Electronic Supplementary Material (ESI) for Chemical Science.

Supporting Information

Enhanced sampling molecular dynamics simulations correctly predict the diverse activities of a series of stiff-stilbene G-quadruplex DNA ligands

Michael P. O'Hagan,[‡,a] Susanta Haldar,[‡,a,b] Juan C. Morales,[c] Adrian J. Mulholland,[*,a,b] and M. Carmen Galan[*,a]

[a] School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS (UK)

[b] Centre for Computational Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS (UK)

[c] Instituto de Parasitología y Biomedicina, López Neyra, CSIC, PTS Granada, Avenida del Conocimiento, 17, 18016 Armilla, Granada (Spain)

‡Authors contributed equally

MATERIALS AND METHODS

Ligand preparation

Samples of ligands **2**-**5** for CD and NMR studies were synthesised using procedures reported previously.[1,2] Stock solutions (10 mM) were prepared in DMSO-d_6, stored at -20 °C and thawed and diluted in appropriate buffer prior to use.

Oligonucleotide sequence

The DNA sequence used was telo22 (5'-AGGGTTAGGGTTAGGGTTAGGG-3'), purchased from Eurogentec (Belgium) purified by RP-HPLC and delivered dry. For the experimental studies, telo22 was diluted using the appropriate buffer (*vide infra*) and the concentration quantified using a UV-vis absorbance spectrometer (NanoDrop 2000, Thermo Scientific, UK). The DNA samples were annealed in the absence of ligand by heat-denaturation at 90 °C for 2 minutes followed by cooling on ice.

Modelling/simulation protocol for investigating ligand binding modes to DNA

To model the binding modes of ligands **2**-**5** to telo22 G4, we used the following workflow:

1. A docking calculation (see below for details) was performed for each ligand to generate possible binding poses. Representative structures were selected for analysis: the ten lowest-energy docked poses returned by the calculations were considered. However, amongst these, several poses were found to be similar in nature. Therefore, only the diverse poses were selected for further study.
2. An MD simulation was performed for each pose selected from the docking simulation. These simulations were run on long timescales (~1 μs) to test the stability of ligand binding modes and also to investigate whether the induced conformational changes, e.g., changes to the DNA secondary structure.
3. WT-MetaD simulations were performed for binding poses found to be stable in MD, in order to simulate the complete binding mechanism and also to ascertain whether alternative possible binding modes were found, for each ligand.
4. In parallel, we obtained experimental data on the structure of the G4/ligand complexes by CD spectroscopy and 1D and 2D ^{1}H NMR spectroscopy, and compared with these experimental structural data: this tested the ability of the simulation methods to predict the G4-interactive behaviour of the ligands. It is important to note that no fitting of the simulations to experiment was performed: the simulation predictions do not depend on experimental data.

Computational molecular docking

We performed molecular docking calculations of ligands **2**-**5** to telo22 G4 DNA (PDB code: 143D)[3] individually in order to predict possible high-affinity binding modes. The geometry of each ligand was first optimized with PerkinElmer Chem3D software using the MM2 forcefield. The target DNA structure, telo22, was prepared using the default build setup in AMBER.[4] Docking calculations were performed using AutoDock Vina[5] with the DNA structure kept fixed in its

original crystal conformation throughout the docking procedures. In order to allow the ligand to explore the whole conformational space, the docking area was centred on the Cartesian coordinates of the centre of mass (COM) of the G4 structure and defined by a box large enough to include the whole receptor macromolecule. The ligand molecules were treated as flexible during docking to allow for structural changes. Docking poses were clustered on the basis of the root mean square deviation (RMSD) between the Cartesian coordinates of the ligand atoms (cut-off = 0.2 nm) and were ranked based on the AutoDock Vina[5] scoring functions. The 20 highest scoring docked poses were saved to check the structural similarity/dissimilarity in the binding poses to the G4 DNA. For molecular dynamics (MD) and metadynamics (MetaD) simulations, structures were selected based on the diversity in their binding modes: i.e., binding either on the top face or towards the G4 groove, because many top face and groove-based binding modes were found to be similar in energy. The selected poses were then subjected to unbiased MD simulations (see below). Stable poses from the MD simulations were then used to begin MetaD simulations in order to investigate binding modes and mechanisms.

Standard (unbiased) molecular dynamics simulations

All the MD and MetaD simulations were performed using the Gromacs-5.0 software package. The parm-BSC1 force field was used for the DNA parameters.[6] The General Amber Force Field (GAFF) was used to generate ligand parameters.[4,7] The atomic charges were calculated using the restrained electrostatic potential (RESP) fitting procedure.[8] The RESP fit was performed onto a grid of electrostatic potential points calculated at the HF/6-31G(d) level as recommended by the force-field designers and recent literature.[7,8] For more information about the system preparation and simulation protocol, we refer the reader to our previous work.[2]

Well-tempered metadynamics simulations

We performed metadynamics (MetaD) simulations of the binding of ligands **3**-**5** to the telo22 G4 DNA.[9] Metadynamics is an enhanced sampling technique that allows sampling of long timescale events, which are not accessible through standard molecular dynamics simulations at a reasonable computational cost. Acceleration of the sampling is achieved by introducing a history-dependent bias potential on some chosen degrees of freedom, which are called the collective variables (CV). The bias potential grows by dropping Gaussian 'hills'. Hence, one can easily overcome the barrier heights and therefore possibly make several re-crossing events, computing the free energy surface (FES) of the process under study.

Here, we have adopted a well-tempered version of MetaD (WT-MetaD) which has a better algorithm to converge the FES within a reasonable timescale.[10] WT-MetaD simulations allow modelling of binding/unbinding events, and can reveal the mechanism in atomic detail. Each WT-MetaD simulation started with a structure generated from the docking poses and then equilibrated by MD simulations. As discussed above, selected docking poses were simulated for 1 μs with standard MD; however, N.B. the starting structure for WT-MetaD simulation was taken after 20 ns of MD simulation, to ensure that the starting structure was close to the crystal structure.

The same systems and MD settings as described above were used for the WT-MetaD simulation. The Plumed 2.3 plugin[11] was used to carry out simulations with the Gromacs-5.0.7 code.[12] The bias potential was calculated according to the WT-MetaD scheme as follows:

$$V(s,t) = \sum_{t'=0,\tau_G,2\tau_{G,\ldots}}^{t'<t} \omega\tau_G \, e^{-V(s(q(t'),t'))/\Delta T} \, e^{-\sum_{i=1}^{2}[(s_i(q)-s_i\left(q(t')\right))^2/2\sigma_i^2]} \quad (1)$$

where the deposition rate, ω, and deposition stride, τ_G, of the Gaussian hills were set to 0.358 kcal.mol^{-1}·ps^{-1} (1.5 kJ.mol^{-1}·ps^{-1}) and 1.0 ps, respectively. The bias factor (T + ΔT)/T was set to 15, and the final free energy surface (FES) was calculated as follows:

$$F(s,t) = -\frac{T+\Delta T}{\Delta T}(V(s,t) - C(T)) \quad (2)$$

where $V(s,t)$ is the bias potential added to the CV and the T is the simulation temperature. ΔT is the difference between the temperature of the CV and the simulation temperature. The bias potential is grown as the sum of the Gaussian hills deposited along the chosen CV space and finally the sampling of particular CV space can be controlled with the tuning of the ΔT parameter.

On the choice of the collective variables for MetaD simulations

The choice of collective variables (CVs) in MetaD is a non-trivial task which usually requires knowledge of the particular system under study and also some prior experience (i.e. several trial and error computational experiments).[13,14] Also, the multidimensional free energy surface can take long simulation times to converge, becaause the computational cost of constructing the free energy grows exponentially with the number of CVs.

