Supporting information

DNA-Binding Mechanism of Spiropyran Photoswitches: the Role of Electrostatics

Davide Avagliano, Pedro A. Sánchez-Murcia* and Leticia González*

Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, Währinger Straße 17, A-1090 Vienna (Austria)

Section S1 Molecular dynamics simulation setup

The initial molecular model of the 12-*mer* dsDNA (poly-dAT)$_2$ was prepared using the *nab* program implemented in AmberTools14, as described elsewhere.[1,2] The MCH form of **3a** was inserted into the dsDNA manually and minimized *in vacuum* following a standard protocol using *sander*. The resulting intercalated molecular model **3a**:(poly-dAT)$_2$ was used as starting geometries for the umbrella sampling MD studies, where both compounds were pulled out to the bulk solvent. Atom-centered ESP charges of **3a, 3b** and **3c** were computed with *antechamber* (AmberTools17)[1] following the standard procedure for classical MD simulation on their *ab initio* optimized geometries using HF/6-31G* with the Gaussian suite.[3] Their GAFF force field parameters were computed using the module *parmchk2*[4] of Amber17[1] and the ff14SB force field parameter set[5] was also used to assign the bonded and non-bonded parameters of compounds **3a, 3b** and **3c** and to describe the nucleotides of the dsDNA. Each MCH:dsDNA complex (e.g. **3a**:dsDNA) was immersed in a cubic box of 30 Å from the solute to the border of the box filled with TIP3P water molecules[6] and 20 Na^+ to ensure electroneutrality. In addition, and in order to reproduce the experimental conditions reported by Andersson *et al.*,[7] a final NaCl concentration of 1 10^{-5} M was achieved by addition of Na^+ and Cl^- atoms. *tleap*[1] was used for the setup of all systems. In all cases, periodic boundary conditions were used, and the electrostatic interactions were computed using the Ewald method[8] with a grid spacing of 1 Å. The cutoff distance for the non-bonded interactions was 10

Å and the SHAKE algorithm[9] was applied to all bonds involving hydrogens. An integration step of 2.0 fs was defined. Before the production phase, all the simulated systems were prepared following a sequential protocol: First, the system was minimized in three steps, where all the hydrogens, the solvent molecules (waters and counter ions) and, finally, the whole system, were sequentially minimized in 20,000 steps. For the initial 10,000 steps, the steepest descendent algorithm was used; the last 10,000 steps were run using the conjugate gradient algorithm. The resulting minimized solvated geometries were heated from 100 to 300 K in 20 ps using the Langevin thermostat[10] with an integration step of 0.2 fs (collision frequency of 1 ps^{-1}) but keeping the position of all the heavy atoms of the solute restrained with a strong harmonic constant (40 $kcal^{-1}$ mol^{-1} $Å^{-2}$). In this step, a random seed was imposed. The Langevin dynamics, as well as the initial velocity for the dynamics are dependent of such random number. After that, the imposed restrains were removed in 6 steps of 20 ps each, were the system was switched from a NVT (40 to 10 $kcal^{-1}$ mol^{-1} $Å^{-2}$ in 80 ps, constant volume) to a NPT (10 to 0 $kcal^{-1}$ mol^{-1} $Å^{-2}$ in 40 ps, constant pressure) ensemble. As standard conditions, each of the MCH:dsDNA complexes were simulated at 300 K and constant pressure (1 atm) for three independent MD simulations of 30, 100 or 300 ns with an integration step of 2.0 ps using the *pmemd.cuda* engine of single precission - fixed precission (SPFP)[11,12] on two GeForce Nvidia GTX 1080Ti GPUs.

