A TDDFT Study of the Optical Response of DNA Bases, Base Pairs, and Their Tautomers in the Gas Phase

Argyrios Tsolakidis* and Efthimios Kaxiras

Department of Physics and Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138

Received: November 18, 2004; In Final Form: January 11, 2005

We present calculations of the optical response of the DNA bases and base pairs both in their normal and tautomeric forms in the gas phase, using time-dependent density functional theory (TDDFT). These calculations are performed in real time within the adiabatic approximation with a basis of local orbitals. Our results for the individual bases are in good agreement with experiment and computationally more demanding calculations of chemical accuracy. The optical response of base pairs indicates that the differences between normal and tautomeric forms in certain cases are significant enough to provide a means of identification.

1. Introduction

Photophysics and photochemistry[1−6] of DNA address the interaction of this biomolecule with light. This interaction (especially with the UV part of the spectrum) is the leading cause of photodamage that can lead to carcinogenesis.[7,8] As a result of the evolutionary process, nature has developed a defense mechanism so that the excited states of the nucleic acid bases are very stable to photochemical decay through very fast decay channels for the electronic excitation energy. Accordingly, comprehension of the photochemistry of DNA begins with a detailed and accurate knowledge of its optical response, which is related to the excited state spectrum of the nucleic acid bases. Early theoretical calculations of this response are reviewed by Callis.[5] Recent increases of computational power have made possible ever more sophisticated calculations of the excitation energy spectrum. Examples of such methods are the single excitation configuration interaction (CIS),[9] which does not take into account any dynamic electron correlations, the complete active space self-consistent field method (CASSCF),[10] and the complete active space second-order perturbation theory (CASPT2).[10,11] In the last example, the dynamic electron correlations are included by means of a multi configurational second-order perturbation theory on a CASSCF wave function. A more accurate method is the multireference perturbation configuration interaction method, known as CIPSI.[12−16] Finally, time dependent density functional theory (TDDFT)[17−19] is a very promising method for the study of DNA bases, not only because of its relative computational simplicity but also because of its accuracy.

There is good agreement between the various computational methods used and experiment regarding the position of the $\pi\pi^*$ transitions, which carry considerable oscillator strength. On the other hand, for the $n\pi^*$ transitions there are great differences between various methods.[20,21] To complicate things further, the $n\pi^*$ transitions are very difficult to detect experimentally because of their very small oscillator strength and because often they are in close proximity with other $\pi\pi^*$ transitions. Despite these difficulties in their detection, knowledge of the exact energy of these transitions is important because they play an important role in the radiationless decay of the nucleic acid bases.[22] In this work we cannot characterize the nature of the different transitions because the implementation of our method is based on the electron density, and therefore we do not have any information about the wave functions. This would have been a major shortcoming only if we were specifically concerned with radiationless decay processes.

The calculation of the optical response of the nucleic acid bases is complicated by the existence of various tautomers and by solvent effects. For instance, in the case of supersonic jet experiments where many of these tautomers are detected,[6] the correct interpretation of the excited-state spectrum of DNA requires the computation of the base tautomer spectra. There exists a large number of possible tautomers, but only few are relevant in condensed phase because the (deoxy)ribosyl-substituted bases found in DNA can only have one form. In most experiments, water or a water-based solution is used as a solvent. For the simulation of the solvent effects, the self-consistent reaction field method (SCRF)[23−25] is most commonly used, and lately a small number of solvent molecules are attached to the base for more realistic representation of solvent effects.[25−27] Regardless of the method of calculation used for the determination of the excited states and the model to incorporate the solvent, it is found that the $\pi\pi^*$ transitions change very little, while the $n\pi^*$ transitions are shifted to higher energies when solvent effects are included. The shift in $n\pi^*$ transition energies increases with increasing solvent polarity and can affect the ordering of the transitions. In addition, the solvent can affect the relative population of the various tautomers that appear in the solution. One example is adenine which in gas phase is in the 9H-adenine form while in aqueous solution 20% is found to be in the 7H-adenine form.[28−31]

Finally, one interesting application of the excited states is the calculation of the geometry when the system is not in its ground state. The task is much more challenging than the determination of the ground-state geometry because the excited states lack analytical derivatives of the energy with respect to atomic displacements; that is, forces on atoms cannot be calculated. In addition, the information that can be extracted

* Corresponding author.

10.1021/jp044729w CCC: $30.25
Published on Web 02/19/2005

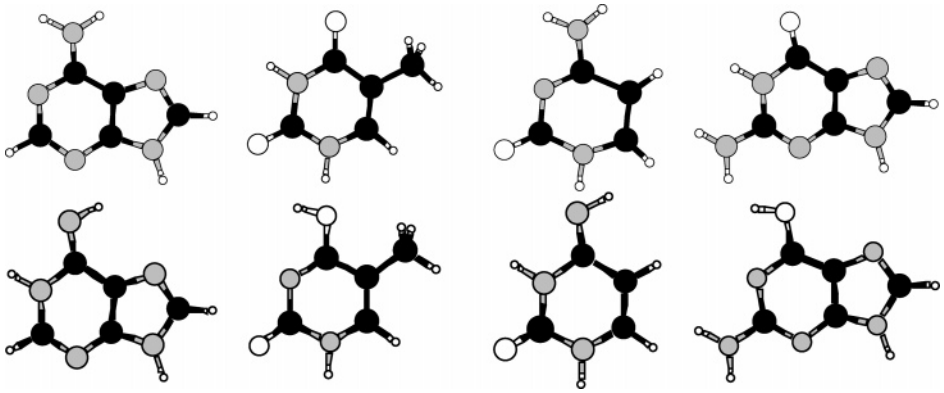

Figure 1. Single-base geometries for the normal A, T, C, G (top row, left to right), and the tautomeric forms A′, T′, C′, G′ (bottom row, left to right). C atoms are denoted with black circles, N atoms with gray, and O with white; the small circles represent H atoms.

from experiment about the excited-state geometry is very limited. From the study of the excited-state geometries it is found that they are nonplanar and that they can affect the ordering of the transitions.[26,27,32−34]

To have a more thorough understanding of the photophysics of the genetic code, it is essential to investigate the optical response of the nucleic acid base pairs. There are few recent computational studies of the base-pair spectra.[35−37] Shukla and Leszczynski have studied the adenine−uracil (AU),[35] adenine−thymine(AT), and cytosine−guanine (CG)[36] base pairs, while Sobolewski and Domcke[37] have studied the CG base pair. The reason for the limited number of base-pair studies is the increasing computational complexity involved. There are numerous other studies that deal with different aspects of the DNA base pairs such as their ground-state energies and geometries,[38−43] transport properties,[44] thermodynamic properties,[45,46] stacking properties,[47,48] vibrational modes,[49] applications to nanotechnology,[50] polarizabilities,[51] proton and charge transfer,[52,53] and their interaction with various metal cations.[54]

