Supplementary Information for

The Mechanism Underlying the Nucleobase-Distinguishing Ability of Benzopyridopyrimidine (BPP)

Michał A. Kochman,*† Andrzej Bil,# R. J. Dwayne Miller†‡

† Max Planck Institute for the Structure and Dynamics of Matter, Bldg. 99 (CFEL), Luruper Chaussee 149, 22761 Hamburg, Germany.
# Faculty of Chemistry, University of Wrocław, F. Joliot-Curie 14, 50-383 Wrocław, Poland
‡ Department of Chemistry and Physics, University of Toronto, 80 St. George Street, Toronto, Ontario M5S 3H6, Canada.

e-mail: michal.kochman@mpsd.mpg.de

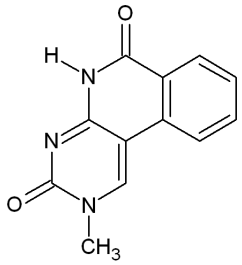
1 Tautomeric preference of BPP

The present section reports an analysis of the tautomerism of BPP in aqueous solution. In experimental studies, BPP has been used in deoxyribonucleoside form, in which its N1 atom is substituted by the C1' atom of the deoxyribose moiety. The deoxyribose substituent reduces the number of available tautomers as compared to the free base, and it may also interact with the π-bonding system of BPP through hyperconjugation. In order to take these effects into account without having to include the large deoxiribose moiety in simulations, the analysis was performed on the molecule 1-methylbenzopyridopyrimidine (1MeBPP) as a reduced model of the deoxyribonucleoside.

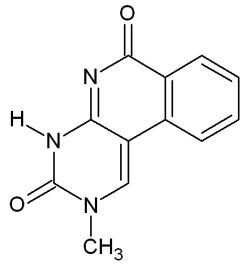
As can be seen in Figure S1, 1MeBPP possesses a total of four tautomers, which can be identified by the atom to which the single labile hydrogen is bonded. The N8-H tautomer is the counterpart of the canonical amino-keto tautomer of cytosine; the other three tautomers are the N3-H, the O7-H, and the O12-H. The equilibrium geometries of the four tautomers were optimized and confirmed to correspond to potential energy minima through the calculation of the normal modes of the molecule. The calculations were performed at the density functional theory (DFT) level of theory within Gaussian 09, Revision D.01.[1] The B3LYP exchange-correlation functional[2,3] was applied in combination with the high-quality def2-TZVPD basis set.[4] The effects of aqueous solvation were included with the use of the continuous surface charge implementation of the polarizable continuum model (CSC-PCM).[5,6]

Table S1 lists the relative energies of the tautomers, and estimated populations at room temperature. The DFT calculations predict that the N8-H tautomer is by far the dominant form of 1MeBPP in aqueous solution, making up around 99.98% of the total population. This result is line with the findings of Okamoto et al., who established that BPP exists as the N8-H tautomer when it is incorporated into double-stranded DNA, regardless of whether it is paired with adenine or guanine.[7] Because of the predominance of the N8-H tautomer under experimentally-relevant conditions, in our

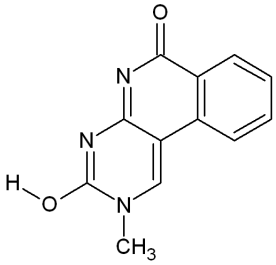
Figure S1: Tautomers of 1MeBPP.



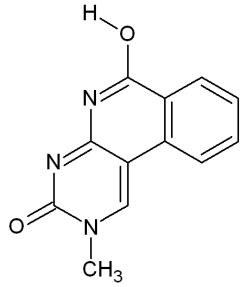
(a) N8-H



(b) N3-H



(c) O7-H



(d) O12-H

simulation study of the photochemistry of BPP we took into account only that tautomer.

Table S1: Energies (E), Gibbs free energies (G) of tautomers of 1MeBPP in aqueous solution at a temperature of 298 K and pressure of 1 atm., and their estimated populations. The energy and Gibbs free energy values are quoted relative to the lowest-energy syn N8-H tautomer. The energy values include zero-point vibrational corrections.

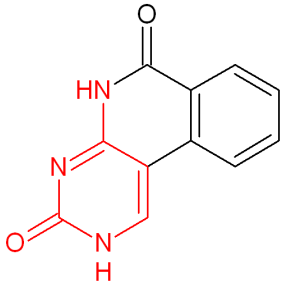
Tautomer	E, kJ/mol	G, kJ/mol	Population
N8-H	0	0	0.9998
N3-H	21.8	21.3	0.0002
O7-H	83.5	82.9	3×10^{-15}
O12-H	54.3	54.7	2×10^{-10}

2 Details of computational methods

2.1 Choice of model chemistry

The foremost question is the choice of model chemistry employed for the excited electronic states of BPP and its base pairs. Calculated excitation energies and potential energy surfaces (PESs) of nucleobases and their analogues can be very sensitive to the level of electronic structure theory; an accurate description of the excited states of the canonical (amino-keto) tautomer of cytosine, in particular, requires high-level electronic structure methods.[8,9] Because BPP incorporates the cytosine chromophore (see Figure S2), only methods which perform well for cytosine are expected to be reliable for BPP. We restrict ourselves here to considering only methods based on coupled cluster theory, as they can be thought of as forming a well-defined hierarchy of approximations according to the level of excitation at which the cluster operator is truncated.

Figure S2: The amino-keto cytosine chromophore – highlighted in red – within the molecule BPP.



The performance of several approximations within the coupled cluster hierarchy for the canonical nucleobases of DNA been benchmarked by Szalay et al.[8] According to this study, methods with up to triple excitations, such as equation of motion coupled cluster including perturbative triple excitations (EOM-CCSD(T)), are capable of achieving an accuracy of around 0.1 eV for vertical excitation energies into all low-lying excited states of cytosine.[8] However, these methods are computationally too demanding for a molecule as large as BPP, and a lower-level method had to be chosen.

Moving down the hierarchy of approximations, EOM-CCSD overestimates vertical excitation energies into states of all types by some 0.1 to 0.3 eV relative to the EOM-CCSD(T) benchmark, but without an apparent bias towards excited states of a specific character.[8] Hence, EOM-CCSD does give a good account of the relative ordering in energy of the low-lying excited states of cytosine. This degree of accuracy, with a moderately large but consistent overestimation of the energies of transitions dominated by single excitations, is believed to be typical of the EOM-CCSD method.[8,10–12]

In turn, the second-order approximate coupled cluster method (CC2), with linear response calculation of excitation energies,[13] performs somewhat erratically. Whereas the energies of transitions into the lowest two $^1\pi\pi^*$-type states coincide closely with the benchmark provided by the EOM-CCSD(T) method, the energies of the lowest two $^1\mathrm{n}\pi^*$-type transitions are severely underestimated (by as much as 0.56 eV in the case of the lowest $^1\mathrm{n_O}\pi^*$ transition).[8] The EOM-CCSD method, with its fairly consistent overestimation of excitation energies, seems more likely to be able to correctly predict the nature and ordering of the excited states of BPP. For this reason, EOM-CCSD is our method of choice for the treatment of the excited electronic states of the isolated BPP molecule.

As mentioned in the main body of the present paper, the large size of the base pairs of BPP

compelled us to resort to the MP2/ADC(2) combination[14] when calculating their excited electronic states. The MP2/ADC(2) method is closely related to the CC2 method,[15] which, as we have seen, gives rise to large errors for the low-lying $^1n\pi^*$-type states of unmodified cytosine. In order to assess the performance of low-level electronic structure methods such as CC2 and MP2/ADC(2) with regard to the excited states of BPP, we compared the structures of the excited electronic states calculated using CC2 with the equation of motion formulation[16] (EOM-CC2) against the benchmarks provided by EOM-CCSD. (Excitation energies calculated with the EOM formalism are identical to those obtained with linear response theory.) Due to the close relationship between the CC2 and MP2/ADC(2) methods, the conclusions from this comparison will apply also to MP2/ADC(2).

The vertical excitation spectra of BPP were calculated using the EOM-CC2 and EOM-CCSD methods as implemented in the program PSI4, version 1.0.0.[17,18] The cc-pVDZ basis set was used in both the EOM-CC2 and EOM-CCSD calculations, and the frozen core approximation was applied. The three lowest excited states each of symmetry species A′ and A″ were calculated.

Focusing first on the electronic ground state, the following procedure was used in order to quantify the degree of resemblance between the ground-state wavefunctions predicted by CC2 and CCSD. Recall that within the coupled cluster Ansatz, the correlated ground-state wavefunction Ψ_0 is generated by the wave operator $\exp(\hat{T})$ acting on the reference determinant Φ_0:

$$\Psi_0 = \exp(\hat{T})\,\Phi_0 \tag{1}$$

In both the CCSD and CC2 methods, the cluster operator $\hat{T}$ is truncated after the term with double excitations, taking the form:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 \tag{2}$$

where

$$\hat{T}_1 = \sum_{i,a} t_i^a\, \hat{a}^\dagger \hat{i} \qquad \text{and} \qquad \hat{T}_2 = \frac{1}{4} \sum_{i,j,a,b} t_{ij}^{ab}\, \hat{a}^\dagger \hat{b}^\dagger \hat{i}\, \hat{j} \tag{3}$$

In the full CCSD method, the single (t_i^a) and double (t_{ij}^{ab}) cluster amplitudes are obtained from the CCSD amplitude equations, which can be written out in the following form:

$$\langle \Phi_i^a \,|\, \tilde{\hat{H}} + [\tilde{\hat{H}}, \hat{T}_2] \,|\, \Phi_0 \rangle = 0 \tag{4}$$

$$\langle \Phi_{ij}^{ab} \,|\, \tilde{\hat{H}} + [\tilde{\hat{H}}, \hat{T}_2] + \tfrac{1}{2}\,[[\tilde{\hat{H}}, \hat{T}_2], \hat{T}_2] \,|\, \Phi_0 \rangle = 0 \tag{5}$$

Here, $\tilde{\hat{H}} = \exp(-\hat{T}_1)\,\hat{H}\exp(\hat{T}_1)$ is the $\hat{T}_1$-transformed Hamiltonian. $\Phi_i^a = \hat{a}^\dagger \hat{i}\,\Phi_0$ denotes the singly excited determinant obtained from the reference determinant by the action of the excitation operator with the annihilator $\hat{i}$ and creator $\hat{a}^\dagger$. Likewise, $\Phi_{ij}^{ab} = \hat{a}^\dagger \hat{b}^\dagger \hat{i}\, \hat{j}\,\Phi_0$. The CCSD energy is calculated as:

$$E = \langle \Phi_0 \,|\, \hat{H} \exp(\hat{T}) \,|\, \Phi_0 \rangle \tag{6}$$

In the CC2 method, the singles equations (4) are retained unmodified, but the doubles equations (5) are approximated to be correct only through first order in the fluctuation potential.[13] The doubles equations thus become:

$$\langle \Phi_{ij}^{ab} \,|\, \tilde{\hat{H}} + [\hat{F}, \hat{T}_2] \,|\, \Phi_0 \rangle = 0 \tag{7}$$

where $\hat{F}$ is the Fock operator.

The quality of the approximation represented by equation (7) can be assessed by comparing the resulting cluster amplitudes with those obtained with the full CCSD method. For example, in the previous studies of Pabst and coworkers[19], and Wang and Wang,[20] failures of the CC2 methods were diagnosed by calculating the magnitudes of the amplitude vectors $\boldsymbol{t}_i^a$ and $\boldsymbol{t}_{ij}^{ab}$, whose elements are, respectively, the single (t_i^a) and double (t_{ij}^{ab}) cluster amplitudes. In the present study, we decided to quantify the agreement between CC2 and CCSD by introducing a singles diagnostic $S^{(1)}$ which was taken as the absolute value of the cosine similarity between the t_i^a vectors obtained with CC2 and with CCSD, and a doubles diagnostic $S^{(2)}$, which was defined analogously for the $\boldsymbol{t}_{ij}^{ab}$ vectors:

$$S^{(1)} = \frac{|\,\boldsymbol{t}_i^{a\,(\mathrm{CC2})} \cdot \boldsymbol{t}_i^{a\,(\mathrm{CCSD})}\,|}{||\,\boldsymbol{t}_i^{a\,(\mathrm{CC2})}\,||\;||\,\boldsymbol{t}_i^{a\,(\mathrm{CCSD})}\,||} = \frac{|\,\sum_{i,a}\, t_i^{a\,(\mathrm{CC2})}\, t_i^{a\,(\mathrm{CCSD})}\,|}{\sqrt{\sum_{i,a}\,\left[t_i^{a\,(\mathrm{CC2})}\right]^2}\,\sqrt{\sum_{i,a}\,\left[t_i^{a\,(\mathrm{CCSD})}\right]^2}} \tag{8}$$

$$S^{(2)} = \frac{|\,\boldsymbol{t}_{ij}^{ab\,(\mathrm{CC2})} \cdot \boldsymbol{t}_{ij}^{ab\,(\mathrm{CCSD})}\,|}{||\,\boldsymbol{t}_{ij}^{ab\,(\mathrm{CC2})}\,||\;||\,\boldsymbol{t}_{ij}^{ab\,(\mathrm{CCSD})}\,||} = \frac{|\,\sum_{i,j,a,b}\, t_{ij}^{ab\,(\mathrm{CC2})}\, t_{ij}^{ab\,(\mathrm{CCSD})}\,|}{\sqrt{\sum_{i,j,a,b}\,\left[t_{ij}^{ab\,(\mathrm{CC2})}\right]^2}\,\sqrt{\sum_{i,j,a,b}\,\left[t_{ij}^{ab\,(\mathrm{CCSD})}\right]^2}} \tag{9}$$

$S^{(1)}$ and $S^{(2)}$ each take values ranging from 0, indicating no resemblance between the CCSD and CC2 amplitude vectors, to 1, meaning that the two methods return identical amplitude vectors. Intermediate values of $S^{(1)}$ and $S^{(2)}$ indicate partial agreement between the two methods. In practice, for the nucleobases and nucleobase analogues to which we applied the $S^{(1)}$ and $S^{(2)}$ diagnostics, in all cases we found $S^{(1)} \approx 0.9$ and $S^{(2)} \approx 0.9$, which suggests the CC2 approximation was consistently adequate for the electronic ground states of those molecules.

