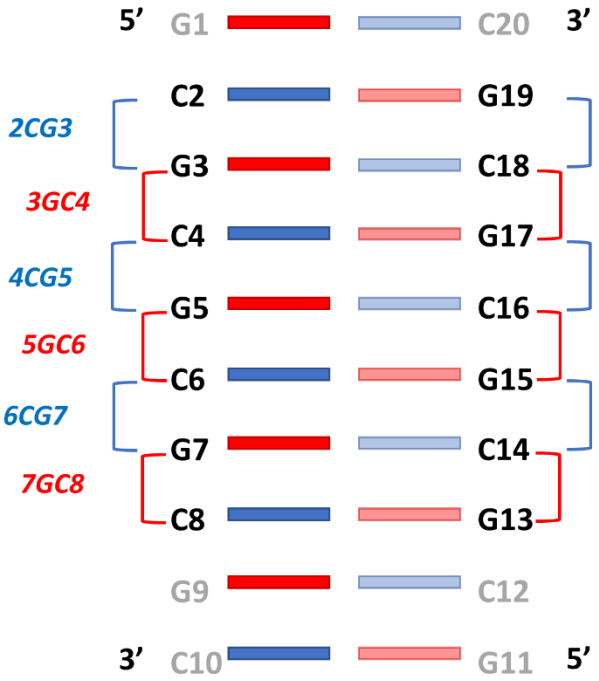
# The photoactivated dynamics of dGpdC and dCpdG sequences in DNA: a comprehensive quantum mechanical study.

## Supporting Information

Lara Martinez-Fernandez, James A. Green, Luciana Esposito, Martha Yaghoubi Jouybari, Yuyuan Zhang, Fabrizio Santoro, Bern Kohler and Roberto Improta

### 1. MD details




**Scheme S1.** Schematic representation of the analyzed CG and GC steps in the $d(GC)_5$ structures extracted from the MD trajectory. The nucleotides in grey have been neglected.

**Table S1**. Base stacking and interbase distances for the three CG and GC steps.

| | | Stacking (Area) | | | | | | | | COM ring atms distance[2] | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **2CG3** | | Overlap Area[1] | | | | | | | | *C1G1* | *C2G2* |
| | **t (ns)** | 5'-C2/G3-3' | | 5'-C2/C18-5' | | 3'-G3/G19-3' | | 5'-C18/G19-3' | | | |
| | | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | | |
| | 0 | 0 | 0 | 0 | 0 | 0.42 | 0 | 1.51 | 0 | 4.95 | 4.09 |
| | 400 | 0.63 | 0 | 0 | 0 | 1.79 | 0 | 0.16 | 0 | 4.22 | 4.45 |
| | 800 | 0.01 | 0 | 0 | 0 | 0.25 | 0 | 2.08 | 0 | 4.90 | 4.13 |
| | 1200 | 6.44 | 2.49 | 0 | 0 | 0.01 | 0 | 0 | 0 | 3.82 | 5.15 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1600 | 5.4 | 1.48 | 0 | 0 | 0.02 | 0 | 0 | 0 | 3.82 | 4.68 |
| | 2000 | 3.3 | 1.19 | 0 | 0 | 0.67 | 0 | 0.8 | 0 | 4.09 | 4.45 |
| **4CG5** | | | | | | | | | | | |
| | **t (ns)** | **5'-C4/G5-3'** | | **5'-C4/C16-5'** | | **3'-G5/G17-3'** | | **5'-C16/G17-3'** | | | |
| | | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | | |
| | 0 | 4.35 | 0.55 | 0 | 0 | 0 | 0 | 0 | 0 | 4.14 | 4.96 |
| | 400 | 0 | 0 | 0 | 0 | 0 | 0 | 5.79 | 1.56 | 4.90 | 3.56 |
| | 800 | 0 | 0 | 0 | 0 | 2.45 | 0.23 | 0.33 | 0 | 5.02 | 4.47 |
| | 1200 | 0.21 | 0 | 0 | 0 | 0.14 | 0 | 1.51 | 0 | 4.60 | 4.15 |
| | 1600 | 7.63 | 3.47 | 0 | 0 | 0 | 0 | 0 | 0 | 3.39 | 5.16 |
| | 2000 | 1.27 | 0 | 0 | 0 | 0.21 | 0 | 2.7 | 0.02 | 4.39 | 4.01 |
| **6CG7** | | | | | | | | | | | |
| | **t (ns)** | **5'-C6/G7-3'** | | **5'-C6/C14-5'** | | **3'-G7/G15-3'** | | **5'-C14/G15-3'** | | | |
| | | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | | |
| | 0 | 2.36 | 0.09 | 0 | 0 | 0.21 | 0 | 2.11 | 0 | 4.08 | 4.20 |
| | 400 | 0.51 | 0 | 0 | 0 | 0.92 | 0 | 0.64 | 0 | 4.49 | 4.45 |
| | 800 | 0 | 0 | 0 | 0 | 0.01 | 0 | 5.53 | 1.3 | 5.03 | 3.91 |
| | 1200 | 0 | 0 | 0 | 0 | 0.22 | 0 | 5.03 | 2.02 | 4.93 | 4.00 |
| | 1600 | 0 | 0 | 0 | 0 | 2.42 | 0.36 | 0.81 | 0.04 | 5.01 | 4.30 |
| | 2000 | 1.04 | 0 | 0 | 0 | 1.43 | 0 | 0.54 | 0 | 4.49 | 4.50 |

Table S1. *continued*

| | | **Stacking (Area)** | | | | | | | | **COM ring atms distance[2]** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **3GC4** | | **Overlap Area[1]** | | | | | | | | ***G1C1*** | ***G2C2*** |
| | **t (ns)** | **5'-G3/C4-3'** | | **5'-G3/G17-5'** | | **3'-C4/C18-3'** | | **5'-G17/C18-3'** | | | |
| | | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | | |
| | 0 | 4.98 | 2.02 | 0 | 0 | 0 | 0 | 5.18 | 2.38 | 3.66 | 4.02 |
| | 400 | 4.87 | 1.74 | 0 | 0 | 0 | 0 | 0.4 | 0 | 3.85 | 4.43 |
| | 800 | 5.78 | 3.18 | 0 | 0 | 0 | 0 | 5.06 | 2.54 | 3.62 | 3.63 |
| | 1200 | 2.64 | 0.82 | 0 | 0 | 0 | 0 | 5.45 | 3.17 | 3.75 | 3.46 |
| | 1600 | 0.02 | 0 | 0.15 | 0 | 0 | 0 | 3.33 | 0.42 | 4.82 | 3.81 |
| | 2000 | 4.74 | 1.98 | 0 | 0 | 0 | 0 | 1.49 | 0.15 | 3.68 | 4.12 |
| **5GC6** | | | | | | | | | | | |
| | **t (ns)** | **5'-G5/C6-3'** | | **5'-G5/G15-5'** | | **3'-C6/C16-3'** | | **5'-G15/C16-3'** | | | |
| | | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | | |
| | 0 | 0.76 | 0 | 0.06 | 0 | 0 | 0 | 1.6 | 0.09 | 3.94 | 3.96 |
| | 400 | 3.18 | 1.22 | 0 | 0 | 0 | 0 | 3.67 | 1.57 | 3.74 | 3.60 |
| | 800 | 5.07 | 2.84 | 0 | 0 | 0 | 0 | 2.26 | 0.67 | 3.72 | 3.91 |
| | 1200 | 5.84 | 3.59 | 0 | 0 | 0 | 0 | 3.19 | 1.05 | 3.51 | 3.81 |
| | 1600 | 1.44 | 0.2 | 0 | 0 | 0 | 0 | 4.11 | 1.73 | 4.19 | 3.68 |
| | 2000 | 6.07 | 3.15 | 0 | 0 | 0 | 0 | 4.17 | 1.36 | 3.64 | 4.11 |
| **7GC8** | | | | | | | | | | | |

| | t (ns) | 5'-G7/C8-3' | | 5'-G7/G13-5' | | 3'-C8/C14-3' | | 5'-G13/C14-3' | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | ring+ exocycl atoms | only ring atoms | | |
| | 0 | 2.55 | 0.82 | 0 | 0 | 0 | 0 | 4.2 | 1.82 | 3.85 | 3.47 |
| | 400 | 5.28 | 3.14 | 0 | 0 | 0 | 0 | 5.65 | 3.55 | 3.41 | 3.51 |
| | 800 | 5.9 | 3.29 | 0 | 0 | 0 | 0 | 1.78 | 0.27 | 3.47 | 4.04 |
| | 1200 | 5.58 | 2.91 | 0 | 0 | 0 | 0 | 4.22 | 1.9 | 4.29 | 3.84 |
| | 1600 | 4.93 | 1.83 | 0 | 0 | 0 | 0 | 0.75 | 0 | 3.66 | 4.17 |
| | 2000 | 5.99 | 3.93 | 0 | 0 | 0 | 0 | 5.34 | 3.37 | 3.50 | 4.11 |

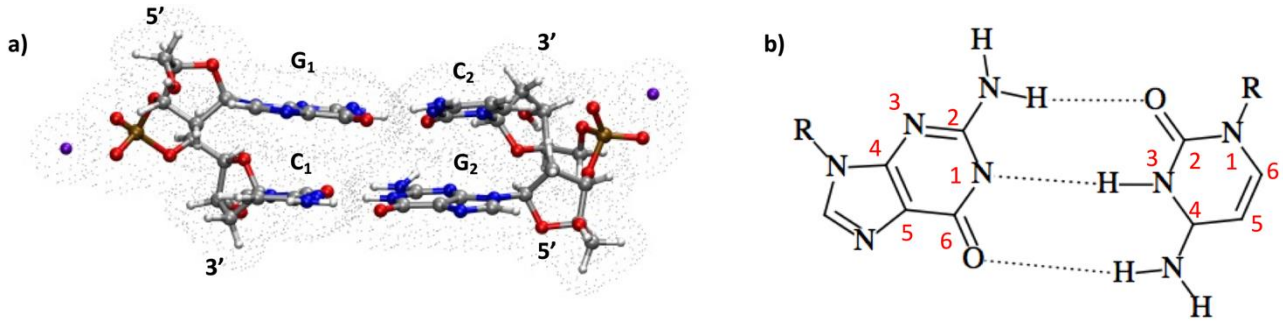
[1] In 3DNA, base-stacking interactions are assessed from planar projections of the ring and exocyclic atoms in consecutive bases or base pairs; Overlap area includes a distance cutoff of 4.5 Å. If the shortest distance of any two-pair between the two sets of atoms is larger than the cutoff, then the stacking area is set to 0.

[2] The interbase distance is calculated between the centers of mass of ring atoms on strand1 (*G1C1* and *C1G1*) and on strand2 (*G2C2* and *C2G2*).

## 2. QM details

### 2.1 Models

In order to simulate a $d(GC)_9$ duplex we have adopted two computational models, i.e. 2GC (depicted in Scheme S2) and 3•GC (see Figure 1 in the main text), containing two or three G•C base pairs arranged to yield an alternating sequence along each strand. As a starting structure we used the requisite number of base pairs extracted from the crystal structure of a CG rich duplex (1CGC.pdb).[1] Since the experiments we use as reference were performed in $D_2O$, all of the exchangeable hydrogen atoms (those bonded to nitrogen and oxygen atoms) have been substituted by deuterium atoms in our IR calculations.




**Scheme S2. a)** Schematic picture of the 2•GC duplex used to model $(GC)_9$ in most of our calculations. The cavity used in PCM calculations is also schematically depicted. The view is looking into the major groove. Color code for the different atoms: Gray=carbon; red=oxygen, blue=nitrogen; white=hydrogen (or deuterium); ochre=phosphorous; purple=sodium. **b)** guanine and cytosine atom labeling.

### 2.2 Methods

See section in the main text.

## 3. Additional results

3.1 FC region in 2GC and 2CG

In the crystal structure of the GC-rich DNA sequence used as the starting geometry for our geometry optimization,[1] the two strands exhibit some differences, which are maintained also in the PCM/M052X optimized geometry of 2GC. In one strand (hereafter labeled as strand 1) the bases are more closely stacked than in the other one (strand 2). Moreover, in strand 2 the planes of the two bases are slightly tilted with respect to the average WC base pair in order to maximize an intrastrand HB between the amino group of $\mathbf{G_2}$ and the CO group of $\mathbf{C_2}$. This is reflected in an unusual combination of propeller parameters for the two consecutive base pair. We have exploited this asymmetry to get some insights on the possible effect on the duplex structural inhomogeneity on its vibrational behavior. The lowest energy excited states (Table S2) of 2GC, up to 6 eV, derive from the combination of the $\pi\pi^*$ bright excited states of G ($L_a$ and $L_b$) and C, with G→C CT states.[2] Due to the close stacking between the bases, there is a strong mixing between the different excited states. Nonetheless, it is possible to qualitatively recognize their predominant diabatic character, as detailed in Table S2 below.

**Table S2**. Vertical absorption energy ($\Delta E$), oscillator strength ($f$) and CT character of the 7 lowest energy singlet excited states of 2GC and 2CG in the FC region.[2]

| **2GC** | **Character** | **$\Delta E$ (eV)** | ***f*** | **CT** |
|---|---|---|---|---|
| $S_1$ | $G_1 \rightarrow C_1$ ($CT^{intra}$-1) | 4.96 | 0.004 | 0.80 |
| $S_2$ | $G_2 \rightarrow C_2$ ($CT^{intra}$-2) | 5.04 | 0.024 | 0.60 |
| $S_3$ | $\pi\pi^*$ ($G_1$) | 5.23 | 0.063 | 0.05 |
| $S_4$ | $\pi\pi^*$ ($C_2$)[a] | 5.27 | 0.131 | 0.10 |
| $S_5$ | $\pi\pi^*$ ($C_1$)[a] | 5.30 | 0.145 | 0.02 |
| $S_6$ | $\pi\pi^*$ ($G_2$)[a] | 5.35 | 0.313 | 0.12 |
| $S_7$ | $G_1 \rightarrow C_2$ ($CT^{inter}$) | 5.64 | 0.002 | 0.77 |
| **2CG** | **Character** | **$\Delta E$ (eV)** | ***f*** | **CT** |
| $S_1$ | $G_1 \rightarrow C_1$ ($CT^{intra}$-1) | 5.11 | 0.030 | 0.41 |
| $S_2$ | $\pi\pi^*$ ($C_2$+$C_1$) | 5.23 | 0.387 | 0.03 |
| $S_3$ | $\pi\pi^*$ ($C_1$+$C_2$) | 5.24 | 0.007 | 0.09 |
| $S_4$ | $\pi\pi^*$ ($G_2$)[b] | 5.30 | 0.122 | 0.16 |
| $S_5$ | $\pi\pi^*$ ($G_1$)[b] | 5.35 | 0.124 | 0.17 |
| $S_6$ | $\pi\pi^*$ ($G_2$)[b] | 5.62 | 0.609 | 0.10 |
| $S_7$ | $G_2 \rightarrow C_2$ | 5.65 | 0.338 | 0.62 |
| $S_8$ | $\pi\pi^*$ ($G_1$)[b] | 5.71 | 0.283 | |
| $S_{11}$ | $G_1 \rightarrow C_2$ | 5.83 | 0.003 | 0.44+ |

| | | | | |
|---|---|---|---|---|
| | $G_2 \rightarrow C_1$ ($CT^{inter}$) | | | 0.30 |

[a]Strong mixing between the lowest energy $\pi\pi^*$ states on C1 and C2 and $L_a$ state on G2.

[b]Strong mixing between the $\pi\pi^*$ states of G's and the intra-strand CT

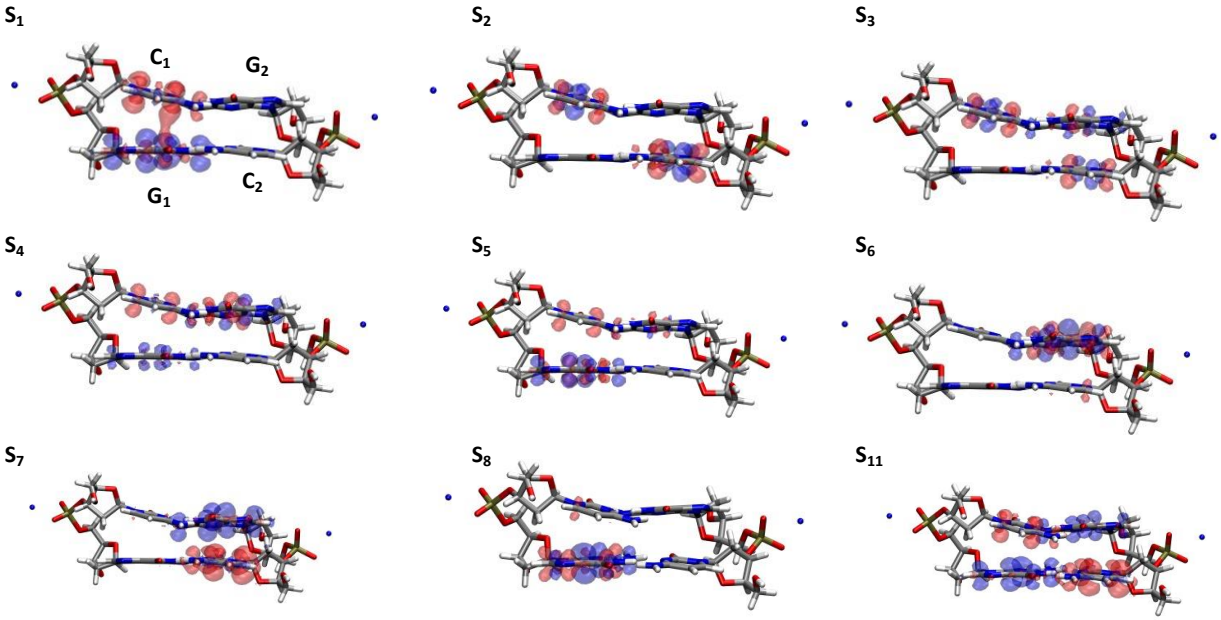



**Figure S1**. Excited state density difference with respect $S_0$ computed for the 11 lowest energy excited state of 2CG PCM/TD-M052X/6-31G(d) calculations. A decrease of the electron density is depicted in blue, an increase in red.