In this paper we propose that the base-pair absorption spectrum can be used to distinguish between different base pairs and to make the distinction between the regular (AT and CG) base pairs and the pairs that involve tautomers of nucleic acid bases. Here, by "regular pairs" we mean those in which each base exists in the normal state, which is referred as *keto*, when the H atom is part of an NH group, or *amino*, when the H atom is part of an NH_2 group. The tautomeric forms of bases, which have slightly higher energies than the normal ones, are metastable and are referred as *enol* and *imino* forms: In the enol form the H atom has left the NH group and it is attached to an O atom, forming a OH group, while in the imino form the H atom has left the NH_2 group and moved to a N atom, forming a NH group. All these forms are illustrated in Figure 1. The tautomers form hydrogen-bonded pairs with the normal bases, but because of their slightly different structure they form wrong pairs: imino-cytosine with amino-adenine (C′A), amino-cytosine with imino-adenine (CA′), keto-thymine with enol-guanine (TG′), and enol-thymine with keto-guanine (T′G). These base pairs are very similar structurally in terms of the number and position of hydrogen bonds to the regular ones. Their structures are illustrated in Figure 2. The wrong pairs, due the presence of tautomers, can lead to a wrong genetic message when DNA is transcribed.

The paper is organized as follows: In section 2, we describe the method of calculation. In section 3, we present our results for the optical absorption of isolated normal and tautomeric forms of the DNA bases and DNA base pairs. We also make comparisons with all relevant calculations and with the experimental data where available. In section 4, we give the conclusions.

2. Method of Calculation

The calculation of the optical response of the system is closely tied to the calculation of the polarizability $\alpha(\omega)$. The polarizability describes the distortion of the charge cloud caused by the application of an external electric field. It is an important response function because it is directly related to electron−electron interactions and correlations. In addition, it determines the response to charged particles and optical properties. A quantity of particular interest, which is used for the presentation of the results in this work, is the dipole strength function, $S(\omega)$, which is directly related to the frequency-dependent linear polarizability, $\alpha(\omega)$, by

$$\alpha(\omega) = \frac{e^2\hbar}{m}\int_0^\infty \frac{S(\omega')\,d\omega'}{\omega'^2 - \omega^2} \tag{1}$$

By taking the imaginary part of eq 1 we obtain the dipole strength function as

$$S(\omega) = \frac{2m}{\pi e^2 \hbar}\omega \mathrm{Im}[\alpha(\omega)] \tag{2}$$

The dipole strength function is proportional to the photoabsorption cross section, $\sigma(\omega)$, measured by most experiments and, therefore, allows direct comparison with experiment. In addition, the integration of S over energy gives the number of electrons, N_e, (f-sum rule), i.e.,

$$\int_0^\infty dE S(E) = \sum_i f_i = N_e \tag{3}$$

where f_i are the oscillator strengths. This sum rule is very important because it provides an internal consistency test of the calculations, indicating the completeness and adequacy of the basis set used for the computation of the optical response.

Our method[55] involves the description of the electronic states using linear combination of atomic orbitals (LCAO). Because the size of the LCAO basis is small for the elements involved in the nucleic acid bases, the TDDFT calculations can be done efficiently. Our scheme is based on the SIESTA[56−58] code, which is used to compute the initial wave functions and the Hamiltonian matrix for each time step. Core electrons are

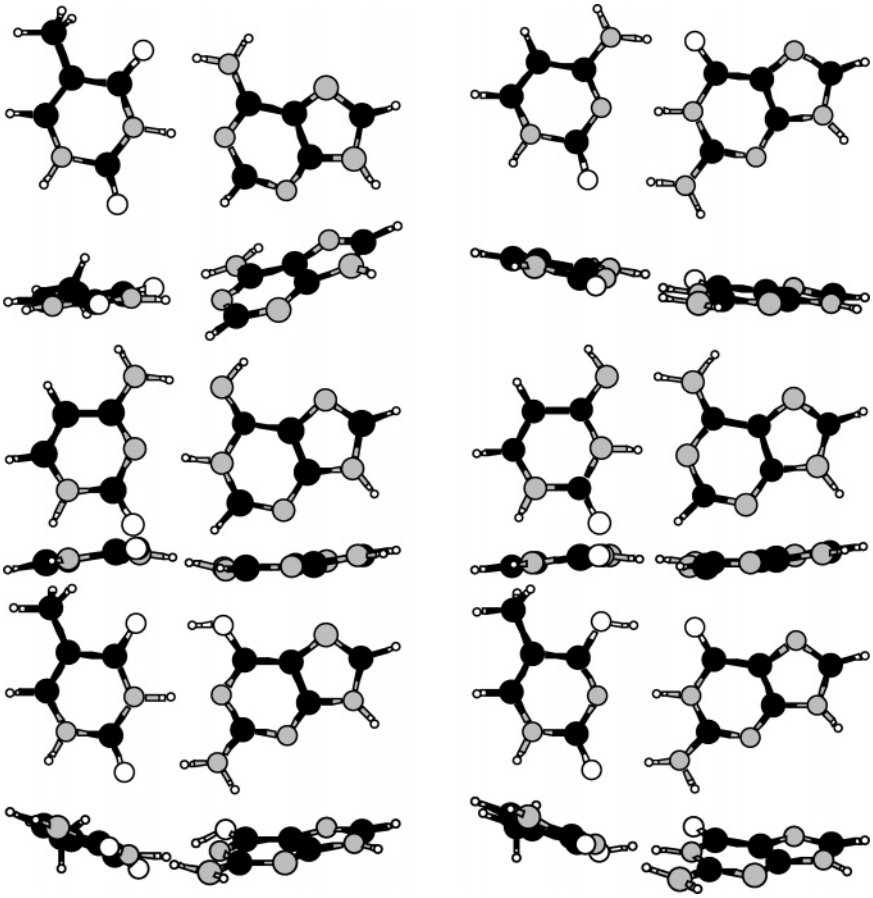

Figure 2. Top and side views of the geometries of AT and CG (first and second rows), CA′ and C′A (third and fourth rows), and TG′ and T′G (fifth and sixth rows), showing the propeller and planar geometries of the AT and CG base pairs, the shifted planar geometries of the CA′ and C′A base pairs, and the propeller geometries of the TG′ and T′G base pairs. Symbols are the same as in Figure 1.

replaced by norm-conserving pseudopotentials[59] in the fully nonlocal Kleinman−Bylander[60] form, and the basis set is a general and flexible linear combination of numerical atomic orbitals (NAOs), constructed from the eigenstates of the atomic pseudopotentials.[57,61] The NAOs are confined, being strictly zero beyond a certain radius.

