HOKKAIDO UNIVERSITY

Title	Photophysics of cytosine tautomers : new insights into the nonradiative decay mechanisms from MS-CASPT2 potential energy calculations and excited-state molecular dynamics simulations
Author(s)	Nakayama, Akira; Harabuchi, Yu; Yamazaki, Shohei et al.
Citation	Physical chemistry chemical physics, 15(29), 12322-12339 https://doi.org/10.1039/c3cp51617b
Issue Date	2013-08-07
Doc URL	https://hdl.handle.net/2115/56717
Type	journal article
File Information	PCCP_15_12322.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Photophysics of Cytosine Tautomers: New Insights into the Nonradiative Decay Mechanisms from MS-CASPT2 Potential Energy Calculations and Excited-State Molecular Dynamics Simulations

Akira Nakayama*, Yu Harabuchi, Shohei Yamazaki†, and Tetsuya Taketsugu

Department of Chemistry, Faculty of Science, Hokkaido University
Sapporo 060-0810, Japan

*To whom correspondence should be addressed. Electronic mail: akira-n@sci.hokudai.ac.jp
†Present address: Department of Frontier Materials Chemistry, Graduate School of Science and Technology, Hirosaki University, Hirosaki 036-8561, Japan

Keywords

cytosine, MS-CASPT2, cytosine tautomers, nonradiative decay, excited-state molecular dynamics

Abstract

A comprehensive picture of the ultrafast nonradiative decay mechanisms of three cytosine tautomers (amino-keto, imino-keto, and amino-enol forms) is revealed by high-level *ab initio* potential energy calculations using the multistate (MS) CASPT2 method and also by on-the-fly excited-state molecular dynamics simulations employing the CASSCF method. To obtain a reliable potential energy profile along the deactivation pathways, the MS-CASPT2 method is employed even for the optimization of minimum energy structures in the excited state and conical intersection (CI) structures between the ground and excited states. In the imino (imino-keto) form, we locate a new CI structure involving the twisting of the imino group, and the decay pathway leading to this CI is found to be barrierless, suggesting a remarkably efficient deactivation of imino cytosine. In the keto (amino-keto) form, the MS-CASPT2 calculations exhibit an efficient decay path to the ethylene-like CI involving the twisting of C-C double bond in the six-membered ring, with a barrier of ~0.08 eV from the minimum of the $^1\pi\pi^*$ state. In the enol (amino-enol) form, three types of CIs are identified for the first time. Among them, the ethylene-like CI with a similar molecular structure to the keto form provides the most preferred deactivation pathway of enol cytosine. This pathway exhibits a higher barrier of ~0.22 eV and a higher energy of CI than those of keto cytosine. Nonadiabatic molecular dynamics simulations provide a time-dependent picture of the deactivation processes, including the excited-state lifetime of each tautomer. In particular, the decay time of the imino tautomer is predicted to be only ~100 fs. Our computational results are in remarkably good agreement with the experimental findings in recent femtosecond pump-probe photoionization spectroscopy [J. Am. Chem. Soc. **131**, 16939 (2009); J. Phys. Chem. A **115**, 8406 (2011)], supporting the coexistence of more than one tautomer in the photophysics of isolated cytosine and that each tautomer exhibits a different excited-state lifetime.

I. Introduction

A detailed understanding of photochemical properties of nucleic acid bases is fundamental due to its implications for the photostability and photodamage in DNA/RNA, and numerous efforts have been made to elucidate the underlying mechanisms of these processes at the molecular level. The DNA bases are likely to possess intrinsic mechanisms of extremely efficient nonradiative deactivation, and this is an essential factor for photostability against UV radiation.[1,2] In experiments, this ultrafast deactivation process is supported by very short excited-state lifetime in the subpicosecond and picosecond range, which is established by time-resolved spectroscopic studies of DNA bases in various environments, such as the gas phase, solution, and fully assembled DNA strands.[1,2] On the theory side, conical intersections (CIs) between potential energy surfaces of the ground and first singlet excited (S_0 and S_1) states have been proposed as a funnel for the particularly efficient nonradiative decay.[3]

The photophysics of isolated cytosine has been the subject of extensive experimental studies with time-resolved spectroscopy.[4-8] Cytosine in the gas phase is of special interest among the nucleic acid bases in various environments, because several tautomeric forms are likely to coexist. Experimental[9-17] and theoretical[11-13,18] studies have revealed the coexistence of at least three lowest-energy tautomers of isolated cytosine: keto (amino-keto), imino (imino-keto), and enol (amino-enol) forms. Figure 1 shows the molecular structure of these tautomers. The theoretical studies also suggest that the enol form is energetically the most stable tautomer in the gas phase and that the keto form, which is the canonical tautomer found in DNA, is the second most stable. Because of this, the enol form is found to be the dominant tautomer in experiments of isolated cytosine. For example, Bazsó *et al.*[16] estimated the population ratio of keto:imino:enol as 0.22:0.08:0.70 based on the matrix-isolation IR spectrum. They also reported that more than one tautomer should be taken into account in the spectral simulations in order to qualitatively reproduce the experimental UV spectrum. These findings are in contrast to cytosine in aqueous solution, where the canonical keto form is calculated to be much more stable than other tautomers[12,19] and only this form is likely to contribute to the experimental spectrum.[20,21]

Recently, Kosma *et al.*[7] observed appreciable dependence of decay profiles on the excitation wavelength in the femtosecond time-resolved pump-probe ionization spectroscopy of isolated cytosine, and attributed this remarkable observation to the coexistence of the keto, imino, and enol tautomers. With the excitation wavelengths of 290, 280, 270, and 267 nm, the observed transients were decomposed into three components with lifetimes of $\tau_1 < 0.25$ ps, $\tau_2 = 1.1 - 2.3$ ps, and $\tau_3 \geq 19$ ps. With the help of quantum chemical calculations of excitation energies by Tomić *et al.*,[22] the transients with excitation at the former two wavelengths were assigned to the keto tautomer, while those at the latter two were assigned to a mixture of the keto and enol or imino tautomers. With an excitation at 260 nm, on the other hand, only two transients were identified with lifetime constants of $\tau_1 = 0.12$ ps and $\tau_2 = 3.8$ ps. The dominant component for this wavelength was determined as the enol form, with a minor contribution of the imino form. The authors also proposed that the long-time constant of τ_3, varied from 19 to more than 150 ps, corresponds to a long-lived state of the imino form populated by the excited-state tautomerization from the keto form.

More recently, Ho *et al.*[8] further clarified the dependence of decay profiles on the excitation wavelength (260 to 300 nm) in the femtosecond pump-probe photoionization spectra of cytosine, 1-methylcytosine, and 5-fluorocytosine in the gas phase, where these molecules have different population ratio of their tautomers.[8] After detailed analysis, the authors concluded that the long time constant of 3 to 45 ps observed at excitation wavelengths between 260 and 285 nm corresponds to the decay from the enol tautomer and that the subpicosecond time constant for excitation wavelengths between 260 and 300 nm is assigned to the decay from the keto and imino tautomers. For the enol tautomer, they estimated the barrier height for the decay process to be ~0.15 eV from the minimum of

the S_1 state based on the RRKM (Rice-Ramsperger-Kassel-Marcus) model. They observed a certain consistency with the experiments by Kosma *et al.*,[7] but also reported a discrepancy that the long-time component corresponding to τ_3 up to $\geq$ 150 ps was not detected.

These experimental studies by Kosma *et al.*[7] and Ho *et al.*[8] motivate us to theoretically investigate the excited-state potential energy surfaces of the keto, imino, and enol tautomers of cytosine on the equal footing and scrutinize the proposed scenarios on the multiple components of the observed spectra. Most of the previous computational studies have focused on the keto form of cytosine and its derivatives such as 5-fluorocytosine because of its biological significance.[22-38] For the keto tautomer, three types of CIs between the S_0 and S_1 states are well defined, denoted as $(^1\pi\pi^*/gs)_{CI}$, $(^1\pi_{N3}\pi^*/gs)_{CI}$, and $(^1n_O\pi^*/gs)_{CI}$ in this work (details of these CIs are explained in Section III). Potential energy profiles for the decay pathways connecting the Franck-Condon (FC) region to these CIs have been investigated at the various levels of theory. On-the-fly excited-state dynamical simulations including the nonadiabatic transitions have been also performed in recent years to elucidate the deactivation mechanisms from a dynamical point of view.[34-38]

However, theoretical studies on the imino and enol tautomers of cytosine are still very scarce. The vertical excitation energies of cytosine tautomers have been calculated by several groups[16,22,39,40], but to the best of our knowledge only one literature is available which explored the excited-state potential energy profiles of the imino and enol tautomers of cytosine.[22] In the work by Tomić *et al.*,[22] the role of so-called the ethylene-like CI between the ground and $^1\pi\pi^*$ states, which involves the twisting of the C-C double bond in the six-membered ring was investigated for the keto and imino tautomers, but the other types of CIs were not examined. For the enol form, in particular, they proposed that low-lying vibrational levels in the $^1\pi\pi^*$ state decay by normal internal conversion involving no CIs. In the experimental report by Ho *et al.*,[8] however, the efficient deactivation mechanism through CI was suggested even for the enol tautomer.

In this work, we perform high-level *ab initio* calculations employing the multistate complete active space second-order perturbation theory (MS-CASPT2) method for exploring the excited-state potential energy surfaces of all three cytosine tautomers. As seen in the following sections, the MS-CASPT2 treatment is essential for providing a reliable description of the deactivation pathways from the FC region to CIs. Besides revisiting the deactivation pathways of the keto form, we propose new decay pathways of the imino and enol forms leading to CIs that are located for the first time. We also carry out on-the-fly excited-state molecular dynamics (MD) simulations including the nonadiabatic transitions in order to complement the proposed scenario based on the potential energy surfaces. The present study provides a comprehensive picture of the photoinduced dynamics of cytosine tautomers and will be of great help to explain the experimentally observed multicomponent lifetimes.

This paper is organized as follows. Section II describes the computational details, and the results and discussion including comparisons with the experimental results are given in Sec. III. Conclusions are given in section IV, which also includes the future directions of research.

II. Computational Details

The equilibrium geometry in the electronic ground state is determined by the Møller-Plesset second-order perturbation (MP2) method while the geometry optimization in the singlet excited states is performed by the MS-CASPT2 method without any symmetry constraints. After the geometry optimization, single-point energy calculation is carried out at the MS-CASPT2 level for the ground and excited states. For comparison purposes, the single-state (SS) CASPT2 and SA-CASSCF (state-averaged complete active space self-consistent field) methods are also used for the calculation

of some potential energy profiles. A level shift with a value of 0.2 is applied for all SS- and MS-CASPT2 calculations.[41] The Sapporo-DZP basis set[42] (a natural orbital based correlating contracted basis set in the segmented form) is employed throughout this study (simply denoted as DZP for conciseness), and the MOLPRO 2008.1 package is used for all quantum chemical calculations.[43]

In modeling the nonradiative deactivation pathways of DNA bases, it is important to note that CASPT2 rather than CASSCF is required in many cases for the excited-state geometry optimization as well as for energy calculation. Although the CASSCF method has been frequently used for the excited-state geometry optimization in previous theoretical studies, the CASSCF-optimized geometries may be less accurate even qualitatively due to the lack of dynamic electron correlation.[44] Actually, in our recent study of uracil, thymine, and 5-fluorouracil, the CASPT2 and CASSCF methods predict completely different structures for the $^1\pi\pi^*$ minimum.[45] We also note that the MS-CASPT2 method[46] is more desirable than the SS-CASPT2 method[47,48] for the excited-state calculations of DNA bases.[45] In SS-CASPT2, the electronic energies are improved from SA-CASSCF by perturbation, but no mixing of the perturbed states is considered, which often leads to an inadequate description around avoid crossings and CIs.[49] In particular, the SS-CASPT2 method often predicts spurious CIs due to the nonorthogonality of the relevant wavefunctions.[46,49] The MS-CASPT2 method corrects this artifact by mixing these perturbed states via an effective Hamiltonian approach. In Ref.[45], for example, the MS-CASPT2 potential energy curves for the decay path of uracil derivatives are almost barrierless and consistent with experimental observations, while the SS-CASPT2 ones show an artificial barrier with substantial height due to the lack of state mixing. In this study, due to a significant role of CIs in the photodynamics of DNA bases, the MS-CASPT2 method is used for geometry optimization in order to provide a reliable description of the multiple electronic-state potential energy surfaces. The significance of MS-CASPT2 treatment is demonstrated below in some cases by comparing with the SA-CASSCF and SS-CASPT2 results.

In geometry optimization by the MS-CASPT2 method, the active space and the number of states included in the averaging procedures for the reference CASSCF wavefunctions are determined by a judicious choice for each structure in order to reduce the computational cost. The details are provided in the respective part of Section III. The notation of SA(l)-CASSCF(m,n) is used throughout this study, where l is the number of states included equally in the averaging procedure, and m and n are the numbers of electrons and orbitals in the active space, respectively. The notation of MS(l)-CASPT2(m,n) is also used, in which case the SA(l)-CASSCF(m,n) wavefunction is used as the reference and the l states are mixed after the perturbation. After the geometry is determined, the potential energies are recalculated by the MS-CASPT2 method with a larger active space for the reference CASSCF wavefunctions in order to provide improved energetics for a given geometry: MS(4)-CASPT2(12,9) for keto and imino, and MS(4)-CASPT2(10,8) for enol cytosine.

The minimal energy conical intersections (MECIs) are defined as local minima on the intersection seam and provide a useful characteristics of the nonadiabatic transitions. In this work, deactivation pathways connecting the FC region or S_1 minimum to MECIs between the S_1 and S_0 states are extensively investigated. The MECI search is also performed at the MS-CASPT2 level since the MECI points determined by SA-CASSCF or SS-CASPT2 are often quite different from those obtained by MS-CASPT2. The MECI search between the i- and j-th electronic states ($i = j+1$) is performed by a penalty-function approach proposed by Levine *et al*,[50] where the following function is minimized

$$F_{ij}(\mathbf{R};\sigma,\alpha) = \frac{E_i(\mathbf{R}) + E_j(\mathbf{R})}{2} + \sigma \frac{(E_i(\mathbf{R}) - E_j(\mathbf{R}))^2}{E_i(\mathbf{R}) - E_j(\mathbf{R}) + \alpha}. \qquad (2.1)$$

Here $E_i(\mathbf{R})$ ($E_j(\mathbf{R})$) is the electronic energy of the i-th (j-th) state at the molecular coordinate $\mathbf{R}$. The parameters of α = 0.02 hartree and σ = 3.5 are employed in this work by following the previous

report.[50] It has been confirmed that at all MECIs reported in this work the energy differences between the two relevant electronic states are less than ~0.1 eV at the MS-CASPT2 level used to determine the MECI geometries.

The excited-state transition state (TS) structure separating the S_1 minimum and S_1/S_0 MECI is determined in the following procedure. As a first step, we calculate the excited-state potential energy curve along the linearly interpolated internal coordinate (LIIC) points connecting the minimum and MECI. When the geometries of the two end-points are significantly different, however, potential energies may be seriously overestimated at the interpolated coordinates around the middle of the two structures. Therefore, as a second step, we pick up the highest energy point on the LIIC path and re-optimize this geometry in the S_1 state with a fixed driving coordinate representing the molecular distortion that leads to the target MECI (usually it is the twisting of a C-C or C-N double bond). Finally, the deactivation pathway is determined by connecting the optimized geometry in the second step to the minimum and MECI points using LIIC. The molecular structure of the highest energy point on the new LIIC path calculated in the final step (generally it is deviated from the optimized structure in the second step) is regarded as an approximation to the TS structure. The barrier height at the TS structure determined in this procedure can be viewed as an upper bound for that of the minimum energy pathway.

The on-the-fly *ab initio* MD simulations are performed at the SA-CASSCF level of theory including the nonadiabatic transitions based on the Tully's fewest switches algorithm.[51] The accuracy of SA-CASSCF potential energies is examined prior to the dynamical simulations by comparing with the MS-CASPT2 results. The gradients and nonadiabatic coupling vectors are evaluated analytically during the course of simulations. The equation-of-motion for nuclei is integrated by the velocity Verlet algorithm with a time step of 0.2 fs. At this time step, the total energy is well conserved and its standard deviation is less than 0.01 eV. The equations for the time evolution of electronic amplitudes are solved with a smaller time step of 0.04 fs using a unitary propagator, and the electronic energies and nonadiabatic coupling vectors required at the smaller time step are evaluated by interpolation using values obtained at the time steps of nuclei. When a potential energy difference between the two adjacent electronic states is less than 0.01 hartree, the time step for nuclei is reduced to 0.05 fs because values of the nonadiabatic coupling vectors vary rapidly around this region. In this case, the time step for electronic amplitude is also reduced to 0.01 fs.

In order to prepare the initial conditions for the excited-state MD simulations, the constant temperature MD simulation is performed in the electronic ground state at the MP2 level of theory. The temperature is controlled at 300 K by the Nosé-Hoover thermostat. The initial coordinates and velocities for the excited-state MD are taken from the ground-state MD run by picking up coordinates and velocities at every 50 fs. The excited-state MD simulations are initiated from the lowest $^1\pi\pi^*$ state for each tautomer. After the hopping event takes place, the component of velocity in the direction of the nonadiabatic coupling vector is adjusted to conserve a total energy. Trajectories that switch to the ground state and remain for more than 50 fs are terminated.

III. Results and Discussion

The ground-state equilibrium structures of the three tautomers of cytosine at the MP2/DZP level are shown in Figure 2, along with bond lengths and the atomic numbering. There are two equilibrium structures (rotamers) for the imino and enol tautomers depending on the direction of the NH bond in the imino group and the OH bond in the hydroxyl group, respectively, and the figure shows only the lower-energy rotamer. The structures of the higher-energy rotamers for imino and enol tautomers are given in the Electronic Supplementary Information (ESI) (Figure S1). In the keto and enol forms, the molecular structure is almost planar except for a slight pyramidal shape of the

amino group. In the imino form, the molecule exhibits a completely planar structure belonging to the C_s point group. The enol form is energetically more stable than the canonical keto form by 0.074 eV at the MP2 level, while the imino form is less stable than the keto form by 0.027 eV. The energetic order of enol < keto < imino is consistent with previous theoretical results by Yang *et al.*[18] and Kobayashi[52]. The equilibrium structure is designated as $(S_0)_{min}$ for each tautomer hereafter.

The photo-deactivation pathways of each cytosine tautomer are discussed in the following subsections. For the imino and enol forms, only the lower-energy rotamer is considered. The decay pathways of keto cytosine were extensively calculated in previous studies, but it is nevertheless essential to report our MS-CASPT2 results of this tautomer basically for two reasons. First, the potential energy profiles of the keto form should be compared with the imino and enol forms on the equal footing using a highly accurate *ab initio* method, in particular, in order to discuss the proposed scenario on the coexistence of tautomers in the experimental observations. Second, even when considering only the keto tautomer, comprehensive theoretical studies that have examined all three types of deactivation pathways are still limited. For imino and enol cytosine, we emphasize that the present work is the first *ab initio* study of potential energy profiles for the decay pathways and also the first study of the on-the-fly nonadiabatic molecular dynamics simulations.

A Keto Cytosine

For the keto tautomer of cytosine, vertical excitation energies at $(S_0)_{min}$ and potential energy profiles between the S_1 minimum and S_1/S_0 MECI points are calculated at the MS(4)-CASPT2(12,9) level of theory, where the S_0, first $^1\pi\pi^*$, and two $^1n\pi^*$ states are averaged with equal weights in the reference SA(4)-CASSCF(12,9) calculation. The active space is comprised of seven π orbitals (four of them are doubly-occupied in the closed-shell configuration) and two lone-pair orbitals on the O7 and N3 atoms. The π orbital localized on the N8 atom is excluded from the active space since inclusion of this orbital (which in this case is (14,10)) induces only ~0.01 eV differences in the excitation energies. The active orbitals at $(S_0)_{min}$ in the SA(4)-CASSCF(12,9) calculation are shown in the ESI (Figure S2).

Vertical excitation energies

The vertical excitation energies of keto cytosine are summarized in Table I, where the results of MS(4)-CASPT2(12,9), SS-CASPT2(12,9), and SA-CASSCF(12,9) using the DZP basis set are shown, along with previously reported theoretical and experimental values. Our calculations predict that the S_1 state is of $^1\pi\pi^*$ character and the S_2 and S_3 states are $^1n\pi^*$, as seen in most of the previous theoretical studies. In the present work, the MS(4)-CASPT2(12,9) excitation energy of the $^1\pi\pi^*$ state is calculated to be 4.48 eV, which agrees well with the experimental value of ~4.65 eV[53] in the gas phase as well as those in the other environments.[54-56] According to the oscillator strengths at the SA(4)-CASSCF(12,9) level, the $^1\pi\pi^*$ state is expected to be primarily populated by UV absorption. Our MS-CASPT2 calculations also exhibit that the two $^1n\pi^*$ states are characterized by modest mixing of the excitation from lone-pair orbitals centered on the O7 and N3 atoms, denoted as $^1n_O\pi^*$ and $^1n_N\pi^*$, respectively. Overall, the agreement with the previous CASPT2 calculations[24,33] is fairly well. The multireference configuration interaction (MR-CISD+Q) calculations predicted higher excitation energies for all states,[38] which could be attributed to the smaller active space employed in the MR-CISD calculations (6 electrons in 5 orbitals).