3.2 FC region in 3GC

In Table S3 we report the main features of the 14 lowest energy excited states of 3GC in the FC region. The general picture is similar to that described above for 2GC. The three lowest energy excited states have a clear intra-strand CT character. On the other hand, due to the conformational asymmetry in the stacking geometry and to increase of the possible intra-strand CT transitions, we observe that several excited states have some CT character and involve more than one base. For example, $S_3$, with a predominant G2'→C3' (see Figure S2 for base labeling) intra-strand CT character, also has a small G2'→C2' inter-strand CT character. The 6 excited states between $S_4$ and $S_9$ derive from the mixing of the lowest energy bright $\pi\pi^*$ excited states of C and G ($L_a$-like) and their CT character is smaller than that found in the three lowest ones. Then, we find in $S_{10}$ another excited state with strong intra-strand CT character (G3→C2). It falls at 5.42 eV, i.e., 0.3~0.4 eV higher in energy than the other ones, confirming that the increase of the chain length leads to a larger mixing between bright and intra-strand CT states. The three following excited states are mainly localized on the Gs and can be assigned to their $L_b$ bright $\pi\pi^*$ states. Then we find the first excited state ($S_{14}$), with significant inter-strand CT states, G2'→C2, falling at 5.67 eV, i.e. > 0.5 eV less stable than the lowest energy intra-strand CT states.

We end this section by reminding that in case of a stiff GC DNA structure where all the bases adopt an almost ideal B-DNA structure, delocalized High-energy Emitting Long-lived Mixed (HELM) excited states, extended across both strands, can be populated.[3] They arise from mixing between cytosine Frenkel excitons and

guanine-to-cytosine CT states and emit at energies higher than ππ* states localized on single bases. On the other hand, the importance of these states is decreased by DNA fluctuations and, due to their small population (< 1%), they are not expected to affect the Difference IR (DIR) spectra of $(GC)_9$.

**Table S3**. Vertical absorption energy ($\Delta E$), oscillator strength ($f$) and CT character of the 14 lowest energy excited states of 3GC in the FC region.

| | Character | $\Delta E$ (eV) | $f$ | CT |
|---|---|---|---|---|
| $S_1$ | G2→C2 | 5.01 | 0.0429 | 0.42 |
| $S_2$ | G1→C1 | 5.05 | 0.0785 | 0.51 |
| $S_3$ | G2→C2/C2’ | 5.17 | 0.0309 | 0.47 |
| $S_4$ | ππ* (C1) | 5.19 | 0.0654 | 0.24 |
| $S_5$ | ππ* (G1’) | 5.23 | 0.0712 | 0.03 |
| $S_6$ | ππ* (C2’) + CT G2-C2’ | 5.27 | 0.2026 | 0.20 |
| $S_7$ | ππ* (C2) mixed | 5.29 | 0.0659 | 0.15 |
| $S_8$ | Delocalized | 5.33 | 0.1256 | 0.25 |
| $S_9$ | Delocalized | 5.37 | 0.1037 | 0.36 |
| $S_{10}$ | G1’→C1 | 5.42 | 0.0564 | 0.45 |
| $S_{11}$ | ππ* (G2) | 5.60 | 0.9257 | 0.08 |
| $S_{12}$ | ππ* (G1’) | 5.64 | 0.2207 | 0.07 |
| $S_{13}$ | ππ* (G1) | 5.65 | 0.3130 | 0.13 |
| $S_{14}$ | G2→C1 | 5.67 | 0.0055 | 0.60 |

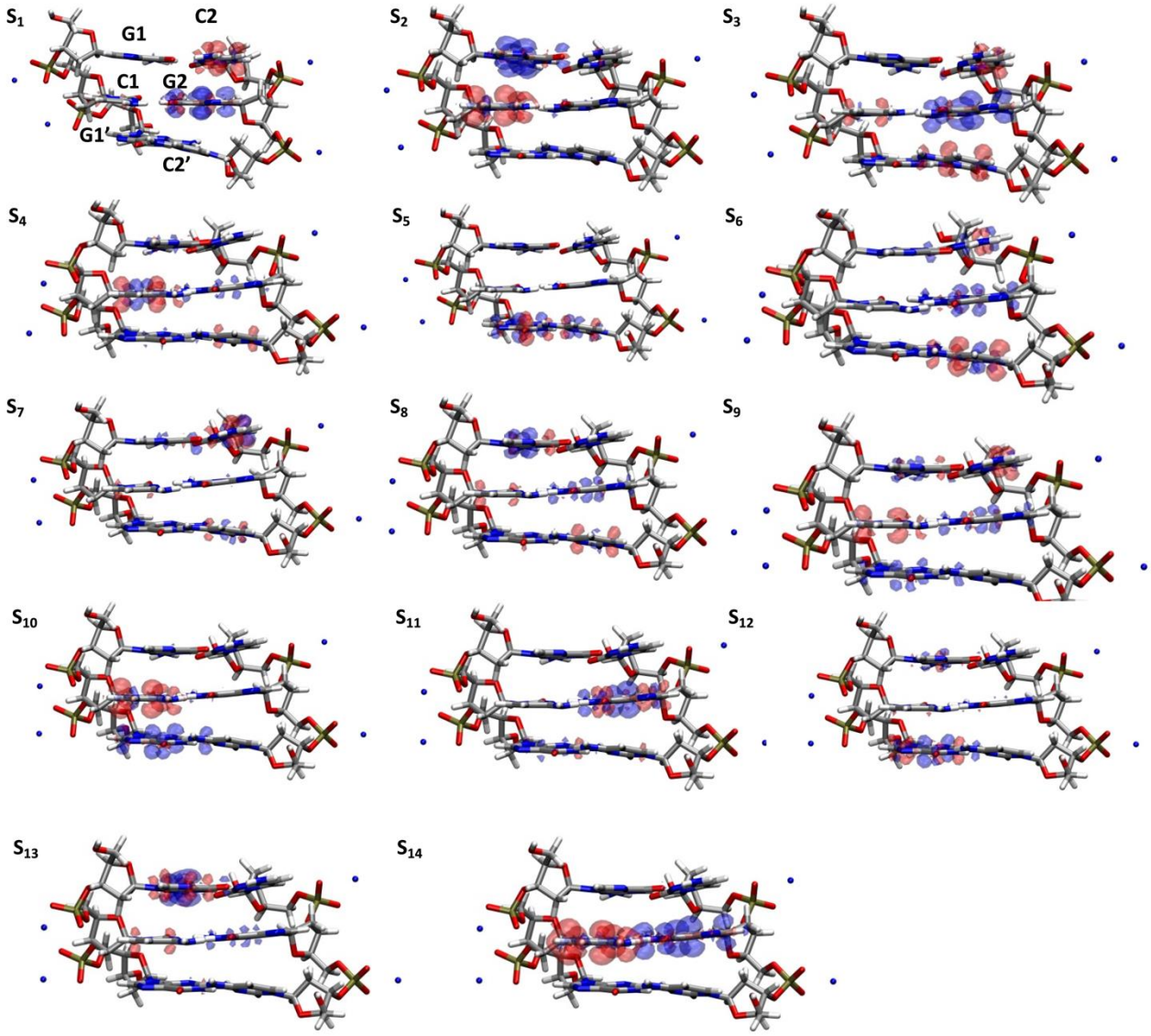



**Figure S2**. Excited state density difference with respect $S_0$ computed for the 14 lowest energy excited state of 3GC PCM/TD-M052X/6-31G(d) calculations. A decrease of the electron density is depicted in blue, an increase in red.

3.3 Ground state IR spectrum

In Figure 2 of the main text, we report the computed ground state spectrum 3GC and 2GC. Our calculations predict two close lying bands at 1650~1670 $cm^{-1}$. The high energy peak is more intense and it can be assigned to the CO stretching of the G bases, while that on the red to the CO stretching modes of C bases. It is important to remind, however, that an assignment based on local vibrational modes is only approximate. First of all, as previously highlighted, there is a strong intra-base vibrational coupling: for example, the CO stretching modes are strongly mixed with the ring stretching and amino-modes (stretching and scissoring). Then, HB and stacking interactions lead to a significant inter-base vibrational mixing. With this caveat, we can then assign the two features at ~1570 $cm^{-1}$ and ~1500 $cm^{-1}$ at G and C ring modes, respectively, in both cases coupled with the stretching/scissoring modes of the amino groups. The computed spectra nicely compare with the available experimental results, but for an almost uniform red-shift of 50~60 $cm^{-1}$, mainly due to the neglect of anharmonic effects. The most significant discrepancy with the experimental spectrum of GC-DNA concerns the relative intensity of the ring G band which, according to calculations, it is almost twice as large as that associated to Ring C stretching, whereas in the experimental spectra their intensity is similar. Being based on a single minimum and not considering the conformational heterogeneity of GC DNA, we can underestimate the energy difference between 'analogous' vibrational modes on different bases and overestimate their coupling, potentially over-enhancing resonance effects. This bias could be more acute for some peaks, as the intense one at 1570 $cm^{-1}$ which is assigned to the combination of the two ring-stretching of the two G's. The two Gs, whose amino groups strongly overlap, are more stacked than Cs, explaining why we overestimate their inter-base couplings. Interestingly, in 3GC, the bases exhibit a larger variety of stacking arrangement than in the smaller 2GC, where geometry optimizations can more easily maximize the stacking interactions between the two WC pairs, and the relative intensity of this band is in better agreement with the experimental one. The overestimation of the intensity of this band is expected to be larger for the ground state, where two Gs are 'equivalent', i.e., they are both in their ground electronic state, with a symmetric arrangement, than for the excited states, where only one of them is excited. In many of the electronic states investigated, indeed, at least of one of the G bases is excited, and, in any case, the duplex is less symmetric. As a consequence, we could overestimate the intensity of the negative bleaching band at 1570 $cm^{-1}$ in the computed DIR spectra.

3.4 FrD-LVC Calculations

3.4.1 GC

**Table S4** Diabatic energies $E_{ii}$ and constant couplings $E_{ij}$ from FrD-LVC model of GC at its ground state minimum.

| | G1(La) | G1(Lb) | C1(ππ*1) | C1(ππ*2) | G2(La) | G2(Lb) | C2(ππ*1) | C2(ππ*2) | G1→C1 CT | G1→C2 CT | G2→C1 CT | G2→C2 CT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G1(La) | 5.304 | | | | | | | | | | | |
| G1(Lb) | 0.071 | 5.775 | | | | | | | | | | |
| C1(ππ*1) | -0.027 | 0.007 | 5.311 | | | | | | | | | |
| C1(ππ*2) | 0.021 | 0.013 | 0.101 | 6.198 | | | | | | | | |
| G2(La) | 0.023 | 0.011 | -0.022 | 0.046 | 5.290 | | | | | | | |
| G2(Lb) | 0.002 | 0.017 | 0.016 | -0.005 | 0.078 | 5.754 | | | | | | |
| C2(ππ*1) | -0.022 | 0.015 | 0.005 | -0.001 | 0.009 | 0.009 | 5.354 | | | | | |
| C2(ππ*2) | 0.047 | -0.005 | -0.002 | 0.007 | 0.052 | 0.035 | 0.111 | 6.207 | | | | |
| G1→C1 CT | 0.039 | 0.024 | 0.036 | 0.126 | 0.000 | -0.001 | 0.000 | 0.000 | 5.244 | | | |
| G1→C2 CT | 0.058 | 0.007 | 0.000 | 0.000 | -0.001 | -0.001 | 0.021 | -0.007 | -0.001 | 5.929 | | |
| G2→C1 CT | 0.000 | -0.001 | 0.029 | -0.014 | 0.073 | 0.001 | 0.000 | 0.000 | -0.020 | 0.000 | 5.865 | |
| G2→C2 CT | -0.002 | 0.000 | 0.000 | 0.000 | 0.102 | 0.041 | -0.075 | -0.200 | 0.000 | -0.024 | 0.000 | 5.329 |

**Table S5** Averaged diabatic energies $E_{ii}$ and constant couplings $E_{ij}$ from FrD-LVC models of GC from the molecular dynamics snapshots.

| | G1(La) | G1(Lb) | C1(ππ*1) | C1(ππ*2) | G2(La) | G2(Lb) | C2(ππ*1) | C2(ππ*2) | G1→C1 CT | G1→C2 CT | G2→C1 CT | G2→C2 CT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G1(La) | 5.306 | | | | | | | | | | | |
| G1(Lb) | 0.062 | 5.770 | | | | | | | | | | |
| C1(ππ*1) | 0.021 | 0.009 | 5.312 | | | | | | | | | |
| C1(ππ*2) | 0.033 | 0.013 | 0.092 | 6.214 | | | | | | | | |
| G2(La) | 0.028 | 0.005 | 0.020 | 0.046 | 5.320 | | | | | | | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G2(Lb) | 0.005 | 0.013 | 0.016 | 0.003 | 0.064 | 5.775 | | | | | | |
| C2($\pi\pi$*1) | 0.021 | 0.016 | 0.004 | 0.002 | 0.018 | 0.007 | 5.318 | | | | | |
| C2($\pi\pi$*2) | 0.047 | 0.003 | 0.003 | 0.007 | 0.041 | 0.015 | 0.103 | 6.206 | | | | |
| G1→C1 CT | 0.063 | 0.030 | 0.086 | 0.100 | 0.001 | 0.002 | 0.001 | 0.001 | 5.291 | | | |
| G1→C2 CT | 0.070 | 0.007 | 0.001 | 0.001 | 0.001 | 0.001 | 0.024 | 0.016 | 0.002 | 5.878 | | |
| G2→C1 CT | 0.001 | 0.000 | 0.023 | 0.014 | 0.064 | 0.005 | 0.000 | 0.000 | 0.023 | 0.000 | 5.896 | |
| G2→C2 CT | 0.001 | 0.002 | 0.001 | 0.001 | 0.059 | 0.033 | 0.045 | 0.084 | 0.000 | 0.024 | 0.002 | 5.308 |

