Electronic Supplementary Information for:

A 'bottom up', *ab initio* computational approach to understanding fundamental photophysical processes in nitrogen containing heterocycles, DNA bases and base pairs

B. Marchetti, T.N.V. Karsili,* M.N.R. Ashfold * and W. Domcke

University of Bristol, School of Chemistry, Bristol, BS8 1TS

*Corresponding authors: tolga.karsili@bristol.ac.uk, mike.ashfold@bris.ac.uk

COMPUTATIONAL METHODOLOGY

Ground State Keto-Enol Tautomerism

2-hydroxypyridine, 2-hydroxypyrimidine and 2,4-hydroxypyrimidine exhibit keto-enol tautomerism in both the ground and electronically excited states. Using the Gaussian 09 [1] computational package the ground state tautomerism was explored by initially optimising the ground state geometries of the keto/enol tautomers using the Møller-Plesset second order perturbation theory (MP2) along with Dunning's augmented correlation consistent basis set of triple-ξ quality: aug-cc-pVTZ [2] (henceforth AVTZ). Once optimised, these structures were used as initial guesses when calculating the transition state between the keto-enol tautomers. using the Synchronous Transit guided *quasi*-Newton method (QST2) at the MP2/AVTZ level of theory. A minimum energy pathway (MEP) was then constructed using the intrinsic reaction coordinate (IRC) algorithm embedded within Gaussian 09 using the CAM-B3LYP/AVTZ level of theory.

Conical Intersection (CI) searches and potential energy scans

Monomers

Using the Gaussian 09 computational package, the ground and first electronically excited state geometries of all the monomer species were optimised using complete active space self-

consistent field (CASSCF) theory coupled with the 6-31G(*d*) Pople basis set.[3] The geometries of the lowest energy $^1\pi\pi^*/S_0$ CI in each molecule was optimised at the SA2-CASSCF/6-31G(*d*) level of theory using the Coupled Perturbed-MCSCF formulism embedded within Gaussian 09. Table S.1 summaries the active spaces employed for each molecule studied in this Perspective.

Table S.1 – Active spaces, state averaging and basis sets utilised in the various CASSCF and CASPT2 calculations for the species discussed in sections 3.2 and 3.3.

Molecule	**CASSCF optimisations***	**CASPT2 Energies**		
	Orbital space	Orbital space	State Average	Basis Set
Phenol	(8/7)	(10/10)	Lowest 3 singlet states	AVDZ
2-hydroxypyridine	(10/8)	(12/11)	Lowest 4 singlet states	AVDZ
2-pyridone	(10/8)	(12/11)	Lowest 4 singlet states	AVDZ
4-hydroxypyrimidine	(10/8)	(10/9)	Lowest 4 singlet states† Lowest 3 singlet states§	AVDZ
4-pyrimidone	(10/8)	(10/9)	Lowest 6 singlet states† Lowest 3 singlet states§	AVDZ
2,4-dihydroxypyrimidine	(10/8)	(14/10)	Lowest 6 singlet states	AVDZ
Uracil	(12/9)	(14/10)	Lowest 6 singlet states	AVDZ
Thymine	(12/9)	(14/10)	Lowest 6 singlet states	AVDZ
Cytosine	(10/8)	(14/10)	Lowest 4 singlet states	VDZ
Indole	(10/8)	(10/10)	Lowest 4 singlet states	VDZ
7-azaindole	(10/8)	(12/10)† (10/10)§	Lowest 6 singlet states† Lowest 4 singlet states§	VDZ
5,7-azaindole	(10/8)	(12/10)† (10/10)§	Lowest 6 singlet states† Lowest 5 singlet states§	VDZ
Purine	(10/8)	(12/10)† (10/10)§	Lowest 6 singlet states† Lowest 5 singlet states§	VDZ
Adenine	(10/8)	(14/10)† (10/10)§	Lowest 6 singlet states$^{\dagger, \S}$	VDZ
Guanine	(10/8)	(14/10)	Lowest 4 singlet states	VDZ

*These optimisations are of the ground state, first excited S_1 state, conical intersections and, where present, TS geometries. Self-consistent field (SCF) convergence thresholds were fixed at 10^{-8} for ground state, 10^{-7} for the excited state and 10^{-6} for the CI and TS optimisations.

† Level of theory used for the single point energy calculations reported in Tables 2 and 3

§ Level of theory used to calculate the LIIC along the Q_{oop} coordinate.

Using the Molpro 2010.1 computational package, two potential energy curves were constructed. The first comprised a CASPT2/aug-cc-pVDZ (AVDZ) rigid body (unrelaxed) scan along the R_{X-H} (where X = O or N) coordinate (for phenol, 2-hydroxypyridine and 2-pyridone), whilst freezing all other internal degrees of freedom at the ground state CASSCF geometry. At this same level of theory, a second PEC was constructed along the linearly interpolated internal coordinate (LIIC) that connected the optimised geometry of the ground state to the optimised geometry of the CI (*v.* CASPT2 energies in table S.I.). The latter scan provided an initial guess for any likely transition states (TSs) that might exist between the vertical/optimised $^1\pi\pi^*$ state and the $^1\pi\pi^*/S_0$ CI. Such TS initial guesses were optimised using the Berny optimisation method in Gaussian 09 at the CASSCF/6-31G(*d*) level of theory. Single point energy calculations of the optimised TS and S_1 minimum energy structures were undertaken using CASPT2/AVDZ or CASPT2/cc-pVDZ (VDZ) levels of theory (*v.* CASPT2 energies in table S.I for details). All CASPT2 calculations required an imaginary level shift of 0.5 E_h in order to avoid intruder state effects.

DNA and RNA nucleosides

Gaussian 09 was used to compute ground state minimum energy geometries for four ribonucleosides: 5-methyluridine, cytidine, adenosine and guanosine. Table S.2 summaries the active spaces employed for each of these DNA/RNA nucleosides.

Table S.2 – Active spaces, state averaging and basis sets utilised in the various CASSCF and CASPT2 calculations for the nucleosides discussed in section 3.4.

Molecule	**Optimisations***		**CASPT2 Energies****	
	S_0	$^1\pi\pi^*/S_0$ CI	Orbital space	State Average
5-methyluridine	MP2	CASSCF(8,8)	(12/9)	Lowest 6 singlets
Cytidine	MP2	CASSCF(8,8)	(14/10)	Lowest 4 singlets
Adenosine	MP2	CASSCF(8,8)	(10/10)	Lowest 6 singlets
Guanosine	MP2	CASSCF(10,8)	(12/11)	Lowest 6 singlets

* Optimisations of the S_0 and $^1\pi\pi^*/S_0$ CI were performed using the 6-31G(*d*) basis set. Self-consistent field (SCF) convergence thresholds were fixed at 10^{-8} for the ground state (except for 5-methyluridine, where it was fixed to 10^{-6}) and at 10^{-6} for CI optimisation.

** CASPT2 energies were calculated using the VDZ basis set.

The PECs along the electron driven hydrogen transfer coordinate between the O-H methanolic group on the ribose sugar and the N atom in the ring of the DNA bases were explored using ADC(2)/VDZ in Turbomole[4]. The S_0 state was scanned by extending $R_{O–H}$ and relaxing the rest of the molecular parameters; the energies of the lowest excited states were then calculated at the relaxed S_0 geometries. A relaxed scan was also constructed for the $S_1(^1CT)$ state, and the S_0 state energies calculated at each relaxed ^1CT state geometry. The S_0 and S_1 adiabatic potential energy profiles were then calculated along the LIIC connecting the S_0 minimum geometry and the relaxed ^1CT structure at the shortest R_{OH} distance.

DNA base pairs.

The minimum energy geometry of the ground state A-T base pair was optimised at the MP2/6-31G(*d*) level of theory. The out-of-plane $^1\pi\pi^*/S_0$ CI was optimised at the CASSCF(8,8)/6-31G(d) level. Excited state optimisations were more challenging and required that we use somewhat less computationally demanding theories. As with the nucleosides, CASPT2(10,10)/VDZ PECs (on a SA8-CASSCF reference wavefunction) were computed along the relevant out-of-plane deformations. For both A-T and G-C isolated pairs ADC(2)/VDZ PECs along the EDPT coordinate ($R_{N\text{-}H\text{--}N(O)}$) were constructed, both of which lead to low energy $^1\pi\pi^*/S_0$ CIs.

PECs along the EDPT driving coordinate were calculated for the neutral and radical-anion G-C and A-T base pairs at the TD-DFT/CAM-B3LYP/VDZ level of theory. The S_0 (neutral base pair) was scanned as a function of the relevant $R_{N–H\cdots N(O)}$ stretch coordinate by fixing $R_{N–H\cdots N(O)}$ at various values and allowing the remaining internal degrees of freedom to relax to their respective minima, and the energies of the D_0 state (radical-anion pair) then computed at these relaxed S_0 geometries. Relaxed scans along the EDPT driving coordinate were also computed for the D_0 state. Potential energy profiles along the LIIC connecting the respective S_0 minima at $R_{N\text{-}H}$ ~ 1.0 Å and the D_0 minima at $R_{N\text{-}H}$ ~ 1.2 Å (in the case of the G-C base pair) and ~1.3 Å (for the A-T base pair) were computed in order to assess the feasibility of the reaction path from the neutral to the radical-anion state.

RESULTS

PECs along the relevant Q_{oop} coordinate for each of the nitrogen containing heterocycles featured in this Perspective, along with illustrations of the geometry of the $^1\pi\pi^*/S_0$ MECI in each case. The meanings of the filled and open points are is in figs. 8-10 of the main text.

