Intersystem Crossing in 5-Azacytosine: Time-Resolved Photoelectron and Quantum-Chemical Insights into the Effect of Aza Substitution on Cytosine

Bijay Duwal,[1] Antonio Carlos Borin,[2*] and Susanne Ullrich[1*]

[1] Department of Physics and Astronomy, University of Georgia, Athens, Georgia 30602, USA.
[2] Department of Fundamental Chemistry, Institute of Chemistry, University of São Paulo, Av. Prof. Lineu Prestes, 748, São Paulo, 05508-000, São Paulo, Brazil.

ORCIDs: BD (0009-0003-6475-9948), ACB (0000-0003-3047-2044), SU (0000-0002-1828-2777)

Email: ancborin@iq.usp.br, ullrich@uga.edu

Supplementary Information

Contents

S1. Active space and orbital transitions

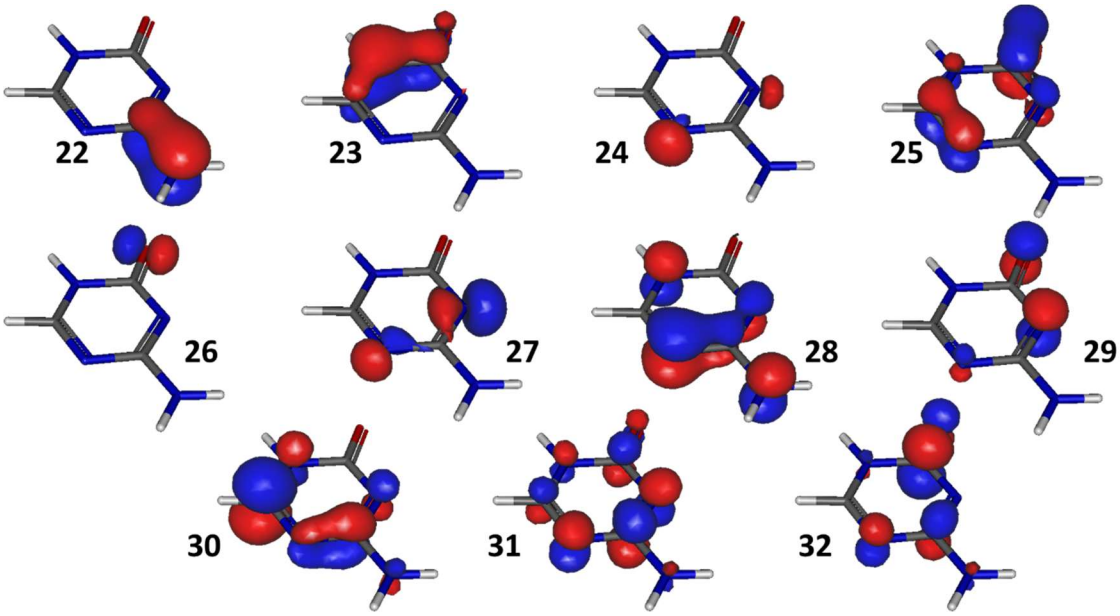

Figure S1: Active space used for the keto tautomer of 5AC in terms of state-averaged natural orbitals, computed at the SA(6S;6T)-CASSCF(16,11) level, in the ground-state equilibrium geometry and with the cc-pVDZ basis sets.

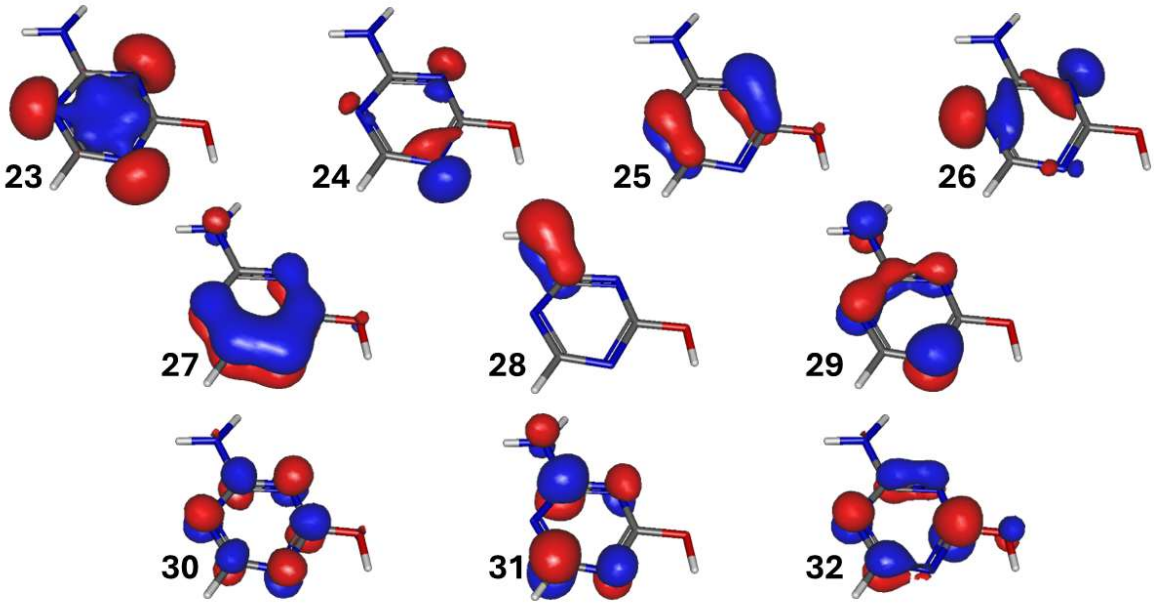

Figure S2: Active space used for the enol tautomer of 5AC in terms of state-averaged natural orbitals, computed at the SA(7S;7T)-CASSCF(14,10) level, in the ground-state equilibrium geometry and with the cc-pVDZ basis sets.