To describe the different ligand binding conformations during the metadynamics simulations, after initial testing and exploration, we used two collective variables (CV): (i) the distance (D) between the centre of mass (COM) of the middle G-tetrads (base counts) and the heavy atoms of each ligand and (ii) a torsion angle (T) is used. The torsion is defined between two points in the ligand to two points in the G-tetrad. In the ligand, the two points are considered as the COM of the two indane residues.

The distance CV (D) is used to separate the ligand from DNA, whereas the torsion CV (T) is used to differentiate the conformational change in the ligand binding to DNA. The combination of distance and torsion CVs has been demonstrated to be effective in previous studies.[2,15,16]

To analyse ligand binding, we plotted the two-dimensional free energy surface (FES) of D vs T. However, from this FES alone it is not possible to determine the ligand binding mode unambiguously; it is not clear from the D/T FES alone whether the ligand sits on an external face of the G4 or resides in one of the grooves. The FES is therefore reweighted in terms of the distance from axis (DFA) and position on axis (POA) collective variables in order to determine the precise binding poses of ligands **4** and **5**.

We note that the binding free energy is calculated simply as the free energy difference between bound and unbound states on the resulting free energy surfaces. No correction is applied for standard states or concentration effects. The statistical and methodological limitations as discussed previously (e.g. lack of changes in electronic polarization in the invariant atomic charge representation)[17] should also be considered. For thse reasons, we do not claim that the approaches here provide exact calculation of binding affinities, but the agreement with experiment (e.g. with experimental K_d) is encouraging and indicicates that the models here provide a good description of the thermodynamics of binding.

Circular dichroism titrations

Circular dichroism (CD) experiments were conducted using a spectropolarimeter (J-810, Jasco, Japan) fitted with a Peltier temperature controller. Measurements were taken in a quartz cuvette with a path length of 5 mm at 20 °C, at a 100 nm/min scanning speed at 1 nm intervals with a 1 nm bandwidth. The DNA concentration was 4.2 μM in 100 mM sodium phosphate buffer, pH 7.4. The CD spectra were recorded over the range 550 – 200 nm and baseline corrected for the buffer used. During the titrations, aliquots of ligand were added from a 1 mM stock solution in buffer (containing 10% DMSO for solubility). The sample was mixed thoroughly and the CD spectrum acquired immediately. The reported spectrum for each titration point represents the average of 3 scans. Data processing was carried out using Prism 7 with an 8-point second order smoothing polynomial applied to all spectra. Observed ellipticities were converted to molar ellipticity.

NMR titrations

All NMR spectra were recorded at 298 K on a 600 MHz Varian VNMRS spectrometer equipped with a triple resonance cryogenically cooled probe head. The DNA concentration was 185 μM (1D titration experiments) or 300 μM (2D NOESY experiments) in 25 mM sodium phosphate buffer, pH 7.0, containing 70 mM sodium chloride and 10% D_2O. All experiments employed sculpted excitation water suppression. Samples were referenced to the triethylammonium counterion, δ CH_3 = 1.297. During the 1D ^1H titrations, aliquots of ligand were added from a 10 mM stock solution in DMSO-d_6 to give the required titration points. The sample was mixed thoroughly and the NMR spectrum acquired immediately. NOESY spectra were recorded with a 300 ms mixing time, on telo22 alone and in the presence of 2 equiv. ligand, with the ligand added in the same manner as for 1D experiments. The imino (H1), aromatic (H2/H6/H8) and carbohydrate (H1', H2', H2'') resonances of telo22 in the absence of ligand were assigned by comparison with previously published data.[3] All spectra were processed with MestReNova software. The dissociation constants for ligands **3** and **4** were estimated by fitting the chemical shift perturbations of the indicated imino resonances to a standard 1:1 binding model appropriate to NMR spectroscopy using BindFit software.[18,19]

SUPPLEMENTARY FIGURES

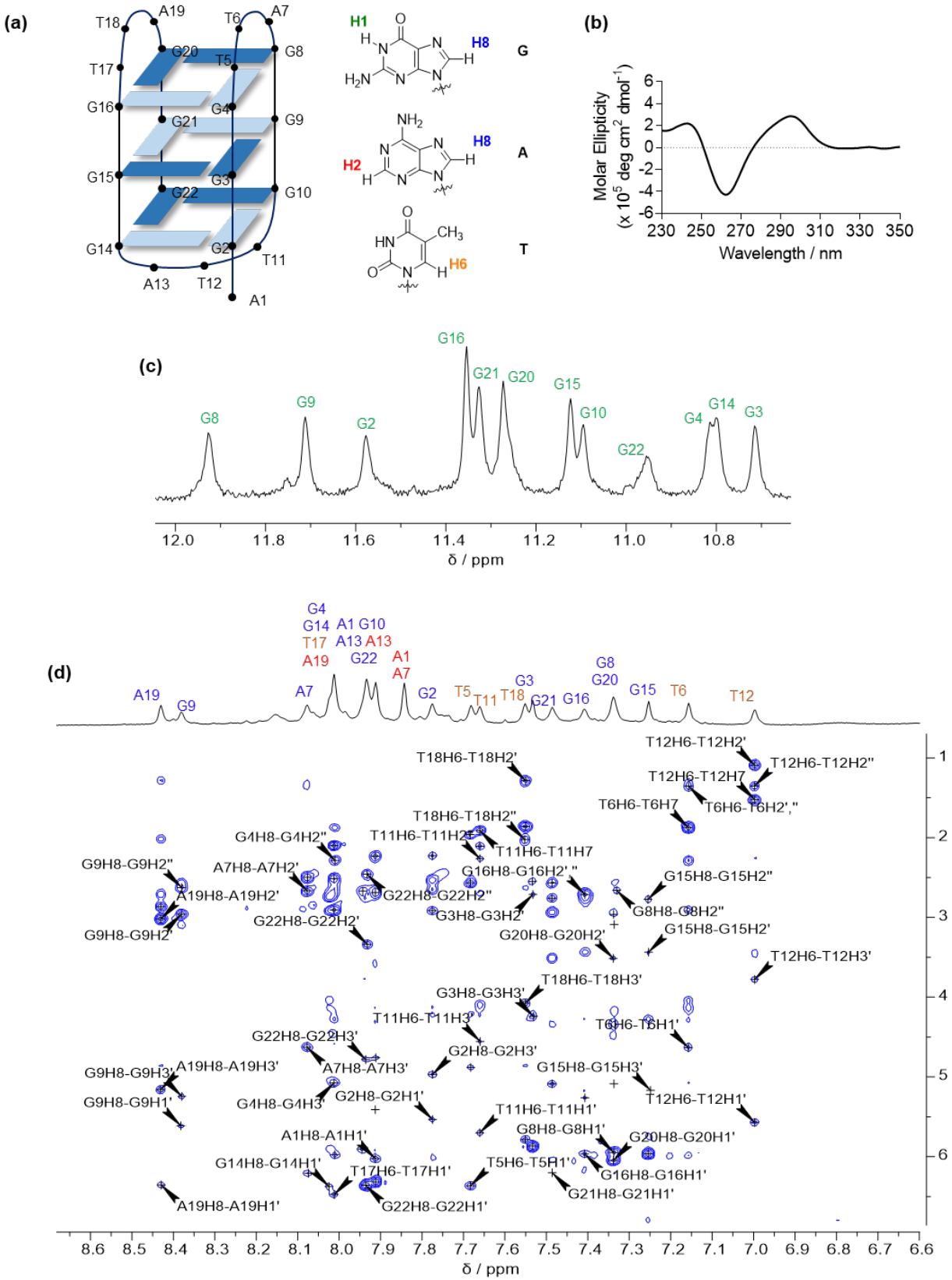

Figure S1: assignment of spectral data of antiparallel-basket telo22. (a) cartoon representation of the antiparallel fold showing base numbering and alternating *syn*- (dark blue) and *anti*- (light blue) conformation of guanosine residues, and colour-coded labelling of G, C and T nucleobase proton resonances; (b) CD spectrum (100 mM sodium phosphate buffer, pH 7.4); (c) assigned ^{1}H 1D NMR spectrum of imino region; (d) assigned ^{1}H 1D and 2D NOESY spectrum of aromatic/carbohydrate spectral regions.