Our approach is to carry out the calculations in the time domain, explicitly evolving the wave functions. We consider a bounded system in a finite electric field, i.e., the Hamiltonian includes a perturbation $\Delta\mathscr{H} = -\mathbf{E}\cdot\mathbf{x}$. For the linear response calculations in the present work we have set the value of this field to 0.1 V/Å. The system is solved for the ground state using standard time independent DFT. Then we switch off the electric field at time $t = 0$, and for every subsequent time step we propagate the occupied Kohn−Sham eigenstates by solving the time-dependent Kohn−Sham equation

$$i\hbar\frac{\partial\Psi}{\partial t} = \mathscr{H}\Psi \tag{4}$$

where $\mathscr{H}$ is the time-dependent Hamiltonian given by

$$\mathscr{H} = -\frac{\hbar^2}{2m}\nabla^2 + V_{\text{ext}}(\mathbf{r}, t) + e^2\int\frac{\rho(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|}\,\mathrm{d}\mathbf{r}' + V_{\text{xc}}[\rho](\mathbf{r}, t) \tag{5}$$

where $V_{\text{ext}}(\mathbf{r}, t)$ is the external (ionic) potential, $\rho(\mathbf{r}, t)$ is the electron density, and $V_{\text{xc}}[\rho](\mathbf{r}, t)$ is the exchange-correlation potential. The calculation of the exchange-correlation potential is done using the adiabatic local density approximation (ALDA), which is local both in time and space. For every time step we solve eq 4, and from the new wave functions we construct the new density matrix. The electron density is then obtained and used for the calculation of the Hamiltonian in the new cycle.

From the electron density at every time step we calculate the dipole moment $\mathbf{D}(t)$. This defines the response to all orders and the frequency dependent response is found by the Fourier transform

$$\mathbf{D}(\omega) \equiv \int e^{i\omega t - \delta t}\mathbf{D}(t)\,\mathrm{d}t \tag{6}$$

In our case, we Fourier transform the dipole moment only for $t > 0$. It is necessary to include a damping factor δ in order to perform the Fourier transform. This damping factor gives the minimum width of the peaks of the imaginary part of the response. Physically, it can be regarded as an approximate way to account for broadening. To linear order, the polarizability is given by $\mathbf{D}(\omega) = \alpha(\omega)\mathbf{E}(\omega)$, so that

$$\mathrm{Im}[\alpha(\omega)] = \omega\frac{\mathrm{Re}[D(\omega)]}{E} \tag{7}$$

where the field is given by $E(t) = E\ \theta(-t)$. After Fourier transforming the dipole moment, we obtain the elements of the frequency-dependent polarizability tensor $\alpha_{ij}(\omega)$. We repeat the calculation with the electric field along different axes unless the symmetry is high enough that this is not needed. The average linear polarizability is given by

$$\langle\alpha(\omega)\rangle = \frac{1}{3}\mathrm{Tr}\{\alpha_{ij}(\omega)\} \tag{8}$$

TABLE 1: Selected Calculated and Observed Excitation Energies of A

Av. exp.	TDDFT[a]	CASSCF[b]	CASPT2[b]	CIPSI[c]	TDDFT[c]	TDDFT[d]	TDDFT[e]
4.59	4.51	5.73	5.1	4.97	5.08	5.09	4.94
4.78	4.95	6.48	5.2	5.34	5.35		5.21
5.38	5.58						
5.91	5.79	7.80	6.2				5.93
6.26	6.28	8.30	6.7				6.12
	6.63						6.16
6.81	6.92	8.77	7.0				
	7.47						
7.73	7.81	9.29	7.6				

[a] This work. [b] Reference 23. [c] Reference 21. [d] Reference 22. [e] Reference 70.

The choice of the coordinate system does not affect the average polarizability because of the rotational invariance of the trace.

In all the calculations in this paper we let the systems evolve for a total time of $T = 260$ fs. The energy resolution corresponding to this choice is $\Delta\omega = \pi/T = 0.05$ eV. The time step is 15.18×10^{-3} fs, and the damping factor used in the Fourier transform is 0.0025 eV, which corresponds to a temperature of 290 K. Troullier−Martins pseudopotentials[59] and an auxiliary real-space grid[56] equivalent to a plane-wave cutoff of 50 Ry are also used in this calculation. For every C atom the basis set includes 13 NAOs: two radial shapes to represent the 2s states with confinement radii $r_s = 5.12$ au, and two additional 2p shells plus a polarization[57] p shell with confinement radii $r_p = r_p^{Pol} = 6.25$ au. For every H atom we have 5 NAOs: two radial shapes for the 1s orbital and a polarization s orbital with confinement radii $r_s = r_s^{Pol} = 6.05$ au. For the N atom we have 13 orbitals: two 2s shells, two 2p shells, and a polarization p orbital with radii $r_s = 4.50$ au and $r_p = r_p^{Pol} = 5.50$ au, respectively. Finally, the number of orbitals for every O atom is 13, with two 2s shells, two 2p shells, and a p polarization shell with confinement radii $r_s = 3.93$ au and $r_p = r_p^{Pol} = 4.93$ au, respectively. For the systems under study it was found that inclusion of additional shells does not alter noticeably the results. The completeness of the basis is also evident from the fact that the *f*-sum rule for all the systems is in the 96−98% range.

3. Discussion of Results

A. Single Bases. Although there are quite a few calculations for the DNA bases[20−27,62−72] even at the TDDFT level, we briefly include these results in order to demonstrate the level of accuracy of our method and to compare it with other established approaches and with experiment.[20,73−87] One new aspect of our work is the calculation of the optical response for the whole energy range beyond 8 eV, which was the upper limit in all other studies until now. This is made possible by the real-time implementation of TDDFT. The results at very high energies must be viewed with caution because of the use of a local orbital basis.

Changes in the geometry of DNA bases can affect the oscillator strength and the position of the peaks. Therefore, before presenting such results the geometry of the bases under consideration has to be carefully described. In our case the bases were relaxed at the LDA level without any constraint, and their geometry is found to be planar. In other calculations it is found that the amino group of the DNA bases is not planar with the rest of the base. In addition, detailed knowledge of the spectrum of the various tautomers is very important for the explanation of the experimental results. In this paper we present the results for the imino-adenine (A′) and -cytosine (C′), and for the enol-guanine (G′) and -thymine (T′), for which a limited number of calculations are available.[25−27,72]

Extensive discussions of the experimental results are provided by Fülscher et al.[67] for C and its derivatives, Lorentzon et al.[68] for T and its derivatives, and Fülscher et al.[23] for A, G, and their derivatives. In this work only a brief discussion of the absorption spectrum of the bases is included.

