

# Electronic Supplementary Information (ESI) for: UV-induced hydrogen transfer in DNA base pairs promoted by dark $n\pi^*$ states<sup>†</sup>

Kinga E. Szkaradek,<sup>a</sup> Petr Stadlbauer,<sup>b,c</sup> Jiří Šponer,<sup>b,c</sup> Robert W. Góra<sup>\*,a</sup> and Rafał Szabla<sup>\*,c,d</sup>

October 13, 2019

## 1 Computational methods

All investigated closed-shell systems were optimized in the ground state, assuming planar ( $C_s$ ) symmetry, using the second order Møller-Plesset perturbation theory (MP2)<sup>1</sup> and the correlation-consistent cc-pVTZ basis set.<sup>2</sup> Calculations of the 20 lowest-lying vertical excitations in  $C_s$  symmetry were performed in Turbomole 7.3,<sup>3</sup> using the algebraic diagrammatic construction to the second order method [ADC(2)]<sup>4–6</sup>, as well as the second-order approximate coupled-cluster model (CC2)<sup>7</sup> with the cc-pVTZ basis set. The resolution-of-the-identity (RI) approximation was employed in calculations of the electron repulsion integrals. The outlined levels of theory are further denoted as ADC(2)/cc-pVTZ and CC2/cc-pVTZ, respectively. Solvation effects exerted by bulk solvents were mimicked with the Conductor-like Screening Model (COSMO).<sup>8</sup> The vertical excitation energies in solution were calculated on top of the  $S_0$  geometries (optimized with the MP2/cc-pVTZ method in the gas phase) assuming the non-equilibrium COSMO post-SCF solvation model implemented in the TURBOMOLE 7.3 package<sup>9</sup> with the slow part of the reaction field equilibrated for the  $S_0$  state. In addition, ADC(2)-COSMO energies of the  $S_1$  minima were calculated with the equilibrium approach, i.e. with the reaction field equilibrated for the  $S_1$  state. The solvent with dielectric constant  $\epsilon=78.34$  and  $\epsilon=4.80$  were taken into consideration, representing polar (water) and apolar (chloroform) solvents, respectively. The Natural Transition Orbitals (NTOs)<sup>10</sup> analysis was performed in order to determine character of electronic transitions. The fraction of electron transferred between the interacting nucleobases was estimated based on electron-hole analysis implemented in the TheoDore 1.5.1 package.<sup>11–13</sup> The CT numbers presented in the main article correspond to the  $\Omega_{AB}$  values, i.e. charge transfer from fragment A (purine base) to B (pyrimidine base), and were computed based on the default Mulliken type analysis.

The potential energy surface (PES) cross-sections were constructed for the ground and low-lying excited singlet states by linear interpolation in internal coordinates (LIIC) between the key stationary points (FC region,  $S_1$  minima and state crossings), to illustrate the nonradiative deactivation pathway. Stationary points on excited-state potential energy surfaces were located using ADC(2)/cc-pVTZ method. Minimum-energy conical intersection (MECI) points were located employing the sequential penalty constrained optimization implemented by Levine et al. in the CIOpt package.<sup>14</sup> These MECIs were optimized using the energies and analytical energy gradients obtained at the ADC(2)/MP2 level and the cc-pVTZ basis set. For this purpose we interfaced the CIOpt package<sup>14</sup> with the TURBOMOLE 7.3 program.<sup>3</sup>

Benchmark XMS-CASPT2/SA-2-CASSCF(6,6)/cc-pVDZ calculations were performed to validate the accuracy of the ADC(2) method. In this case the complete active space for the GC base pair consisted of 6 electrons correlated in 6 molecular orbitals (2 occupied  $\pi$ , 1 occupied  $n$  and 3 virtual  $\pi^*$  orbitals). All the XMS-CASPT2 calculations were performed using the BAGEL<sup>15</sup> software.

**Classical MD simulations.** The B-form of the DNA [d(GGGCCC)]<sub>2</sub> and d(GGGCCC).(GGHCCC) double helices was built using the Avogadro software. The molecules were solvated in a cubic box of

---

SPC/E waters with the minimum distance of 15 Å between solute and the periodic box border.<sup>16</sup> The system was either neutralized by addition of Na<sup>+</sup> cations or neutralized by K<sup>+</sup> with excess 0.15M KCl.<sup>17</sup> The topology was constructed using the parameters from the parmOL15 DNA force field.<sup>18–22</sup> All the MD simulations described here were performed with the AMBER16 program.<sup>23</sup>

The system was subjected to equilibration with the following protocol:

500 steps of steepest descent minimization was followed by 500 steps of conjugate gradient minimization with 25.0 kcal · mol<sup>-1</sup> · Å<sup>-2</sup> position restraints imposed on the DNA atoms. The system was then heated from 0 to 300 K during preliminary dynamics lasting 100 ps with position restraints of 25.0 kcal · mol<sup>-1</sup> · Å<sup>-2</sup> on the DNA fragment and with constant volume. Afterwards, the geometry of the system was minimized with 5.0 kcal · mol<sup>-1</sup> · Å<sup>-2</sup> restraints on the DNA atoms for 500 steps with the steepest descent method followed by 500 steps using the conjugate gradient approach. Continuing in position restraints on DNA atoms of 5 kcal · mol<sup>-1</sup> · Å<sup>-2</sup>, Afterwards, the system was equilibrated for 50 ps maintaining the position restraints of 5.0 kcal · mol<sup>-1</sup> · Å<sup>-2</sup> on the DNA atoms, and keeping a constant temperature and pressure of 300 K and 1 atm., respectively. An analogous series of alternating minimisations and equilibrations was performed, consecutively using decreasing position restraints of 4.0, 3.0, 2.0 and 1.0 kcal · mol<sup>-1</sup> · Å<sup>-2</sup>. Finally, equilibration using position restraints of 0.5 kcal · mol<sup>-1</sup> · Å<sup>-2</sup> and starting velocities from the previous equilibration step, followed by a short 50 ps unrestrained MD run was applied.

After equilibration, a production run was commenced. Constant temperature and pressure (300 K and 1 atm.) were maintained using the Berendsen weak-coupling thermostat and barostat, respectively.<sup>24</sup> A time step of 4 fs, using SHAKE,<sup>25</sup> SETTLE<sup>26</sup> and hydrogen mass repartitioning was applied.<sup>27</sup> Each system was run for 10 ns. The average structure was calculated by taking a frame every 10 ps from the whole simulation, resulting in 1000 frames. The average structure of [d(GGGCCC)]<sub>2</sub> in Na<sup>+</sup> was subjected to QM calculations. This structure and the KCl average structure, as well as the G9H mutant average structures from Na<sup>+</sup> and KCl simulations were used as reference points for subsequent MD simulations of d(GGGTCC).(GGGCCC) and d(GGGTCC).(GGHCCC), respectively (see below).