Figure S.1: Phenol

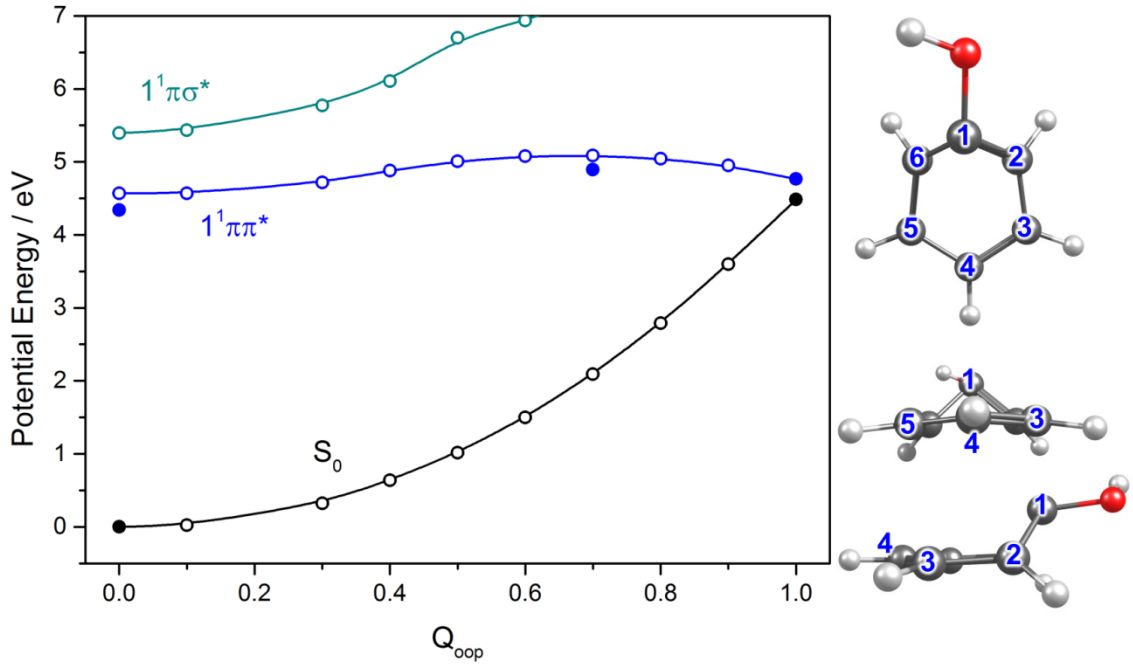


Figure S.2: 2-hydroxypyridine

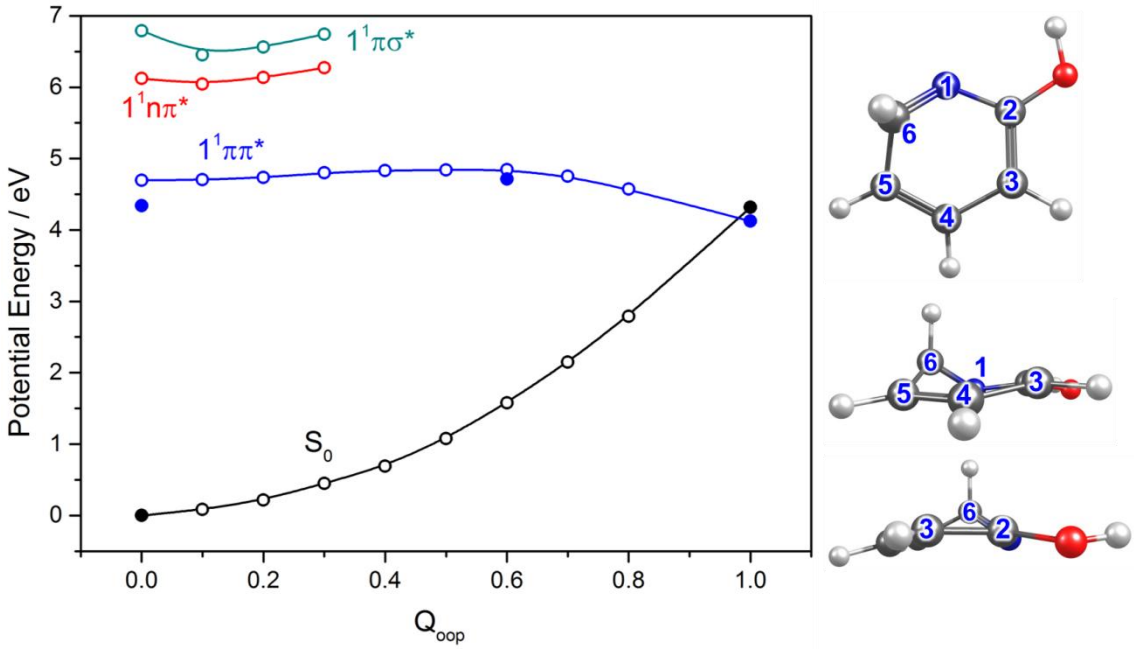


Figure S.3: 2-pyridone

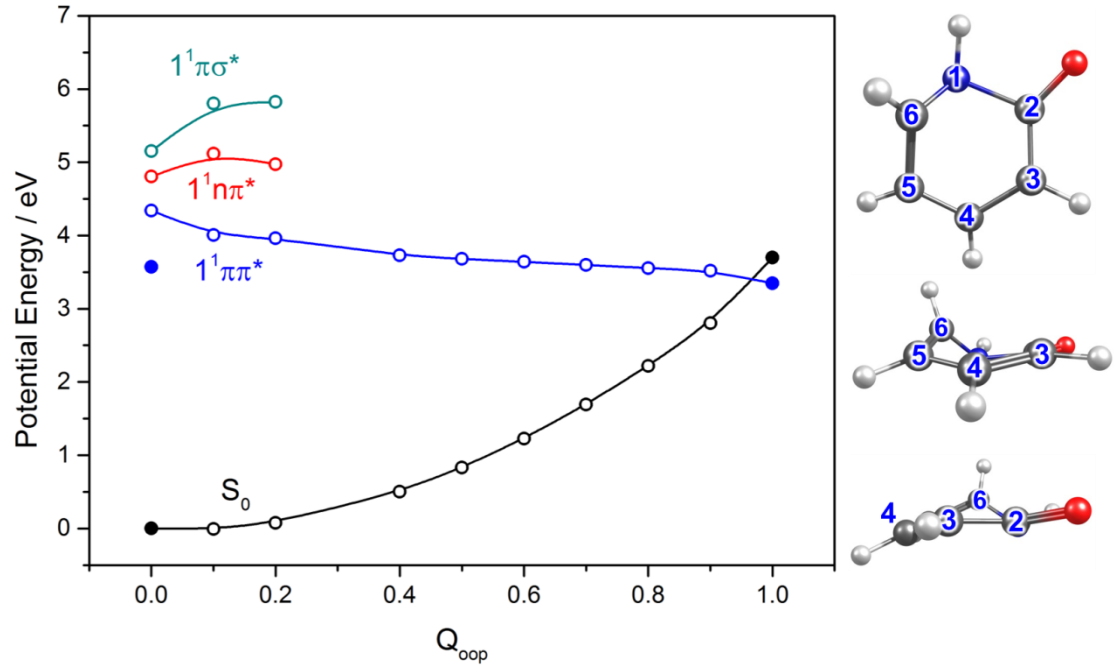


Figure S.4: 4-hydroxypyrimidine

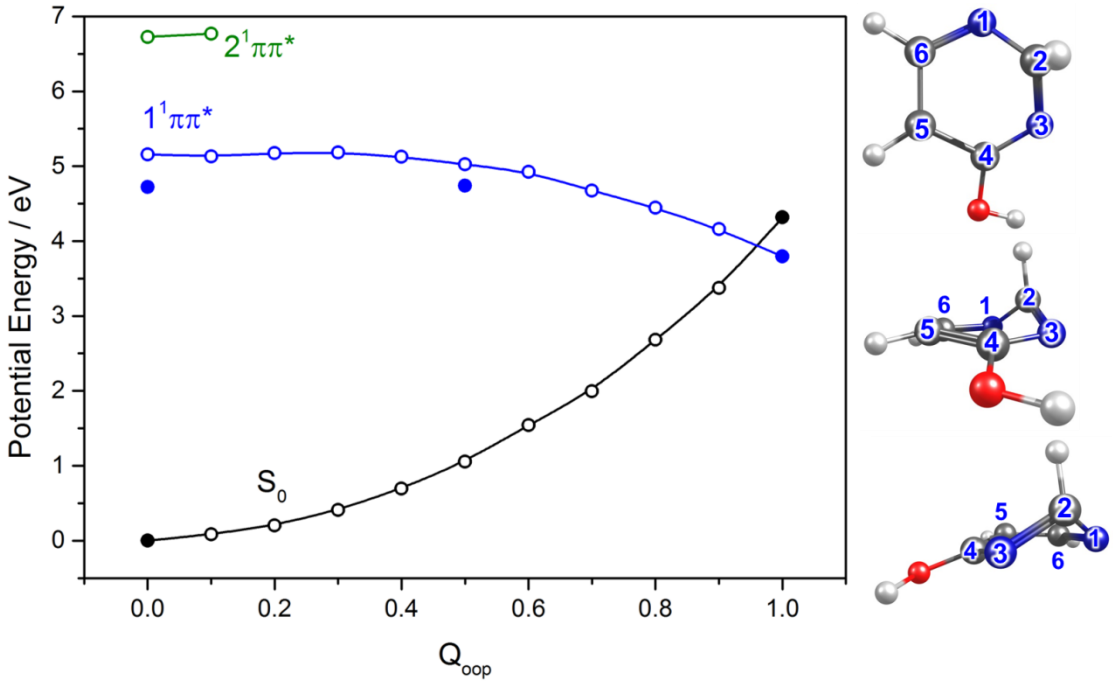


Figure S.5: 4-pyrimidone

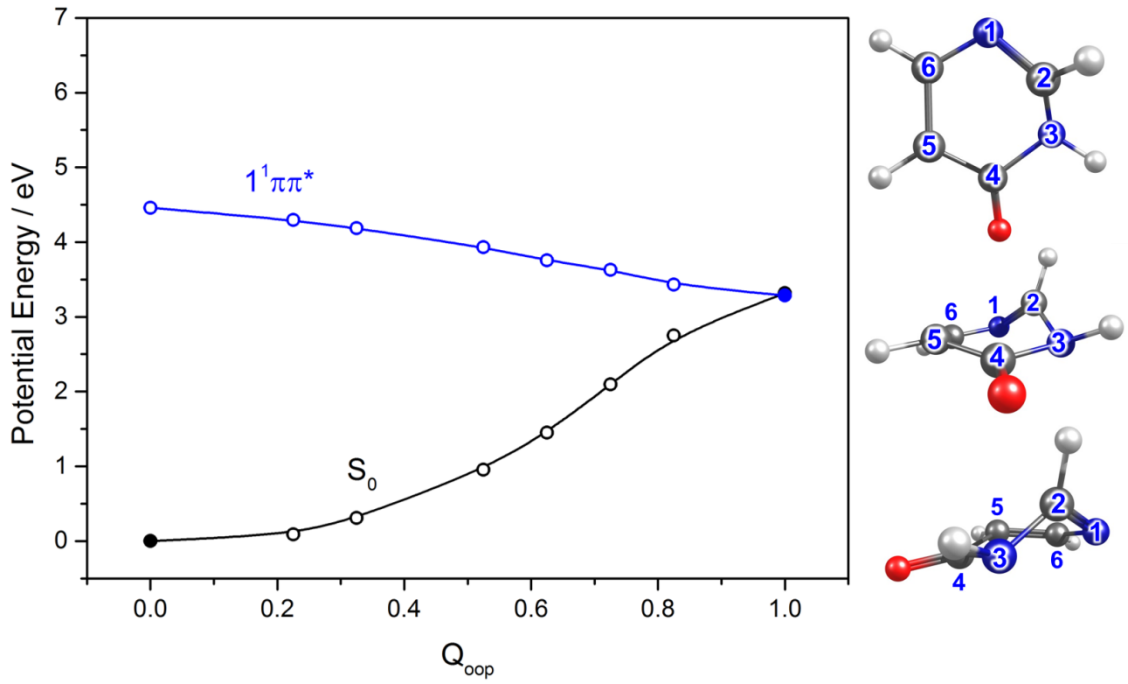


Figure S.6: 2,4-dihydroxypyridine

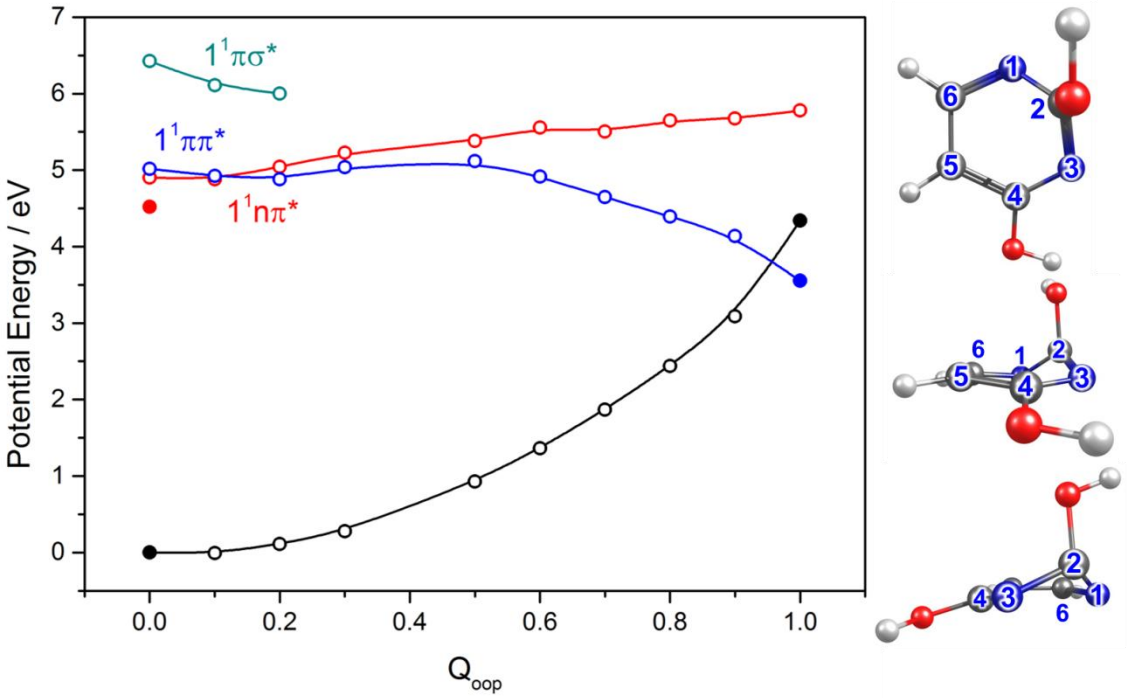


Figure S.7: uracil