As seen in Table I, the SA(4)-CASSCF(12,9) method significantly overestimates the excitation energies of both the $^1\pi\pi^*$ and $^1n\pi^*$ states primarily due to the lack of dynamic electron correlation. The SS-CASPT2(12,9) excitation energies are much smaller, and in particular, the $^1\pi\pi^*$

excitation energy is lower than that of SA(4)-CASSCF(12,9) by 0.82 eV, and it is underestimated compared to experimental values and other theoretical predictions. The MS(4)-CASPT2(12,9) calculation exhibits the $^1\pi\pi^*$ excitation energy larger than SS-CASPT2(12,9) by 0.30 eV owing to the mixing between electronic states, which results in the good agreement with experimental value mentioned above.

Table I. Vertical Excitation Energies (in eV) of Keto Cytosine for the Three Lowest Singlet Excited States.

	$^1\pi\pi^*$	$^1n\pi^*$	ref.
MS(4)-CASPT2(12,9)/DZP	4.48	4.74, 5.26	this work
SS-CASPT2(12,9) /DZP	4.18	4.68, 5.01	this work
SA(4)-CASSCF(12,9) /DZP	5.00 (0.083)[a]	5.14 (0.001)[a], 5.52 (0.003)[a]	this work
Previous calculations			
DFT-MRCI/TZVP	4.83	5.02, 5.50	22
EOM-CCSD/cc-pCVTZ	5.06	5.51, 6.14	57
CR-EOM-CCSD(T)/cc-pVDZ	4.76	5.24	58
CC3/aug-cc-pVDZ	4.71	–	16
CC2/aug-TZVP	4.61	4.87, 5.27	38
CC2/aug-cc-pVDZ	4.56	4.80, 5.26	40
SS-CASPT2(12,10)/ANO	4.39	5.00	59
SS-CASPT2(12,9)/6-31G**	4.50	4.88, 5.23	24
SS-CASPT2(12,11)/6-31+G*	4.46	5.26	31
SS-CASPT2(12,12)/6-31+G*	4.51	5.59	31
MS(5)-CASPT2(12,9)/6-31G*	4.56	5.14, 5.68	34
CASPT2(14,10)/ANO-S	4.41	4.95, 5.06	33
MRCI(12,9)/cc-pVDZ	5.14	5.29, 5.93	30
MR-CISD+Q(6,5)/6-31G*	5.39	5.55, 6.80	38
Experiments			
exp. in the gas phase	4.65	5.5, 6.2, 6.7	53
exp. in trimethyl phosphate	4.48	5.23, 6.08, 6.70	54
exp. in aqueous solution	4.65	5.44, 6.30,	55
exp. in crystal	4.70	5.33, 5.63, 6.26	56
exp. in acetonitrile	4.58	–	40

[a] Number in parentheses represents oscillator strength.

S_1 minimum energy structure

Our MS-CASPT2 calculations predict that the S_1 global minimum of keto cytosine is characterized by the $^1\pi\pi^*$ state. The SA-CASSCF geometry optimization results in a minimum energy structure of the $^1n_O\pi^*$ state, but the MS-CASPT2 optimization starting from this structure leads a minimum of the $^1\pi\pi^*$ state; the detailed discussion is found in the ESI. This result also suggests the necessity of CASPT2 rather than CASSCF for the excited-state geometry optimization. It is noteworthy that the $^1\pi\pi^*$ character of the S_1 minimum was also suggested by previous theoretical studies using quantum chemical methods that sufficiently take into account dynamic electron correlation.[22,30,31]

Figure 3(a) shows the optimized minimum energy structure in the $^1\pi\pi^*$ state (labeled by

($^1\pi\pi^*$)$_{min}$ hereafter). This structure is determined by the MS(2)-CASPT2(8,7) method, where the active space is comprised of only π orbitals. The π orbital on the N8 atom is excluded from the active space. As seen in the figure, planarity of the six-membered ring is maintained at ($^1\pi\pi^*$)$_{min}$. The geometrical change from (S_0)$_{min}$ involves the bond inversion of the ring; in particular, the N3-C4, C4-C5, and C5-C6 bond lengths are changed by more than 0.05 Å. The MS(4)-CASPT2(12,9) energy at ($^1\pi\pi^*$)$_{min}$ is calculated to be 3.98 eV from (S_0)$_{min}$. This energy agrees well with the DFT/MRCI and MRCI energies of 4.06 and 4.31 eV, respectively.[22,30] It also agrees quite well with the experimental adiabatic excitation energy of 3.95 eV estimated from the resonance-enhanced multiphoton ionization (REMPI) spectrum.[10,60]

It has been proposed previously that the minimum energy path (MEP) approach is appropriate for the determination of the deactivaiton pathway since it provides information about the accessibility of the conical intersection.[29,61-63] There is the possibility that it leads to the seam of CIs before reaching the ($^1\pi\pi^*$)$_{min}$ structure, thus providing the most relevant photochemical deactivation pathway. Therefore, the MEP computation in the $^1\pi\pi^*$ state was performed at the MS(2)-CASPT2(8,7) level starting from the (S_0)$_{min}$ structure, and it was confirmed that MEP leads to ($^1\pi\pi^*$)$_{min}$ for the keto form (the details are provided in the ESI). The potential energy profiles from (S_0)$_{min}$ to ($^1\pi\pi^*$)$_{min}$ are shown in Figure 4(a) and the $^1\pi\pi^*$ state exhibits a smooth downhill potential. Therefore the molecule is expected to initially relax toward ($^1\pi\pi^*$)$_{min}$ after photoexcitation to the $^1\pi\pi^*$ state. The involvement of the $^1n_O\pi^*$ state in the early dynamics was suggested in previous theoretical studies, [17,28,30,32] but this speculation comes mainly from less accurate CASSCF potential energy profiles. Our MS-CASPT2 calculations predict that the contribution of the $^1n_O\pi^*$ state is small in the excited-state dynamics of the keto tautomer (see also discussion in the ESI).

S_1/S_0 MECI structures

Three types of MECIs are located for the keto form at the MS-CASPT2 level. Figure 3(b-d) shows the optimized geometry of these MECIs, which are named as ($^1\pi\pi^*$/gs)$_{CI}$, ($^1\pi_{N3}\pi^*$/gs)$_{CI}$, and ($^1n_O\pi^*$/gs)$_{CI}$ from the electronic character of the excited state at each structure.

First, ($^1\pi\pi^*$/gs)$_{CI}$ is the ethylene-like CI involving the twisting of the C5-C6 double bond, which was first identified by Sobolewski and Domcke in their work on the guanine-cytosine base pair.[64] We performed a MECI search between the ground and $^1\pi\pi^*$ states at the MS(2)-CASPT2(8,7) level, where the active space is the same as that used in the geometry optimization of the $^1\pi\pi^*$ minimum. As seen in Figure 3(b), the MECI geometry exhibits a significant puckering of the C6 atom, and the amino group possesses an almost planar structure. The dihedral angle of *d*(N1-C6-C5-H5) representing the twisting of the C5-C6 double bond is 66.5 degrees. The $^1\pi\pi^*$ state exhibits a diradical character which is described by a primary excitation from the *p* orbitals on C5, N3, and O7 to the *p* orbital on C6. The natural orbitals relevant to this excitation are given in the ESI (Figure S6). Owing to the $^1\pi\pi^*$ excitation, the C5-C6 bond distance is elongated to 1.475 Å. The C2-O7 bond length is only 1.218 Å at ($^1\pi\pi^*$/gs)$_{CI}$ and it shrinks slightly from 1.276 Å at ($^1\pi\pi^*$)$_{min}$, suggesting that the π orbital of this bond plays a minor role in the excitation. Our structural data are consistent with MRCI results[30], where the C5-C6 and C2-O7 bond lengths are reported as 1.464 and 1.206 Å, respectively. The C2-O7 bond length reported by Merchán and Serrano-Andrés[24] is 1.428 Å and it exhibits a large discrepancy with our result. This is possibly because the CASSCF method is used for their MECI optimization.

Second, ($^1\pi_{N3}\pi^*$/gs)$_{CI}$ involves the puckering of the N3 atom with the other five ring atoms remaining coplanar, which is often referred to as the sofa conformation and is characterized by the twisting of the N3-C4 double bond. This structure was first identified by Ismail *et al.*[23] The MECI point is located at the MS(2)-CASPT2(8,7) level, where the active space is comprised of six π orbitals and a lone-pair orbital localized on the N3 atom. The occupied two π orbitals, which are mainly

localized on the N1 and N8 atoms, are excluded from the active space since the occupation numbers of the corresponding natural orbitals are more than 1.96 by SA(2)-CASCSF(12,9) calculation around this CI. The optimized structure is shown in Figure 3(c) and it involves a significant deformation of the amino group from the molecular plane in addition to the puckering of the N3 atom. The C4 atom is strongly pyramidalized due to the deformation of the amino group. The dihedral angle of *d*(C2-N3-C4-N8) representing the twisting of the N3-C4 double bond is 71.9 degrees. This MECI exhibits the $^1\pi\pi^*$ state of a diradical character, which is well described by an excitation from the *p* orbital on N3 to the *p* orbital on C4 (see Figure S7 for the relevant orbitals), as has been discussed in refs.[30,38] Therefore this state is labeled as $^1\pi_{N3}\pi^*$ hereafter. It is noted that in some other previous works this CI was denoted by $(^1n_N\pi^*/gs)_{CI}$[23,24] possibly because the $^1n_N\pi^*$ configuration has nonnegligible contribution along the deactivation pathway, but the MECI point itself is well characterized by the $^1\pi_{N3}\pi^*$ configuration.

Finally, $(^1n_O\pi^*/gs)_{CI}$ involves the $^1n_O\pi^*$ state, which was also identified by Ismail *et al.*[23] The search for this MECI is performed at the MS(3)-CASPT2(6,5) level, where the four π orbitals (two occupied and two unoccupied) and one lone-pair orbital of the O7 atom are included in the active space. This size of the active space is determined based on the occupation numbers of natural orbitals in SA(2)-CASSCF(12,9) calculation around this CI. The optimized structure is shown in Figure 3(d) and it involves a significant elongation of the C2-O7 bond to 1.528 Å and also bond inversion of the ring. The pyramidalizations of the C6 and N1 atoms, where *d*(N1-C6-C5-H6) = 127.3° and *d*(C2-N1-C6-H1) = −152.4°, are observed. The planarity of the ring is almost maintained and therefore this structure is called semi-planar CI in Ref.[38]. The ground and first two excited states at this point exhibit a strong mixing of the three configurations: closed-shell, $^1\pi\pi^*$, and $^1n_O\pi^*$. This mixing is indicated in the eigenvectors of the MS-CASPT2 effective Hamiltonian matrix (see Table SI and Table SII in the ESI). Barbatti *et al.*[38] labeled this type of MECI as $((^1n_O\pi^*+cs)/^1\pi\pi^*)_{CI}$ to represent the electronic state mixing, where cs stands for the closed-shell configuration.

The MS(4)-CASPT2(12,9) energies of the S_1 state at the three MECIs, relative to the ground-state energy of $(S_0)_{min}$, are also given in Figure 3(b-d). The $(^1\pi\pi^*/gs)_{CI}$ structure exhibits the $^1\pi\pi^*$ energy of 3.80 eV, which is 0.18 eV below the energy at $(^1\pi\pi^*)_{min}$. The S_1 energy at $(^1\pi_{N3}\pi^*/gs)_{CI}$ is calculated to be 4.07 eV. This energy is 0.09 eV higher than the energy at $(^1\pi\pi^*)_{min}$, but considerably lower than the vertical excitation energy of the $^1\pi\pi^*$ state at the $(S_0)_{min}$ structure (4.48 eV, see Table I). The S_0 energy at $(^1\pi_{N3}\pi^*/gs)_{CI}$ is 3.29 eV and indicates a relatively large energy difference from the $^1\pi_{N3}\pi^*$ state. This is mainly due to smaller active space used in optimizing the MECI structure. For $(^1n_O\pi^*/gs)_{CI}$, the MS(4)-CASPT2(12,9) energies of the lowest three states are 5.17 (S_0), 5.20 (S_1), and 5.30 eV (S_2), in reasonable agreement with those reported in Ref.[34] (4.75 (S_0), 4.91 (S_1), and 5.06 eV (S_2)) obtained by a similar methodology of MS(5)-CASPT2(12,9)/6-31G*. The MS-CASPT2 energy of $(^1n_O\pi^*/gs)_{CI}$ is much higher than the $^1\pi\pi^*$ vertical excitation energy, which indicates that the deactivation through this CI is quite unlikely. This is in accord with previous theoretical calculations.[23,29,30] Since the potential energies of the three states are very close to one another at $(^1n_O\pi^*/gs)_{CI}$, three-state conical intersections[25,30-32] are likely to be located near this MECI.

Deactivation pathway to $(^1\pi\pi^*/gs)_{CI}$

The potential energy profiles along the deactivation pathway from $(^1\pi\pi^*)_{min}$ to $(^1\pi\pi^*/gs)_{CI}$ are shown in Figure 4(b). The driving coordinate for optimizing the highest energy point along LIIC is chosen as the dihedral angle of *d*(N1-C6-C5-H5) which represents the twisting of the C5-C6 double bond. The MS(2)-CASPT2(8,7) method is employed for this optimization, where the dihedral angle is fixed at *d* = 134.0°. The S_1 state is well represented by the $^1\pi\pi^*$ excitation within the C5-C6 double bond all along the decay path.

As seen in the figure, the $^1\pi\pi^*$ potential energy curve is almost flat with a very small barrier

of ~0.08 eV from $(^1\pi\pi^*)_{min}$ at around d = 100°. The energy of the barrier (4.06 eV) is much lower than the vertical excitation energy of the $^1\pi\pi^*$ state (4.48 eV), which suggests very efficient nonradiative deactivation through $(^1\pi\pi^*/gs)_{CI}$. Moreover, the $^1\pi\pi^*$ potential energies at the barrier and MECI for this decay path are lower than those for other paths through $(^1\pi_{N3}\pi^*/gs)_{CI}$ or $(^1n_O\pi^*/gs)_{CI}$ (see also below). Therefore, the population decay through $(^1\pi\pi^*/gs)_{CI}$ is expected to occur preferentially in the deactivation process of keto cytosine. Our calculated barrier height agrees well with previous theoretical results of 0.1 ~ 0.14 eV[24,27,30,31,64].

It is shown in the present work that the MS-CASPT2 method is appropriate for the geometry optimization of $(^1\pi\pi^*/gs)_{CI}$ as well as for the potential energy calculations along the decay path. A more detailed discussion of this point is given in the ESI (Figures S9 and S10).

Deactivation pathway to $(^1\pi_{N3}\pi/gs)_{CI}$

The potential energy profiles from $(^1\pi\pi^*)_{min}$ to $(^1\pi_{N3}\pi^*/gs)_{CI}$ are shown in Figure 4(c), where the highest energy point along LIIC is optimized at the MS(2)-CASPT2(8,7) level by fixing a dihedral angle of d(C2-N3-C4-N8) at 106.4 degrees. While the S_1 state is dominated by the $^1\pi\pi^*$ excitation around $(^1\pi\pi^*)_{min}$, the $^1n_N\pi^*$ excitation from the in-plane lone-pair orbital on the N3 atom also has nonnegligible contribution as the dihedral angle d(C2-N3-C4-N8) is decreased below 150 degrees. When the dihedral angle is decreased further, the configuration for the single excitation from the out-of-plane π_{N3} orbital starts to mix, and at the MECI point $(^1\pi_{N3}\pi^*/gs)_{CI}$, as discussed above, the electronic character is mainly described by the excitation from the π_{N3} orbital.

The barrier height for the decay to $(^1\pi_{N3}\pi^*/gs)_{CI}$ is estimated to be ~0.18 eV from $(^1\pi\pi^*)_{min}$, which is higher than that for the decay to $(^1\pi\pi^*/gs)_{CI}$. Therefore the deactivation through $(^1\pi_{N3}\pi^*/gs)_{CI}$ is expected to play a comparatively minor role in the photophysics of keto cytosine. This finding is partly supported by the experimental result that cytidine at low pH still has an excited-state lifetime in the subpicosecond range despite the loss of the $^1n_N\pi^*$ state by protonation at N3.[65] Our calculated barrier height is consistent with previously reported theoretical values of 0.14 eV by Kistler *et al.*[30] and 0.2 eV by Blancafort[31]. The barrier height reported by Merchán and Serrano-Andrés[24] is relatively high (0.52 eV), which is possibly attributed to the fact that the CASSCF method is used in geometry optimization.

It is noted that the potential energy of $(^1\pi_{N3}\pi^*/gs)_{CI}$ is higher than that of $(^1\pi\pi^*/gs)_{CI}$ by 0.27 eV. This result implies that the pathway through $(^1\pi\pi^*/gs)_{CI}$ has wider energy range of intersection seam accessible with a given amount of energy (for discussions on the role of the extended range of the seam, see for example Refs.[66,67]). The lower energy of MECI thus supports the higher efficiency of the decay through $(^1\pi\pi^*/gs)_{CI}$.

Deactivation pathway to $(^1n_O\pi^*/gs)_{CI}$

The potential energy profiles from $(^1\pi\pi^*)_{min}$ to $(^1n_O\pi^*/gs)_{CI}$ along LIIC points are shown in Figure 4(d). The S_1 potential energy monotonically increases until it reaches $(^1n_O\pi^*/gs)_{CI}$, exhibiting the sloped topology of the CI. As discussed above, the deactivation through $(^1n_O\pi^*/gs)_{CI}$ is unlikely to occur because of the high energy of MECI. The S_1 and S_2 states exhibit very similar energies along the decay path as well as a significant mixing of closed-shell, $^1n_O\pi^*$, and $^1\pi\pi^*$ configurations.

We demonstrate here that the MS-CASPT2 method is required rather than the SA-CASSCF and SS-CASPT2 methods for a quantitative description of the decay path to $(^1n_O\pi^*/gs)_{CI}$. Figure 5 shows the SA(4)-CASSCF(12,9), SS-CASPT2(12,9), and MS(4)-CASPT2(12,9) potential energy profiles along the LIIC path from $(^1\pi\pi^*)_{min}$ to $(^1n_O\pi^*/gs)'_{CI}$, where in this case MECI between the

$^1n_O\pi^*$ and S_0 states is located by the SA(3)-CASSCF(10,8) method (indicated by prime). In this MECI search, the S_0, $^1n_O\pi^*$, and $^1\pi\pi^*$ states are included with equal weights in the average, and the (10,8) active space is comprised of 7 π orbitals and lone pair orbital localized on the O7 atom. The optimized structure of $(^1n_O\pi^*/gs)'_{CI}$ (see Figure S11) exhibits pyramidalizations of the C6 and N1 atoms as seen in $(^1n_O\pi^*/gs)_{CI}$. The electronic structure at $(^1n_O\pi^*/gs)'_{CI}$ is also characterized by a strong mixing of the closed-shell, $^1n_O\pi^*$, and $^1\pi\pi^*$ configurations (the configuration-interaction coefficients are listed in Table SIII). However, as seen in Figure 5(c), the MS-CASPT2 potential energy curves of the ground and first excited states are separated by more than 1.5 eV even at $(^1n_O\pi^*/gs)'_{CI}$. This is quite different from the SA-CASSCF and SS-CASPT2 results in Figure 5(a,b), which predict very small energy gap between the S_0 and S_1 states at $(^1n_O\pi^*/gs)'_{CI}$ and an efficient decay path toward this MECI. The SS-CASPT2 calculations predict that the three states are relatively close in energy at $(^1n_O\pi^*/gs)'_{CI}$, but the inclusion of the mixing of the states by MS-CASPT2 yields quite different energetics. These results strongly suggest that the geometry optimization for $(^1n_O\pi^*/gs)_{CI}$ should be performed using the MS-CASPT2 method.

Recent on-the-fly dynamical simulations based on the CASSCF potential energies predicted that the population decay dominantly takes place via $(^1n_O\pi^*/gs)_{CI}$.[36,38] As shown in Figure 5, however, this is primarily due to artificial stabilization of the pathway to $(^1n_O\pi^*/gs)_{CI}$ in the CASSCF calculation. Our CASSCF dynamical simulations also exhibit this tendency, which will be discussed below.

B Imino Cytosine

The vertical excitation energies and potential energy profiles of the imino tautomer are calculated at the MS(4)-CASPT2(12,9)/DZP level of theory, where twelve active electrons are distributed in eight π orbitals plus the lone-pair orbital on the N8 atom. The lone-pair orbital on the O7 atom is excluded from the active space, since the $^1n\pi^*$ state involving the excitation from this orbital lies quite high in energy. The active orbitals at $(S_0)_{min}$ in the SA(4)-CASSCF(12,9) calculation are shown in the ESI (Figure S12).