Table S1: Orbital transitions of enol 5AC

States	Character	Transitions	Weights
S_1	$^1n_N\pi^*$	26 → 31	37%
S_2	$^1\pi\pi^*$	29 → 31, 29 → 32, 26 → 31	20%, 15%, 23%
S_3	$^1n_N\pi^*$	24 → 31, 24 → 32	39%, 37%
S_4	$^1n_N\pi^*$	26 → 32, 24 → 31	38%, 21%
S_5	$^1\pi\pi^*$	31 → 29, 31 → 32	35%, 26%
S_6	$^1n_N\pi^*$	26 → 32, 24 → 32	39%, 34%
T_1	$^3\pi\pi^*$	29 → 31	51%
T_2	$^3n_N\pi^*$	26 → 31	70%
T_3	$^3\pi\pi^*$	29 → 32	36%
T_4	$^3n_N\pi^*$	24 → 31	36%
T_5	$^3n_N\pi^*$	25 → 31	55%
T_6	$^3n_N\pi^*$	26 → 32, 25 → 32	40%, 36%

S2. UV-vis absorption spectra and FTIR spectra

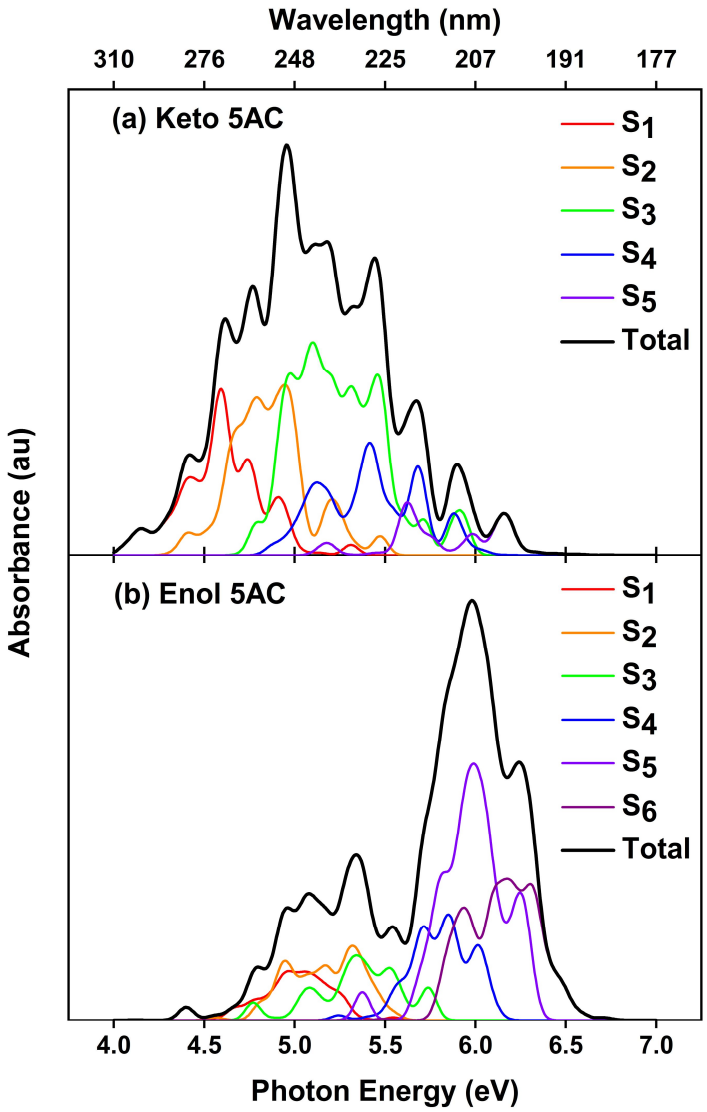

Figure S3: Simulated UV-vis absorption spectra of keto and enol 5AC with contributions from different states. The labels S_1-S_6 refer to the ordering of the singlet excited states according to vertical excitation energies for a distribution of 100 geometries around the ground state minimum. The two lowest singlet excited states in 5AC are close in energy and expected to swap orbital character depending on the starting geometry. The orbital characters provided in Table 1 in the main manuscript are for vertical excitation from the ground state minimum geometry only and don't necessarily apply here.

