## Supporting information for "Femtosecond stimulated Raman spectroscopy of the cyclobutane thymine dimer repair mechanism: A computational study"

Hideo Ando,<sup>†</sup> Benjamin P. Fingerhut,<sup>‡</sup> Konstantin E. Dorfman,<sup>†</sup> Jason D. Biggs,<sup>†</sup> and Shaul Mukamel<sup>\*,†</sup>

<sup>†</sup>Department of Chemistry, 1102 Natural Sciences II, University of California, Irvine, CA 92697-2025, USA

<sup>‡</sup>Max-Born-Institut für Nichtlineare Optik und Kurzzeitspektroskopie, Max Born Strasse 2A, 12489 Berlin-Adlershof, Germany

Email: smukamel@uci.edu

Phone: +1 949 824 7600. Fax: +1 949 824 4759

Table S1: Potential energies (kcal/mol) of optimized geometries in the TT<sup>•-</sup>. The geometries were obtained with the DFT(M05-2X) method, and their energies were also calculated with the CCSD and the CCSD(T) method. The 6-31G\* basis set was employed for all methods. The energies for the closed forms are set to be zero.

	DFT(M05-2X)	CCSD	$\operatorname{CCSD}(T)$
closed	0.0	0.0	0.0
$TS1^{\bullet-}$	0.7	2.2	1.8
$INT1^{\bullet-}$	-10.5	-10.1	-9.3
$TS2^{\bullet-}$	-9.6	-10.5	-8.4
$INT2^{\bullet-}$	-17.8	-16.0	-15.8
$TS3^{\bullet-}$	-12.4	-11.1	-10.7
open <sup>a</sup>	-10.9	-12.5	-11.1

<sup>a</sup> For the open form, partial optimization was done with three constraints,  $C_5-C'_5 = 4.18$  Å,  $C_6-C'_6 = 4.46$  Å, and  $C_7-C_5-C'_5-C'_7 = 35.3^{\circ}$ . These constraints were used in Ref. 50.

## The synchronous concerted mechanism

We study cyclobutane thymine dimer (i.e., TT). Geometry optimization was performed with the unrestricted DFT(M05-2X) method and the 6-31G\* basis set. While full optimization was carried out for the closed (dimerized) form, partial optimization was done for the open form with three constraints,  $C_5-C'_5 = 4.18$  Å,  $C_6-C'_6 =$ 4.46 Å, and  $C_7-C_5-C'_5-C'_7 = 35.3^\circ$ . These constraints were also used in Ref. 50. Frequency analysis clarified that the optimized geometries have no imaginary mode in the (constrained) coordinate space. Also in the radical anion system (i.e.,  $TT^{\bullet-}$ ), partial optimization was performed in a similar fashion. Three constraints with the same geometric parameters of TT were used both for closed and open form optimizations. Note that the partially-optimized geometry in the closed form is close to the fully-optimized one; the difference in geometry and in energy is only 0.42 Å and 0.1 kcal/mol.

To study the synchronous concerted mechanism, a relevant one-dimensional coordinate was constructed that directly connects the closed and the open form. The three constraints used above were linearly interpolated simultaneously and the other degrees of freedom were fully relaxed both for the TT and for the  $TT^{\bullet-}$ . The groundstate potential energy curves and their vibrational frequencies were evaluated along the reaction coordinates. The resulting potential energy curves are shown in Figure S1. The barrier at the  $3^{\rm rd}$  geometry from the left along the coordinate is higher in the TT (56.7 kcal/mol) than in the TT<sup>•-</sup> (20.4 kcal/mol). Even in the TT<sup>•-</sup>, the barrier height is much higher than those evaluated in the IRC path, as seen in Figure 2. This indicates that the synchronous concerted mechanism does not apply in this case; however, the results are useful to understand fundamental characteristics of frequency variations.