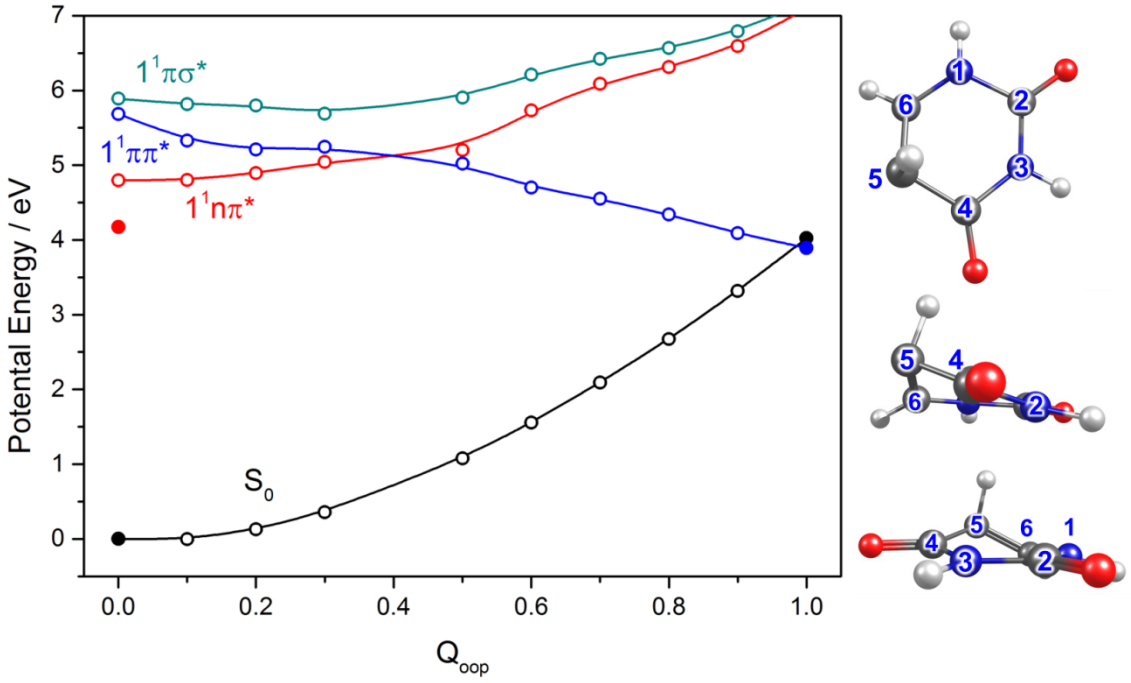


Figure S.8: thymine

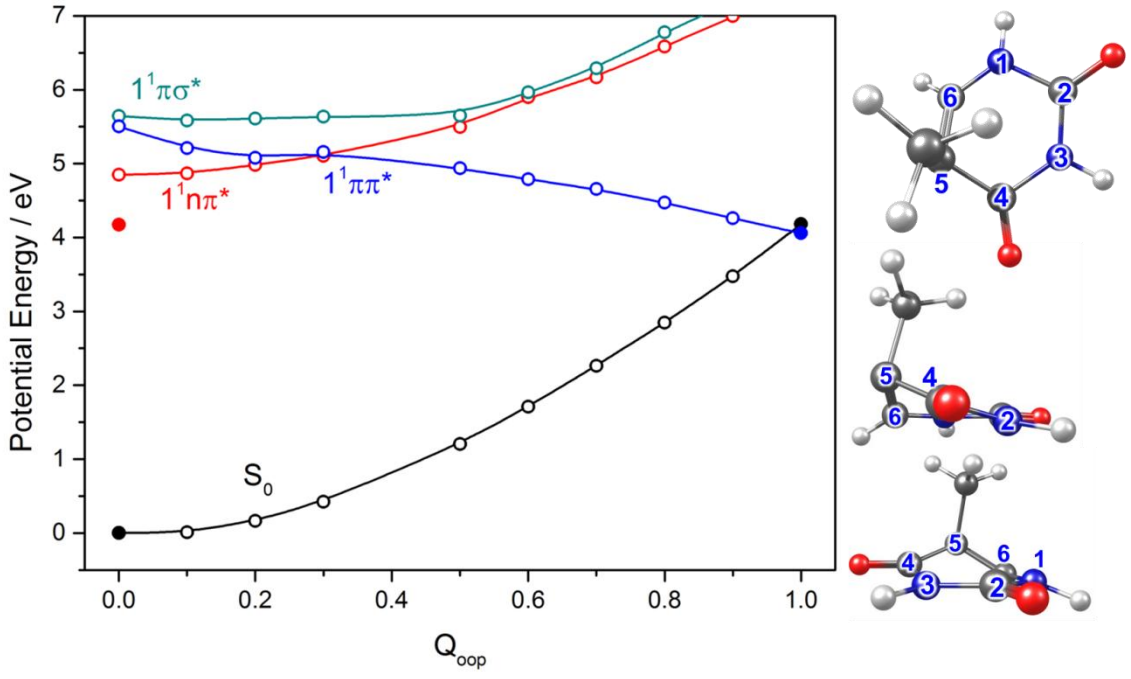


Figure S.9: cytosine

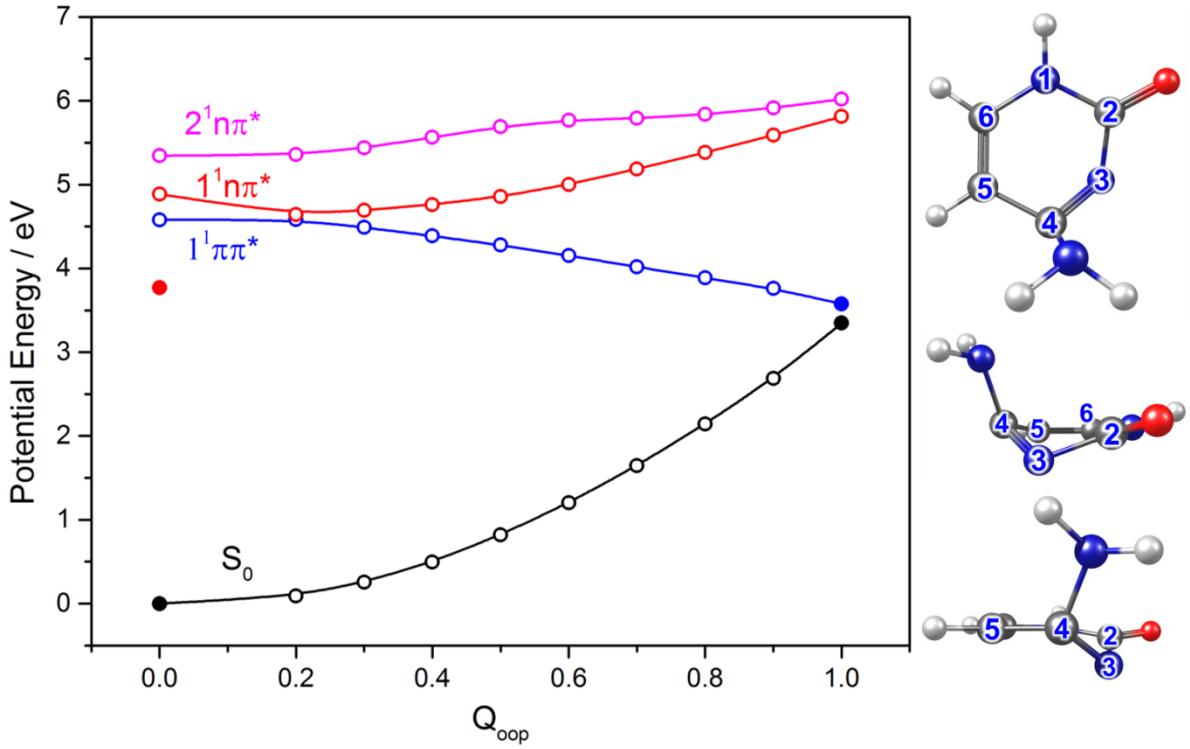


Figure S.10: indole

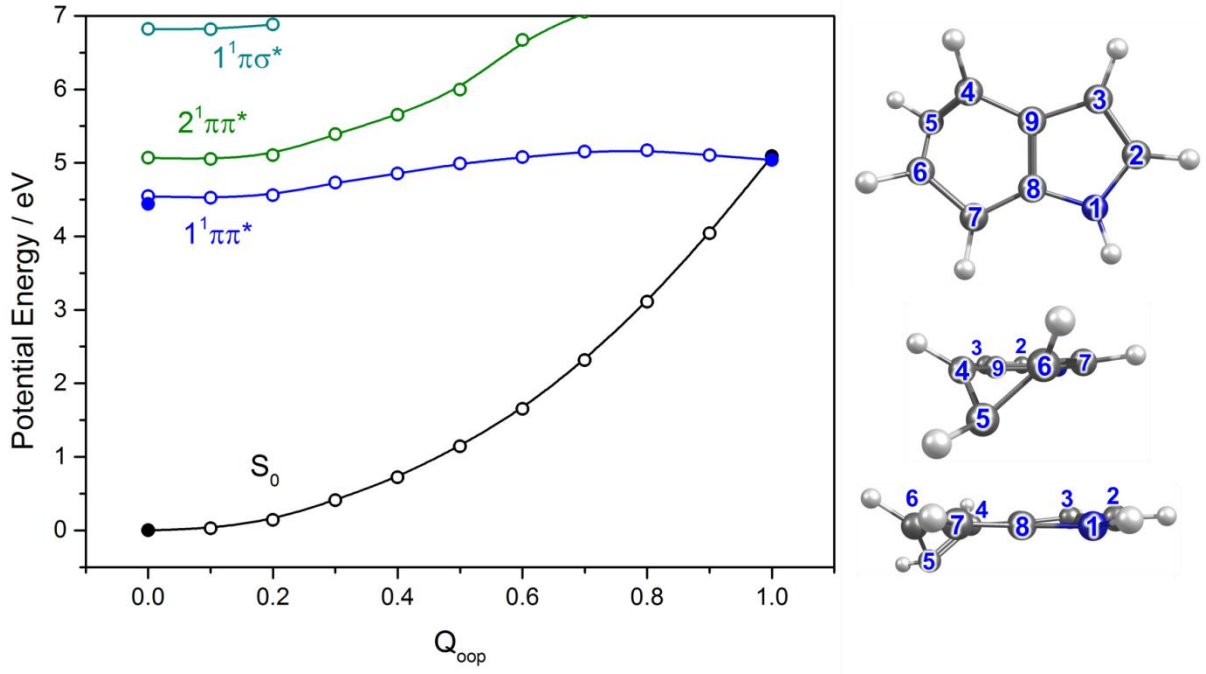


Figure S.11: 7-azaindole

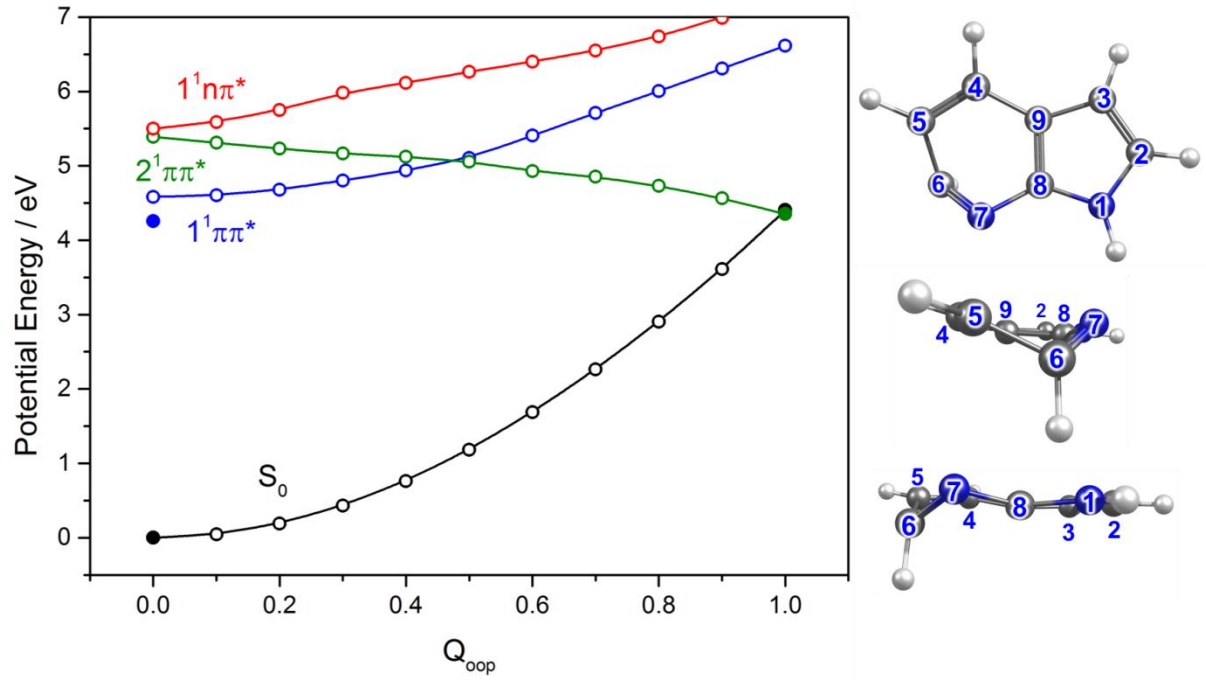


Figure S.12: 5,7-azaindole

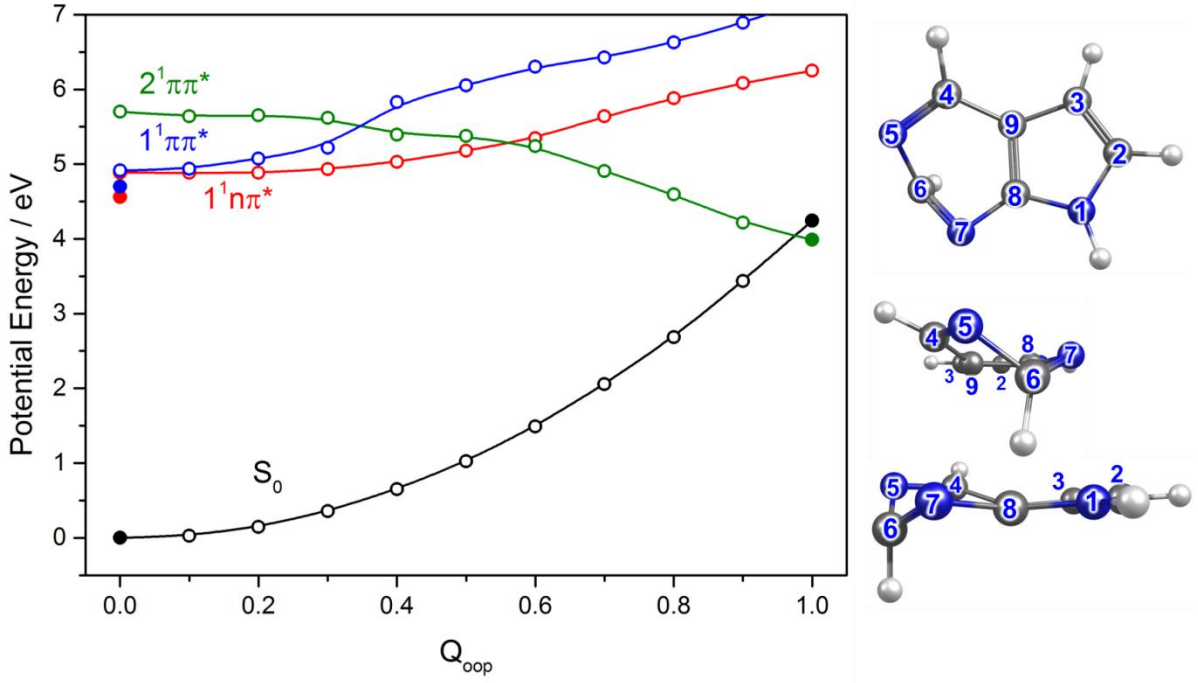


Figure S.13: purine

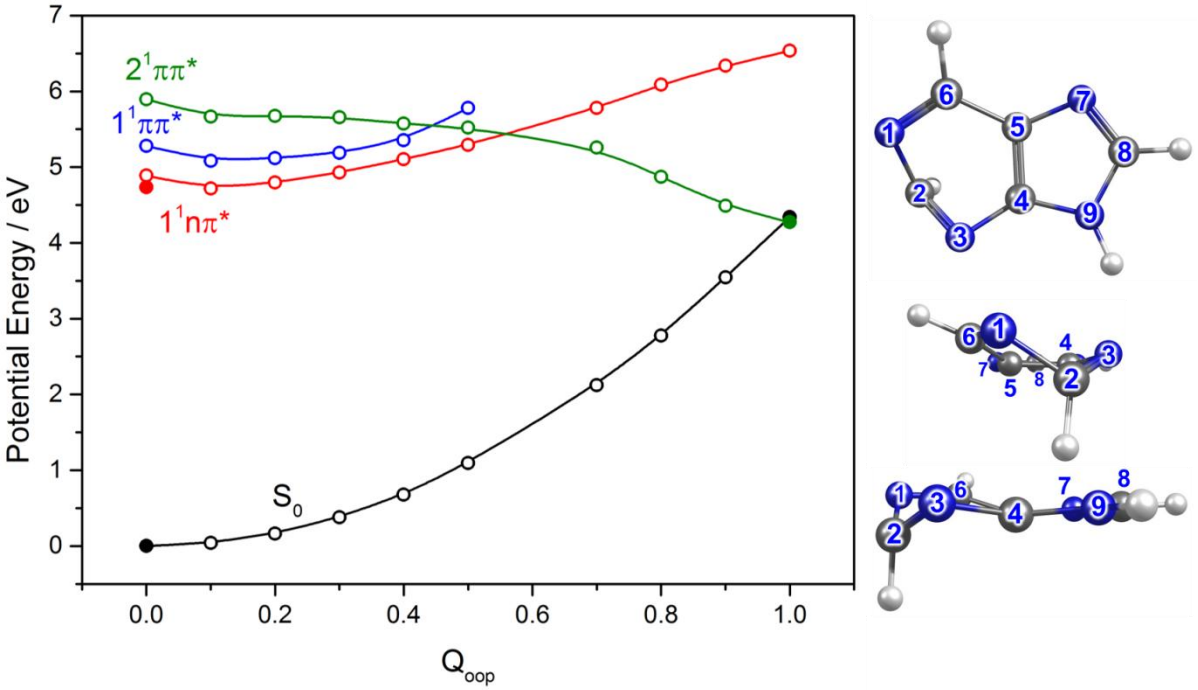


Figure S.14: adenine