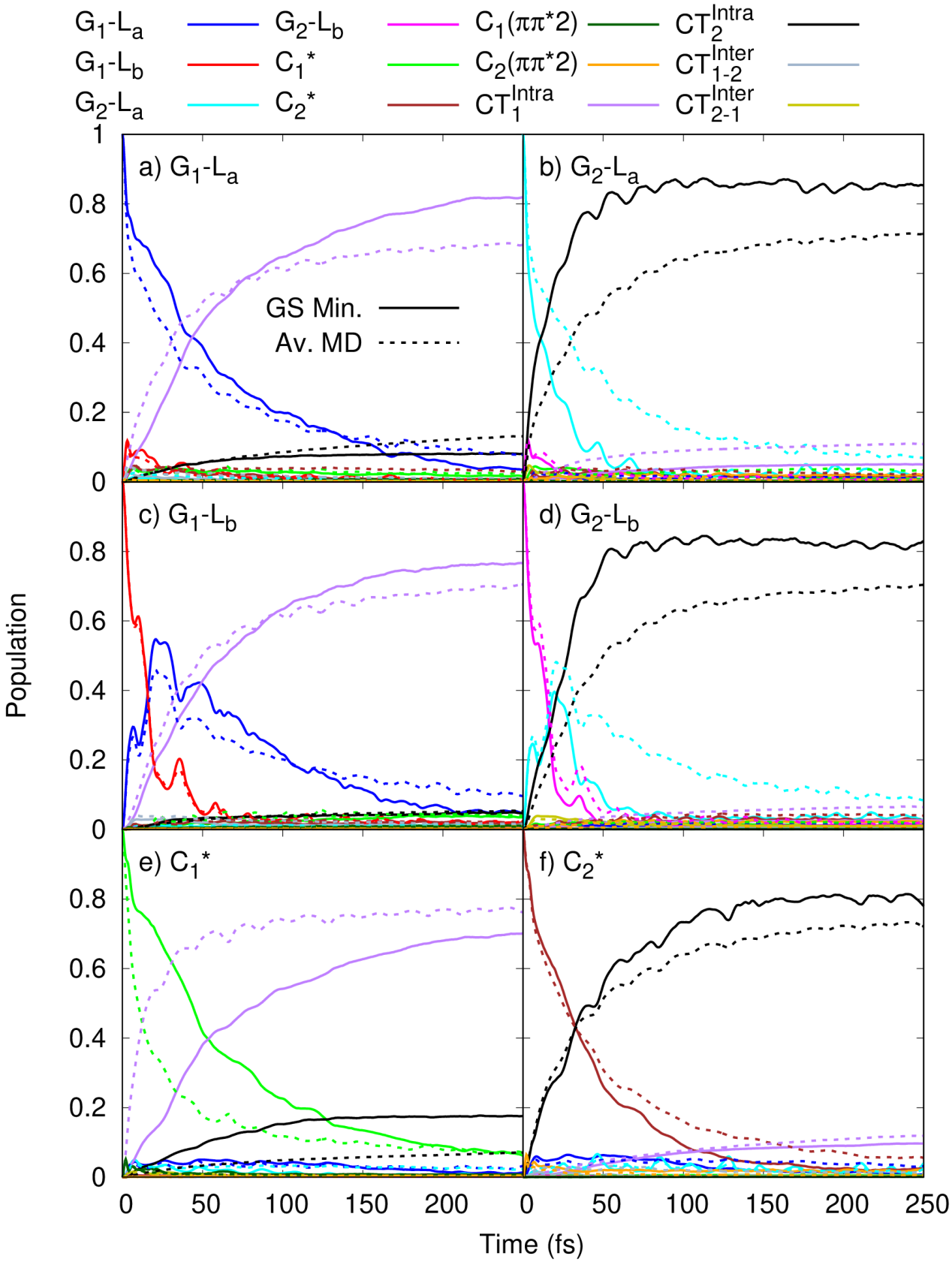



**Figure S3:** Full version of Figure 3 from the main text, including all diabatic states.

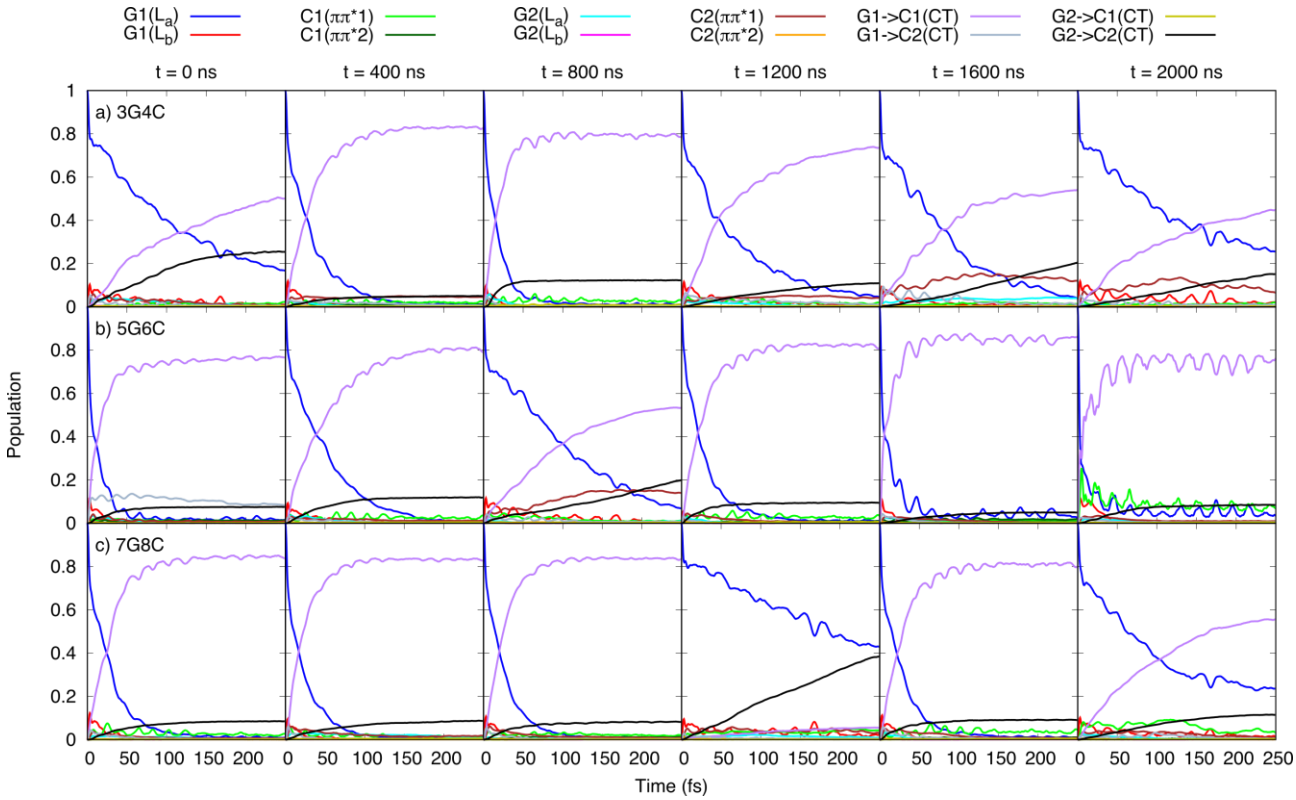



**Figure S4:** Populations following initial excitation to G1(La) for each of the GC tetrad MD snapshots.

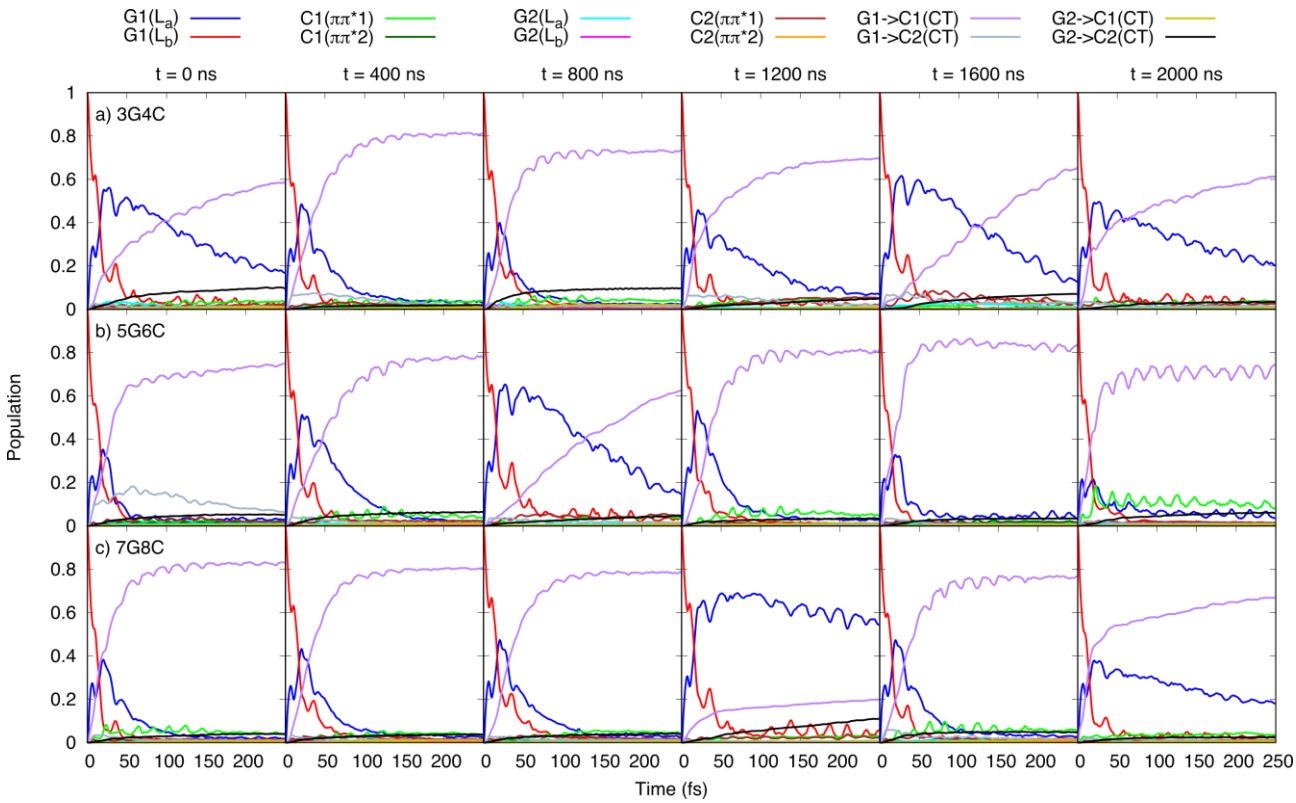



**Figure S5:** Populations following initial excitation to G1(Lb) for each of the GC tetrad MD snapshots.

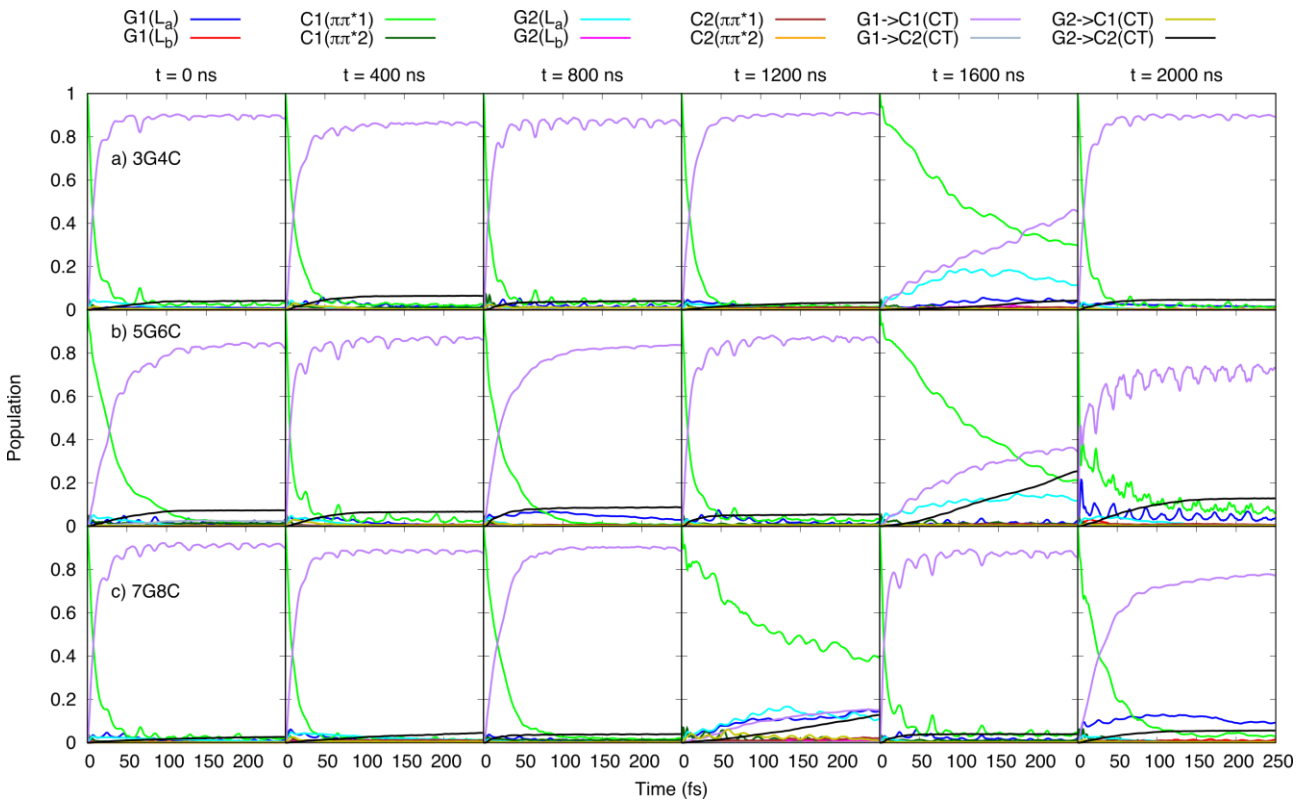



**Figure S6:** Populations following initial excitation to C1(ππ*1) for each of the GC tetrad MD snapshots.

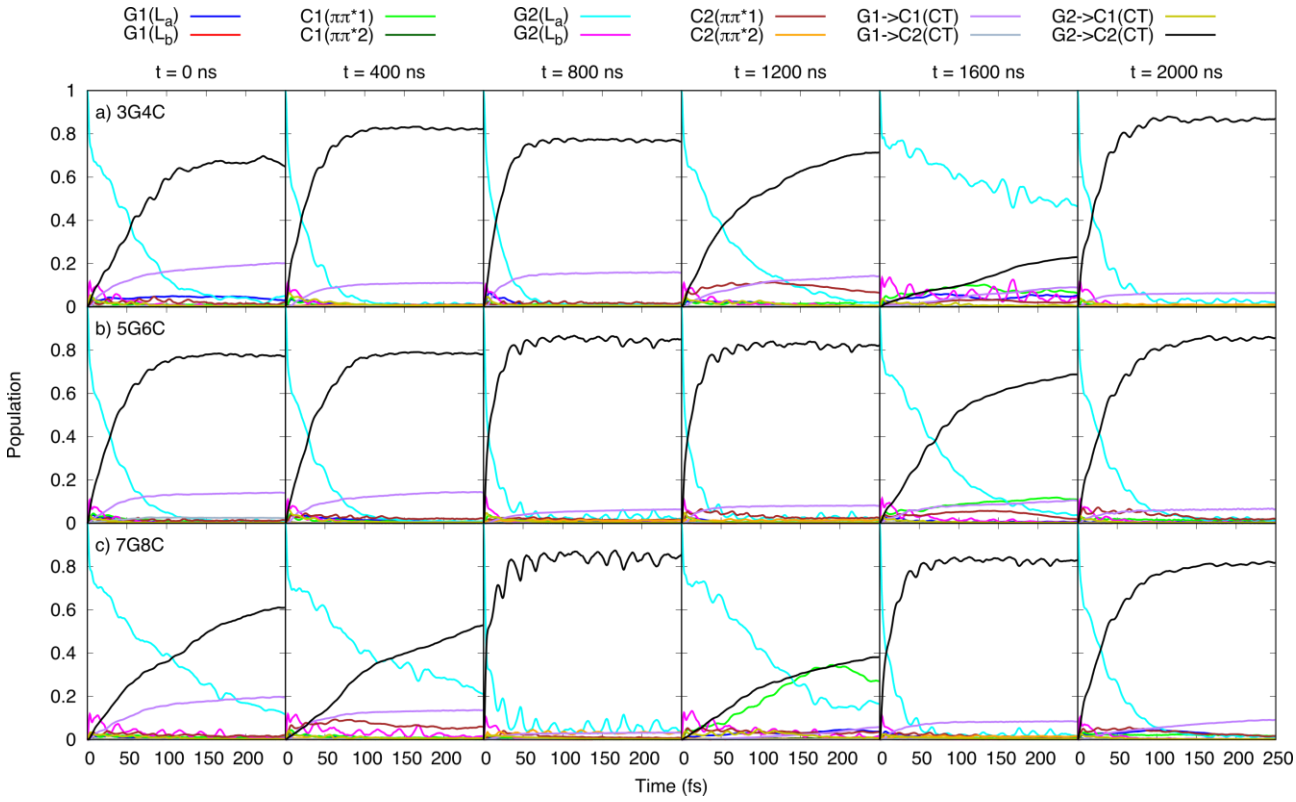



**Figure S7:** Populations following initial excitation to G2(La) for each of the GC tetrad MD snapshots.

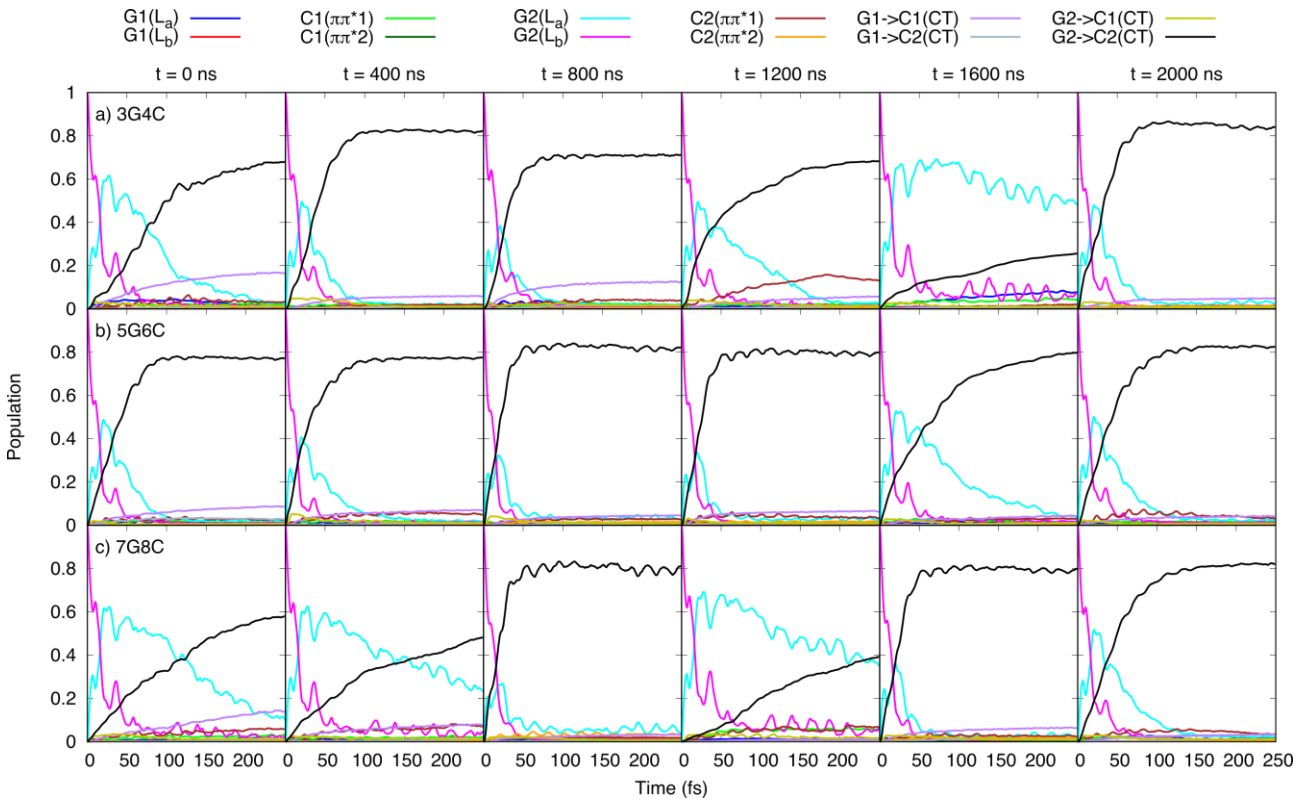



**Figure S8:** Populations following initial excitation to G2(Lb) for each of the GC tetrad MD snapshots.

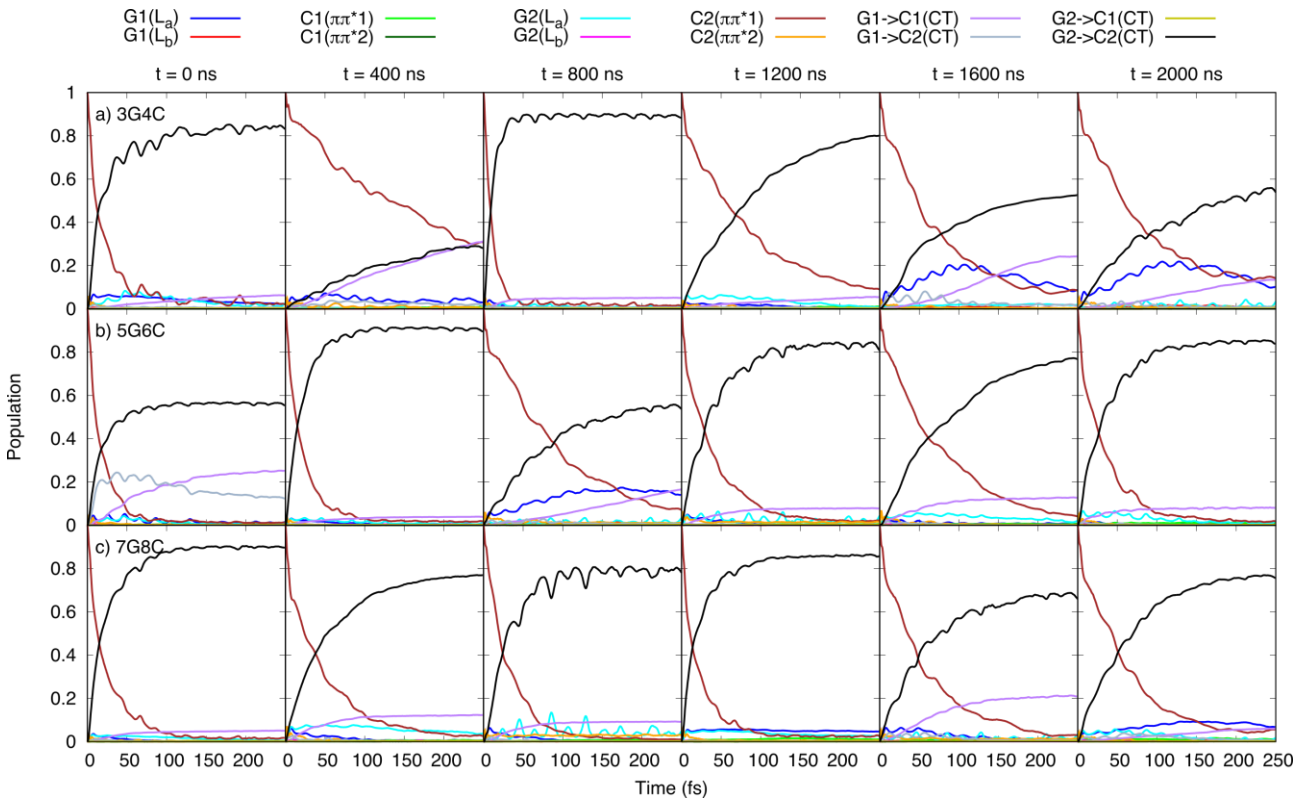



**Figure S9:** Populations following initial excitation to C2(ππ*1) for each of the GC tetrad MD snapshots.

Figures S4-S9 collect the results of the QD calculations on the MD snapshots, following excitation to the lowest 6 bright states. Rows a) b) and c) in each of the figures correspond to three different tetrads extracted from the $(GC)_5$ MD simulation, whilst the columns are different time points from the MD simulation. It can be seen that the majority of the simulations result in effective $CT^{intra}$ population on the same strand that the initially excited bright state is located. However, a few of the snapshots have a much smaller population of $CT^{intra}$, namely: 3GC4 at 1600 ns; 5GC6 at 1600 ns; and 7GC8 at 1200 ns. As can be seen in Table S1, these correspond to geometries in which there is either a small stacking overlap of the G/C bases, or greater inter-base separation. In these cases, the $CT^{intra}$ is disfavoured, and a greater proportion of the excited state population remains in the bright state. In a number of the snapshots there is also an appreciable amount of population transfer to the $CT^{intra}$ state on the other strand to the one initially excited, whilst only one snapshot (5GC6 at 0 ns) shows any transfer to a $CT^{inter}$ state.