Within the framework of the equation of motion (EOM) formalism, the excitation energy E_k into the k-th excited state can be obtained from the solution of the so-called right-hand eigenvalue problem:

$$\overline{H}\,\hat{R}(k)\,|\,\Phi_0\,\rangle = E_k\,\hat{R}(k)\,|\,\Phi_0\,\rangle \tag{10}$$

Here, $\hat{R}(k)$ is the excitation operator for the k-th excited state. In the EOM-CCSD and EOM-CC2 methods, the operator $\hat{R}(k)$ is truncated after the term with double excitations, taking the form:

$$\hat{R}(k) = \hat{R}_0(k) + \hat{R}_1(k) + \hat{R}_2(k) \tag{11}$$

$\hat{R}_0(k) = r_0(k)$ is a scalar number, whose value is specific to the k-th state, and whose purpose is to change the coefficient in front of the reference determinant Φ_0 to that appropriate to the wavefunction of the k-th state. $\hat{R}_1(k)$ and $\hat{R}_2(k)$ are, respectively, the single and double excitations operators:

$$\hat{R}_1(k) = \sum_{i,a}\, r_i^a(k)\,\hat{a}^\dagger\,\hat{i} \tag{12}$$

$$\hat{R}_2(k) = \sum_{i<j\,,\,a<b}\, r_{ij}^{ab}(k)\,\hat{a}^\dagger\,\hat{b}^\dagger\,\hat{i}\,\hat{j} \tag{13}$$

Lastly, $\overline{H} = e^{-\hat{T}}\,\hat{H}\,e^{-\hat{T}}$ is the $\hat{T}$-transformed Hamiltonian.

For transitions dominated by single excitations from Φ_0, the diabatic character of the excited state can be said to be described by the coefficients $r_i^a(k)$ which appear in the single excitation operator $\hat{R}(k)$. Accordingly, for each excited state k we construct a vector $\mathbf{r}_{\mathrm{S}}(k)$ whose elements are the coefficients $r_i^a(k)$. The resemblance (S_{mn}) between an excited state m calculated using the EOM-CC2 method, and an excited state n calculated using EOM-CCSD, can then be quantified

as the absolute value of the cosine similarity between the vectors $\mathbf{r}_S(m)$ and $\mathbf{v}_S(n)$ obtained in the respective calculations:

$$S_{mn} = \frac{|\,\mathbf{r}_S(m) \cdot \mathbf{r}_S(n)\,|}{||\,\mathbf{r}_S(m)\,||\;||\,\mathbf{r}_S(n)\,||} = \frac{|\,\sum_{i,a}\, r_i^a(m)\, r_i^a(n)\,|}{\sqrt{\sum_{i,a}\,[r_i^a(m)]^2}\,\sqrt{\sum_{i,a}\,[r_i^a(n)]^2}} \tag{14}$$

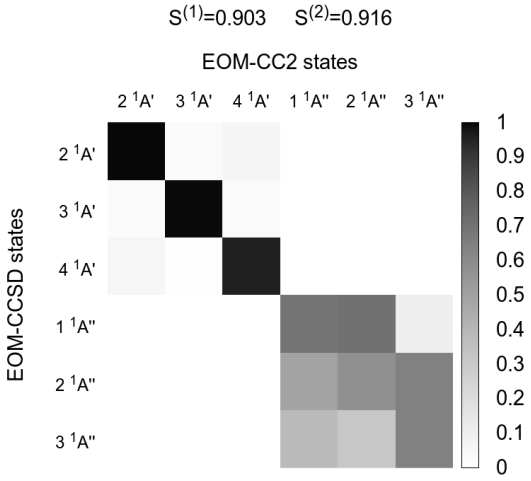
Thus, the diagnostic S_{mn} provides a state-by-state comparison between the diabatic character of excited states predicted the EOM-CCSD and EOM-CC2 methods. Intermediate values of S_{mn} (which is to say, significantly higher than 0, but also appreciably lower than 1) turn out to be common, meaning that often, there is not a one-to-one correspondence between a state obtained with EOM-CC2, and any single state calculated using the full EOM-CCSD method.

A case in point is the BPP molecule. The vertical excitation energies calculated with EOM-CCSD and with EOM-CC2 are listed in Table S2, while Figure S3 provides a visual representation of values of cosine similarity between single excitations vectors of the EOM-CCSD and EOM-CC2 states. Both the EOM-CCSD and EOM-CC2 calculations detect three $\pi\pi^*$-type states of A′ symmetry, and three $n\pi^*$-type states of A″ symmetry. As is generally the case for the canonical nucleobases,[8] the vertical excitation energies of the $\pi\pi^*$-type states calculated with the EOM-CCSD method are higher by a few tenths of an electronvolt than those calculated with EOM-CC2. This is especially true of the $n\pi^*$-type states. (NB due to the small size of the cc-pVDZ basis set, the excitation energies cannot be expected to be close to the complete basis set limit, but a qualitative comparison of the excitation energies calculated with the two methods is nevertheless possible.) For the singlet ground state, we obtain $S^{(1)} = 0.903$ and $S^{(2)} = 0.916$, which suggests that in this regard, CC2 is in excellent agreement with full CCSD.

Table S2: Vertical excitation energies of BPP, calculated using the EOM-CCSD and EOM-CC2 methods.

Level of theory	State	ΔE, eV
EOM-CCSD/cc-pVDZ	$2\ ^1\mathrm{A}'$ ($\pi\pi^*$)	4.248
	$3\ ^1\mathrm{A}'$ ($\pi\pi^*$)	4.601
	$4\ ^1\mathrm{A}'$ ($\pi\pi^*$)	5.308
	$1\ ^1\mathrm{A}''$ ($n\pi^*$)	4.911
	$2\ ^1\mathrm{A}''$ ($n\pi^*$)	5.178
	$3\ ^1\mathrm{A}''$ ($n\pi^*$)	5.808
EOM-CC2/cc-pVDZ	$2\ ^1\mathrm{A}'$ ($\pi\pi^*$)	4.052
	$3\ ^1\mathrm{A}'$ ($\pi\pi^*$)	4.508
	$4\ ^1\mathrm{A}'$ ($\pi\pi^*$)	4.982
	$1\ ^1\mathrm{A}''$ ($n\pi^*$)	4.463
	$2\ ^1\mathrm{A}''$ ($n\pi^*$)	4.752
	$3\ ^1\mathrm{A}''$ ($n\pi^*$)	5.089

Figure S3: Visual representation of cosine similarity between single excitations vectors of low-lying singlet excited states of BPP calculated at the EOM-CCSD/cc-pVDZ level of theory, and at the EOM-CC2/cc-pVDZ level. Values of the S_{mn} diagnostic are plotted in matrix form; shading represents the value of S_{mn} for a given pair of excited states obtained with EOM-CC2 and EOM-CCSD.



The situation becomes more interesting for the excited states. The mutual similarity of the $\pi\pi^*$-type states calculated with the EOM-CCSD and EOM-CC2 methods is described by the upper left 3×3 subblock of the S_{mn} matrix. The diagonal elements of this subblock take values close to unity, while all off-diagonal elements are close to zero. This indicates that there is a clear-cut one-to-one correspondence between the $\pi\pi^*$-type states calculated with the two methods, and that both methods predict the same energy ordering of the corresponding $\pi\pi^*$-type states.

In turn, the lower right 3×3 subblock of the S_{mn} matrix characterizes the mutual similarity of nπ^*-type states calculated with EOM-CCSD and EOM-CC2. It is immediately evident that for the nπ^*-type states, the EOM-CC2 method deviates strongly from the benchmark set by EOM-CCSD. Within each column of this subblock, there are multiple appreciably large values of the cosine similarity, meaning that we do not find a one-to-one correspondence between the nπ^*-type states calculated with the EOM-CCSD and EOM-CC2 methods. For example, both the 1 ^{1}A$''$ and 2 ^{1}A$''$ EOM-CC2 states are mainly similar to the 1 ^{1}A$''$ EOM-CCSD state, while also bearing some resemblance to the 2 ^{1}A$''$ and 3 ^{1}A$''$ EOM-CCSD states. Clearly, then, EOM-CC2 is unreliable for three lowest nπ^*-type states of BPP, to the extent that it does not predict the same diabatic character for any of these states as the EOM-CCSD method. At least some of the nπ^*-type states predicted by the EOM-CC2 method seem to be too low in energy. For instance, according to EOM-CC2, the 1 ^{1}A$''$ (nπ^*) state is located quite close in energy to the emissive 2 ^{1}A$'$ ($\pi\pi^*$) state, while the EOM-CCSD method predicts a significantly larger separation between these states.

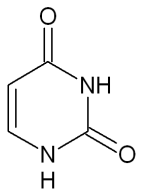
Finally, values of the cosine similarity between EOM-CCSD and EOM-CC2 states of different symmetries are found in the upper right and lower left 3×3 subblocks of the S_{mn} matrix. As expected, all are vanishingly small.

In summary, although for the electronic structures of the low-lying $\pi\pi^*$-type states of BPP, EOM-CC2 is in good agreement with full EOM-CCSD, it fails badly for the nπ^*-type states. It

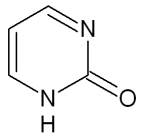
would be outside the scope of the present study to attempt to pinpoint this mode of failure to a specific feature of the approximation represented by EOM-CC2. However, we note here that it is certainly not general to all molecules which possess multiple nπ^*-type states. In order to demonstrate this point, we carried out analogous analyses of the performance of EOM-CC2 versus EOM-CCSD for a set of four molecules relevant for the photochemistry of DNA: uracil, the fluorescent cytosine analogue zebularine, and the damaged nucleobases (etheno adducts) 3,N^4-ethenocytosine and 1,N^6-ethenoadenine (see Figure S4 for molecular structure). All are heteroaromatic molecules, each with two or more lone electron pairs lying in the plane of the aromatic ring.

The results are shown in Figure S5. In qualitative terms, for the $\pi\pi^*$-type (A′-symmetric) states of the four molecules, the EOM-CC2 method consistently achieves good agreement with the EOM-CCSD benchmark. For the nπ^*-type (A″-symmetric) states, in turn, on the whole the performance of EOM-CC2 is slightly worse than for the $\pi\pi^*$-type states, but still reasonable. For each molecule, the largest element of every row and every column of the S_{mn} matrix is the diagonal element, meaning that a one-one-one correspondence is found between the electronic excited states calculated with EOM-CC2 and those obtained with EOM-CCSD. Thus, the general conclusion is that the performance of EOM-CC2 for nπ^* states is generally erratic, but only occassionally so bad as to lead to a qualitatively incorrect description of their electronic structures, as happens in the case of BPP. In systems where such states are relevant, the application of this method should ideally be preceded by testing against higher-level methods such as full EOM-CCSD. This recommendation can be extended to related methods such as MP2/ADC(2).

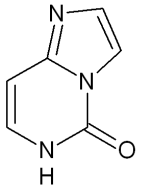
Figure S4: A four-molecule test set for the EOM-CC2 and EOM-CCSD methods.



