Toward the understanding of DNA fluorescence: The singlet excimer of cytosine

Gloria Olaso-González, Daniel Roca-Sanjuán, Luis Serrano-Andrés, and Manuela Merchán[a)]
Instituto de Ciencia Molecular, Universitat de València, Apartado 22085, ES-46071 Valencia, Spain

(Received 16 October 2006; accepted 16 November 2006; published online 18 December 2006)

By using the multiconfigurational second-order perturbation method CASPT2, including corrections for the basis set superposition error, the lowest-singlet excited state of the face-to-face π-stacked cytosine homodimer is revealed to be bound by about half an eV, being the source of an emissive feature consistent with the observed redshifted fluorescence. © *2006 American Institute of Physics*. [DOI: 10.1063/1.2408411]

Determination of the photophysical and photochemical attributes of nucleic acid bases and related biopolymers has been, and remains, a subject of active research.[1–4] Quest of the underlying mechanisms implied in the photoinduced damage to biological material has most probably been one of the major motivations for the outstanding development seen in this area, which may have a great impact in the development of new biotechnologies.[5]

One intriguing aspect of ultraviolet (UV)-irradiated DNA is the appearance of redshifted long-lived emissive states not found in base monomers.[6,7] Whereas the DNA absorption spectra closely resembles that of the building-blocks monomers, apart from the well-documented increase of intensity (hypochromism), the fluorescence spectra of DNA and the constituent nucleotides are qualitatively different.[8] Interestingly, the redshifted fluorescence is observed for both the single- and double-stranded forms of polynucleotides. It is normally denoted in the literature as *excimer fluorescence*, a term first proposed by Eisinger *et al.*,[6] reflecting the relevant role assumed to be played by the corresponding excited dimer (excimer) of the biopolymer. The assignment of the excimer origin of the redshifted fluorescence rests indeed upon the similarity between the emission from polynucleotides and dinucleotides.

The recent time- and wavelength-resolved fluorescence study on different oligonucleotides reported by Plessow *et al.*[7] using 80 picoseconds (ps) excitation pulses makes readily apparent the longer-decay components and redshifted emission that it was assumed to arise from excimer formation. In particular, for the cytosine (C) oligonucleotide 15-mer $d(C)_{15}$ a decay component of several nanoseconds (ns) is clearly observed as compared to the less intense feature of the dimer $d(C)_2$, and to the mononucleotide CMP, the latter showing a short instrument-limited decay. Because of the slow rate of energy relaxation, these long-lived states associated to excimer-like states have been suggested as the precursors of the DNA photolesions, including photodimers.[3,9] On the other hand, Crespo-Hernández *et al.*[10] have recently shown by using femtosecond transient absorption spectroscopy that excimers are formed in high yields in a variety of synthetic DNA oligonucleotides and conclude that excited-state dynamics of A·T DNA is controlled by base stacking.

Despite the existence of excimer- and exciplex-like excited states of nucleobases being invoked widely in experimental literature, as far as we know there is no *ab initio* study supporting it. In this scenario, the performance of accurate (predictive) quantum-chemical computations on the excimers of nucleobases seems timely. We address in this communication the study of the C-excimer in vacuo by using a well-established quantum-chemical *ab initio* method, namely the complete active space self-consistent-field second-order perturbation theory (CASPT2)[11–13] as implemented in the MOLCAS 6.0 software,[14] in conjunction with extended one-electron basis sets.

As a first step toward the characterization of the low-lying singlet excimers of cytosine, the potential energy curves (PECs) with respect to the intermolecular separation (R) of two cytosine molecules kept at the ground-state equilibrium geometry (see Fig. 1) have been built at the CASPT2 level. Unless otherwise stated the one-electron basis set of atomic natural orbital (ANO) type with the primitive set C,N,O(10*s*6*p*3*d*)/H(7*s*3*p*), the ANO-S set,[15] contracted to C,N,O[3*s*2*p*1*d*]/H[2*s*1*p*] was used (hereafter basis set A).