3.4.2 CG

**Table S6** Diabatic energies $E_{ii}$ and constant couplings $E_{ij}$ from FrD-LVC model of CG at its ground state minimum.

| | C1(ππ*1) | C1(ππ*2) | G1(La) | G1(Lb) | C2(ππ*1) | C2(ππ*2) | G2(La) | G2(Lb) | G1→C1 CT | G1→C2 CT | G2→C1 CT | G2→C2 CT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C1(ππ*1) | 5.304 | | | | | | | | | | | |
| C1(ππ*2) | 0.104 | 6.221 | | | | | | | | | | |
| G1(La) | -0.021 | 0.034 | 5.305 | | | | | | | | | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G1(Lb) | 0.054 - | -0.029 | 0.082 | 5.775 | | | | | | | | |
| C2(ππ*1) | 0.012 - | -0.026 | -0.023 | 0.014 | 5.302 | | | | | | | |
| C2(ππ*2) | -0.002 | 0.012 | 0.049 | -0.000 | 0.097 | 6.258 | | | | | | |
| G2(La) | -0.021 | 0.045 | 0.021 | 0.006 | -0.002 | 0.016 | 5.335 | | | | | |
| G2(Lb) | 0.016 | -0.003 | 0.018 | 0.017 | 0.028 | -0.008 | 0.054 | 5.744 | | | | |
| G1→C1 CT | 0.011 | 0.066 | -0.093 | 0.106 | -0.001 | 0.003 | 0.001 | -0.000 | 5.301 | | | |
| G1→C2 CT | 0.000 | -0.004 | 0.080 | -0.005 | 0.036 | -0.013 | -0.001 | 0.000 | 0.054 | 5.869 | | |
| G2→C1 CT | 0.007 | 0.003 | -0.001 | 0.001 | 0.001 | 0.008 | 0.047 | 0.020 | -0.006 | 0.000 | 5.877 | |
| G2→C2 CT | -0.001 | 0.003 | 0.000 | -0.001 | -0.075 | -0.117 | -0.037 | -0.017 | 0.000 | -0.004 | 0.540 | 5.712 |

**Table S7** Averaged diabatic energies $E_{ii}$ and constant couplings $E_{ij}$ from FrD-LVC models of CG from the molecular dynamics snapshots.

| | G1(La) | G1(Lb) | C1(ππ*1) | C1(ππ*2) | G2(La) | G2(Lb) | C2(ππ*1) | C2(ππ*2) | G1→C1 CT | G1→C2 CT | G2→C1 CT | G2→C2 CT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C1(ππ*1) | 5.329 | | | | | | | | | | | |
| C1(ππ*2) | 0.104 | 6.238 | | | | | | | | | | |
| G1(La) | 0.015 | 0.015 | 5.309 | | | | | | | | | |
| G1(Lb) | 0.037 | 0.025 | 0.063 | 5.771 | | | | | | | | |
| C2(ππ*1) | 0.009 | 0.014 | 0.022 | 0.014 | 5.323 | | | | | | | |
| C2(ππ*2) | 0.013 | 0.033 | 0.050 | 0.003 | 0.103 | 6.223 | | | | | | |
| G2(La) | 0.021 | 0.047 | 0.023 | 0.008 | 0.015 | 0.013 | 5.319 | | | | | |
| G2(Lb) | 0.015 | 0.002 | 0.006 | 0.008 | 0.041 | 0.029 | 0.069 | 5.747 | | | | |
| G1→C1 CT | 0.019 | 0.076 | 0.049 | 0.041 | 0.000 | 0.001 | 0.001 | 0.001 | 5.623 | | | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G1→C2 CT | 0.001 | 0.002 | 0.073 | 0.007 | 0.026 | 0.018 | 0.001 | 0.001 | 0.028 | 5.979 | | |
| G2→C1 CT | 0.027 | 0.016 | 0.001 | 0.000 | 0.001 | 0.002 | 0.069 | 0.004 | 0.042 | 0.000 | 5.945 | |
| G2→C2 CT | 0.000 | 0.001 | 0.001 | 0.001 | 0.025 | 0.080 | 0.054 | 0.049 | 0.000 | 0.040 | 0.028 | 5.585 |