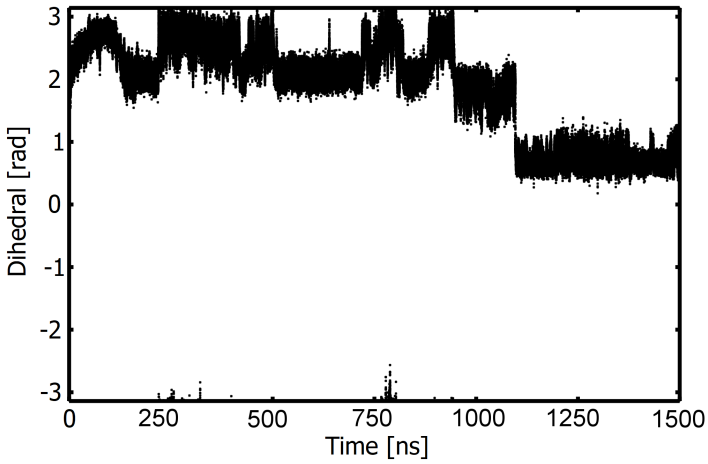

Figure S2. Plot of dihedral distribution of ligand **2** versus simulation time, corresponding to the histogram displayed in Figure 2a.

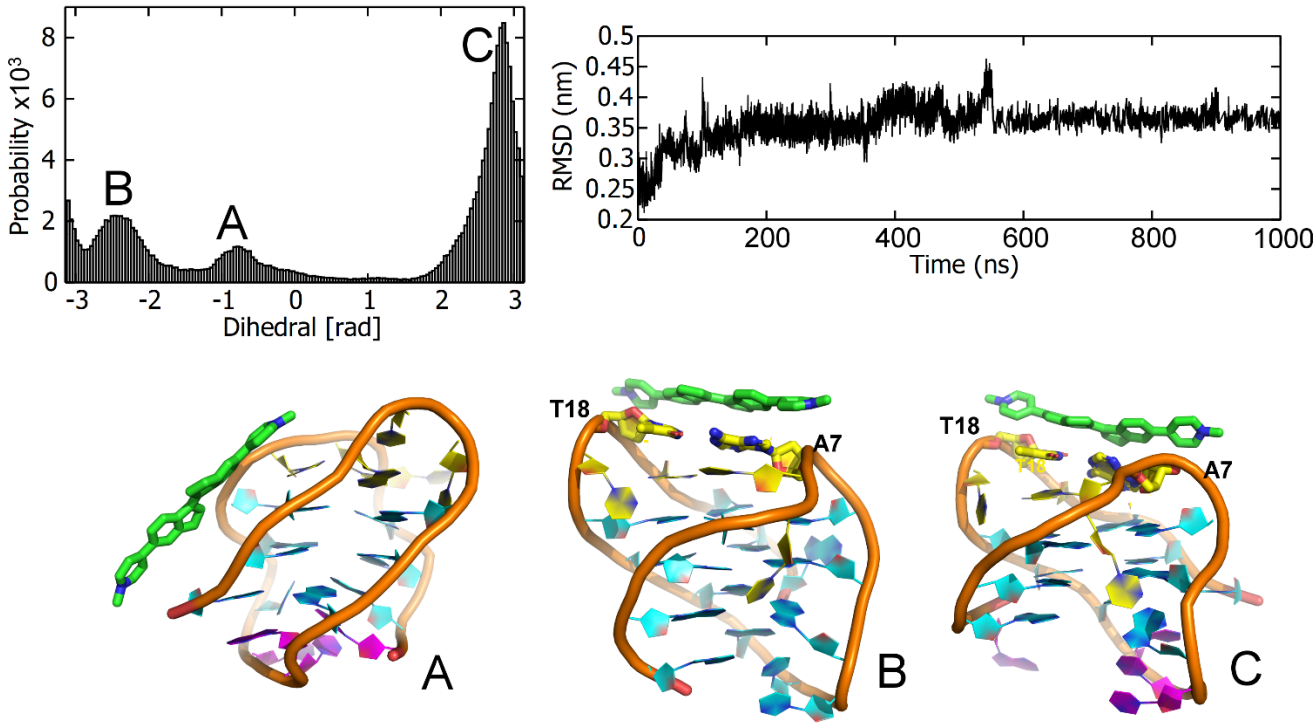

Figure S3. Analysis of the second docked pose of the binding of **2** to antiparallel telo22 investigated by MD simulations. (a) dihedral angle distributions sampled in the MD simulations, (b) representations of the principal binding poses sampled in the MD simulations, key binding residues are indicated, (c) RMSD of the DNA backbone co-ordinates over the time course of the MD simulations.

The figure depicts the results of the binding conformations sampled from the second docked pose of **2** to telo22 within the 1 μs unbiased MD simulation. Similar to the MD studies of the first docked pose, the calculations predict another major groove binding conformation where **2** is bound in another site of the telo22 major groove relative to the first docked pose (see main text). The resulting dihedral angle distribution shows three peaks which confirms three stable binding conformations (A, B and C). Pose A confirms the docking conformation, suggesting that it is well recovered by the ligand dynamics. As the simulation progresses, **2** shifts position from the groove site to the top face of the telo22, interacting mainly with top face bases such as T18 and A7. The binding conformation then remains same until the end of the simulation, with a slight change in the ligand orientation (Pose C).

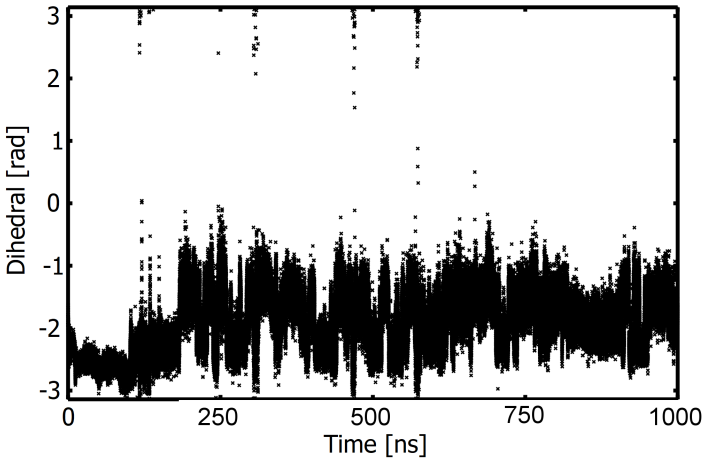

Figure S4. Plot of dihedral distribution of ligand **3** versus simulation time, corresponding to the histogram displayed in Figure 3a.