Section S2 Umbrella sampling

The intercalation pathways were calculated by umbrella sampling MD simulations using Amber17.[1] The reaction pathway was divided in 50 steps, every 0.5 Å, each one obtained with Steered Molecular Dynamics (SMD) simulations. The SMD simulation was run for 0.1 ns and a harmonic constant of 50 $kcal^{-1}$ mol^{-1} $Å^{-2}$. For the umbrella sampling MD simulation the force constant was increased to 1000 $kcal^{-1}$ mol^{-1} $Å^{-2}$ and each of the 50 windows was allowed to oscillate around the anchor potion for 5 ns with a

time step of 1 fs. The energy analysis of the potential mean force was performed with the variational free energy profile method.[13]

Section S3 Binding energy analysis

The binding energy analysis was carried out with the program MMPBSA.py implemented in AmberTools17,[1] which allows to calculate the binding free energy with the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA).[14] The entropy contribution, useful to obtain a better description of the order of magnitude of interaction between the two molecules, was also computed with the same program through normal mode analysis.[15]

The MM-PBSA analysis provided us: (i) a total electrostatic energy term (ΔE^{elec}), with includes the difference between the ΔE^{EEL} term (non-bonded electrostatic energy + 1,4-electrostatic energy) in gas phase and the electrostatic contribution in the complex probe:DNA to the solvation free energy calculated by Poisson Boltzmann (ΔE^{PB}), (ii) the non-electrostatic terms: a van der Waals term (ΔE^{vdW}) and the nonpolar contribution to the solvation free energy (ΔE^{nonpol}), (iii) an entropy term computed using Normal Mode Analysis using the same frames (ΔS), (iv) and finally, a total solvation free energy term (ΔG^{bind}).

For the binding energy analysis, in each of the MCH:dsDNA complexes we selected a window of 20 ns (400 snapshots) from their trajectories where the root-mean squared distance (RMSD) value (Å) for the complex was constant along the MD simulation. In order to obtain the electrostatic map on the dsDNA surface, we followed the procedure included in the CHARMM software,[16,17,18] solving numerically the Poisson-Boltzmann equation and obtaining the electrostatic free energy on the macromolecule surface.

Section S4 Trajectories analysis and graphical representation

The 3D representation of the complex was studied with PyMOL 1.8.6.[19] The binding modes are interpreted with CPPTRAJ,[20] a program of AmberTools17, which allows to extract important values from the MD trajectories, such as distances between DNA and dye contacts.