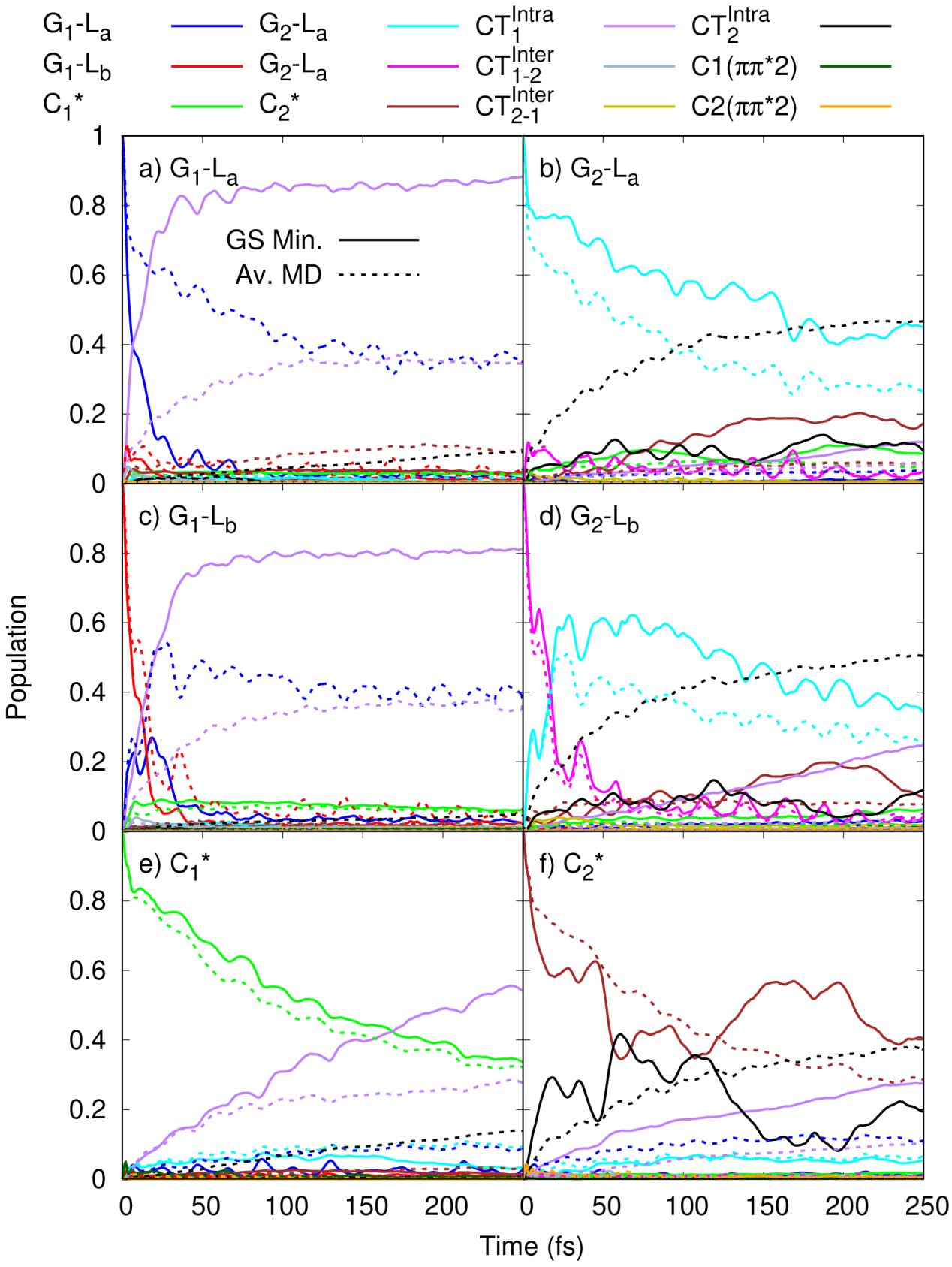



**Figure S10:** Full version of Figure 4 from the main text, including all diabatic states.

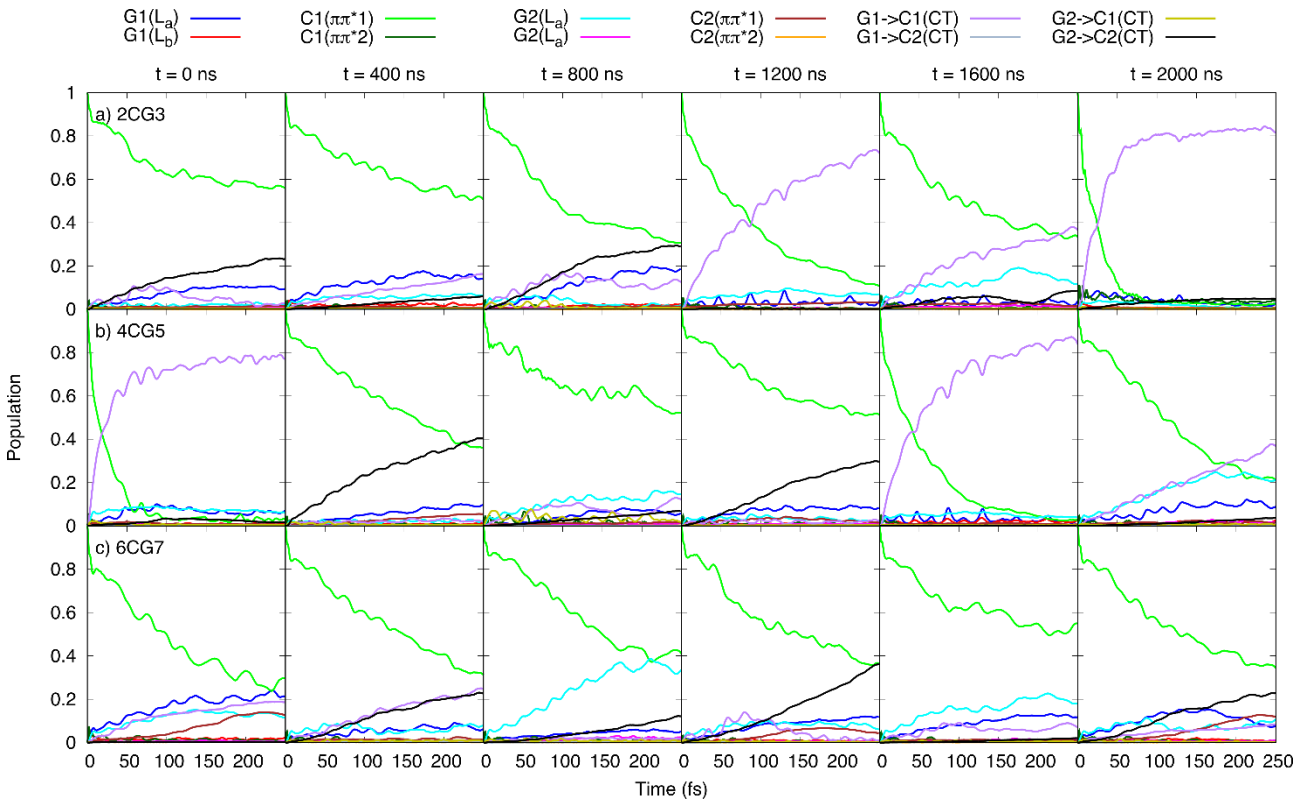



**Figure S11:** Populations following initial excitation to C1(ππ*1) for each of the CG tetrad MD snapshots.

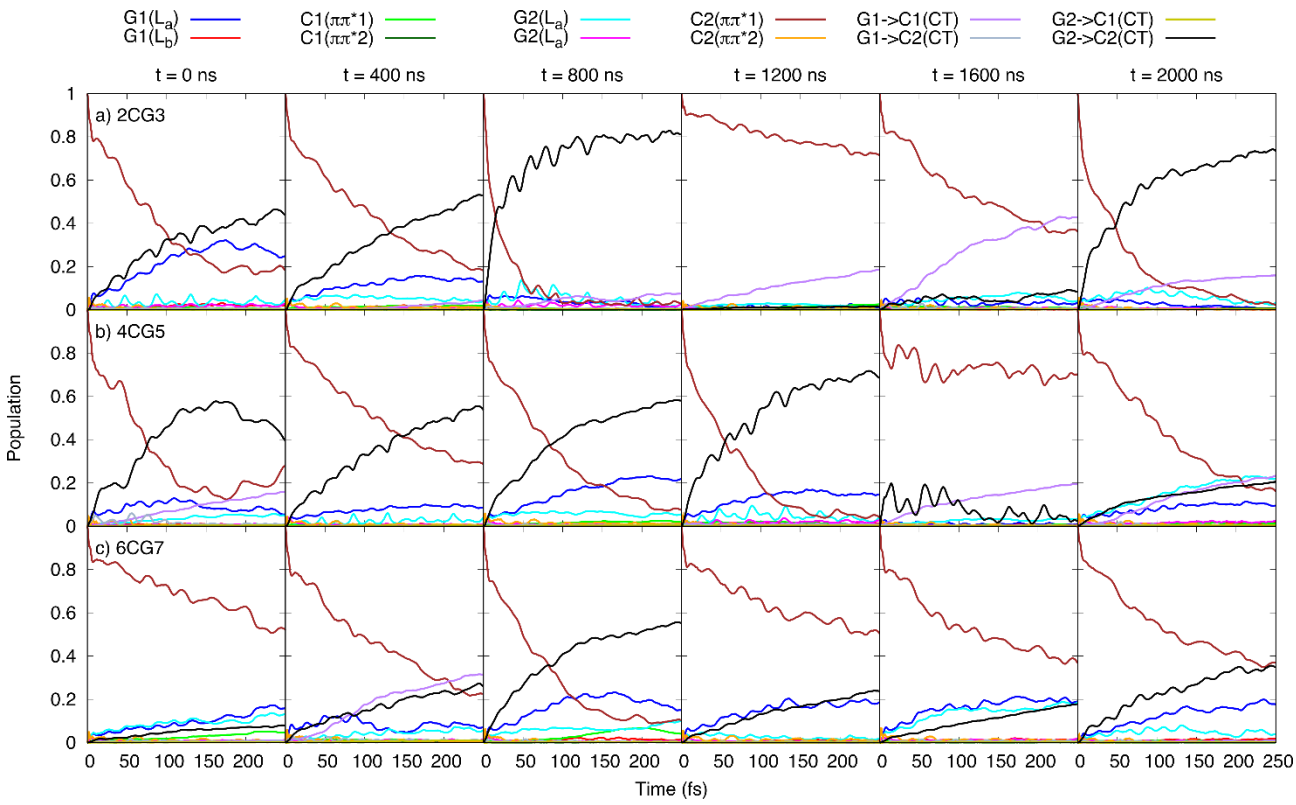



**Figure S12:** Populations following initial excitation to C2(ππ*1) for each of the CG tetrad MD snapshots.

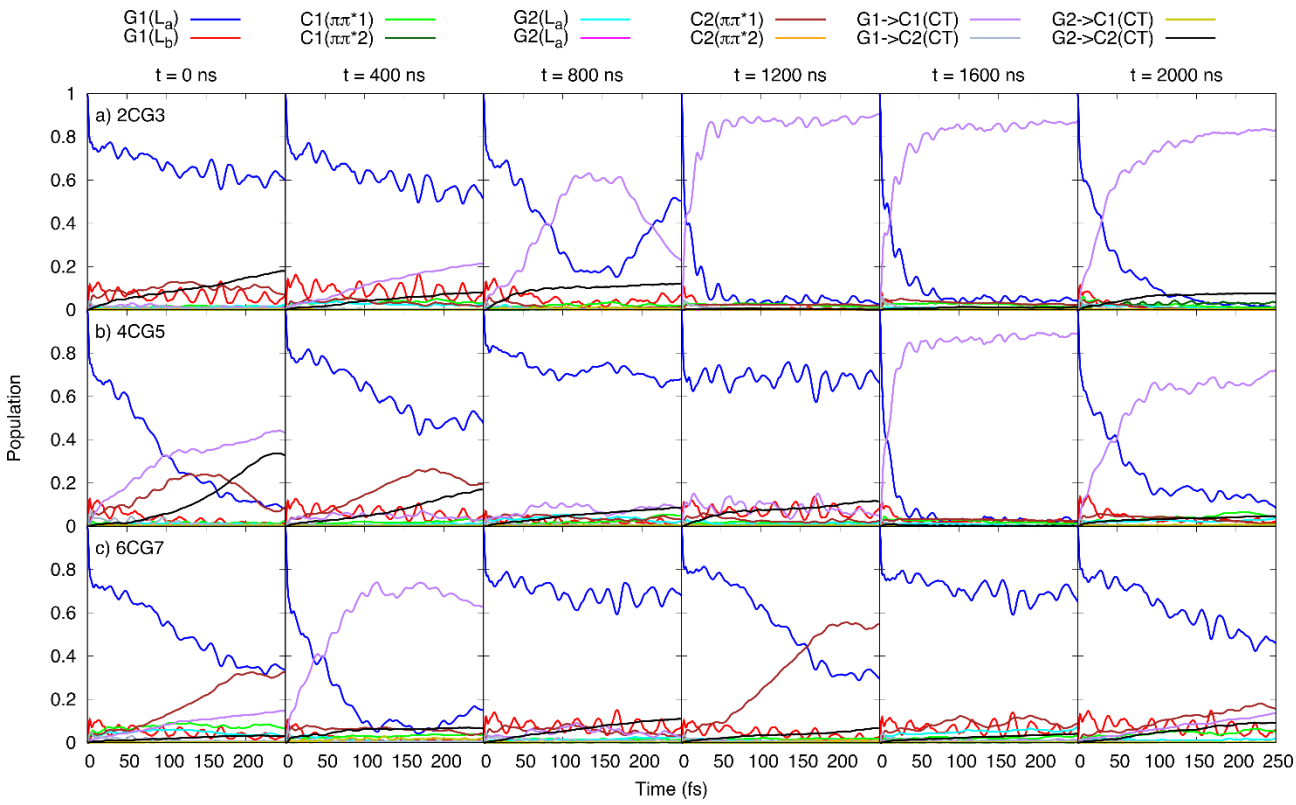



**Figure S13:** Populations following initial excitation to G1(La) for each of the CG tetrad MD snapshots.

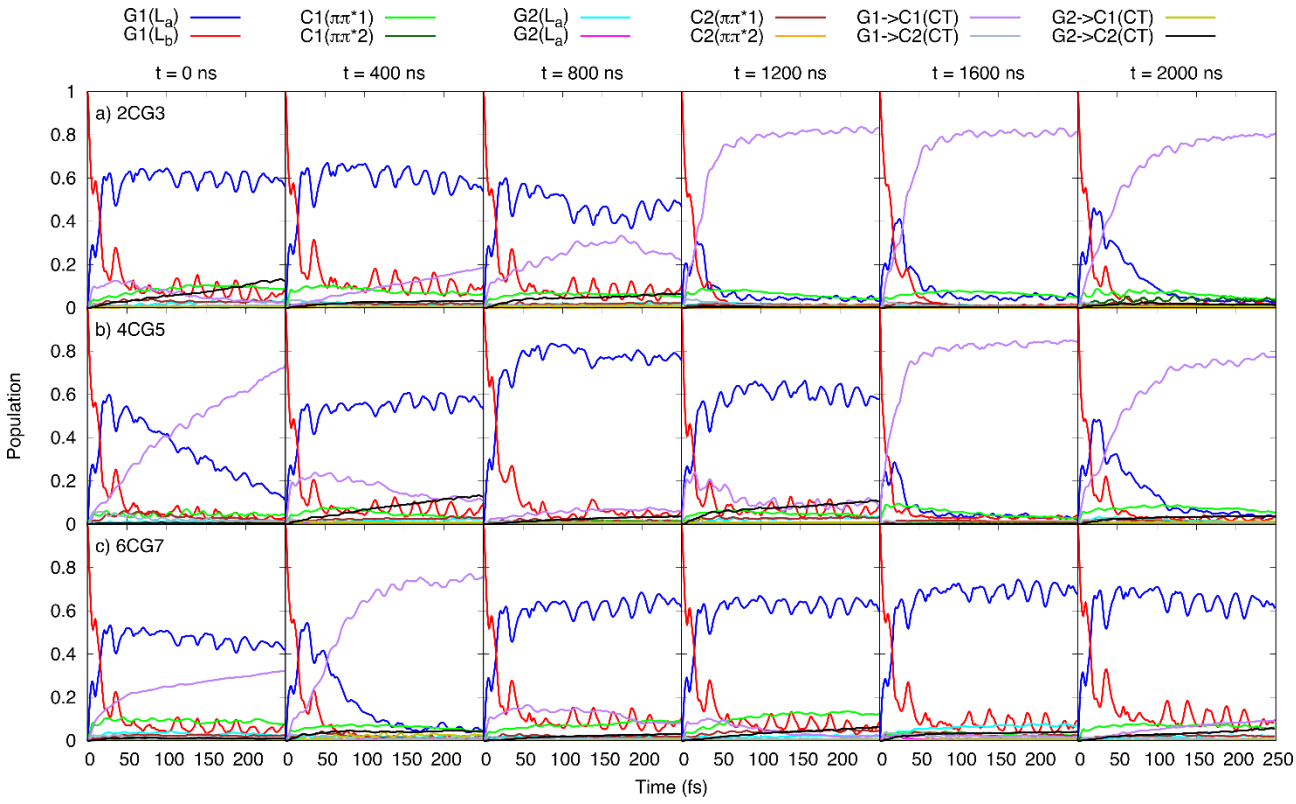



**Figure S14:** Populations following initial excitation to G1(Lb) for each of the CG tetrad MD snapshots.

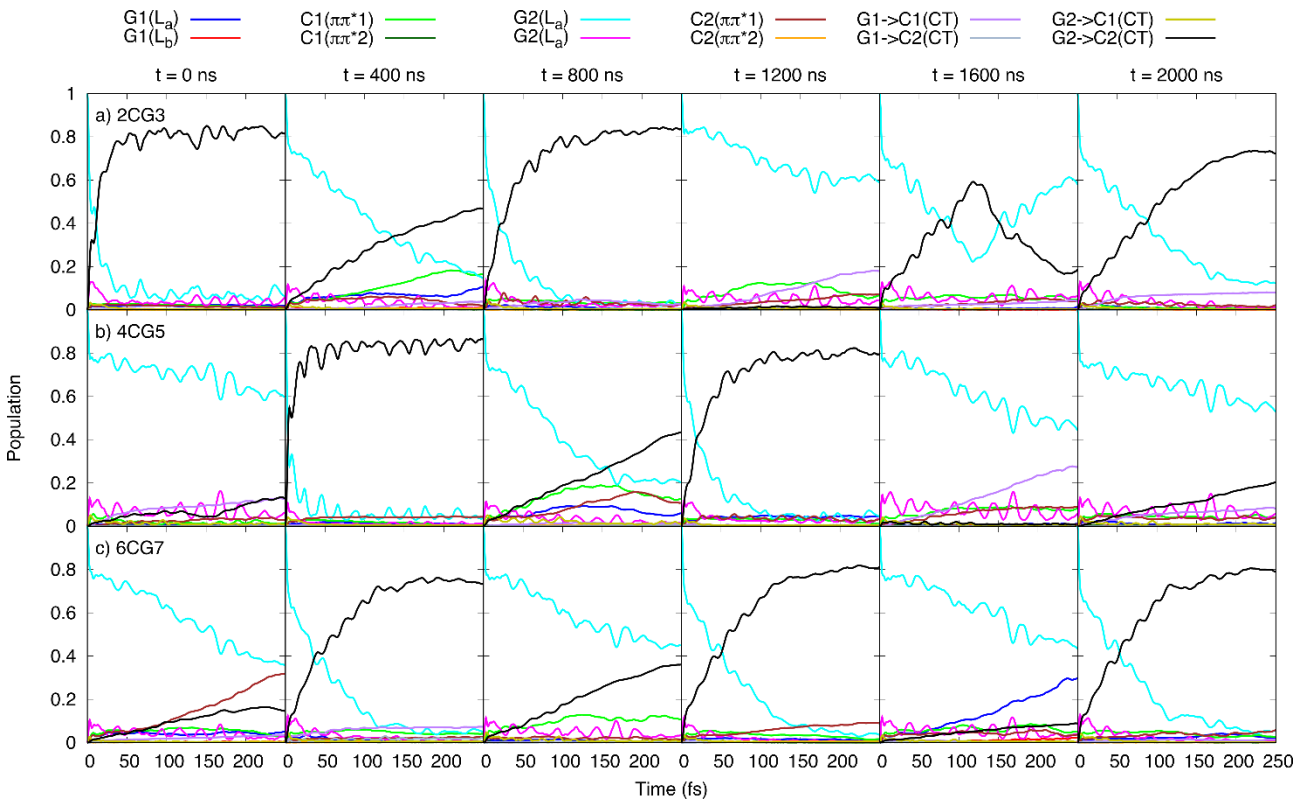



**Figure S15:** Populations following initial excitation to G2(La) for each of the CG tetrad MD snapshots.

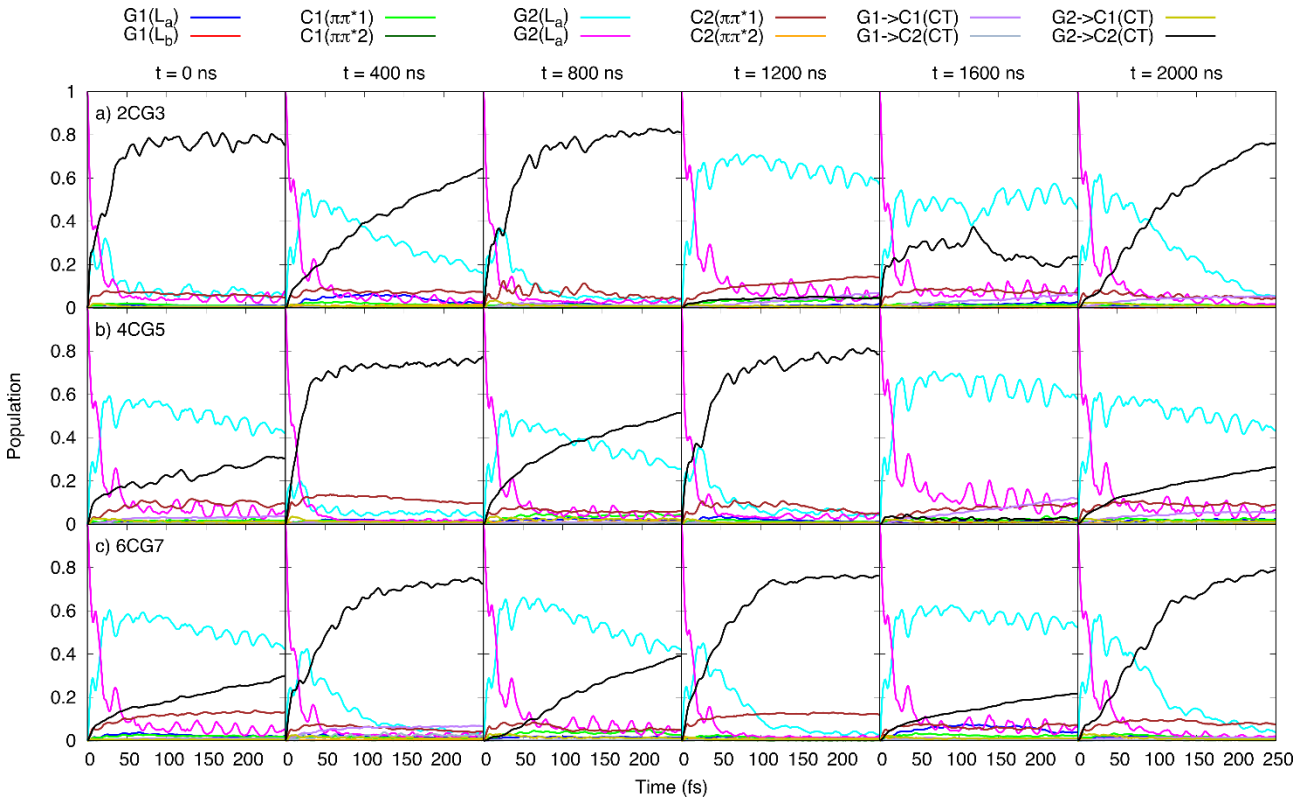



**Figure S16:** Populations following initial excitation to G2(Lb) for each of the CG tetrad MD snapshots.

3.5. Excited state geometry optimizations

$S_1$ /$S_2$ Both for 2GC (Figure 5) and 2CG geometry optimizations (Scheme S3) of $S_1$ and $S_2$ predict that the CT character of the transition further increases with respect to the FC region, until a minimum with strong intra-strand CT character is reached ($CT^{intra}$-min, shown in Figure S18). The relative energies (~0.03 eV) of the two $CT^{intra}$-min are similar, as well as their geometries. The two bases involved in the transition exhibit the geometry of a $G^+$ cation and $C^-$ anion, respectively, and get closer to each other with respect to the GS, to increase the electronic coupling and their electrostatic interactions. A moderate change in the propeller value of the 'nominally' $C^-$ moiety, in order to maximize the HB interaction with the amino-group of the adjacent $G^+$ amino group, is also predicted. At the same time, the 'complementary' strand, i.e. the one not involved in the CT transition, undergoes small geometry shifts to accommodate the electron density of the CT state. The HB interactions where $G^+$ acts as a donor ($NH_2$ and NH) and those where $C^-$ acts as an acceptor become stronger, and vice versa. We therefore observe noticeable changes in lengths of all HBs of the duplex (see Table S8).

Not surprisingly, the most significant differences between the vibrational modes of $S_0$ and $CT^{intra}$-min located by optimizing $S_2$ involve $G^+$ and $C^-$ bases, namely: (i) The blue-shift of the vibrational modes associated to the CO stretching of $G^+$, at 1715 $cm^{-1}$. This value is very close to the features (1703 $cm^{-1}$) appearing in GMP monomer and poly(dGdC) after ionization, indeed assigned to $G^+$.[4] This change is in line with the shorter G(C2-$NH_2$) distance, witnessing the larger delocalization of the amino LP in the charge deficient cation (ii) The blue-shift by 25~30 $cm^{-1}$ of the two vibrational frequencies mainly associated to $G^+$ ring stretching modes, which leads to the appearance of a band at ~1660 $cm^{-1}$. Interestingly, a positive peak just above 1600 $cm^{-1}$ appears in the difference IR spectra associated to G ionization both for isolated guanine and poly(dGdC).[5] (iii) The vibrational modes associated to $C^-$ significantly red-shifts, by ~70 $cm^{-1}$ that associated to C2-CO stretching, whereas several vibrational modes in the range 1400-1600 $cm^{-1}$ have contribution from the C2[C5=C6] stretching. $C^-$ exhibits a significant lengthening of the C5-C6 and, to a lesser extent, C4-O bond distances, since the additional electron occupies an antibonding orbital with respect to those bonds. (iv) Overall, the intensity of the vibrational modes of G and C, which are not involved in the CT states, noticeably decreases. In the ground state, there is a larger coupling between the different vibrational modes, more mixing, and a larger intensity, which decreases in the excited state, where we have 4 different bases (G,C,$G^+$ and $C^-$) instead of two (G and C) with similar IR frequencies. We also observe very small shifts, ~10 $cm^{-1}$, in the frequencies, for example, of the CO stretching modes of C, which is WC bonded with the $G^+$.

Concerning the $CT^{intra}$-min located by optimizing $S_1$, we obtain a similar spectrum to that described above, but with a larger peak intensity at 1600 $cm^{-1}$, associated to the ring stretching modes $G^+$, and larger red shifts (by additional 20 $cm^{-1}$) of the CO stretching modes of C, which is WC bonded with the $G^+$.

$S_3$

Geometry optimization of $S_3$ leads to further localization of the excitation on G (G*). In the first, steep, part of path we observe mainly changes in the bond distances and bond angles, while the purine keeps a planar geometry. Then, a crossing between G* and the close-lying $CT^{intra}$ state is found. If the system stays on G*, now on the $S_1$ PES, distortion of the purine planarity and out-of-plane motion of the amino group is predicted, as it happens for isolated monomer (dG). For dG this motion leads, without any energy barrier, to a Conical Intersection (CI) with $S_0$, in line with an excited state lifetime <1 ps, with a fast, and predominant, component of 0.13 ps. In DNA, instead, a very distorted minimum is found (G*-min, see Figure S17-18). We have verified that, also within the duplex, further pyramidalization at C2 atom of G*-min and out-of-plane motion of the amino group (Figure S17), leads to a crossing region with $S_0$, since the energy gap between G* and $S_0$ is smaller

than 1 eV. In 2GC this distortion can occur in two different directions: the 'external' one, i.e. towards the solvent exposed part of the duplex (see Figure S17 left), or the 'internal one', towards the other WC base pair (see Figure S17 right). In both cases, the crossing region is more stable than G*-min, showing that this process is exo-ergonic as for monomer. An energy barrier is however present in this path. For the external path the barrier is small (≤0.1 eV), whereas the internal path, as it could be expected, is substantial (~0.5 eV). This latter path should be representative of the situation found in $d(GC)_9$, where most of the base pairs are embedded in the duplex. These are only rough estimates, since the energy barrier is expected to strongly depend also on the conformational fluctuations, not included in our treatment, of the duplex, leading, for example, to partial opening of the GC pairs. However, our analysis indicates that the non-radiative decay from G*-min to $S_0$ is possible, but that this decay route is slower, though not dramatically, than for dG in water. As a matter of fact, G*-min is only 0.5 eV more stable than $CT^{intra}$, and also this energy gap would strongly depend on duplex fluctuations. Part of the population trapped on G*-min, because of the barrier experienced within the duplex, could always decay to $CT^{intra}$, not only during their first crossing, but also on a slower timescale.

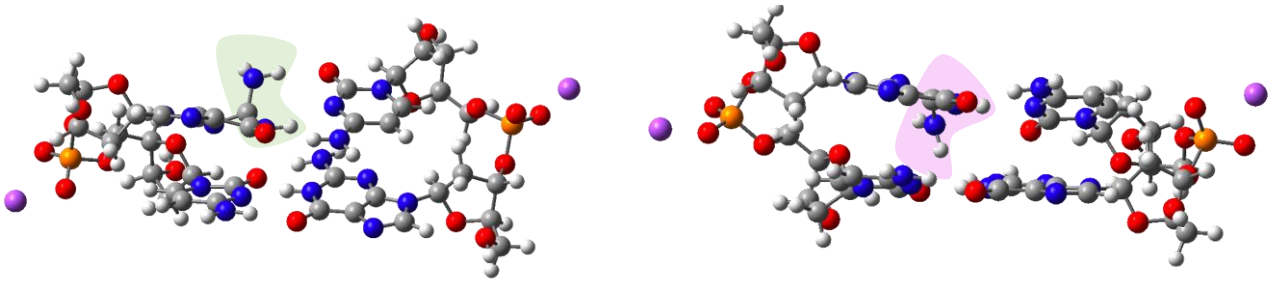


**Figure S17**. Geometries for the degeneracy points between $S_1$ and $S_0$.

Concerning the IR spectrum of G*-min, the small lengthening of CO bond distance is mirrored (i) by a red-shift and a decrease of the intensity of the associated stretching frequency, contributing, coupled with the C's vibrational modes, to an intense peak at 1640 $cm^{-1}$. Analogously (ii) the peaks associated to the coupled ring stretching modes of G's red-shift by ~10 $cm^{-1}$ and decrease in intensity. Finally (iii) another peak with predominant contribution from G1* ring stretching appears at 1510 $cm^{-1}$.