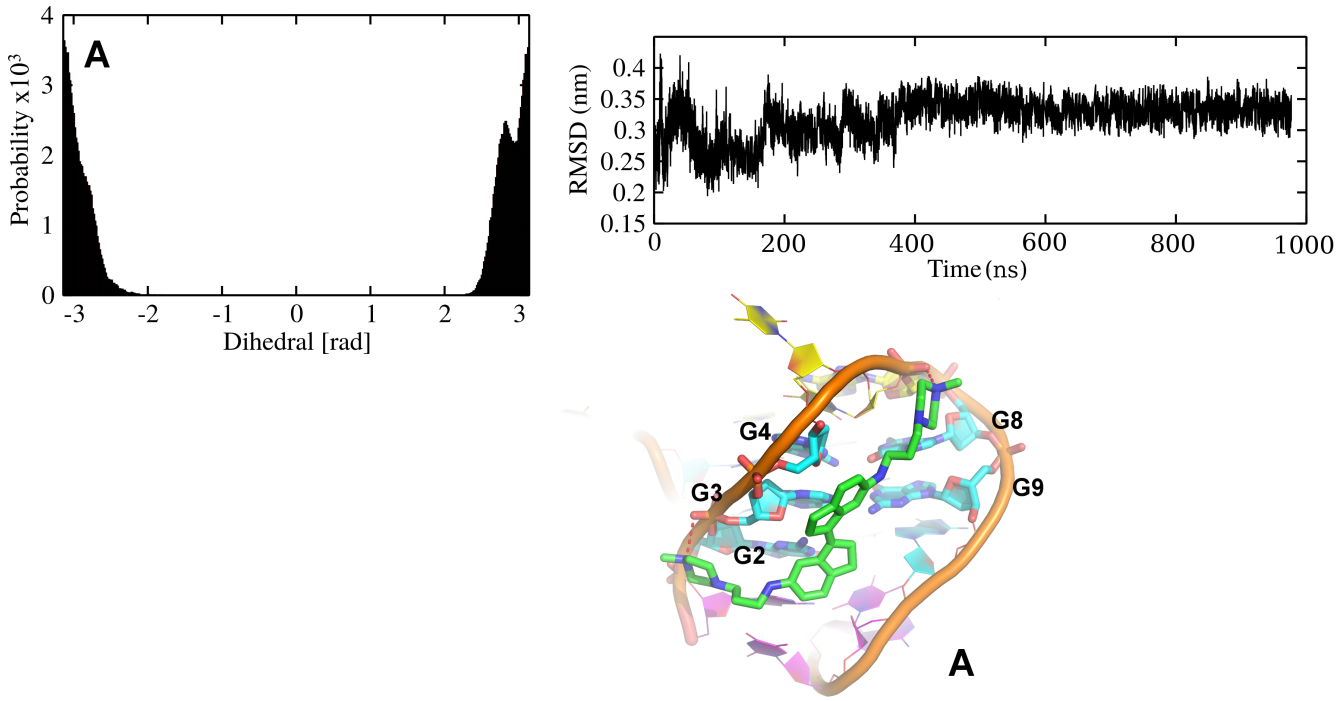

Figure S5. Analysis of the second docked pose of the binding of **3** to antiparallel telo22 investigated by MD simulations. (a) dihedral angle distributions sampled in the MD simulations, (b) representations of the principal binding poses sampled in the MD simulations, key binding residues are indicated, (c) RMSD of the DNA backbone co-ordinates over the time course of the MD simulations.

The figure depicts the MD simulation results of the second docked pose of ligand **3** in which the ligand is bound to the major groove of telo22 G4. The resulting distribution shows a single peak confirming only one single conformation being sampled by **3** during 1 μs long unbiased MD simulation. Pose A shows the interaction between **3** and telo22 G4 where the ligand interacts partially (the edges of the bases and the backbone phosphate groups) with bases from all three G-tetrads and they are G4, G8 and G3, G9 and G2 from the top, middle and bottom G-tetrads, respectively. Moreover, besides these interactions, it is also seen that both the tails of the ligand are involved in H-bond interactions with the sugar phosphate backbone of DNA (shown with red dots). Overall, the MD simulation of both the docked poses show that **3** has higher possibility of binding both the top face and groove sites.

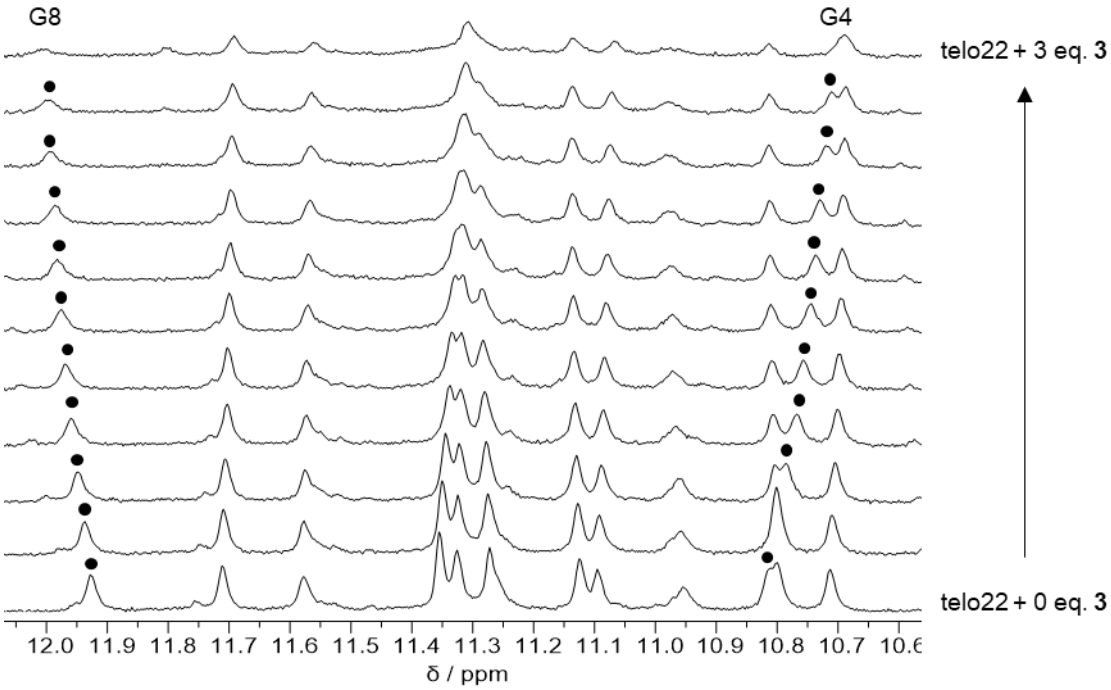

Figure S6: Stacked 1D ^{1}H NMR spectra (imino region) of telo22 titrated with ligand **3**. DNA concentration = 185 μM.

