Theoretical Insight into the Intrinsic Ultrafast Formation of Cyclobutane Pyrimidine Dimers in UV-Irradiated DNA: Thymine versus Cytosine

Juan José Serrano-Pérez, Israel González-Ramírez, Pedro B. Coto, Manuela Merchán,
Luis Serrano-Andrés*

*Instituto de Ciencia Molecular, Universitat de València, Apartado 22085, ES-46071
Valencia, Spain*

Luis.Serrano@uv.es

Supporting Information

---

* Autor to whom correspondence should be addressed: Luis.Serrano@uv.es

Computational details

a. CASSCF and CASPT2 calculations.

Geometry optimizations of the ground-state T<>T dimer and the delocalized excimer $^1(TT)_{exc}$ have been studied by using the complete-active-space self-consistent-field (CASSCF) method comprising as active space a total of 16 π electrons distributed among 14 π molecular orbitals (MOs). The totally symmetric π MO of each thymine was kept inactive. In order to mimic the actual interaction of pyrimidines in DNA, geometry optimization has been carried out within C_s symmetry constraints, allowing so for an effective and natural interaction of two thymine molecules in the biologically relevant cis-syn stereoisomer. At the optimized geometries, the energies 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 electronic singlet transition of the two moieties. It is said that the calculation breaks symmetry when the computed electronic wave function has lower symmetry than that implied by the nuclear coordinates. For the computations in C_1 symmetry, 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. A CASSCF wave function of 12 active π electrons and 12 active π MOs was therefore employed, hereafter denoted as CASSCF(12,12). In this contribution conical intersection crossings were obtained as minimum energy crossing points (MECPs), obtained by using the restricted Lagrange multipliers technique as included in the MOLCAS-6.0 package[1] in which the lowest-energy point was obtained under the restriction of degeneracy between the two considered states.[2] In addition, an standard arrangement has been employed for the B-form DNA, $(TT)_B$.[3]

Using the C_1 state-average CASSCF(12,12) wave functions for three roots, dynamic electron correlation has been subsequently taken into account perturbatively at the second-order level through the CASPT2 method,[4,5,6] labeled as CASPT2(12,12). In order to minimize weakly interacting intruder states, the imaginary level-shift technique, with IMAG=0.2 au, has been employed.[7] The reported CASPT2 $(S_1/S_0)_{CI}$ conical intersection was obtained by exploring a grid of points along the distortions that involve the smallest values for the CASSCF(12,12) gradients in the region of the crossing.[8] The obtained structure is consistent with previously obtained CASSCF results[9] but having larger intermonomer distances, as it occurs in ethene dimer,[8,10] an effect that is directly related to inclusion here of dynamic correlation.

The basis set of Atomic Natural Orbital (ANO) type with the contraction scheme C,N,O[3s2p1d]/H[2s1p] was used throughout.[11] Basis set superposition error (BSSE) was taken into account by using a modified counterpoise (CP) approach based on localized molecular orbitals, specifically designed for correlated approaches (see next sextion).[12] The results discussed shall be generically denoted as CASPT2(12,12)+BSSE and the corresponding CP-corrected binding energy as CP-E_b.

All the computations have been carried out by using the MOLCAS 6.0 quantum-chemical software.[1]

b. Basis Set Superposition Error.

The inclusion of the basis set superposition error (BSSE) is crucial to accurately describe binding energies.[13] Here the effect was taken into account by using the counterpoise correction (CP).[14]

The binding energy (E_b) was obtained as follows:

$$E_b\,(TT^*) = E_T + E_{T^*} - E_{T^*T} \qquad (1)$$

with E_T, E_{T^*}, being the total energies of the ground (T), and excited state (T*) of thymine, both at the equilibrium geometry of the ground-state thymine, and E_{T^*T} representing the total energy of dimer T*T at a given geometry. The corrected counterpoise binding energy (CP-E_b) comes from the expression:

$$\text{CP-}E_b\,(T^*T) = E_b\,(T^*T) - [\text{CP-BSSE}(T^*T)] \qquad (2)$$

where

$$\text{CP-BSSE (T*T)} = E_{T*}\,(T, R=\infty) - E_{T*}\,(T, R=T*T) + E_T\,(T*, R=\infty) - E_T\,(T*, R=T*T). \qquad (3)$$