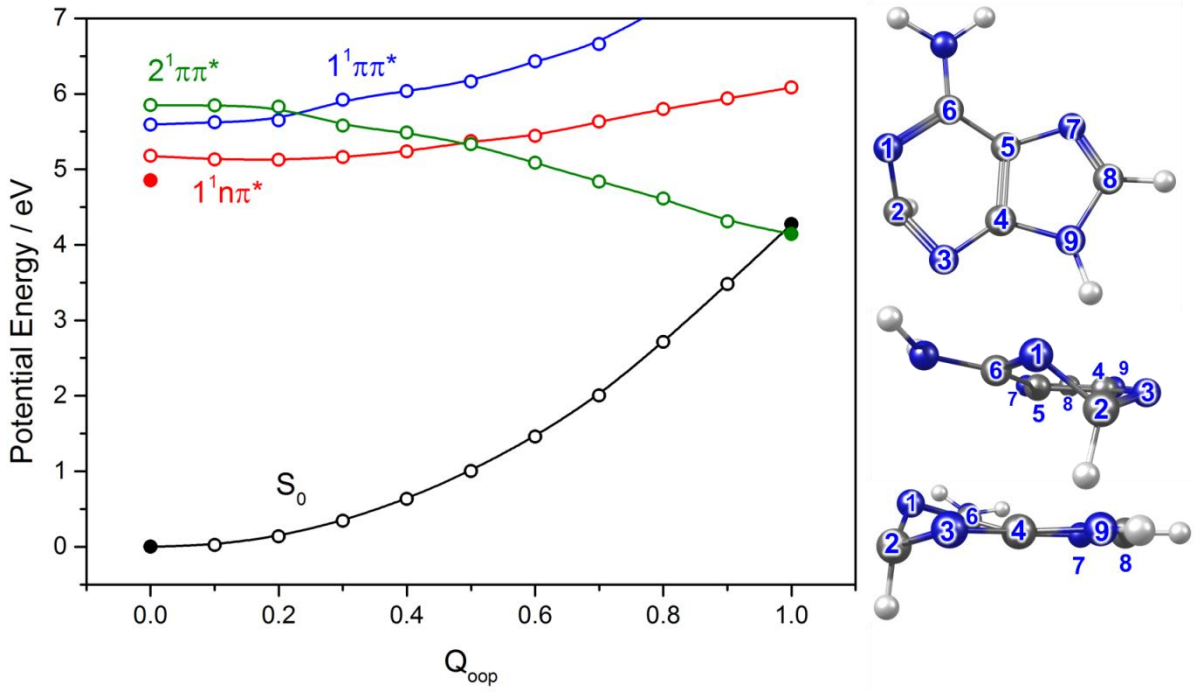
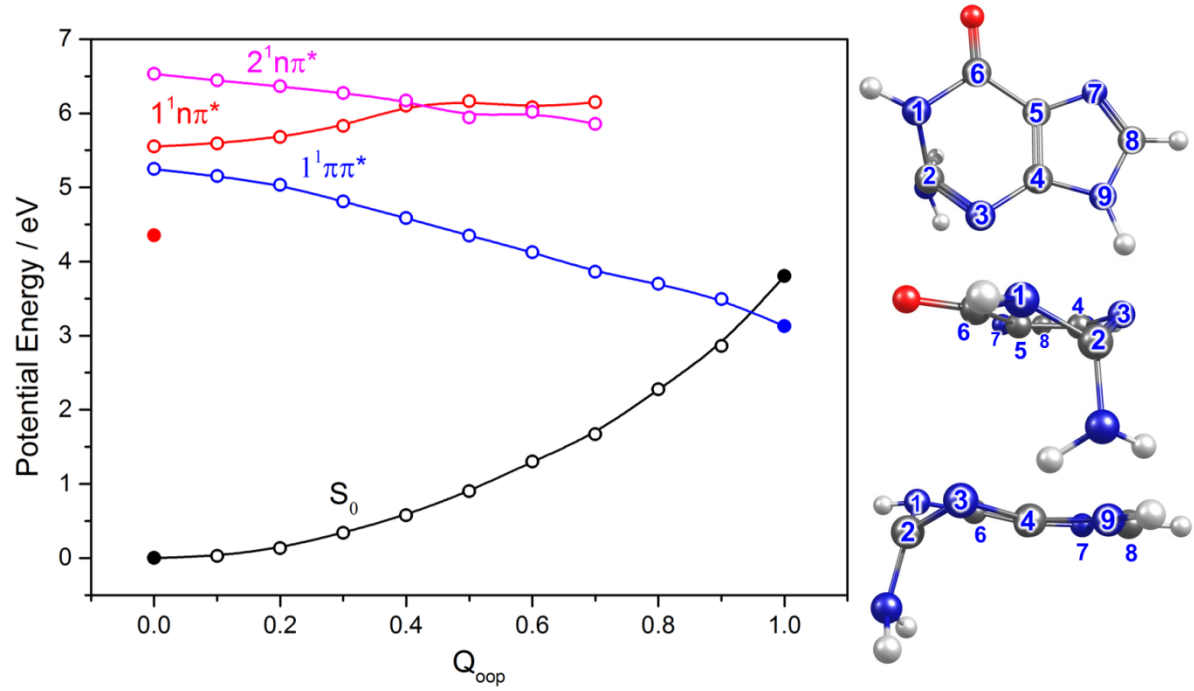
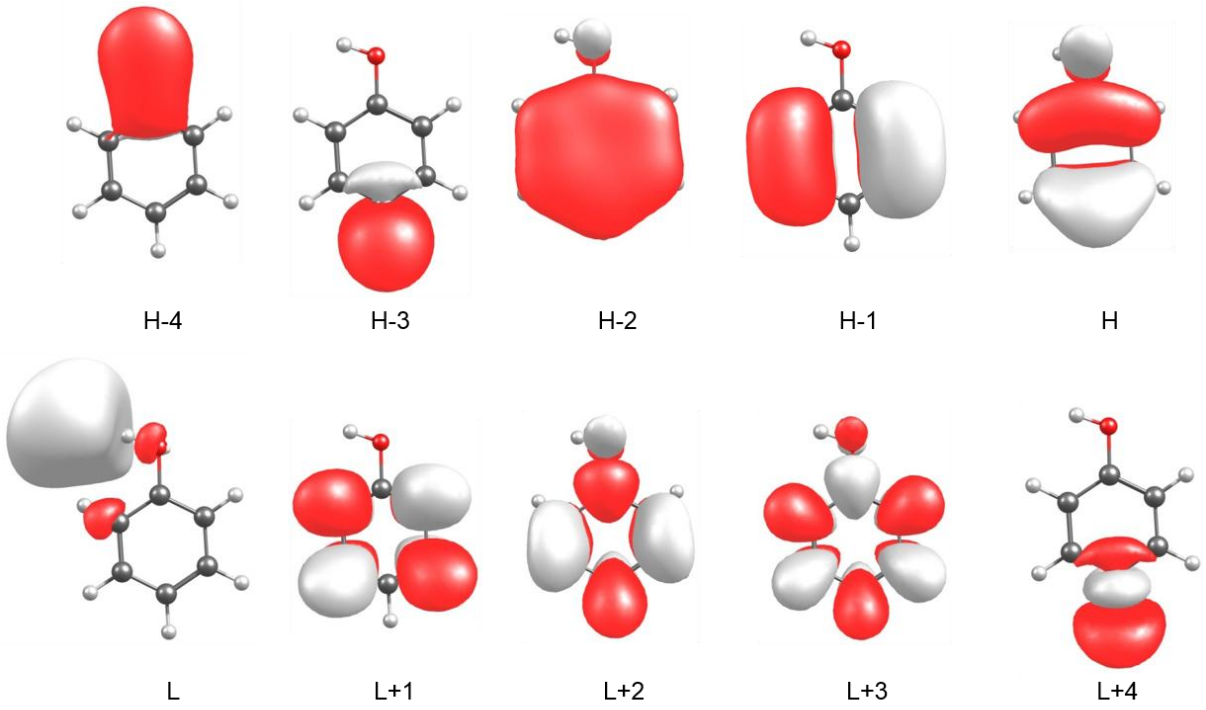


Figure S.15: guanine



Active orbital space used for the CASSCF and CASPT2 calculations and dominant orbital excitations that contribute to the first few excited singlet states of the monomers featured in figs. 1 and 2 (apart from those already shown in fig. 7 of the main text), the nucleosides in fig. 3 and the A-T base pair.