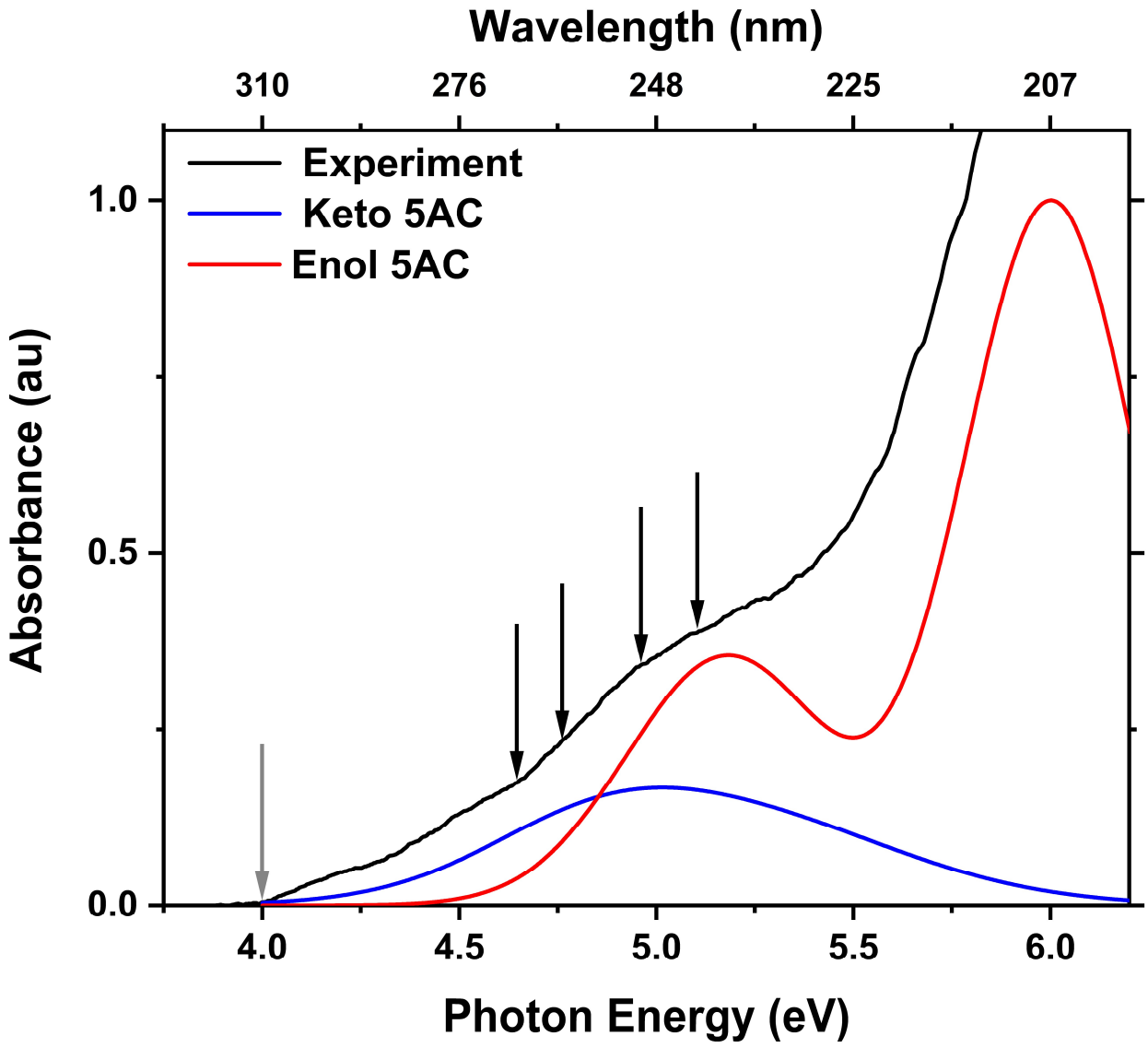

Figure S4: UV-vis absorption spectra of 5-azacytosine. The black trace represents the experimental spectrum. Gaussian fits are performed on the total simulated tautomer spectra in Figure S3. The resulting keto and enol Gaussian curves are then subjected to a composite linear regression fit to the experimental spectrum. The red and blue curves are weighted by the regression coefficients and approximate the contribution of each tautomer to the overall absorption spectrum of 5AC. Black and grey arrows indicate the pump and probe photon energies, respectively.

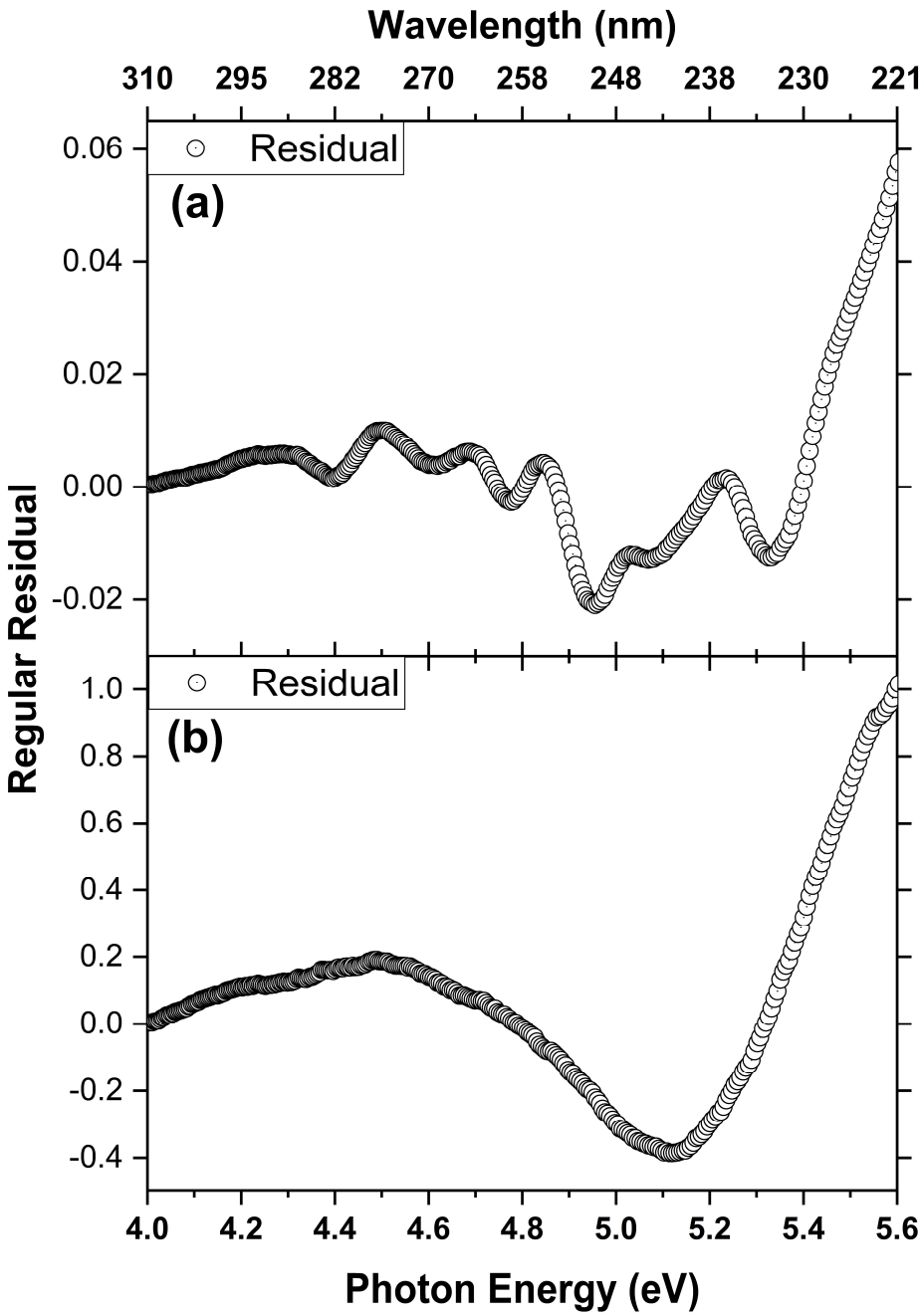

Figure S5: Residuals of the composite linear regression fits. (a) Residuals from the regression fit of the experimental and simulated UV-vis spectra. (b) Residuals from the regression fit of the experimental spectrum and the Gaussian fits to the simulated spectra. Both regression fits are performed in the range 4 – 5.6 eV.