**MD simulations of the d(GGGTCC).(GGGCCC) and d(GGGTCC).(GGHCCC) molecules.** To create the starting structures, we took a frame every 10 ps from the [d(GGGCCC)]<sub>2</sub> and d(GGGCCC).(GGHCCC) simulations, respectively. Average structures in Na<sup>+</sup> and KCl simulations were calculated separately. We then identified three frames with the lowest RMSD (after RMSD fitting) to the average structure for both the DNA sequences in both Na<sup>+</sup> and KCl, therefore we obtained six [d(GGGCCC)]<sub>2</sub> and six d(GGGCCC).(GGHCCC) structures. These frames were extracted from the trajectory including the solvation box. To obtain d(GGGTCC).(GGGCCC), we performed either a C4T or C10T point mutation of [d(GGGCCC)]<sub>2</sub>, keeping T in the same position as C; thus we obtained 12 different, albeit very similar starting structures by this procedure, which was done to ensure non-identical progress in the following MD simulations. Analogously, d(GGGTCC).(GGHCCC) was prepared from d(GGGCCC).(GGHCCC) by a C4T mutation, yielding 6 different starting structures. Each structure was then simulated without prior equilibration at 300 K and 1 atm., with a time step of 2 fs using SHAKE and SETTLE, for 10 ns, followed by 90 ns with a time step of 4 fs and hydrogen mass repartitioning. All systems were then run once more using the recent gHBfix addition to the parmOL15 force field.<sup>28</sup> Hence, we obtained 24 simulations of d(GGGTCC).(GGGCCC) (total simulation time 2400 ns) and 12 simulations of d(GGGTCC).(GGHCCC) (1200 ns in total). gHBfix is an additional potential meant to strengthen H-bonds, because they are known to be under-stabilized in the current parmOL15 DNA force field.<sup>28</sup> The potential is applied to all possible pairs of N–H donor (amino and imino hydrogen) and acceptor oxygen and nitrogen atoms of all the nucleobases. We set that each such pair is stabilized by 0.5 kcal · mol<sup>-1</sup> if the distance between the hydrogen and its

hydrogen bond acceptor is shorter than 2.0 Å. The stabilisation diminishes from 2.0 to 3.0 Å, and at longer distances, the added potential is equal to 0 kcal · mol<sup>-1</sup>. This functional form ensures that only H-bonds that are actually present in the system are stabilised, and those pairs of atoms above the 3.0 Å distance threshold are not affected at all.

## 2 Results

### 2.1 Vertical excitation energies

Vertical electronic excitation spectra of GC and HC were computed assuming the respective ground-state minimum energy structures optimized assuming C<sub>s</sub> point-group symmetry, corresponding to planar geometries of the base pairs. The ADC(2)/cc-pVTZ results discussed in the main article are compared to CC2/cc-pVTZ data. We follow the standard notation of excited states of organic molecules assigned based on analysis of the Natural Transition Orbitals.

**Table 1** Vertical excitation energies (in eV) in Franck–Condon region of guanine-cytosine (G-C) and hypoxanthine-cytosine (H-C) base pairs calculated with imposed C<sub>s</sub> symmetry.

(a) The considered systems were computed using the ADC(2)/cc-pVTZ method, assuming the ground-state minimum energy structures optimized at the MP2/cc-pVTZ level.

(b) The considered systems were computed using the CC2/cc-pVTZ method, assuming the ground-state minimum energy structures optimized at the CC2/cc-pVTZ level.

State / Transition	E <sub>exc</sub> [eV]	f <sub>osc</sub>	λ/[nm]
<i>GC ADC(2) C<sub>s</sub> symmetry</i>			
S <sub>1</sub> (A') ππ* / πGπ <sub>C</sub> <sup>*</sup> (CT)	4.857	6.99 · 10 <sup>-2</sup>	255.3
S <sub>2</sub> (A') ππ*	4.906	5.65 · 10 <sup>-2</sup>	252.7
S <sub>3</sub> (A') πGπ <sub>C</sub> <sup>*</sup> (CT) / ππ*	5.164	2.84 · 10 <sup>-2</sup>	240.1
S <sub>4</sub> (A'') nπ* / nGπ <sub>C</sub> <sup>*</sup> (CT)	5.369	6.45 · 10 <sup>-4</sup>	230.9
S <sub>5</sub> (A') ππ*	5.370	0.234	230.9
S <sub>6</sub> (A') ππ*	5.416	0.407	228.9
S <sub>7</sub> (A'') nπ*	5.583	2.12 · 10 <sup>-4</sup>	222.1
S <sub>8</sub> (A'') nπ*	5.684	5.07 · 10 <sup>-5</sup>	218.1
S <sub>9</sub> (A'') nπ*	6.171	1.60 · 10 <sup>-6</sup>	200.9
S <sub>10</sub> (A'') nπ* / nGπ <sub>C</sub> <sup>*</sup> (CT)	6.312	1.38 · 10 <sup>-3</sup>	196.4
S <sub>11</sub> (A'') nπ* / nGπ <sub>C</sub> <sup>*</sup> (CT)	6.343	9.76 · 10 <sup>-5</sup>	195.5
S <sub>12</sub> (A'') πσ*	6.434	8.80 · 10 <sup>-4</sup>	192.7
<i>HC ADC(2) C<sub>s</sub> symmetry</i>			
S <sub>1</sub> (A') ππ*	4.809	7.60 · 10 <sup>-2</sup>	257.8
S <sub>2</sub> (A') ππ*	5.044	8.70 · 10 <sup>-2</sup>	245.8
S <sub>3</sub> (A'') nπ* / n <sub>H</sub> π <sub>C</sub> <sup>*</sup> (CT)	5.227	2.97 · 10 <sup>-4</sup>	237.2
S <sub>4</sub> (A') ππ*	5.344	0.236	232.0
S <sub>5</sub> (A'') nπ*	5.469	4.24 · 10 <sup>-4</sup>	226.7
S <sub>6</sub> (A') ππ*	5.539	0.224	223.8
S <sub>7</sub> (A'') n <sub>H</sub> π <sub>C</sub> <sup>*</sup> (CT) / nπ*	5.582	2.49 · 10 <sup>-4</sup>	222.1
S <sub>8</sub> (A'') nπ <sub>H</sub> <sup>*</sup>	5.713	2.39 · 10 <sup>-4</sup>	217.0
S <sub>9</sub> (A'') nπ <sub>H</sub> <sup>*</sup>	5.807	5.15 · 10 <sup>-4</sup>	213.5
S <sub>10</sub> (A') π <sub>H</sub> π <sub>C</sub> <sup>*</sup> (CT)	5.912	4.31 · 10 <sup>-3</sup>	209.7
S <sub>11</sub> (A'') nπ <sub>C</sub> <sup>*</sup>	6.010	7.64 · 10 <sup>-6</sup>	206.3
S <sub>12</sub> (A'') nπ <sub>H</sub> <sup>*</sup>	6.350	0.002	195.3