Vertical excitation energies

Table II summarizes the vertical excitation energies of imino cytosine at the MP2 optimized structure of $(S_0)_{min}$, showing the results of MS(4)-CASPT2(12,9), SS-CASPT2(12,9), and SA(4)-CASSCF(12,9) calculations. At the CASSCF level, the lowest singlet excited state involves the excitation from an in-plane lone-pair orbital localized on the N8 atom to the π* orbital, which is hereafter denoted by $^1n_{N8}\pi^*$. The second and third excited states are characterized by the $^1\pi\pi^*$ excitation. The SS-CASPT2(12,9) and MS(4)-CASPT2(12,9) calculations predict that the $^1\pi\pi^*$ state is the lowest excited state and likely to be initially populated by UV absorption. The CASPT2 excitation energies agree fairly well with the DFT/MRCI values,[22] although the energetic order of the first $^1\pi\pi^*$ and $^1n_{N8}\pi^*$ states is opposite in the latter. As far as we are aware, experimental results for vertical excitation energies are not available for the imino tautomer.

Table II Vertical Excitation Energies (in eV) of Imino Cytosine for the Three Lowest Singlet Excited States.

	$^1\pi\pi^*$	$^1n_{N8}\pi^*$	$^1\pi\pi^*$	ref.
MS(4)-CASPT2(12,9)/DZP	4.67	5.59	5.75	this work
SS-CASPT2(12,9)/DZP	4.98	5.53	5.27	this work
SA(4)-CASSCF(12,9)/DZP	6.07 (0.189)[a]	5.92 (0.006)[a]	6.91 (0.167)[a]	this work
Previous calculations				
DFT-MRCI/TZVP	5.26	5.19	5.96	[22]
CC3/aug-cc-pVDZ	5.07	5.89	6.28	[16]
CC2/aug-cc-pVDZ	5.00	5.53	5.77	[40]

[a] Number in parentheses represents oscillator strength.

S_1/S_0 MECI structures

First, we tried to locate the minimum energy structure in the S_1 state with geometry optimization at the MS(2)-CASPT2(6,5) level starting from the FC region. The active space is comprised of four π orbitals (two occupied and two unoccupied) plus one lone-pair orbital belonging to the N8 atom. This size of the active space was determined based on the occupation numbers of natural orbitals in SA(2)-CASSCF(12,9) calculation, and orbitals with the occupation number between 0.05 and 1.95 are included in the active space. The geometry optimization, however, failed to converge since the molecule encountered the CI region, where pronounced out-of-plane twisting of the imino group was observed. Thus, the twisting of the imino group is expected to directly lead to S_1/S_0 MECI rather than S_1 minimum in the MS-CASPT2 calculation.

Therefore, we performed a MECI search at the MS(2)-CASPT2(6,5) level starting from the out-of-plane deformed structure. As expected, the optimization results in the MECI geometry shown in Figure 6, where the imino group is almost perpendicular to the ring and the dihedral angle *d*(N3-C4-N8-H8) is 82.6 degrees.. The excited state at this MECI is characterized by a single excitation from a lone-pair orbital on the N8 atom to a π* orbital of the ring. Since the lone-pair orbital corresponds to a π orbital on the N8 atom at the $(S_0)_{min}$ structure (see Figure S12), this MECI structure is denoted by $(^1\pi_{N8}\pi^*/gs)_{CI}$ hereafter. The relevant natural orbitals are shown in the ESI (Figure S13). Other significant structural changes from $(S_0)_{min}$ are bond elongation of C4-N8 from 1.293 to 1.474 Å, bond shrinkage of C4-C5 from 1.463 to 1.373 Å, and a slight puckering at the N3 atom. The MECI involving the rotation of the imino group was also reported in the 7H-keto-imino tautomer of guanine.[68]

Here we comment on the notation for the excited state of imino cytosine. In the present work, the excited state at the twisted structure is labeled as $^1\pi_{N8}\pi^*$ for the reason stated above. In the previous study by Tomíc *et al.*[22], on the other hand, the excited state at similar structure is referred to as $^1n\pi^*$. The detail of the non-bonding orbital could not be found in Ref. [22], but we believe that this orbital is the same as the π_{N8} orbital, which is a kind of lone-pair orbital on the N8 atom. Also for the imino tautomer of guanine, the excited state at the twisted structure is referred to as $^1n\pi^*$ in Ref. [68], while the non-bonding orbital is produced by the rotation of the π orbital on the imino group from the S_0 minimum structure.

The MS(4)-CASPT2(12,9) energies at $(^1\pi_{N8}\pi^*/gs)_{CI}$ are 2.83 and 3.31 eV from $(S_0)_{min}$ for the S_0 and $^1\pi_{N8}\pi^*$ states, respectively. This $^1\pi_{N8}\pi^*$ energy at $(^1\pi_{N8}\pi^*/gs)_{CI}$ is well below the $^1\pi\pi^*$ energy at $(S_0)_{min}$ (4.67 eV).

Deactivation pathway to $(^1\pi_{N8}\pi/gs)_{CI}$

The MS(4)-CASPT2(12,9) potential energy profiles from $(S_0)_{min}$ to $(^1\pi_{N8}\pi^*/gs)_{CI}$ along the LIIC points are shown in Figure 7. The potential energy curve of the S_1 state clearly exhibits a barrierless pathway toward the MECI. Therefore it is speculated that the molecule would deactivate to the ground state quite efficiently through $(^1\pi_{N8}\pi^*/gs)_{CI}$ after photo-excitation to the $^1\pi\pi^*$ state.

To our knowledge, this is the first time that the $(^1\pi_{N8}\pi^*/gs)_{CI}$ structure and decay path to this MECI have been proposed for imino cytosine. In the previous DFT/MRCI study by Tomíc *et al.*[22], an S_1 minimum rather than S_1/S_0 MECI was located in the twisted structure of the imino group (the S_1 state is called $^1n\pi^*$ in Ref. [22], but this could be $^1\pi_{N8}\pi^*$ in the present notation). The molecular structure of this minimum is close to the $(^1\pi_{N8}\pi^*/gs)_{CI}$ structure, but the S_0 and S_1 energies exhibit a large energy gap of about 1.5 eV. Based on this result, Kosma *et al.*[7] attributed the long-time component of $\tau > 19$ ps observed in their experimental spectra to a population trap in the S_1 minimum. However, our calculations contradict this assumption, and instead show that an extremely efficient deactivation pathway exists for the imino form.

It is anticipated from the study of the keto tautomer that there might be another deactivation pathway which involves ethylene-like CI with a large twisting of the C5-C6 double bond. However, all attempts of MECI search at the MS-CASPT2 level have led to the $(^1\pi_{N8}\pi^*/gs)_{CI}$ region. Instead, MECI of an ethylene-like structure is located by optimization with the SA(2)-CASSCF(10,8) method (performed with the projected gradient method by Bearpark *et al.*[69]), where all π orbitals are included in the active space. The optimized structure is shown in Figure S14, where the dihedral angle *d*(N1-C6-C5-H5) is 61.1 degrees. However, the MS(4)-CASPT2(12,9) energies at this point (3.48 and 4.26 eV for the S_0 and $^1\pi\pi^*$ states, respectively) are considerably higher than the energy of $(^1\pi_{N8}\pi^*/gs)_{CI}$. Therefore, we conclude that the ethylene-like CI plays a very minor role in the deactivation process of imino cytosine. The excited-state MD simulations also show a clear preference for the decay through $(^1\pi_{N8}\pi^*/gs)_{CI}$, which will be described below.

C Enol Cytosine

The vertical excitation energies and potential energy profiles of the enol tautomer are calculated at the MS(4)-CASPT2(10,8)/DZP level of theory, where the active space is composed of π orbitals of the six-membered ring plus lone-pair orbitals belonging to the N1 and N3 atoms. The π orbital on the N8 atom is excluded from the active space as is treated in the calculations of the keto form. Also the lone-pair orbital on the O7 atom is not included in the active space, since the $^1n\pi^*$ state involving an excitation from this orbital lies quite high in energy. As a result, for the enol tautomer we employ a difference size of the active space from the keto and imino tautomers. The active orbitals at $(S_0)_{min}$ in the SA(4)-CASSCF(10,8) calculation are shown in the ESI (Figure S15).

Vertical excitation energies

Table III shows the vertical excitation energies at the $(S_0)_{min}$ structure obtained with the MS(4)-CASPT2(10,8) calculation, along with those at the SS-CASPT2(10,8) and SA(4)-CASSCF(10,8) results. In the CASSCF calculation, the first and third excited states are characterized by the $^1\pi\pi^*$ excitation, while the second excited state is described by the $^1n\pi^*$ excitation. The $^1n\pi^*$ state involves the excitation from a nonbonding orbital represented by a linear combination of the N1 and N3 lone-pair orbitals (orbital #29 in Figure S15). The energetic order of the lowest four states is unchanged upon inclusion of dynamic electron correlation by the SS- and MS-CASPT2

methods. It is noted that the SS- and MS-CASPT2 excitation energies are almost the same values, because the mixing of these excited states are very small at $(S_0)_{min}$. According to the SA(4)-CASSCF(10,8) oscillator strength, the lowest $^1\pi\pi^*$ state is the most likely to be populated by UV absorption.

Table III. Vertical Excitation Energies (in eV) of Enol Cytosine for the Three Lowest Singlet Excited States.

	$^1\pi\pi^*$	$^1n\pi^*$	$^1\pi\pi^*$	ref.
MS(4)-CASPT2(10,8) /DZP	4.81	4.88	5.48	this work
SS-CASPT2(10,8) /DZP	4.81	4.89	5.47	this work
SA(4)-CASSCF(10,8)/DZP	5.20 (0.026)[a]	5.55 (0.013)[a]	6.50 (0.000)[a]	this work
Previous calculations				
DFT-MRCI/TZVP	5.14	5.27	6.13	[22]
CC3/aug-cc-pVDZ	4.88	5.84	6.39	[16]
CC2/aug-cc-pVDZ	4.88	5.12	5.81	[40]

[a] Number in parentheses represents oscillator strength.

S_1 minimum energy structure

In the search of the minimum energy structure in the S_1 state, the MS(2)-CASPT2(8,7) method is employed, where the active space includes the six π orbitals of the ring and one nonbonding orbital. Another nonbonding orbital is not included into the active space since its occupation number is close to 2.0 in SA(2)-CASSCF(10,8) calculation. The optimized structure is shown in Figure 8(a), where the molecule exhibits a slight puckering of the C6 atom represented by a dihedral angle of *d*(C2-N1-C6-C5) = 46.0°. The electronic character of the S_1 state is $^1\pi\pi^*$ at this minimum, and this minimum is denoted as $(^1\pi\pi^*)_{min}$ hereafter. The N1-C6 and C5-C6 bonds are elongated from 1.350 to 1.397 Å and from 1.390 to 1.488 Å, respectively, from the $(S_0)_{min}$ structure. The adiabatic excitation energy for this structure is calculated to be 4.29 eV, which is 0.52 eV below the vertical excitation energy of the lowest $^1\pi\pi^*$ state. This value is in very good agreement with the REMPI experimental value of 4.46 eV[52,53] and the MRCI/DFT result of 4.50 eV.[22] Since the $^1n\pi^*$ and $^1\pi\pi^*$ energies are close in the FC region, we also tried to locate a minimum energy structure in the $^1n\pi^*$ state. However, such structure could not be found at the MS(2)-CASPT2(8,7) level since geometry optimization in the S_1 state always led to the $(^1\pi\pi^*)_{min}$ point.

When the S_1 geometry optimization is performed at the SA-CASSCF level, the minimum energy structure exhibits the $^1n\pi^*$ character. As shown in the ESI (Figure S16), the six-membered ring possesses almost planar structure at the optimized structure, in contrast to the nonplanar structure of $(^1\pi\pi^*)_{min}$ mentioned above. The discrepancy between the MS-CASPT2 and SA-CASSCF results indicates that the inclusion of dynamic electron correlation is critical to determine the S_1 minimum energy structure in enol cytosine.

The potential energy profiles from $(S_0)_{min}$ to $(^1\pi\pi^*)_{min}$ are calculated along the LIIC points and shown in Figure 9(a). The MEP computation in the lowest $^1\pi\pi^*$ state was also performed at the MS(2)-CASPT2(8,7) level starting from the $(S_0)_{min}$ structure and it was confirmed that the MEP lead to $(^1\pi\pi^*)_{min}$ as seen in the keto form (the details are provided in the ESI). Therefore relaxation to $(^1\pi\pi^*)_{min}$ is expected in the early dynamics after photo-excitation.

S_1/S_0 MECI structures

In the enol form, we have located three types of MECIs for the first time, whose structures are shown in Figure 8(b-d). These MECIs are labeled as $(^1\pi\pi^*/gs)_{CI}$, $(^1\pi_{N3}\pi^*/gs)_{CI}$, and $(^1\pi_{N1}\pi^*/gs)_{CI}$. The former two MECIs exhibit structural characteristics similar to the respective MECIs in the keto form with the same label.

First, the ethylene-like CI $(^1\pi\pi^*/gs)_{CI}$ involving the twisting of the C5-C6 bond is located as in the keto tautomer. The MECI search is performed at the MS(2)-CASPT2(8,7) level, where the active space is the same as that used in the S_1 geometry optimization. The optimized structure shown in Figure 8(b) involves the puckering of the C6 atom and a significant distortion of the C5-H5 bond from planar structure. The dihedral angle of *d*(N1-C6-C5-H5) = 51.2° that represents these motions is smaller than the respective dihedral angle of the keto form of 66.5 degrees.

The second CI $(^1\pi_{N3}\pi^*/gs)_{CI}$ involves the puckering of the N3 atom and out-of-plane deformation of the amino group in analogy with keto cytosine. As in the same way in the keto form, the MECI search for $(^1\pi_{N3}\pi^*/gs)_{CI}$ is performed at the MS(2)-CASPT2(8,7) level. The optimized MECI structure shown in Figure 8(c) exhibits the dihedral angle of *d*(C2-N3-C4-N8) = 66.0°, which is slightly smaller than the respective dihedral angle of the keto form of 71.9 degrees.

In the last CI $(^1\pi_{N1}\pi^*/gs)_{CI}$, the N1 atom is strongly puckered and the other five ring atoms remain coplanar. The MECI search is performed also at the MS(2)-CASPT2(8,7) level. The optimized geometry is shown in Figure 8(d) and it involves a large twisting of the N1-C6 bond with the dihedral angle of *d*(C2-N1-C6-H6) = −61.9°. Note that this dihedral angle takes negative value because the C2-N1-C6-H6 torsion in Figure 8(d) is in opposite direction (clockwise or counter-clockwise) to the N1-C6-C5-H5 and C2-N3-C4-N8 torsion in Figure 8(b) and (c), respectively. The excitation at $(^1\pi_{N1}\pi^*/gs)_{CI}$ is characterized by a single-excitation from the π orbital which has a large amplitude on the N1 atom, and therefore this excited state is labeled as $^1\pi_{N1}\pi^*$. The natural orbitals relevant to this excitation are given in the ESI (Figure S18).

As shown in Figure 8(b-d), the MS(4)-CASPT2(10,8) energies of the S_1 state at the three MECIs $(^1\pi\pi^*/gs)_{CI}$, $(^1\pi_{N3}\pi^*/gs)_{CI}$, and $(^1\pi_{N1}\pi^*/gs)_{CI}$ are calculated to be 4.44, 4.63, and 4.72 eV, respectively, from $(S_0)_{min}$. These energies are lower than the $^1\pi\pi^*$ vertical excitation energy at the $(S_0)_{min}$ structure (4.81 eV, see Table III), while they are higher than the $^1\pi\pi^*$ energy at $(^1\pi\pi^*)_{min}$ (4.29 eV).

Deactivation pathway to $(^1\pi\pi/gs)_{CI}$

The MS(4)-CASPT2(10,8) potential energy profiles from $(^1\pi\pi^*)_{min}$ to $(^1\pi\pi^*/gs)_{CI}$ along the LIIC points are given in Figure 9(b). The driving coordinate is chosen as *d*(N1-C6-C5-H5) for the C5-C6 twisting, and the highest energy structure is optimized at *d*(N1-C6-C5-H5) = 95.4º by the MS(2)-CASPT2(8,7) method. The figure shows that the $^1\pi\pi^*$ potential energy curve is very flat as the respective curve of the keto form. The S_1 state is always represented by the $^1\pi\pi^*$ excitation within the C5-C6 double bond along the decay path. The $^1\pi\pi^*$ energy of $(^1\pi\pi^*/gs)_{CI}$ is slightly higher than that of $(^1\pi\pi^*)_{min}$ by 0.15 eV. Also there is a small barrier separating these two structures and the barrier height is estimated to be ~0.22 eV. Since the energy of this barrier is substantially lower than the $^1\pi\pi^*$ energy at $(S_0)_{min}$, efficient deactivation through $(^1\pi\pi^*/gs)_{CI}$ is expected. The barrier height of ~0.22 eV is larger than that for the $(^1\pi\pi^*/gs)_{CI}$ decay path of the keto tautomer (~0.08 eV), while it is lower than the barrier for other decay paths of the enol tautomer (see below).

The MS-CASPT2 results shown above are well supported by the recent experimental work by Ho *et al.*[8], where they proposed that the deactivation of the enol form takes place through CIs in a

similar mechanism to the keto form but with a slightly higher barrier. In particular, the barrier height of ~0.15 eV that was estimated in Ref. [8] is in quite good agreement with our calculated value of ~0.22 eV.

Deactivation pathway to $(^1\pi_{N3}\pi^*/gs)_{CI}$

The MS-CASPT2 potential energy profiles for the decay path from $(^1\pi\pi^*)_{min}$ to $(^1\pi_{N3}\pi^*/gs)_{CI}$ are shown in Figure 9(c), where the driving coordinate is chosen as the dihedral angle of *d*(C2-N3-C4-N8) and the highest energy structure is optimized at *d*(C2-N3-C4-N8) = 98.1º. In analogy with the keto tautomer, the S_1 state exhibits a change of electronic character from $^1\pi\pi^*$ to $^1\pi_{N3}\pi^*$ along the decay path, and in between the $^1n_N\pi^*$ excitation is also involved. In particular, an avoided crossing between the S_1 and S_2 states is observed around *d*(C2-N3-C4-N8) = 130.0°, which exhibits the mixing of the $^1\pi\pi^*$ and $^1n_N\pi^*$ excitation. The S_1 electronic character at the MECI structure is well-characterized by the $^1\pi_{N3}\pi^*$ state. The avoided crossing indicates a barrier with the height of ~0.41 eV from $(^1\pi\pi^*)_{min}$, and the MECI $(^1\pi_{N3}\pi^*/gs)_{CI}$ exhibits higher energy than $(^1\pi\pi^*)_{min}$ by 0.34 eV. Since the S_1 energies of the barrier and MECI are higher than the corresponding values for the pathway to $(^1\pi\pi^*/gs)_{CI}$, the decay through $(^1\pi_{N3}\pi^*/gs)_{CI}$ would play a minor role as is speculated in the keto form.

Deactivation pathway to $(^1\pi_{N1}\pi^*/gs)_{CI}$

Figure 9(d) shows the MS-CASPT2 potential energy curves from $(^1\pi\pi^*)_{min}$ to $(^1\pi_{N1}\pi^*/gs)_{CI}$, with the driving coordinate defined by the dihedral angle of *d*(C2-N1-C6-H6) and the highest energy structure optimized at −97.3 degrees. As seen in the figure, the $^1\pi\pi^*$ energy almost monotonically increases toward $(^1\pi_{N1}\pi^*/gs)_{CI}$, which lies higher in energy than $(^1\pi\pi^*)_{min}$ by 0.43 eV. The barrier with the height of ~0.53 eV is also found near $(^1\pi_{N1}\pi^*/gs)_{CI}$, and its energy (4.82 eV) is slightly higher than the $^1\pi\pi^*$ vertical excitation energy. Therefore this pathway would also play a minor role in the deactivation process of the enol form.

D On-the-fly nonadiabatic MD simulations

The excited-state MD simulations are performed for the three tautomers of cytosine at the SA(4)-CASSCF(*m*,*n*)/DZP level, where the active space (*m*,*n*) is chosen as (12,9) for the keto and imino tautomers and (10,8) for the enol tautomer. A total of 80 (keto), 20 (imino), and 30 (enol) trajectories are launched from the lowest $^1\pi\pi^*$ state of each tautomer. As stated repeatedly, the SA-CASSCF method is not sufficient for a quantitative description of the potential energy surface. Even so, the results of MD simulations could serve to understand the dynamical behavior as long as the accuracy of the SA-CASSCF potential energies is carefully examined and contrasted to that of the MS-CASPT2 potential energies. The relevant SA-CASSCF energies at important structures are shown in the ESI (Figures S19-S21). For the imino tautomer, the SA-CASSCF and MS-CASPT2 potential energy profiles are qualitatively very similar. For the keto and enol tautomers, however, the two computational methods exhibit notable differences. In particular, the S_1 minimum energy structures in the keto and enol forms are characterized by the $^1n\pi^*$ state at the SA-CASSCF level, while they are in the $^1\pi\pi^*$ state at the MS-CASPT2 level. The reader should bear in mind the discrepancies between these two methods in the following discussion.