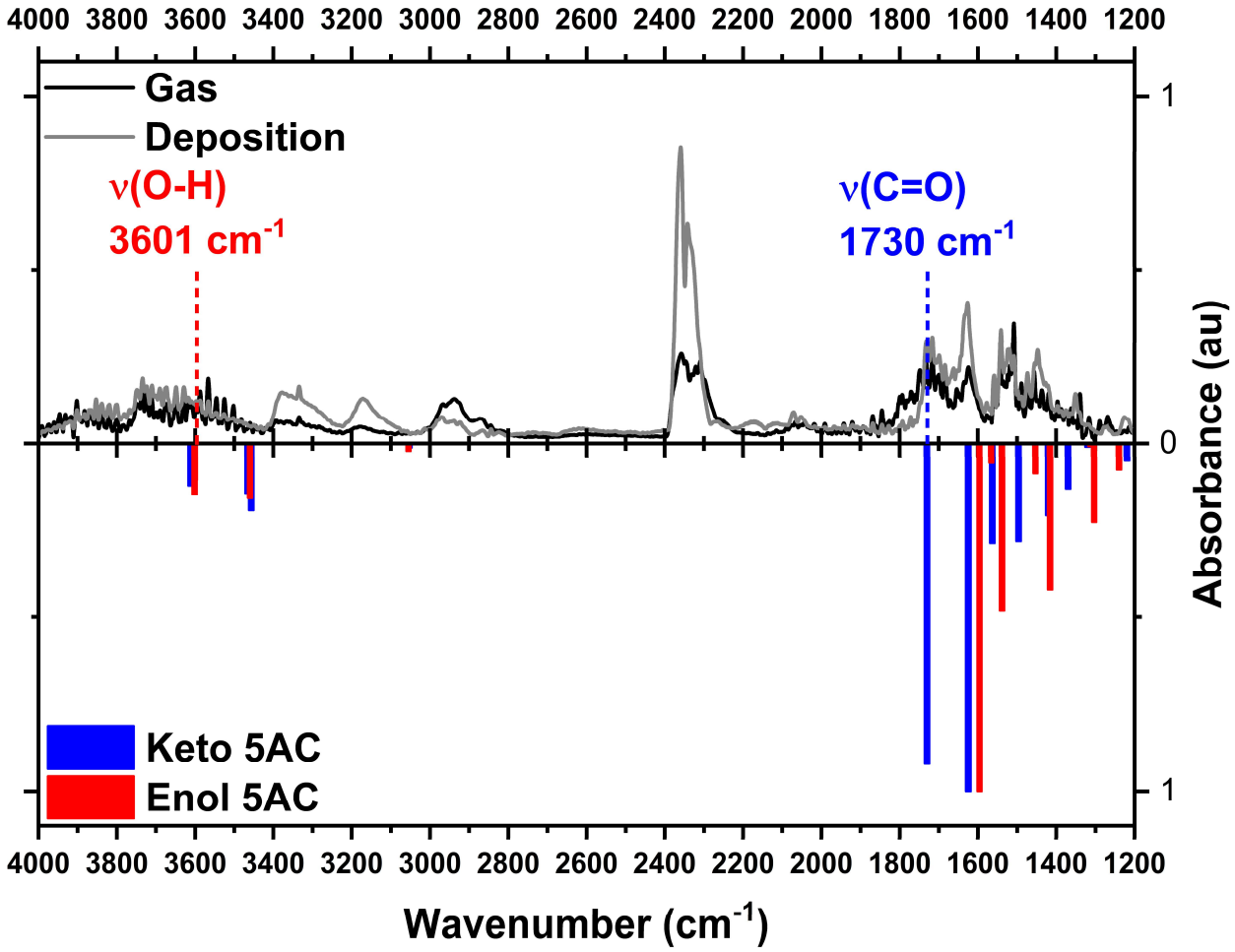

Figure S6: Fourier-transform infrared (FTIR) spectra of 5AC obtained in the gas phase (black) and from deposition on cell windows (grey). The assignments of vibrational peaks are made based on the IR spectra calculated at the MP2/cc-pVTZ level of theory (Tables S4 and S5).

S3. Excited state energy, ionization energy, and electron binding energy

Table S2: Excited state energy, vibrational energy, ionization energy, and electron binding energy at different geometries of the keto and enol tautomers are calculated based on XMS(6S;6T)-CASPT2(16,11)/cc-pVDZ and XMS(7S;7T)-CASPT2(14,10)/cc-pVDZ levels of theory, respectively.

Tautomer	Geometry	Excited State Energy (eV)	Vibrational Energy (eV)	Ionization Energy (eV)	Binding Energy (eV) = Ionization Energy + Vibrational Energy
Keto	**FC ($^1\pi\pi^*$)**	4.92	-	9.01	9.01
	S_2 ($^1\pi\pi^*$) min	4.47	0.45	10.04	10.49
	S_1 ($^1n_N\pi^*$) min	4.00	0.92	10.18	11.10
	T_1 ($^3\pi\pi^*$) min	3.51	1.41	9.79	11.20
Enol	**FC ($^1\pi\pi^*$)**	5.28	-	9.21	9.21
	S_1 ($^1n_N\pi^*$) min	4.21	1.07	10.16	11.23
	T_1 ($^3\pi\pi^*$) min	4.03	1.25	10.24	11.49

The vertical excitation energies of the bright S_2 ($^1\pi\pi^*$) states of the keto (4.92 eV) and enol (5.28 eV) forms are closest to the pump photon energies of 250 nm (4.96 eV) and 243 nm (5.10 eV), respectively. As the excitation wavelength is changed, the TRPES signals are expected to shift along the electron binding energy axis due to varying amounts of excess vibrational energy. These shifts are estimated as the difference between the calculated vertical excitation energy and the experimentally applied pump photon energy.

Table S3: Expected shift in TRPES 2D plots along electron binding energy axis at different pump wavelengths.

Pump wavelength (nm)	Photon energy (eV)	Expected shift in binding energy, keto/enol (eV)
267 nm	4.64	-0.28/-0.64
260 nm	4.76	-0.16/-0.52
250 nm	4.96	+0.04/-0.32
243 nm	5.10	+0.18/-0.18

S4. IR spectra calculations

Table S4: IR spectra of keto 5AC computed at MP2/cc-pVTZ and B3LYP/aug-cc-pVDZ.

