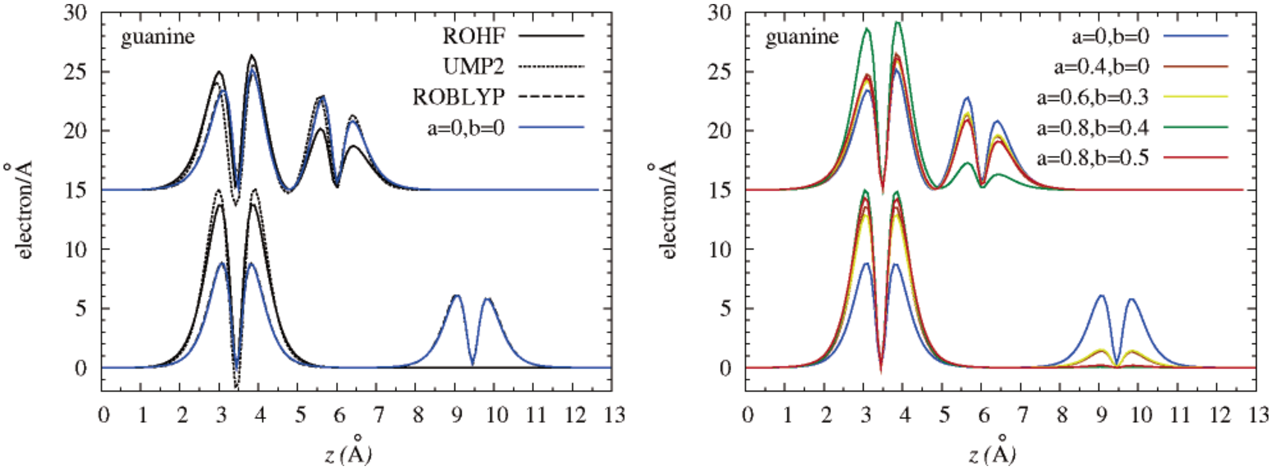

Figure 4. Projection of $m(\mathbf{r})$ for separations of 2.5 and 6.0 Å between a guanine cation and molecule (0.08 Å bin).

remaining eight π electrons would be prohibitive. Upon increasing the gap to 6 Å, the isosurface of m obtained at the ROHF level is nearly identical to the SOMO of an isolated guanine radical cation obtained at the CASSCF(7,5) level upon interchanging the ROHF orbitals 35 and 36 (Figure 3, first and second rows). Again, this result is expected, given a ground state weight of $0.9899^2 = 98\%$ in the CASSCF CI matrix; the energies are $E(\text{ROHF}) = -539.1481472$ au versus $E(\text{CASSCF}) = -539.1675688$ au. Last, the projections of m predicted at the ROHF level of theory are very similar to the UMP2 results (Figure 4). Thus, ROHF calculations are able to describe well the electronic structure of stacked guanines. Although dispersion effects originating in the electron correlation are very important in describing the minimum energy structures obtained with respect to the parameters (such as twist, slide, shift, and roll) defining the orientation of the *neutral* monomers,[33−36] they are less important here, where the predominant electrostatic (e.g., ion−dipole) contributions and $\pi-\pi$ interactions are captured at the ROHF level of theory.

A dramatic failure of DFT-ROBLYP, but spectacular recovery upon applying a SIC, is also observed for this stacked pair. In particular, the electron is highly delocalized at a separation of 6 Å as predicted at the ROBLYP level of theory, which is physically unreasonable (Figure 3, third row). However, upon applying the empirical SIC method with the choice $a = 0.8$ and $b = 0.5$, the electron is now localized exclusively on the molecule with the isolated cation geometry, as shown by the results labeled "ROBLYP-SIC" in Figure 3, fourth row, in excellent agreement with the reference calculations. By integrating the z-projections of m, the fractional electron populations on the guanine cation are, for ROHF, 0.71 (3 Å) and 1.00 (7 Å), while for $a = 0.8$ and $b = 0.5$, they are 0.68 (3 Å) and 0.99 (7 Å). This result is dependent on the choice of a and b, e.g., the structure predicted using $a = 0.8$ and $b = 0.4$ is too localized at a gap of 2.5 Å (Figure 4). The $E_{\text{SIC}}^{\text{PZ}}$ and $E_{\text{SIC}}^{\text{M}}$ schemes are suffering from the same problems as described above, while a "window" of parameter values is once again able to reproduce the m projections of the ROHF or UMP2 calculations, although the optimal values of a and b are somewhat different from those in Figure 2.