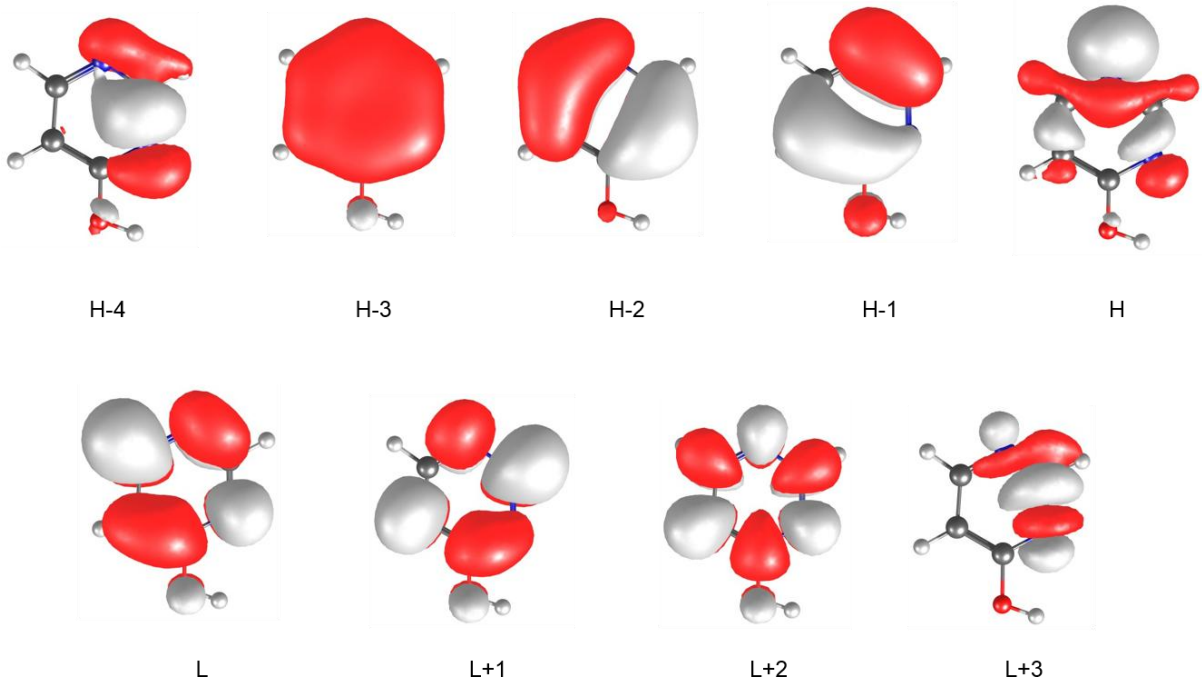
Figure S.16: phenol



$S_1/1^1\pi\pi^*$: H→L+1

$S_2/1^1\pi\sigma^*$: H→L

Figure S.17: 4-hydroxypyrimidine

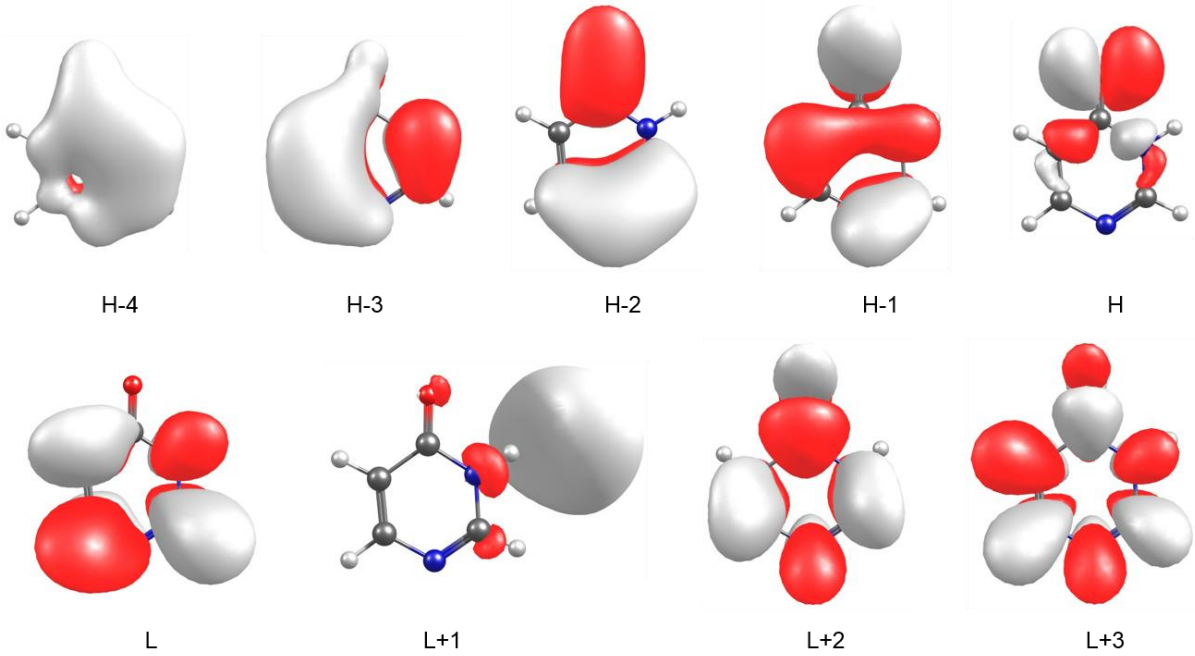


$S_1/1^1n\pi^*$: H→L+1

$S_2/1^1\pi\pi^*$: H→L

$S_3/2^1n\pi^*$: H–1→L

Figure S.18: 4-pyrimidone

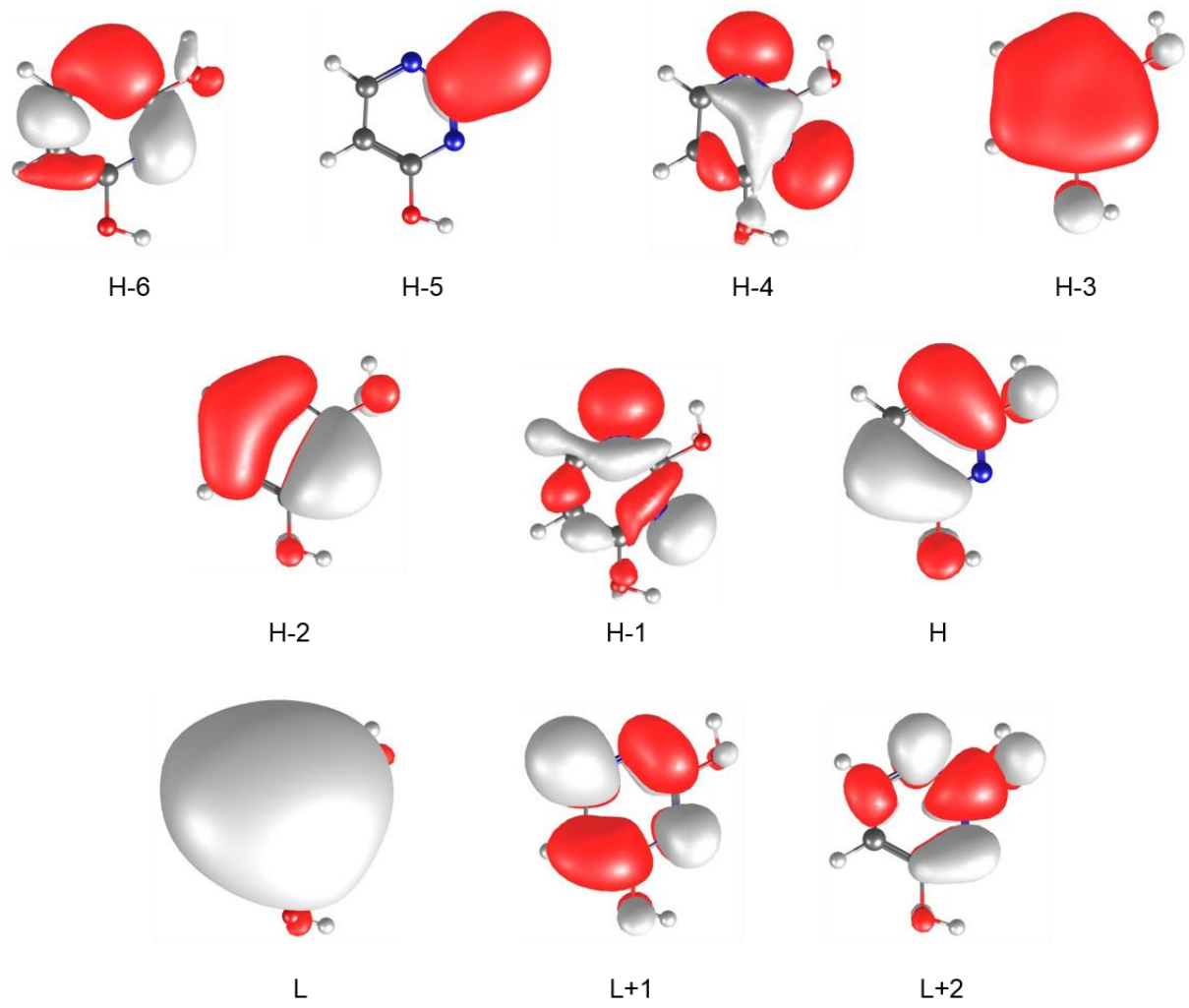


$S_1/1^1\pi\pi^*$: H–1→L

$S_2/1^1n\pi^*$: H→L

$S_3/2^1n\pi^*$: H→L+2

Figure S.19: 2,4-dihydroxypyrimidine

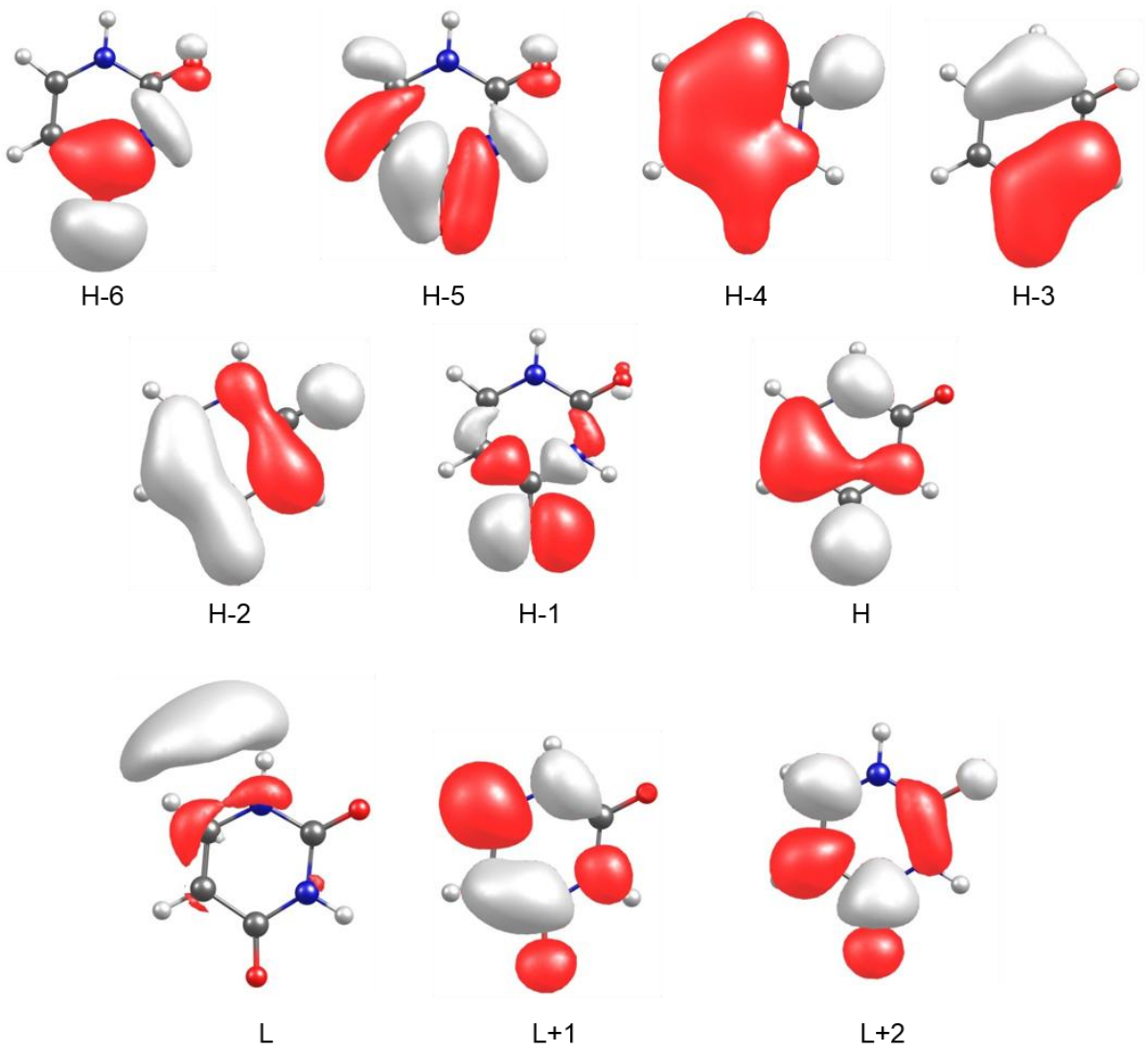


$S_1/1^1n\pi^*$: H–1→L+1

$S_2/1^1\pi\pi^*$: H→L+1

$S_3/1^1\pi\sigma^*$: H→L

Figure S.20: uracil

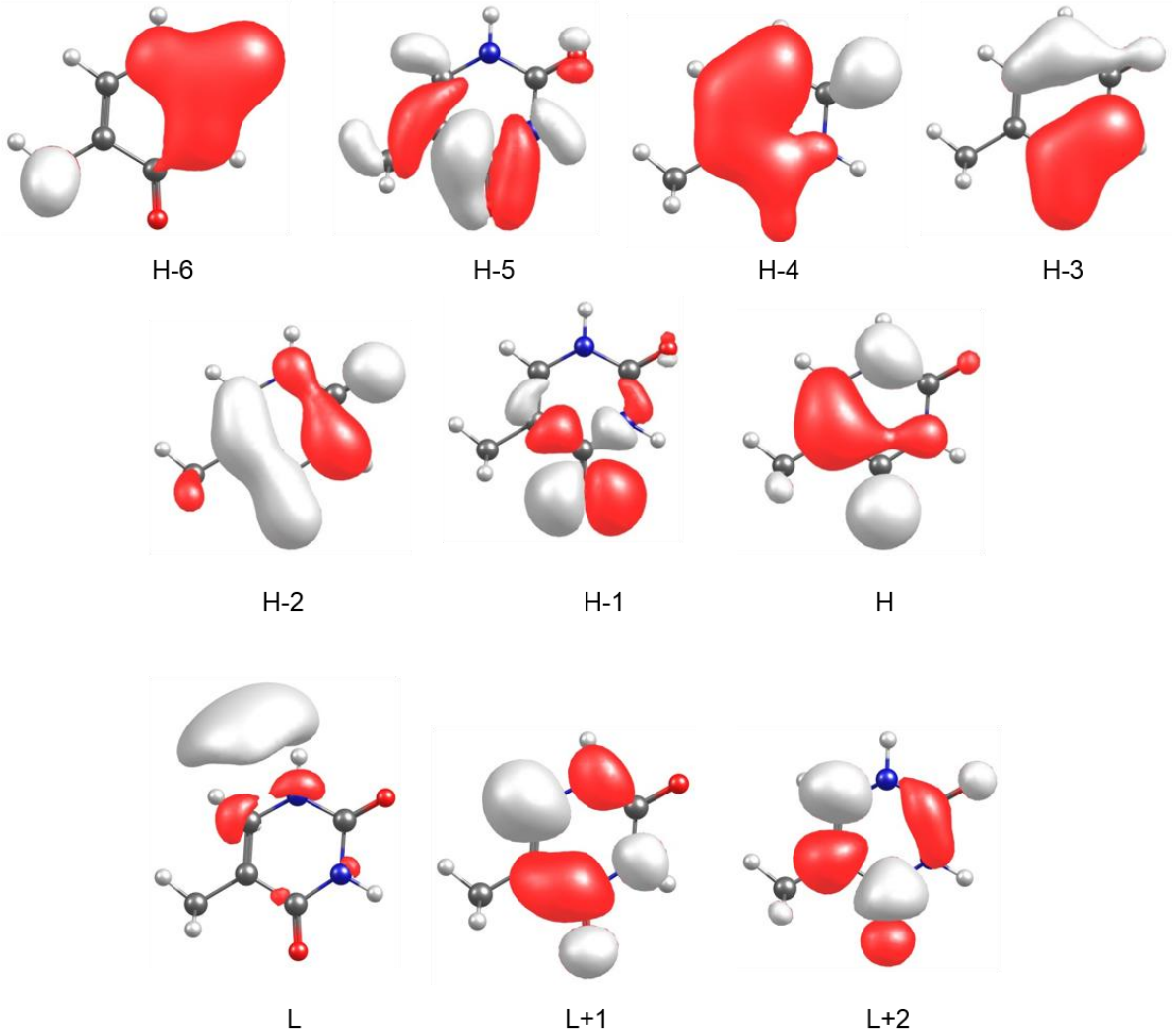


$S_1/1^1n\pi^*$: H–1→L+1

$S_2/1^1\pi\pi^*$: H→L+1

$S_3/1^1\pi\sigma^*$: H→L

Figure S.21: thymine

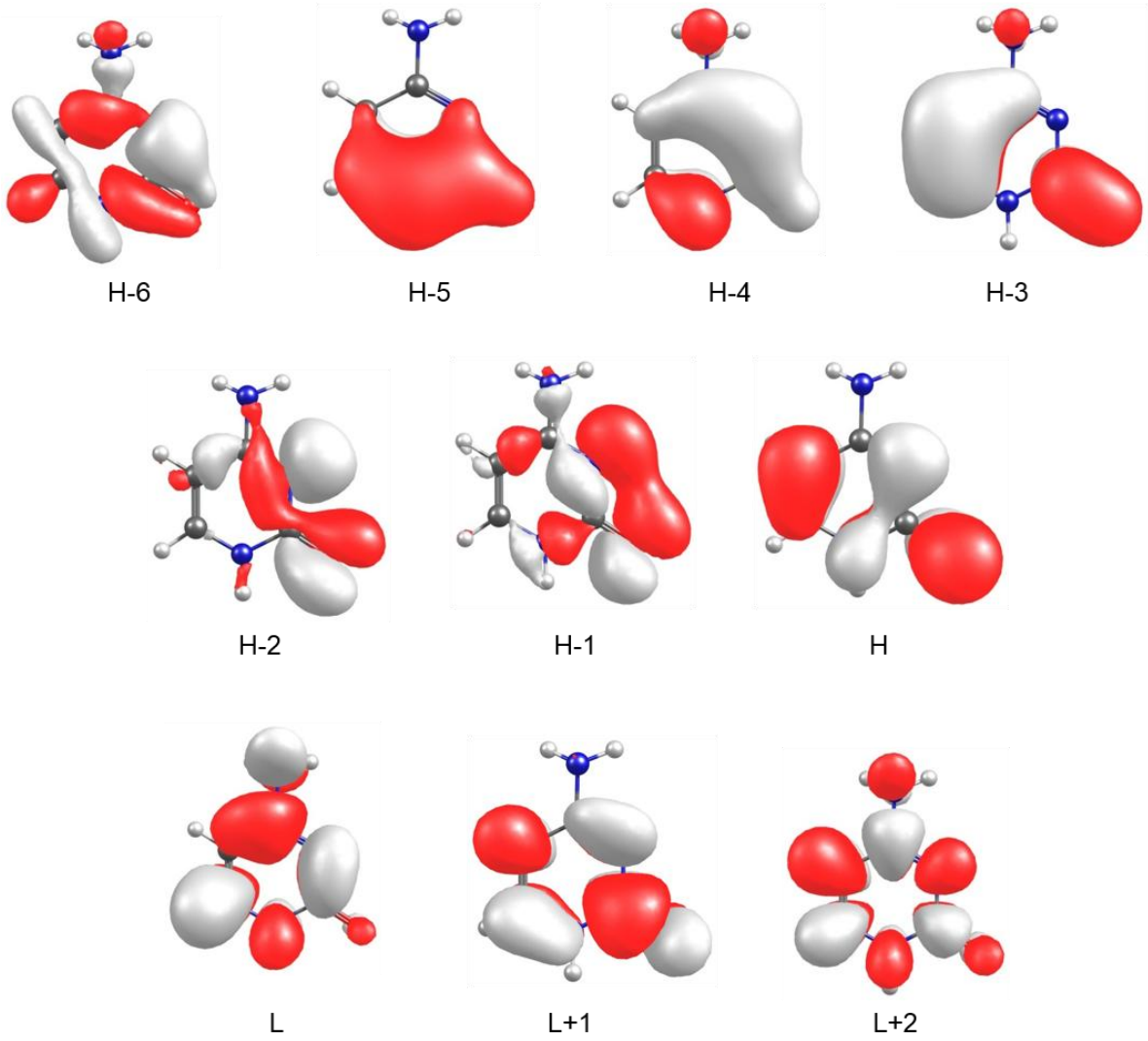


$S_1/1^1n\pi^*$: H–1→L+1

$S_2/1^1\pi\pi^*$: H→L+1