MP2/cc-pVTZ			B3LYP/aug-cc-pVDZ		
Frequency (cm^{-1})	**Intensity (km/mol)**	**Character**	**Frequency (cm^{-1})**	**Intensity (km/mol)**	**Character**
133.747	4.00779	-	137.151	0.72856	-
178.022	3.82181	-	174.791	0.00998	-
350.464	0.556872	-	244.473	27.0311	-
401.674	1.75643	-	347.149	0.57819	-
505.596	0.834271	-	404.387	1.08452	-
535.017	0.132326	-	503.367	0.98405	-
557.605	0.370735	-	534.641	0.24618	-
568.943	0.09775	-	559.084	0.19029	-
640.406	9.42002	-	561.497	0.08583	-
738.1	1.57973	-	635.806	8.416	-
757.724	0.758378	-	731.737	1.80576	-
766.275	1.81003	-	748.22	0.62476	-
894.977	0.074694	-	765.728	3.13246	-
932.573	1.61648	-	885.144	0.00732	-
952.284	0.411714	-	931.703	1.63809	-
1034.37	4.09242	-	942.4	0.25616	-
1120.45	3.01826	-	1025.01	4.3953	-
1219.44	4.94144	C_2-N_3 stretch	1110.54	3.09121	-
1319.95	0.958855	C_6-H bend	1195.29	3.87899	-
1369.73	12.9661	N_1-H & C_6-H bend	1302.85	0.69263	N_1-H bend
1420.44	20.5507	Amino group N-H bending, N_1-H & C_6-H bend	1354.14	14.7881	Amino group N-H bending, N_1-H & C_6-H bend
1496.1	28.1327	Amino group N-H bending, N_3=C_4 stretch, N_1-H bend	1398.15	20.0583	Amino group N-H bending, N_1-H & C_6-H bend
1563.47	28.6888	Amino group N-H bending, N_5=C_6 & N_3=C_4 stretch	1469.43	29.3525	Amino group N-H bending, N_3=C_4 stretch, N_1-H bend
1624.77	100	N_5=C_6 & N_3=C_4 stretch, Amino group N-H bending	1545.93	24.3405	Amino group N-H bending, ring
1730.19	92.2365	C_2=O stretch	1593.21	100	N_5=C_6 & N_3=C_4 stretch, Amino group N-H bending
3052.31	1.19217	C_6-H stretch	1691.43	91.7483	C_2=O stretch
3456.89	19.142	N_1-H stretch, Symmetric stretching of amino group N-H	3008.11	1.31074	C_6-H stretch
3465.82	14.2542	Symmetric stretching of amino group N-H, N_1-H stretch	3418.35	20.4708	Symmetric stretching of amino group N-H, N_1-H stretch
3611.66	12.0936	Asymmetric stretching of amino group N-H	3421.64	3.27019	Symmetric stretching of amino group N-H, N_1-H stretch
			3556.91	9.17051	Asymmetric stretching of amino group N-H

Table S5: IR spectra of enol 5AC computed at MP2/cc-pVTZ and B3LYP/aug-cc-pVDZ.

MP2/cc-pVTZ			B3LYP/aug-cc-pVDZ		
Frequency (cm^{-1})	**Intensity (km/mol)**	**Character**	**Frequency (cm^{-1})**	**Intensity (km/mol)**	**Character**
174.626	0.24044	-	172.159	0.80622	-
193.971	0.38319	-	193.22	0.01994	-
218.948	32.4253	-	205.922	22.6628	-
335.79	1.90526	-	332.652	1.79187	-
450.009	0.4125	-	445.977	0.32718	-
479.48	0.23113	-	478.401	0.27031	-
535.79	0.65618	-	547.19	0.73879	-
548.609	11.4728	-	552.615	9.60375	-
558.877	0.44746	-	556.709	0.11492	-
576.858	0.48413	-	577.419	0.33129	-
729.599	0.09649	-	723.539	0.11727	-
796.09	2.22771	-	795.758	1.15393	-
799.078	3.14011	-	796.508	4.41811	-
944.846	1.65241	-	938.837	1.79832	-
963.651	2.91335	-	953.914	1.8423	-
973.915	0.51117	-	966.14	0.29786	-
1049.81	7.8508	-	1037.8	7.03966	-
1183.33	4.1824	-	1168.17	2.48317	-
1239.65	7.55889	Ring	1223.48	6.88252	O-H bend, ring
1303.35	22.4609	O-H bend, C_1-H bend	1291.42	6.90656	C_1-H bend
1312.24	0.8071	C_1-H bend	1298.69	16.2529	O-H bend
1416.04	42.1365	C_1-H bend, Amino group N-H bending	1389.8	39.6006	C_1-H bend, Amino group N-H bending
1453.34	8.57267	Amino group N-H bending, Ring	1429.97	8.19359	Amino group N-H bending, Ring
1538.77	48.2491	Ring	1505.44	46.8572	Ring
1566.13	5.58433	Amino group N-H bending, ring	1552.05	1.27999	Amino group N-H bending, ring
1596.19	100	Ring	1570.87	100	Ring
3054.69	2.13124	C_6-H stretch	3008.94	2.07918	C_6-H stretch
3459.85	15.7031	Symmetric stretching of amino group N-H	3423.75	11.7316	Symmetric stretching of amino group N-H
3600.56	10.4291	Asymmetric stretching of amino group N-H, O-H stretch	3558.9	7.83826	Asymmetric stretching of amino group N-H
3601.39	14.4502	O-H stretch, Asymmetric stretching of amino group N-H	3568.53	11.2238	O-H stretch

S5. TRPES analysis

The TRPES data are analyzed using a three-exponential sequential decay model at pump wavelenghts: 267 nm, 260 nm, 250 nm, and 243 nm, and probe wavelength: 310 nm. The analysis is illustrated in Figure S7.