To continue the evaluation of the method, the binding energies for the stacked guanine cation and neutral molecule are given in Table 1. At the very short separation distance of 2.5 Å, the charged pair is repulsive, −90.7 kcal/mol, but it is attractive, 12.4 kcal/mol, for a gap of 6.0 Å at the uncorrected DFT-BLYP level. These values are significantly altered by the empirical SIC scheme with the choice of parameters $a = 0.8$ and $b = 0.5$, yielding −101.8 kcal/mol and 0.7 kcal/mol, respectively. The magnitude of the difference is less for the choice $a = 0.2$ and $b = 0.0$ and, surprisingly, for $a = 1.0$ and $b = 1.0$ as well, neither of which is able, however, to reproduce the correct m.

5. Results: Two Guanine Molecules in the Arnott B-DNA Geometry

The excellent agreement described in section 4 is also observed upon examining a pair of guanine bases separated by 3.4 Å and twisted by 36° relative to one another, i.e., neighbors excised from a poly(dG)-poly(dC) strand in the Arnott B-DNA canonical structure. Calculations at both the ROHF and ROBLYP-SIC levels of theory (with $a = 0.8$ and $b = 0.5$) are predicting a high localization of the unpaired electron on the guanine radical cation (Figures 5 and 6), consistent with earlier studies of charged sequences[37,38] as well as studies examining the sensitivity of the HOMO to a change in twist angle between *neutral* guanines.[33,34] In contrast, it is shared among both guanine bases at the uncorrected ROBLYP level. However, these results must be carefully interpreted in regard to the electronic structure of adjacent guanines in a DNA helix. If a "random" configuration is examined at finite temperature in which the helical geometry is distorted from the average X-ray structure, the positive charge may be delocalized among two or three bases, but solvent effects can act to modulate the spatial extent of the charge.[19,39−41] Thus, the most accurate results can only be obtained by inclusion of explicit solvent and counterions, as well as finite-temperature (entropic) effects.

Because both dispersion interactions and hydrogen bonding between bases are essential to preserve the DNA double helix, it is important to ascertain the effect of the empirical SIC scheme on the calculated binding energies between the guanine cation and molecule. A binding energy of 7.6 kcal/mol is calculated at the uncorrected DFT-BLYP level, while that obtained upon applying the SIC with the empirical parameters, $a = 0.8$ and $b = 0.5$, is approximately halved. Interestingly, this result is independent of the choice of a and b for the values shown in Table 1. Thus, the SIC is not influencing the qualitative prediction of a binding interaction between the charged pair, although the behavior of the DNA in solution during a dynamics run could be significantly altered. In this case, a dispersion correction term may be applied.[42]

6. Additional Results and Implications for Modeling DNA

In this section, the feasibility of using the empirical SIC scheme to model the hole wavefunction in DNA base sequences

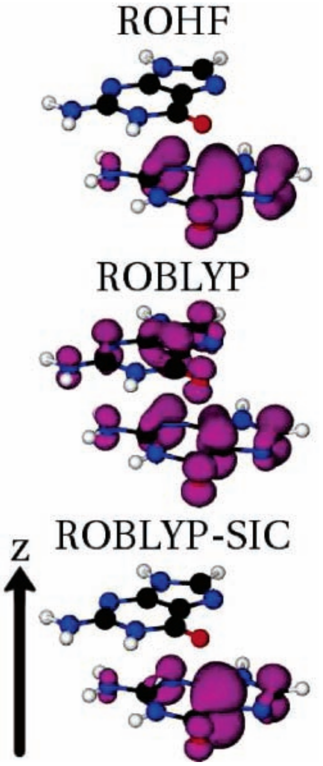

Figure 5. Isosurface contours of $m(\mathbf{r})$ (=0.002) for a separation of 3.4 Å between a guanine cation and molecule in the canonical Arnott B-DNA geometry calculated at different levels of theory. The “ROBLYP-SIC” results were obtained by setting $a = 0.8$ and $b = 0.5$.