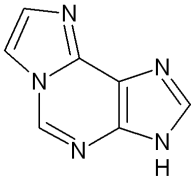
(a) uracil (U)



(b) zebularine

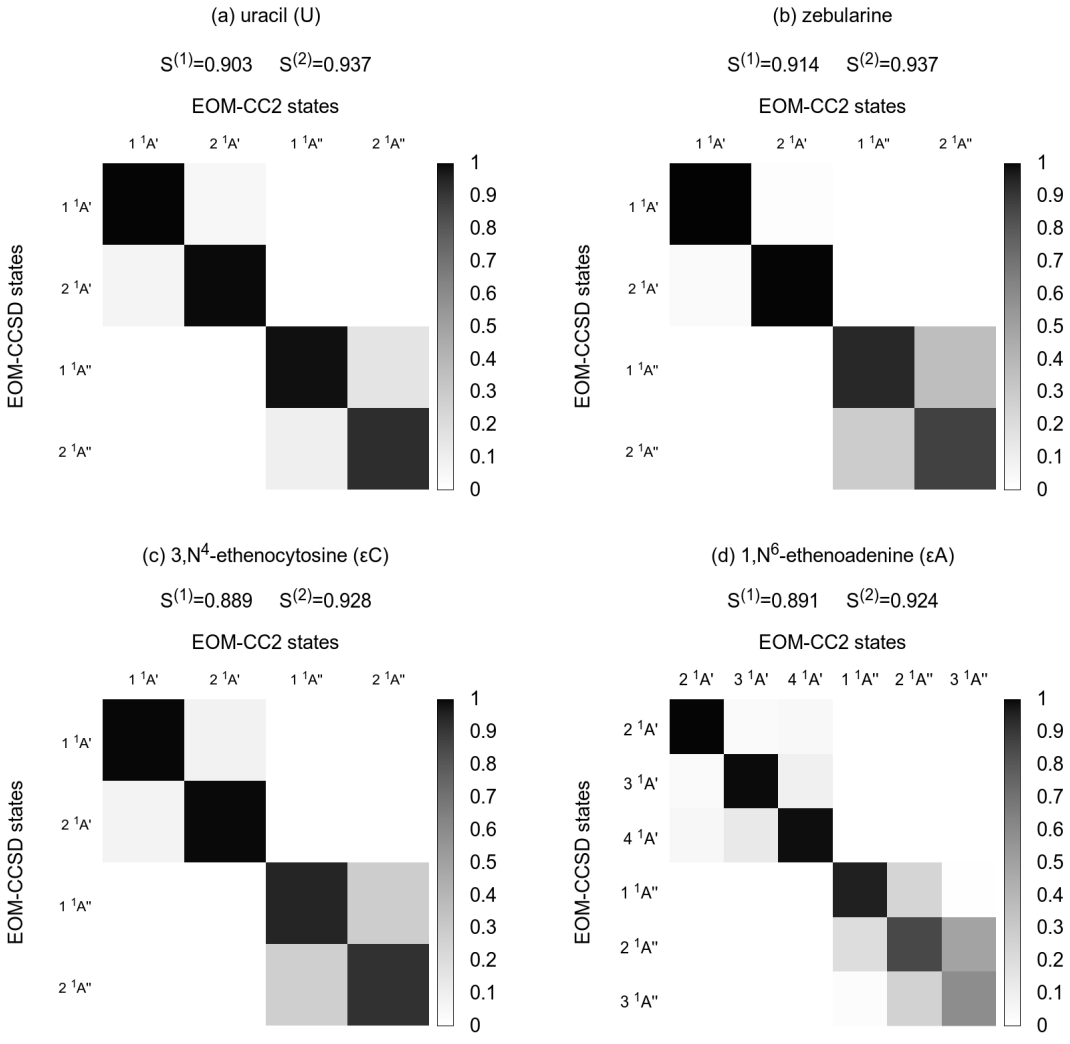


(c) 3,N^4-ethenocytosine (εC)



(d) 1,N^6-ethenoadenine (εA)

Figure S5: Visual representation of cosine similarity values for the four-molecule test set. Shading represents the value of the cosine similarity for a given pair of excited states. In each of the four molecules, the lowest two or three A′ states which were taken into consideration are of the $\pi\pi^*$ type, while the A″ states are $n\pi^*$ in character.



2.2 Excited Electronic States of BPP

The present section expands on Section 2.1 of the main body of our paper.

Except when noted otherwise, all calculations used the jun-cc-pVDZ basis set, which is one of the "calendar basis sets" introduced by Papajak et al.[21] These basis sets are derived through the successive elimination of diffuse basis functions from the standard aug-cc-pVnZ basis sets, starting with the diffuse subshells with the highest angular momentum. In particular, the jun-cc-pVDZ basis set is obtained from aug-cc-pVDZ by deleting the diffuse d subshell from the second-row elements, as well as all diffuse functions from hydrogen. Thus, only the s and p diffuse functions on the second-row elements are retained. Papajak et al.[21] have shown that this augmentation scheme performs very closely to the original aug-cc-pVDZ basis set for hydrogen bond lengths and energies.

The ground-state equilibrium geometry of BPP was optimized with the use of the Møller-Plesset perturbation method (MP2) in combination with the jun-cc-pVDZ basis set, implemented in the computational chemistry software package Gaussian 09, Revision D.01.[1] A restricted Hartree-Fock (RHF) reference determinant was used, and the frozen core approximation was applied. Initially, no symmetry constraints were imposed. The resulting equilibrium geometry was confirmed to correspond to a potential energy minimum by calculating analytically the harmonic vibrational frequencies. The equilibrium geometry of BPP was found to be near-planar, with only a very slight pyramidalization of nitrogen N8. Taking advantage of the near-planarity of the molecule, the ground-state equilibrium geometry was re-optimized while imposing planar symmetry (point group C_s) as a cost-reducing approximation for the calculation of the electronically excited states, and the electron-attached states. The resulting planar equilibrium geometry possesses a single imaginary vibrational frequency of $85.9\mathrm{i}\ \mathrm{cm}^{-1}$, and it is therefore a first-order saddle point on the ground-state PES. However, it is only marginally (by 2×10^{-4} eV) higher in energy than the near-planar minimum, and so the imposition of planar symmetry seems justified.

The vertical excitation spectrum of BPP was calculated at the EOM-CCSD/jun-cc-pVDZ level of theory at the planar equilibrium geometry optimized at the MP2/jun-cc-pVDZ level. The EOM-CCSD calculations were performed within Molpro version 2012.1.[22] An RHF reference determinant was used, and the frozen approximation was applied. Three excited states each of symmetry species A′ and A″ and were calculated.

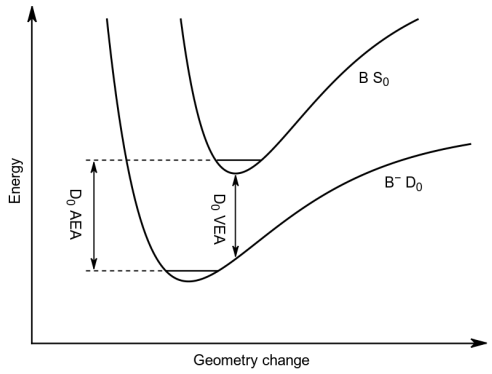
The emissive structure of BPP was characterized by means of geometry optimization on the lowest singlet adiabatic excited state (S_1) as calculated using the EOM-CCSD method. The optimization was performed within the program PSI4, version 1.0.0.[17,18] In order to keep the geometry optimization tractable, the basis set size was reduced to cc-pVDZ. Furthermore, due to program limitatations, all electrons were correlated during the optimization. No symmetry constraints were imposed. Subsequently, the vertical emission energy of BPP at the resulting excited-state equilibrium geometry was computed through a single-point, frozen-core calculation at the EOM-CCSD/jun-cc-pVDZ level of theory, performed using Molpro 2012.1.

2.3 Electron-attached states of BPP

Before describing the technical setup of the simulations which addressed the electron-attached states of BPP, we very briefly review some general information on the electronic structures of molecular anions. For a more comprehensive discussion, the Reader is referred to monographs on systems of this type.[23–27]

Electron attachment to an isolated neutral molecule can result in the formation of anionic

Figure S6: Schematic illustration of the quantities to describe the energies of electron-attached states, on the example of a molecule B whose electronic ground state is a singlet. The ground state of the anion B^- is the doublet D_0 state, whose VAE and AEA are each positive.



states of several types, which can be broadly categorized into valence-bound (VB) and nonvalence states. In a VB state, the excess electron can be described as occupying a valence-type molecular orbital, such that the total electron density of the anion is fairly compact (although usually more diffuse than that of the neutral species). Those anionic states which do not meet this criterion are nonvalence states. In particular, molecules whose electric dipole moments are equal or higher than a critical value of roughly 2.4 D possess dipole-bound (DB) states, in which the excess electron resides in a diffuse region of space at the positive end of the neutral molecule's dipole moment, and it is bound to the molecule through the charge-dipole interaction and electron correlation effects.[23,24] Also known are examples of systems in which the excess electron is bound by higher-order multipoles.[28–30] What is more, certain molecules and clusters which do not possess a permanent electric dipole moment exhibit nonvalence anionic states of the correlation-bound (CB) type, in which the excess electron occupies a diffuse volume of space around the molecule, and is bound via a combination of electron correlation effects and electrostatic interactions.[31–36]

As shown in Figure S6, the energetics of electron attachment is conventionally described by quantities such as the vertical electron affinity (VEA) and the adiabatic electron affinity (AEA). When electron attachment is viewed as an instantaneous process, with no time to allow the relaxation of the nuclear geometry, the appropriate quantity is the VEA, which is defined as the vertical energy difference between the electronic ground state of the neutral molecule, and the electron-attached state, at the equilibrium geometry of the former. Note that a molecule will, in general, possess multiple electron-attached states, possibly of different types, and hence multiple corresponding VEA values.

If a minimum exists on the PES of a given electron-attached state, then the AEA can be defined as the energy difference between the electronic ground state of the neutral molecule, and the electron-attached state, at their respective equilibrium geometries, corrected for the zero-point vibrational energy of each species. Thus, the AEA takes into account the relaxation of the nuclear geometry following electron attachment. A positive AEA implies that the anion is lower in energy than the neutral molecule and the free electron, and that it is stable with respect to electron autodetachment. On the other hand, a negative AEA indicates that the electron-attached state lies within a continuum of states of the neutral molecule and the free electron. In this case, the electron is adiabatically unbound, and the anion can exist, if at all, only as a temporary resonance state that is prone to autodetachment.

Regarding the electron-attached states of BPP, it is safe to assume that in the condensed phase,

the diffuse nature of any nonvalence anionic states causes them to become strongly destabilized relative to the VB states, as is also the case for natural nucleobases.[26] It follows that an accurate description of the nonvalence states of BPP is not essential for our purposes. This is fortunate, because an accurate treatment of nonvalence states requires very extensively augmented basis sets, and calculated electron affinities tend to converge only slowly with respect to augmentation with increasingly diffuse functions.[27,37] In contrast, the VB states are less demanding on the basis set: as a rule, augmentation with standard diffuse basis functions is essential, but the addition of further, increasingly diffuse, basis functions has little effect.[37] We have verified that the reduced augmentation scheme within the jun-cc-pVDZ basis set performs very well for the VB states of cytosine and its analogues. In the case of the FBA F (see Figure 1 (c) in the main body of our paper), for example, an EOM-CCSD/jun-cc-pVDZ calculation underestimates the VEA of the lowest VB state by only 0.072 eV relative to a calculation with the full aug-cc-pVDZ basis set. Clearly, however, the jun-cc-pVDZ basis set is insufficiently flexible to properly describe nonvalence states.

The electron-attached states of BPP were calculated independently using two coupled cluster methods. Firstly, the VEAs of the lowest few electron-attached states were calculated at the EOM-CCSD/jun-cc-pVDZ level in combination with the continuum orbital method of Stanton and Gauss.[38] The continuum orbital method takes advantage of the fact that vertical electron attachment to a molecule can be viewed as a vertical excitation process in which an electron is promoted from a continuum orbital into a virtual orbital localized on the molecule. Then, the VEA is equal in magnitude to the corresponding excitation energy, but with opposite sign. In practice, the continuum orbital is represented by a very diffuse (low-exponent) s-type orbital.[38] This enables the calculation of the VEA using a standard electronic structure method for the calculation of excitation energies, such as EOM-CCSD.[38] An advantage of this method is that it is not restricted to the electronic ground state of the anion; multiple electron-attached states can be calculated.

The EOM-CCSD calculation was performed within Molpro version 2012.1.[22] The continuum orbital was represented by introducing a "dummy" atom positioned at the center of nuclear charge, whose basis set consisted of a single s-type Gaussian primitive with exponent 1×10^{-12}. The following procedure was applied to generate the reference RHF determinant for the EOM-CCSD calculation: firstly, an RHF calculation was performed for the molecule as a neutral system, such that the resulting canonical LUMO had an energy of zero (to within 0.0001 eV) and consisted predominantly of the low-exponent s-type primitive, with minimal contributions from other atomic orbitals. Thus, the canonical LUMO was the continuum orbital; the next-lowest canonical unoccupied orbital (LUMO+1) had a positive energy of 0.814 eV. Then, the canonical LUMO was populated by an extra two electrons without allowing the orbitals to relax. Subsequently, the CCSD and EOM-CCSD calculations were performed as normal. It was verified that in the ground-state CCSD calculation, all singles amplitudes involving the continuum orbital were vanishingly small, i.e. there was no significant relaxation of the continuum orbital at the CCSD stage. The three lowest electron-attached states each of symmetry species A' and A'' and were included in the EOM-CCSD calculation. As in the calculation of the vertical excitation spectrum, the frozen-core approximation was applied.

Secondly, as a verification of the above EOM-CCSD calculation, the VEA of the lowest electron-attached state was re-calculated at the CCSD(T)/jun-cc-pVDZ level of theory.[39] Here, the VEA was calculated directly from the definition as the vertical energy difference between the neutral species and the anion. The same level of theory was also used to calculate the vertical ionization potential (VIP) of BPP. The CCSD(T) calculations were performed with the use of the program ORCA, version 3.0.3.[40] A RHF reference determinant was used for the neutral molecule, whereas for the anion and the cation, ROHF reference determinants were used. The frozen-core approximation was applied.