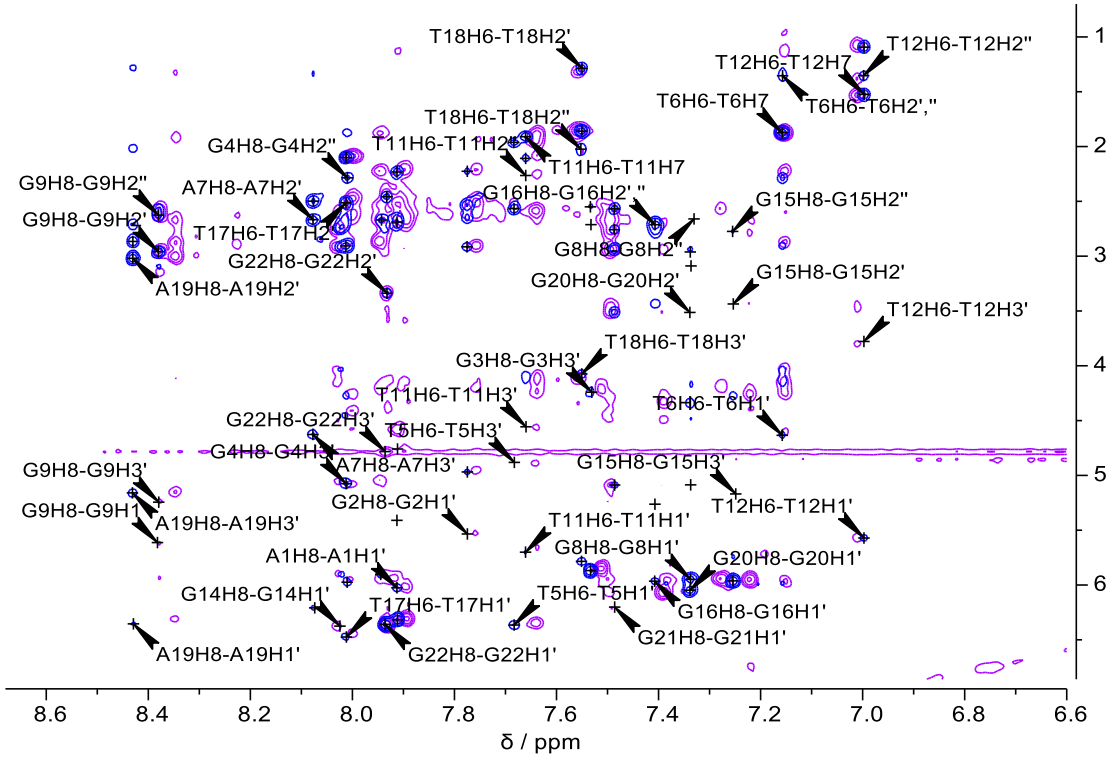

Figure S7: Overlaid 2D NOESY spectra of telo22 alone (blue) and in the presence of 2 equiv. (purple) ligand **3**. DNA concentration = 300 μM.

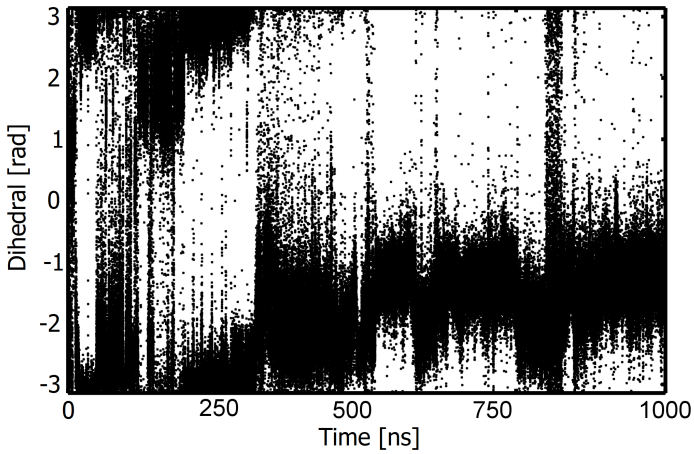

Figure S8. Plot of dihedral distribution of ligand **4** versus simulation time, corresponding to the histogram displayed in Figure 5a.

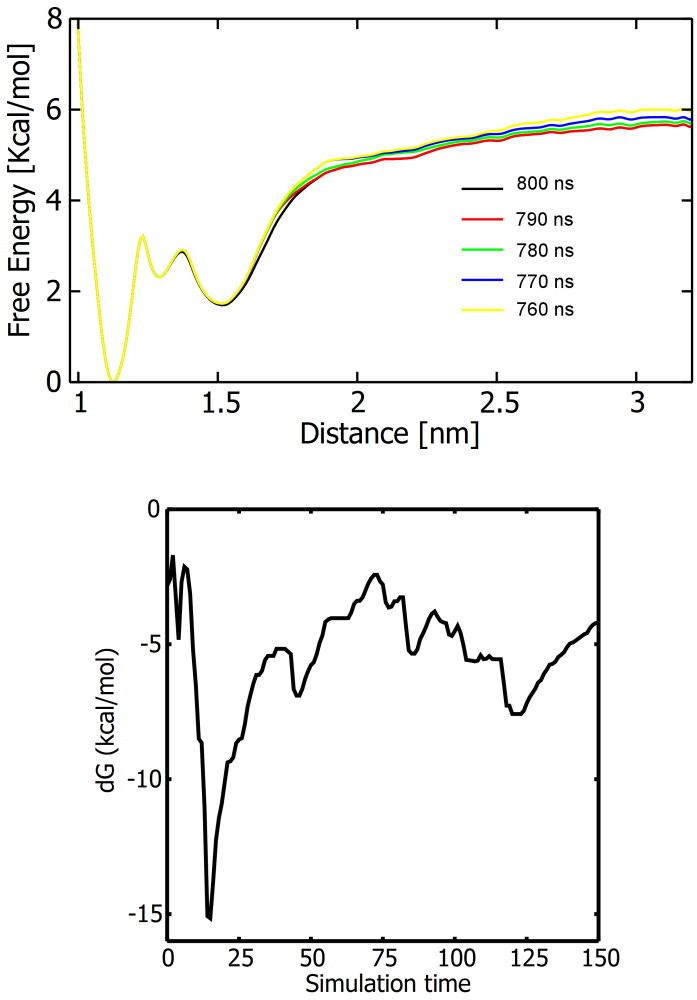

Figure S9: *Upper*: The one-dimensional free energy profile projected on Distance (D) CV of the binding of ligand **4** to the telo22 G4 DNA. The free energy profile is monitored in the last 50 ns of the simulation with a stride of 10 ns. The free energy difference between the 760 and 800 ns is less than 5 kcal/mol indicating that the simulation is well converged. *Lower:* To further demonstrate convergence, the difference in the free energy between bound and unbound states is plotted against the simulation time. Here, the simulation time is divided into 150 blocks of 10 ns and the free energy difference is calculated for each block. The block analysis shows that the free energy is converged after 800 ns of simulation time.

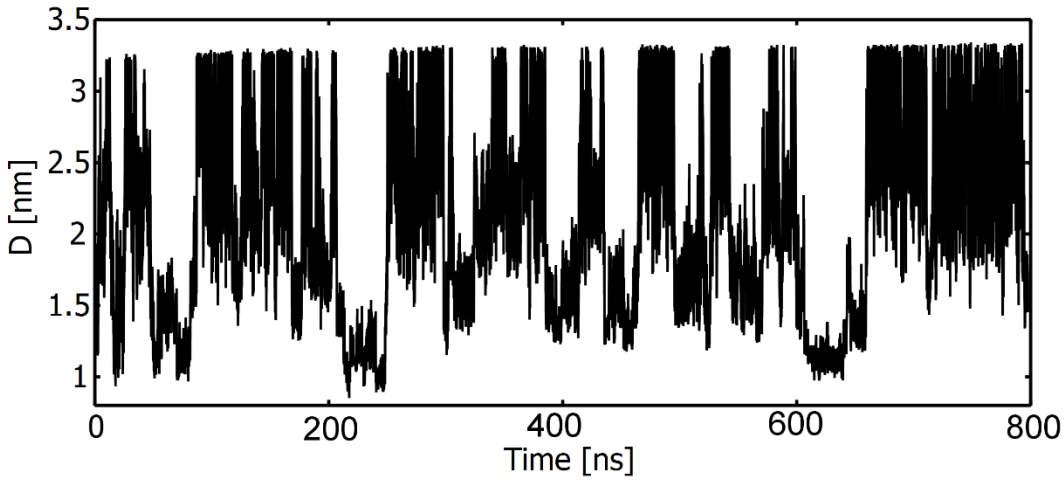

Figure S10: The exploration of the distance (D) collective variable with respect to the simulation time showing the multiple recrossing events from bound to unbound states and visa-versa.