$S_3/1^1\pi\sigma^*$: H→L

Figure S.22: cytosine

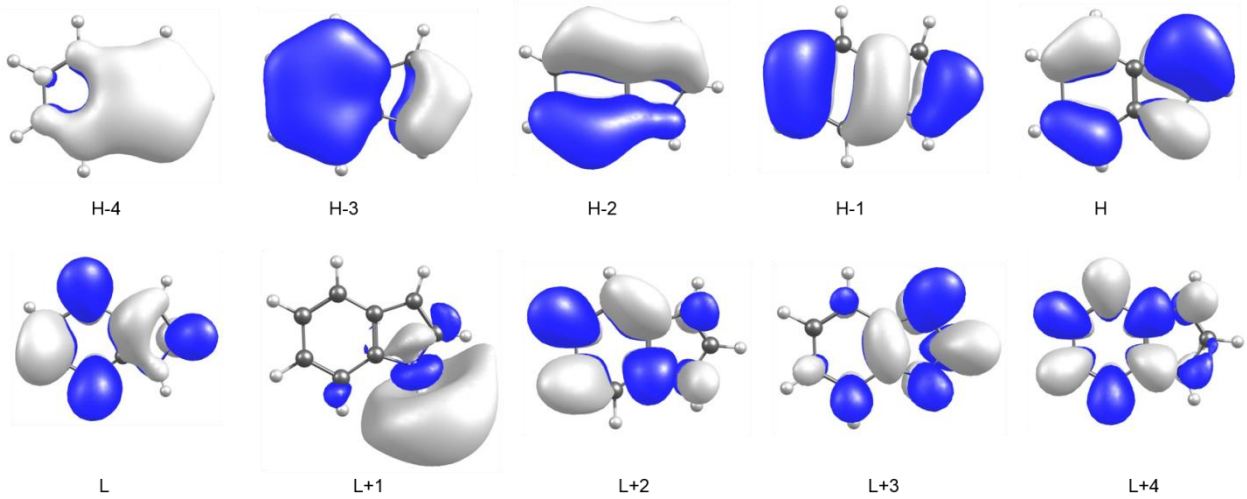


$S_1/1^1\pi\pi^*$: H→L

$S_2/1^1n\pi^*$: H–1→L

$S_3/2^1n\pi^*$: H–2→L

Figure S.23: indole

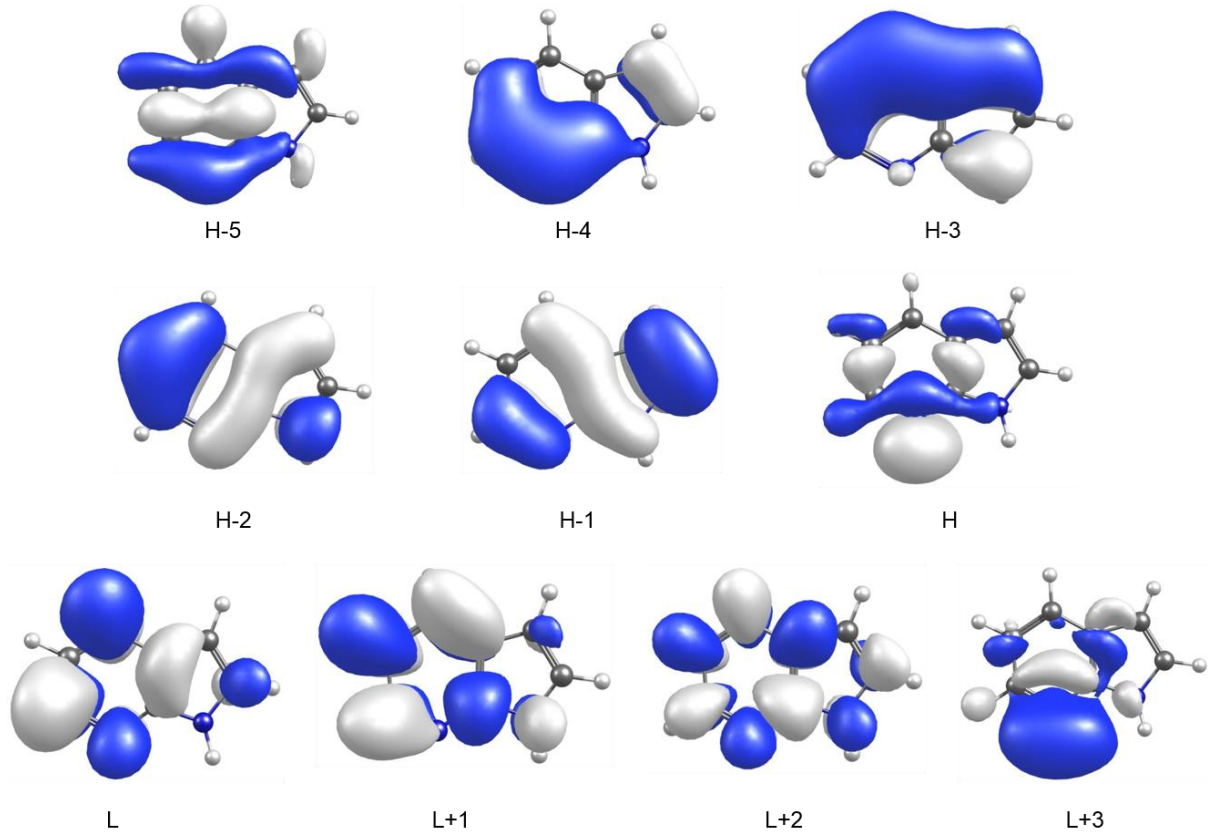


$S_1/1^1\pi\pi^*$: H–1→L and H→L+2

$S_2/2^1\pi\pi^*$: H→L

$S_3/1^1\pi\sigma^*$: H→L+1

Figure S.24: 7-azaindole

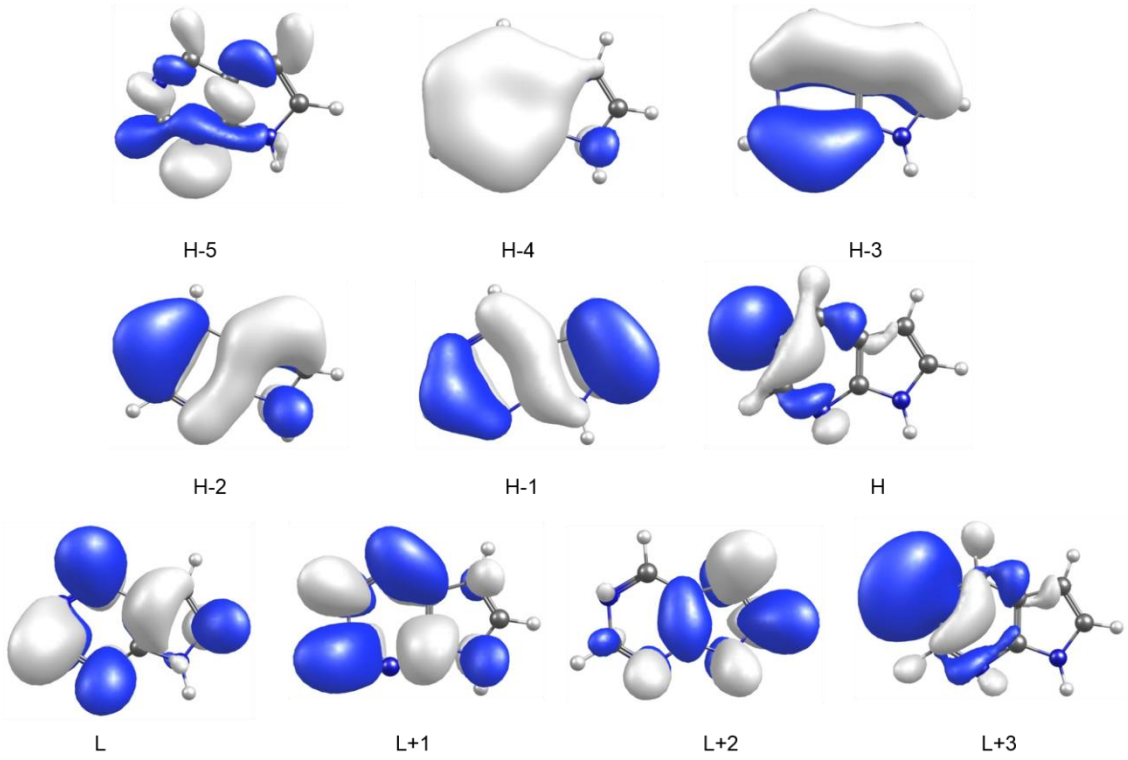


$S_1/1^1\pi\pi^*$: H–1→L+1 and H–2→L

$S_2/2^1\pi\pi^*$: H–1→L

$S_3/1^1n\pi^*$: H→L

***Figure S.25**: 5,7-azaindole*

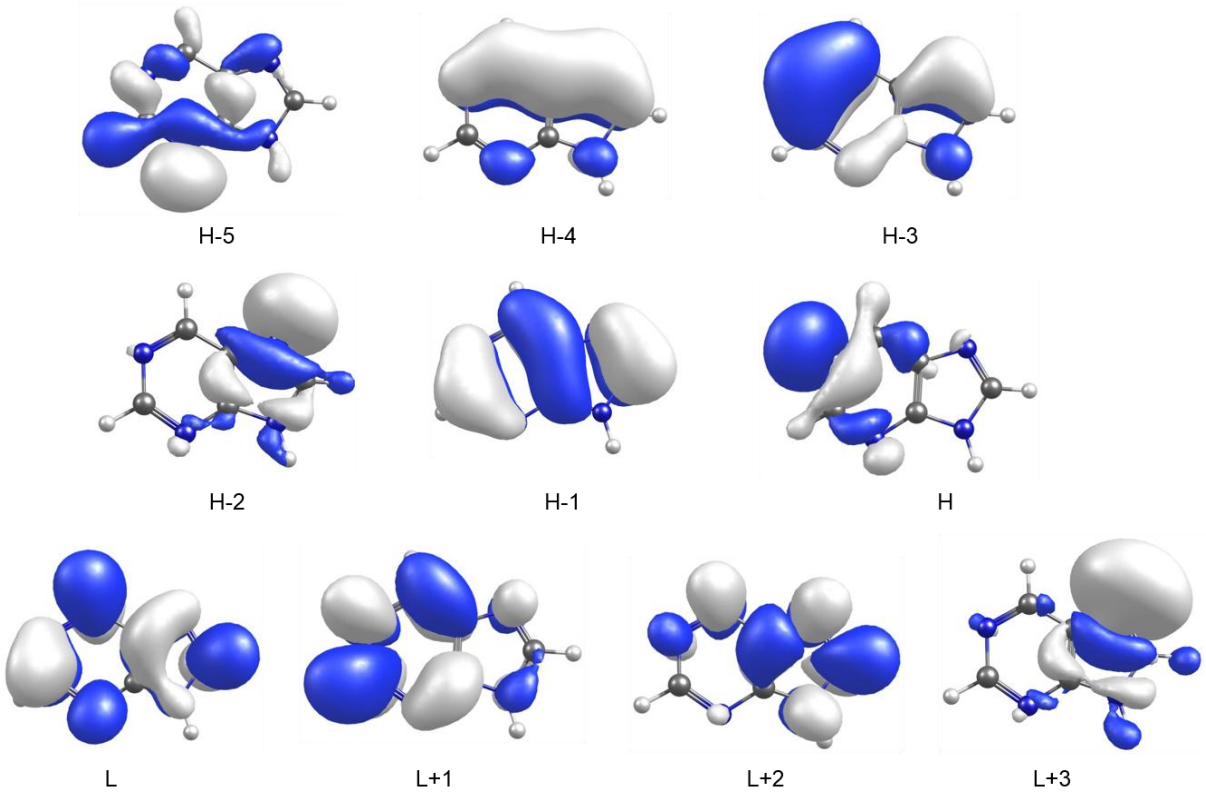


$S_1/1^1n\pi^*$: H→L

$S_2/1^1\pi\pi^*$: H–1→L+1 and H–2→L

$S_3/2^1\pi\pi^*$: H–1→L

Figure S.26: purine

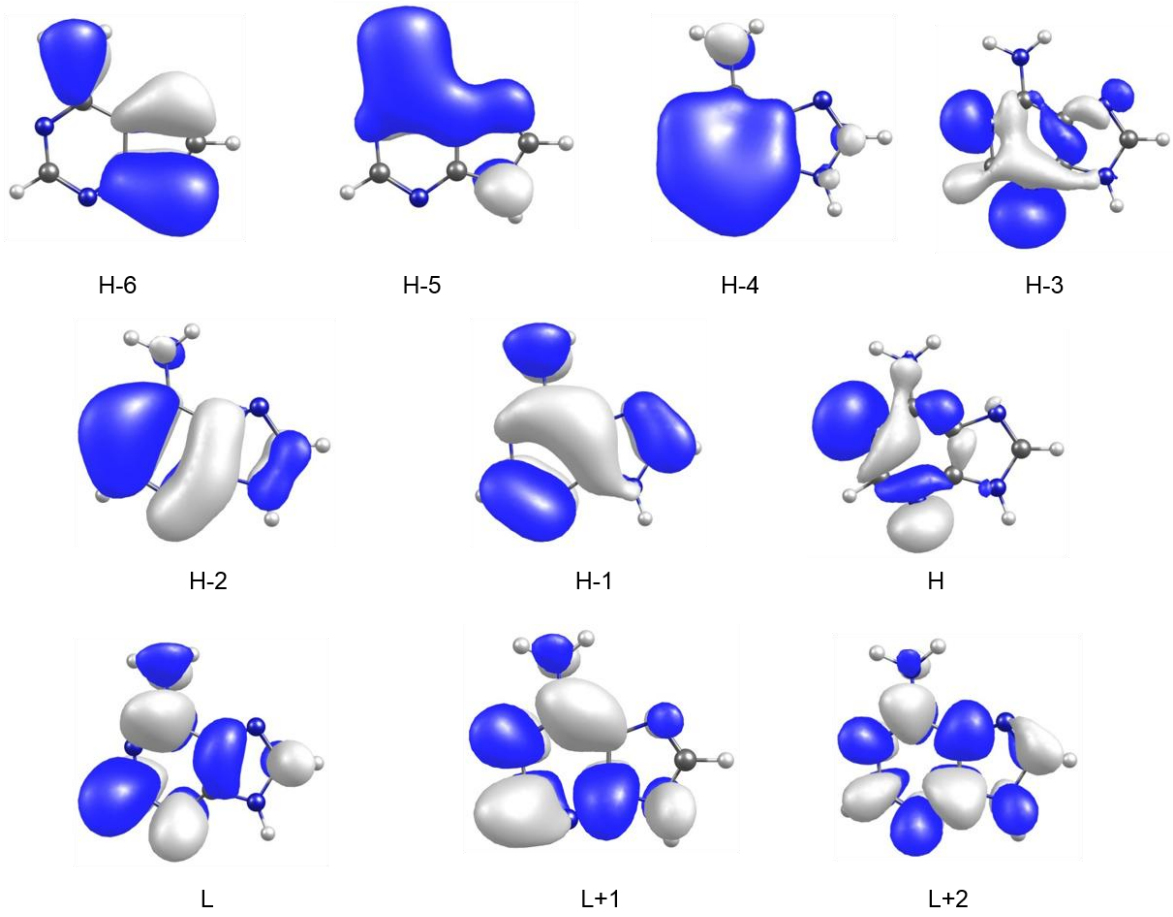


$S_1/1^1n\pi^*$: H→L

$S_2/1^1\pi\pi^*$: H–1→L+1 and H–2→L

$S_3/2^1\pi\pi^*$: H–1→L

Figure S.27: adenine

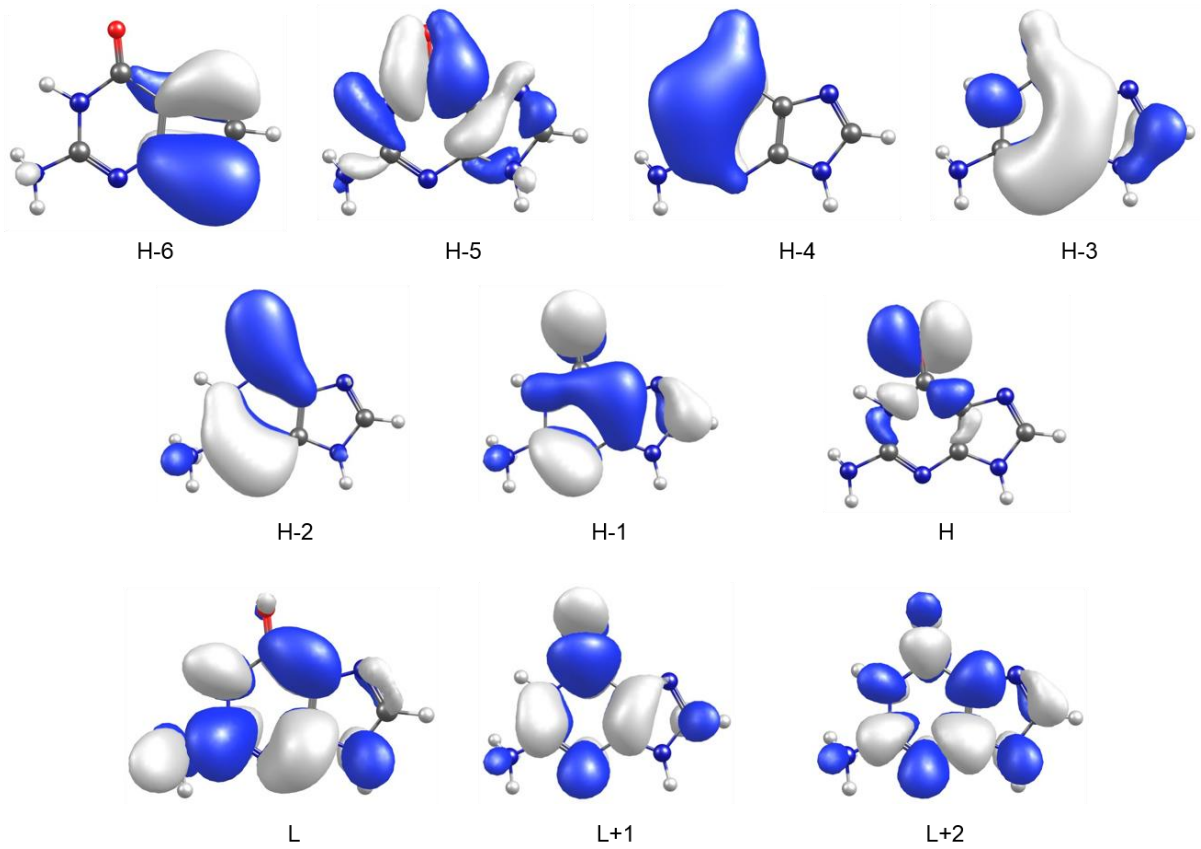