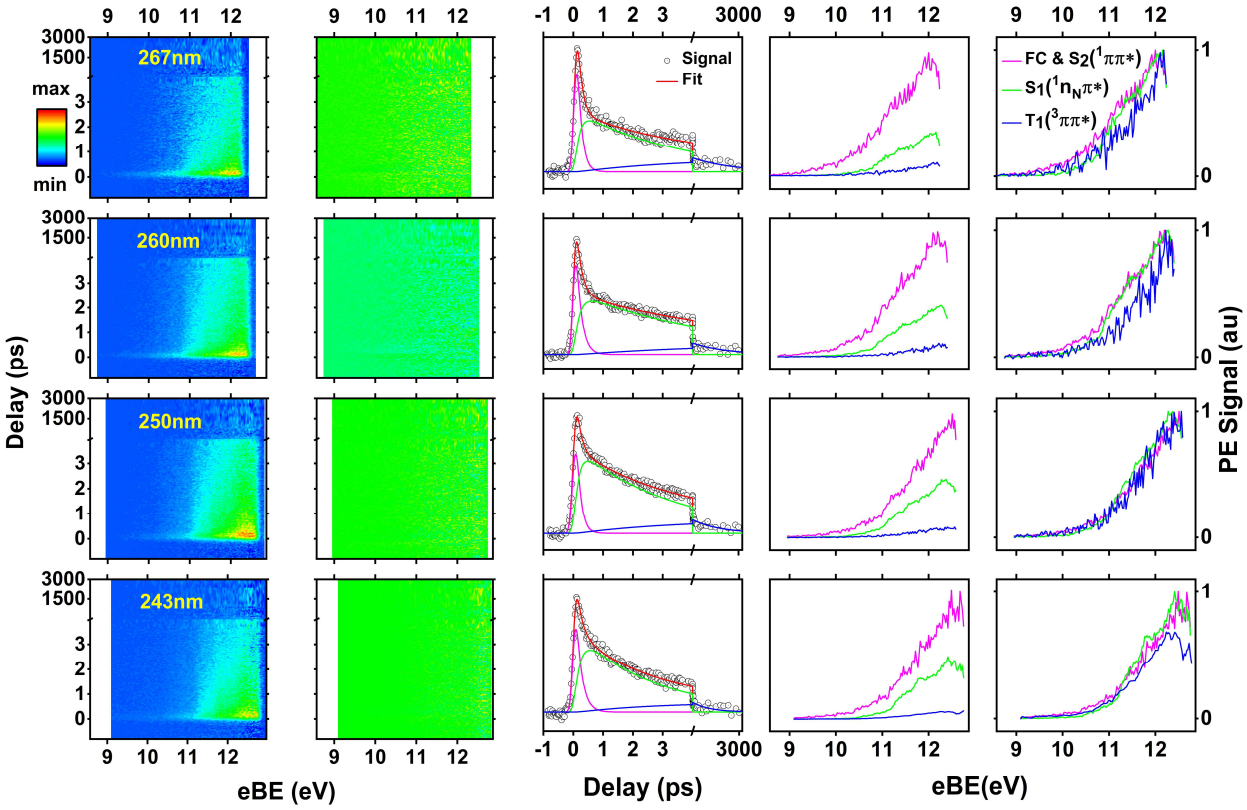

Figure S7: TRPES analysis of 5AC in a three-exponential sequential decay model. Columns 1 through 5 show the TRPES signal, the residuals of the global fit, the energy-integrated time traces, the evolution-associated spectra, and the normalized evolution-associated spectra, respectively. The rows correspond to different pump wavelengths as labeled in the TRPES signal plots.

S6. Fit equations

The TRPES data is analyzed with a three-step sequential exponential decay function, which corresponds to the decay processes: A → B → C → D.

$$
\begin{aligned}
x(t,E) = x(0,E) + \sigma_A(E) e^{-k_1 t} e^{\left(\frac{IRF*k_1}{4\sqrt{ln2}}\right)^2} * \frac{1}{2}\left[1 + \mathrm{erf}\left\{\left(\frac{2\sqrt{ln2}*t}{IRF}\right) - \left(\frac{IRF*k_1}{4\sqrt{ln2}}\right)\right\}\right] \\
+ \sigma_B(E) * \frac{k_1}{k_2 - k_1} * e^{-k_1 t} e^{\left(\frac{IRF*k_1}{4\sqrt{ln}}\right)^2} * \frac{1}{2}\left[1 + \mathrm{erf}\left\{\left(\frac{2\sqrt{ln2}*t}{IRF}\right) - \left(\frac{IRF*k_1}{4\sqrt{ln2}}\right)\right\}\right] \\
+ \sigma_B(E) * \frac{k_1}{k_1 - k_2} * e^{-k_2 t} e^{\left(\frac{IRF*k_2}{4\sqrt{ln2}}\right)^2} * \frac{1}{2}\left[1 + \mathrm{erf}\left\{\left(\frac{2\sqrt{ln2}*t}{IRF}\right) - \left(\frac{IRF*k_2}{4\sqrt{ln2}}\right)\right\}\right] \\
+ \sigma_C(E) * \frac{k_1 k_2}{(k_2 - k_1)(k_3 - k_1)} * e^{-k_1 t} e^{\left(\frac{IRF*k_1}{4\sqrt{ln}}\right)^2} \\
* \frac{1}{2}\left[1 + \mathrm{erf}\left\{\left(\frac{2\sqrt{ln2}*t}{IRF}\right) - \left(\frac{IRF*k_1}{4\sqrt{ln2}}\right)\right\}\right] + \sigma_C(E) * \frac{k_1 k_2}{(k_1 - k_2)(k_3 - k_2)} \\
* e^{-k_2 t} e^{\left(\frac{IRF*k_2}{4\sqrt{ln}}\right)^2} * \frac{1}{2}\left[1 + \mathrm{erf}\left\{\left(\frac{2\sqrt{ln2}*t}{IRF}\right) - \left(\frac{IRF*k_2}{4\sqrt{ln2}}\right)\right\}\right] + \sigma_C(E) \\
* \frac{k_1 k_2}{(k_1 - k_3)(k_2 - k_3)} * e^{-k_3 t} e^{\left(\frac{IRF*k_3}{4\sqrt{ln}}\right)^2} \\
* \frac{1}{2}\left[1 + \mathrm{erf}\left\{\left(\frac{2\sqrt{ln2}*t}{IRF}\right) - \left(\frac{IRF*k_1}{4\sqrt{ln2}}\right)\right\}\right]
\end{aligned}
$$

In the fit equation, $\sigma_A(E)$, $\sigma_B(E)$, and $\sigma_C(E)$ are the evolution-associated spectra, and $k_1 = 1/\tau_1$, $k_2 = 1/\tau_2$, and $k_3 = 1/\tau_3$ are the rate constants of the consecutive decay processes. The instrument response function (IRF) is characterized by the FWHM of the Gaussian cross-correlation.

S7. Justification of fits

To justify the use of a fitting model with 3 exponentials, a fit with only 2 exponentials is provided in Figure S7 for the time trace at 250 nm. The latter fit clearly demonstrates that 2 exponentials are insufficient to describe the decay dynamics over the entire pump-probe delay range. The addition of a third exponential is required to obtain a good fit of the slower dynamics.

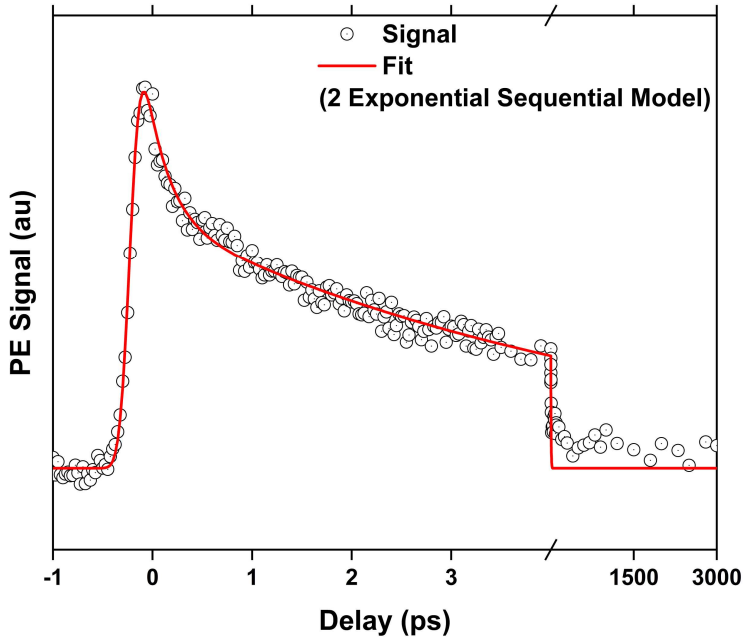

Figure S8: Fit of the energy-integrated time trace at 250 nm using a sequential decay model with two exponentials.