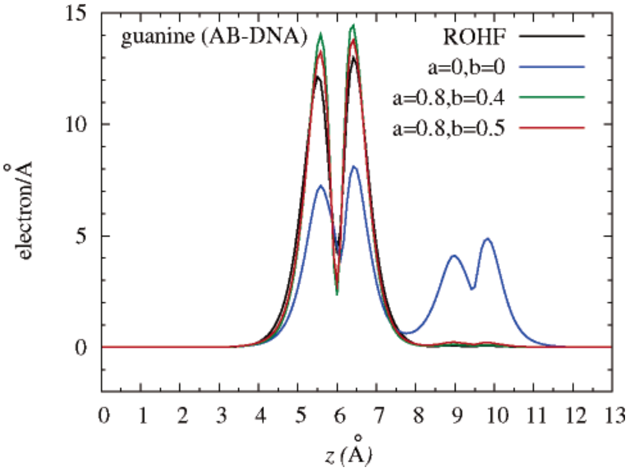

Figure 6. Projection of $m(\mathbf{r})$ between a guanine cation and molecule in the canonical Arnott B-DNA geometry (0.08 Å bin).

is discussed. First, the accuracy of this scheme in reproducing the ionization potentials of benzene and the isolated nucleotides is evaluated, followed by SIE-free DFT results obtained for the remaining stacked DNA bases. Although the calculated IP of cytosine is anomalously low, the method, with $a = 0.8$ and $b = 0.5$, can still be applied to selected DNA fibers.

TABLE 2: Vertical Ionization Potentials (eV) for Benzene and the DNA Bases at the ROBLYP Level of Theory Compared to Currently Accepted Values

		SIC for selected values of *a*,*b*			
molecule	no SIC	0.8,0.4	0.8,0.5	0.8,0.6	ref
guanine	7.56	6.84	7.64	8.37	8.24[a]
adenine	7.94	7.27	8.04	8.73	8.44[a]
cytosine	8.33	5.70	7.00	8.29	8.80[a]
thymine	8.48	7.51	8.29	9.05	~9.1[a]
benzene	8.97	7.57	8.34	9.08	9.448[b]

[a] Experimental values quoted from refs 43 and 44. [b] CCSD(T)/cc-pV∞Z value from ref 45.

An important test of the “scaled SIC” scheme is the calculation of the vertical ionization potential (VIP) for benzene and the DNA nucleotides. The VIP is defined as the difference in energy between the neutral molecule and the radical cation with the geometry of the neutral molecule. With one exception, the relative VIPs obtained with the SIC of $a = 0.8$ and $b = 0.5$ are in reasonable agreement with both the uncorrected results and the literature values[43−45] (Table 2). Unfortunately, the VIP of cytosine (7.00 eV) is substantially underestimated and might compete with guanine (7.64 eV) for the positive charge (electron hole) in model DNA fibers. By employing the ROBLYP geometry of the radical cation and calculating (with the empirical SIC of $a = 0.8$ and $b = 0.5$) the adiabatic ionization potentials (AIPs) of guanine (7.51 eV) and cytosine (7.26 eV), this difference, however, is reduced from 0.64 eV to 0.25 eV. The method could be further evaluated by calculating the VIPs of selected base sequences. For example, the ionization potential of a sequence, such as GGG, is known to be less than that of a single nucleotide[39] and might compete favorably with a cytosine base. Alternatively, this empirical SIC can be applied to a DNA double helix if certain base pairs, i.e., A-T, are treated quantum mechanically, whereas others, i.e., G-C, are modeled with a molecular mechanical force field in a mixed QM/MM computational scheme. (This setup would ensure the integrity of the hydrogen bonds forming the double helix; it may be difficult to capture these interactions in a QM/MM calculation employing a fully QM helix but a molecular-mechanical sister helix.) In Table 2, neither the parameter choice $a = 0.8$ and $b = 0.4$ nor $a = 0.8$ and $b = 0.6$ is able to yield the correct relative values of the guanine and cytosine VIPs.