State / Transition	E <sub>exc</sub> [eV]	f <sub>osc</sub>	λ/[nm]
<i>GC CC2 C<sub>s</sub> symmetry</i>			
S <sub>1</sub> (A') ππ* / πGπ <sub>C</sub> <sup>*</sup> (CT)	4.943	6.17 · 10 <sup>-2</sup>	250.8
S <sub>2</sub> (A') ππ*	5.036	6.20 · 10 <sup>-2</sup>	246.2
S <sub>3</sub> (A') πGπ <sub>C</sub> <sup>*</sup> (CT) / ππ*	5.237	2.29 · 10 <sup>-2</sup>	236.7
S <sub>4</sub> (A'') nπ* / nGπ <sub>C</sub> <sup>*</sup> (CT)	5.487	7.38 · 10 <sup>-4</sup>	226.0
S <sub>5</sub> (A') ππ*	5.518	0.509	224.7
S <sub>6</sub> (A') ππ*	5.524	9.45 · 10 <sup>-2</sup>	224.5
S <sub>7</sub> (A'') nπ*	5.768	3.39 · 10 <sup>-5</sup>	215.0
S <sub>8</sub> (A'') nπ*	5.852	7.48 · 10 <sup>-6</sup>	211.9
S <sub>9</sub> (A'') nπ*	6.356	5.72 · 10 <sup>-4</sup>	195.1
S <sub>10</sub> (A'') πσ*	6.371	1.26 · 10 <sup>-3</sup>	194.6
S <sub>11</sub> (A'') nπ*	6.420	2.85 · 10 <sup>-6</sup>	193.1
S <sub>12</sub> (A') πGπ <sub>C</sub> <sup>*</sup> (CT) / ππ*	6.496	6.10 · 10 <sup>-2</sup>	190.9
<i>HC CC2 C<sub>s</sub> symmetry</i>			
S <sub>1</sub> (A') ππ*	4.948	8.03 · 10 <sup>-2</sup>	250.6
S <sub>2</sub> (A') ππ*	5.168	8.27 · 10 <sup>-2</sup>	239.9
S <sub>3</sub> (A'') nπ*	5.354	3.60 · 10 <sup>-4</sup>	231.6
S <sub>4</sub> (A') ππ*	5.485	0.230	226.0
S <sub>5</sub> (A') nπ*	5.651	0.187	219.4
S <sub>6</sub> (A'') nπ*	5.658	7.57 · 10 <sup>-4</sup>	219.1
S <sub>7</sub> (A'') nπ*	5.741	5.59 · 10 <sup>-5</sup>	216.0
S <sub>8</sub> (A'') nπ*	5.789	3.15 · 10 <sup>-4</sup>	214.2
S <sub>9</sub> (A'') nπ*	5.911	3.85 · 10 <sup>-4</sup>	209.8
S <sub>10</sub> (A') π <sub>H</sub> π <sub>C</sub> <sup>*</sup> (CT)	6.023	4.08 · 10 <sup>-3</sup>	205.9
S <sub>11</sub> (A'') nπ*	6.259	7.84 · 10 <sup>-6</sup>	198.1
S <sub>12</sub> (A'') nπ*	6.375	1.38 · 10 <sup>-3</sup>	194.5

All the states of interests are either of the nπ\* or ππ\* character. In the latter the major contribution is due to electronic transition from the valence π molecular orbital to antibonding π\* orbital while in the former

the electronic transition occurs from the nonbonding lone-electron-pair  $n$  orbital. Unless otherwise noted, all the states have locally-excited character (LE), which means that the transition is localized on one of the bases. In the case of charge-transfer (CT) states or states having CT component, for which the electron is transferred between bases, we indicated localization of the hole and the particle orbitals in the subscript.

## 2.2 Solvation effects on the photochemistry of DNA base pairs

**Table 2** Vertical excitation energies (in eV) in Franck–Condon region of guanine-cytosine (G-C) and hypoxanthine-cytosine (H-C) base pairs calculated with Conductor-like Screening Model (COSMO) and imposed  $C_s$  symmetry.

**(a)** Calculations performed at the ADC(2)-COSMO/cc-pVTZ level considering the model of bulk water with charge distribution equilibrated for the ground state.

**(b)** Calculations performed at the ADC(2)-COSMO/cc-pVTZ level considering the model of bulk chloroform with charge distribution equilibrated for the ground state.

State / Transition	$E_{\text{exc}}[\text{eV}]$	$f_{\text{osc}}$	$\lambda/[\text{nm}]$	
<i>GC ADC(2)-COSMO (<math>\epsilon = 78.34</math>)</i>				
S <sub>1</sub> (A')	$\pi\pi^*$	4.829	0.136	256.7
S <sub>2</sub> (A')	$\pi\pi^*$	4.953	0.161	250.3
S <sub>3</sub> (A')	$\pi\pi^*$	5.390	0.475	230.0
S <sub>4</sub> (A')	$\pi_G\pi_C^*$ (CT)	5.427	$1.2 \cdot 10^{-4}$	228.5
S <sub>5</sub> (A')	$\pi\pi^*$	5.474	0.192	226.5
S <sub>6</sub> (A'')	$n\pi^*$	5.663	$7.9 \cdot 10^{-4}$	218.9
S <sub>7</sub> (A'')	$n\pi^*$	5.744	$3.4 \cdot 10^{-4}$	215.8
S <sub>8</sub> (A'')	$n\pi^*$	6.018	$1.4 \cdot 10^{-4}$	206.0
S <sub>9</sub> (A')	$\pi\pi^*$	6.309	0.274	196.5
S <sub>10</sub> (A'')	$n\pi^*$	6.379	$4.6 \cdot 10^{-5}$	194.4
S <sub>11</sub> (A'')	$n\pi^*$	6.409	$7.5 \cdot 10^{-4}$	193.5
S <sub>12</sub> (A')	$\pi\pi^*$	6.511	$2.5 \cdot 10^{-2}$	190.4
<i>HC ADC(2)-COSMO (<math>\epsilon = 78.34</math>)</i>				
S <sub>1</sub> (A')	$\pi\pi_C^*$	4.909	0.166	252.6
S <sub>2</sub> (A')	$\pi\pi_H^*$	5.066	0.105	244.7
S <sub>3</sub> (A')	$\pi\pi_H^*$	5.379	0.287	230.5
S <sub>4</sub> (A')	$\pi\pi_C^*$	5.510	0.222	225.0
S <sub>5</sub> (A'')	$n\pi_H^* / n\pi_C^*$	5.587	$7.0 \cdot 10^{-4}$	221.9
S <sub>6</sub> (A'')	$n\pi^*$	5.655	$3.3 \cdot 10^{-4}$	219.2
S <sub>7</sub> (A'')	$n\pi_H^*$	5.876	$1.18 \cdot 10^{-3}$	211.0
S <sub>8</sub> (A'')	$n\pi^*$	5.910	$2.6 \cdot 10^{-4}$	209.8
S <sub>9</sub> (A'')	$n\pi^*$	5.969	$3.4 \cdot 10^{-4}$	207.7
S <sub>10</sub> (A'')	$n\pi^*$	6.257	$2.9 \cdot 10^{-5}$	198.2
S <sub>11</sub> (A')	$\pi_H\pi_C^*$ (CT)	6.362	$9.4 \cdot 10^{-2}$	194.9
S <sub>12</sub> (A')	$\pi\pi_C^*$	6.377	0.286	194.4