$S_1/1^1n\pi^*$: H→L

$S_2/1^1\pi\pi^*$: H–1→L+1 and H–2→L

$S_3/2^1\pi\pi^*$: H–1→L

Figure S.28: guanine

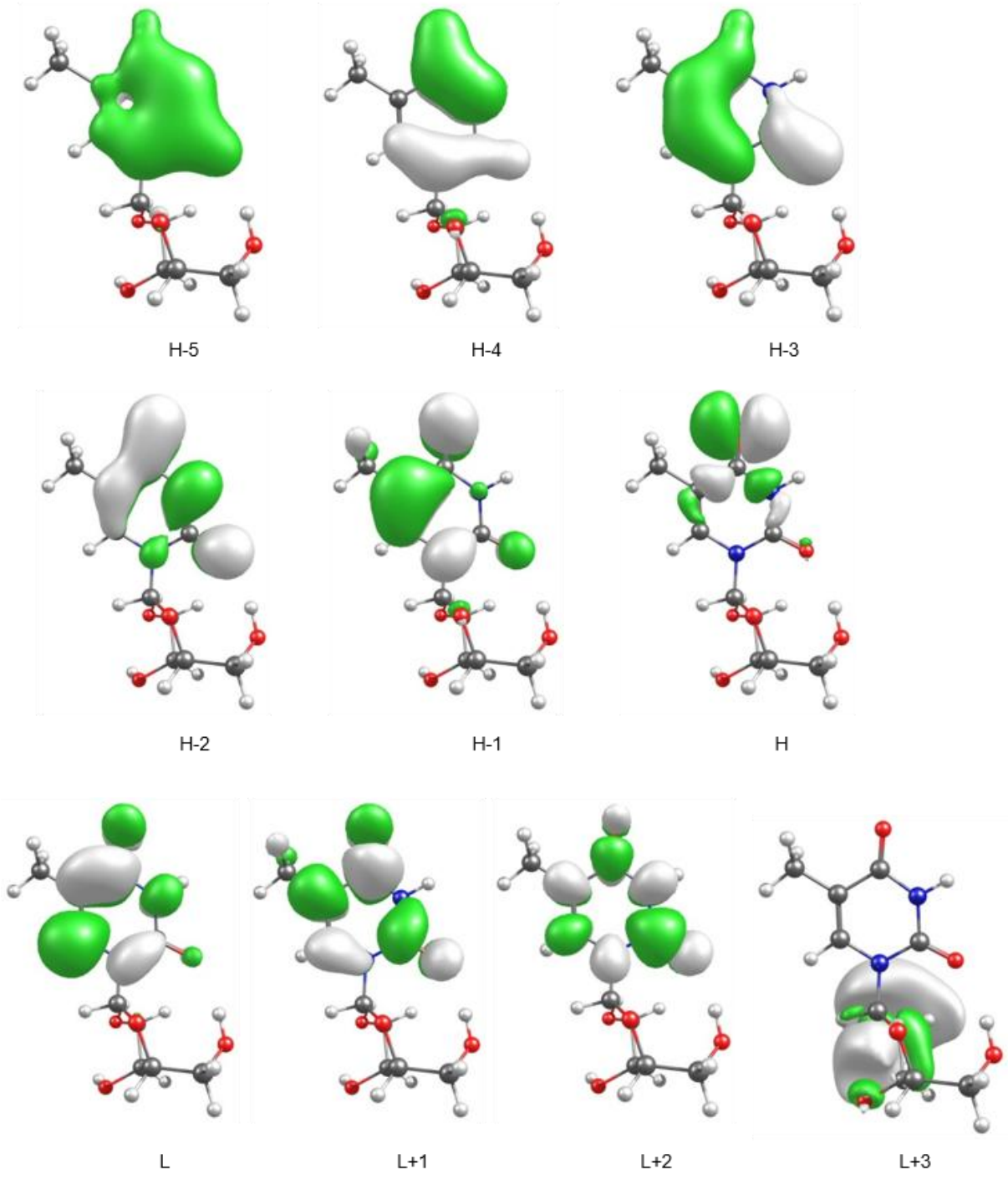


$S_1/1^1\pi\pi^*$: H–1→L

$S_2/1^1n\pi^*$: H→L+1

$S_3/2^1n\pi^*$: H→L

Figure S.29: thymidine

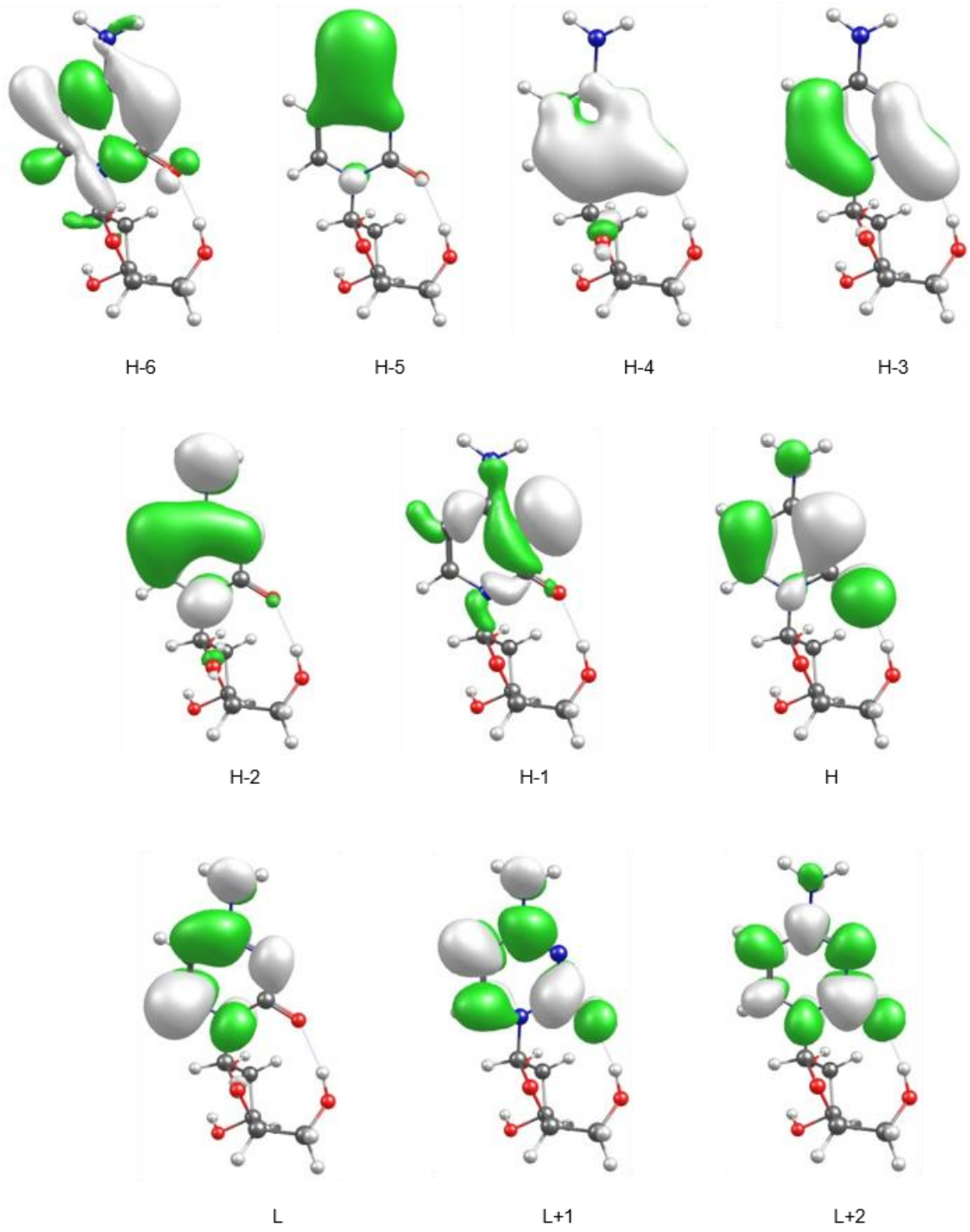


$S_1/1^1n\pi^*$: H→L

$S_2/1^1\pi\pi^*$: H–1→L

$S_3/2^1\pi\pi^*$: H–2→L+1

Figure S.30: cytidine

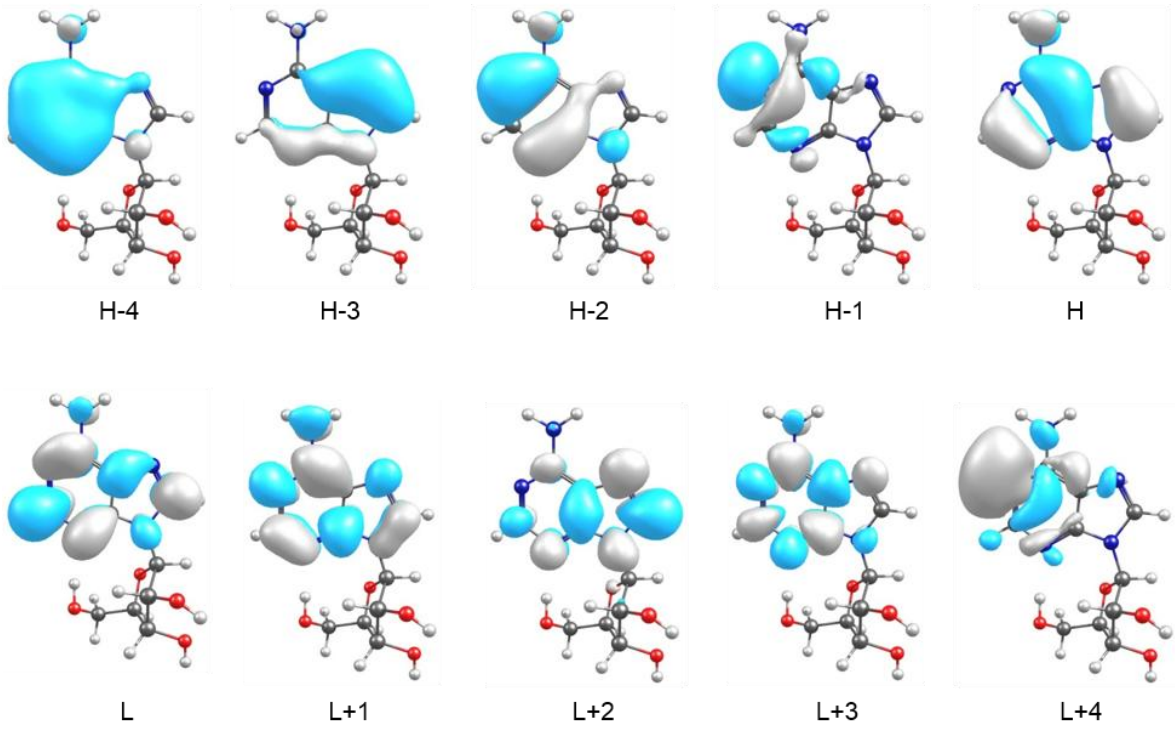


$S_1/1^1\pi\pi^*$: H→L

$S_2/1^1n\pi^*$: H–1→L

$S_3/2^1\pi\pi^*$: H–2→L

Figure S.31: adenosine

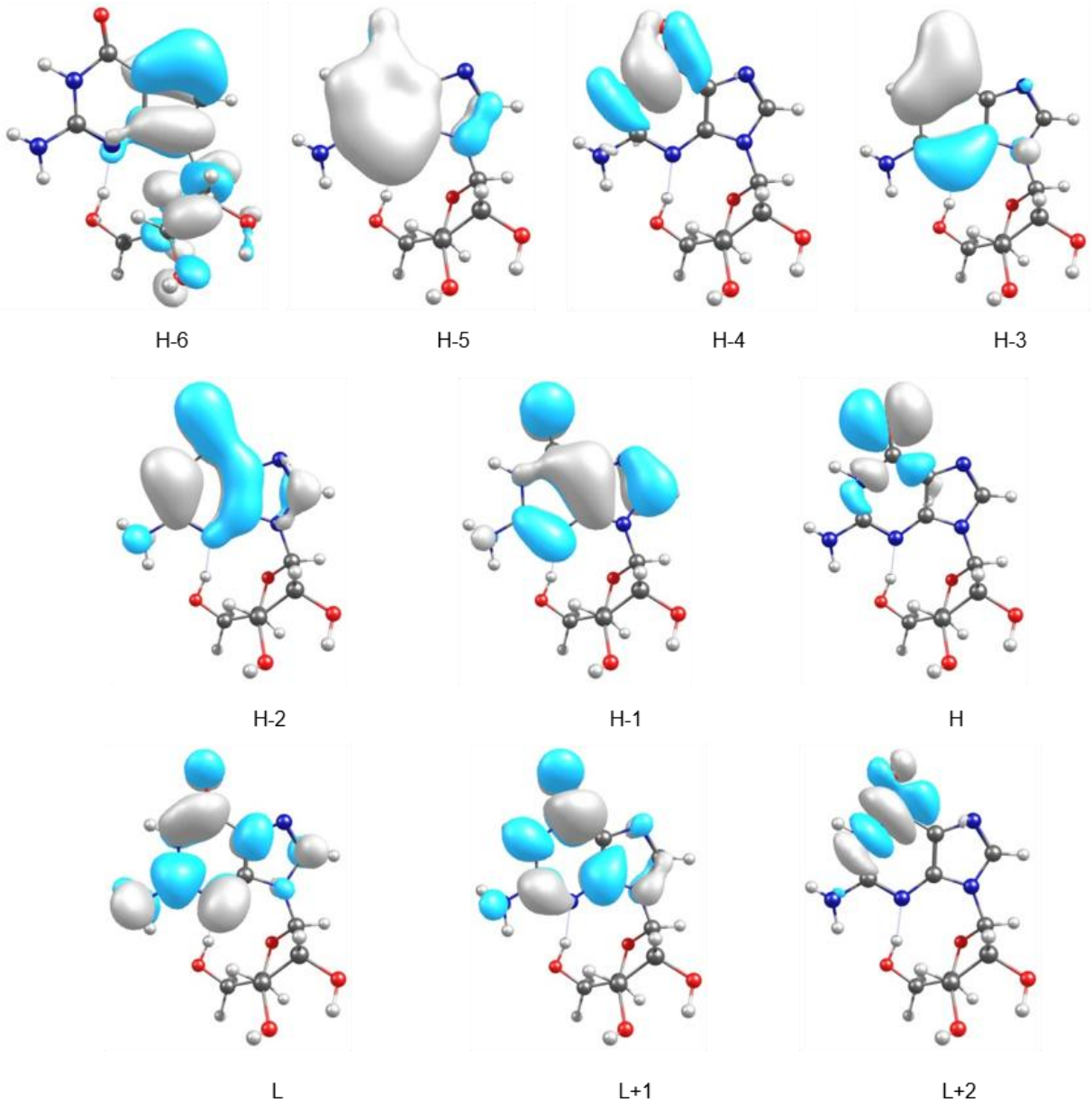


$S_1/1^1n\pi^*$: H–1→L

$S_2/1^1\pi\pi^*$: H–3→L and H→L+1

$S_3/2^1\pi\pi^*$: H→L

Figure S.32: guanosine

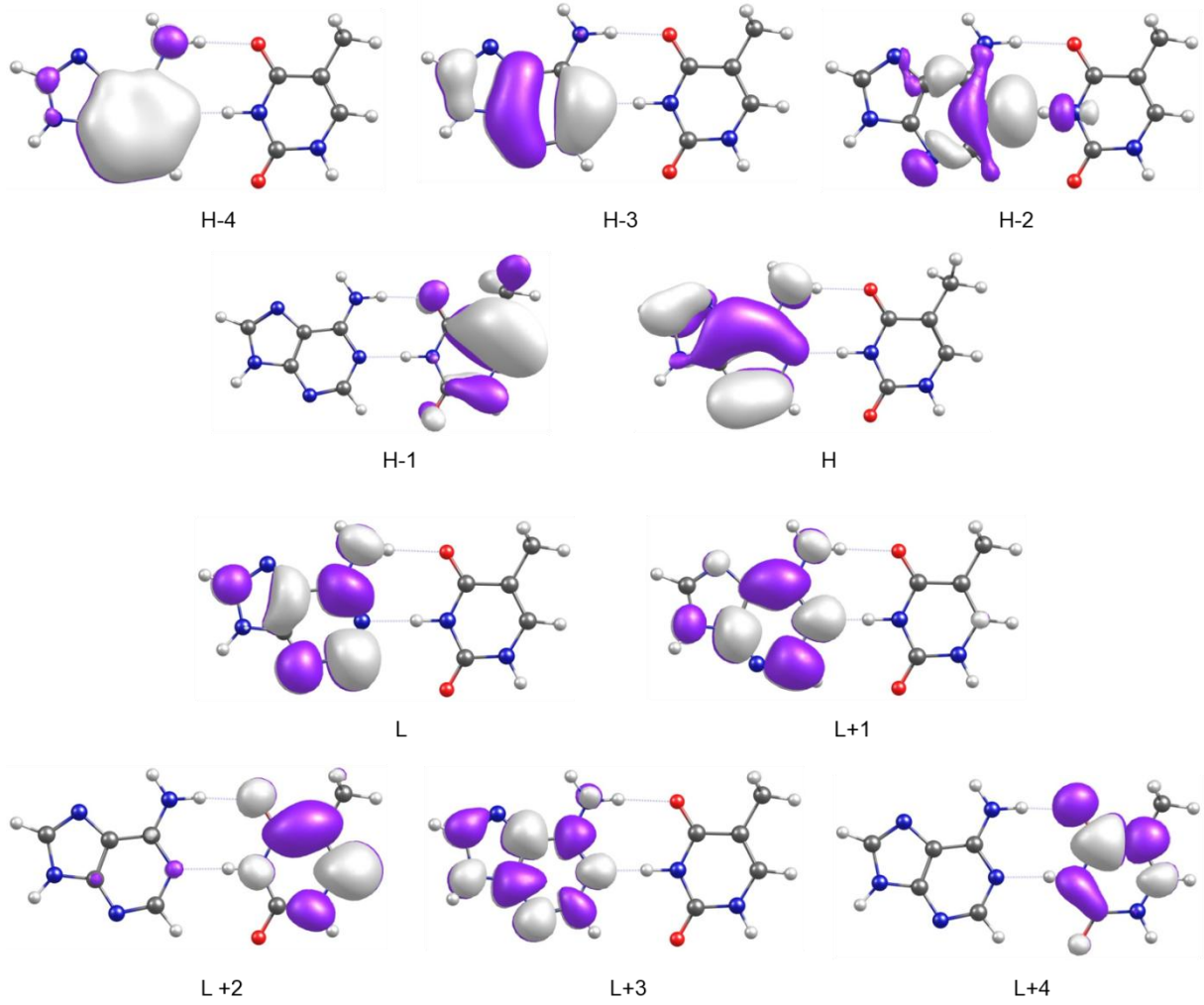


$S_1/1^1\pi\pi^*$: H–1→L

$S_2/1^1n\pi^*$: H→L+1

$S_3/2^1\pi\pi^*$: H–1→L+1

Figure S.33: adenine-thymine base pair