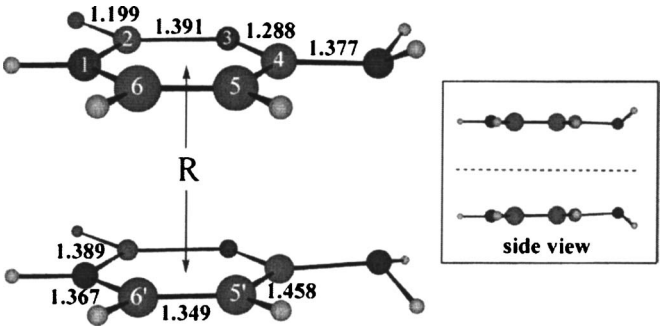

FIG. 1. Labeling for the cytosine dimer. Bond distances (in Å) correspond to the ground-state equilibrium geometry of the monomer computed at the CASSCF(8,7)/ANO-S C,N,O[3*s*2*p*1*d*]/H[2*s*1*p*] level. The homodimer system displays C_s symmetry, being the mirror symmetry plane represented by a dashed line in the side-view inset.

[a)]Author to whom correspondence should be addressed. Electronic mail: Manuela.Merchan@uv.es

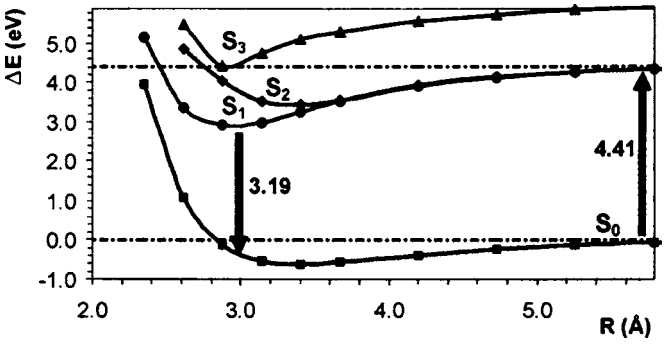

FIG. 2. CASPT2(12,12)/ANO-S C,N,O[3*s*2*p*1*d*]/H[2*s*1*p*] potential energy curves built with respect to the intermolecular distance R(C5–C5′) of two face-to-face π-stacked cytosine molecules.

The optimized structure for ground-state cytosine was determined at the complete active space self-consistent field (CASSCF) level, employing basis set A. The active space comprised eight active π electrons distributed among seven molecular π orbitals (MOs), denoted as CASSCF(8,7). It corresponds to the full π system, except for the deep π MO localized mainly on the nitrogen atom of the NH_2 fragment, which was treated as inactive. The choice is sustained by the fact that when the π-NH_2 MO is allowed to become active, the occupation number of the corresponding CASSCF natural orbital (NO) is close to 2.0. In order to mimic the actual interaction of pyrimidines in DNA in the biologically relevant *cis-syn* stereoisomer, the face-to-face arrangement of the C-dimer was considered.

Although the homodimer displays spatial C_s symmetry, the PECs with respect to R were computed with no symmetry restrictions (C_1 symmetry), since wave function symmetry breaking is a prerequisite to describe correctly the asymptotic limit for the lowest singlet electronic transition of the two moieties.[16] For the computations of the dimer, two additional π MOs were also kept inactive, since the occupation number of the corresponding natural orbitals when they were treated as active was practically 2.0. They correspond to the all-in-phase MOs in each cytosine. The dimer was, therefore, described initially by the CASSCF(12,12) wave function, employing four roots in the averaging procedure. On the other hand, the corresponding results including the second-order corrections shall be labeled as CASPT2(12,12). In order to minimize weakly interacting intruder states, the imaginary level-shift technique, with IMAG=0.2 a.u.,[17] was turned on.