Figure 10 shows the time-dependent average population of the excited states for the three tautomers. The population is the sum of the S_1 and S_2 states (S_3 has no population), but most of the population switches to the S_1 state within 10 fs even if the molecule is photo-excited to the S_2 state at *t*

= 0 (in this case the lowest $^1\pi\pi^*$ state is the second excited state). As seen clearly in the figure, the imino form exhibits a quite efficient deactivation. It is found that all 20 trajectories deactivate via nonadiabatic transition around $(^1\pi_{N8}\pi^*/gs)'_{CI}$ involving the rotation of the imino group (the prime is added to the label to distinguish from the MS-CASPT2-optimized structures) and most of them (17 trajectories out of 20) switch to the ground state within 100 fs. The average of the transition time to the ground state over 20 trajectories is calculated to be 58 fs and the standard deviation is 35 fs. This rapid decay is expected since the reaction pathway toward $(^1\pi_{N8}\pi^*/gs)'_{CI}$ is barrierless and the rotation of the imino group is dominated by the motion of the hydrogen atom that has a very small mass. The MS-CASPT2 potential energy profiles also predict a barrierless decay path to $(^1\pi_{N8}\pi^*/gs)_{CI}$, as shown in Figure 7. Therefore, similar excited-state dynamics and decay time would be expected if the MD simulations were performed under the MS-CASPT2 potentials. No trajectories were observed for the deactivation via ethylene-like CI involving the twisting of the C5-C6 double bond in the CASSCF simulations, and it would be highly unlikely that trajectories of this kind appear in MS-CASPT2 simulations.

Figure 11 shows time evolution of potential energies of low-lying electronic states and geometrical parameters characterizing the motion of the imino group along a representative trajectory of imino cytosine. In this trajectory, a state switch from the S_1 to S_0 state is observed at 28.6 fs. It is clearly seen in Figure 11(b) that the dihedral angle of *d*(N3-C4-N8-H8) representing the twisting of the imino group monotonically decreases from d = ~180° and it exhibits d = 91.9° when the nonadiabatic transition takes place. After the transition, it keeps decreasing to d = ~0° until the end of the trajectory. The bond length of *r*(C4-N8) shown in the same figure indicates a sudden elongation of the bond upon the $^1\pi\pi^*$ excitation at t = 0, which is a typical motion in the excited-state dynamics involving the twisting of a double bond. Then it oscillates around 1.4 Å for a few cycles and exhibits r = 1.42 Å at 28.6 fs. These geometrical changes are totally consistent with the barrierless deactivation pathway leading to $(^1\pi_{N8}\pi^*/gs)_{CI}$ explored in the previous section.

The MD simulations of keto and enol cytosine suggest substantially longer excited-state population decay time compared to imino cytosine, as seen in Figure 10. The decay time of the keto form is in the subpicosecond range, where 76 % of the trajectories (61 out of 80) switch to the ground state within 500 fs. The enol form indicates even longer decay time, where only 10 % of the trajectories (3 out of 30) return to the ground state within 1.0 ps. The order of imino < keto < enol in decay time is apparently in agreement with the energetic order of the barrier height in the deactivation pathways calculated at the MS-CASPT2 level; the decay path of the imino form is barrierless, and the barrier heights of the keto and enol forms for the decay to $(^1\pi\pi^*/gs)_{CI}$ are ~0.08 and ~0.22 eV, respectively. However, the underlying excited-state dynamics of the keto and enol forms observed in the MD simulations is quite different from the dynamics expected from the MS-CASPT2 calculations because of the qualitatively different potential energy surfaces at the CASSCF level.

For the keto form, it is observed that 67 trajectories out of 80 decay through $(^1n_O\pi^*/gs)'_{CI}$ and that only 3 trajectories deactivate through $(^1\pi_{N3}\pi^*/gs)'_{CI}$ within 1.0 ps. The other 10 trajectories remains in the S_1 state until the MD simulations are terminated at 1.0 ps. Since the S_1 minimum energy structure corresponds to the $^1n_O\pi^*$ state at the SA-CASSCF level, the molecule tends to relax toward $(^1n_O\pi^*)'_{min}$ in the early dynamics. In the case of the decay through $(^1n_O\pi^*/gs)'_{CI}$, nonadiabatic transition to the ground state takes place without any significant out-of-plane deformation of the ring, as expected from the SA-CASSCF-optimized structure of this CI. Similar results were reported in other on-the-fly simulations based on the SA-CASSCF method.[34,36,38] However, these results should be challenged because the SA-CASSCF method is strongly biased to stabilize the decay pathway through $(^1n_O\pi^*/gs)'_{CI}$, as examined in Section III.A (see also Figure 5). Note that the MS-CASPT2 calculations favor the pathway through $(^1\pi\pi^*/gs)_{CI}$, as shown in Figure 4. The SA-CASSCF method predicts the highest energy of the barrier and MECI for the decay through $(^1\pi\pi^*/gs)'_{CI}$ among the three types of pathways (see Figure S19), which would explain why no trajectories through this CI were observed in the MD simulations.

In the enol form, two trajectories out of 30 deactivate through $(^1\pi_{N3}\pi^*/gs)'_{CI}$ accompanying the deformation of the amino group at around 500 fs, and one trajectory decays through $(^1\pi_{N1}\pi^*/gs)'_{CI}$ with the puckering of the N1 atom. These three trajectories quench to the S_0 state at around 500 fs, while the other 27 trajectories remain in the S_1 state at 1.0 ps. This slow population decay is expected from the SA-CASSCF potential energy profiles, where the relevant barrier and MECI for all three decay paths are located relatively high in energy in comparison to the keto and imino tautomers (Figure S21). The lowest energy among the three MECIs is at $(^1\pi_{N3}\pi^*/gs)'_{CI}$, and the barrier height from $(^1n\pi^*)'_{min}$ is only ~0.05 eV for the pathway to this MECI. The decay path involving $(^1\pi\pi^*/gs)_{CI}$ is preferred in the MS-CASPT2 calculations, but no trajectories were found for the decay through the corresponding CI in the MD simulations at the SA-CASSCF level possibly because the MECI energy and barrier height are the highest among the three pathways.

E Comparison with experimental findings.

In this subsection, we summarize the comparison of our theoretical results with experimental observations[7,8] from the viewpoint of the nonradiative decay mechanisms of cytosine tautomers. In particular, our results show a good consistency with the observations in the pump-probe photoionization experiments by Ho *et al*.[8] when we assume that the three tautomers coexist in the experimental condition.

In our calculations the vertical excitation energies to the first spectroscopic bright $^1\pi\pi^*$ states of the three tautomers differ by only ~0.3 eV (4.48 eV for keto, 4.67 eV for imino, and 4.81 eV for enol). We speculate that there is a significant spectral overlap among these three forms. The observed time-dependent decay signals of isolated cytosine would therefore reflect a mixture of the transients of these tautomers.

Even so, based on our results significant contribution from the imino tautomer can be ruled out from the time-resolved pump-probe spectra. Ho *et al*.[8] speculated that the imino form deactivates in a similar time scale to the keto form and thus both tautomers may contribute to the decay signal in the subpicosecond range. Our results suggest, however, that the imino form deactivates much more effectively than the keto form with the newly-found mechanism involving the twisting of the imino group. The excited-state MD simulation of imino cytosine predicts that the most of the population decay to the ground state within 100 fs. Therefore, we speculate that the signal from the imino form may be hard to detect experimentally due to its extremely fast decay and would possibly be buried in the initial spike of the time-resolved spectrum.

Moreover, decay signals of the keto and enol tautomers could be distinguished to some extent due to the different adiabatic excitation energies, which are calculated to be 3.98 and 4.29 eV, respectively (see Figures 3(a) and 8(a)). Because the contribution from the imino tautomer can be excluded in the way as discussed above, it is expected from our calculations that with the excitation energies between 3.98 and 4.29 eV, the observed signal would be assigned mainly to the keto form. Above 4.29 eV, all three tautomers could be excited, and the observed decay signal would be a mixture of the keto and enol forms.

In the experiment by Ho *et al*.,[8] a single-exponential decay with the lifetime of ~1 ps was observed at the excitation wavelengths of 300 nm (4.13 eV) and 290 nm (4.28 eV). According to the calculated adiabatic excitation energies, this decay signal would correspond to a nonradiative deactivation of the keto form. The photoreaction pathway leading to $(^1\pi\pi^*/gs)_{CI}$ is responsible for the deactivation, since this pathway exhibits the lowest energies of the barrier and CI among the three paths of keto cytosine examined in this work. The low barrier of ~0.08 eV for the $(^1\pi\pi^*/gs)_{CI}$ decay

path supports the very short excited-state lifetime.

For excitation wavelengths shorter than 280 nm (4.43 eV), Ho *et al.*[8] reported a double-exponential decay of the excited-state signal where a long-time component ranging from 3 to 45 ps appears in addition to the fast subpicosecond component. The long-time component should be assigned to the decay path from the enol form, because this component is not observed at the excitation energies below 4.29 eV. For the decay path from $(^1\pi\pi^*)_{min}$ to $(^1\pi\pi^*/gs)_{CI}$ of enol cytosine, in particular, the calculated barrier height of ~0.22 eV agrees quite well with the experimental estimates of ~0.15 eV based on the RRKM model. Therefore this pathway would be responsible for the observed slow component. With respect to the fast component, two possibilities can be considered for its assignment. First, it may correspond to the decay of the keto form as in the case of longer excitation wavelengths. This assignment may be supported by the lower barrier for the $(^1\pi\pi^*/gs)_{CI}$ decay path of keto cytosine compared to the corresponding path of enol cytosine. Second, the relaxation in the excited state from the FC region to $(^1\pi\pi^*)_{min}$ of the enol form may be responsible for the subpicosecond component, which would be followed by the long-time decay from $(^1\pi\pi^*)_{min}$ to CIs. We observe that enol cytosine exhibits a large out-of-plane deformation of the ring at $(^1\pi\pi^*)_{min}$. This deformation may lead to a rapid decrease of the FC factor upon the $^1\pi\pi^*$ excitation and thus contributes to the decay signal of the subpicosecond range. It would be difficult to distinguish these two contributions based on the current theoretical results. On-the-fly MD simulations using more accurate potential energies than CASSCF would be required to provide more decisive conclusion about the assignment of the subpicosecond component.

The pump-probe ionization spectrum by Kosma *et al.*[7] can also be interpreted basically in the same way. As mentioned in the Introduction, there is one notable difference from the experiments by Ho *et al.*[8]; Kosma *et al.*[7] observed another slow component of the decay with lifetime in the range from 19 up to ≥ 150 ps. They assigned this slow decay to the long-lived $^1n\pi^*$ (or $^1\pi_{N8}\pi^*$) excited state of the imino form populated by the keto-to-imino photo-tautomerization. However, our calculation of imino cytosine clearly excludes this possibility, because it predicts very efficient pathway for the deactivation of the respective excited state. Ho *et al.*[8] suggested that this long-lived component might be attributed to the signal from fragment ions produced by dissociative ionization of excited-state cytosine. It is noted that it does not exclude the possibility of the photo-tautomerization. The photo-tautomerization from the keto to imino form via the hydrogen-atom-transfer was observed at least in the matrix environments for cytosine[70] and also for 1-methylcytosine[71], which have been later supported by the recent theoretical calculations.[72] It is also noted in passing that the ring-opening reactions from keto cytosine to the conjugated isocyanate products were reported[70] (which is known as the Norrish type I (α-bond cleavage) photoreaction[73]). Better quantitative comparison with the experimental results must await the dynamical simulations including these possibilities.

IV. Conclusions and future directions.

We have revealed a comprehensive picture of the ultrafast nonradiative decay of the keto, imino, and enol tautomers of cytosine after photo-excitation to the lowest $^1\pi\pi^*$ state using quantum chemical methods. MS-CASPT2 calculations have been carried out to compare in detail the excited-state potential energy profiles for the deactivation pathways leading to S_1/S_0 CIs of the three tautomers. On-the-fly nonadiabatic MD simulations have also been performed to obtain a time-dependent picture of the decay processes. Our theoretical results strongly suggest the coexistence of more than one tautomer in the photophysics of isolated cytosine, as is supported by the recent experimental observations of femtosecond pump-probe photoionization spectra.[7,8]

In the keto form, three kinds of deactivation pathways involving $(^1\pi\pi^*/gs)_{CI}$, $(^1\pi_{N3}\pi^*/gs)_{CI}$,

and $(^1n_O\pi^*/gs)_{CI}$ have been investigated. The present MS-CASPT2 results suggest that the decay path through $(^1\pi\pi^*/gs)_{CI}$, which involves the ethylene-like twisting of the C5-C6 double bond, is energetically the most favorable among the three pathways. The barrier height for this decay is predicted to be only ~0.08 eV from $(^1\pi\pi^*)_{min}$. Therefore the deactivation via $(^1\pi\pi^*/gs)_{CI}$ would be responsible for the subpicosecond decay observed in experimental pump-probe photoionization spectra at the lowest excitation energy. The decay pathway through $(^1\pi_{N3}\pi^*/gs)_{CI}$, which involves the deformation of the amino group, exhibits a slightly higher barrier of ~0.18 eV. The other MECI $(^1n_O\pi^*/gs)_{CI}$ is calculated to be located quite high in energy, and therefore the decay though this CI would be highly unlikely. In this respect our MS-CASPT2 results contradict the deactivation scenario proposed in previous theoretical studies with the SA-CASSCF and SS-CASPT2 methods, where the decay path through $(^1n_O\pi^*/gs)_{CI}$ is the most favored.

For imino and enol cytosine, we have proposed new mechanisms of nonradiative decay on the basis of MS-CASPT2 potential energy profiles. In the imino form, we have found a barrierless decay pathway leading to $(^1\pi_{N8}\pi^*/gs)_{CI}$ which involves the rotation of the imino group. The MD simulation has revealed that most of the trajectories deactivate to the ground state within 100 fs, which leads us to conclude that the contribution of the imino form to the pump-probe spectra would be very small because of this extremely fast decay. In the enol form, the most efficient decay path has been found for the ethylene-like CI $(^1\pi\pi^*/gs)_{CI}$ involving the twisting of the C5-C6 double bond, as in the case of the keto form. The barrier height for this decay pathway is estimated to be ~0.22 eV, which is quite consistent with the decay time of up to tens of picoseconds observed in pump-probe spectroscopy at higher excitation energies. Other two CIs, $(^1\pi_{N3}\pi^*/gs)_{CI}$ and $(^1\pi_{N1}\pi^*/gs)_{CI}$ involving the puckering of an N atom, are located higher in energy than $(^1\pi\pi^*)_{min}$ by 0.33 eV and 0.42 eV, respectively, and the deactivation through these CIs would play a very minor role.

The present results also provide some important insights into the photostability of cytosine, particularly with respect to the imino and enol forms. For the enol tautomer, the newly proposed deactivation mechanism via ethylene-like CI predicts very short excited-state lifetime, but it would be still longer than the lifetime of the canonical keto form. This finding is consistent with the conjecture that the most photostable tautomer has been selected as the canonical form in each DNA base. For the imino tautomer, on the other hand, the newly-found decay path involving the rotation of the imino group suggests shorter excited-state lifetime than the keto form, which may imply that the imino form is more photostable than the canonical form in the gas phase. It is worth noting, however, that the deactivation via rotation of the imino group would be suppressed in DNA environment, because the imino group is likely to form a hydrogen bond with guanine, which may induce a substantial barrier for the rotation. By contrast, the decay pathway of the keto tautomer is predicted to be still very efficient even under hydrogen bonding with guanine.[64,74,75] It is also interesting to note that imino cytosine is hydrogen bonded with the enol tautomer of guanine in DNA, which would make the guanine-cytosine pair less stable in energy.[76] Of course, more realistic simulations in double-stranded DNA environments in solution phase will be needed to further validate the above conjecture.

The effects of substitutions on the excited-state dynamics in cytosine derivatives have been also the subject of many experimental studies. For example, Kohler *et al*. reported quite different excited-state lifetimes for several C5-substituted compounds of cytosine (5-fluorocytosine and 5-methylcytosine).[1,28] Very recently, Keane *et al*. investigated the effects of N1-substitution and observed the longer lifetimes of the excited-state intermediates for the nucleotide and nucleoside analogues than for 1-methylcytosine.[77] The role of these substitutions on the excited-state dynamics are not still fully understood, and as suggested in the present work, we believe the MS-CASPT2 treatment is required to accurately quantify the differences of the deactivation mechanisms between these cytosine derivatives.

Of course, further dynamical simulations would be required for unraveling the intricate

relaxation dynamics of DNA bases. Here we note again that the accuracy of potential energy surface is an essential factor responsible for a reliable description of the relaxation dynamics. The current work suggests that on-the-fly molecular dynamics simulation of DNA bases should be carried out using a highly correlated multireference method such as MS-CASPT2. However, such simulation is highly demanding and actually almost impractical for currently available computational resources. Computationally inexpensive approaches such as semiemprical, CASSCF, and TD-DFT are now mainly used for dynamical simulations, but the accuracy of these methods should be carefully examined beforehand. The CASSCF method is occasionally not accurate even qualitatively due to the lack of dynamic electron correlation as seen in the present study, and also the TD-DFT method has an intrinsic limitation to describe the potential energy surfaces around CIs. Recent on-the-fly excited-state dynamical simulations[34-38] for keto cytosine show consistency in predicting the decay times of the subpicosecond range, but they proposed different deactivation pathways as the main channel primarily due to the different potential energy surfaces employed in the simulations. In the present work, our simulations at the CASSCF level predict seemingly consistent excited-state lifetimes with experimental findings for the keto and enol tautomers (Figure 10), but the underlying deactivation mechanisms of these forms are completely different from the mechanisms predicted by the MS-CASPT2 calculations. These findings imply that dynamical simulation based on inaccurate potential energy surface may lead to serious misunderstanding of the photophysics of DNA bases. It may sometimes be necessary to reduce the dimensionality of the system for running the dynamical simulations. Even so, given the rapid advances in computational power, a comprehensive picture of the excited-state dynamics of DNA bases will be available soon in the near future with the close interplay between experiments and theoretical calculations.

V. Acknowledgments

A.N. acknowledges the financial support by KAKENHI (Grant-in-Aid for Scientific Research) and the Global COE Program (Project No. B01: Catalysis as the Basis for Innovation in Materials Science) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT, Japan). Y.H. and S.Y. thank the Japan Society for the Promotion of Science (JSPS) for the Research Fellowships for Young Scientists. Part of the calculations was performed on supercomputers at Research Center for Computational Science, Okazaki, Japan.

VI. Figure Caption

Figure 1 Molecular structure of cytosine tautomers.

Figure 2 Equilibrium structures of $(S_0)_{min}$ for three cytosine tautomers optimized by the MP2 calculation. The energies of the imino and enol forms are 0.027 and −0.074 eV, respectively, relative to the keto form. The bond lengths are given in units of Å.

Figure 3 Minimum energy structure of (a) $(^1\pi\pi^*)_{min}$ and MECI structures of (b) $(^1\pi\pi^*/gs)_{CI}$, (c) $(^1\pi_{N3}\pi^*/gs)_{CI}$, and (d) $(^1n_O\pi^*/gs)_{CI}$ for keto cytosine optimized by the MS-CASPT2 calculation. The bond lengths are given in units of Å. The potential energies of the first excited state are also included.

Figure 4 (a) MS(4)-CASPT2(12,9) potential energy profiles for low-lying electronic states of keto cytosine from $(S_0)_{min}$ to $(^1\pi\pi^*)_{min}$ using LIIC points. The $^1\pi\pi^*$ energies at $(S_0)_{min}$ and $(^1\pi\pi^*)_{min}$ are also shown. (b-d) MS(4)-CASPT2(12,9) potential energy profiles for

low-lying electronic states of keto cytosine from $(^1\pi\pi^*)_{min}$ to (b) $(^1\pi\pi^*/gs)_{CI}$, (c) $(^1\pi_{N3}\pi^*/gs)_{CI}$, and (d) $(^1n_O\pi^*/gs)_{CI}$ using LIIC points. Panels (b) and (c) are plotted as a function of the dihedral angles *d*(N1-C6-C5-H5) and *d*(C2-N3-C4-N8), respectively, representing out-of-plane deformation of the six-membered ring to reach the respective MECI. The vertical dotted line represents the dihedral angle where the optimization of the highest energy point along LIIC is performed (134.0 and 106.4 degrees for (b) and (c), respectively). The $^1\pi\pi^*$ energies of the $(^1\pi\pi^*)_{min}$, TS, and MECI structures are also included.

Figure 5 Potential energy profiles for low-lying electronic states of keto cytosine from $(^1\pi\pi^*)_{min}$ to $(^1n_O\pi^*/gs)'_{CI}$ calculated at the (a) SA(4)-CASSCF(12,9), (b) SS-CASPT2(12,9), and (c) MS(4)-CASPT2(12,9) levels using LIIC points.

