Electronic Delocalization, Charge Transfer and Hypochromism in the UV Absorption Spectrum of Polyadenine Unraveled by Multiscale Computations and Quantitative Wavefunction Analysis

Juan J. Nogueira,*[,†] Felix Plasser,*[,†] and Leticia González*

Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, Währinger Straße 17, 1090 Vienna (Austria).
[†]Authors contributed equally to this work.

Corresponding Authors

* Juan J. Nogueira. Email: nogueira.perez.juanjose@univie.ac.at
* Felix Plasser. Email: felix.plasser@univie.ac.at
* Leticia González. Email: Leticia.gonzalez@univie.ac.at

Table of Contents

S1 Sampling the Franck-Condon Region of solvated $(dA)_{20}$ and AMP by Molecular Dynamics

A double-stranded helix $(dA)_{20}(dT)_{20}$ was generated by the nucleic acid builder implemented in AmberTools15.[1] The strand $(dT)_{20}$ was manually removed and the remaining single strand $(dA)_{20}$ was neutralized with Na^+ ions and solvated by a periodic truncated octahedral box of water molecules extended to a distance of 12 Å from any solute atom by the leap module of AmberTools15.[1] This setup results in a system with 46924 atoms: 639 atoms composing the DNA strand, 19 Na^+ ions and 15422 water molecules. Then, the solvated single strand was minimized, heated and equilibrated following the next protocol. First, a minimization employing the steepest descent method for 10000 steps and conjugate gradient method for 10000 additional steps was performed. After minimization, the system was heated to 300 K at constant volume (NVT ensemble), using a Langevin thermostat with a collision frequency of 1 ps^{-1}, for 200 ps with a time step of 2 fs. The coordinates of the strand were restrained with a force constant of 10 kcal/(molÅ^2) during the first 100 ps, and the restrains were removed during the last 100 ps. Once the system was at 300 K, the density of the solvent was equilibrated at constant pressure (NPT ensemble) during a 50 ns simulation with a time step of 2 fs. The Berendsen barostat with a pressure relaxation time of 5 ps was used to maintain a pressure of 1 bar. In the previous protocol of minimization, heating and equilibration, the whole system was described classically by a force field. In particular, DNA was described by the ff14SB set of parameters[2] and the water molecules by the TIP3P model.[3] The Coulomb and van der Waals' interactions were truncated at 10 Å and the particle mesh Ewald (PME) method[4] was employed to compute the Cou-

lomb interactions. Bond distances involving H atoms were restrained by the SHAKE algorithm.[5]

The Franck-Condon region of solvated adenosine monophosphate (AMP) was also explored by classical molecular dynamics. AMP was generated by the leap module of AmberTools15[1] using the sequence {OHE A3} command, which generates an A3 type residue with a protonated 5' phosphate group. Then, the monomer was neutralized with one Na^+ ion and solvated by a periodic truncated octahedral box of water molecules extended to a distance of 12 Å. The solvated AMP was minimized, heated and equilibrated following exactly the same protocol and technical details as the ones employed for solvated $(dA)_{20}$. The only difference is that the NPT simulation was evolved for 5 ns. All classical molecular dynamics simulations were conducted by the GPU module pmemd[6] implemented in Amber14.[1]

S2 Absorption Spectrum for Thermal Ensembles of AMP and $(dA)_{20}$

We selected 100 equidistant snapshots from the last half part of each classical molecular dynamics simulation performed for solvated AMP and $(dA)_{20}$. Then, the excited state energies for the 100 snapshots of each system were calculated by a QM/MM electrostatic scheme. In the case of solvated $(dA)_{20}$, the eight nucleobases (without the sugar-phosphate backbone) in the middle of the strand were included in the QM region and embedded in the electric field created by the atomic charges of the remaining part of the strand, the water molecules and the Na^+ ions. The atomic charges of the environment were taken from the above force field.[2, 3] The hydrogen link atom scheme as implemented in Amber14[1] was employed to cap the covalent bonds that connect the QM and MM regions. The lowest 60 excited states of the eight

nucleobases in the QM region were calculated using full time-dependent density functional theory (TD-DFT) with the CAM-B3LYP[7] functional and the SV(P) basis set.[8] Polarization p-functions were removed from the H atoms for computational efficiency. In case of solvated AMP, the only adenine of the system was included in the QM region and the 10 lowest electronically excited states were computed using the same computational details as for $(dA)_{20}$. All excited-state calculations were performed using the GPU-based TeraChem1.9 package.[9, 10]