In order to better understand the nature of the interactions between the BPP anion and the

purine cation in an inter-base CT state, we evaluated the molecular electrostatic potential[41–44] (MEP) generated by the lowest-energy electron-attached state of BPP. The calculation was performed with the BPP anion at the MP2/jun-cc-pVDZ ground-state equilibrium geometry of the neutral molecule, so as to represent conditions prevalent immediately after an initial photoexcitation of a base pair, while the system remains close to the ground-state equilibrium geometry. The CCSD/jun-cc-pVDZ level of theory was employed as implemented in Gaussian 09, Revision D.01.[1] An UHF reference determinant was used, and the frozen-core approximation was applied. The MEP was computed from the generalized density corresponding to the CCSD energy.[45,46]

2.4 Base Pairs of BPP

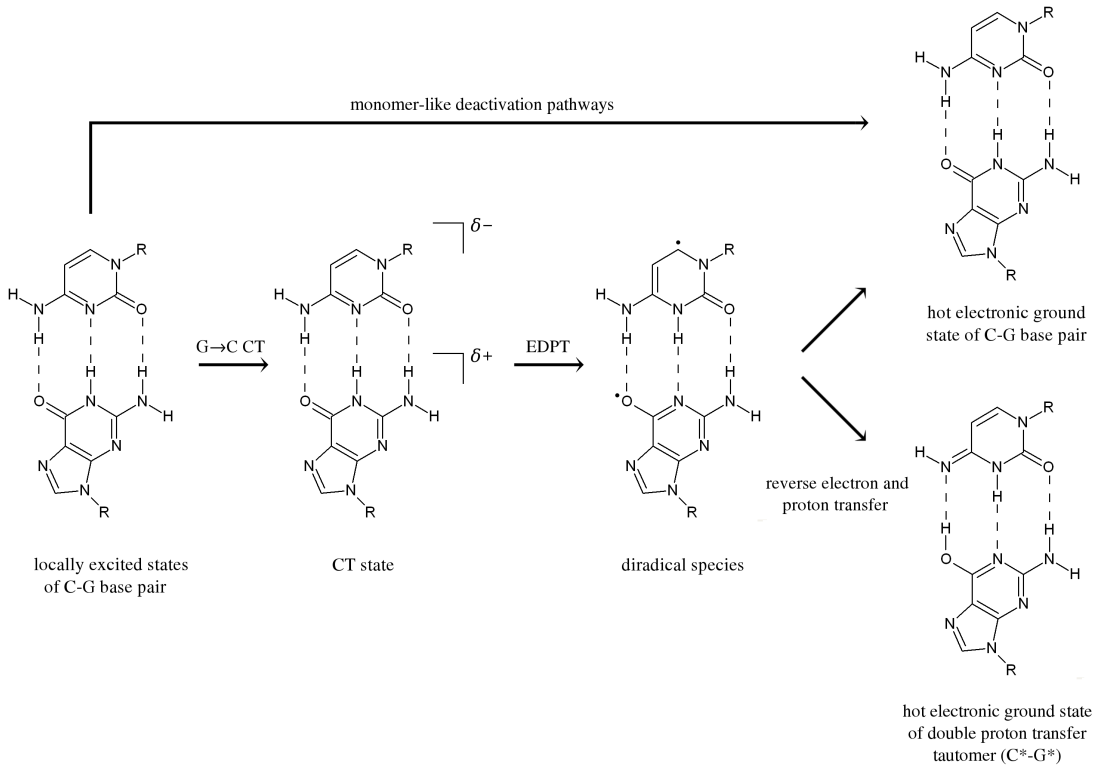
Due to the large size of the base pairs of BPP, we described their electronic structures with the use of the MP2/ADC(2) combination. As discussed earlier, the quantitative accuracy of MP2/ADC(2) for excitation energies and the electronic structures of excited states generally varies depending on the character of the excited state in question; the reasons for this inconsistent behavior are not completely understood.[8,47] Previous studies by Szalay et al.[48] and Aquino et al.[49] have reported that MP2/ADC(2) underestimates the energies of intermolecular CT states in π-stacked molecular clusters, by as much as several tenths of an electronvolt in the worst cases. On the other hand, it seems that in hydrogen-bonded molecular clusters such as base pairs, the most pessimistic case is not realized: in preliminary simulations, we found that for the C-G base pair in the Watson-Crick conformation, the MP2/ADC(2)/jun-cc-pVDZ level of theory places the vertical excitation energy into the G→C CT state at 5.117 eV, which is only 0.25 eV lower than the benchmark value of 5.36 eV obtained by Szalay et al.[48] at the EOM-CCSD(T)/TZVP level. Hence, we believe that MP2/ADC(2) is accurate enough to provide a semiquantitative description of the electronic excitation spectra of the base pairs of BPP. However, true quantitative accuracy certainly cannot be expected.

Ground-state MP2[50,51] and excited-state ADC(2)[52,53] calculations were performed within the computational chemistry software package Turbomole version 6.3.1,[54] taking advantage of the frozen core and resolution of the identity approximations. The default auxiliary basis set for aug-cc-pVDZ[55] was used as the auxiliary basis set for jun-cc-pVDZ.

As with the isolated BPP molecule and its anion, the calculations for the base pairs of BPP were performed *in vacuo*. Consequently, they represent an idealized model which includes the mutual interactions of BPP and its purine partner, but not the interactions between the base pair and its environment. In reality, of course, when incorporated into an ODN, the base pair is partially exposed to the aqueous solvent, which will influence the energies of its excited states – expecially since BPP is much larger than a pyrimidine nucleobase, and is likely jutting out from within the base stack. Neglecting these interactions can, however, be justified in *post hoc* fashion by noting that simulations suggest that interactions within the base pairs are sufficient to explain the selectivity of the quenching process.

By analogy to the Watson-Crick base pairs of canonical nucleobases,[56–60] intermolecular CT processes in the BPP-A and BPP-G base pairs may potentially be followed by electron-driven proton transfer (EDPT), whereby a proton is transferred in the same direction as the electron (see Figure S7). The possible involvement of photoinduced EDPT in the relaxation dynamics of DNA was apparently first recognized by Guallar and coworkers,[56] who explored the photophysics of the base pairs A-T and C-G using the configuration interaction singles (CIS) method. In that study, at the Franck-Condon geometry of either base pair, the CT states were found to be well separated in energy from the lowest $\pi\pi^*$-type states. For this reason, the CT states were concluded to be inaccessible following photoexcitation in the UV range; it was speculated that they may be populated following exposure to high-energy radiation. Crucially, it was noticed that purine→pyrimidine CT states were

Figure S7: Schematic diagram of electron-driven proton transfer (EDPT) in the C-G base pair. The diradical species resulting from EDPT may undergo further relaxation through a reverse transfer of an electron and proton from the cytosine radical to the guanine radical, which leads either to a recovery of the original canonical tautomer in the vibrationally hot ground state, or to the formation of the double-transfer tautomer (C*-G*), in the vibrationally hot ground state.[62]



strongly stabilized by proton transfer from the purine base to the pyrimidine base. The diradical species that are the products of EDPT (see Figure S7) were identified as minima on the PESs of the purine→pyrimidine CT states, from which it was inferred that their lifetimes may be relatively long.

However, later studies,[57,59,60] where the photophysics of the base pairs was addressed using correlated methods such as complete active space second-order perturbation theory (CASPT2) and the second-order approximate coupled cluster (CC2) method, have made substantial revisions to the above model of EDPT. Namely, it was demonstrated that the purine→pyrimidine CT states actually lie low enough in energy to enable population transfer from the lowest $\pi\pi^*$-type states.[57,59,60] Though it was confirmed that the purine→pyrimidine CT states are strongly stabilized by EDPT from the purine base to the pyrimidine base, the EDPT process was found to lead not to a minimum on the PES of the CT state, but to the vicinity of an S_1/S_0 crossing between the CT state and the singlet ground state.[57,59,60] This crossing provides a means for the electronically excited base pair to return to the electronic ground state. These findings led Sobolewski, Domcke, and coworkers to formulate the hypothesis that purine→pyrimidine CT processes followed by EDPT provide a rapid and efficient radiationless deactivation mechanism for DNA/RNA base pairs excited in the UV range.[57,59,60]

Accordingly, EDPT must also be taken into account a possible radiationless deactivation mechanism of the base pairs of BPP. To this end, minima on the S_1/S_0 crossing seam (MXSs) accessible

through EDPT were optimized using the penalty function method of Ciminelli and coworkers,[61] in combination with the limited-memory BFGS optimization algorithm. The penalty function parameters were set to the values recommended in the work just cited: $c_1 = 5$ (kcal/mol)$^{-1}$ and $c_2 = 5$ kcal/mol.

For each resulting MXS, we scanned out the ground- and excited-state PESs along a path connecting the Franck-Condon geometry (i.e. the ground-state equilibrium geometry) to the MXS. The path used in the PES scan was constructed as follows: firstly, the MXS geometry was reoriented so as to minimize the root-mean-square deviation of positions of non-hydrogenic atoms between the MXS geometry, and the Franck-Condon geometry. Then, the scan path was generated via linear interpolation in Cartesian coordinates between the two geometries. Because the linearly interpolated pathway will not, in general, coincide with the true minimum energy pathway connecting the same geometries, any potential energy barrier along the linearly interpolated path is, in general, higher than or equal to the true potential barrier along the minimum energy pathway. What is more, the interpolated path depends on the choice of coordinate system used to carry out the interpolation, and it is therefore a somewhat arbitrary construct. Nonetheless, this technique does provide a useful insight into the topography of the PESs in the intervening region.

We close this section by noting that the role (if any) of EDPT in the relaxation of UV-excited natural DNA/RNA remains under debate, as experimental and theoretical evidence has been presented both for and against the occurrence of this process under biologically relevant conditions.[62–78] On the theoretical side, one of the points of controversy is the nature of the interaction of a base pair undergoing EDPT with its environment. Solvation effects in EDPT in nucleobase pairs were first analyzed by Markwick et al.[63,64] who simulated the relaxation dynamics of the C-G base pair, photoexcited initially into a $\pi\pi^*$ state localized on guanine, in the gas phase and in aqueous solution. The electronic structures of the base pair and the solvent were treated on an equal footing with the use of the restricted open-shell Kohn-Sham (ROKS) method.[79] Under both gas-phase and solution conditions, in the majority of simulated fewest switches surface hopping (FSSH) trajectories the initial photoexcitation was followed by charge transfer and EDPT from guanine to cytosine within a few hundreds of femtoseconds.[64] The rate and the free energy profile for EDPT were found to be only slightly affected by aqueous solvation;[63,64] in particular, the lifetimes fitted to the decay of the excited-state population in gas phase and in solution were the same to within the uncertainty of the fit.[64] Likewise, studies where the environment of nucleobases was modeled using a hybrid QM/MM approach have predicted the occurrence of EDPT in the C-G base pair embedded in DNA,[65] and also in the adenine uracil monophosphate (ApU) in solution,[76] which is not, strictly speaking, a base pair, but it does possess conformations in which the bases form a hydrogen bond.

In contrast, the studies by Biemann et al.[69] and Dargiewicz et al.[71], where the solvent was treated using a variant of the polarizable continuum model (PCM), reported much stronger solvation effects which disfavoured the relaxation of the CT states of both the G-C and A-T base pairs via EDPT. This was attributed to the thermodynamic stabilization of the CT states relative to the diradical species by polar solvation.[69] However, no such effect was evident in the in the FSSH simulations of Markwick et al.,[64] possibly because charge transfer and the subsequent EDPT occurred faster than the relaxation of the solvent to adapt to the charge distribution of the CT state. From the above, we conjecture that explicit solvent simulations are better able to capture the far-from-equilibrium character of solvent dynamics during CT and EDPT processes than PCM methods, and that the explicit solvent simulations are likely correct in predicting that these processes are only slightly affected by polar solvation. This reasoning also lends some support to the validity of the approximation made by performing simulations *in vacuo*, as we have done.

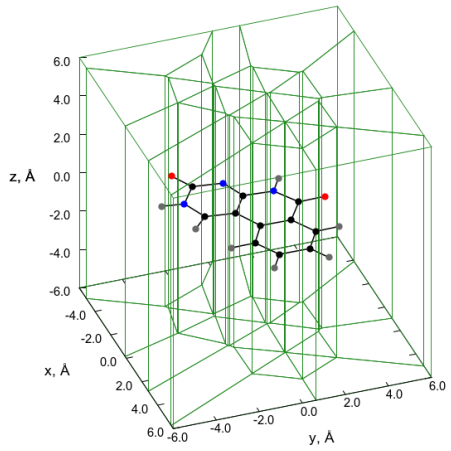
2.5 Analysis of Electronic Structures

The redistribution of electron density in the excited, and electron-attached, electronic states of BPP and its base pairs was analyzed in terms of the electron density difference map (EDDM, denoted $\Delta\rho(\mathbf{r})$), which is defined simply as the difference between the electron density of a given state i ($\rho_i(\mathbf{r})$), and that of the singlet ground state of the neutral system ($\rho_0(\mathbf{r})$).

$$\Delta\rho(\mathbf{r}) = \rho_i(\mathbf{r}) - \rho_0(\mathbf{r}) \tag{15}$$

In order to quantify the shift of electron density from one base to another, a Voronoi-Dirichlet tesselation was performed on the space around the nuclei. This procedure assigns to each atom (which in this context is considered to be a point positioned at the nucleus) a Voronoi-Dirichlet polyhedron (VDP) that contains all points whose distance to that atom is not greater than the distance to any other atom. In this manner, space is partitioned completely and disjointly into VDPs associated with individual atoms. (The partitioning is not disjoint in the rigorous sense, because the surface of each polyhedron is shared with the neighboring polyhedra, but this point can be ignored here.) As an illustration, Figure S8 shows the Voronoi-Dirichlet tesselation of space around the atoms of the isolated BPP molecule.