In [CP-BSSE (T*T)] the geometry of the monomers is kept to that of the dimer. Thus, the notation E_{T*} (T, R=T*T) indicates the energy of T* computed in the ghost orbitals of T at the geometry of T*T, whereas in E_{T*} (T, R= ∞) the ghost MOs of T are at infinity distance of T*. In this manner the influence of the variation of geometry is accounted for in the BSSE treatment. The findings discussed correspond to CASPT2 results with inclusion of BSSE.

c. Geometries

Table S1 compiles the optimized geometries and their computed CASPT2 energies.

References

---

[1] Andersson, K., Barysz, M., Bernhardsson, A., Blomberg, M. R. A., Carissan, Y., Cooper, D. L., Cossi, M., Fülscher, M. P., Gagliardi, L., de Graaf, C., Hess, B., Hagberg, G., Karlström, G., Lindh, R., Malmqvist, P.-Å. , Nakajima, T., Neogrády, P., Olsen, J., Raab, J., Roos, B. O., Ryde, U., Schimmelpfennig, B., Schütz, M., Seijo, L., Serrano-Andrés, L., Siegbahn, P. E. M., Stålring, J., Thorsteinsson, T., Veryazov, V., Widmark, P. -O. *MOLCAS, version 6.0*; Department of Theoretical Chemistry, Chemical Centre, University of Lund, P.O.B. 124, S-221 00 Lund: Sweden, 2006.

[2] De Vico, L.; Olivucci, M.; Lindh, R. *J. Chem. Theory Comp.* **2005**, *1*, 1029-1037.

[3] Lu, X.-J.; Olson, W. K. *Nucleic Acids Res.*, **2003**, *31*, 5108-5121.

[4] Andersson, K., Malmqvist, P. -Å.; Roos, B. O. *J. Chem. Phys.* **1992,** *96*, 1218-1226.

[5] Roos, B. O., Andersson, K., Fülscher, M. P., Malmqvist, P.-Å., Serrano-Andrés, L., Pierloot, K., Merchán, M. *Adv. Chem. Phys.* **1996**, *93*, 219-331.

[6] Merchán, M.; Serrano-Andrés, L. In *Computational Photochemistry,* ed. Olivucci, M., Elsevier, Amsterdam, 2005.

[7] Forsberg, N.; Malmqvist, P.-Å. *Chem. Phys. Lett.* **1997**, *274*, 196-204.

[8] Roca-Sanjuán, D.; Olaso-González, G.; González-Ramírez, I.; Serrano-Andrés, L. *J. Am. Chem. Soc.* **2008**, in press.

[9] Boggio-Pasqua, M.; Groenhof, G.; Schäfer, L. V.; Grubmüller, H.; Robb, M. A. *J. Am. Chem. Soc.* **2007**, *129*, 10996-10997.

[10] Serrano-Andrés, L.; Merchán,M.; Lindh, R. *J. Chem. Phys.* **2005**, *122*, 104107.

[11] Pierloot, K., Dumez, B., Widmark, P.-O.; Roos, B. O. (1995) *Theor. Chim. Acta* **1995**, *90*, 87-114.

[12] Van Duijneveldt, F. B., Van Duijneveldt-Van de Rijdt, J. G. C. M.; Van Lenthe, J. H. *Chem. Rev.* **2004**, ***94***, 1873-1885.

[13] Olaso-González, G.; Roca-Sanjuán, D.; Serrano-Andrés, L.; Merchán, M. *J. Chem. Phys.* **2006**, *125*, 231102.

[14] Boys, S. F.; Bernardi, F. *Mol. Phys.* **2002**, *100*, 65-73.

Table S1. Cartesian coordinates x, y, z (in Å) of the stationary points optimized for the cytosine-cytosine dimer. CASPT2(12,12) total energy (Et) is also included.