One of the goals of the present manuscript was to use the largest possible QM region size in order to properly characterize the electronic delocalization length. In addition, a large QM region also enables an accurate description of the interactions between the excited nucleobases and the neighboring ones. The use of wavefunction-based methods to compute the electronically excited states is unfeasible due to the large size of the QM region (eight adenines). Instead, DFT was employed. As discussed in the main manuscript, charge-transfer (CT) states play a relevant role in the UV absorption spectrum of $(dA)_{20}$, by mixing with Frenkel excitons, and likely in the subsequent excited-state dynamics. Therefore, an appropriate description of these states is crucial to get a correct picture of the photophysics of $(dA)_{20}$. It is well known that DFT functionals without long-range Hartree-Fock exchange correction are not able to properly describe CT states.[11] Thus, the long-range corrected CAM-B3LYP functional was employed in the calculations.

After following the above explained protocol, a total of 6000 excited states for solvated $(dA)_{20}$ (60 states for each of the 100 snapshots) and 1000 states for solvated AMP (10 states for each of the 100 snapshots) were calculated. In order to obtain the absorption spectra shown in Fig. 3a of the main manuscript, the resulting 6000 (1000) excitation energies for $(dA)_{20}$ (AMP) were convoluted with Gaussian functions with a full width at half maximum of 0.20 eV. The height of the Gaussian functions is equal to the oscillator strengths of the electronic states. The intensity of the absorption band of the polymer was scaled to unity, and the intensity of the absorption band of AMP was scaled by the same factor. The calculated absorption spectrum of $(dA)_{20}$ is blue shifted by 0.70 eV with respect to the experimental spectrum.[12, 13] As was previously shown,[14] better agreement with experiments can be achieved using larger basis sets. This will be discussed in more detail in Section S3. For an easier comparison with experiments, all energy calculations performed in this work were red-shifted by 0.70 eV.
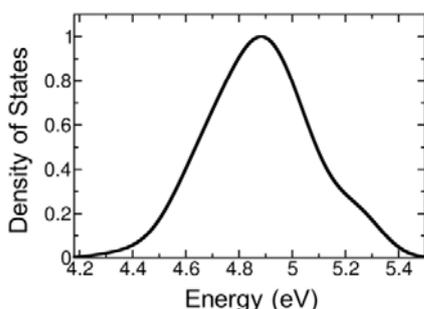

Fig. S1 Density of electronic states of AMP computed for a thermal ensemble of 100 geometries.

The density of states (Fig. S1) for the ensemble of AMP geometries was also obtained by convolution of the 6000 excitation energies with Gaussian functions, but in this case the height of the functions are equal to the unity (and not to the oscillator strength of the states). Finally, the maximum of the density of states band was scaled to unity. As can be seen in Fig. S1 (and is discussed in Section 3.2 of the main manuscript), the density of electronic states is much higher in the center than in the edges of the band. Thus, when exciton states are formed in the polymer, they will be composed mainly by monomer-like excitations that were located in the center of the band. As a consequence, the contribution of monomer-like excitations to the absorption spectrum of the polymer (blue line in Fig. 4a) presents a flat shape.
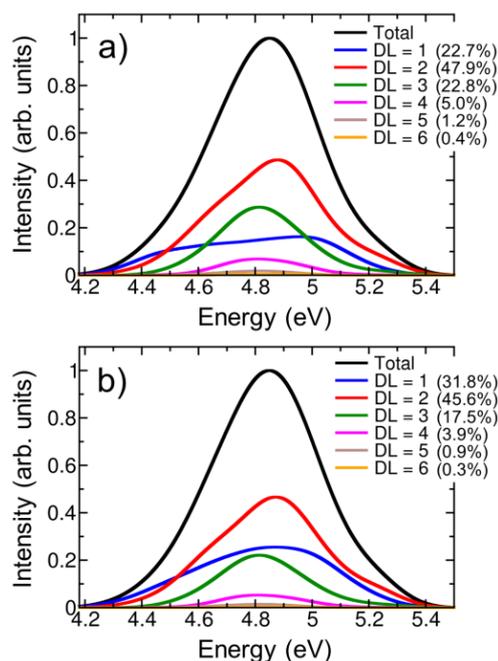

Fig. S2 Decomposition of the lowest-energy band of the UV absorption spectrum for $(dA)_{20}$ into different delocalization length (DL) contributions (a) excluding the edge nucleobases and (b) including all nucleobases in the analysis. The contribution of each delocalization length is indicated in parenthesis.