Figure S8: Perspective view of the Voronoi-Dirichlet tesselation of space around the nuclei of BPP, enclosed in a cubic cell of side length 12 Å. The green lines are the edges of the VDPs associated with the constituent atoms. The BPP molecule is symmetric with respect to reflection in the $z = 0$ plane; consequently, the resulting Voronoi-Dirichlet tesselation is also symmetric with respect to that reflection.



The VDP of an atom may be interpreted as the volume of space that it takes up, and by extension, the electronic charge contained in the VDP may be assigned as belonging to that atom. By extension, the change of the charge of an atom I upon photoexcitation, or electron attachment, can be calculated as a volume integral of the EDDM over the VDP of that atom:

$$\Delta q_I = \int_{\mathrm{VDP}(I)} \Delta\rho(\mathbf{r})\,\mathrm{d}\mathbf{r} \tag{16}$$

Likewise, in calculations for base pairs, the electronic charge belonging to a component molecule was taken as the charge contained in the union of the VDPs of the constituent atoms.

It is also of some interest to have a quantitative measure of the distance over which charge separation occurs. For that purpose, we used the CT excitation length (D_{CT}) introduced by Le Bahers and coworkers.[80] In order to calculate D_{CT} for a given electronic state, its EDDM is partitioned into a non-negative part $\rho_+(\mathbf{r})$ and a non-positive part $\rho_-(\mathbf{r})$:

$$\rho_+(\mathbf{r}) = \max(0\,,\,\Delta\rho(\mathbf{r})) \qquad \text{and} \qquad \rho_-(\mathbf{r}) = \min(0\,,\,\Delta\rho(\mathbf{r})) \tag{17}$$

Then, the CT excitation length is defined as the distance between the barycenter of the non-negative part of the EDDM ($\mathbf{r}_+$), and the barycenter of the non-positive part ($\mathbf{r}_-$).

$$D_{\mathrm{CT}} = |\,\mathbf{r}_+ - \mathbf{r}_-\,| \qquad \text{where} \qquad \mathbf{r}_+ = \frac{\int \mathbf{r}\,\rho_+(\mathbf{r})\,\mathrm{d}\mathbf{r}}{\int \rho_+(\mathbf{r})\,\mathrm{d}\mathbf{r}} \qquad \text{and} \qquad \mathbf{r}_- = \frac{\int \mathbf{r}\,\rho_-(\mathbf{r})\,\mathrm{d}\mathbf{r}}{\int \rho_-(\mathbf{r})\,\mathrm{d}\mathbf{r}} \tag{18}$$

Electronic structure programs such as Turbomole and Molpro print electron densities on three-dimensional regular grids. In order to achieve good accuracy when evaluating numerically the integrals appearing in equations (2) and (4), the EDDMs calculated from the definition (1) were transferred onto finer grids through trilinear interpolation. Subsequently, the integrals were evaluated numerically on the finer grids using the midpoint rule.

3 Charge Transfer in Alternative Base-Pair Geometries

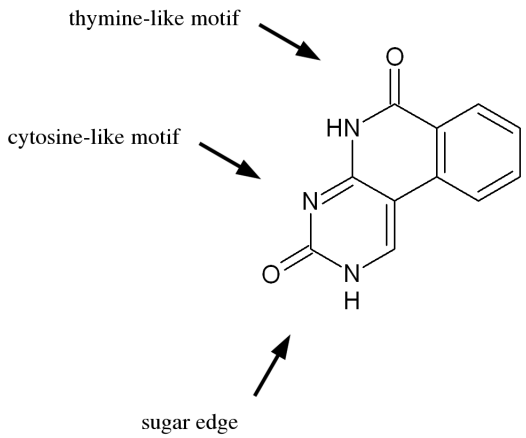
As discussed in Section 3.3 of the main body of the present paper, the molecular electrostatic potential (MEP) generated by the BPP^- anion in its ground electronic state is highly anisotropic. A purine→BPP CT state will benefit from particularly strong stabilization when the purine nucleobase is hydrogen-bonded to the cytosine-like motif along the extended hydrogen-bonding face of BPP – in other words, when the nucleobase in question is guanine, or perhaps adenine in its rare imino tautomeric form. On the other hand, the electrostatic interactions are expected to be weaker when the purine nucleobase is hydrogen-bonded either to the thymine-like motif, or to the sugar edge (SE) of BPP (see Figure S9 below). The latter situation cannot be realized when BPP is incorporated into an oligodeoxynucleotide, but nevertheless it provides an interesting test case of the correctness of our model of the electrostatic interactions between the cationic form of the purine nucleobase and the anionic form of BPP.

Accordingly, in order to verify our understanding of the electrostatic stabilization of CT states in base pairs of BPP, we have optimized, at the MP2/jun-cc-pVDZ level of theory, the ground-state equilibrium geometries of the base pairs BPP(SE)-A and BPP(SE)-G, in which the purine bases are hydrogen-bonded to the sugar edge of BPP. Afterwards, we computed their vertical excitation spectra at the same level of theory (MP2/ADC(2)/jun-cc-pVDZ) as was employed for the biologically relevant BPP-A and BPP-G base pairs.

The resulting equilibrium geometries are shown in Figure S10. Interestingly, as indicated in Figure S10 (a), in the BPP(SE)-G base pair, both hydrogen bonds are very short. The vertical excitation spectra of the base pairs are characterized in Table S3; for easier comprehension, the energy levels of the base pairs and individual component molecules are also presented visually in Figure S11.

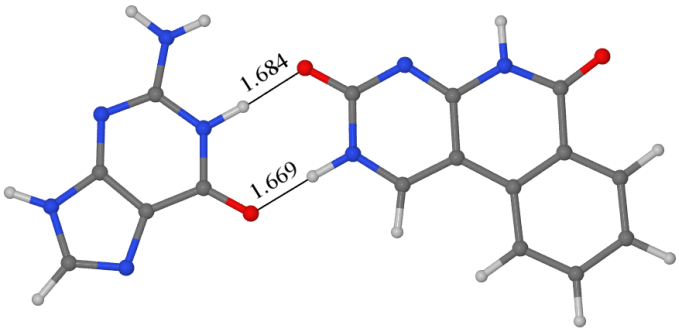
In the BPP(SE)-G base pair, the inter-base CT state is found almost 1 eV higher in energy (at 4.948 eV) than in the "normal" BPP-G base pair (4.000 eV). This is exactly as expected on the basis of the MEP generated by the BPP^- anion – the negative region of the MEP covers little of

Figure S9: Locations of hydrogen-bonding sites of the isolated BPP molecule.

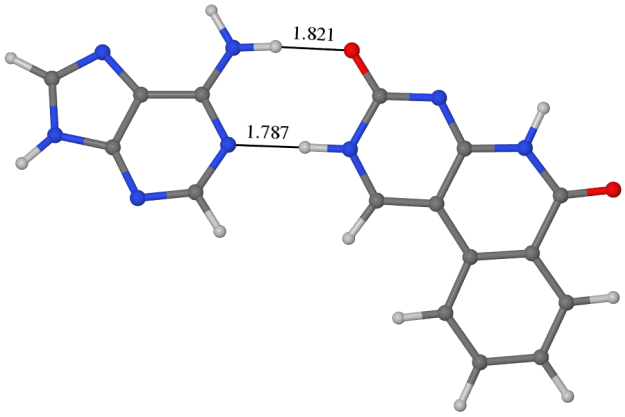


the volume of space occupied by a nucleobase that is hydrogen-bonded to the sugar edge of BPP (see Figure 13 in the main body of the present paper). In the case of the BPP(SE)-A base pair, the inter-base CT state is not even detected among the lowest ten vertically excited states. This suggests that such a state must lie at an energy of over 5 eV, which is again higher than in the "normal" BPP-A base pair. These results demonstrate that the energies of CT states in the base pairs of BPP are highly sensitive to the relative orientations on the purine nucleobase and BPP (or, in other words, on the mode of hydrogen bonding between the purine nucleobase and BPP), and that the energy of the CT state can be predicted in qualitative terms on the basis of the MEP of the BPP^- anion.

Figure S10: Ground-state equilibrium geometries of the BPP(SE)-A and BPP(SE)-G base pairs. The lengths of hydrogen bonds are marked in units of Å.



(a) BPP(SE)-G



(b) BPP(SE)-A

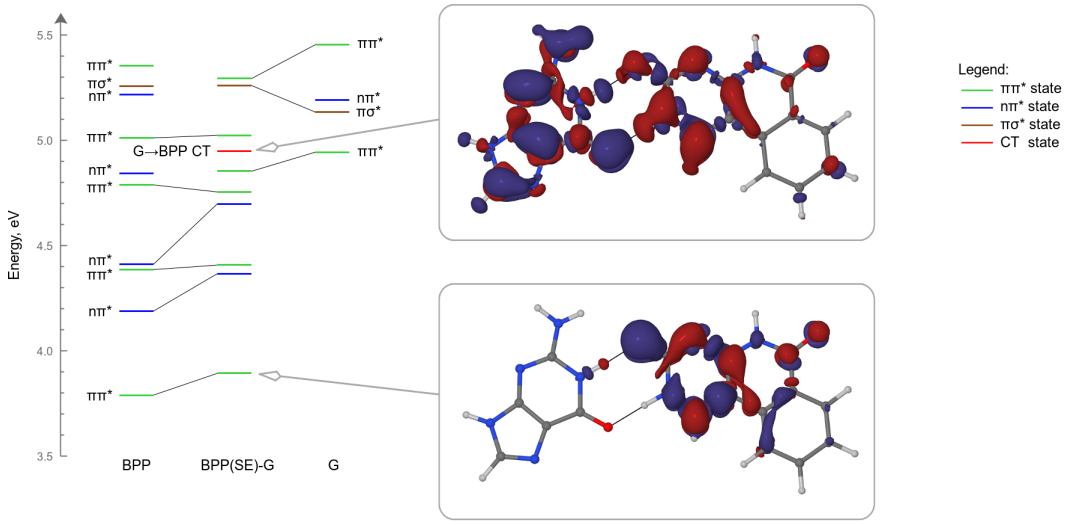
Table S3: Vertical excitation energies (ΔE), oscillator strengths (f), and (orbital-relaxed) electric dipole moment magnitudes (μ) of the base pairs BPP(SE)-A and BPP(SE)-G, computed at the MP2/ADC(2)/jun-cc-pVDZ level of theory. Δq_{BPP} is the amount of charge accepted by the BPP fragment in a given electronic excited state, according to the Voronoi-Dirichlet spatial paritioning scheme, and D_{CT} denotes the charge-transfer length as defined by of Le Bahers and coworkers.[80]

System	State	ΔE, eV	f	μ, D	Δq_{BPP}, e	D_{CT}, Å
BPP(SE)-G	S_0			3.88		
	S_1 (BPP $\pi\pi^*$)	3.892	0.201	5.09	−0.016	0.85
	S_2 (BPP nπ^*)	4.365	0.001	1.34	−0.003	0.94
	S_3 (BPP $\pi\pi^*$)	4.406	0.008	5.20	−0.011	1.53
	S_4 (BPP nπ^*)	4.697	0.002	8.97	−0.092	1.57
	S_5 (BPP $\pi\pi^*$)	4.754	0.112	5.06	−0.022	1.22
	S_6 (G $\pi\pi^*$)[a]	4.852	0.233	13.76	−0.305	3.52
	S_7 (G→BPP CT)	4.948	0.033	22.88	−0.581	4.21
	S_8 (BPP $\pi\pi^*$)	5.023	0.806	2.88	−0.009	1.09
	S_9 (G $\pi\sigma^*$)[b]	5.260	0.001	6.27	0.014	1.76
	S_{10} (G $\pi\pi^*$)	5.293	0.334	2.50	0.004	0.45
BPP(SE)-A	S_0			9.00		
	S_1 (BPP $\pi\pi^*$)	3.797	0.191	9.07	−0.012	0.92
	S_2 (BPP nπ^*)	4.295	0.001	6.24	−0.025	0.84
	S_3 (BPP $\pi\pi^*$)	4.351	0.011	9.22	−0.011	1.78
	S_4 (BPP nπ^*)	4.486	$< 10^{-3}$	7.45	−0.035	1.02
	S_5 (BPP $\pi\pi^*$)	4.740	0.181	8.26	−0.003	1.09
	S_6 (BPP nπ^*)	4.914	0.051	7.72	−0.036	1.15
	S_7 (BPP $\pi\pi^*$)	4.935	0.749	5.23	0.001	1.11
	S_8 (A $\pi\pi^*$)	4.999	0.037	7.19	−0.047	1.01
	S_9 (A $\pi\pi^*$)	5.079	0.239	7.99	−0.035	0.58
	S_{10} (BPP nπ^*)	5.249	$< 10^{-3}$	11.38	−0.071	1.73

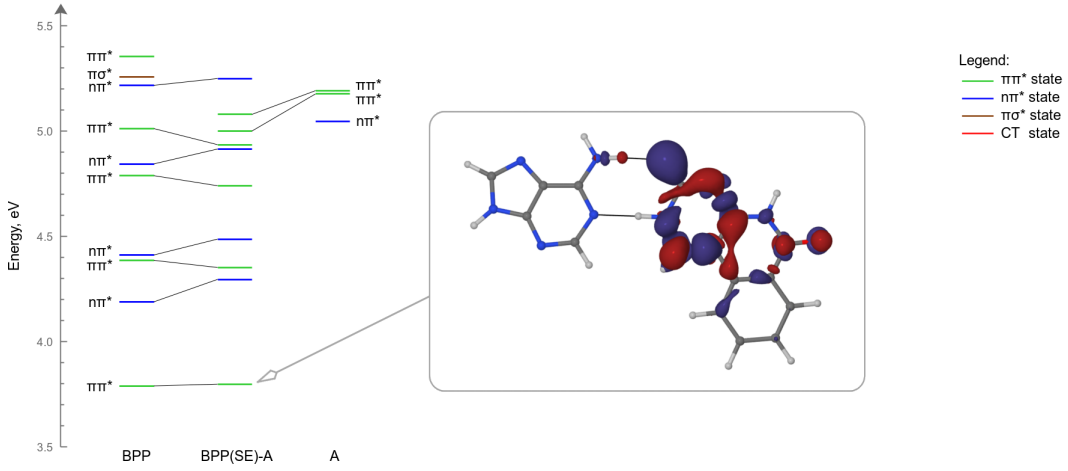
[a] The S_6 (G $\pi\pi^*$) state of the BPP(SE)-G base pair exhibits an admixture of G→BPP CT character.
[b] Rydberg-type state involving the excitation of an electron from a π-type orbital of guanine into a diffuse σ^*-type orbital surrounding that molecule.