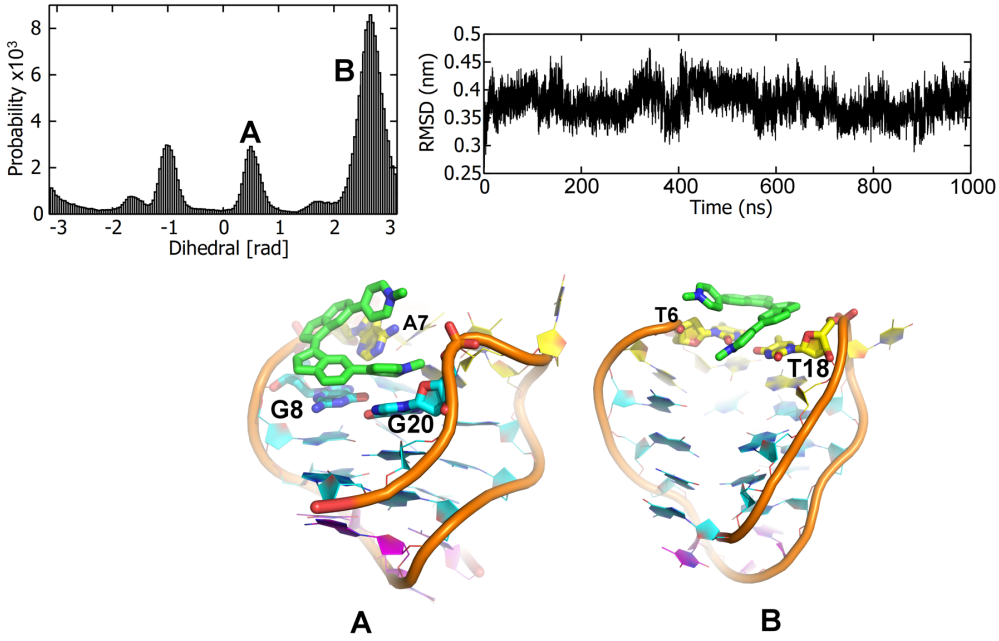

Figure S11: Analysis of the second docked pose of the binding of **4** to antiparallel telo22 G4 DNA investigated by MD simulations. (a) dihedral angle distributions sampled in the MD simulations, (b) representations of the principal binding poses sampled in the MD simulations, key binding residues are indicated, (c) RMSD of the DNA backbone co-ordinates over the time course of the MD simulations

The Figure depicts the MD simulation results of the second docked pose of ligand **4**. All the binding conformations are sampled within the 1 μs unbiased standard MD simulation. The docking calculation predicts that ligand **4** binds with the same residues as predicted in first docked pose but with opposite orientation (see main text, Pose A in Figure 5). The resulting dihedral distribution shows two major peaks A and B where Ligand **4** is bound to the A_7 base with stacking interaction and partially interacting with the **G_8 and G_{20}** bases (Pose A). As the simulation progresses, ligand **4** sampled another major conformation, B. At the start of the simulation, ligand **4** slides down towards major groove and partially interacts with the **G_8 and G_{20}** bases via stacking which is also found in the WT-MetaD simulation (see main text, Basin D, Figure 6). The MD simulation of both the first and second docked poses of ligand **4** predict that it has a favoured tendency to slide to interact with both the **G_8 and G_{20}** bases from the top G-tetrad indicating the possibility for ligand **4** binding to the top G-tetrad. In the next phase of the simulation, ligand **4** slides upward again (as observed in the MD simulation of first docked pose) and samples the top face of the telo22 by stacking with **T_6 and T_{18}** (Pose B). A careful investigation of the whole 1 μs long trajectory suggests that ligand **4** is sliding up and down the helix, with **A7** being as the primary base for binding. At the end of the simulation, the overall RMSD of the DNA backbone is found to be constant at ~0.36 nm (3.6Å) which again suggest that telo22 is stable during the MD simulation.

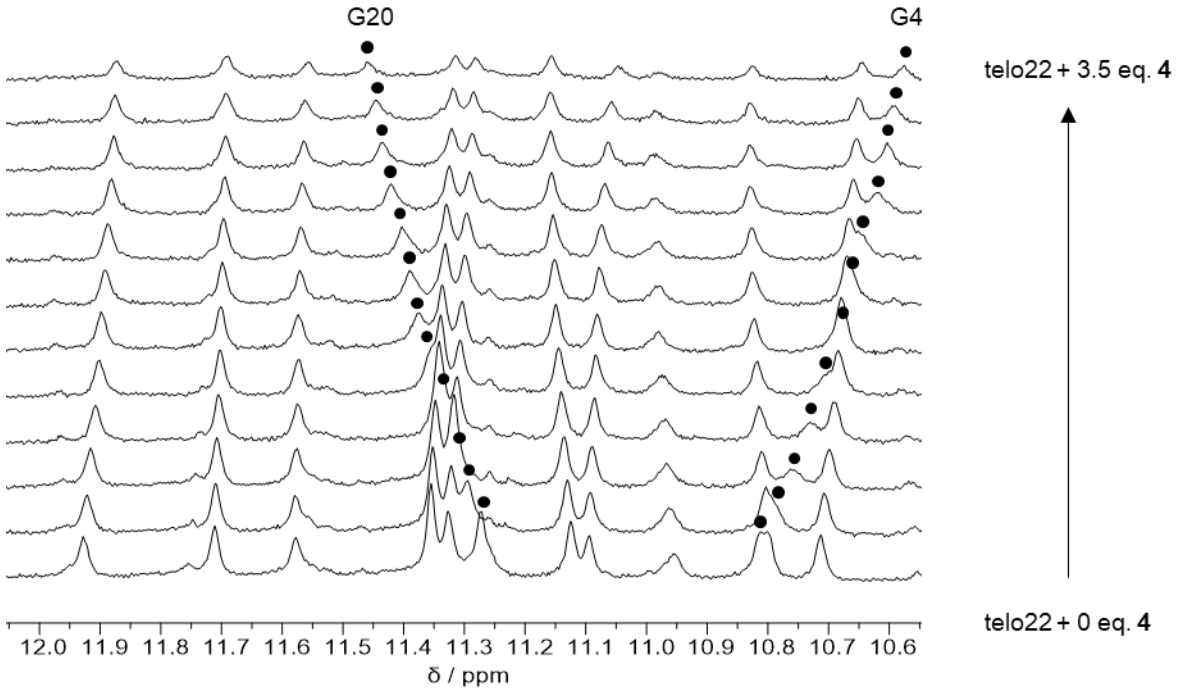

Figure S12: Stacked 1D ^{1}H NMR spectra (imino region) of telo22 titrated with ligand **4**. DNA concentration = 185 μM.