The scree plot and the Singular Value Decomposition (SVD) of the 2D TRPES data at 250 nm also show that 3 exponentials are required for a good fit. Furthermore, adding a fourth exponential does not improve the quality of the fit.

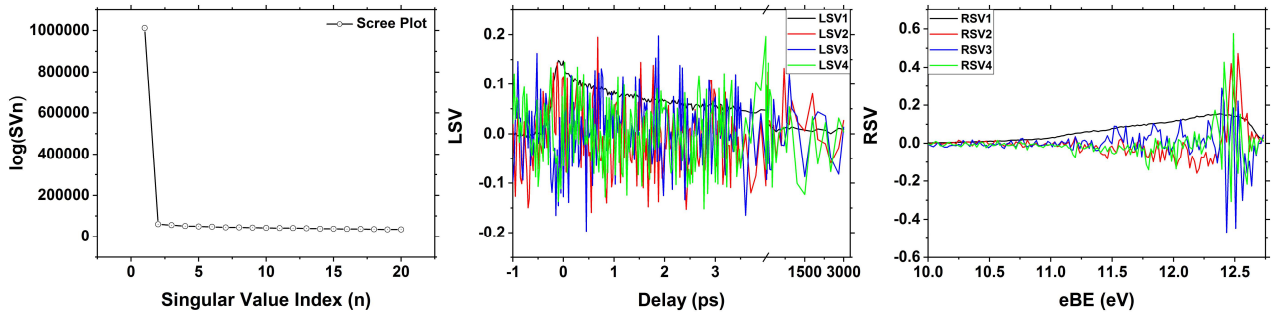

Figure S9: Scree plot, left singular values (LSV) plot, and right singular values (RSV) plot obtained from the global lifetime analysis of the 250 nm TRPES using Glotaran.

S8. Spin-Orbit Coupling

The effective spin-orbit couplings are obtained using Equation 1.

$$\langle \Psi_S | H_{eff}^{SOC} | \Psi_T \rangle = \sqrt{\frac{\langle \Psi_S | H_{SOC} | \Psi_{T_{+1}} \rangle^2 + \langle \Psi_S | H_{SOC} | \Psi_{T_0} \rangle^2 + \langle \Psi_S | H_{SOC} | \Psi_{T_{-1}} \rangle^2}{3}}, \quad (1)$$

where Ψ_S and Ψ_T are the perturbatively modified singlet and triplet state wavefunctions, respectively, and H_{SOC} is the spin-orbit operator.

S9. Cartesian coordinates

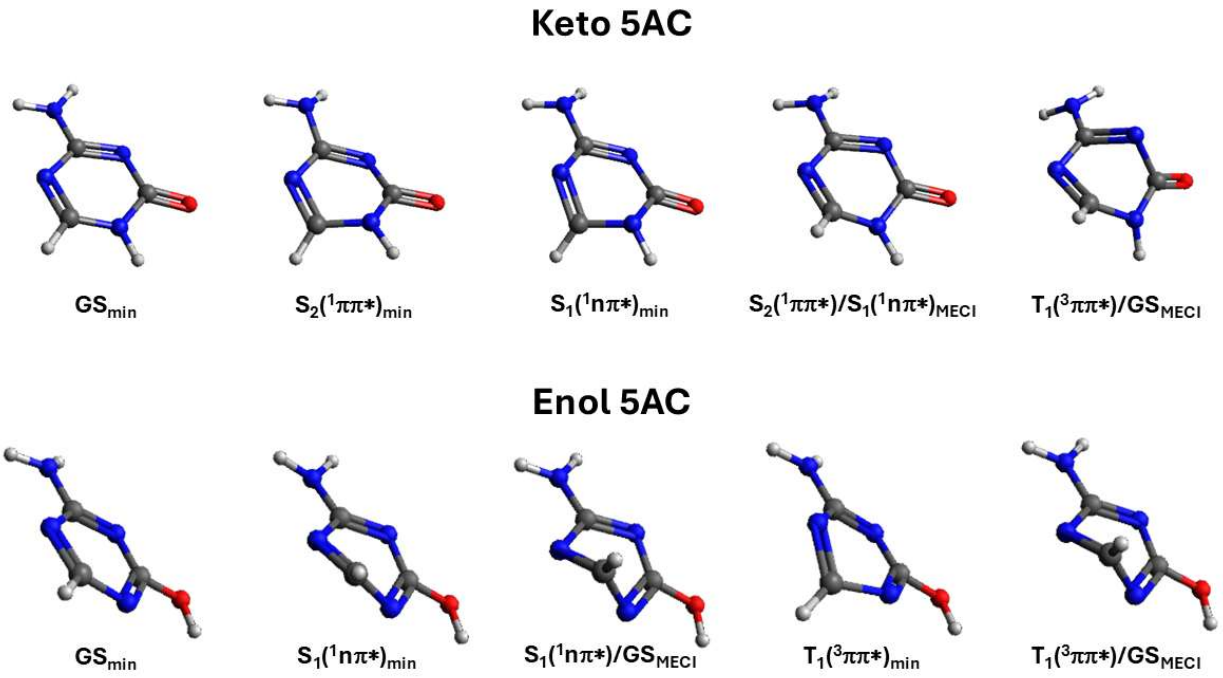

Figure S10: Optimized geometries. See coordinates below.