$S_1/1^1\pi\pi^*$: H→L+1 and H–3→L

$S_2/2^1\pi\pi^*$: H→L

$S_3/3^1\pi\pi^*$: H–1→L+2

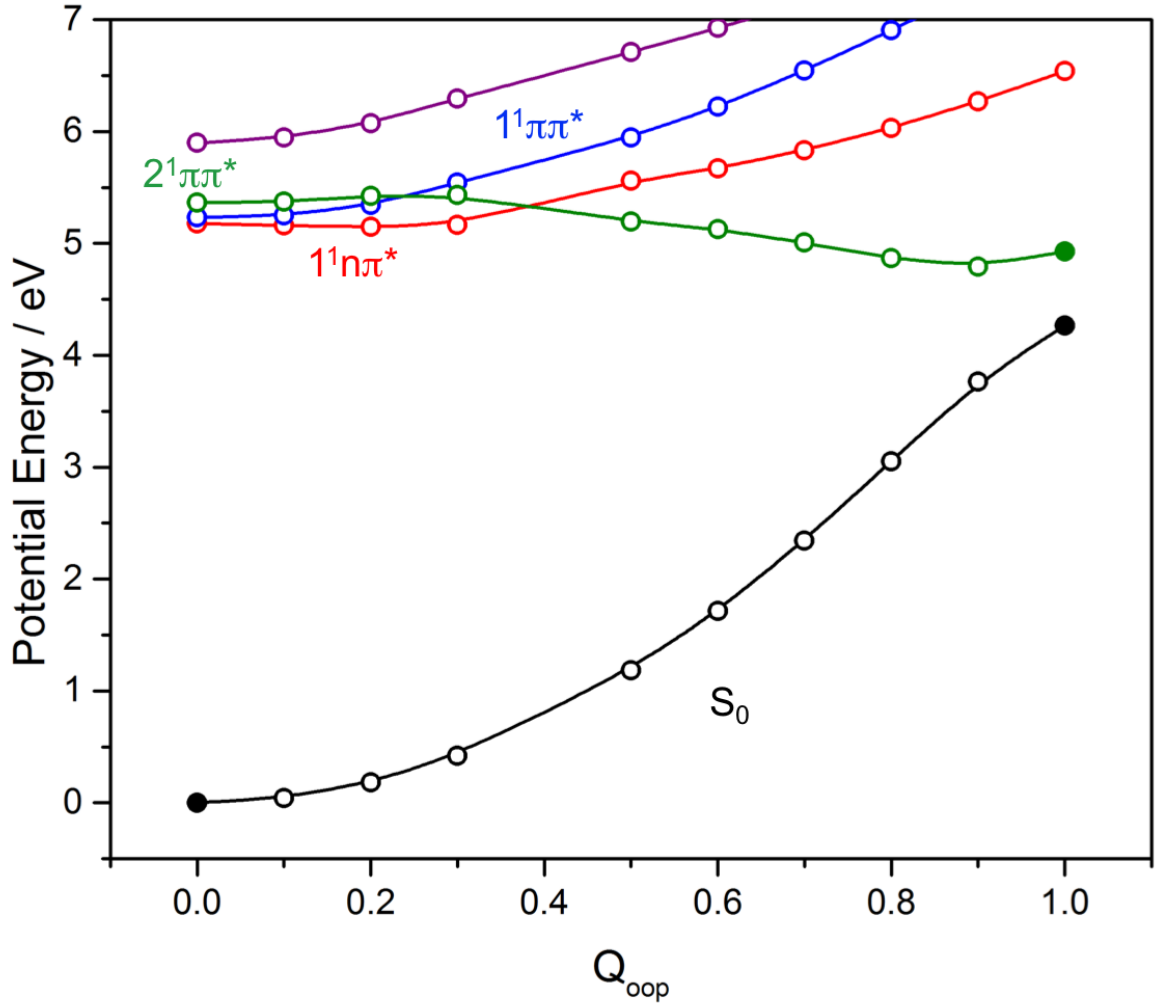
$S_4/1^1n\pi^*$: H–2→L

$S_5/2^1n\pi^*$: H–2→L+1

$S_6/4^1\pi\pi^*$: H–3→L and H→L+1

$S_7/5^1\pi\pi^*$: (H–1→L)+(H→L+2) *(double excitation)*

Figure S.34: PECs for the ground and first few singlet excited states of adenosine along Q_{oop} calculated at the ADC(2)/cc-pVDZ level of theory.



References

1. M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennuchi, G. Petersson, *et al*., Gaussian 09 **r**evision B.01; Gaussian Inc.: Wallingford, CT (2010).

2. T.H. Dunning, Jr., *J. Chem. Phys.* 1989, **90**, 1007-1023.

3. W.J. Hehre, R.F. Stewart and J.A. Pople, *J. Chem. Phys*. 1969, **51**, 2657-2664.

4. Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from www.turbomole.com