1. Amino- and Imino-Adenine. For A in the gas phase the first peak is measured at 4.92 eV,[73] while in solution it is shifted to lower energies around 4.77 eV.[73] It was later realized that the lowest energy peak consisted of two peaks, one centered around 4.59 eV and the other around 4.78 eV. The splitting was observed with linear dichroism (LD)[20,77] and magnetic circular dichroism (MCD)[75,80] but not by the circular dichroism (CD) measurements.[78,79] All calculations of the spectrum of A in the gas phase so far (including the present one), regardless of the method used, have found two peaks in this range. In our calculation the two lowest peaks are at 4.51 and 4.80 eV, in good agreement with the average experimental values of 4.59 and 4.78 eV. In the experiments it is also observed that the first peak carries more oscillator strength than the second one. This is not the case in our calculation nor for the CASPT2 calculation by Fülscher et al.,[23] but we expect the inclusion of hydration effects to shift oscillator strength from the first peak to the second.

The spectrum also contains $n\pi^*$ transitions, but because these carry oscillator strength which is typically 2 orders of magnitude smaller compared to the strength of $\pi\pi^*$ transitions, they are very difficult to detect and are masked by the envelopes of the $\pi\pi^*$ transitions. The peak around 5.4 eV detected by the CD experiments in adenosine[78,79] is attributed to a $n\pi^*$ transition. A band at 5.38 eV is also observed by LD measurements.[20] This band may correspond to the third band appearing in the calculated absorption spectrum at 5.58 eV.

The fourth band observed by experiment at the average position of 5.91 eV corresponds to the fourth band in our calculation at 6.28 eV, while our fifth band located at 6.63 eV may correspond to the band observed at 6.50 eV in the electron scattering experiment.[81] In our calculation, the band at 6.92 eV is by far the most intense and is identified with the experimental peak observed at an average energy of 6.81 eV.

The results for various other ab initio calculations and the values of the average experimental values for A are summarized in Table 1. Here, we chose to include only $\pi\pi^*$ transitions, because the agreement of those transitions between different calculations and with experiment is typically much better than other types of transitions. The optical response for the low and whole energy spectrum for adenine is shown in Figure 3.

We have not found published results (theory or experiment) for the tautomeric form A′. We expect the calculated spectrum shown in Figure 3 to be close to the actual values, judging from the agreement between our results and experiment for A. The number of peaks in the absorption spectrum of both forms of adenine up to 8 eV is the same. In A′ the splitting of the two

TABLE 2: Average Experimental and Selected Calculated Values of the Low Energy Excitations of C

Av. exp.	TDDFT[a]	CASSCF[b]	CASPT2[b]	CASSCF[c]	CASSCF[d]	MRCI/RPA[e]	TDDFT[f]	TDDFT[g]
4.60	4.10	5.18	4.39	5.0	4.81	5.5/6.19	4.65	4.64
5.35	4.90	6.31	5.36	6.6	6.68	6.8/7.40	5.39	5.43
5.89	5.92					6.8/7.54	6.11	
6.26	6.39	7.30	6.16		6.98	7.3/7.83	6.32	
6.62	6.48	7.82	6.74	8.2	7.92		6.46	
7.37	6.88	9.13	7.61					
	7.16							

[a] This work. [b] Reference 67. [c] Reference 64. [d] Reference 62. [e] Reference 66. [f] Reference 70. [g] Reference 27.

TABLE 3: Selected Calculated and Observed Excitation Energies of G

Av. exp.	TDDFT[a]	CASSCF[b]	CASPT2[b]	INDO/SCI[c]	CIS[c]	CIS[d]	CIPSI[e]	TDDFT[f]
4.50	4.46	6.08	4.76	4.07	6.12	4.44	4.76	4.85
4.96	4.71	6.99	5.09	4.69	6.93	4.46		5.11
	5.04			5.08	8.01	4.53		
5.62	5.64	7.89	5.96			4.56	5.64	5.59
6.23	6.23	8.60	6.55	6.16	8.91	4.91		5.83
6.58	6.53	8.69	6.65	6.65	9.05	5.05		
		9.43	6.66					
6.70	6.82	9.76	6.77					
	6.93							
	7.26							

[a] This work. [b] Reference 23. [c] Reference 25. [d] Reference 26. [e] Reference 69. [f] Reference 70.

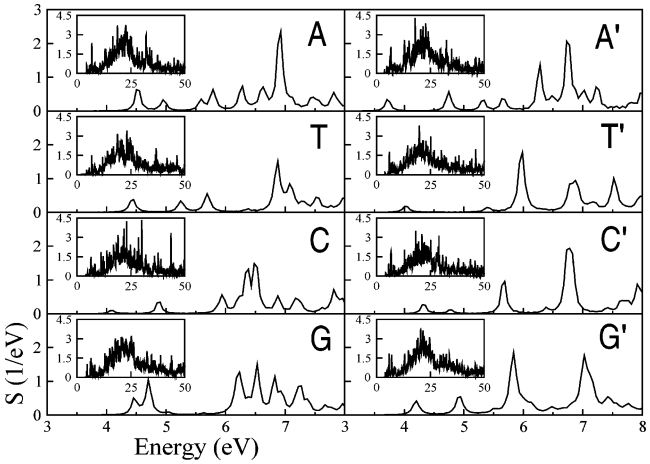

Figure 3. Low- and whole-energy absorption spectra of A, A′, T, T′, C, C′, G, and G′.

lowest energy peaks is very big, close to 1 eV and the lowest energy peak carries less oscillator strength than the second one, as it should be for A. The lowest energy peak is at 3.71 eV considerably lower compared to first peak at 4.51 eV for A, indicating a smaller HOMO−LUMO gap for A′, by 0.8 eV.

2. Amino- and Imino-Cytosine. For C, the average experimental and theoretical results, including ours, are summarized in Table 2. Six bands are observed in experiment at the average energies 4.60, 5.35, 5.89, 6.26, 6.62, and 7.37 eV while in our calculation, for the same energy range we find seven peaks at 4.10, 4.90, 5.92, 6.39, 6.48, 6.88, and 7.16 eV. While agreement between calculation and experiment is not very good at low energies, the situation is somewhat better in the high end of this range. The results for C are the least satisfactory compared to all the other isolated bases we studied.