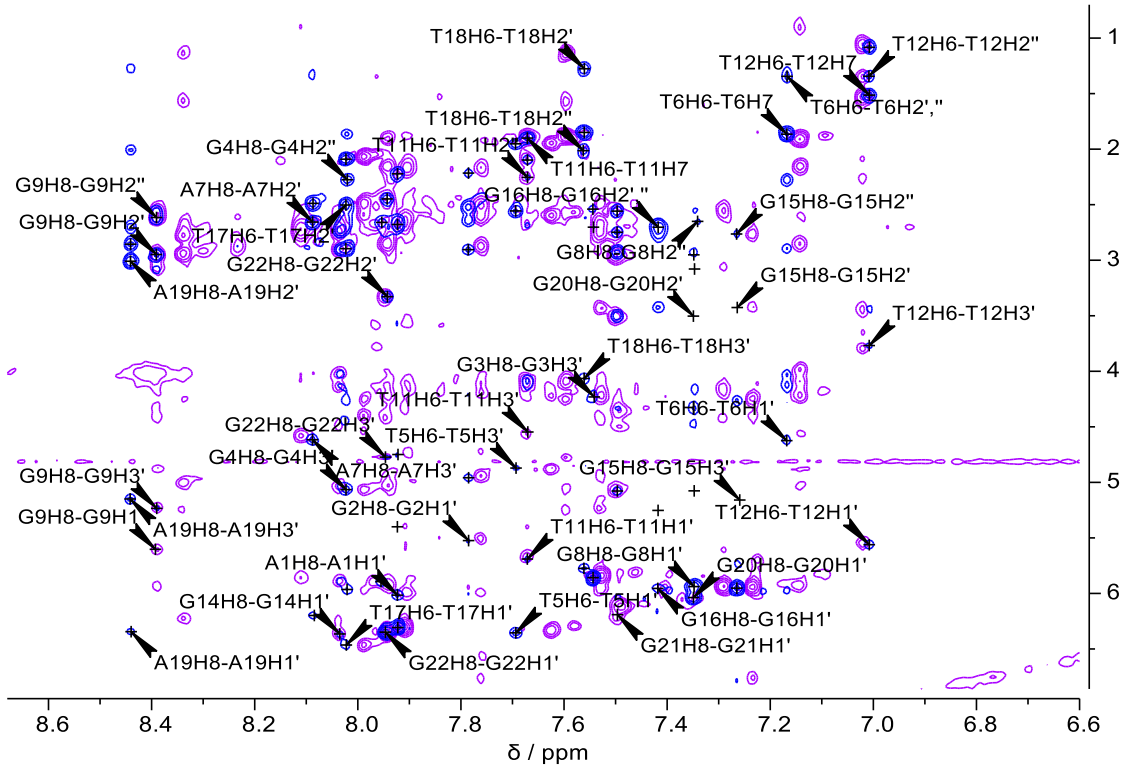

Figure S13: Overlaid 2D NOESY spectra of telo22 alone (blue) and in the presence of 2 equiv. (purple) ligand **4**. DNA concentration = 300 μM.

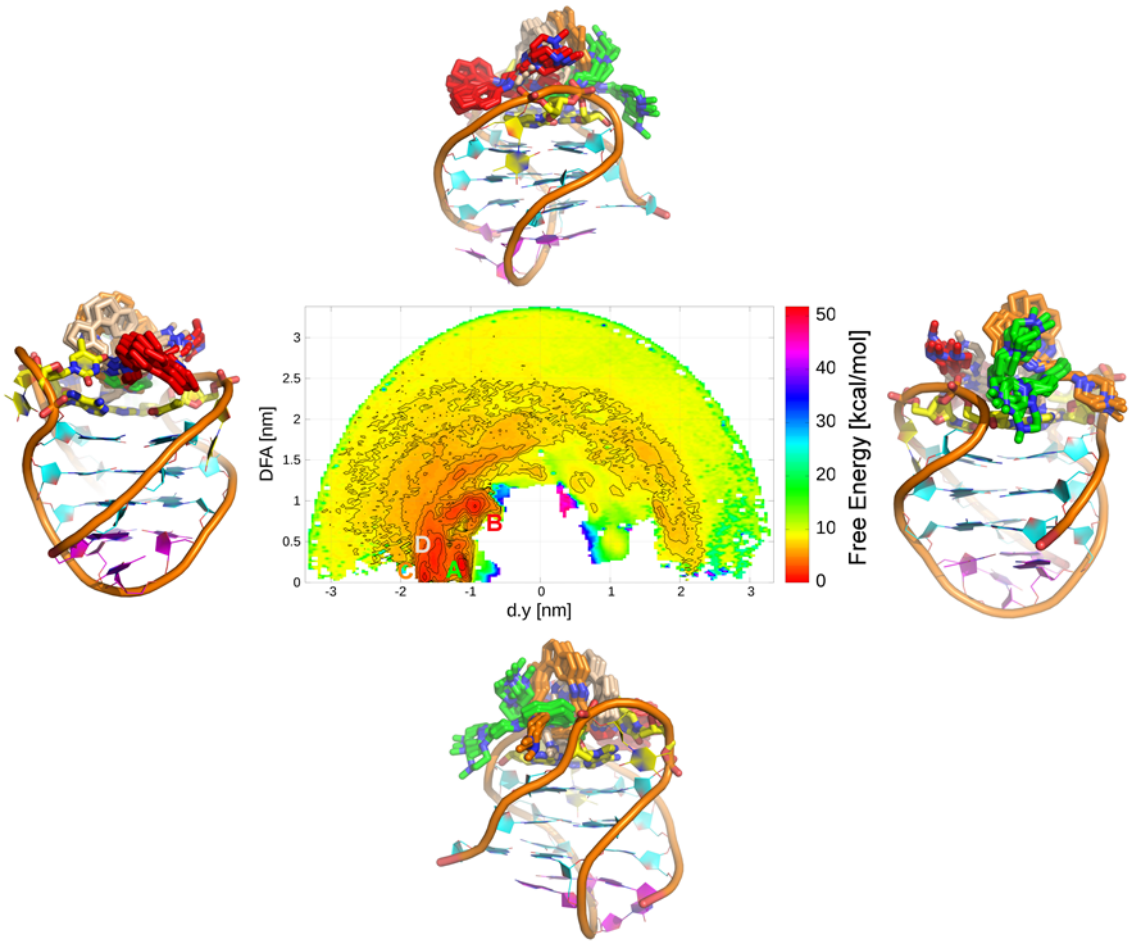

Figure S14: Free energy surface of telo22/ligand **5** binding process computed as a function of DFA and POA collective variables. The WT-MetaD simulation showed that the ligand **5** mainly binds on the top face of the G4 DNA. Moreover, it can interact with the guanine bases from the first G-quartet, such as G20, like ligand **4**. Color codes are showing different free energy minima.