Figure 6 MECI structure $(^1\pi_{N8}\pi^*/gs)_{CI}$ of imino cytosine optimized by the MS-CASPT2 calculation. The bond lengths are given in units of Å. The potential energy of the first excited state is also included.

Figure 7 MS(4)-CASPT2(12,9) potential energy profiles for low-lying electronic states of imino cytosine from $(S_0)_{min}$ to $(^1\pi_{N8}\pi^*/gs)_{CI}$ using LIIC points, plotted as a function of the dihedral angle *d*(N3-C4-N8-H8) representing the rotation of the imino group. The $^1\pi\pi^*$ energies at $(S_0)_{min}$ and $(^1\pi_{N8}\pi^*/gs)_{CI}$ are also shown..

Figure 8 Minimum energy structure of (a) $(^1\pi\pi^*)_{min}$ and MECI structures of (b) $(^1\pi\pi^*/gs)_{CI}$, (c) $(^1\pi_{N3}\pi^*/gs)_{CI}$, and (d) $(^1\pi_{N1}\pi^*/gs)_{CI}$ for enol cytosine optimized by the MS-CASPT2 calculation. The bond lengths are given in units of Å. The potential energies of the first excited state are also included.

Figure 9 (a) MS(4)-CASPT2(10,8) potential energy profiles for low-lying electronic states of enol cytosine from $(S_0)_{min}$ to $(^1\pi\pi^*)_{min}$ using LIIC points. The $^1\pi\pi^*$ energies at $(S_0)_{min}$ and $(^1\pi\pi^*)_{min}$ are also shown. (b-d) MS(4)-CASPT2(10,8) potential energy profiles for low-lying electronic states of enol cytosine from $(^1\pi\pi^*)_{min}$ to (b) $(^1\pi\pi^*/gs)_{CI}$, (c) $(^1\pi_{N3}\pi^*/gs)_{CI}$, and (d) $(^1\pi_{N1}\pi^*/gs)_{CI}$ using LIIC points. Panels (b-d) are plotted as a function of the dihedral angles *d*(N1-C6-C5-H5), *d*(C2-N3-C4-N8), and *d*(C2-N1-C6-H6), respectively, representing out-of-plane deformation of the six-membered ring to reach the respective MECI. The vertical dotted line represents the dihedral angle where the optimization of the highest energy point along LIIC is performed (95.4, 98.1, and −97.3 degrees for (b), (c), and (d), respectively). The $^1\pi\pi^*$ energies of the $(^1\pi\pi^*)_{min}$, TS, and MECI structures are also included.

Figure 10 Time-dependent average population of the excited states (sum of S_1 and S_2 states) for three tautomers of cytosine, estimated from on-the-fly nonadiabatic MD simulation at the SA-CASSCF level.

Figure 11 Time evolution of (a) SA(4)-CASSCF(12,9) potential energies (in eV) and (b) dihedral angle *d*(N3-C4-N8-H8) (in degrees) and bond length *r*(C4-N8) (in Å) of imino cytosine along a representative trajectory of nonadiabatic MD simulations. The thick line with circles represents the current state at each time step and the vertical dotted line indicates the time at which the hopping event from S_1 to S_0 state takes place (28.6 fs).

VII. Associated Content

Electronic supplementary information (ESI) available

Equilibrium structures for higher-energy rotamers of imino and enol tautomers. Active orbitals of three cytosine tautomers at $(S_0)_{min}$ and relevant natural orbitals at CIs. The S_1 minimum energy structure of keto cytosine. The MEP from $(S_0)_{min}$ for keto and enol tautomers. The potential energy profiles from $(S_0)_{min}$ to $(^1n_O\pi^*)'_{min}$ for keto cytosine. The configuration interaction coefficients for S_0, S_1, and S_2 states, and eigenvectors of the effective Hamiltonian matrix at $(^1n_O\pi^*/gs)_{CI}$. The potential energy profiles from $(^1\pi\pi^*)_{min}$ to $(^1\pi\pi^*/gs)'_{CI}$ at the different levels of theory for keto cytosine. The MECI structure $(^1\pi\pi^*/gs)'_{CI}$ of imino cytosine. The minimum energy structure of enol cytosine in the $^1n\pi^*$ state determined by SA(4)-CASSCF(10,8) method, $(^1n\pi^*)'_{min}$. The SA(4)-CASSCF(12,9) electronic energies of the $(S_0)_{min}$, $(S_1)_{min}$ = $(^1n_O\pi^*)'_{min}$, TS, and MECI structures for keto cytosine. The SA(4)-CASSCF(12,9) electronic energies of the $(S_0)_{min}$, $(S_1)_{min}$ = $(^1\pi_{N8}\pi^*)'_{min}$, and $(^1\pi_{N8}\pi^*/gs)_{CI}$ structures for imino cytosine. The SA(4)-CASSCF(10,8) electronic energies of the $(S_0)_{min}$, $(S_1)_{min}$ = $(^1n\pi^*)'_{min}$, TS, and MECI structures for enol cytosine.

VIII. Author Information

Corresponding Author

*akira-n@sci.hokudai.ac.jp

Present Addresses

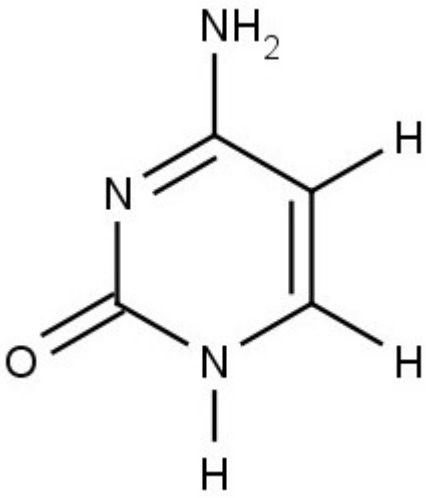
†Department of Frontier Materials Chemistry, Graduate School of Science and Technology, Hirosaki University, Hirosaki 036-8561, Japan.

IX. References

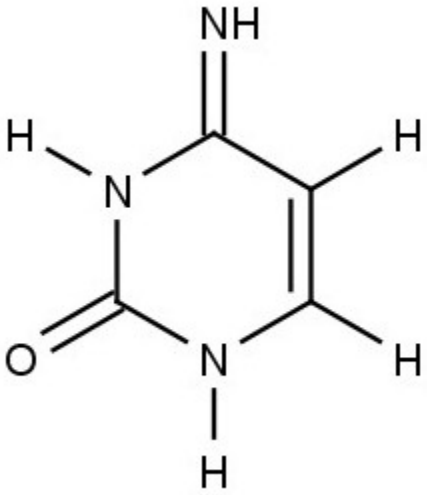
[1] C. E. Crespo-Hernández, B. Cohen, P. M. Hare and B. Kohler, Chem. Rev. **104**, 1977 (2004).
[2] B. Kohler, J. Phys. Chem. Lett. **1**, 2047 (2010).
[3] *Conical Intersections: Theory, Computation and Experiment*. edited by W. Domcke, D. R. Yarkony and H. Köppel (World Scientific, Singapore, 2011).
[4] H. Kang, K. T. Lee, B. Jung, Y. J. Ko and S. K. Kim, J. Am. Chem. Soc. **124**, 12958 (2002).
[5] S. Ullrich, T. Schultz, M. Z. Zgierski and A. Stolow, Phys. Chem. Chem. Phys. **6**, 2796 (2004).
[6] C. Canuel, M. Mons, F. Piuzzi, B. Tardivel, I. Dimicoli and M. Elhanine, J. Chem. Phys. **122**, 074316 (2005).
[7] K. Kosma, C. Schröter, E. Samoylova, I. V. Hertel and T. Schultz, J. Am. Chem. Soc. **131**, 16939 (2009).
[8] J.-W. Ho, H.-C. Yen, W.-K. Chou, C.-N. Weng, L.-H. Cheng, H.-Q. Shi, S.-H. Lai and P.-Y. Cheng, J. Phys. Chem. A **115**, 8406 (2011).
[9] R. D. Brown, P. D. Godfrey, D. McNaughton and A. P. Pierlot, J. Am. Chem. Soc. **111**, 2308 (1989).
[10] E. Nir, M. Müller, L. I. Grace and M. S. de Vries, Chem. Phys. Lett. **355**, 59 (2002).
[11] G. Fogarasi, J. Phys. Chem. A **106**, 1381 (2002).
[12] S. A. Trygubenko, T. V. Bogdan, M. Rueda, M. Orozco, F. J. Luque, J. Sponer, P. Slavicek and P. Hobza, Phys. Chem. Chem. Phys. **4**, 4192 (2002).
[13] J. R. B. Gomes, M. D. M. C. Ribeiro da Silva, V. L. S. Freitas and M. A. V. Ribeiro da Silva, J. Phys. Chem. A **111**, 7237 (2007).
[14] O. Kostko, K. Bravaya, A. Krylov and M. Ahmed, Phys. Chem. Chem. Phys. **12**, 2860 (2010).
[15] V. Feyer, O. Plekan, R. Richter, M. Coreno, M. de Simone, K. C. Prince, A. B. Trofimov, I. L. Zaytseva and J. Schirmer, J. Phys. Chem. A **114**, 10270 (2010).

[16] G. Bazsó, G. Tarczay, G. Fogarasi and P. G. Szalay, Phys. Chem. Chem. Phys. **13**, 6799 (2011).
[17] J. L. Alonso, V. Vaquero, I. Peña, J. C. López, S. Mata and W. Caminati, Angew. Chem. Int. Ed. **52**, 2331 (2013).
[18] Z. Yang and M. T. Rodgers, Phys. Chem. Chem. Phys. **6**, 2749 (2004).
[19] J. R. Sambrano, A. R. de Souza, J. J. Queralt and J. Andrés, Chem. Phys. Lett. **317**, 437 (2000).
[20] T. Ueda and J. J. Fox, J. Am. Chem. Soc. **85**, 4024 (1963).
[21] M. Dreyfus, O. Bensaude, G. Dodin and J. E. Dubois, J. Am. Chem. Soc. **98**, 6338 (1976).
[22] K. Tomić, J. Tatchen and C. M. Marian, J. Phys. Chem. A **109**, 8410 (2005).
[23] N. Ismail, L. Blancafort, M. Olivucci, B. Kohler and M. A. Robb, J. Am. Chem. Soc. **124**, 6818 (2002).
[24] M. Merchán and L. Serrano-Andrés, J. Am. Chem. Soc. **125**, 8108 (2003).
[25] L. Blancafort and M. A. Robb, J. Phys. Chem. A **108**, 10609 (2004).
[26] M. Z. Zgierski, S. Patchkovskii, T. Fujiwara and E. C. Lim, J. Phys. Chem. A **109**, 9384 (2005).
[27] M. Z. Zgierski, S. Patchkovskii and E. C. Lim, J. Chem. Phys. **123**, 081101 (2005).
[28] L. Blancafort, B. Cohen, P. M. Hare, B. Kohler and M. A. Robb, J. Phys. Chem. A **109**, 4431 (2005).
[29] M. Merchán, R. González-Luque, T. Climent, L. Serrano-Andrés, E. Rodríguez, M. Reguero and D. Peláez, J. Phys. Chem. B **110**, 26471 (2006).
[30] K. A. Kistler and S. Matsika, J. Phys. Chem. A **111**, 2650 (2007).
[31] L. Blancafort, Photochem. Photobiol. **83**, 603 (2007).
[32] K. A. Kistler and S. Matsika, J. Chem. Phys. **128**, 215102 (2008).
[33] R. González-Luque, T. Climent, I. González-Ramírez, M. Merchán and L. Serrano-Andrés, J. Chem. Theo. Comp. **6**, 2103 (2010).
[34] H. R. Hudock and T. J. Martínez, ChemPhysChem **9**, 2486 (2008).
[35] Z. Lan, E. Fabiano and W. Thiel, J. Phys. Chem. B **113**, 3548 (2009).
[36] J. González-Vázquez and L. González, ChemPhysChem **11**, 3617 (2010).
[37] A. N. Alexandrova, J. C. Tully and G. Granucci, J. Phys. Chem. B **114**, 12116 (2010).
[38] M. Barbatti, A. J. A. Aquino, J. J. Szymczak, D. Nachtigallova and H. Lischka, Phys. Chem. Chem. Phys. **13**, 6145 (2011).
[39] M. K. Shukla and J. Leszczynski, J. Phys. Chem. A **106**, 11338 (2002).
[40] R. I. Bakalska and V. B. Delchev, J. Mol. Model. **18**, 5133 (2012).
[41] B. O. Roos and K. Andersson, Chem. Phys. Lett. **245**, 215 (1995).
[42] http://setani.sci.hokudai.ac.jp/sapporo/
[43] MOLPRO, version 2008.1, a package of *ab initio* programs, H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz and others
[44] L. Serrano-Andrés and M. Merchán, J. Mol. Struc. THEOCHEM **729**, 99 (2005).
[45] S. Yamazaki and T. Taketsugu, J. Phys. Chem. A **116**, 491 (2012).
[46] J. Finley, P.-Å. Malmqvist, B. O. Roos and L. Serrano-Andrés, Chem. Phys. Lett. **288**, 299 (1998).
[47] K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej and K. Wolinski, J. Phys. Chem. **94**, 5483 (1990).
[48] P. Celani and H.-J. Werner, J. Chem. Phys. **112**, 5546 (2000).
[49] L. Serrano-Andrés, M. Merchán and R. Lindh, J. Chem. Phys. **122**, 104107 (2005).
[50] B. G. Levine, J. D. Coe and T. J. Martínez, J. Phys. Chem. B **112**, 405 (2007).
[51] J. C. Tully, J. Chem. Phys. **93**, 1061 (1990).
[52] R. Kobayashi, J. Phys. Chem. A **102**, 10813 (1998).
[53] R. Abouaf, J. Pommier, H. Dunet, P. Quan, P.-C. Nam and M. T. Nguyen, J. Chem. Phys. **121**, 11668 (2004).
[54] L. B. Clark and I. Tinoco, Jr., J. Am. Chem. Soc. **87**, 11 (1965).
[55] H. Morita and S. Nagakura, Theor. Chim. Acta **11**, 279 (1968).
[56] F. Žaloudek, J. S. Novros and L. B. Clark, J. Am. Chem. Soc. **107**, 7344 (1985).

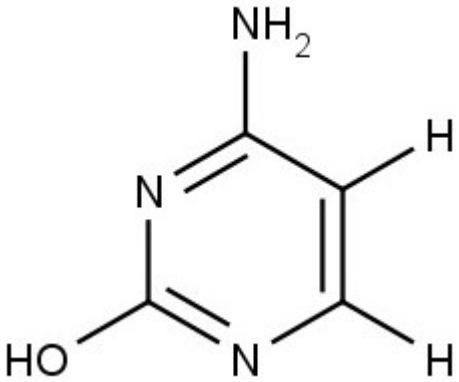
[57] A. Tajti, G. Fogarasi and P. G. Szalay, ChemPhysChem **10**, 1603 (2009).
[58] M. Valiev and K. Kowalski, J. Chem. Phys. **125**, 211101 (2006).
[59] M. P. Fülscher and B. O. Roos, J. Am. Chem. Soc. **117**, 2089 (1995).
[60] E. Nir, I. Hunig, K. Kleinermanns and M. S. de Vries, Phys. Chem. Chem. Phys. **5**, 4780 (2003).
[61] L. Serrano-Andrés, M. Merchán and A. C. Borin, Proc. Natl. Acad. Sci. U.S.A. **103**, 8691 (2006).
[62] L. Serrano-Andrés, M. Merchán and A. C. Borin, J. Am. Chem. Soc. **130**, 2473 (2008).
[63] L. Serrano-Andrés and M. Merchán, J. Photochem. Photobiol. C **10**, 21 (2009).
[64] A. L. Sobolewski and W. Domcke, Phys. Chem. Chem. Phys. **6**, 2763 (2004).
[65] R. J. Malone, A. M. Miller and B. Kohler, Photochem. Photobiol. **77**, 158 (2003).
[66] O. Weingart, A. Migani, M. Olivucci, M. A. Robb, V. Buss and P. Hunt, J. Phys. Chem. A **108**, 4685 (2004).
[67] P. A. Hunt and M. A. Robb, J. Am. Chem. Soc. **127**, 5720 (2005).
[68] S. Yamazaki and W. Domcke, J. Phys. Chem. A **112**, 7090 (2008).
[69] M. J. Bearpark, M. A. Robb and H. B. Schlegel, Chem. Phys. Lett. **223**, 269 (1994).
[70] L. Lapinski, I. Reva, M. J. Nowak and R. Fausto, Phys. Chem. Chem. Phys. **13**, 9676 (2011).
[71] I. Reva, M. J. Nowak, L. Lapinski and R. Fausto, J. Phys. Chem. B **116**, 5703 (2012).
[72] Q. Li and L. Blancafort, Photochem. Photobiol. Sci. (2013).
[73] E. W. G. Diau, C. Kötting and A. H. Zewail, ChemPhysChem **2**, 273 (2001).
[74] S. Yamazaki and T. Taketsugu, Phys. Chem. Chem. Phys. **14**, 8866 (2012).
[75] T. Zelený, M. Ruckenbauer, A. J. A. Aquino, T. Müller, F. Lankaš, T. Dršata, W. L. Hase, D. Nachtigallova and H. Lischka, J. Am. Chem. Soc. **134**, 13662 (2012).
[76] A. Abo-Riziq, L. Grace, E. Nir, M. Kabelac, P. Hobza and M. S. de Vries, Proc. Natl. Acad. Sci. U.S.A. **102**, 20 (2005).
[77] P. M. Keane, M. Wojdyla, G. W. Doorley, G. W. Watson, I. P. Clark, G. M. Greetham, A. W. Parker, M. Towrie, J. M. Kelly and S. J. Quinn, J. Am. Chem. Soc. **133**, 4212 (2011).



keto (amino-keto)

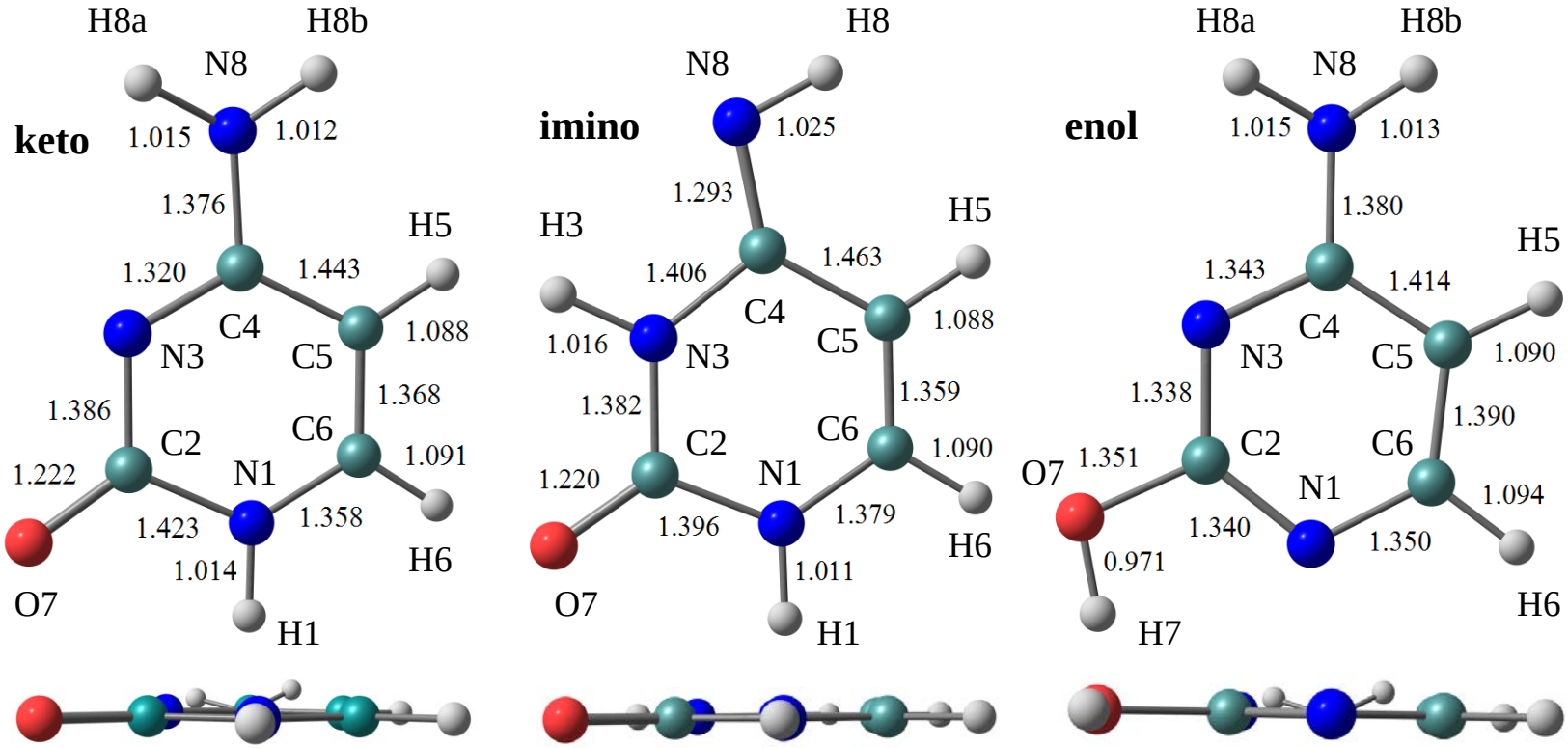


imino (imino-keto)