Frequencies of a free thymine and those of the dimers along the coordinate are given in Table S2 and Figure S2. At the open form of the TT, the  $67^{\text{th}}$  and  $68^{\text{th}}$ modes are  $C_5=C_6$  stretch vibrations, the  $69^{\text{th}}$  and  $70^{\text{th}}$  modes are *anti-symmetric* and *symmetric*  $C_4=O$  stretch vibrations, and the  $71^{\text{st}}$  and  $72^{\text{nd}}$  modes are *anti-symmetric symmetric* and *symmetric*  $C_2=O$  stretch vibrations. At the open form of the  $TT^{\bullet-}$ , the  $68^{\text{th}}-72^{\text{th}}$  modes are certain mixtures of the  $C_5=C_6$ , the  $C_4=O$ , and the  $C_2=O$ stretch vibrations. The  $67^{\text{th}}$  mode does not involve C=C or C=O stretch vibrations because the SOMO is localized in the upper thymine base, as seen in the last panel of Figure S9.

As shown in Figure S2a, the frequencies of the C=O stretch modes get lower along the coordinate. In particular, at the 4<sup>th</sup> geometry from the left, which is close to the TS1 geometry, the frequencies of the 69<sup>th</sup> and 70<sup>th</sup> C<sub>4</sub>=O modes get notably lower. Besides, the 67<sup>th</sup> and 68<sup>th</sup> modes change into  $C_5=C_6$  stretch vibrations due to breaking of the intradimer  $C_5-C'_5$  and  $C_6-C'_6$  bonds, and their frequencies get to lie in a range of 1750–1800 cm<sup>-1</sup>. The mode characters are shown also by  $6\times 6$  sub-block of Duschinsky rotation matrices in Figure S3; the characters of the  $67^{\text{th}}$  and  $68^{\text{th}}$  modes change dramatically between the  $3^{\text{rd}}$  and the  $4^{\text{th}}$  geometry (i.e., small matrix elements in the corresponding  $2\times 2$  sub-block), while those of the  $69^{\text{th}}$  and  $72^{\text{th}}$  modes retain (i.e., almost diagonal  $4\times 4$  sub-block) along the whole coordinate.

In the  $TT^{\bullet-}$ , the frequency changes are more complicated due to an excess electron, as shown in Figure S2b. There are four normal modes in the 1600–1900 cm<sup>-1</sup> frequency range from the 1<sup>st</sup> geometry to the 3<sup>rd</sup> one, six modes from the 4<sup>th</sup> geometry to the 7<sup>th</sup> one, and five modes from the 8<sup>th</sup> geometry to the 12<sup>th</sup> one. This can be understood from Duschinsky rotation matrices and SOMO (or an excess electron) distribution, shown in Figures S4 and S5. In the closed form (i.e., the 1<sup>st</sup> geometry), saturated  $C_5-C'_5$  and  $C_6-C'_6$  bonds are formed; on the other hand, the  $C_5=C_6$  double bonds are broken (see doubly-occupied LUMO in Figure S5) and hence, there are only four C=O stretch modes in the 1600–1900 cm<sup>-1</sup> frequency range. One of the C<sub>4</sub>=O modes has a low frequency at 1615 cm<sup>-1</sup> because of the notable anti-bonding  $\pi^*(C_4-O)$  character of the SOMO (see the left end panel in Figure S5). Between the 3<sup>rd</sup> and the 4<sup>th</sup> geometry, the Duschinsky sub-block of the 67<sup>th</sup> and 68<sup>th</sup> modes has relatively small matrix elements. After the 4<sup>th</sup> geometry, the two normal modes

change into  $C_5=C_6$  stretching and a total of six modes lie in 1600–1900 cm<sup>-1</sup>. Note that the  $C_5=C_6$  stretching frequencies are lower than the common values (e.g., 1774 cm<sup>-1</sup> in Table S2) because of the single electron occupation in delocalized LUMO, shown in Figure S5. Between the 7<sup>th</sup> geometry and the 8<sup>th</sup> one, the 6×6 Duschinsky sub-block has large off-diagonal elements. After the 8<sup>th</sup> geometry, one  $C_5=C_6$  stretch mode disappears and a total of five modes are found in 1600–1900 cm<sup>-1</sup>. This is because the excess electron occupies a LUMO localized in one thymine base, as seen in Figure S5. The above-mentioned characteristics are simpler but similar to those in the IRC path, mentioned in the text.