State / Transition	$E_{\text{exc}}[\text{eV}]$	$f_{\text{osc}}$	$\lambda/[\text{nm}]$	
<i>GC ADC(2)-COSMO (<math>\epsilon = 4.80</math>)</i>				
S <sub>1</sub> (A')	$\pi\pi_G^*$	4.833	0.124	256.5
S <sub>2</sub> (A')	$\pi\pi_C^*$	4.926	0.142	251.7
S <sub>3</sub> (A')	$\pi_G\pi_C^*$ (CT)	5.197	$4.8 \cdot 10^{-4}$	238.6
S <sub>4</sub> (A')	$\pi\pi_G^*$	5.364	0.488	231.14
S <sub>5</sub> (A')	$\pi\pi_C^*$	5.386	0.212	230.2
S <sub>6</sub> (A'')	$n_G\pi_C^*$ (CT) / $n\pi_G^*$	5.586	$1.0 \cdot 10^{-3}$	222.0
S <sub>7</sub> (A'')	$n\pi_G^* / n\pi_C^*$	5.661	$1.5 \cdot 10^{-5}$	219.0
S <sub>8</sub> (A'')	$n\pi_C^*$	5.863	$1.5 \cdot 10^{-4}$	211.5
S <sub>9</sub> (A'')	$n\pi^*$	6.286	$1.4 \cdot 10^{-5}$	197.2
S <sub>10</sub> (A')	$\pi\pi_C^*$	6.357	0.281	195.0
S <sub>11</sub> (A'')	$n\pi^*$	6.368	$1.0 \cdot 10^{-3}$	194.7
S <sub>12</sub> (A')	$\pi\pi_G^*$	6.494	$5.0 \cdot 10^{-3}$	190.9
<i>HC ADC(2)-COSMO (<math>\epsilon = 4.80</math>)</i>				
S <sub>1</sub> (A')	$\pi\pi_C^*$	4.870	0.146	254.6
S <sub>2</sub> (A')	$\pi\pi_H^*$	5.047	0.109	245.7
S <sub>3</sub> (A')	$\pi\pi_H^*$	5.363	0.268	231.2
S <sub>4</sub> (A')	$\pi\pi_C^*$	5.477	0.256	226.4
S <sub>5</sub> (A'')	$n\pi^*$	5.496	$9.23 \cdot 10^{-4}$	225.6
S <sub>6</sub> (A'')	$n\pi^*$	5.571	$2.0 \cdot 10^{-7}$	222.6
S <sub>7</sub> (A'')	$n\pi_C^* / n\pi_H^*$ (CT)	5.809	$8.31 \cdot 10^{-5}$	213.4
S <sub>8</sub> (A'')	$n\pi^*$	5.824	$9.59 \cdot 10^{-4}$	212.9
S <sub>9</sub> (A'')	$n\pi^*$	5.874	$5.67 \cdot 10^{-4}$	211.1
S <sub>10</sub> (A')	$\pi_H\pi_C^*$ (CT)	5.923	$6.0 \cdot 10^{-2}$	209.3
S <sub>11</sub> (A'')	$n\pi_C^*$	6.160	$2.32 \cdot 10^{-5}$	201.3
S <sub>12</sub> (A'')	$n\pi_H^*$	6.396	$1.28 \cdot 10^{-3}$	193.8

We performed additional calculations employing conductor-like screening model (COSMO) to investigate the effects of the environment on the energies of key excited states. Vertical excitation energies calculated within the non-equilibrium model (with the charge distribution equilibrated for the electronic ground state) are presented in Table 2. We selected dielectric constants characteristic for water ( $\epsilon = 78.34$ ), as the most commonly considered polar solvent, and for an apolar solvent chloroform ( $\epsilon = 4.8$ ). Chloroform was previously suggested to be a good representative of the dielectric environment inside a DNA double helix.<sup>29–31</sup>

The vertical excitation energies obtained using the ADC(2)-COSMO approach indicate that the  $n\pi^*$  states are generally blueshifted with respect to the gas phase calculations. In particular, the blueshift of the lowest lying  $n\pi^*$  state in the **GC** base pair amounts to 0.29 eV and 0.22 eV in bulk water and chloroform, respectively. Similarly, the lowest lying  $n\pi^*$  state in the **HC** base pair are blueshifted by 0.36 eV and

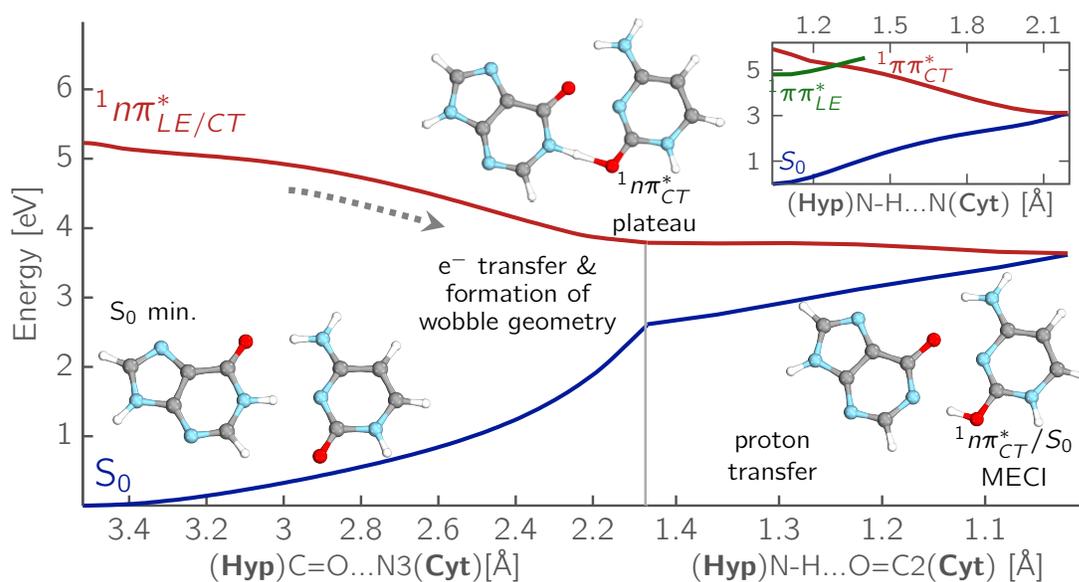
---

0.27 eV in bulk water and chloroform, respectively. Considerably lower blueshifts may be observed for the  $\pi\pi_{CT}^*$  state in the **GC** base pair, namely 0.26 eV and 0.03 eV in bulk water and chloroform, respectively. The energies of the lowest lying optically bright  $\pi\pi^*$  states are generally less affected by the continuum solvation model.