Importantly, the isosurfaces of m obtained for the neutral adenine molecule and adenine radical cation can be fixed by applying the empirical SIC scheme (Figure 7, top panel). With the choice of parameters $a = 0.8$ and $b = 0.5$, the description of the adenine charged pair is consistent with ROHF calculations, whereas the uncorrected ROBLYP calculation is predicting a shared (delocalized) state at a separation of 6 Å. The situation is similar for the thymine cation−molecule pair (Figure 7, middle panel), although an exceptional localization of m is predicted for a separation of 6 Å upon applying the empirical SIC scheme. The cytosine cation−molecule is poorly described by the empirical SIC (Figure 7, bottom panel). The excessive localization for a separation of 2.5 Å, as well as some of the anomalies (i.e., spikes) for the adenine and thymine stacked pairs, is a consequence of the uncorrected ROBLYP geometries employed; the correct projections are obtained for thymine and cytosine systems using the actual optimized geometries for values of $a = 0.8$ and $b = 0.5$. Last, binding energies for these systems are presented in Table 1.

In hindsight, the success of the set of parameters $a = 0.8$ and $b = 0.5$ is not entirely surprising. In these systems, the SIE is manifested at long range, where a slowly decaying effect such as the Coulomb self-repulsion is expected to dominate, requiring a large correction to the Hartree energy in order to compensate, i.e., $a \approx 1$. By contrast, within the first solvation shell, a different set of parameters, $a = 0.2$ and $b = 0.0$, is preferred to correct for the SIE.[15] However, the difference between a and b in both of these cases is about the same, either 0.2 or 0.3. Thus, apparently the total “repulsion” experienced by the unpaired electron, including the self-interaction, exchange with electrons of like spin, and correlation effects, needs to be corrected by

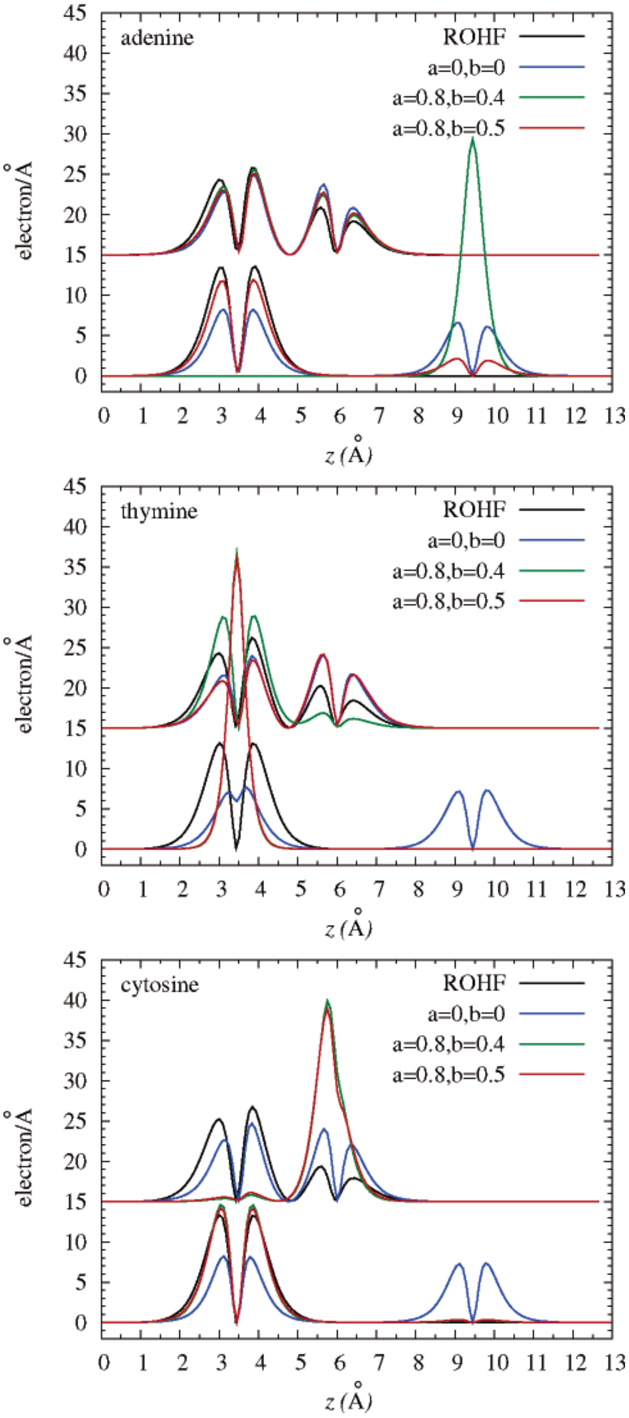

Figure 7. Projection of $m(\mathbf{r})$ for separations of 2.5 and 6.0 Å between (top panel) an adenine cation and molecule, (middle panel) thymine cation and molecule, and (bottom panel) cytosine cation and molecule (0.08 Å bin).