The PEC for the ground state is repulsive at the closed-shell Hartree-Fock and at the CASSCF(12,12) level. Similarly, at the semiempirical level, employing the AM1 parameterization,[9] no stable excimer was found if the two molecules were constrained to be parallel, leading to a small binding energy (3 kcal/mol) upon geometry optimization. The lowest singlet excited state becomes weakly bound at the CASSCF(12,12) level. In contrast, as can be seen in Fig. 2, the CASPT2(12,12) PECs for the ground state (S_0) and the three lowest singlet excited states (S_1, S_2, and S_3) have well-defined minima with binding energies of a few tenths of an eV. In the PECs we have chosen to monitor the intermolecular distance R(C5–C5′), which is particularly relevant in the formation of cyclobutane photodimers. In Fig. 2, energies are referred to two ground-state cytosine molecules separated about 20 a.u. In the asymptotic limit S_1 and S_2 become degenerate. They are related to the equivalent situations C $+C^*$ and C^*+C, where C and C^* represent the ground-state cytosine and its lowest singlet excited state, respectively.

Thus, the absorption $S_0 \rightarrow S_1$ calculated at 20 a.u., corresponds to the monomer absorption. It is here computed at 4.41 eV in agreement with previous findings and experimental evidence.[18–20] The value becomes about half an eV too low if computed in C_s symmetry because localization of the electronic excitation in one monomer is not permitted by symmetry constraints. However, at short intermolecular distances the CASSCF(12,12) wave functions are completely delocalized over the two moieties and the S_1 and S_2 states correlate nicely with the wave functions $1^1A''$ and $2^1A'$ of C_s symmetry, respectively. Therefore, S_1 is described mainly by intermonomer charge-transfer one-electron promotions and S_2 by simultaneous singly excited intra-monomer configurations with leading weights 69% (S_1) and 40% (S_2). Since S_3 is related to a higher excited state of the monomer it shall not be discussed further. On the other hand, the vertical emission from the $(S_1)_{min}$ structure is calculated at 3.19 eV. The agreement with the fluorescence maximum observed in aqueous solution for the dimer $d(C)_2$ and the 15-mer $d(C)_{15}$ (λ_{max}=385 nm;3.22 eV),[7] considerably redshifted as compared to that of the monomer (λ_{max}=313 nm;3.96 eV),[1] is surprising. It is clear that the results might be affected by the basis set superposition error (BSSE) that so far has not been taken into account. In order to analyze in detail these results, the BSSE has been considered for the states of interest (S_0, S_1, and S_2) "*a posteriori*" at the minima of the CASPT2(12,12) PECs, as it is a common practice routinely employed in theoretical computations. The relevant spectroscopic results are listed in Table I.

The binding energy (E_b) has been obtained as

$$E_b(S_n) = E_C + E_{C^*} - E_{CC^*}, \tag{1}$$

with E_C, E_{C^*}, and E_{CC^*} being the total energies of C, C^*, and the dimer CC^*, respectively. The BSSE has been corrected by using the counterpoise correction (CP).[21] Therefore, the CP-E_b is defined as follows:

$$\text{CP-}E_b(S_n) = E_C(C^*) + E_{C^*}(C) - E_{CC^*}, \tag{2}$$

where $E_C(C^*)$ and $E_{C^*}(C)$ represent the energy of C in the presence of the ghost orbitals of C^* and the energy of C^* employing also the orbitals of C, respectively, computed at the CC^* structure considered. Conversely, the CP-BSSE for a given state at a fixed geometry can be seen as the difference

$$\text{CP} - \text{BSSE}(S_n) = E_b(S_n) - (\text{CP} - E_b(S_n)). \tag{3}$$

For the ground state, the C^* in these expressions has, of course, to be replaced by C. The energies for $E_C(C^*)$ and $E_{C^*}(C)$ have been obtained at the CASPT2(6,6) level, where the two-roots state-average CASSCF(6,6) reference wave function comprised the equivalent active MOs as in the respective CASSCF(12,12) computation of the dimer.