Figure S11: Energy level diagrams for the BPP(SE)-G and BPP(SE)-A base pairs and the isolated component molecules BPP, A, and G. All calculations were performed at the MP2/ADC(2)/jun-cc-pVDZ level of theory, at the ground-state equilibrium geometry of the given system. The horizontal bars representing the excited electronic states are colored according to their character: green for $\pi\pi^*$ states, blue for $n\pi^*$ states, brown for $\pi\sigma^*$ states, and red for CT states. The black lines connect states of the same nature found in the isolated BPP molecule, and in its base pairs. The insets on the right-hand side show EDDMs for selected excited states of the base pairs, plotted in the form of isosurfaces with isovalues of $\pm 0.002\ {a_0}^{-3}$.

(a) The BPP(SE)-G base pair



(b) The BPP(SE)-A base pair



4 Molecular geometries

This section provides the equilibrium geometries of BPP and its Watson-Crick-type base pairs with adenine and guanine. For the base pairs, we also list the geometries of minima along the S_1/S_0 crossing seams (MXSs) which arise from electron-driven proton transfer.

All geometries are given in terms of Cartesian coordinates in units of ångström (Å).

Ground-state equilibrium geometry of BPP as optimized at the MP2/jun-cc-pVDZ level of theory, within the C_s point group.

```
23

N    -3.0644110   -0.3808519    0.0000000
C    -3.2636208   -1.7514666    0.0000000
N    -2.0804342   -2.5451399    0.0000000
C    -0.8185893   -2.0486522    0.0000000
C    -0.6133370   -0.6829228    0.0000000
C    -1.8306315    0.0935762    0.0000000
N    -1.7105120    1.4704281    0.0000000
C    -0.5281252    2.2100167    0.0000000
C     0.6914000   -0.0141177    0.0000000
C     0.7261753    1.4036760    0.0000000
C     1.9546484    2.0965457    0.0000000
C     3.1592224    1.3845490    0.0000000
C     3.1397358   -0.0268627    0.0000000
C     1.9209736   -0.7158663    0.0000000
H     1.9323709   -1.8106107    0.0000000
H     1.9343510    3.1903989    0.0000000
H     4.1137766    1.9205406    0.0000000
H     4.0787938   -0.5901663    0.0000000
H    -2.2403628   -3.5514094    0.0000000
H    -0.0076838   -2.7819026    0.0000000
H    -2.5836398    2.0014519    0.0000000
O    -4.3547148   -2.3095015    0.0000000
O    -0.5553856    3.4382875    0.0000000
```

Equilibrium geometry of BPP in its 2 ^{1}A$'$ ($\pi\pi^*$) state, as optimized at the EOM-CCSD/cc-pVDZ level of theory with no symmetry constraints.

23

C	-1.138517447	1.409587918	0.000300990
C	-1.190812719	-0.016437274	0.000241254
C	0.038569573	-0.766434224	-0.000063252
C	1.268481339	-0.071810047	-0.000299258
N	1.268788807	1.309953173	-0.000209873
C	0.159336786	2.139639610	0.000067392
C	-2.323242676	2.163947077	0.000584239
C	-2.470863361	-0.646895521	0.000478281
N	2.512805337	-0.639577100	-0.000607215
C	2.554912297	-1.983702456	-0.000595025
N	1.352916772	-2.760522472	-0.000301160
C	0.099714840	-2.220326526	-0.000161958
H	-0.757433786	-2.890125615	-0.000164405
O	3.590152824	-2.664594182	-0.000870897
O	0.263675103	3.358946855	0.000147191
H	2.181013087	1.759439906	-0.000375916
H	1.503801990	-3.765617222	-0.000408228
C	-3.569557354	1.527760822	0.000799642
C	-3.636510141	0.116300476	0.000745323
H	-4.489583973	2.119980499	0.001012511
H	-2.236371136	3.253605383	0.000622297
H	-2.544245967	-1.737681277	0.000443544
H	-4.610191172	-0.383930033	0.000917162

Equilibrium geometry of the BPP^{-} anion in its 1 $^{2}A''$ (VB) state, as optimized at the UMP2/jun-cc-pVDZ level of theory with no symmetry constraints.

23

N	-3.060013000	-0.378400000	-0.062209000
C	-3.243316000	-1.742208000	-0.042339000
N	-2.103481000	-2.554361000	0.137323000
C	-0.781854000	-2.106523000	-0.050116000
C	-0.598547000	-0.692682000	0.019796000
C	-1.791561000	0.071422000	-0.010789000
N	-1.675603000	1.459697000	0.026752000
C	-0.521313000	2.224410000	0.027233000
C	0.662229000	-0.006186000	-0.000617000
C	0.717734000	1.428234000	0.034076000
C	1.957348000	2.107008000	0.031161000
C	3.160888000	1.395807000	0.013942000
C	3.128906000	-0.022744000	-0.020481000
C	1.910615000	-0.703980000	-0.021267000
H	1.910513000	-1.798718000	-0.060927000
H	1.938131000	3.202038000	0.050432000
H	4.119110000	1.927964000	0.014511000
H	4.066281000	-0.592618000	-0.042379000
H	-2.306474000	-3.540543000	-0.002255000
H	0.001406000	-2.820805000	0.215530000
H	-2.560227000	1.968994000	0.005913000
O	-4.353078000	-2.297062000	-0.134329000
O	-0.577693000	3.471257000	0.033040000

Ground-state equilibrium geometry of the BPP-A base pair, as optimized at the MP2/jun-cc-pVDZ level of theory

38

N	1.8317392	4.8384013	0.1865725
C	1.7771532	3.4574905	0.2385123
N	3.0343453	2.7878089	0.2608051
C	4.2372850	3.4118089	0.2320902
C	4.2985627	4.7913875	0.1988560
C	3.0102342	5.4453044	0.1624174
N	2.9758324	6.8242254	0.1186306
C	4.0756175	7.6495320	-0.0057829
C	5.5276422	5.5849523	0.1342569
C	5.4113877	6.9958201	0.0594803
C	6.5592817	7.8144987	0.0057747
C	7.8323445	7.2359009	0.0498091
C	7.9629766	5.8323182	0.1332773
C	6.8247195	5.0191759	0.1854795
H	6.9519257	3.9338641	0.2529085
H	6.4223199	8.8980327	-0.0570543
H	8.7248068	7.8690438	0.0137934
H	8.9568260	5.3736322	0.1675997
H	2.9785375	1.7708243	0.2896983
H	5.1203273	2.7672387	0.2462983
H	2.0224430	7.2797889	0.0601038
O	0.7479240	2.7895728	0.2634761
O	3.9423158	8.8746985	-0.1418129
N	0.4373129	8.0788086	-0.0977399
C	-0.6441410	7.2587881	-0.2313721
N	-1.9206243	7.6006753	-0.4603369
C	-2.0501126	8.9402249	-0.5603390
C	-1.0349679	9.9072203	-0.4421628
C	0.2738560	9.4186586	-0.1970164
N	-1.5220289	11.1935092	-0.6122406
C	-2.8244499	11.0003620	-0.8256364
N	-3.1956993	9.6742239	-0.8038943
N	1.3351171	10.2449442	-0.0077000
H	-3.5498089	11.7950115	-1.0033089
H	-4.1282440	9.2944721	-0.9413221
H	-0.4210373	6.1908161	-0.1355134
H	1.2097476	11.2144932	-0.2821182
H	2.2767426	9.8417411	-0.0590988

Structure of minimum along the S_1/S_0 crossing seam (MXSs) arising from electron-driven proton transfer (EDPT) in the BPP-A base pair.

38

N	1.8079527	4.8312843	0.1848808
C	1.7722645	3.4477749	0.2393962
N	3.0109405	2.7795662	0.2642802
C	4.2423434	3.4117604	0.2309691
C	4.3035353	4.7900927	0.1989920
C	3.0118632	5.4458189	0.1607606
N	2.9917882	6.8086482	0.1101277
C	4.1276743	7.6170971	0.0211627
C	5.5198977	5.5968224	0.1316187
C	5.3926681	7.0314290	0.0655792
C	6.5748665	7.8330794	0.0015261
C	7.8284526	7.2391234	0.0485589
C	7.9712824	5.8234297	0.1330680
C	6.8230624	5.0318793	0.1871357
H	6.9432141	3.9420651	0.2526596
H	6.4452248	8.9168469	-0.0574818
H	8.7217031	7.8718576	0.0129865
H	8.9666560	5.3664298	0.1691662
H	2.9626777	1.7640319	0.2904004
H	5.1241142	2.7668918	0.2459166
H	2.0611860	7.2385753	0.0658522
O	0.7240468	2.7872103	0.2625861
O	3.9765015	8.9771321	-0.1978054
N	0.4150686	8.0965671	-0.0797170
C	-0.6162693	7.2595039	-0.2256236
N	-1.9211999	7.6339350	-0.4643366
C	-2.0703299	8.9365609	-0.5637528
C	-1.0398084	9.9338046	-0.4530120
C	0.2455374	9.4385618	-0.1995590
N	-1.5283031	11.2037721	-0.6150495
C	-2.8332939	10.9916707	-0.8259839
N	-3.2135884	9.6762162	-0.8054855
N	1.3514775	10.2144081	-0.0290824
H	-3.5497559	11.7916926	-1.0033308
H	-4.1492196	9.2984729	-0.9449709
H	-0.4039650	6.1863065	-0.1356459
H	1.1519287	11.2040570	-0.2678501
H	3.0972353	9.2977739	0.0621038

Ground-state equilibrium geometry of the BBP-G base pair, as optimized at the MP2/jun-cc-pVDZ level of theory

```
39

N     1.7312378    4.9514721   -0.0427256
C     1.6691930    3.5799457   -0.0586574
N     2.8949476    2.8928180    0.0108547
C     4.1037642    3.5117372    0.0723140
C     4.1830560    4.8895176    0.0931449
C     2.9119276    5.5746518    0.0188060
N     2.8862546    6.9393430    0.0286146
C     4.0179868    7.7688233    0.0322331
C     5.4289143    5.6612335    0.1315491
C     5.3363682    7.0746272    0.1279661
C     6.4998340    7.8713146    0.1691797
C     7.7595029    7.2662376    0.2410147
C     7.8637810    5.8582349    0.2561191
C     6.7105917    5.0656417    0.2115765
H     6.8172919    3.9760923    0.2248904
H     6.3846364    8.9592345    0.1536296
H     8.6633176    7.8829648    0.2777716
H     8.8471364    5.3795253    0.3114646
H     2.8291946    1.8764401   -0.0041908
H     4.9784750    2.8578771    0.1113014
H     1.9533174    7.4086244   -0.0933195
O     0.6156694    2.9321416   -0.1276831
O     3.9021765    8.9875068   -0.0256988
N    -0.8900853    6.2655422   -0.0384906
C    -2.0841653    5.5820632    0.0394240
N    -3.2698570    6.1454862   -0.1395145
C    -3.1553140    7.4775375   -0.4054381
C    -2.0021434    8.2751570   -0.5100694
C    -0.7255070    7.6345835   -0.3265541
N    -2.3166837    9.5913650   -0.8051413
C    -3.6449557    9.5882290   -0.8756867
N    -4.1983869    8.3432034   -0.6437096
O     0.4129039    8.1317286   -0.3915068
N    -1.9960887    4.2578679    0.3929822
H    -0.0005106    5.7383158    0.0384685
H    -4.2638578   10.4591455   -1.0917266
H    -5.1834409    8.0947769   -0.6418377
H    -1.1262456    3.7631860    0.1660753
H    -2.8520670    3.7407471    0.2122906
```

Structure of the first (MXS1) of two minima along the S_1/S_0 crossing seam arising from electron-driven proton transfer (EDPT) in the BPP-G base pair.

```
39

N     1.7606379    4.9108087   -0.0265693
C     1.6668019    3.5331989   -0.0474005
N     2.8601190    2.8566345    0.0049790
C     4.1295052    3.4707752    0.0618716
C     4.1834544    4.8899496    0.1009098
C     2.9859823    5.5855373    0.0346077
N     2.9109531    6.9714112    0.0397232
C     4.0251484    7.7996015    0.0355639
C     5.4120623    5.6707201    0.1303089
C     5.3288398    7.0929735    0.1299161
C     6.5014359    7.8801506    0.1685647
C     7.7577457    7.2716757    0.2408410
C     7.8525390    5.8603508    0.2553878
C     6.6993873    5.0737649    0.2112842
H     6.7967058    3.9836670    0.2226102
H     6.3877657    8.9686435    0.1533614
H     8.6645230    7.8834667    0.2775475
H     8.8345826    5.3773832    0.3117548
H     2.7800852    1.8440433   -0.0156269
H     4.9917093    2.8112317    0.1173683
H     1.9936686    7.4299378   -0.0924807
O     0.5611976    2.9498205   -0.1052842
O     3.9055885    9.0315685   -0.0278119
N    -0.8213788    6.2399727   -0.0066162
C    -1.9974966    5.6053505    0.0371424
N    -3.2716733    6.1672337   -0.1353430
C    -3.1950886    7.4542860   -0.4028680
C    -2.0169083    8.2602101   -0.5123279
C    -0.7324125    7.5854872   -0.3095282
N    -2.3139747    9.5551807   -0.8005770
C    -3.6580545    9.5647741   -0.8713534
N    -4.2332602    8.3450398   -0.6436149
O     0.3497577    8.1832588   -0.4045982
N    -2.0034190    4.2850002    0.3021449
H     0.8567618    5.4171342    0.0378239
H    -4.2532914   10.4532858   -1.0887948
H    -5.2267332    8.1200229   -0.6491060
H    -1.1093553    3.7722775    0.2150280
H    -2.8813418    3.7841512    0.2156315
```

Structure of the second (MXS2) of two minima along the S_1/S_0 crossing seam arising from electron-driven proton transfer (EDPT) in the BPP-G base pair.