Figure S1: Potential energy curves of TT (black) and  $TT^{\bullet-}$  (red) for synchronous concerted dissociation. The three constraints,  $(C_5-C'_5, C_6-C'_6, C_7-C_5-C'_5-C'_7)$ , were linearly interpolated simultaneously between (4.18 Å, 4.46 Å, 35.3°) and (1.57 Å, 1.56 Å, 30.4°), while the other degrees of freedom were fully relaxed. The resulting energies are plotted along a mass-weighted coordinate, where the distance between two neighboring points is defined in a *xyz* frame so as to satisfy the Eckart condition. The energies and the coordinates at the closed forms are set to be zero.



Figure S2: Frequencies of C=C and C=O stretch modes of (a) TT and (b)  $TT^{\bullet-}$  along the synchronous dissociation coordinate. In TT, the  $67^{\text{th}}$  and  $68^{\text{th}}$  modes (black and brown lines) correspond to C=C stretch and the  $69^{\text{th}}-72^{\text{th}}$  modes (the other lines) correspond to C=O stretch. In  $TT^{\bullet-}$ , the six modes are certain mixtures of the C=C and C=O stretch modes. In a lower frequency region, not depicted here, the  $67^{\text{th}}$  or the  $68^{\text{th}}$  mode significantly involves or completely changes into other bending vibrations (e.g., C–H bending of the methyl groups).



Figure S3: Sub-block of Duschinsky rotation matrices between neighboring geometries along the synchronous concerted dissociation coordinate for TT. In the last panel (between the 11<sup>th</sup> and the 12<sup>th</sup> geometry), the 67<sup>th</sup> and 68<sup>th</sup> modes are  $C_5=C_6$  stretching, the 69<sup>th</sup> and 70<sup>th</sup> modes are  $C_4=O$  stretching, and the 71<sup>st</sup> and 72<sup>nd</sup> modes are  $C_2=O$  stretching. The matrix elements are depicted in absolute value from 0 (white) to 1 (black).



Figure S4: Sub-block of Duschinsky rotation matrices between neighboring geometries along the synchronous concerted dissociation coordinate for  $TT^{\bullet-}$ . In the last panel (between the 11<sup>th</sup> and the 12<sup>th</sup> geometry), the 67<sup>th</sup> mode is C–H bending and the 68<sup>th</sup>-72<sup>nd</sup> modes are certain mixtures of one C=C stretch and four C=O stretch vibrations. The matrix elements are depicted in absolute value from 0 (white) to 1 (black).



Figure S5: Singly-occupied natural orbitals and schematic orbital energy diagrams of  $TT^{\bullet-}$ . From left to right, the orbitals at the 1<sup>st</sup>, 2<sup>nd</sup>, 4<sup>th</sup>, and 8<sup>th</sup> points on the synchronous concerted dissociation coordinate (i.e., 0.0, 7.7, 15.7, and 30.2 Bohr amu<sup>1/2</sup> in Figure S1) are shown. The orbital at the 2<sup>nd</sup> point is close to the HOMO of a free thymine, while those at the 4<sup>th</sup> and 8<sup>th</sup> points are the LUMO.

Mode <sup>a</sup>	Neutral	Radical anion
$C_5 = C_6$	1774	
$C_4=O$ $C_2=O$	$1851 \\ 1891$	1761 (with notable mixing of $C_4=C_5$ stretching) 1806

Table S2: Frequencies  $(\rm cm^{-1})$  of C=C and C=O stretch modes of a neutral and a radical anion thymine.

<sup>a</sup> The  $C_4=O$  is next to the methyl group, while the  $C_2=O$  is opposite to the group.



Figure S6: Molecular properties of the TT (black lines) and the  $TT^{\bullet-}$  (red lines) along the IRC paths: (a) the  $C_5-C'_5$  and the  $C_6-C'_6$  bond lengths (solid and broken, respectively) and (b) the dihedral angle  $C_7-C_5-C'_5-C'_7$  and that between thymine planes (solid and broken). The thymine plane was defined by using  $C_5$  and two N atoms. Filled circles represent the optimized geometries.



Figure S7: Raman activities of the C=C and C=O stretch vibrations, whose frequencies are between 1580–1900 cm<sup>-1</sup>, in TT<sup>•-</sup> along the IRC path: the 67<sup>th</sup> (black), 68<sup>th</sup> (brown), 69<sup>th</sup> (cyan), 70<sup>th</sup> (blue), 71<sup>st</sup> (pink), 72<sup>nd</sup> (red) modes. Filled circles represent the optimized geometries.