S_0 Thymine CASSCF(8e,7MOs)/ANO-S N,C,O[3s2p1d]/H[2s1p]
C_1 symmetry
Two thymine molecules at about 22 Å, Et: -905.842746 au

N	-0.732356	1.031478	1.499145
N	-1.118134	-1.240959	1.499349
C	-1.663092	0.015776	1.499908
C	0.247451	-1.477747	1.500391
C	1.154751	-0.484077	1.499508
C	2.641891	-0.702205	1.500204
C	0.648709	0.902545	1.499442
O	1.352132	1.876155	1.500610
O	-2.843758	0.219643	1.500798
H	0.529155	-2.517058	1.501808
H	-1.101588	1.961794	1.499721
H	-1.768089	-1.997177	1.498473
H	2.871350	-1.765157	1.494060
H	3.095037	-0.250930	2.380649
H	3.096774	-0.240693	0.626048

S_0 T<>T CASSCF(16e,14MOs)/ANO-S N,C,O[3s2p1d]/H[2s1p]
C_s symmetry ($1^1A'$)
symmetry distinct nuclear coordinates
Et: -905.871904 au

N	-0.663032	1.034929	1.687450
N	-1.060511	-1.250889	1.493713
C	-1.557050	-0.016821	1.808581
C	0.165186	-1.500364	0.797693
C	1.199282	-0.351630	0.818750
C	2.523897	-0.707456	1.493337
C	0.689844	0.949306	1.423407
O	1.409690	1.882208	1.629656
O	-2.676235	0.161052	2.191298
H	0.589352	-2.426619	1.176232
H	-0.991511	1.910444	2.049657
H	-1.757302	-1.967895	1.502049
H	2.913191	-1.650289	1.121174
H	2.366217	-0.807822	2.566969
H	3.259216	0.073234	1.327812

Table S1 (continuation).

^{1}(TT)$_{exc}$ CASSCF(16e,14MOs)/ANO-S N,C,O[3s2p1d]/H[2s1p]
C$_s$ symmetry (1^{1}A")
Symmetry distinct nuclear coordinates. Et: -905.763511 au

N	-0.690402	1.005415	1.678946
N	-1.038640	-1.279440	1.503768
C	-1.589768	-0.014225	1.726182
C	0.255607	-1.509223	1.190208
C	1.213985	-0.445747	1.324791
C	2.653161	-0.755307	1.632312
C	0.714226	0.872395	1.634770
O	1.386410	1.849858	1.854621
O	-2.762095	0.109610	1.924902
H	0.573793	-2.537725	1.277865
H	-1.050281	1.922076	1.863143
H	-1.701012	-2.029394	1.541463
H	2.959491	-1.696647	1.181048
H	2.817442	-0.833152	2.711560
H	3.298194	0.033500	1.258784

(S$_1$/S$_0$)$_{CI}$ CASPT2(12e,12MOs)/ANO-S N,C,O[3s2p1d]/H[2s1p]
C$_1$ symmetry
Et: -787.615430 au

N	-1.586509	-0.762859	1.035520
N	0.601105	1.836724	0.935071
N	-1.574258	-1.085655	-1.268716
N	-0.474168	1.805148	-1.129034
C	-2.303789	-0.761344	-0.136622
C	-0.461589	2.268135	0.160892
C	-0.204711	-1.201709	-1.278956
C	0.443879	0.886483	-1.662139
C	0.448362	-1.590449	-0.009889
C	1.471430	0.376338	-0.788221
C	1.422267	-2.747310	0.033472
C	2.722856	-0.162422	-1.417427
C	-0.328674	-1.359972	1.209969
C	1.611573	0.972227	0.556133
O	0.040338	-1.639826	2.320221
O	2.538846	0.742898	1.281807
O	-3.466261	-0.490408	-0.202444
O	-1.303787	2.995501	0.600121
H	0.208146	-1.604691	-2.185655
H	0.656099	0.994170	-2.716163
H	-2.125853	-0.604634	1.865804
H	0.638885	2.201934	1.866729
H	-2.058402	-0.923889	-2.127567
H	-1.285539	2.066501	-1.651431
H	1.953028	-2.851814	-0.907793
H	3.333424	0.685195	-1.736233
H	0.902850	-3.688475	0.233672
H	2.498767	-0.762334	-2.297095
H	2.147820	-2.596394	0.829002
H	3.292271	-0.746922	-0.708417