Considering that the blueshift of  $n\pi^*$  excitations in the Franck-Condon region might have an effect on their contribution to the photochemistry and photodeactivation mechanisms, we also performed equilibrium single point ADC(2)-COSMO energy calculations on the  $S_1(n\pi_{CT}^*)$  minimum-energy geometries derived from the gas phase (see Fig. 2 and 4 in the main article presenting the  $S_1$  minimum energy geometries). As discussed in the main article, the  $S_1$ - $S_0$  energy gaps ( $\Delta E_{S_1-S_0}$ ) calculated at the ADC(2)/cc-pVTZ level in the gas phase amount to 0.50 eV and 0.54 eV for **GC** and **HC**, respectively. The corresponding  $\Delta E_{S_1-S_0}$  calculated in bulk chloroform amounts to 0.77 eV for both studied base pairs, while analogous calculations performed in bulk water yield the  $\Delta E_{S_1-S_0}$  of 0.86 eV for **GC** and 0.82 eV for **HC**. These results demonstrate that the solvatochromic shifts of 0.2 to 0.4 eV are retained even outside the Franck-Condon region, and the photodeactivation mechanisms discussed based on the ADC(2) simulations performed in the gas phase remain qualitatively valid in polar and apolar environments represented by the COSMO model. It is worth noting that this blueshift does not change the order of the  $n\pi_{CT}^*$  states in the near proximity to their corresponding minima. The  $S_1(n\pi_{CT}^*)$  minima are energetically well separated from higher energy electronic excitations, *i.e.* by at least 1.4 eV. This indicates that the  $S_1(n\pi_{CT}^*)$  minima may also be populated in the solution phase and that the sequential EDPT mechanism triggered by the  $n\pi_{CT}^*$  may operate in the rather apolar environment typical for the interior of a DNA double helix represented here by the COSMO model of chloroform.

It is worth noting that the implicit solvation model is not capable of reproducing all of the environmental effects exerted by the interior of DNA double helix. Therefore, to provide a more comprehensive description of the effects of the surroundings on the two-stage EDPT process, it is essential to include stacking interactions with the neighbouring bases in future studies. However, such calculations performed withing the QM/MM framework would require the inclusion of 6 nucleobases in the QM region to properly account for stacking (the base pair participating in EDPT and two adjacent base pairs). These simulations would require substantial computational effort considering that currently the largest QM/MM calculation employing the ADC(2) approach included only 4 nucleobases.<sup>32</sup> The inclusion of the sugar-phosphate backbone would be even more demanding computationally. It is worth mentioning though, that our preliminary simulations performed on several different nucleobases and nucleosides indicate that sugar substitution stabilizes  $n\pi^*$  states. Therefore, we anticipate that the EDPT process occurring on the  $S_1(n\pi_{CT}^*)$  surface should be available in a DNA double helix.

### 2.3 Electron-driven proton transfer in HC involving the $n\pi_{CT}^*$ state and formation of the wobble exciplex



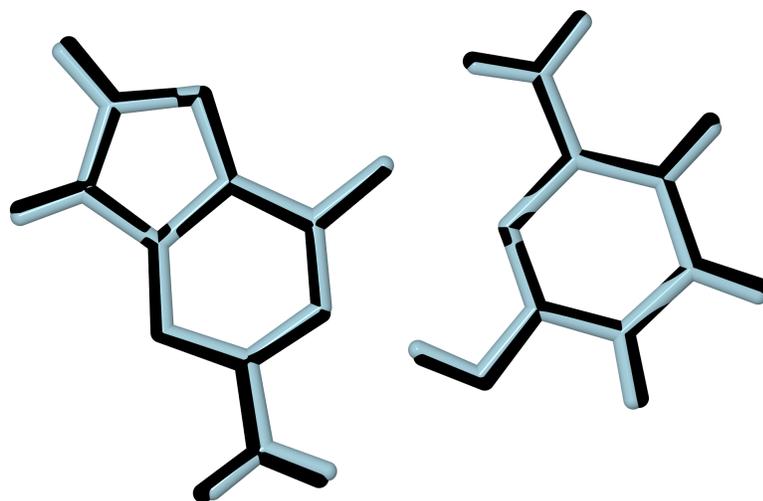
**Fig. 1** Potential energy surface cuts along the LIIC coordinate: formation of wobble-paired HC and the transfer of proton. The inset presents the comparable PES referring to WC-type HC geometry

Potential energy profile along the proton transfer coordinate calculated for the hypoxanthine-cytosine (H-C) wobble base pair is shown in Figure 1. To construct this profile we performed a linear interpolation in internal coordinates (LIIC), between three stationary points, namely the  $S_0$  minimum-energy structure, the  $S_1(n\pi_{CT}^*)$  plateau and the  $n\pi_{CT}^*/S_0$  MECI. Since geometry optimization performed on the  $S_1(n\pi_{CT}^*)$  hyper-surface, first leads to the formation of the wobble **HC** geometry and then to the subsequent proton transfer, we anticipate that the associated EDPT process is sequential. To optimize a geometry that corresponds to the **HC** wobble interaction in the  $n\pi_{CT}^*$  state (formed before the actual proton transfer) we kept the N3–H bond frozen during the optimization procedure. This nonradiative deactivation mechanism is virtually identical to the analogous presented in Fig. 3 and discussed in the main article. The hole-electron population analysis revealed that in the Franck–Condon region only 0.08 electron is transferred between the bases in the  $^1n\pi_{CT}^*$  excited state. This is due to a large admixture of locally excited (LE) configuration. However, in the region of  $n\pi_{CT}^*$  plateau, the fraction of transferred electron grows to 0.46 indicating the charge-transfer character of this state.

In the inset we show the potential energy profile for EDPT deactivation channel mediated by the  $\pi\pi_{CT}^*$  state which could occur in the canonical structure of the Watson–Crick **HC** base pair. In this case, the  $\pi\pi_{CT}^*$  state connects the optically accessible locally-excited (LE) bright  $\pi\pi^*$  state with the electronic ground state via two conical intersections: LE/CT and CT/ $S_0$ . Although the latter MECI in the Watson–Crick **HC** is lying 0.52 eV below the above discussed  $n\pi_{CT}^*/S_0$  MECI in the Wobble **HC** exciplex, the energy of  $\pi\pi_{CT}^*$  state in the Franck–Condon region is clearly higher (by about 0.69 eV) than that of the  $n\pi_{CT}^*$  state. This indicates that the  $\pi\pi_{CT}^*$  state might be available at much higher excitation energies than in **GC**, while the EDPT process on the dark  $n\pi_{CT}^*$  surface involving the formation of the wobble **HC** exciplex may be accessed in a barrierless manner upon the population of the  $n\pi_{LE/CT}^*$  state.