For C′ there exist calculations by Shukla et al.[27] and experiments by Kaito et al.[72] Shukla et al.[27] studied the excitation spectrum of both the isolated and the hydrated C′ and they observed minor changes of the energies of the $\pi\pi^*$ transitions and a blue shift of the $n\pi^*$ transitions. The experiment of Kaito et al.[72] measured two peaks, one at 4.64 eV and the other at 5.59 eV, which is close to the peak at 5.68 eV predicted by our calculation. The optical response of the C′ is dominated by two peaks at 5.68 and 6.77 eV, which carry most of the oscillator strength of the low energy spectrum, as shown in Figure 3.

3. 9NH Keto- and Enol-Guanine. For guanine there is consensus between different experimental measurements about the position of the observed peaks in optical absorption: two peaks observed in the low energy part of the spectrum, one at 4.50 eV, the other at 4.96 eV. Our calculations predict two peaks at 4.46 and 4.71 eV, with the correct relative oscillator strengths. From 5 to 6 eV two quite weak peaks appear, one at 5.04 eV and the other at 5.64 eV, but only the second peak is observed experimentally. At higher energies two strong peaks are observed (at 6.11 and 6.59 eV), in very good agreement with our calculated values (at 6.23 and 6.53 eV). At still higher energies our calculated spectrum predicts peaks at 6.82, 6.93, and 7.26 eV. The entire calculated spectrum is shown in Figure 3. In Table 3 the average experimental values are given and various calculations are summarized.

Guanine has four forms (9NH enol and keto, and 7NH enol and keto), all of which are detected in supersonic jet experiments.[88−90] For the 9NH enol form of guanine (G′), four peaks were measured in the experiment of Mons et al.,[90] one at 4.31 eV and three other peaks between 4.51 and 4.52 eV. Theoretical results have been reported by Broo and Holmén[25] using the semiempirical INDO/S-CI method and by Shukla et al.[26] using CIS. In our calculations we obtain the lowest peak at 4.20 eV (the same value as in the calculation of Broo and Holmén), which is in reasonable agreement with the experimentally observed peak[90] at 4.31 eV. Shukla et al. have argued that the NH7 keto form of guanine will be dominant in aqueous solution while the NH9 keto (G) form will be dominant in the gas phase, and that the enol forms will not be in abundance; these authors calculated the lowest peak at 4.54 eV, which is very close to the experimentally observed peaks at 4.51−4.52 eV. These results are summarized in Table 5.

The spectra of G and G′ are actually quite distinct from each other, as is evident in Figure 3. In particular, there are six prominent peaks in the spectrum of *G* in the range up to 7.5 eV, but only four peaks for G′, two of which are dominant.

4. Keto- and Enol-Thymine For T there is also consensus between different experimental measurements about the position

TABLE 4: Selected Calculated and Observed Excitation Energies of T

Av. exp.	TDDFT[a]	CASSCF[b]	CASPT2[b]	TDDFT[c]	INDO[d]
4.67	4.45	6.75	4.88	4.96	4.98
5.30	5.24				
5.85	5.68	7.15	5.88	5.95	5.81
6.20	6.38	8.33	6.10	6.19	
				6.50	6.55
6.80	6.86	8.62	7.13	6.86[e]	
	7.07				7.17
	7.28				
	7.52				

[a] This work. [b] Reference 68. [c] Reference 70. [d] Reference 25. [e] There are 15 closely spaced transitions in the range 6.86−7.37 eV, which are not included in this table.

of the observed peaks in the spectrum: the first peak appears at 4.8 eV in the gas phase[73] but is shifted to lower energies, around 4.5−4.7 eV in solution.[74,76,87] Our results are in very good agreement with experiment for the first band (4.5−4.8 eV in experiment vs 4.45 eV), the second band (5.7−6.1 eV in experiment vs 5.68 eV), the third band (6.2−6.5 eV in experiment vs 6.38 eV), and the fourth band (6.8−7.0 eV in experiment vs 6.86 eV). We find two additional peaks, at 5.24 and 7.07 eV, the latter in good agreement with the band calculated at 7.13 by Lorentzon et al.[68] The peak at 5.24 eV could be related to an $n\pi^*$ transition measured in circular dichroism experiments of thymidine,[75] although the large oscillator strength we find is not characteristic of $n\pi^*$ transitions. A summary of selected calculations and the average experimental values for *T* is given in Table 4. A large number of the transitions calculated by Shukla et al.[70] around 7 eV is unlikely to be observed by experiment.

For *T*′ we find a spectrum with little oscillator strength up to relatively high energies (6 eV). The first peak appears at 4.00 eV and the second at 5.39 eV, both quite weak, while the first peak of considerable oscillator strength appears at 5.98 eV.

B. DNA Base Pairs. The study of the isolated DNA bases is only the first step toward the understanding of the photophysics of living matter. In the natural environment, the bases are hydrogen bonded in Watson−Crick pairs,[91,92] they are covalently bonded to the backbone sugars,[93−96] and they interact by van der Waals forces.[97,98] In this part we will study the effect of the hydrogen bonding between different base pairs by calculating the optical response of the pairs. The computational requirements for this case increase considerably relative to single bases, especially for methods of chemical accuracy which scale very unfavorably with increasing size of the system. As a result, there are very few theoretical calculations for the spectrum of DNA base pairs.[36,37]

The calculation of Shukla and Leszczynski[36] for the adenine−thymine (AT) and cytosine−guanine (CG) pairs was done at the CIS level, and their main finding was that the electronic excitations are localized at one of the bases comprising the pair, with the exception of some states which involve charge transfer. The $\pi\pi^*$ transitions were unaffected by the hydrogen bonding, while the $n\pi^*$ transitions were shifted to higher energies.[5,99] The ground-state geometries of the pairs were found to be planar and the formation of the pair affects the oscillator strength of the transitions. Sobolewski and Domcke[37] studied the photophysics of the CG pair using the CASPT2 method, which is more sophisticated and accurate than CIS. They reported the energies of the three lowest transitions, one of which they argued is a charge transfer transition. According to these authors not only is the energy of $\pi\pi^*$ transitions changed by −0.17 eV for the first transition and +0.40 eV for the second one, but also the oscillator strength is affected by hydrogen bonding. Finally, the CG was found to have a propeller structure in this calculation.