$S_4$

As anticipated above, in the FC region, $S_4$ (and $S_5$) are mainly localized on the C bases, in an excited state similar to the lowest energy $\pi\pi^*$ state of isolated (C*), a HOMO/LUMO transition, which is bonding/antibonding with respect to the C5=C6 double bond. Also in this case, geometry optimization of $S_4$ predicts a more complete localization on a single C and, consequently, a decay path similar to that found in 'free' C. First, a steep path leads to a low-energy gradient region, where the pyrimidine ring keeps a structure close to planarity and the main distortion with respect to the FC geometry involves significant lengthening of the C5-C6 bond. A representative point in this planar plateau (energy gradient < 0.001 a.u.), hereafter labeled as C*-min, is shown in Figure S18. C*-min belongs to the PES of the $S_2$ adiabatic state, and it is very close in energy to $CT^{intra}$. The non-radiative ground state recovery from C*-min in DNA then follows a mechanism similar to that of isolated cytosine in water. H6 and H5 atoms move out from the molecular plane, and then pyramidalization at C6 leads to a CI with $S_0$. In free C the energy barrier associated to this path is very small, allowing ground state recovery within ~1 ps. An exact estimate of the energy barrier in DNA is not possible, due to the proximity of the CT state, but our analysis of the PES shows that this path is not hindered by the duplex and that, therefore, the energy barrier is not significantly larger than for the monomer.

The most peculiar features of C*-min IR spectrum, when compared to $S_0$, are (i) the peak at 1611 $cm^{-1}$ associated to C* CO stretching and (ii) a less intense one at 1430 $cm^{-1}$ associated to C* ring stretching. These peaks are strongly red-shifted and less intense with respect to those (falling at ~1650 $cm^{-1}$) associated to C vibrations.

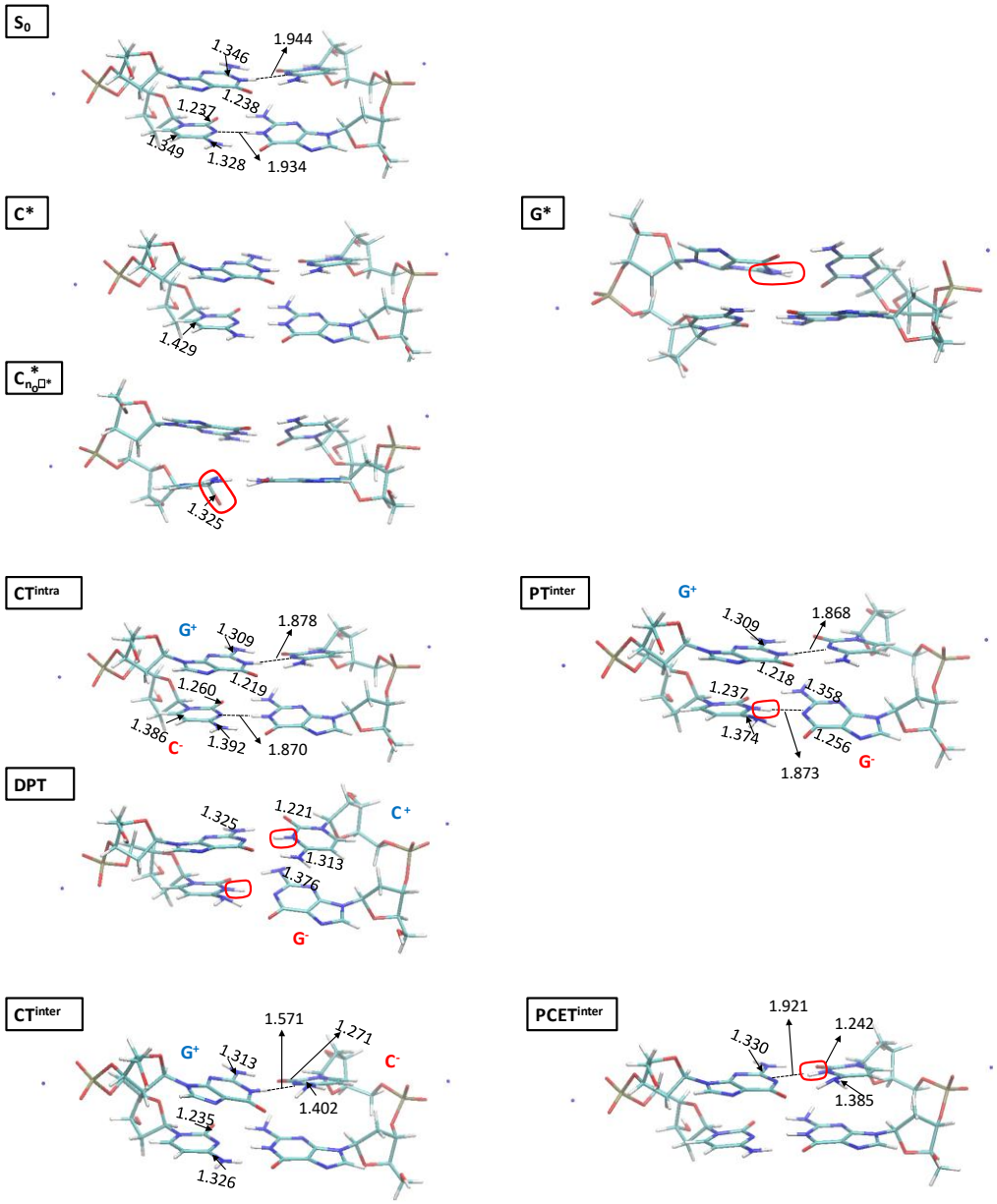
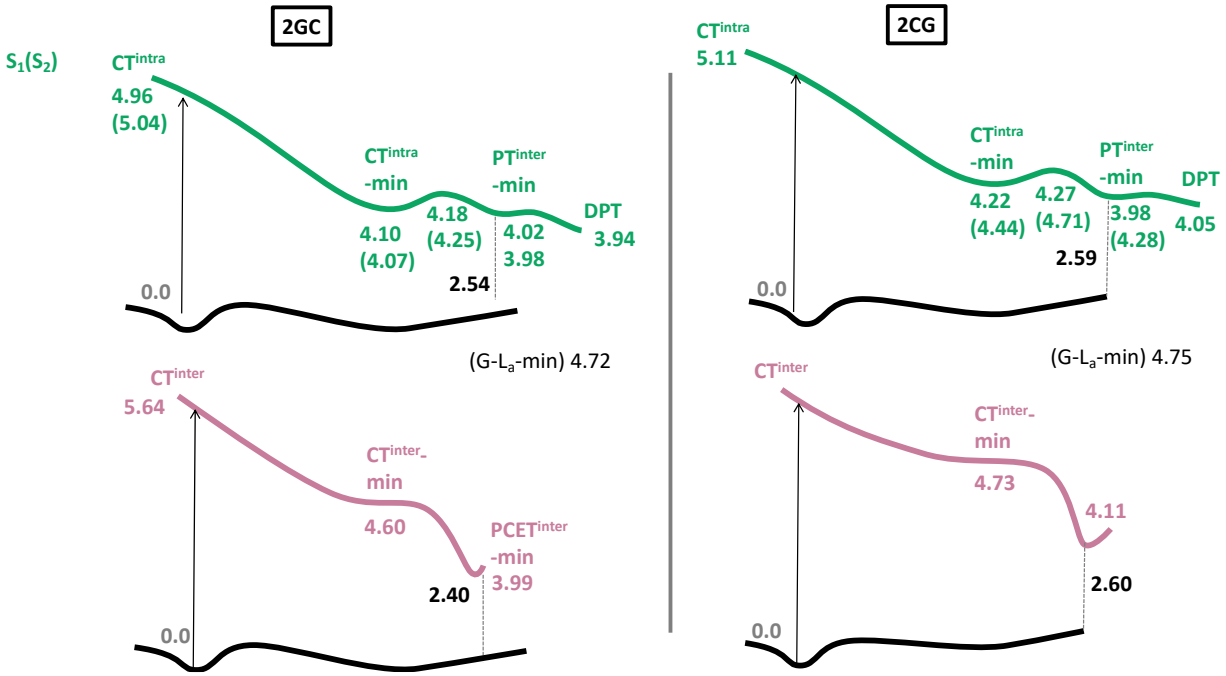



**Figure S18.** Optimized geometries for the ground and excited states minima. Distances in angstroms.




**Scheme S3**. Schematic description of the potential energy surfaces for the CT and PT reactions in 2GC and 2CG. Adiabatic energies in green and pink (in eV) and Vertical Energies (in eV) in black. LR-PCM. For The $CT^{intra}$ mechanism the energy values outside the parenthesis correspond to $S_{1min}$ reached when optimizing the $S_1$ at FC, whereas the values inside the parenthesis correspond to $S_{1min}$ reached when optimizing the $S_2$ at FC.

*$S_5$-$S_6$* For these two states the shape of the PES is very similar to that described above for $S_3$ and $S_4$, and our geometry optimizations predict that excitation localizes on a single base (G-$L_a$ or C* excited states), whose bond lengths/angle exhibit the only significant geometry shifts. In addition to the IR spectra of the G-$L_a$-min and C*-min, we computed also the IR spectrum of the n$\pi$* state, involving the L.P. carbonyl of **C**, which has been shown to be populated in isolated **C**.[32]

*$S_7$*. In 2GC $S_7$ is the lowest energy state in the FC region with significant inter-strand CT character ($CT^{inter}$). Its geometry optimization first leads to a region where two bases on the opposite strands have geometries similar to a $G^+$ cation and $C^-$ anion, stabilized by an intra-strand hydrogen bond between the carbonyl group of $C^-$ and the amino group of the adjacent G. The gradient in this region is very low (< 0.001 a.u.in $CT^{inter}$-min in Figure S17), but, as discussed below, its geometry optimization eventually leads directly to the PT reaction sketched in Figure 1 in the main text. We have located the CT-inter-min also in 2CG, also in this case, there is a portion of the PES with very low gradient (<0.001 a.u.). We have used this structure ($CT^{inter}$-min) to estimate the IR vibrational feature associated to this excited state. As discussed in the SI, in 3GC CT-inter–min is a minimum of the PES (gradient 0.0006). This minimum involves the above-mentioned changes towards the $C^-$ and $G^+$ geometries and although the PT has not yet taken place, the distance between the C and H-G decreases to ~1.7 Å.

*Inter-strand PT reactions.* As discussed in detail in ref.[26], starting from $CT^{intra}$-min, in 2GC a relaxed path along the PT coordinate involving the proton of the G (H1) hydrogen bonded with the N3 nitrogen atom of anionic $C^-$ leads to a stable minimum $PT^{inter1}$-min.[26] In $PT^{inter1}$ G has lost a proton $[G\text{-}H1]^-$ and acquired a formal negative charge, whereas C has acquired a proton $[C+H3]^{\bullet}$. The paths in the two strands are similar, and are characterized by a small energy barrier, ~0.2 eV (at the LR-PCM level) or ~0.3 eV (at the State Specific PCM

level).[26] This process is slightly exoergonic as $PT^{inter1}$-min is more stable than $CT^{intra}$-min by 0.05~0.1 eV, depending on the level of calculation and on the model adopted. The energetics of this process are not significantly modulated by changing the effective dielectric constant experienced by the duplex, decreasing it up to $\varepsilon$=4, to mimic a pair embedded in a larger duplex. On the other hand, we expect that this PT reaction, which involves the shift of a charge between the two strands, can be affected by the local conformational behavior of the duplex, e.g., inter alia, the presence of the charged moieties (phosphate groups, counter-ions etc.).

Some of these considerations apply also to another PT reaction that can occur in $CT^{intra}$ and $PT^{inter1}$, i.e. the one involving the positively charged $G^+$ base, which can lose its N1 proton to the neutral C in the same base pair. This process also leads to a stable minimum, where the positive charge shifts to the C base. If we start from $PT^{inter1}$, we end in a minimum where a double proton transfer reaction (DPT) has occurred following a single ET event. $PT^{inter1}$-min and DPT-min have a very similar energy, their relative stability depending on the level of calculation. At the LR-PCM level DPT-min is slightly more stable (by ~0.1 eV), whereas the opposite is predicted by SS-PCM calculations (which should be more accurate). In 2CG we obtain a similar picture, starting from CT-intra1 $PT^{inter1}$-min is more stable by 0.2 eV than $CT^{intra}$-min, and by only 0.05 eV than DPT-min.

The predictions of our calculations are fully consistent with estimates based on the experimental $pK_a$ of the different bases, with $G + C^{\bullet -} \rightarrow G(-H1)^- + C(+H3)^\bullet$ being strongly exergonic ($K = 10^{+3.4}$), while the lower driving force for PT in $G^{\bullet +} + C \rightarrow G(-H1)^\bullet + C(+H3)^+$ ($K = 10^{+0.55}$) is expected to result in an equilibrium between reactants and products.

As anticipated above, starting from $CT^{inter}$-min, LR-PCM geometry optimization predicts a barrierless $G^+ \rightarrow C^-$ PT reaction, involving the H1 atom of G, leading to $PT^{inter2}$-min (Figure 1). $PT^{inter2}$-min is significantly more stable than $CT^{inter}$-min and has a similar energy to $PT^{inter1}$-min. However, this PT reaction leads to the formation of two neutral radicals, and the large dipole moment of $CT^{inter}$-min strongly decreases. This reaction is thus affected by proper inclusion of dynamical solvation effect, which we do via by SS-PCM/TD-M052X calculations. At this latter level, an energy barrier (0.20~0.40 eV) is associated to PT. The PT driving force also decreases. At the SS-PCM/TD-M052X level, full equilibration of solvent degrees of freedom makes $PT^{inter2}$-min very close to the Ground state (energy gap ~0.4 eV), suggesting the proximity of a crossing with GS.[26]

At the LR-PCM/TD-DFT level, in 3GC the $PT^{inter2}$-min is also more stable compared to $CT^{inter}$–min by ~0.6 eV, and is only ~2 eV less stable than GS. At the SS-PCM/TD-DFT level the energy difference between the two minima decreases up to 0.27 eV and $PT^{inter2}$-min is only 0.6 eV less stable than GS, i.e. close to a crossing region between the two states. However, already at the LR-PCM level, $CT^{inter}$ does not decay directly to $PT^{inter2}$, suggesting that, in this case, the inclusion of the GC pair duplex can slow down the reaction, which, in any case, would be severely affected by the slow degrees of freedom of the duplex.

**4. Additional 2•GC and 3•GC spectra**

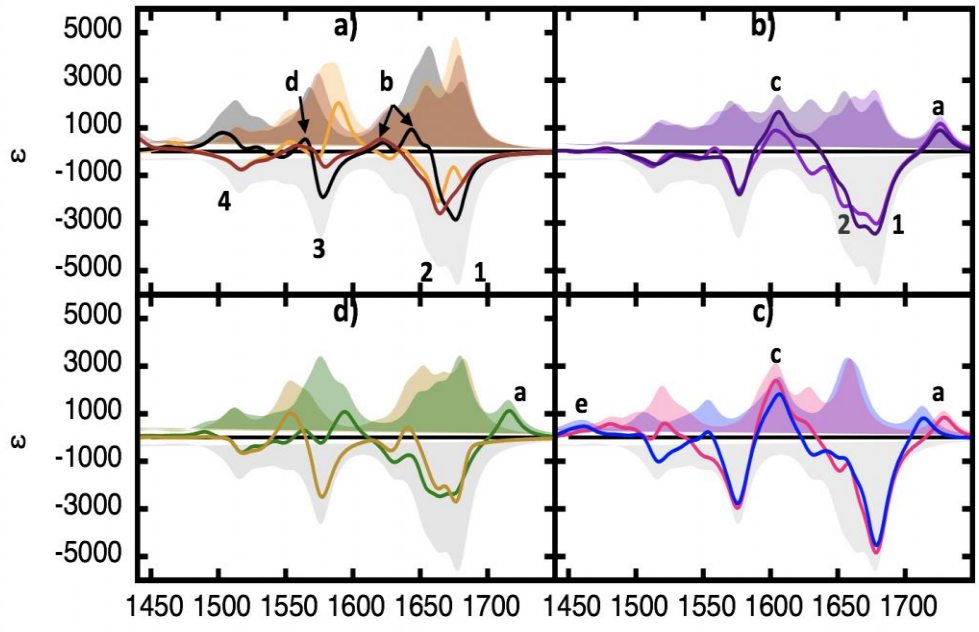



**Figure S19.** Excited state (shaded curves) and difference spectra ($S_n$-$S_0$, solid lines; the inverted $S_0$ spectrum is the light grey shaded curve) computed in **2GC** for a) G* (black), C* (brown) and C*$n_O\pi$* (orange) b) $CT^{intra}$-1 (dark purple) and $CT^{intra}$-2 light purple), c) $PT^{inter1}$ (pink) and DPT (blue), d) $CT^{inter}$ (green) and $PT^{inter2}$ (ochre).