Figure S8: Singly-occupied natural orbitals of the  $TT^{\bullet-}$ . Same as Figure 4 but in side view. From upper left to lower right, the orbitals at the closed form,  $TS1^{\bullet-}$ ,  $INT1^{\bullet-}$ ,  $TS2^{\bullet-}$ ,  $INT2^{\bullet-}$ ,  $TS3^{\bullet-}$ , and open form are shown.



















Figure S9: Singly-occupied natural orbitals of  $TT^{\bullet-}$ . From upper left to lower right, the orbitals at the 1<sup>st</sup> to the 13<sup>th</sup> geometry are shown. See Figure S10 for each geometry number.



Figure S10: Frequencies of the  $67^{\text{th}}-72^{\text{th}}$  normal modes, which involve C=C and C=O stretch vibrations, of  $\text{TT}^{\bullet-}$  along the IRC path. The boundary of regions I and II lies between the  $10^{\text{th}}$  and the  $11^{\text{th}}$  geometry, and that of regions II and III lies between the  $11^{\text{th}}$  and the  $12^{\text{th}}$  geometry. The geometry numbers from  $1^{\text{st}}$  to  $13^{\text{th}}$  are referred to in Figures S9 and S11. Filled circles represent the optimized geometries, while open circles do not.



Figure S11: Sub-block of Duschinsky rotation matrices between neighboring geometries along the IRC path for  $TT^{\bullet-}$ . See Figure S10 for each geometry number. The matrix elements are depicted in absolute value from 0 (white) to 1 (black).



Figure S12: Frequencies of two C=C stretch modes (black and brown lines, the  $67^{\text{th}}$  and  $68^{\text{th}}$  modes) and four C=O ones (the other lines, the  $69^{\text{th}}-72^{\text{th}}$  modes) of TT along the IRC path.



Figure S13: Molecular properties of TpT<sup>•-</sup> along the IRC path: (a) the  $C_5-C'_5$  and the  $C_6-C'_6$  bond lengths (solid and broken lines), (b) the dihedral angle of  $C_7-C_5-C'_5-C'_7$  and that between thymine planes (solid and broken lines), and (c) frequencies of the  $152^{th}-157^{th}$  normal modes, which include C-C/C=C and C=O stretch vibrations. The thymine plane was defined by using  $C_5$  and two N atoms. Filled circles represent the optimized geometries.



Figure S14: Singly-occupied natural orbitals of  $TpT^{\bullet-}$ . From upper left to lower right, the orbitals at the closed form,  $TS1^{\bullet-}$ ,  $INT1^{\bullet-}$ ,  $TS2^{\bullet-}$ ,  $INT2^{\bullet-}$ ,  $TS3^{\bullet-}$ , and open form are shown.



Figure S15: Variation of FSRS signals of TpT<sup>•-</sup> with different delay times T. Two different trajectories were employed: (a) trajectory A with  $\tau_0 = 500$  fs and (b) trajectory B with  $\tau_0 = 100$  ps. The red dots mark instantaneous frequencies. The stick spectra on the horizontal bottom (top) axis represent the initial (final) frequencies.

## FSRS signals for the stochastic Liouville equation

The FSRS signals of intradimer bond splitting of cyclobutane thymine dimer in a small DNA model (TpT<sup>•-</sup>) are obtained by using the stochastic Liouville equation (Eq. 7).<sup>64,65</sup> The coherent part  $\hat{\mathcal{L}}_S = -(i/\hbar) [H_S, \ldots]$  vanishes for the  $|aa\rangle\rangle$  and  $|cc\rangle\rangle$ blocks,  $[\hat{\mathcal{L}}_S]_{aa,aa} = [\hat{\mathcal{L}}_S]_{cc,cc} = 0$ . The remaining blocks of  $\hat{\mathcal{L}}_S$  read

$$[\hat{\mathcal{L}}_S]_{ac,ac} = i \begin{pmatrix} \omega_{ca,closed} & 0 & 0 & 0\\ 0 & \omega_{ca,INT1} & 0 & 0\\ 0 & 0 & \omega_{ca,INT2} & 0\\ 0 & 0 & 0 & \omega_{ca,open} \end{pmatrix}.$$
 (S1)