The CASPT2(12,12) ground-state binding energy is substantial, 0.62 eV, but the system becomes unbound by

TABLE I. Binding energy (E_b), basis set superposition error (BSSE) obtained through the counterpoise method (CP-BSSE), and the corrected binding energy (CP-E_b), computed at the CASPT2(12,12)/ANO-S C,N,O[3*s*2*p*1*d*]/H[2*s*1*p*] level at the PECs minima of ground and lowest singlet excited states of the cytosine dimer represented in Fig. 1. Distances in Å and energies in eV.

State	R(C5–C5′)	E_b	CP-BSSE	CP-E_b
		At the $(S_0)_{min}$ structure		
S_0	3.416	0.62	0.77	−0.15
		At the $(S_1)_{min}$ structure		
S_1[a]	2.954	1.51	0.97	0.54
S_0	2.954	0.29	0.97	−0.68
		At the $(S_2)_{min}$ structure		
S_2	3.376	0.99	0.74	0.25

[a]The CASPT2(12,12) vertical emission (fluorescence) including the CP-BSSE correction leads to 3.19 eV, as the result of (4.41 eV−0.54 eV−0.68 eV).

−0.15 eV when the BSSE is included, that is, the ground-state dimer at 3.416 Å is 0.15 eV above the sum of two ground-state monomers. The CP-BSSE corrections seem to be large. However, as discussed below, it is just an indication of the diffuseness of the basis set employed, flexibility which is required to treat successfully the excited states of the monomer and the dimer. Table II, which compiles the results for the ground state at the SCFPT2 level (equivalent to MP2 but with IMAG=0.2 a.u.), dramatically makes this point. First, as expected, the CP-E_b at the CASPT2(12,12) and SCFPT2 levels employing basis set A are similar, −0.15 and −0.12 eV, respectively. A similar result, −0.11 eV, is also obtained using the same contraction scheme with the larger primitive set implied in the ANO-L type.[22] The SCFPT2 convergence on CP-E_b is certainly slow. A comparable pattern has been recently reported at the CASPT2 level for the benzene dimer.[23] With the larger basis set (H) the dimer interaction is repulsive by 0.04 eV (0.94 kcal/mol), consistent with previous findings for an equivalent structure (2.45 kcal/mol).[24] By inspection of Table II one can conclude that the CP-E_b result using basis set B is too repulsive, by 0.07 eV, as compared to that obtained with basis set H. Since an analogous performance can be presumed for the excited states as regards the influence of the basis set, we expect the computed CP-E_b to be accurate within ±0.1 eV. On the other hand, the segmented basis set L, much less diffuse but of comparable quality to basis set A, yields a similar corrected result, −0.15 eV, even if the E_b and CP-BSSE values are totally different with respect to the ANOs results.

TABLE II. Convergence pattern with respect to the increase of the one-electron basis set for the binding energy (E_b), basis set superposition error (BSSE) obtained through the counterpoise method (CP-BSSE), and the corrected binding energy (CP-E_b) computed at the SCFPT2 level for the ground state (S_0) of the cytosine dimer at the intermolecular distance R(C5–C5′) =3.416 Å. Energies are given in eV.

Basis set	N[a]	E_b	CP-BSSE	CP-E_b
Generally contracted ANO-S-type scheme[b]				
A:C,N,O[3*s*2*p*1*d*]/H[2*s*1*p*]	274(848)	0.64	0.76	−0.12
Generally contracted ANO-L-type scheme[c]				
B:C,N,O[3*s*2*p*1*d*]/H[2*s*1*p*]	274(1176)	0.67	0.78	−0.11
C:C,N,O[4*s*3*p*2*d*]/H[2*s*1*p*]	418(1176)	0.41	0.47	−0.06
D:C,N,O[4*s*3*p*2*d*]/H[3*s*2*p*]	458(1176)	0.42	0.47	−0.05
E:C,N,O[5*s*4*p*2*d*]/H[2*s*1*p*]	482(1176)	0.30	0.36	−0.06
F:C,N,O[5*s*4*p*2*d*]/H[3*s*2*p*]	522(1176)	0.30	0.35	−0.05
G:C,N,O[5*s*4*p*2*d*1*f*]/H[2*s*1*p*]	594(1512)	0.21	0.26	−0.05
H:C,N,O[4*s*3*p*2*d*1*f*]/H[3*s*2*p*1*d*]	620(1662)	0.35	0.39	−0.04
Additional ANO-L contractions checked[c]				
I:C,N,O[4*s*3*p*1*d*]/H[2*s*1*p*]	338(1176)	0.34	0.44	−0.10
J:C,N,O[5*s*4*p*1*d*]/H[2*s*1*p*]	402(1176)	0.17	0.27	−0.10
K:C,N,O[6*s*5*p*1*d*]/H[2*s*1*p*]	466(1176)	0.12	0.21	−0.09
For comparison, a segmented basis set[d]				
L:cc-pVDZ[3*s*2*p*1*d*/2*s*1*p*]	274(486)	0.05	0.20	−0.15