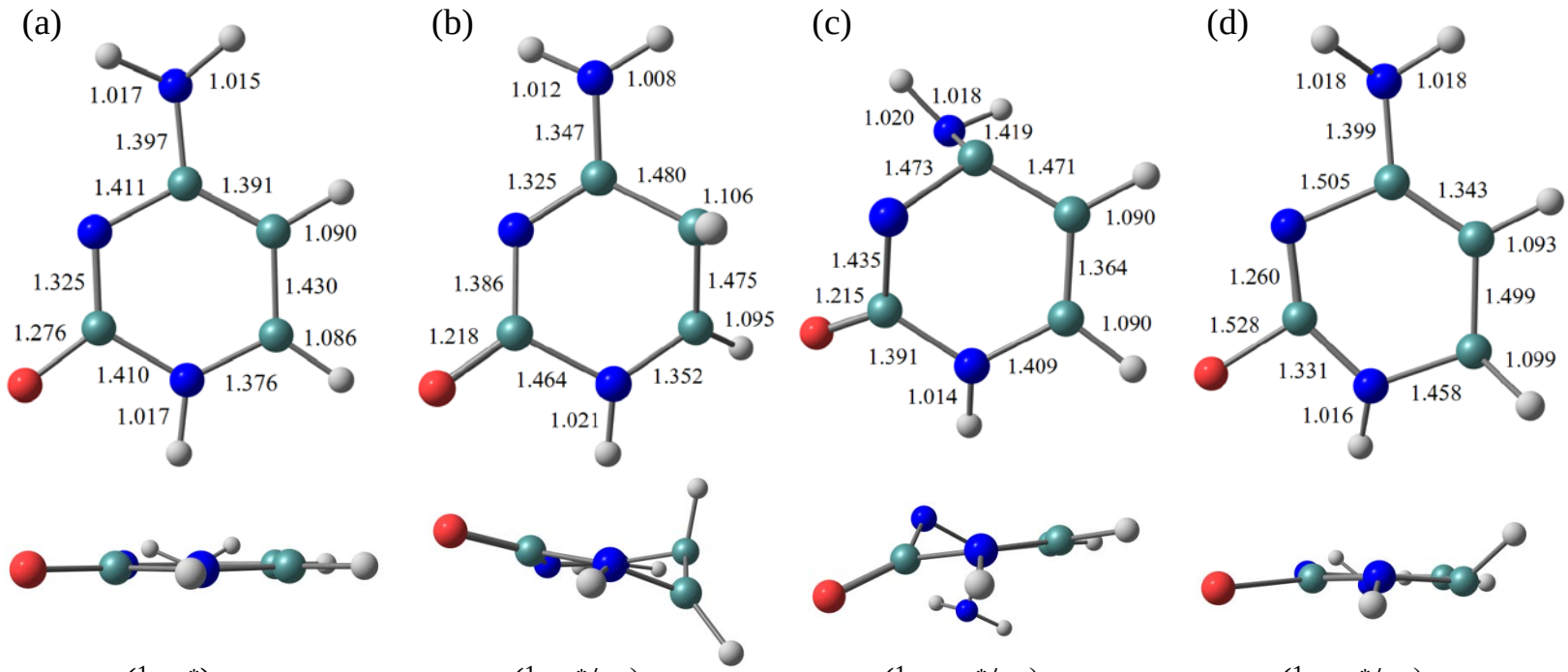


enol (amino-enol)

Nakayama, Fig. 1



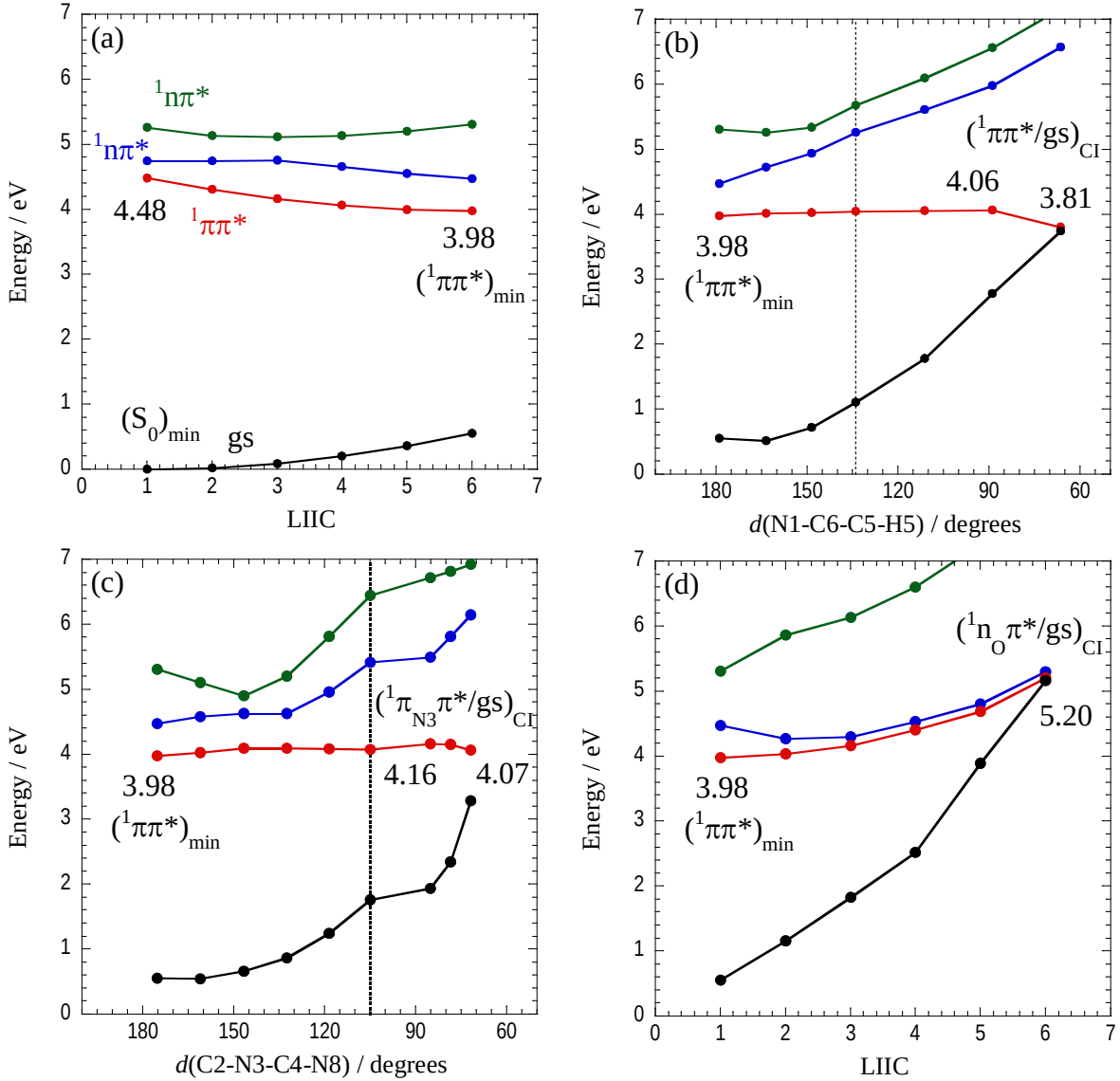
Nakayama, Fig. 2



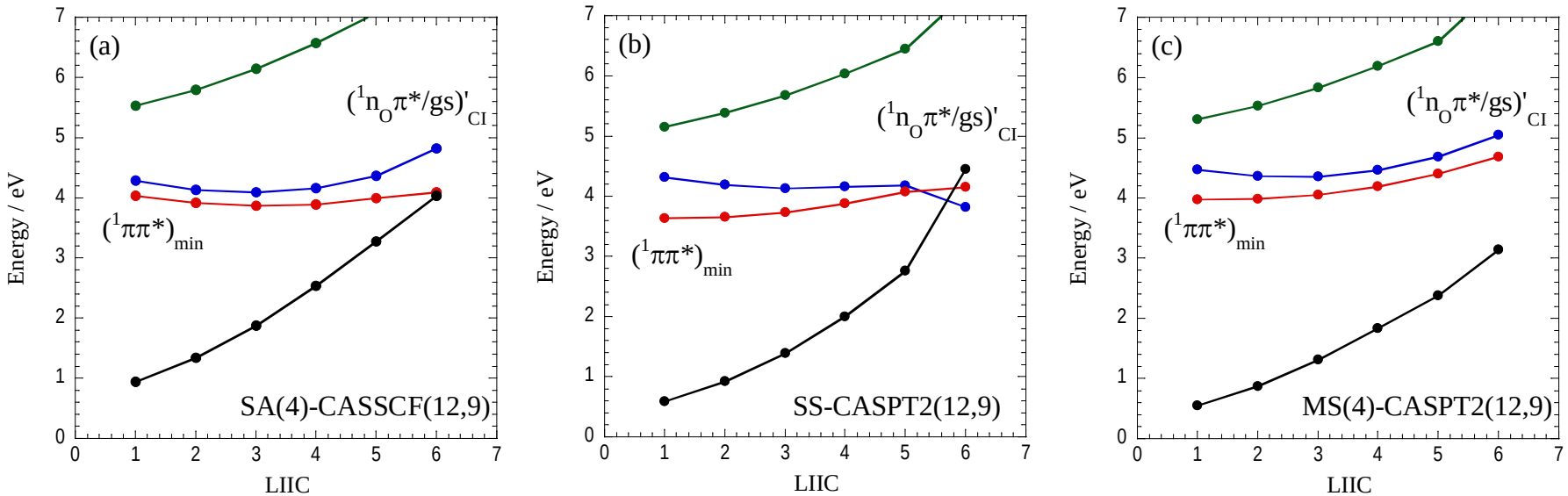
$(^1\pi\pi^*)_{min}$ $(^1\pi\pi^*/gs)_{CI}$ $(^1\pi_{N3}\pi^*/gs)_{CI}$ $(^1n_O\pi^*/gs)_{CI}$

$E(^1\pi\pi^*) = 3.98$ eV $E(^1\pi\pi^*) = 3.80$ eV $E(^1\pi_{N3}\pi^*) = 4.07$ eV $E(^1n_O\pi^*) = 5.20$ eV

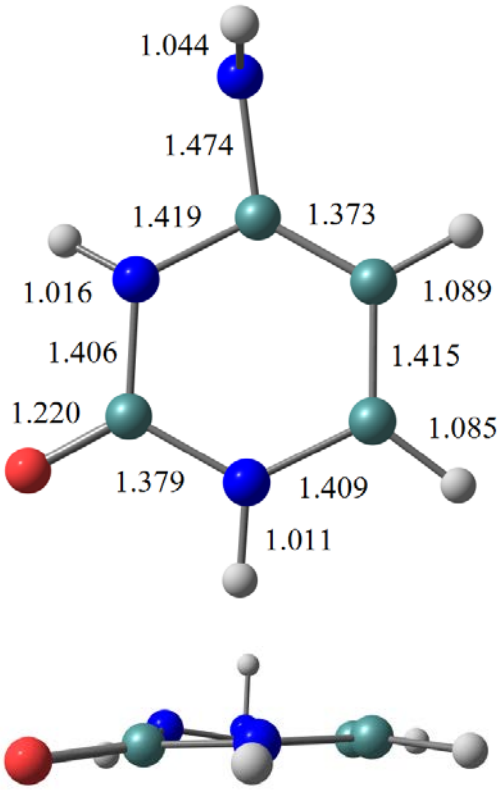
Nakayama, Fig. 3



Nakayama, Fig. 4



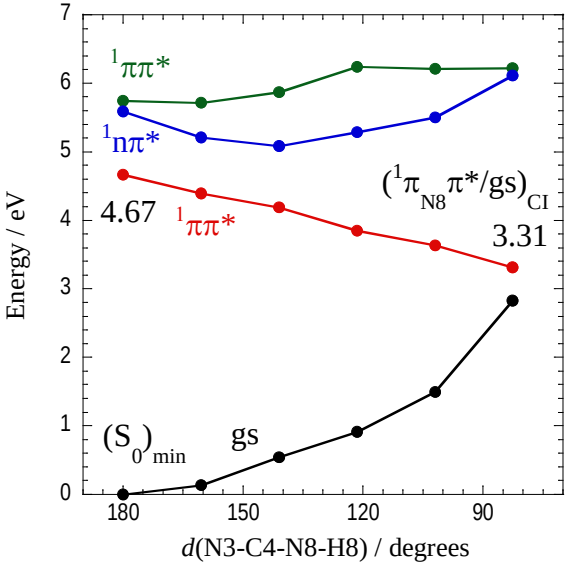
Nakayama, Fig. 5



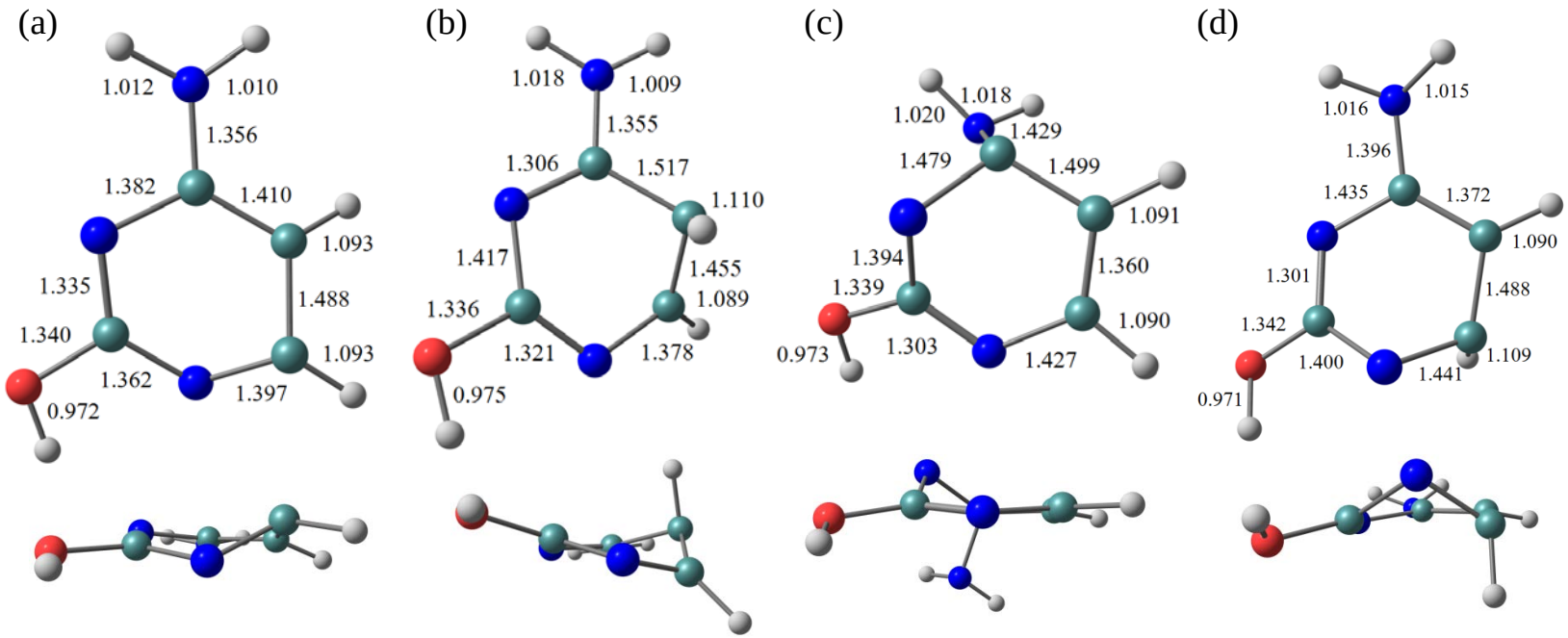
$({}^1\pi_{N8}\pi^*/gs)_{CI}$

$E({}^1\pi_{N8}\pi^*) = 3.31$ eV

Nakayama, Fig. 6



Nakayama, Fig. 7



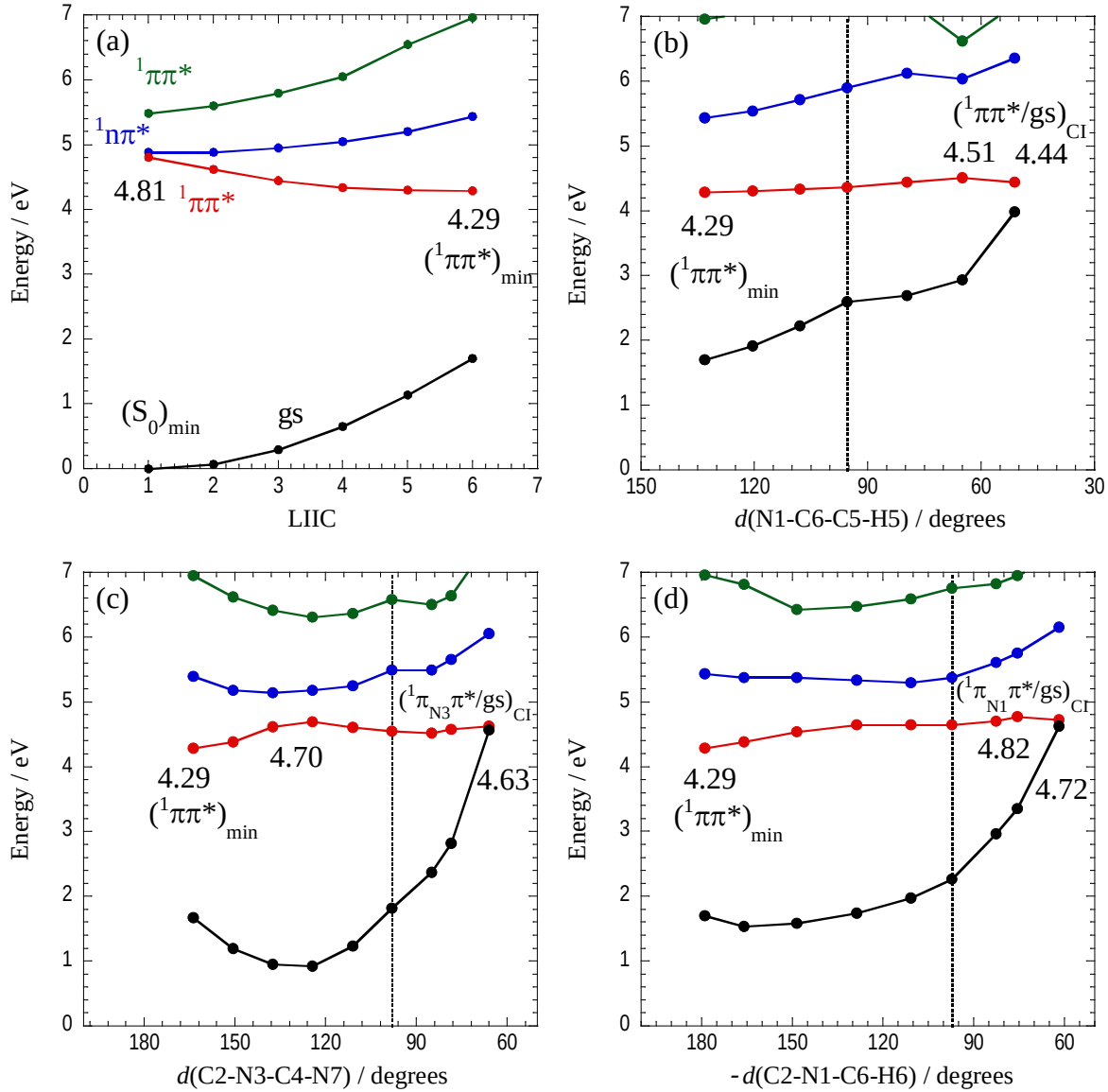
$(^1\pi\pi^*)_{min}$ $E(^1\pi\pi^*) = 4.29$ eV

$(^1\pi\pi^*/gs)_{CI}$ $E(^1\pi\pi^*) = 4.44$ eV

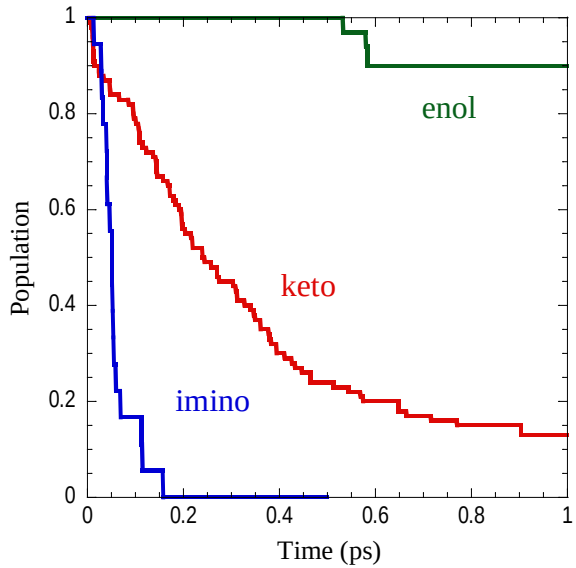
$(^1\pi_{N3}\pi^*/gs)_{CI}$ $E(^1\pi_{N3}\pi^*) = 4.63$ eV