39

N	1.6917724	4.9410402	-0.0255475
C	1.7177665	3.6228684	-0.0262454
N	2.8501493	2.8764277	0.0044251
C	4.1332191	3.4663118	0.0425705
C	4.1709886	4.8889688	0.0964674
C	2.9607715	5.5708525	0.0250164
N	2.9106980	6.9555294	0.0355699
C	4.0192960	7.7907219	0.0335498
C	5.4038551	5.6652022	0.1298347
C	5.3246146	7.0874518	0.1305220
C	6.4973771	7.8749925	0.1688772
C	7.7540369	7.2685045	0.2410806
C	7.8485296	5.8566286	0.2556055
C	6.6948409	5.0681944	0.2122819
H	6.7990562	3.9786783	0.2227650
H	6.3837539	8.9631089	0.1531107
H	8.6607234	7.8814659	0.2771342
H	8.8325315	5.3770268	0.3114082
H	2.7500235	1.8654587	-0.0014233
H	4.9885634	2.7976806	0.1166218
H	1.9939703	7.4138729	-0.0905127
O	0.5963270	2.8906638	-0.0584054
O	3.8993472	9.0230774	-0.0269350
N	-0.8647024	6.2785352	-0.0137421
C	-2.0476759	5.5471611	0.0439867
N	-3.2790799	6.1512819	-0.1521887
C	-3.1803974	7.4572812	-0.4069906
C	-2.0178180	8.2671689	-0.5067782
C	-0.7139626	7.6243546	-0.3189053
N	-2.3126636	9.5616442	-0.7970830
C	-3.6571270	9.5676818	-0.8713255
N	-4.2207717	8.3436042	-0.6463941
O	0.3713794	8.1837101	-0.4094768
N	-1.9270694	4.2647650	0.2968512
H	0.0329245	5.7416164	0.0429300
H	-4.2564753	10.4533045	-1.0888597
H	-5.2112602	8.1093227	-0.6478507
H	-0.2089179	3.4420992	0.1437740
H	-2.8572348	3.8226810	0.2712311

References

[1] Gaussian 09, Revision D.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.

[2] Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange *J. Chem. Phys.* **1993**, 98, 5648-5652, DOI: 10.1063/1.464913.

[3] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *Ab Initio* Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields *J. Phys. Chem.* **1994**, 98, 11623-11627, DOI: 10.1021/j100096a001.

[4] Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations *J. Chem. Phys.* **2010**, 133, 134105, DOI: 10.1063/1.3484283.

[5] York, D. M.; Karplus, M. A Smooth Solvation Potential Based on the Conductor-Like Screening Model *J. Phys. Chem. A* **1999**, 103, 11060, DOI: 10.1021/jp992097l.

[6] Scalmani, G.; Frisch, M. J. Continuous Surface Charge Polarizable Continuum Models of Solvation. I. General Formalism. *J. Chem. Phys.* **2010**, 132, 114110, DOI: 10.1063/1.3359469.

[7] Okamoto, A.; Tainaka, K.; Saito, I. Clear Distinction of Purine Bases on the Complementary Strand by a Fluorescence Change of a Novel Fluorescent Nucleoside *J. Am. Chem. Soc.* **2003**, 125, 4972-4973, DOI: 10.1021/ja034090u.

[8] Szalay, P. G.; Watson, T.; Perera, A.; Lotrich, V. F.; Bartlett, R. J. Benchmark Studies on the Building Blocks of DNA. 1. Superiority of Coupled Cluster Methods in Describing the Excited States of Nucleobases in the Franck-Condon Region *J. Phys. Chem. A* **2012**, 116, 6702-6710, DOI: 10.1021/jp300977a.

[9] Nakayama, A.; Harabuchi, Y.; Yamazaki, S.; Taketsugu, T. Photophysics of Cytosine Tautomers: New Insights Into the Nonradiative Decay Mechanisms From MS-CASPT2 Potential Energy Calculations and Excited-State Molecular Dynamics Simulations *Phys. Chem. Chem. Phys.* **2013**, 15, 12322-12339, DOI: 10.1039/c3cp51617b.

[10] Szalay, P. G.; Watson, T.; Perera, A.; Lotrich, V.; Fogarasi, G.; Bartlett, R. J. Benchmark Studies on the Building Blocks of DNA. 2. Effect of Biological Environment on the Electronic Excitation Spectrum of Nucleobases *J. Phys. Chem. A* **2012**, 116, 8851-8860, DOI: 10.1021/jp305130q.

[11] Szalay, P. G.; Watson, T.; Perera, A.; Lotrich, V.; Bartlett, R. J. Benchmark Studies on the Building Blocks of DNA. 3. WatsonCrick and Stacked Base Pairs *J. Phys. Chem. A* **2013**, 117, 3149, DOI: 10.1021/jp3100975.

[12] Kánnár, D.; Szalay, P. G.; Benchmarking Coupled Cluster Methods on Singlet Excited States of Nucleobases *J. Mol. Model.* **2014**, 20, 2503, DOI: 10.1007/s00894-014-2503-2.

[13] Christiansen, O.; Koch, H.; Jørgensen, P. The Second-Order Approximate Coupled Cluster Singles and Doubles Model CC2 *Chem. Phys. Lett.* **1995**, 243, 409-418, DOI: 10.1016/0009-2614(95)00841-Q.

[14] Trofimov, A. B.; Schirmer, J. An Efficient Polarization Propagator Approach to Valence Electron Excitation Spectra *J. Phys. B: At., Mol. Opt. Phys.* **1995**, 28, 2299-2324, DOI: 10.1088/0953-4075/28/12/003.

[15] Hättig, C. Structure Optimizations for Excited States with Correlated Second-Order Methods: CC2 and ADC(2). In *Advances in Quantum Chemistry*; Jensen, H. J. Å.., Ed.; Academic Press: New York, 2005; Vol. 50; pp 37-60. DOI: 10.1016/S0065-3276(05)50003-0.

[16] Hohenstein, E. G.; Kokkila, S. I. L.; Parrish, R. M.; Martínez, T. J. Tensor Hypercontraction Equation-of-Motion Second-Order Approximate Coupled Cluster: Electronic Excitation Energies in $\mathcal{O}(N^4)$ Time *J. Phys. Chem. B* **2013**, 117, 12972-12978, DOI: 10.1021/jp4021905.

[17] Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L. et al., PSI4: an Open-Source *Ab Initio* Electronic Structure Program *WIREs Comput. Mol. Sci.* **2012**, 2, 556-565, DOI: 10.1002/wcms.93.

[18] Parrish, R. M.; Burns, L. A.; Smith, D. G. A.; Simmonett, A. C.; DePrince, A. E.; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M. et al., Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability *J. Chem. Theor. Comput.* **2017**, 13, 3185-3197, DOI: 10.1021/acs.jctc.7b00174.

[19] Pabst, M.; Köhn, A.; Gauss, J.; Stanton, J. F. A Worrisome Failure of the CC2 Coupled-Cluster Method When Applied to Ozone *Chem. Phys. Lett.* **2010**, 495, 135-140, DOI: 10.1016/j.cplett.2010.06.023.

[20] Wang, Z.; Wang, F. Analysis of a Failure of the CC2 CoupledCluster Method for Bond Lengths of SnO and PbO *Theor. Chem. Acc.* **2014**, 133, 1579, DOI: 10.1007/s00214-014-1579-1.

[21] Papajak, E.; Zheng, J.; Xu, X.; Leverentz, H. R.; Truhlar, D. G. Perspectives on Basis Sets Beautiful: Seasonal Plantings of Diffuse Basis Functions *J. Chem. Theory Comput.* **2011**, 7, 3027-3034, DOI: 10.1021/ct200106a.

[22] MOLPRO is a package of *ab initio* programs written by H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklaß, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang. See http://www.molpro.net

[23] Jordan, K. D.; Wang, F. Theory of Dipole-Bound Anions *Annu. Rev. Phys. Chem.* **2003**, 54, 367-396, DOI: 10.1146/annurev.physchem.54.011002.103851.

[24] Simons, J. Molecular Anions *J. Phys. Chem. A* **2008**, 112, 6401-6511, DOI: 10.1021/jp711490b.

[25] Gu, J.; Leszczynski, J.; Schaefer, H. F. III Interactions of Electrons with Bare and Hydrated Biomolecules: From Nucleic Acid Bases to DNA Segments *Chem. Rev.* **2012**, 112, 5603-5640, DOI: 10.1021/cr3000219.

[26] Storoniak, P.; Wang, H.; Ko, Y. J.; Li, X.; Stokes, S. T.; Eustis, S.; Bowen, K. H.; Rak, J. Valence Anions of DNA-Related Systems in the Gas Phase: Computational and Anion Photoelectron Spectroscopy Studies In *Practical Aspects of Computational Chemistry III*; Leszczynski, J.; Shukla, M. K., Eds.; Springer US; Boston, MA, 2014; pp 323-392. DOI: 10.1007/978-1-4899-7445-7.

[27] Herbert, J. M. The Quantum Chemistry of Loosely-Bound Electrons *Rev. Comput. Chem.* **2015**, 28, 391-517, DOI: 10.1002/9781118889886.ch8.

[28] Gutowski, M.; Skurski, P.; Li, X.; Wang, L.-S. $(MgO)_n^-$ $(n = 1 - 5)$ Clusters: Multipole-Bound Anions and Photodetachment Spectroscopy *Phys. Rev. Lett.* **2000**, 85, 3145-3148, DOI: 10.1103/PhysRevLett.85.3145.

[29] Compton, R. N.; Hammer, N. I.; Multipole-Bound Molecular Anions In *Advances in Gas-Phase Ion Chemistry, Volume 4*; Adams, N.; Babcock, L., Eds.; Elsevier, New York, 2001; pp. 257-291.

[30] Sommerfeld, T.; Dreux, K. M.; Joshi, R. Excess Electrons Bound to Molecular Systems with a Vanishing Dipole but Large Molecular Quadrupole *J. Phys. Chem. A* **2014**, 118, 7320-7329, DOI: 10.1021/jp411787w

[31] Sommerfeld, T.; Bhattarai, B.; Vysotskiy, V. P.; Cederbaum, L. S. Correlation-Bound Anions of NaCl Clusters *J. Chem. Phys.* **2010**, 133, 114301, DOI: 10.1063/1.3488228.

[32] Bezchastnov, V. G.; Vysotskiy, V. P.; Cederbaum, L. S. Anions of Xenon Clusters Bound by Long-Range Electron Correlations *Phys. Rev. Lett.* **2011**, 107, 133401, DOI: 10.1103/PhysRevLett.107.133401.

[33] Voora, V. K.; Cederbaum, L. S.; Jordan, K. D. Existence of a Correlation Bound s-Type Anion State of C_{60} *J. Phys. Chem. Lett.* **2013**, 4, 849-853, DOI: 10.1021/jz400195s.

[34] Voora, V. K.; Jordan, K. D. Nonvalence Correlation-Bound Anion State of C_6F_6: Doorway to Low-Energy Electron Capture *J. Phys. Chem. A* **2014**, 118, 7201-7205, DOI: 10.1021/jp408386f.

[35] Voora, V. K.; Jordan, K. D. Nonvalence Correlation-Bound Anion States of Polycyclic Aromatic Hydrocarbons *J. Phys. Chem. Lett.* **2015**, 6, 3994-3997, DOI: 10.1021/acs.jpclett.5b01858.

[36] Bull, J. N.; Verlet, J. R. R. Observation and Ultrafast Dynamics of a Nonvalence Correlation-Bound State of an Anion *Sci. Adv.* **2017**, 3, e1603106, DOI: 10.1126/sciadv.1603106.