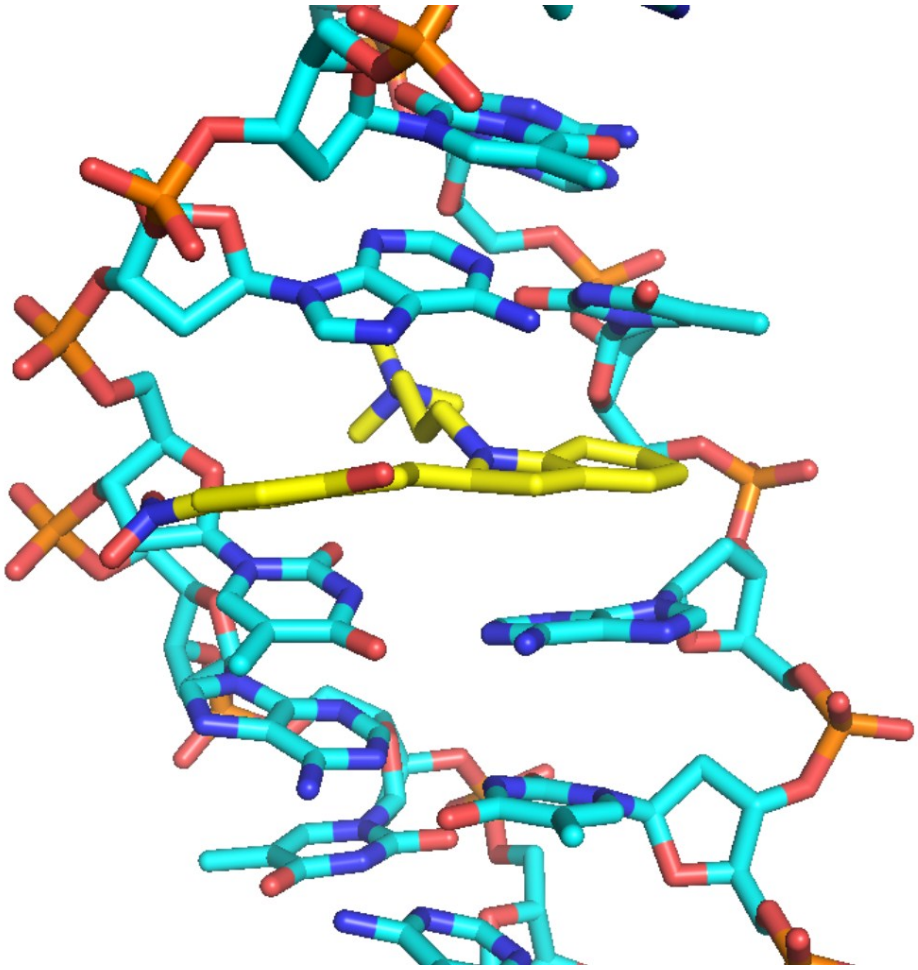

Figure S1: Intercalative geometry of **3a** (sticks, carbon atoms colored in yellow) in the binding mode M3. The intercalation of **3a** from the major groove (path M), introduces a strong perturbation into the dsDNA. As result, the state M3 is metastable.