The two Liouville space Green's functions (i.e., the solution of Eq. 7) are thus given

by

$$\mathcal{G}_{aa,aa}(t) = -\frac{i}{\hbar}\theta(t)\exp\left[[\hat{L}_S]t\right]$$
  
=  $-\frac{i}{\hbar}\theta(t)U\exp\left[[\hat{L}_S]^{\text{diag}}t\right]U^{-1},$  (S2)

$$\mathcal{G}_{ac,ac}(t) = -\frac{i}{\hbar}\theta(t)\exp\left[([\hat{L}_S] + [\hat{\mathcal{L}}_S]_{ac,ac})t\right]$$
  
$$= -\frac{i}{\hbar}\theta(t)V\exp\left[[\hat{\mathcal{L}}]_{ac,ac}^{\text{diag}}t\right]V^{-1},$$
(S3)

where U and V are transformation matrices, which diagonalize the matrices in the exponents.

The time domain FSRS signal on the Stokes side is given by

$$S_{FSRS}(\omega,T) = \Im \int_{-\infty}^{\infty} \frac{d\Delta}{2\pi} \mathcal{E}_3^*(\omega) \mathcal{E}_3(\omega+\Delta) \tilde{S}_{FSRS}^{(i)}(\omega,T;\Delta), \qquad (S4)$$

where  $\tilde{S}^{(i)}_{FSRS}(\omega,T;\Delta)$  can be recast in Liouville space as follows:

$$\tilde{S}_{FSRS}^{(i)}(\omega,T;\Delta) = \frac{2}{\hbar} \int_{-\infty}^{\infty} dt \int_{-\infty}^{t} d\tau_3 |\mathcal{E}_1|^2 |\mathcal{E}_2|^2 e^{-i\Delta(\tau_3 - T)}$$

$$\times e^{i(\omega - \omega_2)(t - \tau_3)} \mathcal{F}(t - \tau_3, \tau_3),$$
(S5)

where by using the Green's functions in Eqs. S2 and S3, the matter correlation function

$$\begin{aligned} \mathcal{F}(t_1, t_2) &\text{ is given by} \\ \mathcal{F}(t_1, t_2) &= -\frac{i}{\hbar} \sum_{a,c} \alpha_{ac}^2 |V_{ag}|^2 \left\langle \langle I | \mathcal{G}_{ac,ac}(t_1) \mathcal{G}_{aa,aa}(t_2) | \rho_0 \right\rangle \right\rangle_S \\ &= -\left(\frac{i}{\hbar}\right)^3 \sum_{a,c} \alpha_{ac}^2 |V_{ag}|^2 \theta(t_1) \theta(t_2) e^{-\gamma_a(t_1+2t_2)} \\ &\times (1, 1, 1, 1) V \exp\left[ [\hat{\mathcal{L}}]_{ac,ac}^{\text{diag}} t_1 \right] V^{-1} U \exp\left[ [\hat{\mathcal{L}}_S]^{\text{diag}} t_2 \right] U^{-1} \begin{pmatrix} 1\\ 0\\ 0\\ 0 \end{pmatrix}. \end{aligned}$$
(S6)

Here, the initial state is the closed form,

$$|\rho_0\rangle\rangle_S = |aa\rangle\rangle \begin{pmatrix} 1\\0\\0\\0 \end{pmatrix}, \tag{S7}$$

and we traced over the final state  $\langle \langle I | = (1, 1, 1, 1)$ Tr where Tr =  $\langle \langle aa | + \langle \langle cc |$ . Vibrational dephasing terms have been added in Eq. S6;  $e^{-\gamma_a t}$  is added to  $\mathcal{G}_{ac,ac}$  and  $e^{-2\gamma_a t}$ to  $\mathcal{G}_{aa,aa}$ , where the time  $\gamma_a^{-1}$  is 532 fs. We set the prefactor  $|\mathcal{E}_1|^2 |\mathcal{E}_2|^2 \alpha_{ac}^2 |V_{ag}|^2 \sigma^2 = 1$ , and then performed analytical integrations to get  $S_{FSRS}(\omega, T)$ .