roughly 20−30% relative to the Hartree energy, assuming that the energy terms, $E_{\mathrm{H}}[m(\mathbf{r})]$ and $E_{\mathrm{XC}}[m(\mathbf{r}),0]$, which are opposite in sign, are also similar in magnitude. That is, the relative magnitude of the SIC appears to be approximately constant, but the forces contributing to it are exhibiting a complex distance dependence, in some way tied to the exchange-correlation (BLYP) functional.

7. Conclusions

A recently developed empirical self-interaction correction scheme is a viable method to describe the electronic structure of selected molecular pairs at both small and large separation distances. Although a truly optimal set of parameters for all π systems discussed above is not available, a reasonable compromise can be made with the selection of $a = 0.8$ and $b = 0.5$. For the benzene, guanine, and adenine stacked pairs, this choice is compensating for the shortcomings of uncorrected DFT-BLYP calculations and is correctly predicting a *localized* unpaired electron on the radical cation at large distances ranging from 6 to 7 Å, consistent with expectations. At shorter separations (2.5−3 Å), the positive charge (electron hole) is shared between eclipsed molecules. However, it is localized for the case of two guanine molecules in vacuo arranged in the "ideal" (0 K) geometry of the Arnott B-DNA, in agreement with earlier studies. This method will facilitate future studies of model DNA fibers and is currently being applied to the investigation of poly(dA)-poly(dT) using hybrid simulation techniques.

Acknowledgment. One of us (Y.A.M.) gratefully acknowledges support from a NSF postdoctoral fellowship. In addition, we would like to thank Dr. Joost VandeVondele for helpful discussions. Some figures were prepared using VMD.[46]

Supporting Information Available: A tabulation of the energies used to calculate the binding energies in Table 1. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Perdew, J. P.; Zunger, A. *Phys. Rev. B* **1981**, *23*, 5048.

(2) Gritsenko, O. V.; Ensing, B.; Schipper, P. R. T.; Baerends, E. J. *J. Phys. Chem. A* **2000**, *104*, 8558.

(3) Grafenstein, J.; Kraka, E.; Cremer, D. *Phys. Chem. Chem. Phys.* **2004**, *6*, 1096.

(4) Goedecker, S.; Umriger, C. J. *Phys. Rev. A* **1997**, *55*, 1765.

(5) Zhang, Y.; Yang, W. *J. Chem. Phys.* **1998**, *109*, 2604.

(6) Sit, P. H.-L.; Cococcioni, M.; Marzari, N. *Phys. Rev. Lett.* **2006**, *97*, 028303.

(7) Wu, Q.; Voorhis, T. V. *J. Chem. Theory Comput.* **2006**, *2*, 765.

(8) Pacchioni, G.; Frigoli, F.; Ricci, D.; Weil, J. A. *Phys. Rev. B* **2001**, *63*, 054102.

(9) Laegsgaard, J.; Stokbro, K. *Phys. Rev. Lett.* **2001**, *86*, 2834.

(10) Gruning, M.; Gritsenko, O. V.; van Gisbergen, S. J. A.; Baerends, E. J. *J. Phys. Chem. A* **2001**, *105*, 9211.

(11) Vassilev, P.; Louwerse, M. J.; Baerends, E. J. *Chem. Phys. Lett.* **2004**, *398*, 212.

(12) Becke, A. D. *J. Chem. Phys.* **2003**, *119*, 2972.

(13) Mori-Sánchez, P.; Cohen, A. J.; Yang, W. *J. Chem. Phys.* **2006**, *124*, 091102.

(14) d'Avezac, M.; Calandra, M.; Mauri, F. *Phys. Rev. B* **2005**, *71*, 205210.