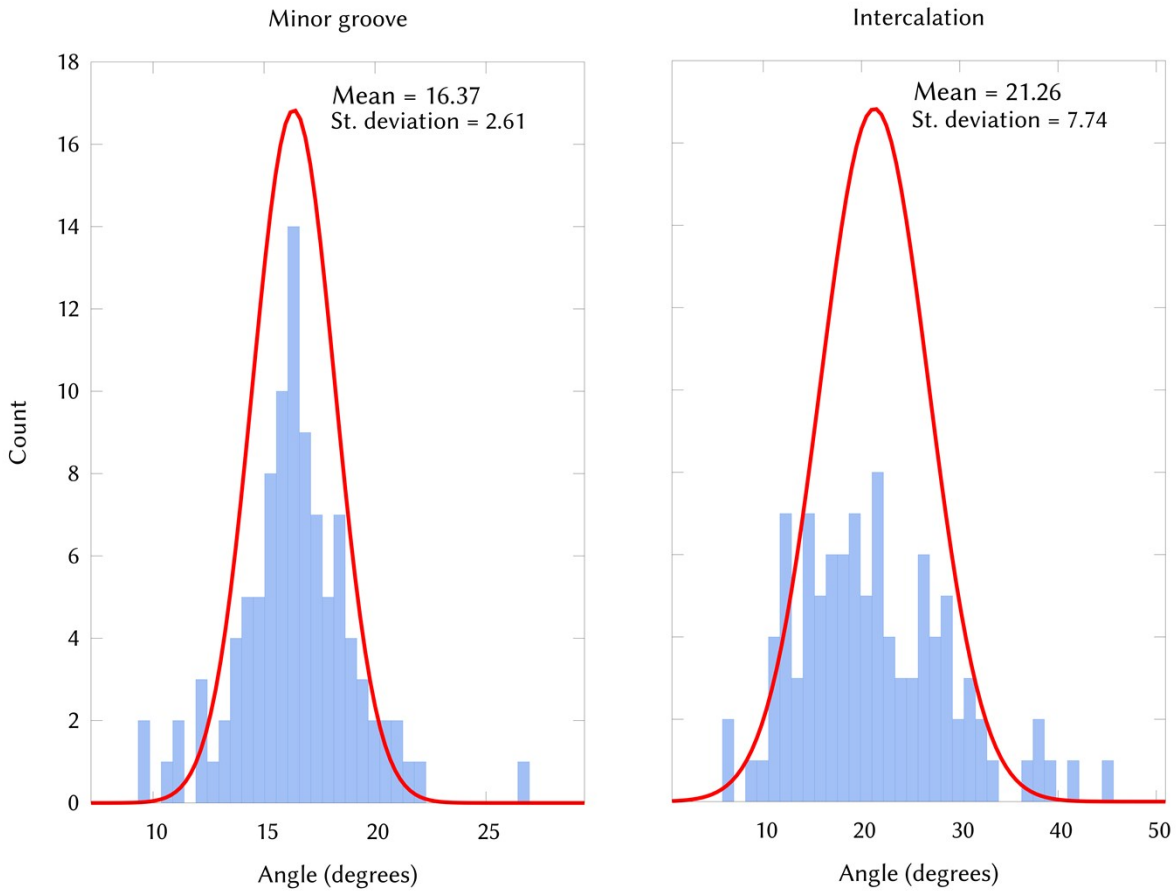

Figure S2: Distribution of the dihedral angle (º) between the two rings of **3a** in the m2 (left) and m3 (right) states along the 100ns-unrestrained MD simulations.

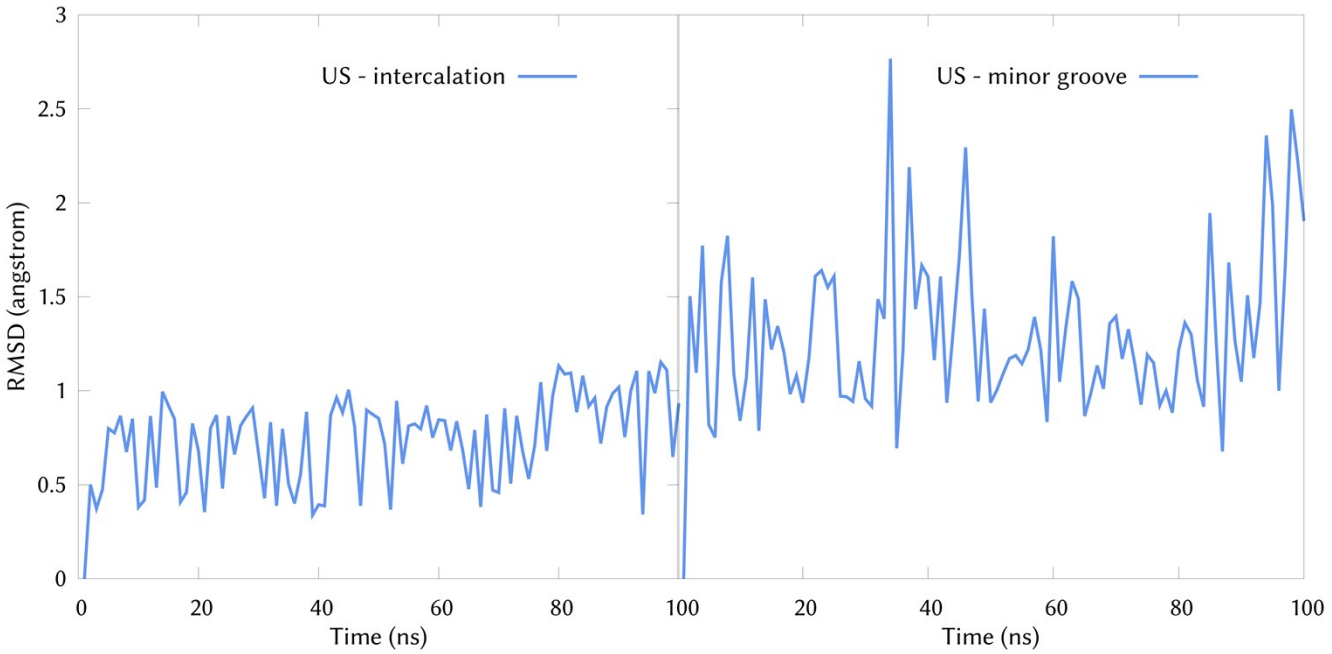

Figure S3: RMSD value (Å) of **3b** along the 100ns-unrestrained MD simulations in the m3 (left) and in the m2 (right) states. The geometries of **3a** in m3 and m2 are taken as reference and only the heavy atoms have been considered. The RMSD values are under 2.0 Å, which indicates that **3b** shares the same binding spots of **3a**.