Evaluating time integrals in Eq. S5 we obtain

$$\tilde{S}_{FSRS}^{(i)}(\omega,T;\Delta) = \frac{-2i}{\hbar^2} |\mathcal{E}_1|^2 |\mathcal{E}_2|^2 \sum_{a,c} \alpha_{ac}^2 |V_{ag}|^2 e^{i\Delta T} \langle \langle I | \mathcal{G}_{ac,ac}(\omega-\omega_2) \mathcal{G}_{aa,aa}(-\Delta) | \rho_0 \rangle \rangle_S.$$
(S8)

Following Eq. 6 we introduce the species-dependent population of the state a after interaction with actinic pulse:

$$\rho_{aa}^{(s)}(t) = |\mathcal{E}_1|^2 |V_{ag}|^2 \mathcal{G}_{aa,aa}(t) |\rho_0\rangle\rangle_S.$$
(S9)

Substituting Eq. S9 into the signal S8 and S4 we obtain Eq. 9.



Figure S16: Same as Figure 8 but for longer delay times T. (a) The stochastic Liouville equation (Eq. 9) and (b) static average (Eq. 10). The signals at T = 2 fs are highlighted in red. After T = 100 fs, time intervals are 500 fs from T = 500 fs up to 5 ps, 1 ps up to 10 ps, and 5 ps later. The stick spectra on the horizontal bottom (top) axis represent the initial (final) frequencies.

## Sequential photoreactivation kinetics and repair quantum yeild of $\mathrm{Tp}\mathrm{T}^{\bullet-}$

We extended the kinetic analysis of photoreactivation in the Supporting information of Ref. 6 to consider the linear sequence of intradimer bond splittings with back reactions (see Eq. 4 in text). Upon excitation of the enzyme-substrate complex at time zero, the time-dependent concentrations of every species were solved with the following rate equations and initial conditions:



$$\frac{d[(\mathbf{T}\ \mathbf{T})]}{dt} = k_{\mathrm{ER}} \left[ (\mathbf{T}^{\bullet-}\ \mathbf{T}) \right]$$
(S15)

$$\left[\mathrm{FADH}^{-*}\right]_{t=0} = n_0 \tag{S16}$$

$$\left[ (\mathbf{T}=\mathbf{T})^{\bullet-} \right]_{t=0} = \left[ (\mathbf{T}=\mathbf{T})^{\bullet-} \right]_{t=0} = \left[ (\mathbf{T}-\mathbf{T})^{\bullet-} \right]_{t=0} = \left[ (\mathbf{T}=\mathbf{T}) \right]_{t=0} = \left[ (\mathbf{T}=\mathbf{T}) \right]_{t=0} = 0.0$$
(S17)

The reciprocals of rate constants of forward electron transfer  $(k_{\text{FET}})$ , lifetime emission  $(k_{\text{LT}})$ , back electron transfer  $(k_{\text{BET}})$ , and electron return  $(k_{\text{ER}})$ , employed here, are 0.25, 1.41667, 2.4, and 0.7 ns, respectively.<sup>6</sup> The reciprocal of  $k_{\text{LT}}$ , close to 1.3 ns in Ref. 15, was determined from the forward electron transfer yield of 0.85.<sup>6</sup> For intradimer bond splittings,  $k_{\pm 1}$ ,  $k_{\pm 2}$ , and  $k_{\pm 3}$  are the same values as those obtained from the transition state theory (Table S3), namely  $(k_1^{-1}, k_2^{-1}, k_3^{-1}, k_{-1}^{-1}, k_{-2}^{-1}, k_{-3}^{-1})=(9.539\times10^{-2}, 1.391, 14.09, 3.959\times10^9, 18.23, 1.590\times10^{13})$  in ps. The above rate equations were treated in a similar way to Eq. 6, and the resulting time-dependent concentrations are shown in Figure 7b. The total repair quantum yield of photoreactivation,

$$\Phi_{\rm repair} = \frac{\lim_{t \to \infty} \left[ (T \ T) \right]}{n_0}, \tag{S18}$$

was calculated at 0.844.

Table S3: Rate constants  $(s^{-1})$  calculated with the transition state theory at room temperature.

$k_1$	$k_2$	$k_3$	$k_{-1}$	$k_{-2}$	$k_{-3}$
$1.05\times 10^{13}$	$7.19\times10^{11}$	$7.10\times10^{10}$	$2.53\times 10^2$	$5.49\times10^{10}$	$6.29\times10^2$