The delocalization length and CT analyses discussed in Sections 3.2 and 3.3 of the main manuscript were performed excluding the excitations involving the first and eighth nucleobases of the QM region. In this way, artifacts due to edge effects are avoided. As shown in Fig. S2, inclusion of the two edge nucleobases (panel b) induces an increase of the contribution of monomer-like excitations and a decrease of the contribution of delocalized excited states. This is not surprising since excited states in which the nucleobases located at the end of the QM region are involved cannot be delocalized towards the neighboring MM nucleobases. As a consequence, their delocalization length is underestimated. This analysis indicates that QM/MM calculations describing only two nucleobases in the QM region would provide quite unrealistic results since both nucleobases are located at the edges of the

QM region. Thus, a large number of nucleobases is required in the QM region to obtain converged results.

S3 Absorption Spectrum for Optimized Geometries of AMP and (dA)$_{20}$

The effect of conformational sampling on the UV absorption spectra and electronic properties of AMP and (dA)$_{20}$ is discussed along the main manuscript. To do such an analysis, excited-state calculations for AMP and (dA)$_{20}$ were carried out at the optimized geometries and compared with the ensemble calculations. The last snapshot from each of the previous classical molecular dynamics simulations for solvated AMP and (dA)$_{20}$ were selected. These geometries were classically optimized for 10000 steps using the steepest descent method, and for additional 10000 steps using the conjugate gradient method. The same technical details used during the minimization explained in Section S1 were used here. Then, the lowest 60 and 10 excited states for (dA)$_{20}$ and AMP, respectively, were computed, and convoluted with Gaussian functions with a full width at half maximum of 0.20 eV and with height equal to the oscillator strength of the electronic states. The resulting spectra are plotted in Fig. 3a (dashed lines).
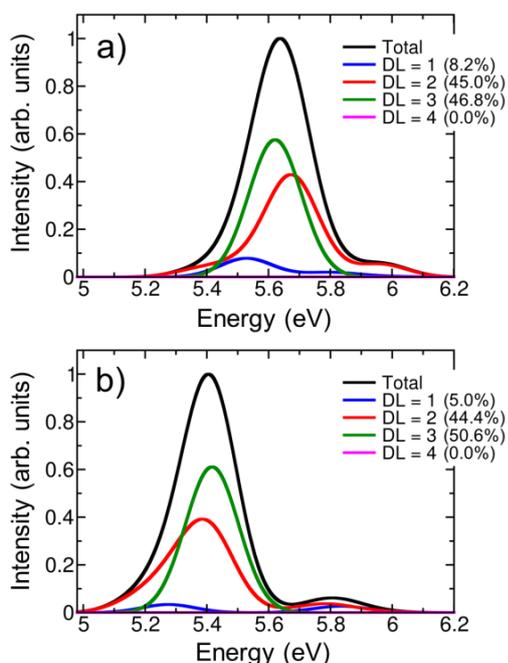

Fig. S3 Decomposition of the lowest-energy band of the UV absorption spectrum for (dA)$_{20}$ into different delocalization length (DL) contributions calculated using the (a) SV(P) and (b) 6-311+G(2d,p) basis sets.

As mentioned in Section S2, the maximum of the calculated absorption spectrum of (dA)$_{20}$ for the thermal ensemble is blue shifted by 0.70 eV with respect to the experimental spectrum.[12, 13] In the case of the spectrum calculated at the optimized geometry the energy overestimation is larger (0.85 eV). Previous calculations showed that the use of a more extended basis set would significantly improve the agreement with experiment by shifting the spectrum to lower energies.[14] However, the use of larger