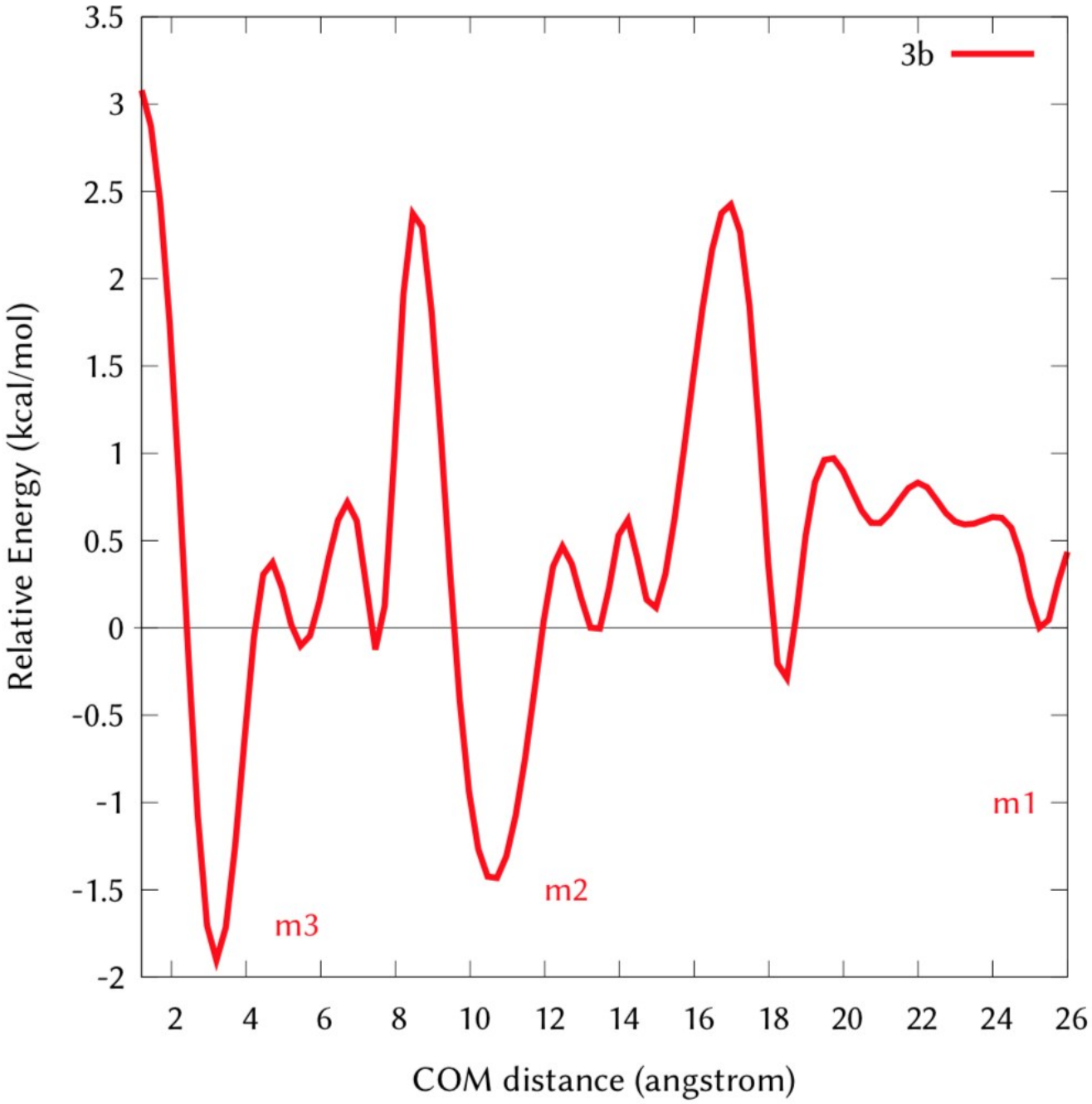

Figure S4: Free energy profiles (kcal mol^{-1}) of path m (minor) of **3b** computed with US MD simulations.