5AC-Keto Optimized Geometries XMS(6)-CAS(16,11)

GS_{min} (-409.87980896998715)

C	0.97793928	0.84978369	-0.00291899
C	-0.26290745	-1.10858179	0.00509840
C	-1.46358205	0.78760827	0.00422975
N	-0.31052514	1.48871203	0.00405787
N	0.93650516	-0.52419613	0.00005495
N	-0.29777056	-2.45562555	0.00595916
N	-1.52865160	-0.52375813	0.00727538
O	1.97141910	1.55576537	-0.00917280
H	-2.39329948	1.36839036	0.00076109
H	-0.30610354	2.50659569	0.00140135
H	-1.19154491	-2.92747423	0.01190843
H	0.57491419	-2.96785459	0.00778242

$S_2(^1\pi\pi^*)/S_1(^1n\pi^*)_{MECI}$ (-409.72326548512780)

C	0.93533391	0.85100004	-0.01923254
C	-0.34852398	-1.14982571	-0.07404406
C	-1.47485895	0.84884964	-0.19872821
N	-0.28913085	1.49572267	0.22293045
N	0.86521181	-0.48830181	-0.26510521
N	-0.26520164	-2.51345977	-0.02639719
N	-1.50793552	-0.52993047	-0.05522570
O	1.98716800	1.50437721	0.01266783
H	-2.40274273	1.40321536	-0.03673297
H	-0.23694566	2.51277286	0.25464038
H	-1.08823744	-2.98256148	0.33986752
H	0.63302104	-2.87857955	0.27640469

$S_1(^1n\pi^*)_{min}$ (-409.73275243167063)

C	0.93478293	0.86885357	0.00783224
C	-0.31071326	-1.16757563	-0.05210962
C	-1.44396029	0.84998444	0.20983049
N	-0.31545014	1.53399829	-0.27311042
N	0.86576161	-0.51394300	0.05686890
N	-0.29329675	-2.52881622	-0.16988285
N	-1.43930019	-0.48054755	-0.26139574
O	1.94545927	1.53501731	0.13057173
H	-2.40065924	1.37800068	0.24109570
H	-0.26419801	2.53912504	-0.09338521
H	-1.14645386	-3.00661424	0.10266424
H	0.57442093	-2.95811771	0.13745757

$T_1(^3\pi\pi^*)/GS_{MECI}$ (-409.73473980688209)

C	0.84290559	0.75706676	-0.07754029
C	-0.50978321	-1.10198539	-0.20032945
C	-1.46373340	0.87991358	-0.38575032
N	-0.26605080	1.52801834	0.07227873
N	0.43125455	-0.33341644	-0.95144584
N	-0.16996371	-2.36870587	0.15194680
N	-1.63289962	-0.49778446	0.00409982
O	1.96560185	0.83849177	0.39130886
H	-2.35716573	1.50381351	-0.40549285
H	-0.26126495	2.25330935	0.78880826
H	-0.70720906	-2.77441270	0.91191585
H	0.81656850	-2.60185844	0.12348042

$S_2(^1\pi\pi^*)_{min}$ (-409.73016249557105)

C	0.88233820	0.80473792	0.04341154
C	-0.35840852	-1.11907981	0.06418150
C	-1.47600702	0.84721151	0.24504917
N	-0.29831424	1.47405568	-0.23688946
N	0.90148562	-0.49756835	0.30565971
N	-0.24396961	-2.45726608	-0.19587117
N	-1.53079850	-0.53372554	0.11334596
O	1.96586319	1.48668523	0.01068616
H	-2.39681599	1.42907104	0.18193590
H	-0.23594236	2.48547871	-0.33157442
H	-1.10127117	-2.99134413	-0.08615127
H	0.60310741	-2.88781217	0.16172238

5AC-Enol Optimized Geometries XMS(7)-CAS(14,10)

GS_{min} (-409.88251885011840)

C	0.30521957	1.05786623	-1.04405332
C	0.09936424	0.00712370	0.91448594
C	-0.04370793	-1.15822197	-1.00028160
N	0.13132310	-0.07220209	-1.77565637
N	0.30351157	1.18721473	0.28312259
N	0.11116088	0.03051665	2.26918087
N	-0.08217429	-1.21881946	0.33569397
O	0.49800982	2.18540025	-1.73586060
H	-0.17845587	-2.10828021	-1.53061041
H	-0.22013252	-0.79586941	2.75218639
H	0.06112392	0.93448181	2.72408371
H	0.46742150	1.90331076	-2.66614318

$S_1(^1n\pi^*)/GS_{MECI}$ (-409.70888916751909)

C	0.31208481	1.06980495	-0.99166232
C	0.04403081	0.06995750	0.89533259
C	0.31172985	-1.24162041	-1.00811347
N	-0.19669207	-0.03386574	-1.61645557
N	0.52902562	1.21275819	0.32088870
N	-0.00904785	0.03260220	2.23212640
N	-0.47183811	-0.98559798	0.20060295
O	0.55694537	2.12743481	-1.76352603
H	0.02842525	-2.16370925	-1.53249687
H	-0.34116810	-0.79866503	2.70530288
H	0.33593330	0.83091069	2.75279578
H	0.35323512	1.83251105	-2.66864706

$S_1(^1n\pi^*)_{min}$ (-409.72773202969876)

C	0.30626036	1.06356294	-1.01922030
C	0.02830620	0.03603664	0.91171825
C	0.20252330	-1.27588427	-1.04333452
N	0.00645891	-0.04856373	-1.70537552
N	0.41514807	1.21096277	0.31779479
N	-0.10472786	0.07319107	2.25761227
N	-0.28037887	-1.10274832	0.27898377
O	0.49773611	2.18065304	-1.73466950
H	-0.12279036	-2.18089672	-1.56564109
H	-0.19283855	-0.79856500	2.76765527
H	0.27272567	0.89110598	2.72431181
H	0.42424101	1.90366657	-2.66368723

$T_1(^3\pi\pi^*)/GS_{MECI}$ (-409.72213166881835)

C	0.28203776	1.08320993	-0.99161119
C	0.00820065	0.08364074	0.89562147
C	0.33626826	-1.25523525	-1.00351863
N	-0.14057654	-0.03908254	-1.63406429
N	0.41682309	1.25819299	0.32629418
N	-0.01466775	0.03315315	2.23173056
N	-0.42713569	-1.01458407	0.21817457
O	0.54695326	2.13246155	-1.76412236
H	-0.02171532	-2.15242406	-1.52981525
H	-0.26763493	-0.82806122	2.70120085
H	0.32808634	0.83328280	2.75142165
H	0.40602490	1.81796694	-2.67516357

$T_1(^3\pi\pi^*)_{min}$ (-409.73453776957820)

C	0.30971137	1.06994491	-1.01015472
C	0.11656769	0.03483388	0.91152965
C	-0.33804236	-1.19368995	-1.06264218
N	0.27964761	-0.09036547	-1.69093305
N	0.22666092	1.26147558	0.31412078
N	0.09864484	0.03439462	2.25883215
N	0.06877991	-1.15949538	0.29837632
O	0.48260305	2.17633478	-1.74227018
H	-0.18845431	-2.15892339	-1.56039526
H	-0.11198442	-0.82390958	2.75481117
H	-0.00000520	0.93244470	2.71961012
H	0.50853489	1.86947630	-2.66473681