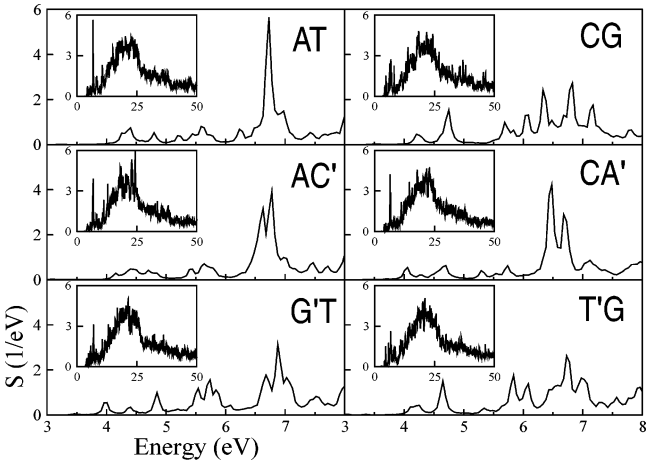

Figure 4. Low- and whole-energy absorption spectra of AT, CG, AC′, CA′, G′T, and T′G base pairs.

In the optical response of DNA base pairs we observe peaks with very small oscillator strength at very low energies. We believe that these peaks correspond to charge transfer states. The energy of these states is greatly underestimated in TDDFT when local exchange-correlation functionals are used, as proven by Dreuw et al.[100]

*1. Amino-Adenine−Keto-Thymine (*AT*) Base Pair.* The relaxed, without any constraints, structure of AT was found to have a propeller structure, shown in Figure 2. Our results indicate that there is a correspondence between the transitions of the isolated adenine and thymine and the transitions that appear at the excitation spectrum of the AT pair up to 8 eV (Figure 4). The excitations of the AT pair differ in energy by at most 0.2 eV (usually by just 0.1 eV) from the corresponding excitations in adenine and thymine. A very striking feature of

TABLE 5: Selected Calculations and Experimental Values of the Low Energy Excitations of the Tautomeric Forms of the DNA Bases

A′	T′	C′			G′			
this work	this work	this work	ref 27	ref 72	this work	ref 25	ref 26	ref 90
3.71	4.00	4.33	4.77	4.64	4.20	4.20	4.54	4.31
4.74	5.39	4.77			4.94	4.62	4.58	4.51
5.34	5.98	5.68	5.97	5.59	5.50	5.37	4.83	4.52
5.65	6.79	6.38	6.52		5.84	6.00	4.83	4.52
6.28	6.87	6.77	6.85		6.13	6.40	4.92	
6.48	7.17	7.43			6.48		4.94	
6.73	7.51				6.68			
6.87					7.02			
7.02					7.42			
7.22								

TABLE 6: Position of the Peaks at the Optical Response of the Normal and Tautomeric Forms of DNA Base Pairs

AT	AC′	G′T	CG	CA′	T′G
3.22	3.09	2.61	2.37	2.82	2.71
4.05	3.63	3.48	3.53	3.70	3.02
4.26	4.15	3.99	4.21	4.05	3.36
4.40	4.42	4.38	4.75	4.27	4.11
4.80	4.69	4.84	5.52	4.68	4.25
5.20	4.85	5.54	5.69	5.28	4.65
5.45	5.41	5.73	5.84	5.52	5.33
5.60	5.64	5.84	6.07	5.74	5.48
5.76	5.78	6.07	6.33	6.04	5.65
6.24	6.27	6.69	6.47	6.24	5.84
6.48	6.62	6.87	6.68	6.47	6.08
6.73	6.77	7.51	6.83	6.68	6.43
6.97	6.98		7.18	7.12	6.57
	7.17		7.28	7.37	6.74
	7.46				6.97
					7.29

the AT absorption spectrum is the single dominant peak at 6.73 eV, which can serve as a signature for the AT base pair. Very minor features appear in the spectrum at lower energies (3.22 and 4.05 eV), which most probably correspond to charge transfer states.

2. Amino-Cytosine−Keto-Guanine (CG) Base Pair. For the *CG* base pair, we find an almost planar structure after full relaxation (Figure 2). The correspondence between the transitions of the pair and the constituting monomers is not as clear as in the case of the AT base pair. A possible reason could be the larger number of hydrogen bonds in the CG pair which might affect the spectrum. At the very low end of the spectrum two very weak peaks appear, at 2.37 and 3.53 eV, which most probably correspond to charge-transfer states. Comparing our results with the CASPT2 results of Sobolewski and Domke,[22] we find very similar features (our peaks at 4.21 and 4.75 eV correspond to their peaks at 4.35 and 4.75 eV). We note, however, that their structure for the CG pair was a propeller in contrast to our almost planar geometry (Figure 2), which can lead to shifts in the transition energies. Finally, we consider the optical response of base-pairs that involve tautomers. Differences in the spectra of the regular and tautomeric forms may serve as a means for detecting the presence of tautomers in DNA.

3. Amino-Cytosine−Imino-Adenine (CA′) and Imino-Cytosine−Amino-Adenine (C′A) Base Pairs. The optical absorption spectra of CA′ and C′A base pairs exhibit quite a few similarities and the actual structure of these pairs involves the same shifted-plane geometry (shown in Figure 2). Previous studies for the normal base pairs have shown that most of the excitations are localized at the constituting monomers and their optical absorption is practically the superposition of the optical absorption of the monomers.[26,37] For CA′ and C′A, the majority of peaks that appear in the absorption spectra (Table 6) correspond to features of the constituting bases, slightly shifted to higher or lower energies (not more than 0.2 eV) and with modified oscillator strength. A prominent feature of both spectra that might be used for the identification of these base pairs is the split peak of large oscillator strength centered around 6.5 eV (see Figure 4).

4. Keto-Thymine−Enol-Guanine (TG′) and Enol-Thymine−Keto-Guanine (T′G) Base Pairs. The *TG′* and *T′G* tautomer base pairs exhibit the propeller structure shown in Figure 2. The peaks in the optical response (Table 6) are in even closer correspondence to peaks of the monomers than for the CA′ and C′A pairs. A possible explanation of this feature is that the propeller structure makes transitions from one monomer to the other more difficult because of the twisted geometry. The spectra of TG′ and T′G exhibit a prominent set of peaks in the region 5.5−6.0 eV and another set of split peaks in the region 6.5−7.3 eV.

4. Conclusions

In this work we reported a thorough investigation of the optical response of the isolated DNA base pairs, their tautomers, and the regular and tautomeric pairs using a real-time implementation of TDDFT. For the isolated DNA base pairs, our results compare very well with experiment and with other calculations based on computationally more demanding methodologies. Both the regular and tautomeric forms of the base pairs exhibit features in their spectra which can be traced to features in the spectra of their constituting monomers. Certain of the prominent features of their optical response could be useful in distinguishing between the regular and tautomeric forms.