[a]Number of basis functions (number of primitives).
[b]Primitive set: C,N,O(10*s*6*p*3*d*)/H(7*s*3*p*) (Ref. 15).
[c]Primitive set: C,N,O(14*s*9*p*4*d*3*f*)/H(8*s*4*p*3*d*) (Ref. 22).
[d]Primitive set: C,N,O(9*s*4*p*1*d*)/H(4*s*1*p*) (Ref. 25).

We conclude that at the highest level of theory, with inclusion of the BSSE, both S_1 and S_2 are bound (cf. Table I). According to their nature, the S_2 state displays a minimum at a larger R(C5–C5′) distance than that computed for S_1. Because of the cancellation of BSSE corrections, the vertical emission remains like the direct CASPT2(12,12) result once that the BSSE has been taken into account. For this reason, and since the basis set A already bears enough flexibility to describe the excited states, further improvement of the basis set does not significantly change the computed vertical emission (0.04 eV with the basis set I). One question still remains: Would inclusion of the BSSE change the equilibrium intermolecular distance? In order to answer to this question, the PECs have been recalculated including the BSSE. The results are depicted in Fig. 3, which represents our best esti-

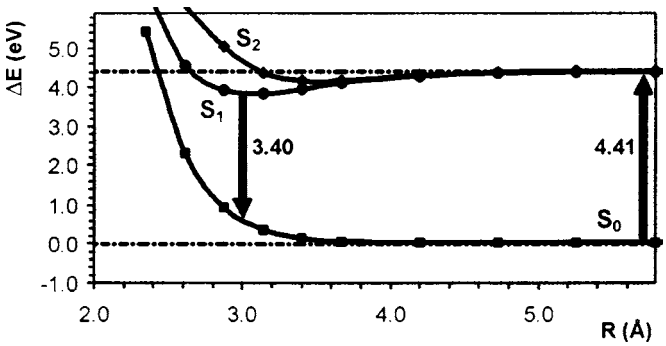

FIG. 3. BSSE-CASPT2(12,12)/ANO-S C,N,O[3*s*2*p*1*d*]/H[2*s*1*p*] potential energy curves built with respect to the intermolecular distance R(C5–C5′) of two face-to-face π-stacked cytosine molecules.

mate for the cytosine dimer. The minimum for S_1 is obtained at $R(C5-C5')=3.076$ Å, with a vertical emission of 3.40 eV, and a binding energy CP-E_b of 0.58 eV. Thus, our best estimate has a difference of 0.18 eV with respect to the emission maximum datum recorded recently.[7] It is worth recalling that the computed vertical transition does not have experimental counterpart, and for a truly correct comparison with experiment, vibrational resolution of the band should be computed in order to determine the band maximum. In this respect, the calculated CASPT2(12,12) result for the vertical emission without (3.19 eV) and with inclusion of the BSSE (3.40 eV) are equally reasonable as compared to the available experimental data for dinucleotides, polynucleotides, and DNA (3.2–3.4 eV).[6,7]

In summary, the main prediction of the present research is the existence of cytosine excimers, which can be regarded as an intrinsic property of the C-dimer. The computed vertical emission supports the excimer origin of the redshifted fluorescence observed in cytosine oligonucleotides.[6,7] Parallel work involving the remaining DNA/RNA nucleobases is currently addressed in our group.