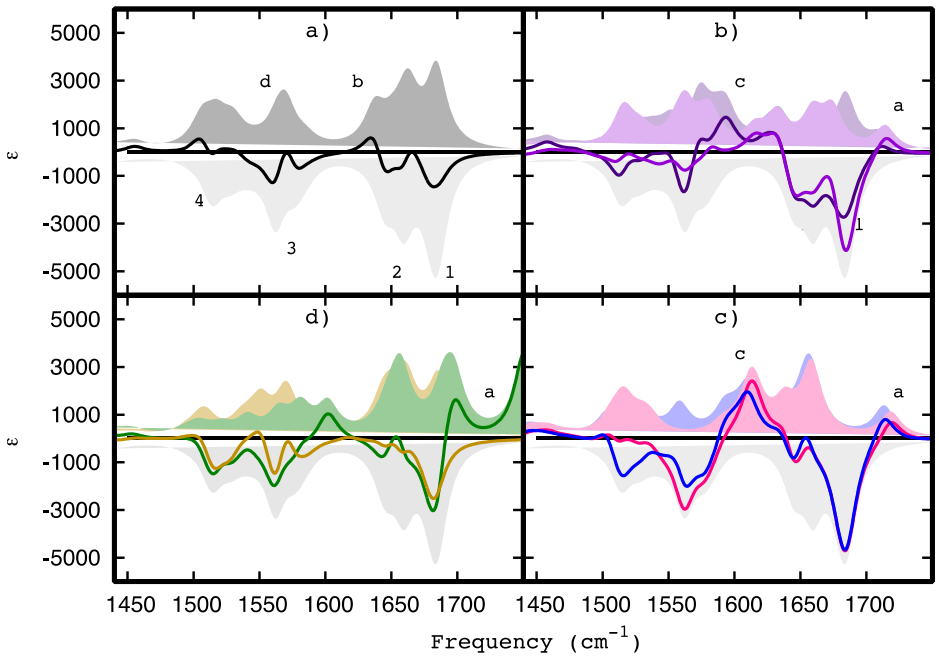



**Figure S20.** Excited state (shaded curves) and difference spectra ($S_n$-$S_0$, solid lines; the inverted $S_0$ spectrum is the light grey shaded curve) computed in **2CG** for a) G* (black), C* (brown) and C*$n_O\pi$* (orange) b) $CT^{intra}$-1 (dark purple) and $CT^{intra}$-2 light purple), c) $PT^{inter1}$ (pink) and DPT (blue), d) $CT^{inter}$ (green) and $PT^{inter2}$ (ochre).

4.1 Unshifted spectra

The unshifted IR spectra computed for all the main minima optimized are shown in the Figures below.

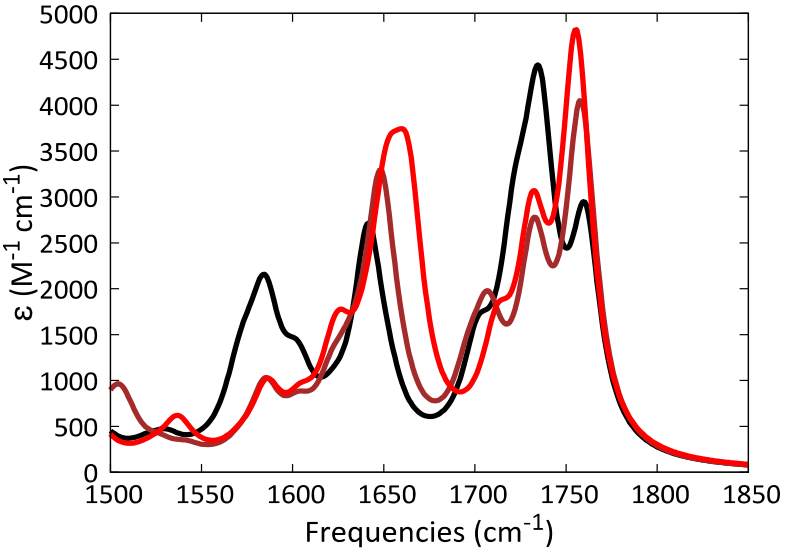



**Figure S21.** Unshifted IR spectra computed for G* (black), C* (brown) and C*$n_O\pi^*$ (red) in 2GC.

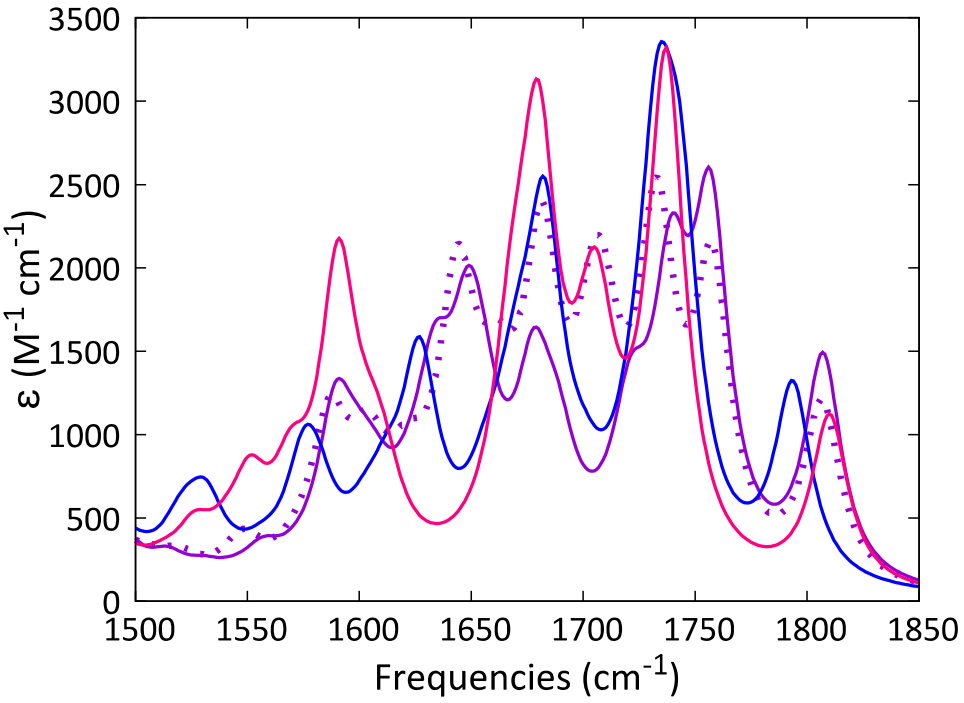



**Figure S22**. Unshifted IR spectra computed for $CT^{intra}$ ($S_2$ solid, $S_1$ dashed, purple), $PT^{inter1}$ (pink) and DPT (blue) in 2GC.

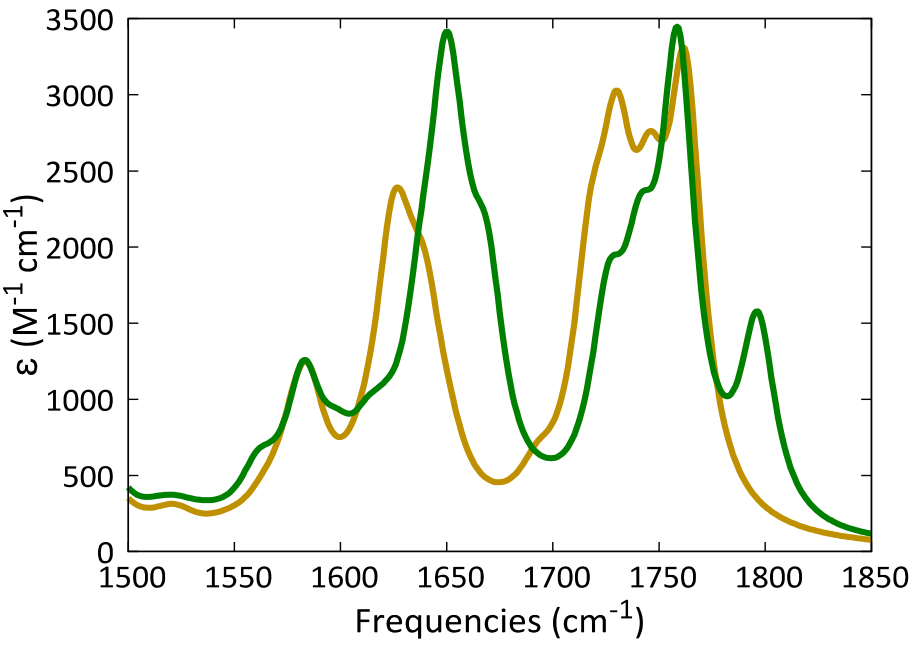



**Figure S23.** Unshifted IR spectra computed for $CT_{inter}$ (green) and $PT^{inter2}$ (ochre) in 2GC.

4.2 Comparison between 2•GC and 3•GC

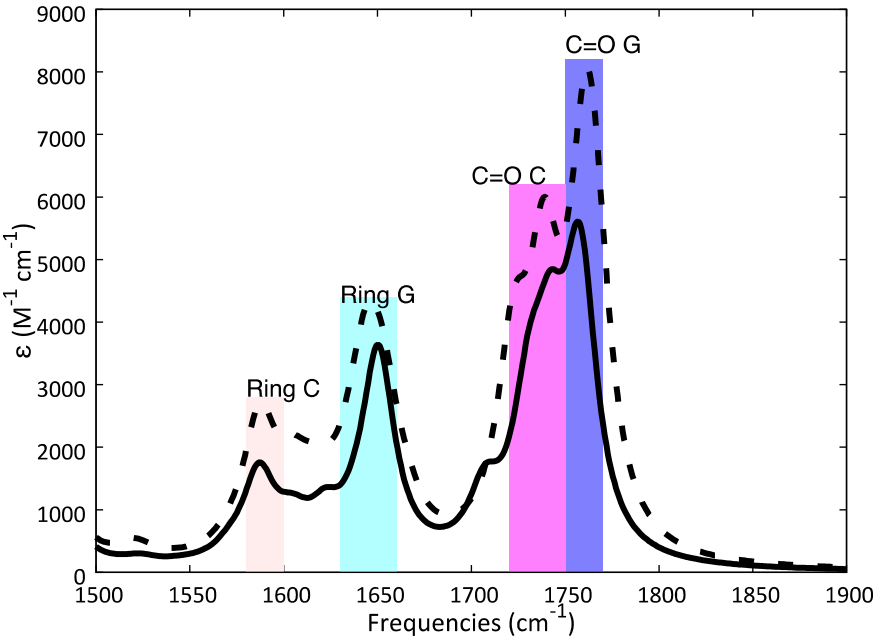



**Figure S24.** Ground State (unshifted) spectra for 2GC (solid line) and 3GC (dashed line). PCM/M052X/6-31G(d) calculations.

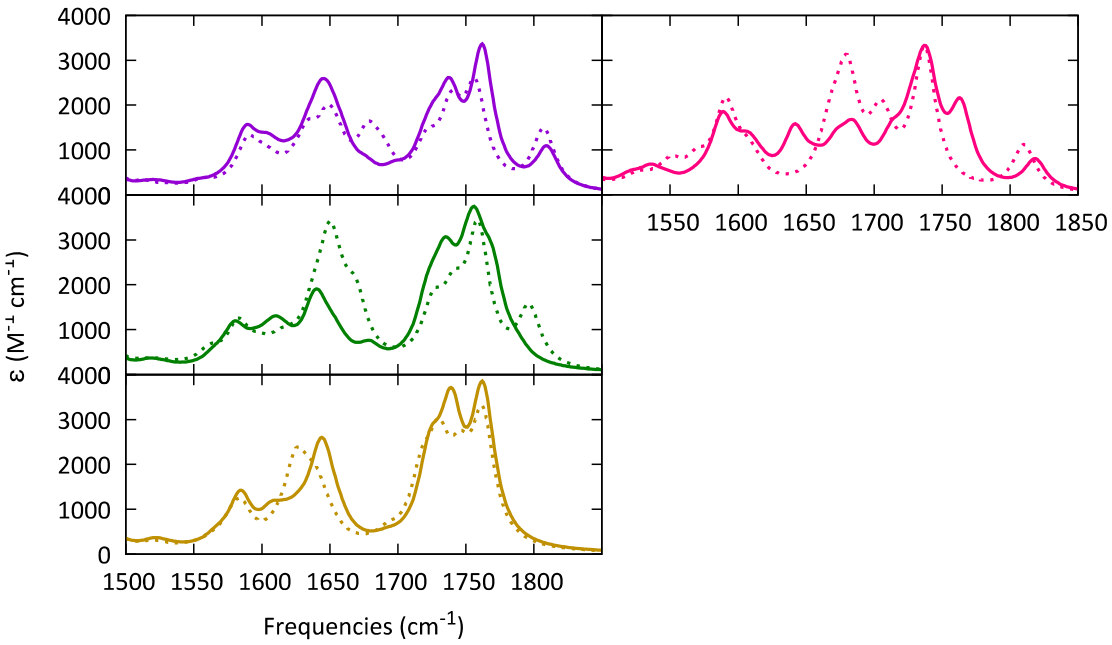



**Figure S25.** Comparison between 2GC (dashed) vs 3GC (solid) unshifted $S_1$ minima spectra for $CT^{intra}$ (purple), $PT^{inter1}$ (pink), $CT^{inter}$ (green) and $PT^{inter2}$ (ochre)

The IR spectra computed for $CT^{intra}$ (purple) $CT^{inter}$ (green) and $PT^{inter2}$ in 2GC and in 3GC are rather similar (see Figure S26), justifying our analysis, which is focused mainly on 2GC. The most significant differences do not concern the position of the 'signature peaks' of the bases involved in the electronic transitions, but their intensity, which, on average, is smaller in 3GC. This result could be, at least partially, due to the overestimation of the IR intensities by continuum solvation models.[6] In 3GC the excited bases are more embedded in the duplex and, therefore, less sensitive to this problem than in 2GC. For $CT^{intra}$, the peaks associated with the $G^+$ stretching fall at similar frequencies in 2GC and 3GC, just above 1600 $cm^{-1}$, but in the former system their intensities are twice as large. For $CT^{inter}$, the feature at 1707 $cm^{-1}$ associated with the CO stretching of $G^+$ is shifted by only 7 $cm^{-1}$ with respect to that found in 2GC, but it is three times less intense and a similar trend is found for the $G^+$ stretching at 1600 $cm^{-1}$. As a consequence, the DIR spectra computed for 3GC (Figure S26) are qualitatively similar to the corresponding ones of 2GC, but the relative intensity of the positive feature is smaller. Moreover, in 3GC the ratio between bases involved in the electronic transition and those in the ground electronic state is smaller than in 2GC.

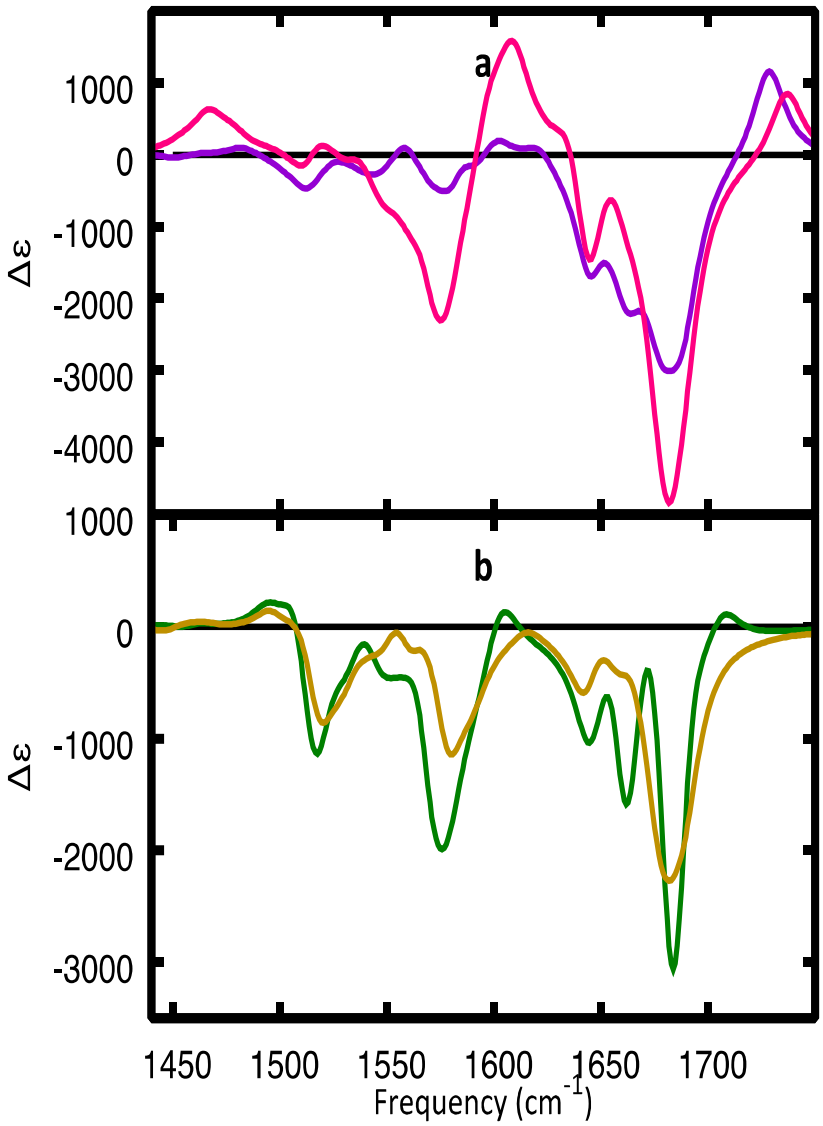



**Figure S26.** Differential spectra ($S_n$-$S_0$, solid line) computed for a) $CT^{intra}$ (purple) and $PT^{inter1}$ (pink) and b) $CT^{inter}$ (green) and $PT^{inter2}$ (ochre) in 3GC.

4.3 Additional spectra

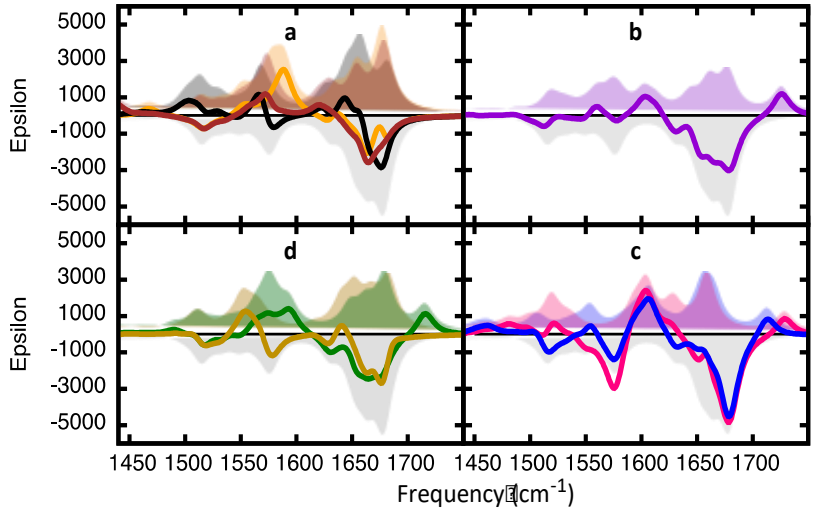



**Figure S27.** Excited state (in color shading) and differential spectra ($S_n$-$S_0$, solid line) computed for a) G* (black), C* (brown) and $C^*_{n O\pi^*}$ (orange) b) $CT^{intra}$ (purple), c) $PT^{inter1}$ (pink) and DPT (blue) and d) $CT^{inter}$ (green) and $PT^{inter2}$ (ochre) with the intensity of the $S_0$ divided by 2 at ~1570 $cm^{-1}$.