---

## 2.4 Benchmark calculations for the EDPT process in the $n\pi_{CT}^*$ state of GC



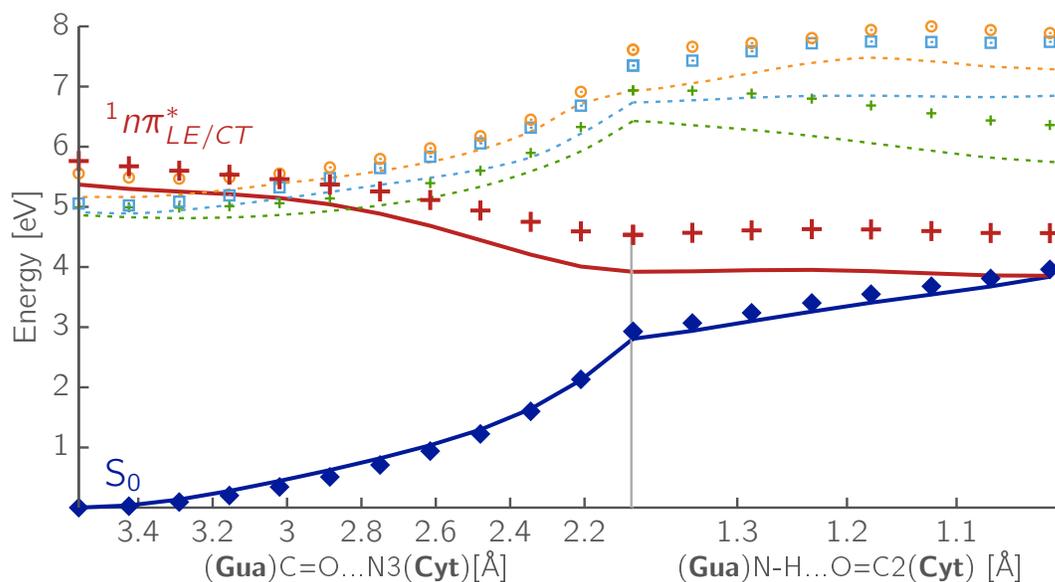
**Fig. 2** Comparison of the  $S_1(n\pi_{CT}^*)/S_0$  state crossing geometries optimized at the ADC(2)/MP2 (light blue) and CASPT2/SA-2-CASSCF(6,6) (black) levels of theory.

Minimum energy geometries of the  $S_1(n\pi_{CT}^*)/S_0$  conical intersection obtained at the ADC(2)/MP2 and XMS-CASPT2/SA-2-CASSCF(6,6) levels are aligned and presented in Fig. 2. This result clearly demonstrates that the state crossing geometry obtained with the more approximate ADC(2)/MP2 approach is highly consistent with the XMS-CASPT2 geometry. In fact, the only visible (but still insignificant) differences are related to the N $\cdots$ O and N–H distances selected as the key molecular coordinates for presenting the potential energy (PE) surface in Fig. 3 of the main article.

We also performed single-point energy calculations at the XMS-CASPT2/SA-2-CASSCF(6,6) level on the key stationary points considered in the preparation of the PE profile in Fig. 3 of the main article. In particular, CASPT2 calculations performed on the  $S_1(n\pi_{CT}^*)/S_0$  conical intersection geometry optimized at the ADC(2)/MP2 level indicated that the two states are separated by 0.80 eV. Nevertheless, we emphasize that only slight change of the key distances discussed above is sufficient to lead to the  $S_1/S_0$  degeneracy in the XMS-CASPT2 results. The vertical  $S_1$ - $S_0$  energy gap calculated on the ADC(2)-optimized  $S_1(n\pi_{CT}^*)$  plateau geometry amounts to 1.55 eV and 1.11 eV, at the XMS-CASPT2 and ADC(2)/MP2 levels, respectively. Consistently with the ADC(2) results the XMS-CASPT2 method also reproduces the flat topography of the  $S_1$  PE surface upon the formation of the wobble  $n\pi_{CT}^*$  exciplex in GC. Considering the excellent qualitative agreement between the ADC(2) and XMS-CASPT2 results we infer that ADC(2) method accurately describes the studied phenomenon, *i.e.* the formation of the wobble  $n\pi_{CT}^*$  exciplex, and the associated photodeactivation mechanism.

The recent benchmark calculations indicated that the spin-component-scaled (SCS) variants of the CC2 method largely eliminate the biases of this method for states having charge-transfer, Rydberg and/or  $n\pi^*$  character, resulting in underestimation of their excitation energies.<sup>9</sup> Since the ADC(2) method may suffer from similar problems we performed the SCS-ADC(2)/cc-pVTZ PE profile calculations for GC base pair. The results shown in Fig. 3 indicate that even though the ADC(2) method systematically underestimates

the energy of the  $n\pi_{CT}^*$  state, the ordering of all the low-lying states does not change when the SCS-ADC(2) variant is used and the PE surface cuts imply identical conclusions. Also the geometry of  $n\pi_{CT}^*/S_0$  MECI located using SCS-ADC(2) method is very similar to the ones obtained with ADC(2) or XMS-CASPT2 methods.



**Fig. 3** Potential energy surface cuts along the LIIC coordinate: formation of wobble-paired GC and the transfer of proton. A comparison of ADC(2)/cc-pVTZ, plotted with lines, and SCS-ADC(2)/cc-pVTZ results, which are plotted using symbols having the same colour as the corresponding state in the ADC(2) results. The SCS-ADC(2) energies were computed at the same geometries as the ADC(2) ones

## 2.5 MD simulations of the d(GGGTCC).(GGGCCC) initiated from the [d(GGGCCC)]<sub>2</sub> structure.

**Table 3** MD simulations of d(GGGTCC).(GGGCCC)

Mutation	Extraction time (ps) <sup>1</sup>	Force field	Ionic conditions	Transition time to GT wobble (ps) <sup>2</sup>
C4T	2980	OL15	Na <sup>+</sup>	—
C4T	2980	OL15+gHBfix	Na <sup>+</sup>	51,000
C4T	7930	OL15	Na <sup>+</sup>	—
C4T	7930	OL15+gHBfix	Na <sup>+</sup>	<1.0
C4T	9390	OL15	Na <sup>+</sup>	16,000
C4T	9390	OL15+gHBfix	Na <sup>+</sup>	4.0
C4T	330	OL15	KCl	16,000
C4T	330	OL15+gHBfix	KCl	34,000
C4T	1280	OL15	KCl	<0.5
C4T	1280	OL15+gHBfix	KCl	<1.0
C4T	5660	OL15	KCl	55,000
C4T	5660	OL15+gHBfix	KCl	58,000
C10T	2980	OL15	Na <sup>+</sup>	<1.5
C10T	2980	OL15+gHBfix	Na <sup>+</sup>	<0.5
C10T	7930	OL15	Na <sup>+</sup>	<1.0
C10T	7930	OL15+gHBfix	Na <sup>+</sup>	<1.0
C10T	9390	OL15	Na <sup>+</sup>	<0.5
C10T	9390	OL15+gHBfix	Na <sup>+</sup>	<0.5
C10T	330	OL15	KCl	29,000
C10T	330	OL15+gHBfix	KCl	<0.5
C10T	1280	OL15	KCl	<1.0
C10T	1280	OL15+gHBfix	KCl	<1.0
C10T	5660	OL15	KCl	<0.5
C10T	5660	OL15+gHBfix	KCl	<1.0