Table S1. Total binding energy (ΔG^{bind}, kcal mol^{-1}) and individual terms used for its calculation.

cmp (charge)	bind. mode	ΔE^{EEL}	ΔE^{PB}	ΔE^{elec} (ΔE^{EEL} + ΔE^{PB})	ΔE^{vdW}	$\Delta E^{non\text{-}pol}$	$\Delta E^{vdW+non\text{-}pol}$	ΔG^{MMPBSA}	$-T\Delta S$	**Total** ΔG^{bind}
3a (+2)	**m3**	-1112.69± 1.04	1128.5 ± 1.03	15.81 ± 0.14	-55.17 ± 0.19	-4.32 ± 0.01	-59.49 ± 0.19	-43.69 ± 0.32	-21.73 ± 0.09	**-21.96 ± 0.24**
3b (+3)		-1643.37 ± 2.06	1653.47 ± 2.02	10.1 ± 0.40	-55.78 ± 0.26	-4.62 ± 0.40	-60.4 ± 1.33	-50.31± 0.36	-20.96 ± 0.17	**-29.35 ± 0.33**
3c (+4)		-2155.88 ± 4.39	2126.97 ± 4.35	11.09 ± 0.59	-50.71 ± 0.49	-4.57 ± 0.03	-55.28 ± 0.49	-44.19 ± 0.61	-26.42 ± 0.37	**-17.77 ± 0.74**
3a (+2)	**m2**	-997.49 ± 1.09	1008.31 ± 1.17	10.82 ± 0.22	-37.72 ± 0.25	-3.23 ± 0.01	-40.95 ± 0.25	-30.27 ± 0.24	-20.97 ± 0.02	**-9.29 ± 0.47**
3b (+3)		-1448.37 ± 1.78	1447.35 ± 1.88	-1.02 ± 0.60	34.16 ± 0.23	-3.22 ± 0.01	-37.38 ± 0.23	-38.40 ± 0.34	-23.88 ± 0.13	**-14.52 ± 0.36**
3c (+4)		-2032.87 ± 7.62	2020.81 ± 7.63	-12.07 ± 0.39	-29.91 ± 0.42	-3.19 ± 0.02	-33.1 ± 0.42	-45.17 ± 0.74	-24.71 ± 0.26	**-20.46 ± 0.61**