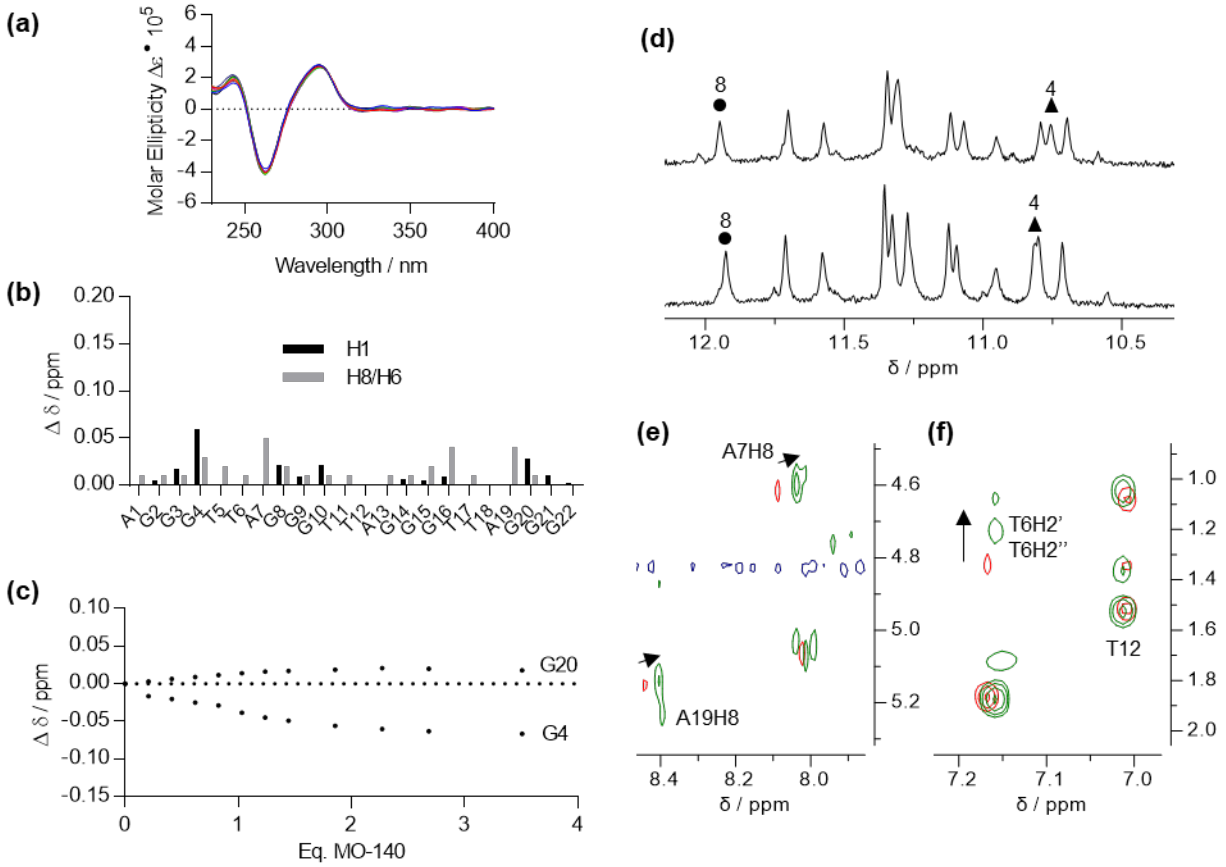

Figure S15: Binding of ligand **5** to antiparallel telo22 investigated by circular dichroism and NMR spectroscopy. (a) circular dichroism spectra of telo22 and increasing equivalents of ligand up to 7 eq. **5**, (b) changes in chemical shifts of imino an aromatic resonances of telo22 upon adding 2 eq. ligand **5**, (c)-(d) NMR titration of telo22 with ligand **5** showing shifts of key imino resonances, (e)-(f) partial NOESY spectra showing shifts of key (e) aromatic/anomeric (f) sugar aliphatic resonances.

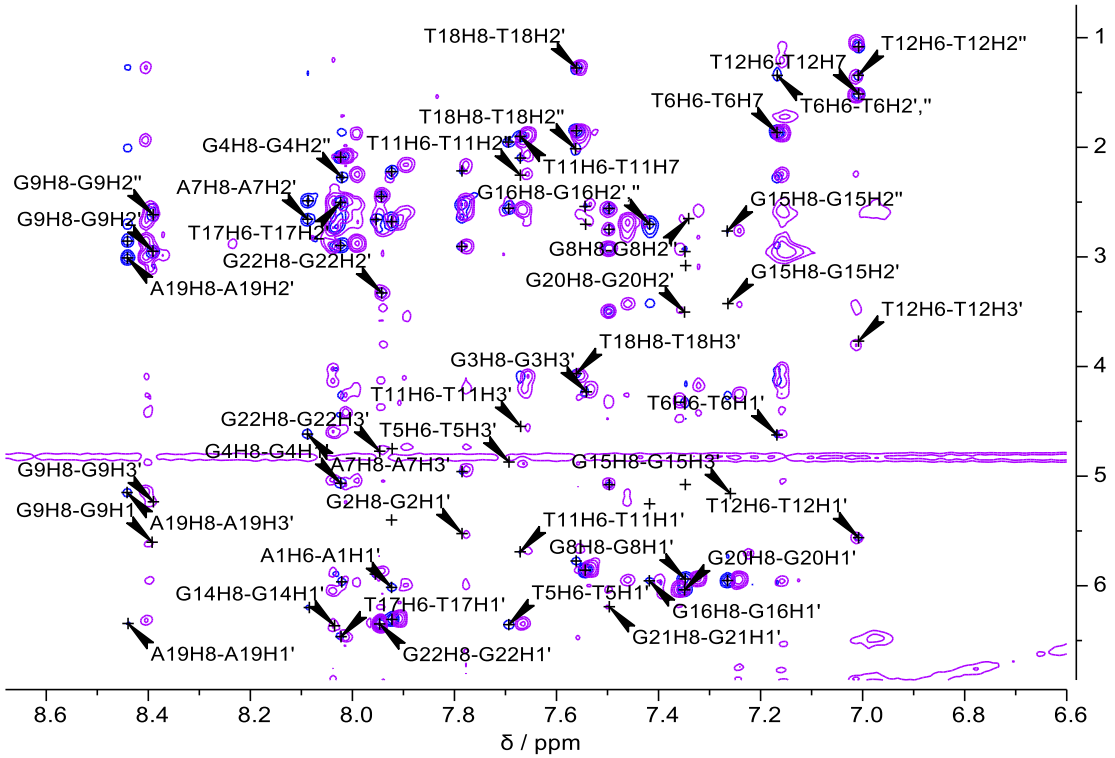

Figure S16: Overlaid 2D NOESY spectra of telo22 alone (blue) and in the presence of 2 equiv. (purple) ligand **5**. DNA concentration = 300 μM.

REFERENCES

[1] M. P. O'Hagan, P. Peñalver, R. S. L. Gibson, J. C. Morales, M. C. Galan, *Chem. – A Eur. J.* **2020**, *26*, 6224–6233.

[2] M. P. O'Hagan, S. Haldar, M. Duchi, T. A. A. Oliver, A. J. Mulholland, J. C. Morales, M. C. Galan, *Angew. Chemie Int. Ed.* **2019**, *58*, 4334–4338.

[3] Y. Wang, D. J. Patel, *Structure* **1993**, *1*, 263–282.

[4] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[5] O. Trott, A. J. Olson, *J. Comput. Chem.* **2009**, *31*, NA-NA.

[6] I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J. L. Gelpí, C. González, M. Vendruscolo, C. A. Laughton, S. A. Harris, D. A. Case, M. Orozco, *Nat. Methods* **2015**, *13*, 55–58.

[7] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

[8] C. I. Bayly, P. Cieplak, W. D. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269–10280.

[9] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci.* **2002**, *99*, 12562–12566.

[10] A. Barducci, G. Bussi, M. Parrinello, *Phys. Rev. Lett.* **2008**, *100*, 1–4.

[11] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, *Comput. Phys. Commun.* **2014**, *185*, 604–613.

[12] B. Hess, C. Kutzner, D. Van Der Spoel, E. Lindahl, *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

[13] O. Valsson, P. Tiwary, M. Parrinello, *Annu. Rev. Phys. Chem.* **2016**, DOI 10.1146/annurev-physchem-040215-112229.

[14] A. Barducci, M. Bonomi, M. Parrinello, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, DOI 10.1002/wcms.31.

[15] F. Moraca, J. Amato, F. Ortuso, A. Artese, B. Pagano, E. Novellino, S. Alcaro, M. Parrinello, V. Limongelli, *Proc. Natl. Acad. Sci.* **2017**, *114*, E2136–E2145.

[16] V. Limongelli, M. Bonomi, M. Parrinello, *Proc. Natl. Acad. Sci.* **2013**, *110*, 6358–6363.

[17] S. Haldar, F. Comitani, G. Saladino, C. Woods, M. W. van der Kamp, A. J. Mulholland, F. L. Gervasio, *J. Chem. Theory Comput.* **2018**, *14*, 6093–6101.

[18] D. Brynn Hibbert, P. Thordarson, *Chem. Commun.* **2016**, *52*, 12792–12805.

[19] *http://supramolecular.org*