Acknowledgment. This work was supported by the Nanoscale Science and Engineering Center of Harvard University which is funded by the National Science Foundation. A.T. is grateful for the hospitality of the Department of Materials Science and Engineering, University of Ioannina, Greece, where part of the work was carried out.

References and Notes

(1) Ruzsicska, B. P.; Lemaire, D. G. E. in *CRC Handbook of Organic Photochemistry and Photobiology*; Horspool, W. M., Song, P. S.; Eds.; CRC Press: Boca Raton, FL, 1995; p 1289.
(2) Vigny, P.; Duquesne, M. in *Excited States of Biological Molecules*; Birks, J. B.; Wiley: New York, 1976; Vol. 3, p 167.
(3) Daniels, M. in *Photochemistry and Photobiology of Nucleic Acids*; Wang, S. Y., Ed.; Academic Press: New York, 1976; Vol. 1, p 23.
(4) Daniels, M. *Photochem. Photobiol.* **1983**, *37*, 691.
(5) Callis, P. R. *Annu. Rev. Phys. Chem.* **1983**, *37*, 691.
(6) Crespo-Hernández, C. E.; Cohen, B.; Hare, P. M.; Kohler, B. *Chem. Rev.* **2004**, *104*, 1977.
(7) Miller, D. L.; Weinstock, M. A. *J. Am. Acad. Dermatol.* **1994**, *30*, 774.
(8) Kraemer, K. H. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11.
(9) Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *J. Phys. Chem.* **1992**, *96*, 135.
(10) Anderson, K.; Malmqvist, P. A.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218.
(11) Fülscher, M. P.; Anderson, K.; Roos, B. O. *J. Phys. Chem.* **1992**, *96*, 9204.
(12) Huron, B.; Malrieu, J.-P.; Rancurel, P. *J. Chem. Phys.* **1973**, *58*, 5745.
(13) Evangelisti, S.; Daudey, J.; Malrieu, J.-P. *Chem. Phys.* **1983**, *75*, 91.
(14) Spiegelmann, F.; Malrieu, J.-P. *J. Phys. B* **1984**, *17*, 1235.
(15) Cimiraglia, R. J. *J. Chem. Phys.* **1985**, *83*, 1746.
(16) Cimiraglia, R.; Persico, M. *J. Comput. Chem.* **1987**, *8*, 39.
(17) Runge, E.; Gross, E. K. U. *Phys. Rev. Lett.* **1984**, *52*, 997.
(18) Gross, E. K. U.; Ullrich, C. A.; Gossmann, U. J. in *Density Functional Theory*; Gross, E. K. U., Dreizler, R. M.; Eds.; Plenum Press: New York, 1995.
(19) Gross, E. K. U.; Dobson, J. F.; Petersilka, M. in *Topics in Current Chemistry*; Nalewajski, R. F., Ed.; Springer-Verlag: Berlin, 1996; Vol. 181.
(20) Holmén, A.; Broo, A.; Albinsson, B.; Nordén, B. *J. Am. Chem. Soc.* **1997**, *119*, 12240.
(21) Mennucci, B.; Toniolo, A.; Tomasi, J. *J. Phys. Chem.* **2001**, *105*, 4749.
(22) Sobolewski, A. L.; Domcke, W. *Eur. Phys. J. D* **2002**, *20*, 369.
(23) Fülscher, M. P.; Serano-Andres, L.; Roos, B. O. *J. Am. Chem. Soc.* **1997**, *119*, 6168.
(24) Mishra, S. K.; Shukla, M. K.; Mishra, P. C. *Spectrochim. Acta Part A* **2000**, *56*, 1355.
(25) Broo, A.; Holmén, A. *J. Phys. Chem. A* **1997**, *101*, 3589.
(26) Shukla, M. K.; Mishra, S. K.; Kumar, A.; Mishra, P. C. *J. Comput. Chem.* **2000**, *21*, 826.
(27) Shukla, M. K.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 11338.
(28) Lührs, D. C.; Viallon, J.; Fischer, I. *Phys. Chem. Chem. Phys.* **2001**, *3*, 1827.