$(^1\pi_{N1}\pi^*/gs)_{CI}$ $E(^1\pi_{N1}\pi^*) = 4.72$ eV

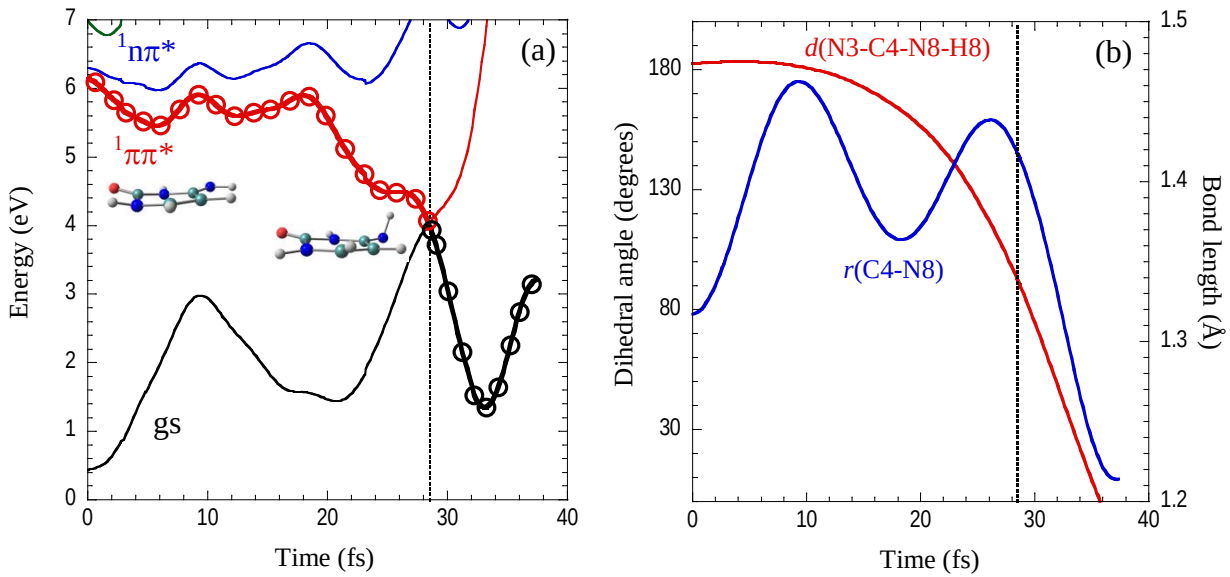
Nakayama, Fig. 8



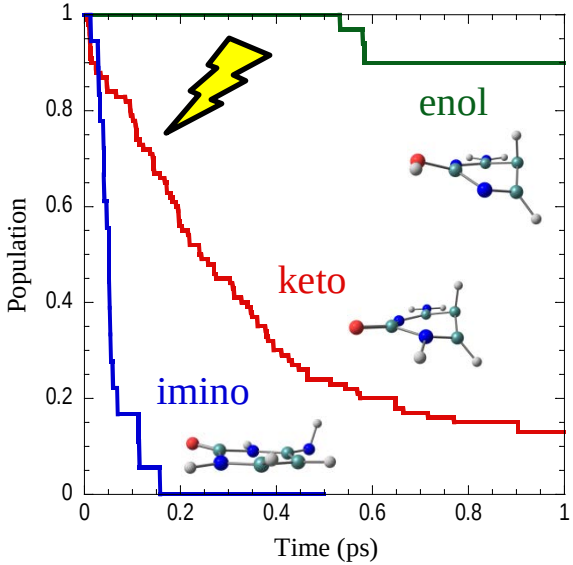
Nakayama, Fig. 9



Nakayama, Fig. 10



Nakayama, Fig. 11



Nakayama, Graphical Abstract

Photophysics of Cytosine Tautomers: New Insights into the Nonradiative Decay Mechanisms from *Ab Initio* Potential Energy Calculations and Excited-state Molecular Dynamics Simulations

Akira Nakayama*, Yu Harabuchi, Shohei Yamazaki, and Tetsuya Taketsugu

Department of Chemistry, Faculty of Science, Hokkaido University
Sapporo 060-0810, Japan

Supporting Information

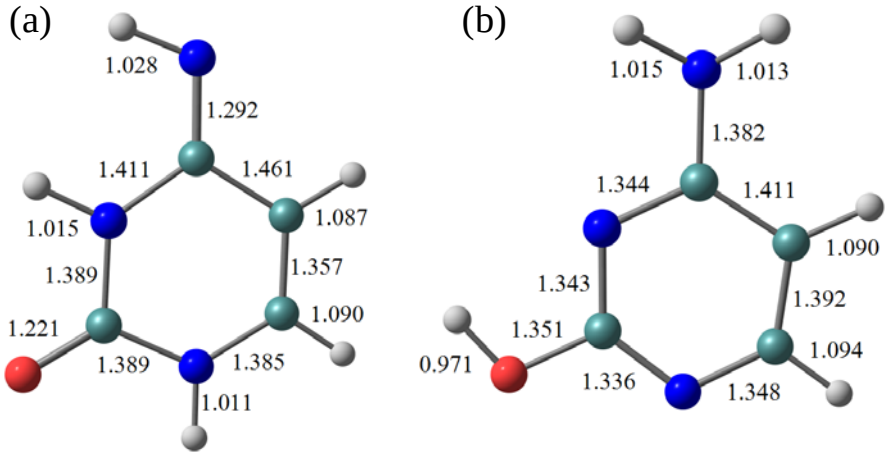


Figure S1. Equilibrium structure in the S_0 state for higher-energy rotamers of (a) imino and (b) enol tautomers. The energies of the imino and enol forms are 0.097 and −0.045 eV relative to the keto form, respectively. The bond lengths are given in units of Å.

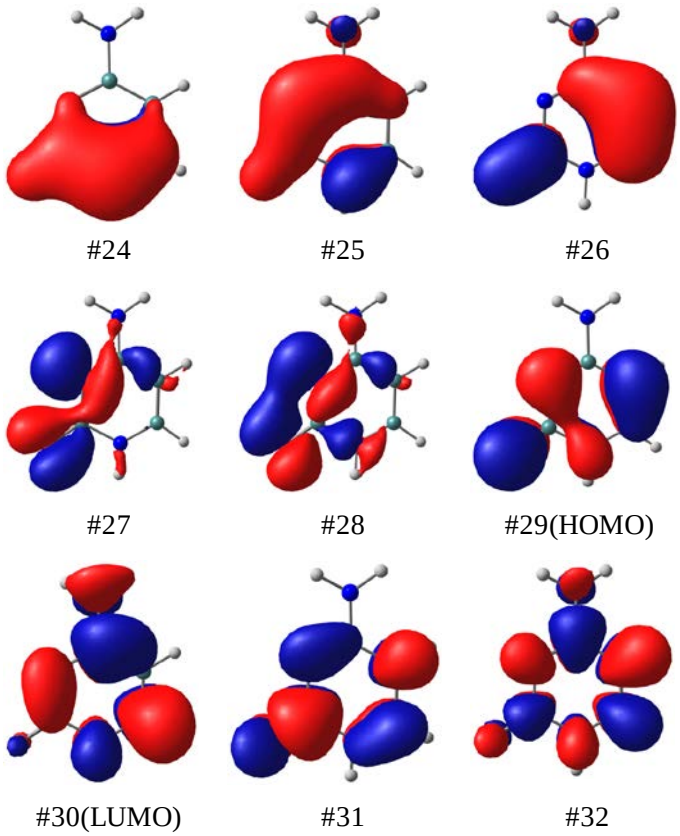


Figure S2. Active orbitals of keto cytosine at $(S_0)_{min}$ in the SA(4)-CASSCF(12,9) calculation.

S_1 minimum energy structure of keto cytosine

When we perform the geometry optimization of keto cytosine in the S_1 state at the SA(2)-CASSCF(12,9) level starting from $(S_0)_{min}$, it is found that the S_1 minimum energy structure is characterized by $^1n_O\pi^*$. It involves the bond inversion of the ring and the elongation of the C2-O7 bond length from 1.222 to 1.362 Å, and also the H1 atom is displaced from the molecular plane. Hereafter this structure is denoted by $(^1n_O\pi^*)'_{min}$ and its structure is given in Figure S3(a). The prime indicates that the structure is optimized by the SA-CASSCF method. The geometry optimization in the $^1\pi\pi^*$ state is also carried out by SA(2)-CASSCF(8,7), where the active space is comprised of only π orbitals. The π orbital that is localized on the N8 atom is excluded from the active space. The optimized geometry is shown in Figure S3(b) and denoted by $(^1\pi\pi^*)'_{min}$. In contrast to $(^1n_O\pi^*)'_{min}$, planarity of the molecule is maintained. The geometrical change from the $(S_0)_{min}$ structure involves the bond inversion of the ring and also the elongation of C2-O7 bond length from 1.222 to 1.327 Å.

The SA(4)-CASSCF(12,9) energy of the S_1 state is 3.64 and 3.81 eV at $(^1n_O\pi^*)'_{min}$ and $(^1\pi\pi^*)'_{min}$, respectively, suggesting that the former exhibits the S_1 global minimum at the SA-CASSCF level. However, the MS(4)-CASPT2(12,9) calculation at the same structures predicts that the S_1 energies at $(^1n_O\pi^*)'_{min}$ and $(^1\pi\pi^*)'_{min}$ relative to $(S_0)_{min}$ are 4.22 and 4.10 eV, respectively, and it is found that $(^1\pi\pi^*)'_{min}$ lies lower than $(^1n_O\pi^*)'_{min}$ at the MS-CASPT2 level.

The optimized structure at the MS(2)-CASPT2(8,7) level is shown in Figure 3(a) in the main document and denoted by $(^1\pi\pi^*)_{min}$. The notable differences between $(^1\pi\pi^*)_{min}$ and $(^1\pi\pi^*)'_{min}$ are the bond lengths of N1-C2, C2-N3, and C2-O7. The MS(4)-CASPT2(12,9) energy at $(^1\pi\pi^*)_{min}$ is calculated to be 3.98 eV relative to $(S_0)_{min}$. We also tried to locate the minimum energy structure in the $^1n_O\pi^*$ state at the MS(3)-CASPT2(6,5) level, starting from the $(^1n_O\pi^*)'_{min}$ structure. The active space is comprised of the four π orbitals (two occupied and two unoccupied) and one lone-pair orbital. The S_1 state at $(^1n_O\pi^*)'_{min}$ is characterized by the $^1n_O\pi^*$ state at MS(3)-CASPT2(6,5) level, but the geometry optimization in the S_1 state leads to the close proximity to $(^1\pi\pi^*)_{min}$ and the S_1 state becomes the $^1\pi\pi^*$ state. Therefore we conclude that the S_1 minimum is characterized by the $^1\pi\pi^*$ state at the MS-CASPT2 level.

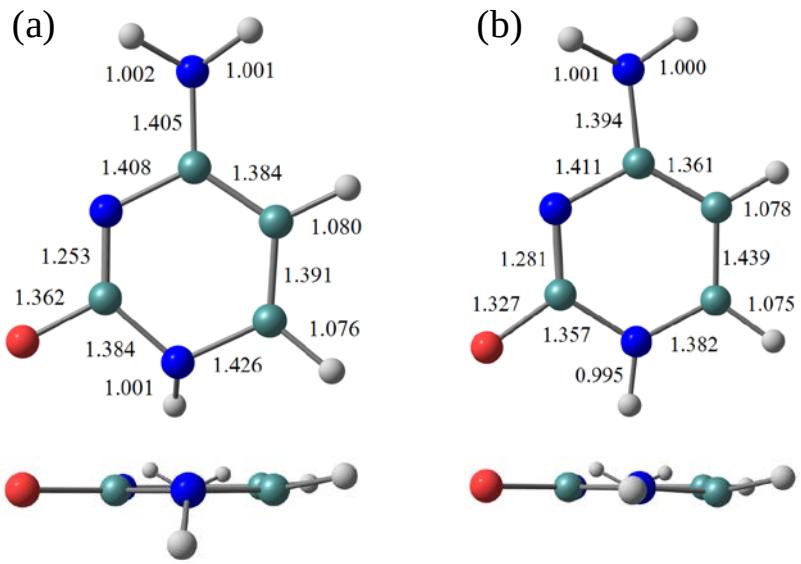


Figure S3. Minimum energy structures of keto cytosine (a) in the $^1n_O\pi^*$ state determined by SA(2)-CASSCF(12,9) method, $(^1n_O\pi^*)'_{min}$, and (b) in the $^1\pi\pi^*$ state determined by SA(2)-CASSCF(8,7) method, $(^1\pi\pi^*)'_{min}$. The bond lengths are given in units of Å.

MEP from $(S_0)_{min}$ for keto cytosine

The MEP computation of keto cytosine in the $^1\pi\pi^*$ state from $(S_0)_{min}$ was performed at the MS(2)-CASPT2(8,7) level of theory with a stepsize of 0.05 bohr·amu$^{1/2}$ and the energetics are reproduced at the MS(4)-CASPT2(12,9) level at selected points along MEP. The MEP calculation was terminated when the plateau region was observed, which was at 1.0 bohr·amu$^{1/2}$ in this case.

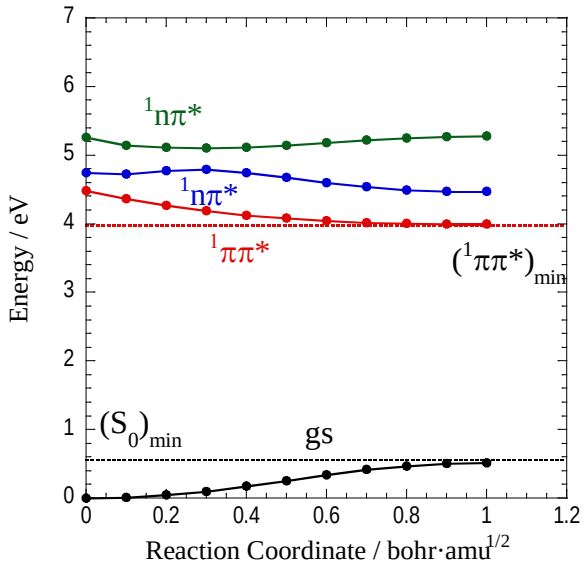


Figure S4. Potential energy profiles of keto cytosine from $(S_0)_{min}$ along the MEP coordinates at the MS(4)-CASPT2(12,9) level. The horizontal dotted lines represent the potential energies of the ground and $^1\pi\pi^*$ states at $(^1\pi\pi^*)_{min}$.

Potential energy profiles from $(S_0)_{min}$ to $(^1n_O\pi^*)'_{min}$

The MS(4)-CASPT2(12,9) potential energy profiles from $(S_0)_{min}$ to $(^1n_O\pi^*)'_{min}$ along LIIC are shown in Figure S5. Clearly it exhibits a minimum before accessing to $(^1n_O\pi^*)'_{min}$ and the electronic structure around this minimum is characterized by the $^1\pi\pi^*$ state. After passing through this minimum, the $^1\pi\pi^*$ and $^1n\pi^*$ states start to mix and then it reaches to $(^1n_O\pi^*)'_{min}$. This behavior is expected since the C2-O7 bond length at $(^1n_O\pi^*)'_{min}$ is longer than that at $(^1\pi\pi^*)_{min}$, and also it involves the out-of-plane displacement of the H1 atom in order to reach $(^1n_O\pi^*)'_{min}$. From these results, it is expected that after photoexcitation to the $^1\pi\pi^*$ state, the molecule will relax toward $(^1\pi\pi^*)_{min}$. It is also seen that the electronic characters of the two $^1n\pi^*$ states (S_2 and S_3) are interchanged along this LIIC points, indicating that the first $^1n\pi^*$ state is rather characterized by the $^1n_N\pi^*$ state at $(S_0)_{min}$.

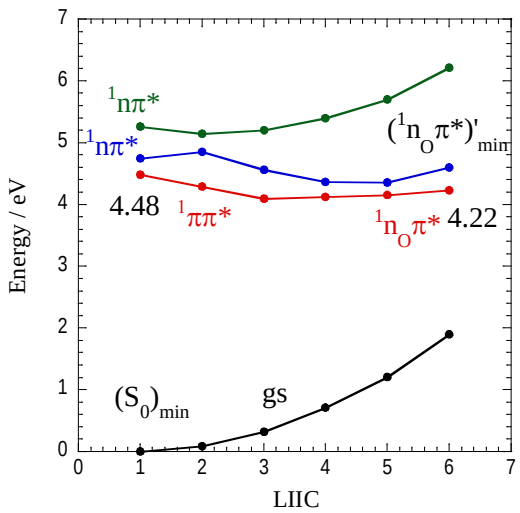
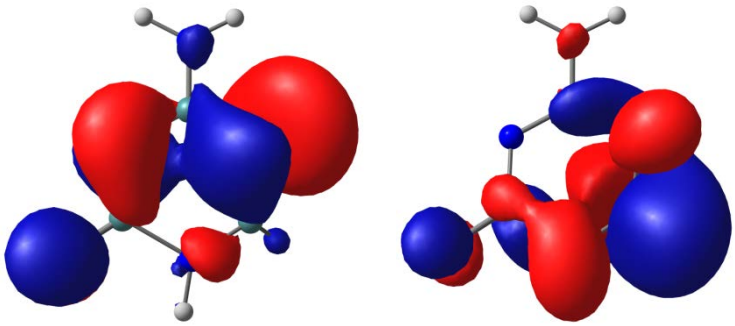
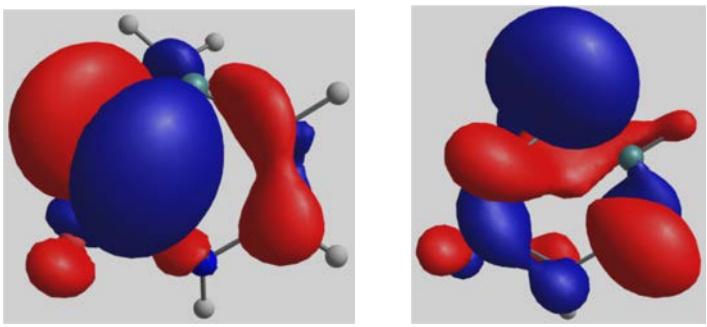


Figure S5. Potential energy profiles for low-lying electronic states of keto cytosine from $(S_0)_{min}$ to $(^1n_O\pi^*)'_{min}$ using LIIC points at the MS(4)-CASPT2(12,9) level.



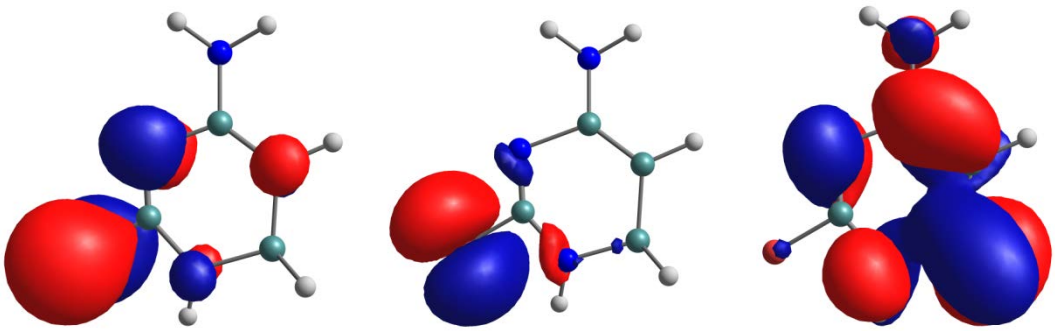
#29(HOMO) #30(LUMO)

Figure S6. SA(2)-CASSCF(8,7) natural orbitals at $(^1\pi\pi^*/gs)_{CI}$ of keto cytosine that are relevant to the excitation.



#29(HOMO) #30(LUMO)

Figure S7. SA(2)-CASSCF(8,7) natural orbitals at $(^1\pi_{N3}\pi^*/gs)_{CI}$ of keto cytosine that are relevant to the excitation.



#28(HOMO−1) #29(HOMO) #30(LUMO)

Figure S8. SA(3)-CASSCF(6,5) natural orbitals at $(^1n_O\pi^*/gs)_{CI}$ of keto cytosine that are relevant to the excitation.

Table SI. Configuration Interaction Coefficients for S_0, S_1, and S_2 States in SA(4)-CASSCF(12,9) Calculation at $(^1n_O\pi^*/gs)_{CI}$. CS Stands for Closed Shell Configuration.

	S_0	S_1	S_2
CS	0.053	−0.069	0.682
$^1\pi\pi^*$	−0.660	0.619	0.043
$^1n_O\pi^*$	0.627	0.669	0.015

Table SII. Eigenvectors of the Effective Hamiltonican Matrix in MS(4)-CASPT2(12,9) Calculation at $(^1n_O\pi^*/gs)_{CI}$.