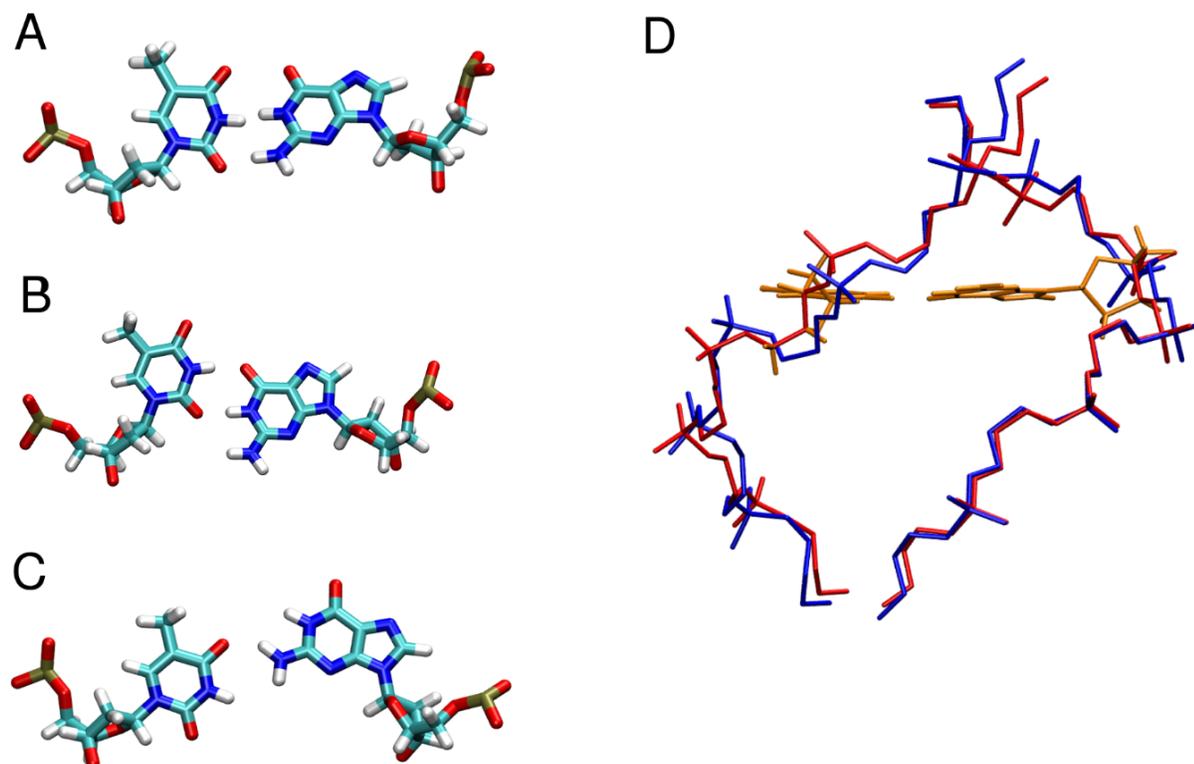
<sup>1</sup>Time of the [d(GGGCCC)]<sub>2</sub> simulation at which the system was taken as the starting point for subsequent d(GGGTCC).(GGGCCC) simulations.

<sup>2</sup>Time required to finish the transition to wobble GT base pair, counted from the simulation start.

The purpose of the simulations was to see what kind of transitions the system would undergo right after the mutation of a middle **GC** base pair into **GT**. The wobble **GT** interaction mimics the structure of the wobble exciplex of **GC** in the  $n\pi^*_{CT}$  excited state. Therefore, our main goal was to test whether such structural changes could occur in a B-DNA double helix on a picosecond timescale, *i.e.* within the lifetime of dark  $n\pi^*$  states in pyrimidine nucleosides.

We started the simulations from six selected snapshot geometries with the lowest RMSD to the averaged structure of [d(GGGCCC)]<sub>2</sub> subjected to either C4T, or C10T mutation (see Table 3 and Methods section for more details). Fourteen out of twenty-four simulations resulted in the formation of a wobble GT base pair in under **2 ps**, which indicates that such structural transitions are ultrafast and entail only slight changes within the sugar phosphate backbone. In the remaining ten simulation, a competing **GT** base pair with only one hydrogen bond was quickly formed instead of the wobble pair (Fig. 4). Nevertheless, eventually only two systems of the twenty four simulations did not attain the wobble base pair by the simulation end (100 ns).

The data in Table 3 show that the transition time necessary to reach the wobble geometry after mutation is either extremely fast (picosecond timescale), or relatively slow (thousands of picoseconds). Apparently, there is statistically significant discrepancy in distribution of the transition times between theoretically identical mutations C4T and C10T. We anticipate that the difference originates from non-identical backbone shape of the two strands of the double helix in the initial structures. In case of C4T, there is a small clash between the ad hoc created methyl group of **T** (in the C5 position) with ribose from the preceding residue in all six initial structures, while this strain is non-existent, except for one case, in the C10T mutants. This clash pushes the **T** into the alternative **GT** base pairing, and it then takes relatively long time to attain the wobble geometry.



**Fig. 4** **A)** Starting orientation of the GT base pair; **B)** GT wobble base pair; **C)** Non-Watson-Crick GT base pair with only one hydrogen bond formed in the simulations without fast GT wobble base pair formation; **D)** Comparison of backbone geometries. The starting backbone geometry in a simulation with the C4T mutation (blue) and the average backbone geometry calculated from one of the C4T simulations with GT wobble base pair formation (red). The wobble base pair is orange. The two structures are RMSD-fitted by all DNA heavy atoms.

Note that all six different starting geometries of the mutants were extracted from just two short (10 ns) simulations of the  $[d(\text{GGGCC})]_2$  sequence and that the structures were not taken randomly, but based on similarity to the average structure of the respective simulations. Since in both the  $[d(\text{GGGCC})]_2$  simulations the helices developed into similar structure (just by chance), the backbone shape was similar in all of them and the differences in behaviour between C4T and C10T mutants are consistent in all the simulations. Nevertheless, we emphasize here that these problems are irrelevant for the **GC** exciplex structure, because no such clash could be present in the **GC** exciplex owing to the absence of the C5 methyl group in **C**. Therefore we expect that the **GC** exciplex would behave like the C10T mutation, rather than the C4T one.

The causality between the clash and the direction of **T** movement is not the only element that affected

the final statistics. Other factors contribute to the process, too, which results in some C4T systems transiting fast into the wobble geometry, and one C10T system transiting into the alternative geometry. These less significant factors are: 1) exact mutual position of two clashing hydrogen atoms in the **GT** pair [Fig. 4], 2) differences in its surroundings (neighbouring base pairs, solvation), 3) slightly different force field, and 4) random distribution of initial atom velocities.