(29) Cohen, B.; Hare, P. M.; Kohler, B. *J. Am. Chem. Soc.* **2003**, *125*, 13594.
(30) Dreyfus, M.; Dodin, G.; Bensaude, O.; Dubois, J. E. *J. Am. Chem. Soc.* **1975**, *97*, 2369.
(31) Holmén, A.; Broo, A. *Int. J. Quantum Chem.* **1995**, *22*, 113.
(32) Langer, H.; Doltsinis, N. L. *J. Chem. Phys.* **2003**, *118*, 5400.
(33) Broo, A. *J. Phys. Chem. A* **1998**, *102*, 526.
(34) Shukla, M. K.; Mishra, P. C. *Chem. Phys.* **1999**, *240*, 319.
(35) Shukla, M. K.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 1011.
(36) Shukla, M. K.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 4709.
(37) Sobolewski, A. L.; Domcke, W. *Phys. Chem. Chem. Phys*. **2004**, *6*, 2763.
(38) Kumar, A.; Knapp-Mohammady, M.; Mishra, P. C.; Suhai, S. *J. Comput. Chem.* **2004**, *25*, 1047.
(39) Joubert, L.; Popelier, P. L. A. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4353.
(40) Podolyan, Y.; Rubin, Y. V.; Leszczynski, J. *J. Phys. Chem. A* **2000**, *104*, 9964.
(41) Komarov, V. M. *J. Biol. Phys.* **1999**, *24*, 167.
(42) Šponer, J.; Hobza, P. *Chem. Phys.* **1996**, *204*, 365.
(43) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965.
(44) Yanov, I.; Leszczynski, J. *J. Quantum Chem.* **2004**, *96*, 436.
(45) Hobza, P.; Šponer, J. *Chem. Phys. Lett.* **1996**, *261*, 379.
(46) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302.
(47) Sivanesan, D.; Sumathi, I.; Welsh, W. J. *Chem. Phys. Lett.* **2003**, *367*, 351.
(48) Jurečka, P.; Nachtigall, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578.
(49) Florián, J.; Leszczynski, J.; Johnson, B. G. *J. Mol. Struct.* **1995**, *349*, 421.
(50) Seeman, N. C. *Nature* **2003**, *421*, 427.
(51) Jasíen, P. G.; Fitzgerald, G. *J. Chem. Phys.* **1990**, *93*, 2554.
(52) Guallar, V.; Douhal, A.; Moreno, M.; Lluch, J. M. *J. Phys. Chem. A* **1999**, *103*, 6251.
(53) Scheiner, S.; Kern, C. W. *J. Am. Chem. Soc.* **1979**, *101*, 4081.
(54) Burda, J. V.; Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. B* **1997**, *101*, 9670.
(55) Tsolakidis, A.; Sánchez-Portal, D.; Martin, R. M. *Phys. Rev. B* **2002**, *66*, 235416.
(56) Sánchez-Portal, D.; Ordejón, P.; Artacho, E.; Soler, J. M. *Int. J. Quantum Chem.* **1997**, *65*, 453.
(57) Artacho, E.; Sánchez-Portal, D.; Ordejón, P.; García, A.; Soler, J. M. *Phys. Status Solidi B* **1999**, *215*, 809.
(58) Ordejón, P. *Phys. Status Solidi B* **2000**, *217*, 335.
(59) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993.
(60) Kleinman, L.; Bylander, D. M. *Phys. Rev. Lett.* **1982**, *48*, 1425.
(61) Sankey, O. F.; Niklewski, D. J. *Phys. Rev. B* **1989**, *40*, 3979. Sánchez-Portal, D.; Artacho, E.; Soler, J. M. *J. Phys.: Condens. Matter* **1996**, *8*, 3859.
(62) Fülscher, M. P.; Malmqvist, P. Å; Roos, B. O. Ab Initio Quantum Chemical Calculations of Excitation Energies and Transition Moments for the Nucleic Acid Base Monomers, in *Time-Resolved Laser Spectroscopy in Biochemistry II*; Lakowicz, J., Ed.; SPIE: Belingham, WA, 1990; p 1204.
(63) Jensen, H. J. A.; Koch, H.; Jørgensen, P.; Olsen, J. *Chem. Phys.* **1998**, *119*, 297.
(64) Matos, J. M. O.; Roos, B. O. *J. Am. Chem. Soc.* **1988**, *110*, 7664.
(65) Petke, J. D.; Maggiora, G. M.; Christoffersen, R. E. *J. Am. Chem. Soc.* **1990**, ***112***, 5452.
(66) Petke, J. D.; Maggiora, G. M.; Christoffersen, R. E. *J. Phys. Chem.* **1992**, *96*, 6992.
(67) Fülscher, M. P.; Roos, B. O. *J. Am. Chem. Soc.* **1995**, *117*, 2089.
(68) Lorentzon, J.; Fülscher, M. P.; Roos, B. O. *J. Am. Chem. Soc.* **1995**, *117*, 9265.
(69) Mennucci, B.; Toniolo, A.; Tomasi, J. *J. Phys. Chem.* **2001**, *105*, 7126.
(70) Shukla, M. K.; Leszczynski, J. *J. Comput. Chem*. **2004**, *25*, 768.
(71) Merchán, M.; Serrano-Andrés, L. *J. Am. Chem. Soc*. **2003**, *125*, 8108.
(72) Kaito, A.; Hatano, M.; Ueda, T.; Shibuya, S. *Bull. Chem. Soc. Jpn.* **1980**, *53*, 3073.
(73) Clark, L. B.; Peschel, G. G.; Tinoco, I., Jr. *J. Phys. Chem.* **1965**, *69*, 3615.
(74) Voet, D.; Gratzer, W. B.; Cox, R. A.; Doty, P. *Biopolymers* **1963**, *1*, 193.
(75) Voelter, W.; Records, R.; Bunnenberg, E.; Djerassi, C. *J. Am. Chem. Soc.* **1968**, *90*, 6163.
(76) Yamada, T.; Fukutome, H. *Biopolymers* **1968**, *6*, 43.
(77) Matsuoka, Y.; Nordén, B. *J. Phys. Chem.* **1982**, *86*, 1378.
(78) Sprecher, C. A.; Johnson, W. C., Jr. *Biopolymers* **1977**, *16*, 2243.
(79) Brunner, W. C.; Maestre, M. F. *Biopolymers* **1975**, *14*, 555.
(80) Sutherland, J. C.; Griffin, K. *Biopolymers* **1984**, *23*, 2715.
(81) Isaacson, M. *J. Chem. Phys.* **1972**, *56*, 1803.
(82) Zaloudek, F.; Novros, J. S.; Clark, L. B. *J. Am. Chem. Soc.* **1985**, *107*, 7344.
(83) Raksányi, K.; Földváry, I.; Fidy, J.; Kittler, L. *Biopolymers* **1978**, *17*, 887.
(84) Morita, H.; Nagakura, S. *Theor. Chim. Acta* **1968**, *11*, 279.
(85) Clark, L. B. *J. Am. Chem. Soc*. **1977**, *99*, 3934.
(86) Miles, D. W.; Hahn, S. J.; Robins, R. K.; Eyring, H. *J. Phys. Chem.* **1968**, *72*, 1483.
(87) Clark, L. B.; Tinoco, I., Jr. *J. Am. Chem. Soc.* **1965**, *87*, 11.
(88) Chin, W.; Mons, M.; Dimicoli, I.; Piuzzi, F.; Tardivel, B.; Elhanine, M. *Eur. Phys. J. D* **2002**, 20, 347.
(89) Nir, E.; Plutzer, C.; Kleinermanns, K.; de Vries, M. *Eur. Phys. J. D* **2002**, *20*, 317.
(90) Mons, M.; Dimicoli, I.; Piuzzi, F.; Tardivel, B.; Elhanine, M. *J. Phys. Chem. A* **2002**, *106*, 5088.
(91) Watson, J. H. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.
(92) Löwdin, P. O. *Adv. Quantum Chem.* **1965**, *2*, 213.
(93) Nikogosyan, D. N.; Angelov, D.; Soeb, B.; Lindqvist, L. *Chem. Phys. Lett*. **1996**, *252*, 322.
(94) Pecourt, J.-M. L.; Peon, J.; Kohler, B. *J. Am. Chem. Soc.* **2001**, *123*, 10370.
(95) Peon, J.; Zewail, A. H. *Chem. Phys. Lett.* **2001**, *348*, 255.
(96) Onidas, D.; Markovitsi, D.; Marguet, S.; Sharonov, A.; Gustavsson, T. *J. Phys. Chem. B* **2002**, *106*, 11367.
(97) Hall, D. B.; Holmlin, R. E.; Barton, J. K. *Nature* **1996**, *384*, 731.
(98) Steenken, S. *Chem. Rev.* **1989**, *89*, 503.
(99) Brealey, G. J.; Kasha, M. *J. Am. Chem. Soc.* **1955**, *77*, 4462.
(100) Dreuw, A.; Weisman, J. L.; Head-Gordon, M. *J. Chem. Phys.* **2003**, *119*, 2943.