As discussed in the main text and in section 2.3 above, our calculations overestimate the intensity of the peak at ~1560 $cm^{-1}$ for GC-DNA in the GS, especially for 2GC model. Actually, the one discrepancy between the experimental long-time TRIR spectrum and the computed $PT^{inter1}$ one is the intensity of the negative feature at 1560 $cm^{-1}$ (Figure 2 of the main text), exaggerated in the computed DIR. Moreover, the GSB band at that energy in the TRIR spectra, though present, is rather shallow, also on a < 5 ps time-scale, less intense than in most of the computed DIR. The error made in estimating the relative intensity of the $S_0$ band at ~1560 $cm^{-1}$ thus makes more difficult the interpretation of the experimental TRIR. In order to further investigate this issue, in Figure S27 we show the difference spectra computed by artificially decreasing the intensity of the GS band at 1570 cm-1, which, as discussed above, is overestimated by our calculations. Actually, though obviously this procedure can give only qualitative insights on the effect of this overestimation, it is noteworthy that both the G*-min and the $CT^{intra}$-min spectra are in better agreement with the experimental one on the 1~4 ps time scale. For example, the positive feature at 1550 $cm^{-1}$ is more clearly visible.

**5. Comparison between $S_1$ and '$S_0$ sum' results**

5.1 Comparison of the difference IR spectra.

DFT is nowadays routinely used to compute ground state IR spectra. As a consequence, whenever possible, we compared the computed Excited State spectra with those estimated by summing up the contributions of their components (cation and anion), computed in their ground state. The two sets of spectra are in good agreement, further supporting the solidity of the conclusions made on the ground of the excited state spectra. However, it is important that the effect of the duplex is considered also in this ground state calculations, i.e. we have not used the results obtained on isolated bases in solution (e.g. bases at different pH) to interpret the results obtained in the duplex. In DNA many geometrical parameters and, consequently, vibrational modes are affected by the presence of a charge moiety. In other words, using the spectra of $(GC)_2^-$ and $(GC)_2^+$ to model that of the CT state is a more accurate approximation than using those of $G^+$ and $C^-$ in water.

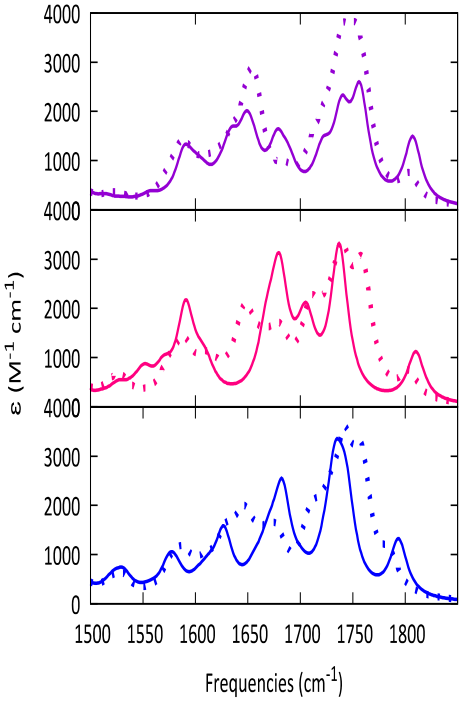



**Figure S28.** Comparison between unshifted spectra simulated as sum of the cationic/anionic species (dashed) vs "true" $S_1$ minima (solid) for $CT^{intra}$ (purple) $PT^{intra}$ (pink) and DPT (blue) in 2•GC. .

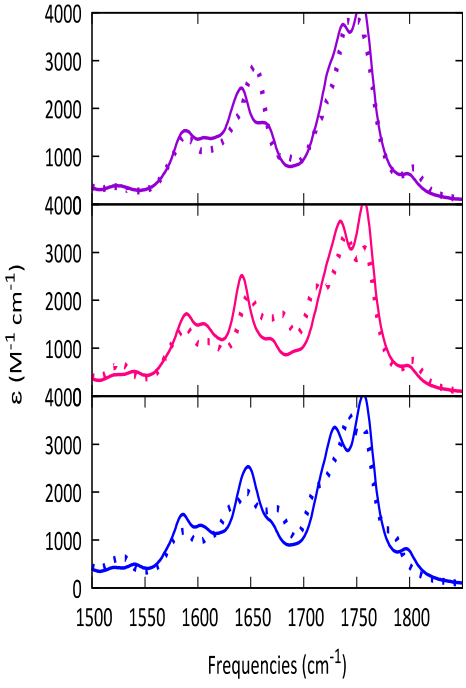



**Figure S29.** Comparison between 2•GC (dashed) vs 3•GC (solid) unshifted spectra simulated as sum of the cationic/anionic species for $CT^{intra}$ (purple) $PT^{intra}$ (pink) and DPT (blue).

**Table S8**: Selected bong lengths (in Å) in the main excited state minima of 2•GC according to PCM/TD-M052/6-31G(d) geometry optimizations.

| | $S_0$ | $CT^{intra}$ | $PT^{inter1}$ | DPT | $CT^{inter\ a}$ | $PT^{inter2}{}_{do}$ | C* | G* |
|---|---|---|---|---|---|---|---|---|
| **G1** | | | | | | | | |
| CO | 1.238 | 1.219 | 1.218 | 1.237 | 1.222(1.220) | 1.235 | 1.237 | 1.247 |
| C-NH2 | 1.346 | 1.309 | 1.310 | 1.325 | 1.312(1.303) | 1.329 | 1.343 | 1.335 |
| CO HB | 1.891 | 2.028 | 1.950 | 1.807 | 2.011(2.120) | 1.967 | 1.878 | 1.886 |
| NH2 HB | 1.922 | 1.739 | 1.750 | 1.958 | 1.649(1.446) | 1.868 | 1.905 | 1.858 |
| NH HB | 1.944 | 1.874 | 1.869 | 1.757[b] | 1.574(1.747) | 1.920[b] | 1.941 | 1.932 |
| **C1** | | | | | | | | |
| CO | 1.237 | 1.262 | 1.237 | 1.238 | 1.236 | 1.235 | 1.237 | 1.240 |
| C-NH2 | 1.328 | 1.394 | 1.374 | 1.378 | 1.326 | 1.327 | 1.356 | 1.330 |
| C5-C6 | 1.349 | 1.385 | 1.388 | 1.385 | 1.351 | 1.350 | 1.429 | 1.354 |
| St(C5-C5)[c] | 3.787 | 3.720 | 3.558 | 3.643 | 3.831 | 3.725 | 3.666 | 3.708 |
| **G2** | | | | | | | | |
| CO | 1.237 | 1.236 | 1.256 | 1.254 | 1.237 | 1.237 | 1.236 | 1.237 |
| C-NH2 | 1.343 | 1.341 | 1.358 | 1.376 | 1.350 | 1.350 | 1.345 | 1.345 |
| CO-HB | 1.875 | 1.966 | 1.829 | 1.857 | 1.884 | 1.879 | 1.892 | 1.892 |
| NH2 HB | 1.876 | 1.790 | 1.952 | 1.995 | 1.912 | 1.937 | 1.937 | 1.937 |
| NH –HB | 1.935 | 1.879 | 1.872[b] | 1.869[b] | 1.979 | 1.975 | 1.917 | 1.916 |
| **C2** | | | | | | | | |
| CO | 1.244 | 1.240 | 1.245 | 1.222 | 1.272(1.271) | 1.243 | 1.244 | 1.246 |
| C-NH2 | 1.331 | 1.330 | 1.331 | 1.314 | 1.402(1.409) | 1.386 | 1.330 | 1.330 |
| C5-C6 | 1.351 | 1.349 | 1.350 | 1.351 | 1.380(1.377) | 1.379 | 1.351 | 1.351 |
| St(C5-C5)[c] | 4.745 | 3.551 | 4.527 | 3.866 | 4.666 | 4.094 | 4.739 | 4.74 |

Notes: a) Not a minimum of the PES. In parentheses the results obtained for 3•GC . B) The proton has been transferred to the HB base. C) Stacking distance

**6. Additional considerations on the possible deactivation channels**

As anticipated in the main text, a full assignment of the transient IR spectra of $(GC)_9$ is not a trivial task. The system is large and conformationally disordered with many different excited states that can, in principle, be involved in its photoactivated dynamics. It is thus not surprising that we are in the presence of very congested spectral signals. From the computational point of view, in addition to the usual sources of errors (related to the density functional, to the incomplete basis set, to the inaccuracies in the treatment of solvent effect by a purely continuum model, etc.), some effects are missing in our treatment. In particular, we do not consider anharmonic effects (which affect both the position and the intensity of the peaks) and the conformational fluctuations of the duplex. In other words, we compute the spectra in a single minimum for each electronic state (as in a 'standard' molecular system). The possible importance of these latter effects can be appreciated by comparing the $CT^{intra}$ spectra computed optimizing the CT state involving the two different strands. Though the two bases involved in the CT reach a similar geometry, some differences are found in the 'opposite strand', due to the different stacking geometry of the two strands in $S_0$. As a consequence, the difference spectra computed for $CT^{intra}$ are different (See Figure 6 in the main text). The most significant features are however similar. Furthermore, our computational predictions are always consistent with the experimental data (e.g., the calculated ground state spectrum compares well with the experimental FTIR spectrum of GC-DNA shown in Figure 2). Similar considerations hold also for the other pillar of our treatment, i.e. the characterization of the PES of the main paths.

Shortly, though we cannot expect a perfect match between experimental and computed spectra for any species, our calculations could provide a reliable basis to interpret the most relevant experimental spectral features and/or to sketch which is the most likely deactivation mechanism.

For what concerns the experimental results, a first important indication is that, within the experimental error, no rise in the ESA signal is observed after the first ps. On the other hand, several positive and negative features of the TR-IR spectrum, in different regions, decay with different time-constants, suggesting a complex deactivation mechanism involving different excited states, in line with the computational indications. One possibility is that we are in the presence of two parallel deactivation channels, with different time-constants. It is however possible that: (i) the rise in the ESA in some region is masked by the decay of other states which also absorb in that region; (ii) many excited states absorb at similar frequencies, which makes monitoring the population transfer more difficult. Our calculations indicate that both alternatives are likely. In the following we shall discuss in more details the main experimental results and the possible scenarios.

1) CT and PCET

Experiments clearly show that already at 1 ps a strong absorption above 1700 $cm^{-1}$ is found. This broad peak is present also in the long living component. According to our calculations, this peak can be associated to the excited states involving a Guanine base with partial positive charge (with the blue-shifted CO stretching), i.e. $CT^{intra}$, $PT^{inter1}$, $CT^{inter}$) and to DPT (because of the presence of a C base with a partial positive charge). According to our calculations, $CT^{intra}$ is a better candidate than $CT^{inter}$ for being responsible of this feature, since it is very stale in the FC region and strongly coupled with the bright electronic states (those resulting from the coupling between C* and G*). On the other hand, it is in principle possible that, once formed, $CT^{intra}$

decays completely to $PT^{inter1}$ and/or DPT. As $PT^{inter1}$ is slightly more stable than $CT^{intra}$ and the energy barrier separating them is not very large, it is in principle possible that all CT population disappears in less than 1 ps. This scenario would imply that in the first ps a certain percentage of the bright excited state decays to $PT^{inter1}$ or DPT. However, several elements suggest that a complete disappearance of $CT^{intra}$ population in 1 ps is unlikely: i) A very large body of experimental and computational results indicate that UV absorption by oligonucleotide leads to the formation of exciton states delocalized over multiple bases. These exciton states then decay to more localized excited states (monomer-like, excimer/exciplexes involving two bases etc). Our geometry optimizations confirm that this picture is valid also for $(GC)_9$. This 'exciton trapping' process requires some time. In $(dA)_{20}$ a fast (~100 fs) intra-band scattering is followed by formation of excimer and CT states, which are fully populated after ~3 ps.[7] This picture is confirmed by semiclassical dynamical calculations on the same system, which indicate that fast formation of monomer-like excited state is accompanied by decay to excimer and CT states on a longer time-scale (200 fs~1 ps).[8] For what concerns $(GC)_9$, the larger rigidity with respect to a single strand could make the exciton trapping processes faster, but, at the same time, more localized excited states could be formed. Actually, the formation of mixed Frenkel-CT states extending over the two strands is well documented in GC duplexes.[9] Moreover, in poly(GC) a strong decrease of the fluorescence anisotropy under to 0.1, usually associated to the formation of CT states, is observed only after 500 fs. [10] As a consequence, though we do not have direct information on the time associated to exciton trapping in $(GC)_9$ and the subsequent formation of CT states, this process will require some time, at least for a part of the $CT^{intra}$ population. ii) Analogously, we need that the amount of charge separation in $CT^{intra}$ makes the PT transfer feasible iii) The driving force towards the PT is modest (0.1 eV) iv) There is a stable minimum for $CT^{intra}$ separated by a non-zero energy barrier (0.2~0.3 eV) from $PT^{inter1}$ v) The shape of the PES does not suggest any incipient PT transfer vi) The conformation heterogeneity of DNA also makes a complete transfer in 1 ps unlikely. There could be conformation for which the driving force towards PT increases and the energy barrier decreases, but also there could be situations where the opposite occurs. vii) $PT^{inter1}$ is a long-living state, i.e. the possibility that the population of the minima of $PT^{inter1}$ and $CT^{intra}$ is in dynamic equilibrium for many dozens of fs cannot be discarded.

These considerations suggest that the population of DPT within 1 ps is even less likely, because a second PT reaction is required.

If we thus accept that at 1ps part of the excited state population is still on $CT^{intra}$, what would be the following steps? Our calculations indicate that $PT^{inter1}$ is more stable than $CT^{intra}$, in line with the $pK_A$ of G and $C^-$, though the energy difference between the two minima (~0.1 eV) is one half of the value predicted for an isolated G-C- WC pair, showing that the energetics of this reaction is affected by the inclusion of the base in the duplex and, likely, to the possible partial CT nature of $CT^{intra}$. Moreover, the energy barrier separating $CT^{intra}$-min and $PT^{inter1}$-min (0.2~0.3 eV) is not large enough to prevent the PT reactions, especially when considering the excess of energy deposited on the system by UV excitation. Our calculations thus strongly indicate that at least a part of the $CT^{intra}$ population is transferred to $PT^{inter1}$. On the analysis of the IR spectra does not provide firm indications on the occurrence and on the time-scale of this process. The DIR spectra of $CT^{intra}$ and $PT^{inter1}$ are fairly similar, the largest difference in the investigated spectral window concerns the relative intensity of the bleaching band **1** and **2**. In $PT^{inter1}$ bleaching band 2 is less intense than band 1, whereas in the $CT^{intra}$ spectrum the intensity of these two bands is more similar. We could thus expect that $CT^{intra} \rightarrow PT^{inter1}$ is mirrored by a variation in the relative intensity of these two bands. Actually, the ratio between the intensity of the 2 and 1 bleaching bands decreases by 10~15% when comparing the 'fast' and the 'slow' DADS extracted by the experimental TRIR. However, considering that the normalized ground state recovery at 1657 (maxima of the bleaching band 2) and 1686 $cm^{-1}$ (bleaching band 1) recovery proceeds at the same rate, we are

hesitant to reach a conclusion based on the small difference in the DADS. In summary, a (even partial) $CT^{intra} \rightarrow PT^{inter1}$ population transfer is likely and, especially if a large part occurs in the first ps, consistent with the experimental TRIR spectra, which, however, does not allow to reach unambiguous conclusion on its extent and time-scale.

2) $PT^{inter2}$

There is no indication in the transient IR spectrum of a significant involvement of $PT^{inter2}$ route. Moreover, $CT^{inter}$ is significantly less stable than $CT^{intra}$ in the FC region and less coupled with the bright states, hindering its initial population. On the other hand, it cannot be excluded that $CT^{inter}$ provides some contribution to the peaks, especially in the first ps after excitation. Indeed, it could be populated from G*-min and C*-min following the same path suggested to occur in isolated GC base pairs. Moreover, the short lifetime of $PT^{inter2}$ state, due to the proximity of a CI with $S_0$, would make its population at a given time rather small and, therefore, not easily recognizable in the spectra, also because no significant ESA is associated to $PT^{inter2}$. However, a major involvement of this decay route is ruled out by the long persistence of the 1700 $cm^{-1}$ signal, since our computed spectra indicate that $CT^{inter} \rightarrow PT^{inter2}$ decay would have been associated to disappearance of that signal. The analysis of the electronic transient absorption spectra (Zhang Chem Comm) with a species showing the typical signature of the $G^+$ and/or (G-H) radical persisting on the dozens of ps time-scale, confirms this conclusion.

**References:**


[1] U. Heinemann, C. Alings, M. Bansal, *EMBO J.* **1992**, *11*, 1931-1939.

[2] a) L. Martinez-Fernandez, R. Improta, *Faraday Discuss.* **2018**, *207*, 199-216;

[3] I. Vayá, J. Brazard, M. Huix-Rotllant, A. K. Thazhathveetil, F. D. Lewis, T. Gustavsson, I. Burghardt, R. Improta, D. Markovitsi, *Chem. Eur. J.* **2016**, *22*, 4904-4914.

[4] M. K. Kuimova, A. J. Cowan, P. Matousek, A. W. Parker, X. Z. Sun, M. Towrie, M. W. George, *Proc. Natl. Acad. Sci.* **2006**, *103*, 2150-2153.

[5] A. W. Parker, C. Y. Lin, M. W. George, M. Towrie, M. K. Kuimova, *J. Phys. Chem. B* **2010**, *114*, 3660-3667.

[6] S. A. Katsyuba, E. E. Zvereva, T. I. Burganov, *J. Phys. Chem. A* **2013**, *117*, 6664-6670.

[7] R. Borrego-Varillas, G. Cerullo, D. Markovitsi, *J. Phys. Chem. Lett.* **2019**, *10*, 1639-1643.

[8] L. M. Ibele, P. A. Sánchez-Murcia, S. Mai, J. J. Nogueira, L. González, *J. Phys. Chem. Lett.* **2020**, *11*, 7483-7488.

[9] a) M. Huix-Rotllant, J. Brazard, R. Improta, I. Burghardt, D. Markovitsi, *J. Phys. Chem. Lett.* **2015**, *6*, 2247-2251; b) I. Vayá, F.-A. Miannay, T. Gustavsson, D. Markovitsi, *ChemPhysChem* **2010**, *11*, 987-989.

[10] F.-A. Miannay, Á. Bányász, T. Gustavsson, D. Markovitsi, *J. Am. Chem. Soc.* **2007**, *129*, 14574-14575.