In case of fast formation of wobble GT interaction, the backbone geometry was nearly identical to the initial geometry of the [d(GGGCCC)]<sub>2</sub> backbone for several hundreds of femtoseconds. In addition, we calculated the average structure, taking every 10<sup>th</sup> ps in the range between 2 and 10 ns of each of the simulations with fast wobble base pair formation. The backbone from this averaged structure was still similar to the initial backbone geometry, though some smaller changes accommodated during the longer simulation period of several nanoseconds were present (Fig. 4).

## 2.6 MD simulations of the double-helix with G9H and C4T mutations.

After the simulations confirmed that GT wobble base pair can be reached on the (sub)picosecond time scale in d(GGGTCC).(GGGCCC), we decided to study whether HT wobble pair can be formed likewise. To test this hypothesis, we took a system analogous to the previous double-helix, specifically its mutated variant having guanine at position 9 mutated into hypoxanthine, d(GGGTCC).(GGHCCC), and proceeded as with d(GGGTCC).(GGGCCC) (see Methods for more details). The twelve independent simulations showed even higher ratio of fast transitions into the wobble base pair. In nine of them, HT wobble base pair was formed in under 2 ps, again with virtually identical backbone conformation as the starting structure. The other three simulations attained the wobble base pair on the nanosecond time scale (Table 4). These simulations thus support our hypothesis about efficient formation of the  $n\pi_{CT}^*$  exciplex in **HC**.

**Table 4** MD simulations of d(GGGTCC).(GGGHCC)

Extraction time (ps) <sup>1</sup>	Force field	Ionic conditions	Transition time to HT wobble (ps) <sup>2</sup>
1380	OL15	Na <sup>+</sup>	2.0
1380	OL15+gHBfix	Na <sup>+</sup>	3,000
4210	OL15	Na <sup>+</sup>	5,000
4210	OL15+gHBfix	Na <sup>+</sup>	18,000
6480	OL15	Na <sup>+</sup>	<0.5
6480	OL15+gHBfix	Na <sup>+</sup>	<0.5
2380	OL15	KCl	<0.5
2380	OL15+gHBfix	KCl	<0.5
6960	OL15	KCl	<0.5
6960	OL15+gHBfix	KCl	<0.5
9310	OL15	KCl	<1.0
9310	OL15+gHBfix	KCl	<1.0

<sup>1</sup>Time of the d(GGGCCC).(GGHCCC) simulation at which the system was taken as the starting point for subsequent d(GGGTCC).(GGHCCC) simulations.

<sup>2</sup>Time required to finish the transition to wobble HT base pair, counted from the simulation start.

---

## References

- 1 C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618–622.
- 2 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.
- 3 *TURBOMOLE V7.3 2018, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from*  
<http://www.turbomole.com>.
- 4 J. Schirmer, *Phys. Rev. A*, 1982, **26**, 2395–2416.
- 5 A. Dreuw and M. Wormit, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **5**, 82–95.
- 6 C. Hättig, in *Advances in Quantum Chemistry*, ed. H. J. A. Jensen, Academic Press, 2005, vol. 50 of Response Theory and Molecular Properties (A Tribute to Jan Linderberg and Poul Jørgensen), pp. 37–60.
- 7 O. Christiansen, H. Koch and P. Jørgensen, *Chem. Phys. Lett.*, 1995, **243**, 409–418.
- 8 A. Klamt and G. Schüürmann, *Journal of the Chemical Society, Perkin Transactions 2*, 1993, 799–805.
- 9 A. Tajti and P. G. Szalay, *J. Chem. Theory Comput.*, 2019.
- 10 R. L. Martin, *J. Chem. Phys.*, 2003, **118**, 4775–4777.
- 11 F. Plasser, “*TheoDORE 1.4: A package for theoretical density, orbital relaxation, and exciton analysis*”,  
<http://theodore-qc.sourceforge.net> (2018).
- 12 F. Plasser, M. Wormit and A. Dreuw, *J. Chem. Phys.*, 2014, **141**, 024106.
- 13 F. Plasser, S. A. Bächler, M. Wormit and A. Dreuw, *J. Chem. Phys.*, 2014, **141**, 024107.
- 14 B. G. Levine, J. D. Coe and T. J. Martínez, *J. Phys. Chem. B*, 2008, **112**, 405–413.
- 15 T. Shiozaki, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2018, **8**, e1331.
- 16 H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *J. Phys. Chem.*, 1987, **91**, 6269–6271.
- 17 I. S. Joungh and T. E. Cheatham, *J. Phys. Chem. B*, 2008, **112**, 9020–9041.
- 18 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
- 19 A. Pérez, I. Marchán, D. Svozil, J. Šponer, T. E. Cheatham, C. A. Laughton and M. Orozco, *Biophys. J.*, 2007, **92**, 3817–3829.
- 20 M. Krepl, M. Zgarbová, P. Stadlbauer, M. Otyepka, P. Banáš, J. Koča, T. E. Cheatham, P. Jurečka and J. Šponer, *J. Chem. Theory Comput.*, 2012, **8**, 2506–2520.
- 21 M. Zgarbová, F. J. Luque, J. Šponer, T. E. Cheatham, M. Otyepka and P. Jurečka, *J. Chem. Theory Comput.*, 2013, **9**, 2339–2354.
- 22 M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo and P. Jurečka, *J. Chem. Theory Comput.*, 2015, **11**, 5723–5736.
- 23 D. Case, R. Betz, D. Cerutti, T. Cheatham III, T. Darden, R. Duke, T. Giese, H. Gohlke, A. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. Lee, S. Legrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. Merz, G. Monard, H. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, D. Roe, A. Roitberg, C. Sagui, C. Simmerling, W. Botello-Smith, J. Swails, R. Walker, J. Wang, R. Wolf, X. Wu, L. Xiao and P. Kollman, *AMBER 16*, 2016.
- 24 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
- 25 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
- 26 S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952–962.

- 
- 27 C. W. Hopkins, S. Le Grand, R. C. Walker and A. E. Roitberg, *J. Chem. Theory Comput.*, 2015, **11**, 1864–1874.
- 28 P. Kührová, V. Mlýnský, M. Zgarbová, M. Krepl, G. Bussi, R. B. Best, M. Otyepka, J. Šponer and P. Banáš, *J. Chem. Theory Comput.*, 2019, **15**, 3288–3305.
- 29 K. Röttger, H. J. Marroux, M. P. Grubb, P. M. Coulter, H. Böhnke, A. S. Henderson, M. C. Galan, F. Temps, A. J. Orr-Ewing and G. M. Roberts, *Angewandte Chemie International Edition*, 2015, **54**, 14719–14722.
- 30 K. Siriwong, A. A. Voityuk, M. D. Newton and N. Rösch, *The Journal of Physical Chemistry B*, 2003, **107**, 2595–2601.
- 31 J. Mazur and R. Jernigan, *Biopolymers: Original Research on Biomolecules*, 1991, **31**, 1615–1629.
- 32 R. Szabla, H. Kruse, P. Stadlbauer, J. Šponer and A. L. Sobolewski, *Chem. Sci.*, 2018, **9**, 3131–3140.