The research reported in this communication has been supported by projects MEC-FEDER CTQ2004-01739 and Generalitat Valenciana GV06-192. G.O.G. and D.R.S. thank MEC-FPU grants.

[1] P. R. Callis, Annu. Rev. Phys. Chem. **34**, 329 (1983).
[2] J. Cadet and P. Vigny, in *Bioorganic Photochemistry*, edited by H. Morrison (Wiley, New York, 1990).
[3] C. E. Crespo-Hernández, B. Cohen, P. M. Hare, and B. Kohler, Chem. Rev. (Washington, D.C.) **104**, 1977 (2004).
[4] L. Serrano-Andrés, M. Merchán, and A. C. Borin, Proc. Natl. Acad. Sci. U.S.A. **103**, 8691 (2006).
[5] P. N. Prasad, *Introduction to Biophotonics* (Wiley & Sons, Inc., New Jersey, 2003).
[6] J. Eisinger, M. Guéron, R. G. Shulman, and T. Yamane, Proc. Natl. Acad. Sci. U.S.A. **55**, 1015 (1966).
[7] R. Plessow, A. Brockhinke, W. Eimer, and K. Kohse-Höinghaus, J. Phys. Chem. B **104**, 3695 (2000).
[8] J. Eisinger and R. G. Shulman, Science **161**, 1311 (1968).
[9] V. I. Danilov, O. N. Slyusarchuk, J. L. Alderfer, J. J. P. Stewart, and P. R. Callis, Photochem. Photobiol. **59**, 125 (1994).
[10] C. E. Crespo-Hernández, B. Cohen, and B. Kohler, Nature (London) **436**, 1141 (2005).
[11] K. Andersson, P.-Å. Malmqvist, and B. O. Roos, J. Chem. Phys. **96**, 1218 (1992).
[12] B. O. Roos, M. P. Fülscher, P.-Å. Malmqvist, L. Serrano-Andrés, K. Pierloot, and M. Merchán, Adv. Chem. Phys. **93**, 219 (1996).
[13] M. Merchán and L. Serrano-Andrés, in *Computational Photochemistry*, edited by M. Olivucci (Elsevier, Amsterdam, 2005).
[14] K. Andersson, M. Barysz, A. Bernhardsson *et al.*, MOLCAS, version 6.0, Department of Theoretical Chemistry, Chemical Centre, University of Lund, Lund, Sweden, 2004.
[15] K. Pierloot, B. Dumez, P.-O. Widmark, and B. O. Roos, Theor. Chim. Acta **90**, 87 (1995).
[16] M. Merchán, R. Pou-Amérigo, and B. O. Roos, Chem. Phys. Lett. **252**, 405 (1996).
[17] N. Forsberg and P.-Å. Malmqvist, Chem. Phys. Lett. **274**, 196 (1997).
[18] M. Merchán and L. Serrano-Andrés, J. Am. Chem. Soc. **125**, 8108 (2003).
[19] M. Merchán, L. Serrano-Andrés, M. A. Robb, and L. Blancafort, J. Am. Chem. Soc. **127**, 1820 (2005).
[20] M. P. Fülscher and B. O. Roos, J. Am. Chem. Soc. **117**, 2089 (1995).
[21] S. F. Boys and F. Bernardi, Mol. Phys. **100**, 65 (2002).
[22] P.-O. Widmark, P.-A. Malmqvist, and B. O. Roos, Theor. Chim. Acta **77**, 291 (1990).
[23] T. Rocha-Rinza, L. De Vico, V. Veryazov, and B. O. Roos, Chem. Phys. Lett. **426**, 268 (2006).
[24] P. Jurecka, J. Sÿponer, and P. Hobza, J. Phys. Chem. B **108**, 5466 (2004).
[25] T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).