	S_0 (MS-CASPT2)	S_1 (MS-CASPT2)	S_2 (MS-CASPT2)
S_0 (SS-CASPT2)	0.951	−0.215	0.222
S_1 (SS-CASPT2)	−0.308	−0.613	0.727
S_2 (SS-CASPT2)	0.020	0.760	0.650
S_3 (SS-CASPT2)	−0.000	−0.013	−0.013

Potential energy profiles from $(^1\pi\pi^*)_{min}$ to $(^1\pi\pi^*/gs)'_{CI}$ at the different levels of theory

Potential energy profiles from $(^1\pi\pi^*)_{min}$ to $(^1\pi\pi^*/gs)'_{CI}$ are calculated by the SA(4)-CASSCF(12,9), SS-CASPT2(12,9), and MS(4)-CASPT2(12,9) methods, where the MECI point $(^1\pi\pi^*/gs)'_{CI}$ is determined at the SA(2)-CASSCF(8,7) level (see Figures S9 and S10). The potential energies are evaluated at geometries determined by fixing the dihedral angle *d*(N1-C6-C5-H5) at selected values and optimizing the other degrees of freedom in the $^1\pi\pi^*$ state at the SA(2)-CASSCF(8,7) level. Although the SA(4)-CASSCF(12,9) energies of the $^1\pi\pi^*$ and ground states differ by only ~0.2 eV at $d < 90°$, the MS-CASPT2 profile predicts that these two states are separated by more than 1 eV. Also, the MS(4)-CASPT2(12,9) energy of the $^1\pi\pi^*$ state at $(^1\pi\pi^*/gs)'_{CI}$ is calculated to be 4.24 eV, which is higher than that at $(^1\pi\pi^*)'_{min}$ by 0.14 eV. This result is in contrast to the energy profiles given in the main document where $(^1\pi\pi^*/gs)_{CI}$ lies lower than $(^1\pi\pi^*)_{min}$. Another point worth noting is that a crossing of the SS-CASPT2 energies at $d < 90°$ is observed, but it is an artifact since it disappears when the mixing between electronic states are taken into account by MS-CASPT2. These results indicate that the MS-CASPT2 method is more appropriate for both geometry optimization and energetics.

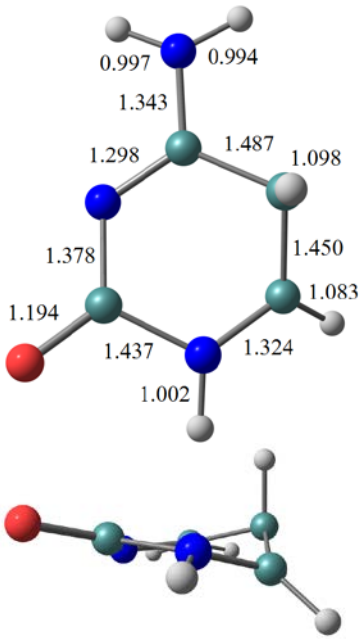


Figure S9. MECI structure $(^1\pi\pi^*/gs)'_{CI}$ of keto cytosine determined by SA(2)-CASSCF(8,7) method. The bond lengths are given in units of Å.

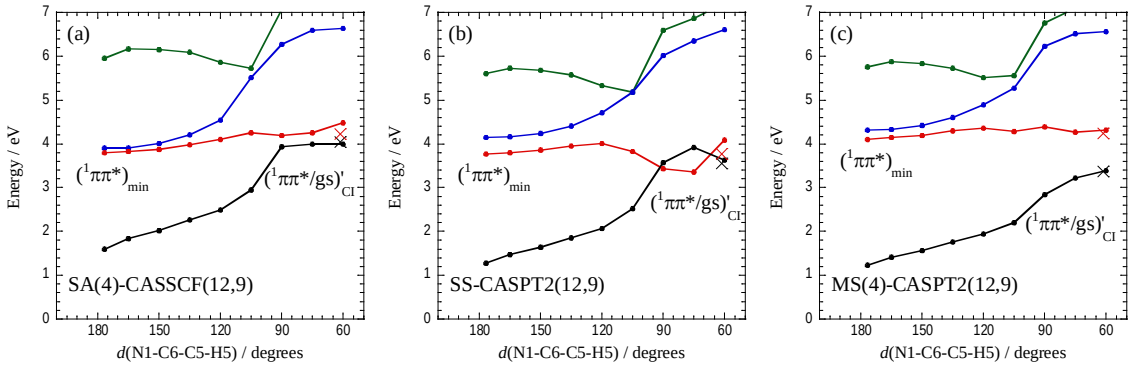


Figure S10. Potential energy profiles for low-lying electronic states of keto cytosine from $(S_0)_{min}$ to $(^1\pi\pi^*/gs)'_{CI}$ at the (a) SA(4)-CASSCF(12,9), (b) SS-CASPT2(12,9), and (c) MS(4)-CASPT2(12,9) levels, as a function of dihedral angle *d*(N1-C6-C5-H5). The potential energies at $(^1\pi\pi^*/gs)'_{CI}$ are shown as cross marks at the respective level.

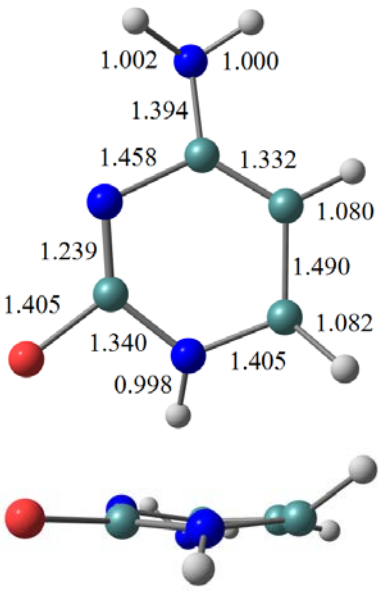


Figure S11. MECI structure $(^1n_O\pi^*/gs)'_{CI}$ of keto cytosine determined by SA(3)-CASSCF(10,8) method. The bond lengths are given in units of Å.

Table SIII Configuration Interaction Coefficients for S_0, S_1, and S_2 States at $(^1n_O\pi^*/gs)'_{CI}$ in SA(4)-CASSCF(12,9) Calculation. CS Stands for Closed Shell Configuration.

	S_0	S_1	S_2
CS	0.120	−0.397	0.671
$^1\pi\pi^*$	0.471	0.718	0.297
$^1n_O\pi^*$	0.777	−0.376	−0.319

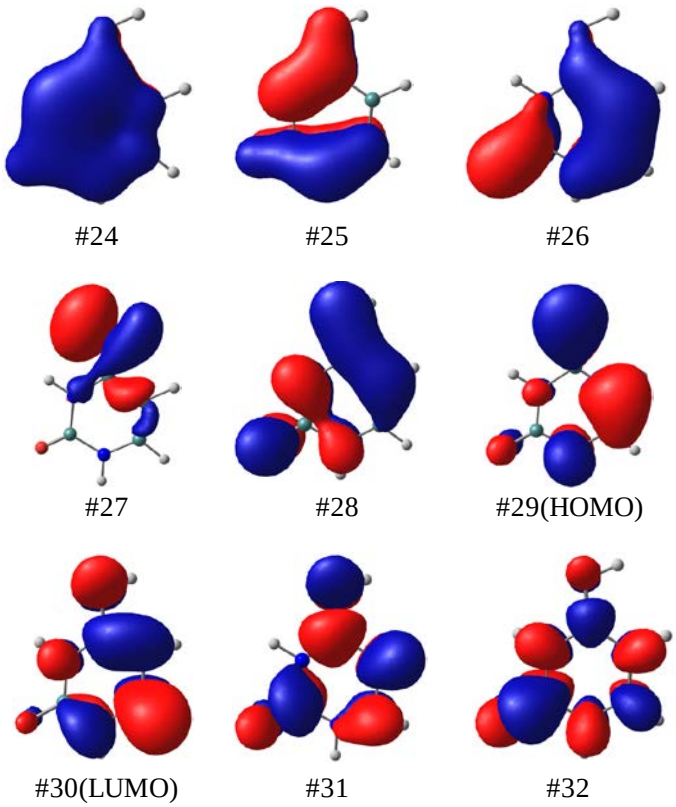


Figure S12. Active orbitals of imino cytosine at $(S_0)_{min}$ in the SA(4)-CASSCF(12,9) calculation.

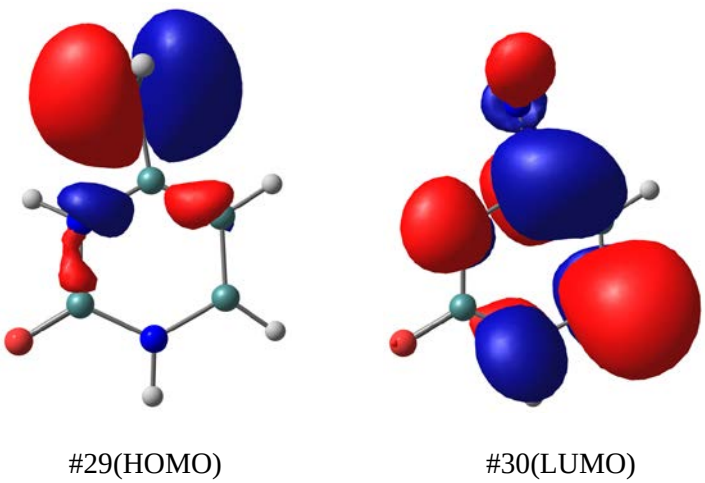


Figure S13. SA(2)-CASSCF(6,5) natural orbitals at $(^1\pi_{N8}\pi^*/gs)_{CI}$ of imino cytosine that are relevant to the excitation.

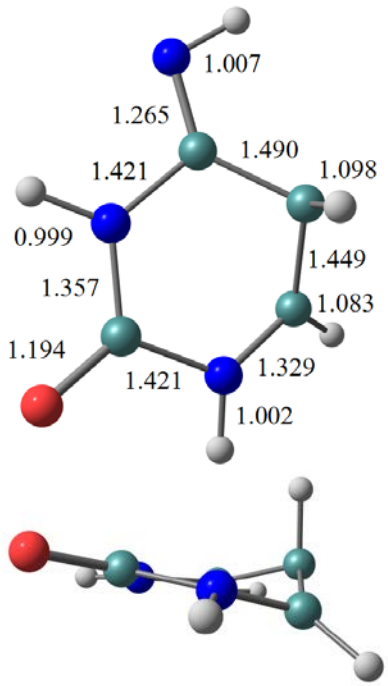


Figure S14. MECI structure $(^1\pi\pi^*/gs)'_{CI}$ of imino cytosine determined by SA(2)-CASSCF(10,8) method. The bond lengths are given in units of Å.

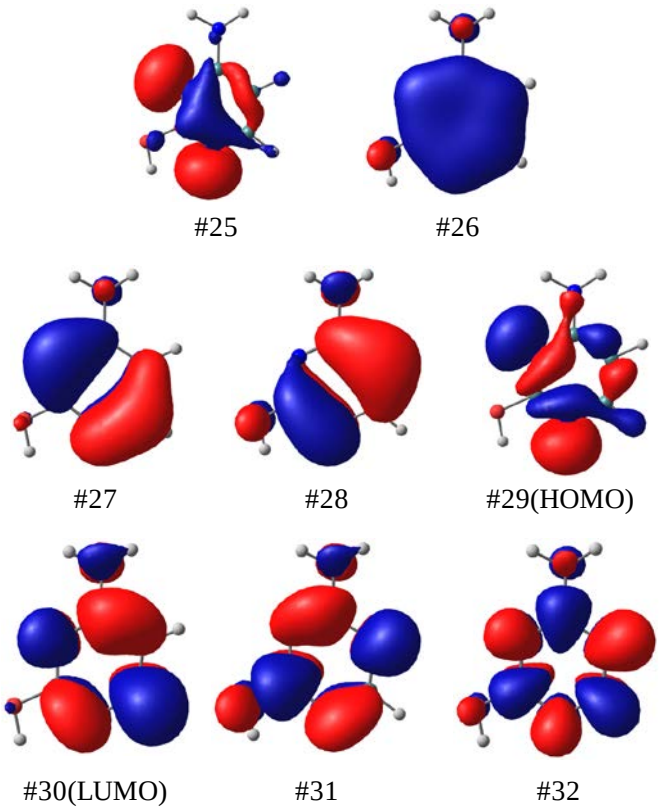


Figure S15. Active orbitals of enol cytosine at $(S_0)_{min}$ in the SA(4)-CASSCF(10,9) calculation.

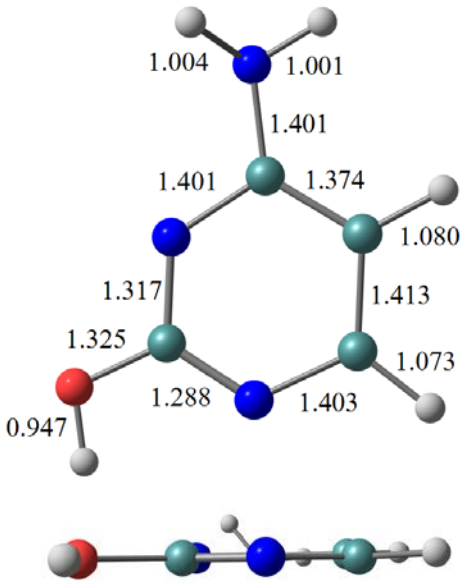


Figure S16. Minimum energy structure of enol cytosine in the $^1n\pi^*$ state determined by SA(4)-CASSCF(10,8) method, $(^1n\pi^*)'_{min}$. The bond lengths are given in units of Å.

MEP from $(S_0)_{min}$ for enol cytosine

The MEP computation of enol cytosine in the $^1\pi\pi^*$ state from $(S_0)_{min}$ was performed at the MS(2)-CASPT2(8,7) level of theory with a stepsize of 0.05 bohr·amu$^{1/2}$ and the energietics are reproduced at the MS(4)-CASPT2(10,8) level at selected points along MEP. The MEP reached a plateau region in the planar structure and thus the calculation was terminated at 1.0 bohr·amu$^{1/2}$. From this structure, the geometry was shifted very slightly toward the direction of the $(^1\pi\pi^*)_{min}$ and the MEP computation was reinitiated. After that, the MEP led to the plateau region again and was terminated at 1.8 bohr·amu$^{1/2}$. It was confirmed that the structure at this point was very close to that at $(^1\pi\pi^*)_{min}$.

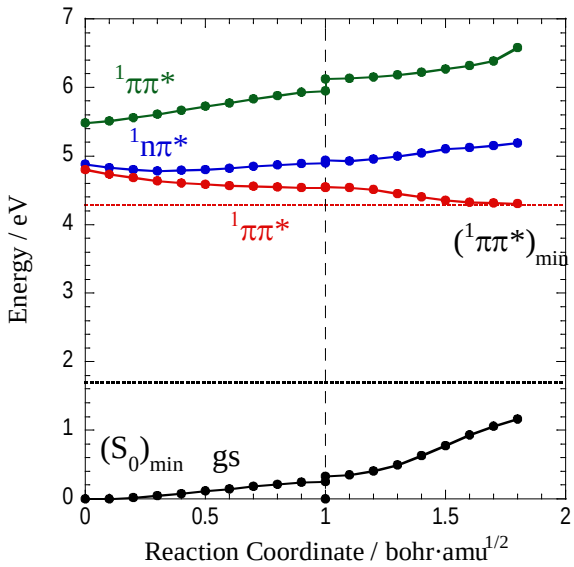
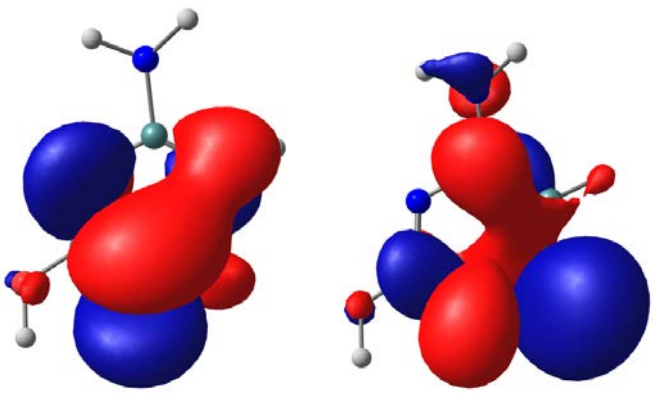


Figure S17. Potential energy profiles of enol cytosine from $(S_0)_{min}$ along the MEP coordinates at the MS(4)-CASPT2(10,8) level. The horizontal dotted lines represent the potential energies of the ground and $^1\pi\pi^*$ states at $(^1\pi\pi^*)_{min}$. The vertical dashed line represents the reaction coordinate at which the MEP was terminated once since it reached a plateau region in the planar structure.



#29(HOMO) #30(LUMO)

Figure S18. SA(2)-CASSCF(8,7) natural orbitals at $(^1\pi_{N1}\pi^*/gs)_{CI}$ of enol cytosine that are relevant to the excitation.

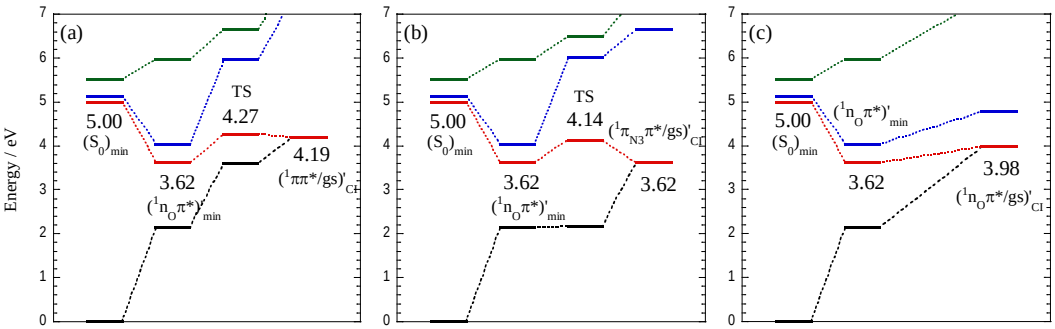


Figure S19. SA(4)-CASSCF(12,9) electronic energies of the $(S_0)_{min}$, $(S_1)_{min}$ = $(^1n_O\pi^*)'_{min}$, TS, and MECI structures for keto cytosine, where the latter three structures are optimized at the SA(4)-CASSCF(12,9) level (indicated by prime). The TS structure is determined by the highest point along the reaction path, which is obtained by excited-state geometry optimizations at fixed values of the same driving coordinate as used in the MS-CASPT2 optimization. For the MECI optimization, the projected gradient method by Bearpark *et al*. (Ref. 69 of the main document) is applied.

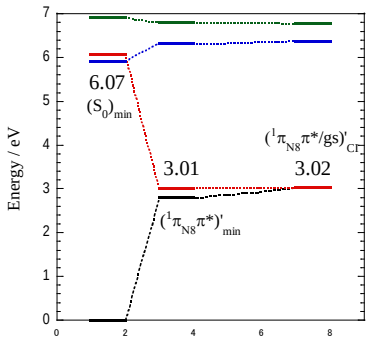


Figure S20. SA(4)-CASSCF(12,9) electronic energies of the $(S_0)_{min}$, $(S_1)_{min}$ = $(^1\pi_{N8}\pi^*)'_{min}$, and $(^1\pi_{N8}\pi^*/gs)_{CI}$ structures for imino cytosine, where the latter two structures are optimized at the SA(4)-CASSCF(12,9) level (indicated by prime).

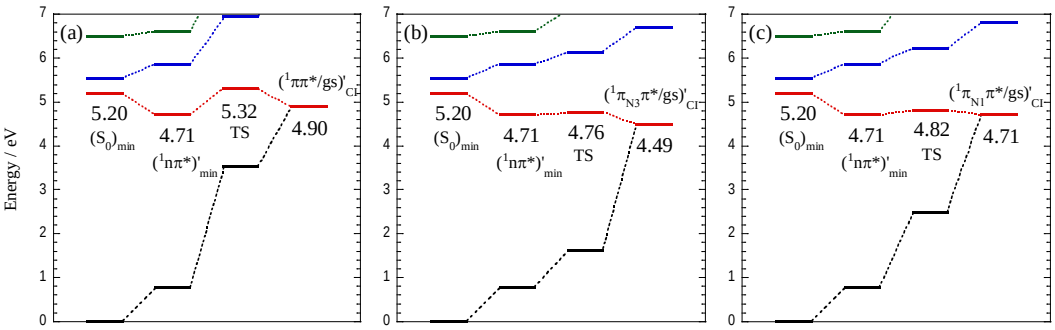


Figure S21. SA(4)-CASSCF(10,8) electronic energies of the $(S_0)_{min}$, $(S_1)_{min}$ = $(^1n\pi^*)'_{min}$, TS, and MECI structures for enol cytosine, where the latter three structures are optimized at the SA(4)-CASSCF(10,9) level (indicated by prime).