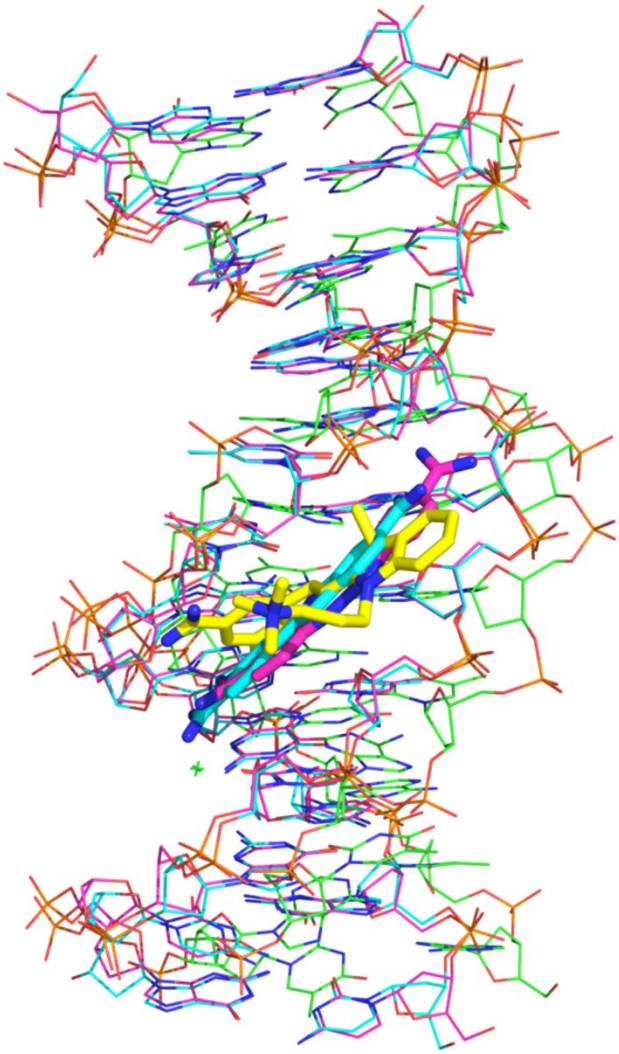

Figure S5: Structural superimposition of the intercalated **3c** (carbon atoms colored in yellow) into 12-mer (poly-dAT)$_2$ and the crystal structures of minor groove binders berenil (C-atoms colored in blue, PDB id. 1D30) and DAPI (C-atoms colored in magenta, PDB id. 2DBE).[21,22]

References

1. D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman 2018, *AMBER 2018*, University of California, San Francisco.

2. J. J. Nogueira and L. González, *Biochem.* 2014, **53**, 2391–2412.

3. Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, *Gaussian, Inc*., Wallingford CT, 2016.

4. J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graph. Model.* 2006, **25**, 247–260.

5. J. W. Ponder and D. A. Case., *Adv. Prot. Chem.* 2003, **66**, 27-85.

6. W. Junmei, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.* 2005, **25**, 1157–1174.

7. J. Andersson, S. Li, P. Lincoln and J. Andréasson, *J. Am. Chem. Soc.* 2008, **130**, 11836–11837.

8. P. P. Ewald, *Ann. Phys.* 1921, **369**, 253–287.

9. J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.* 1977, **23**, 327–341.

10. X. Wu, B. R. Brooks and E. Vanden-Eijnden, *J. Comput. Chem.* 2016, **37**, 595–601.

11. A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.* 2012, **8**, 1542–1555.

12. R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.* 2013, **9**, 3878–3888.

13. T. Lee, B. K. Radak, A. Pabis and D. M. York, *J. Chem. Theory Comput.* 2013, **9**, 153-164.

14. N. Homeyer and H. Gohlke, *Mol. Inform.* 2012, **31**, 114–122.

15. B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *J. Chem. Theory Comput.* 2012, **8**, 3314–3321.

16. S. Jo, T. Kim, V. G. Iyer and W. Im, *J. Comput. Chem.* 2008, **29**, 1859–1865.

17. S. Jo, M. Vargyas, J. Vasko-Szedlar, B. Roux and W. Im, *Nucleic Acids Res.* 2008, **36**, W270–W275.

18. W. Im, D. Beglov and B. Roux, *Comput. Phys. Commun.* 1998, **111**, 59–75.

19. DeLano, W. L. The PyMOL Molecular Graphics System, version 1.8.6 Schrödinger, LLC. 2015.

20. D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.* 2013, **9**, 3084–3095.

21. Brown D.G., Sanderson M.R., Skelly J.V., Skelly J.V., Jenkins T.C., Brown T., Garman E., Stuart D.I. and Neidle S., *EMBO J.* 1990, **9**, 1329-1334.

22. Larsen T.A., Goodsell D.S., Cascio D., Grzeskowiak K and Dickerson R.E., *J. Biomol. Struct. Dyn.* 1989, **7**, 477-491.