[37] Fortenberry, R. C.; Morgan, W. J.; Enyard, J. D. Predictable Valence Excited States of Anions *J. Phys. Chem. A* **2014**, 118, 10763-10769, DOI: 10.1021/jp509512u.

[38] Stanton, J. F.; Gauss, J. A Simple Scheme for the Direct Calculation of Ionization Potentials with Coupled-Cluster Theory that Exploits Established Excitation Energy Methods *J. Chem. Phys.* **1999**, 111, 8785-8788, DOI: 10.1063/1.479673.

[39] Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A Fifth-Order Perturbation Comparison of Electron Correlation Energies *Chem. Phys. Lett.* **1989**, 157, 479-483, DOI: 10.1016/j.cplett.2013.08.064.

[40] Neese, F. The ORCA Program System *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, 2, 73-78, DOI: 10.1002/wcms.81.

[41] Bonaccorsi, R.; Pullman, A.; Scrocco, E.; Tomasi, J. The Molecular Electrostatic Potentials for the Nucleic Acid Bases: Adenine, Thymine, and Cytosine *Theoret. Chim. Acta* **1972**, 24, 51-60, DOI: 10.1007/BF00528310.

[42] Scrocco E., Tomasi J. The Electrostatic Molecular Potential as a Tool for the Interpretation of Molecular Properties In *Topics in Current Chemistry Fortschritte der Chemischen Forschung book series (TOPCURRCHEM, volume 42)*; Springer, Berlin, Heidelberg, 1973; pp 95-170.

[43] Politzer, P.; Laurence, P. R.; Jayasuriya, K. Molecular Electrostatic Potentials: An Effective Tool for the Elucidation of Biochemical Phenomena *Environmental Health Perspectives* **1985**, 61, 191-202, DOI: 10.2307/3430072.

[44] Orozco, M.; Luque, F. J. Generalization of the Molecular Electrostatic Potential for the Study of Noncovalent interactions In *Molecular Electrostatic Potentials: Concepts and Applications*; Murray, J. S.; Sen, K., Eds.; Elsevier; Amsterdam, 1996; pp 181-218. DOI: 10.1016/S1380-7323(96)80044-6.

[45] Handy, N. C.; Schaefer, H. F. On The Evaluation of Analytic Energy Derivatives for Correlated Wave Functions *J. Chem. Phys.* **1984**, 81, 5031-5033, DOI: 10.1063/1.447489.

[46] Wiberg, K. B.; Hadad, C. M.; LePage, T.; Breneman, C. M.; Frisch, M. J. Analysis of the Effect of Electron Correlation on Charge Density Distributions *J. Phys. Chem.* **1992**, 96, 671-679, DOI: 10.1021/j100181a030.

[47] Tajti, A.; Szalay, P. G. Investigation of the Impact of Different Terms in the Second Order Hamiltonian on Excitation Energies of Valence and Rydberg States *J. Chem. Theory Comput.* **2016**, 5477-5482, DOI: 10.1021/acs.jctc.6b00723.

[48] Szalay, P. G.; Watson, T.; Perera, A.; Lotrich, V.; Bartlett, R. J. Benchmark Studies on the Building Blocks of DNA. 3. Watson-Crick and Stacked Base Pairs *J. Phys. Chem. A* **2013**, 117, 3149, DOI: 10.1021/jp3100975.

[49] Aquino, A. J. A.; Borges Jr, I.; Nieman, R.; Köhn, A.; Lischka, H. Intermolecular Interactions and Charge Transfer Transitions in Aromatic Hydrocarbon-Tetracyanoethylene Complexes *Phys. Chem. Chem. Phys.* **2014**, 16, 20586-20597, DOI: 10.1039/C4CP02900C.

[50] Haase, F.; Ahlrichs, R. Semi-direct MP2 Gradient Evaluation on Workstation Computers: The MPGRAD Program *J. Comp. Chem.*, **1993**, 14, 907-912, DOI: 10.1002/jcc.540140805.

[51] Weigend, F.; Häser, M. RI-MP2: First Derivatives and Global Consistency *Theor. Chem. Acc.*, **1997**, 97, 331-340, DOI: 10.1007/s002140050269.

[52] Hättig, C.; Weigend, F. CC2 Excitation Energy Calculations on Large Molecules Using the Resolution of the Identity Approximation *J. Chem. Phys.*, **2000**, 113, 5154-5161, DOI: 10.1063/1.1290013.

[53] Köhn, A.; Hättig, C. Analytic Gradients for Excited States in the Coupled-Cluster Model CC2 Employing the Resolution-of-the-Identity Approximation *J. Chem. Phys.*, **2003**, 119, 5021-5036, DOI: 10.1063/1.1597635.

[54] TURBOMOLE V6.3.1 2011, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.

[55] Weigend, F.; Köhn, A.; Hättig, C. Efficient Use of the Correlation Consistent Basis Sets in Resolution of the Identity MP2 Calculations *J. Chem. Phys.*, **2002**, 116, 3175-3183, DOI: 10.1063/1.1445115.

[56] Guallar, V.; Douhal, A.; Moreno, M.; Lluch, J. M. DNA Mutations Induced by Proton and Charge Transfer in the Low-Lying Excited Singlet Electronic States of the DNA Base Pairs: A Theoretical Insight *J. Phys. Chem. A* **1999**, 103, 6251-6256, DOI: 10.1021/jp9908496.

[57] Sobolewski, A. L.; Domcke, W. *Ab Initio* Studies on the Photophysics of the Guanine-Cytosine Base Pair *Phys. Chem. Chem. Phys.* **2004**, 6, 2763-2771, DOI: 10.1039/B314419D.

[58] Abo-Riziq, A.; Grace, L.; Nir, E.; Kabelac, M.; Hobza, P.; de Vries, M. S. Photochemical Selectivity in Guanine-Cytosine Base-Pair Structures *PNAS* **2004**, 102, 20-23, DOI: 10.1073/pnas.0408574102.

[59] Sobolewski, A. L.; Domcke, W.; Hättig, C. Tautomeric Selectivity of the Excited-State Lifetime of Guanine/Cytosine Base Pairs: the Role of Electron-Driven Proton-Transfer Processes *PNAS* **2005**, 102, 17903-17906, DOI: 10.1073/pnas.0504087102.

[60] Perun, S.; Sobolewski, A. L.; Domcke, W. Role of Electron-Driven Proton-Transfer Processes in the Excited-State Deactivation of the Adenine-Thymine Base Pair *J. Phys. Chem. A* **2006**, 110, 9031-9038, DOI: 10.1021/jp061945r.

[61] Ciminelli, C.; Granucci, G.; Persico, M. The Photoisomerization Mechanism of Azobenzene: A Semiclassical Simulation of Non-adiabatic Dynamics *Chem. - Eur. J.* **2004**, 10, 2327-2341, DOI: 10.1002/chem.200305415.

[62] Röttger, K.; Marroux, H. J. B.; Grubb, M. P.; Coulter, P. M.; Böhnke, H.; Henderson, A. S.; Galan, M. C.; Temps, F.; Orr-Ewing, A. J.; Roberts, G. M. Ultraviolet Absorption Induces Hydrogen-Atom Transfer in G·C Watson-Crick DNA Base Pairs in Solution *Angew. Chem. Int. Ed.* **2015**, 54, 14719-14722, DOI: 10.1002/anie.201506940.

[63] Markwick, P. R. L.; Doltsinis, N. L.; Schlitter, J. Probing Irradiation Induced DNA Damage Mechanisms Using Excited State Car-Parrinello Molecular Dynamics *J. Chem. Phys.* **2007**, 126, 045104, DOI: 10.1063/1.2431177.

[64] Markwick, P. R. L.; Doltsinis, N. L. Ultrafast Repair of Irradiated DNA: Nonadiabatic *Ab Initio* Simulations Of The Guanine-Cytosine Photocycle *J. Chem. Phys.* **2007**, 126, 175102, DOI: 10.1063/1.2728897.

[65] Groenhof, G.; Schäfer, L. V.; Boggio-Pasqua, M.; Goette, M.; Grubmüller, H.; Robb, M. A. Ultrafast Deactivation of an Excited Cytosine-Guanine Base Pair in DNA *J. Am. Chem. Soc.* **2007**, 129, 6812-6819, DOI: 10.1021/ja069176c.

[66] Schwalb, N. K.; Michalak, T.; Temps, F. Ultrashort Fluorescence Lifetimes of Hydrogen-Bonded Base Pairs of Guanosine and Cytidine in Solution *J. Phys. Chem. B* **2009**, 113, 16365-16376, DOI: 10.1021/jp904883n.

[67] de La Harpe, K.; Crespo-Hernández, C. E.; Kohler, B. Excited-State Lifetimes in a G·C DNA Duplex are nearly Independent of Helix Conformation and Base Pairing Motif *Chemphyschem.* **2009**, 10, 1421-1425, DOI: 10.1002/cphc.200900004.

[68] de La Harpe, K.; Crespo-Hernández, C. E.; Kohler, B. Deuterium Isotope Effect on Excited-State Dynamics in an Alternating GC Oligonucleotide *J. Am. Chem. Soc.* **2009**, 131, 17557-17559, DOI: 10.1021/ja9076364

[69] Biemann, L.; Kovalenko, S. A.; Kleinermanns, K.; Mahrwald, R.; Markert, M.; Improta, R. Excited State Proton Transfer Is Not Involved in the Ultrafast Deactivation of Guanine-Cytosine Pair in Solution *J. Am. Chem. Soc.* **2011**, 133, 19664-19667, DOI: 10.1021/ja2089734.

[70] Yamazaki, S.; Taketsugu, T. Photoreaction Channels of the Guanine-cytosine Base Pair Explored By Long-Range Corrected TDDFT Calculations *Phys. Chem. Chem. Phys.* **2012**, 14, 8866-8877, DOI: 10.1039/c2cp23867e.

[71] Dargiewicz, M.; Biczysko, M.; Improta, R.; Barone, V. Solvent Effects on Electron-Driven Proton-Transfer Processes: Adenine-Thymine Base Pairs *Phys. Chem. Chem. Phys.* **2012**, 14, 8981-8989, DOI: 10.1039/c2cp23890j.

[72] Bucher, D. B.; Schlueter, A.; Carell, T.; Zinth, W. Watson-Crick Base Pairing Controls Excited-State Decay in Natural DNA *Angew. Chem. Int. Ed.* **2014**, 53, 11366-11369, DOI: 10.1002/anie.201406286.

[73] Röttger, K.; Temps, F. Detection of the G(–H)˙ Radical in the Electronic Deactivation of the G-C Watson-Crick Base Pair In *19th International Conference on Ultrafast Phenomena, OSA Technical Digest* (online), Optical Society of America, 2014, paper 07.Mon.P1.20, DOI: 10.1364/UP.2014.07.Mon.P1.20.

[74] Zhang, Y.; de La Harpe, K.; Beckstead, A. A.; Improta, R.; Kohler, B. UV-Induced Proton Transfer between DNA Strands *J. Am. Chem. Soc.* **2015**, 137, 7059-7062, DOI: 10.1021/jacs.5b03914.

[75] Röttger, K.; Marroux, H. J. B.; Böhnke, H.; Morris, D. T. J.; Voice, A. T.; Temps, F.; Roberts, G. M.; Orr-Ewing, A. J. Probing the Excited State Relaxation Dynamics of Pyrimidine Nucleosides in Chloroform Solution *Faraday Discuss.* **2016**, 194, 683-708, DOI: 10.1039/C6FD00068A.

[76] Li, Q.; Giussani, A.; Segarra-Martí, J.; Nenov, A.; Rivalta, I.; Voityuk, A. A.; Mukamel, S.; Roca-Sanjuán, D.; Garavelli, M,.; Blancafort, L. Multiple Decay Mechanisms and 2D-UV Spectroscopic Fingerprints of Singlet Excited Solvated Adenine-Uracil Monophosphate *Chem. Eur. J.* **2016**, 22, 7497-7507, DOI: 10.1002/chem.201505086.

[77] Röttger, K.; Marroux, H. J. B.; Chemin, A. F. M.; Elsdon, E.; Oliver, T. A. A.; Street, S. T. G.; Henderson, A. S.; Galan, M. C.; Orr-Ewing, A. J.; Roberts, G. M. Is UV-Induced Electron-Driven Proton Transfer Active in a Chemically Modified A·T DNA Base Pair? *J. Phys. Chem. B* **2017**, 121, 4448-4455, DOI: 10.1021/acs.jpcb.7b02679.

[78] Jie, J.; Liu, K.; Wu, L.; Zhao, H.; Song, D.; Su, H. Capturing the Radical Ion-Pair Intermediate in DNA Guanine Oxidation *Sci. Adv.* **2017**, 3, e1700171, DOI 10.1126/sciadv.1700171.

[79] Frank, I.; Hutter, J.; Marx, D.; Parrinello, M. Molecular Dynamics in Low-Spin Excited States *J. Chem. Phys.* **1998**, 108, 4060-4069, DOI: 10.1063/1.475804.

[80] Le Bahers, T.; Adamo, C.; Ciofini, I. A Qualitative Index of Spatial Extent in Charge-Transfer Excitations *J. Chem. Theory Comput.* **2011**, 7, 2498-2506, DOI: 10.1021/ct200308m.