basis sets in our model is hard because of the large size of the QM region and the large number of snapshots taken from the dynamics for the excited-state calculations. An important question is whether the use of a more extended basis set would only induce a shift in the spectrum or also a significant modification of the electronic wavefunctions, altering the electronic properties. To investigate this, the absorption spectrum of (dA)$_{20}$ was calculated using the CAM-B3LYP functional[7] and two different basis sets, namely SV(P)[8] and 6-311+G(2d,p).[15, 16] Four nucleobases were included in the QM region and the 40 lowest excited states were computed. The spectra calculated with the two basis sets, and their delocalization length decomposition are displayed in Fig. S3. The maxima of the spectra peak at 5.64 and 5.41 eV for SV(P) and 6-311+G(2d,p) basis sets, respectively, i.e., the spectrum is red shifted by 0.23 eV by increasing the basis set. The use of even larger basis sets would induce further energy lowering. The decomposition of the absorption bands into different delocalization length contributions is very similar for both basis sets. Delocalization lengths over two and three nucleobases clearly dominate the band with similar contributions and monomer-like excitations present a very small intensity. Therefore, the red shift of the band is the main modification introduced by the use of a more extended basis set, while the electronic properties of the excited-state wavefunctions remain almost unaltered. The level of theory employed along the calculations of the main manuscript is thus well justified.

S4 Variation of Electronic Properties of Adenine Dimer with Interbase Separation

In order to get more insight into the hypochromism observed in the lowest-energy UV band, excited-state calculations for an adenine dimer were performed and analyzed. The results are presented in Section 3.4. It has been shown along the manuscript that analyses performed at optimized geometries provide misleading conclusions. Conformational sampling was partially considered in our dimer model by constructing the system based on two randomly selected geometries from the molecular dynamics simulation of solvated AMP. The sugar-phosphate backbone and water molecules were removed, and the remaining two adenine nucleobases were spatially oriented to mimic the structure of the classically optimized single-strand (dA)$_{20}$. At the initial geometry, the rise separation, which is the distance between the two planes that define the aromatic moieties of the nucleobases, is 3.6 Å. This separation was elongated considering the next distances: 3.6, 3.8, 4.0, 4.2, 4.4, 4.6, 4.8, 5.0, 5.4, 5.8, 6.2, 6.6, 7.0, 8.0, 9.0, 10.0 and 30.0 Å. For each separation the 20 lowest electronically excited states were calculated in the gas phase at CAM-B3LYP/SV(P) level of theory, removing the polarization p-functions from the H atoms. Based on the spectrum of Fig. 4, electronic states lying at energies lower than 5.3 eV were considered to belong to the first absorption band, and states lying between 5.3 and 6.3 eV to the second band. Then, the excitation energy, delocalization length and CT contribution of each band were calculated as oscillator-strength-weighted averages of each individual state. The oscillator strength of the bands was calculated as the sum of the oscillator strengths of the electronic states. The variation of these electronic properties with the interbase separation is plotted in Fig. 7.

S5 REFERENCES

1. D.A. Case, J. T. Berryman, R. M. Betz, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz et al. AMBER 2015, University of California, San Francisco., Editon edn., 2015.
2. J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696-3713.
3. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
4. M. F. Crowley, T. A. Darden, T. E. Cheatham III and D. W. Deerfield II, *J. Supercomput.*, 1997, **11**, 255-278.
5. S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952–962
6. R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2013, **9**, 3878-3888.
7. T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51-57.
8. A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571-2577.
9. I. S. Ufimtsev and T. J. Martinez, *J. Chem. Theory Comput.*, 2009, **5**, 2619-2628.
10. TeraChem v 1.9, PetaChem, LLC (2009,2015), Editon edn.
11. A. W. Lange and J. M. Herbert, *J. Am. Chem. Soc.*, 2009, **131**, 3913-3922.
12. A. Banyasz, T. Gustavsson, D. Onidas, P. Changenet-Barret, D. Markovitsi and R. Improta, *Chem. Eur. J.*, 2013, **19**, 3762-3774.
13. A. Banyasz, I. Vayá, P. Changenet-Barret, T. Gustavsson, T. Douki and D. Markovitsi, *J. Am. Chem. Soc.*, 2011, **133**, 5163-5165.
14. P. Caruso, M. Causà , P. Cimino, O. Crescenzi, M. D'Amore, R. Improta, M. Pavone and N. Rega, *Theor. Chem. Acc.*, 2012, **131**, 1-12.
15. R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, *J. Chem. Phys.*, 1980, **72**, 650-654.
16. M. J. Frisch, J. A. Pople and J. S. Binkley, *J. Chem. Phys.*, 1984, **80**, 3265-3269.