(15) VandeVondele, J.; Sprik, M. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1363.

(16) Schuster, G. B., Ed. *Long-Range Charge Transfer in DNA I*; Topics in Current Chemistry, Vol. 236; Springer-Verlag: Heidelberg, 2004.

(17) Schuster, G. B., Ed. *Long-Range Charge Transfer in DNA II*; Topics in Current Chemistry, Vol. 237; Springer-Verlag: Heidelberg, 2004.

(18) Gervasio, F. L.; Carloni, P.; Parrinello, M. *Phys. Rev. Lett.* **2002**, *89*, 108102.

(19) Gervasio, F. L.; Laio, A.; Parrinello, M.; Boero, M. *Phys. Rev. Lett.* **2005**, *94*, 158103.

(20) 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.; Salvador, P.; Dannenberg, J. J.; 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 03*, revision D.01; Gaussian, Inc.: Wallingford, CT, 2004.

(21) For example, a CASSCF(7,8) calculation consists of a configuration interaction (CI) calculation on the highest occupied orbitals containing $X = 7$ electrons (for a doublet, totaling four) and the $Y - 4 = 4$ lowest virtual orbitals.

(22) Hegarty, D.; Robb, M. A. *Mol. Phys.* **1979**, *38*, 1795.

(23) Wang, J.; Becke, A. D.; Smith, V. H., Jr. *J. Chem. Phys.* **1995**, *102*, 3477.

(24) VandeVondele, J.; Krack, M.; Mohamad, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103.

(25) VandeVondele, J.; Hutter, J. *J. Chem. Phys.* **2003**, *118*, 4365.

(26) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703.

(27) Hartwigsen, C.; Goedecker, S.; Hutter, J. *Phys. Rev. B* **1998**, *58*, 3641.

(28) Martyna, G. J.; Tuckerman, M. E. *J. Chem. Phys.* **1999**, *110*, 2810.

(29) Applegate, B. E.; Miller, T. A. *J. Chem. Phys.* **2002**, *117*, 10654.

(30) Waller, M. P.; Robertazzi, A.; Platts, J. A.; Hibbs, D. E.; Williams, P. E. *J. Comput. Chem.* **2006**, *27*, 491.

(31) Sinnokrot, M. O.; Sherill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200.

(32) Crespo-Hernández, C. E.; Arce, R.; Ishikawa, Y.; Gorb, L.; Leszczynski, J.; Close, D. M. *J. Phys. Chem. A* **2004**, *108*, 6373.

(33) Sugiyama, H.; Saito, I. *J. Am. Chem. Soc.* **1996**, *118*, 7063.

(34) Sponer, J.; Leszczyński, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590.

(35) Hobza, P.; Sponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802.

(36) Cysewski, P.; Czyznikowska-Balcerak, Z. *J. Mol. Struct.: THEOCHEM* **2005**, *757*, 29.

(37) Saito, I.; Nakamura, T.; Nakatani, K.; Yoshioka, Y.; Yamaguchi, K.; Sugiyama, H. *J. Am. Chem. Soc.* **1998**, *120*, 12686.

(38) Yoshioka, Y.; Kitagawa, Y.; Takano, Y.; Yamaguchi, K.; Nakamura, T.; Saito, I. *J. Am. Chem. Soc.* **1999**, *121*, 8712.

(39) Kurnikov, I. V.; Tong, G. S. M.; Madrid, M.; Beratan, D. N. *J. Phys. Chem. B* **2002**, *106*, 7.

(40) Voityuk, A. A. *J. Phys. Chem. B* **2005**, *109*, 10793.

(41) Conwell, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 8795.

(42) Zimmerli, U.; Parrinello, M.; Koumoutsakos, P. *J. Chem. Phys.* **2004**, *120*, 2693.

(43) Close, D. M. *J. Phys. Chem. A* **2004**, *108*, 10376.

(44) Sevilla, M. D.; Besler, B.; Colson, A.-O. *J. Phys. Chem.* **1995**, *99*, 1060.

(45) Deleuze, M. S.; Claes, L.; Kryachko, E. S.; Francois, J.-P. *J. Chem. Phys.* **2003**, *119*, 3106.

(46) VMD 1.8.2, Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.