Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Electronic Supplementary Information

DNA triplex structure, thermodynamics, and destabilisation: insight from molecular simulations

Belinda J. Boehm, Charles Whidborne, Alexander L. Button, Tara L. Pukala, David M. Huang

Contents

S1 DNA triplex structure

S1.1 Overall structure

DNA triple helices are formed when a short nucleic acid oligomer, known as a triplex-forming oligomer (TFO) binds within the major groove of a duplex. This binding divides the major groove into two asymetric groove, the Watson–Hoogsteen (W–H) and Crick–Hoogsteen (C–H) grooves as shown in Fig. S1.

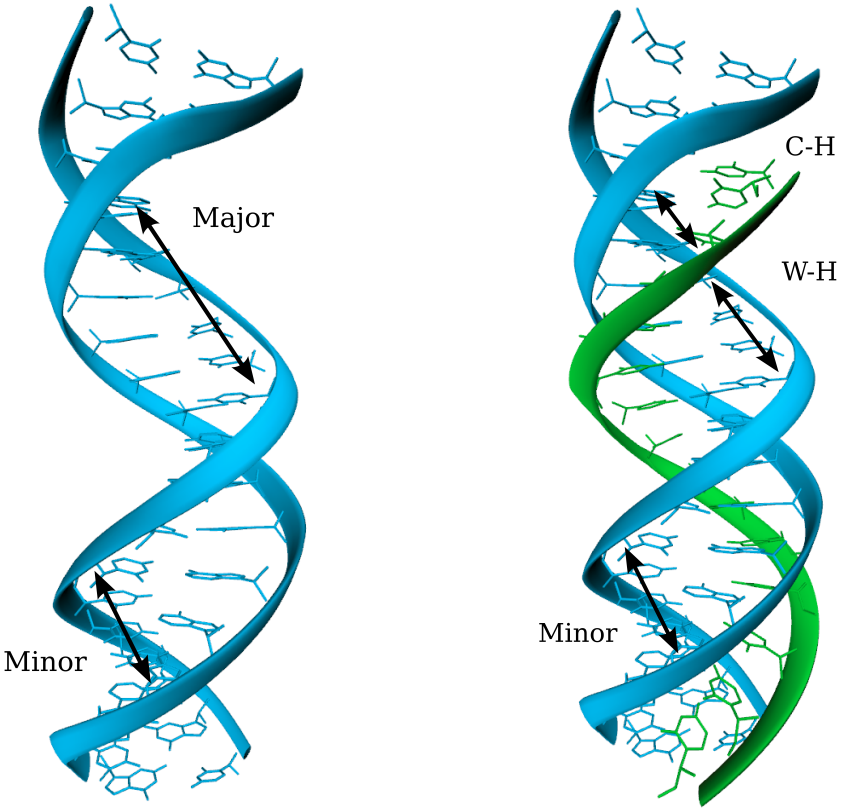


Fig. S1 Structure of a double helix (left) and triple helix (right). The TFO (green) binds in the major groove of the duplex.

S1.2 Hydrogen-bonding motifs

Hydrogen bonding between the TFO and purine (A/G) duplex strand can follow either the Hoogsteen or reverse-Hoogsteen motif, which differ by the atoms involved in the hydrogen bond. These, along with the general Watson-Crick motif present in DNA duplexes, are outlined in Fig. S2.

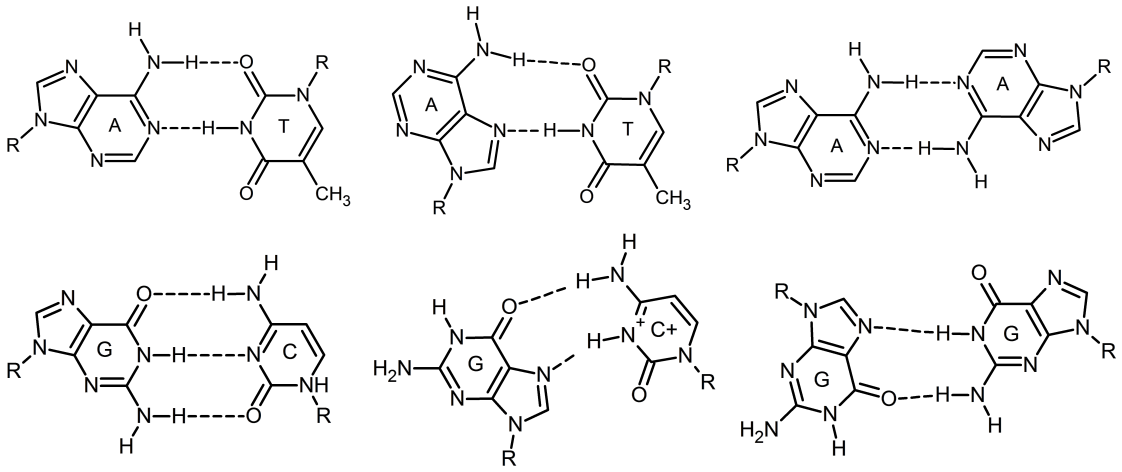


Fig. S2 Three different types of hydrogen bonding between DNA nucleobases. Left: Watson–Crick; centre: Hoogsteen; right: reverse-Hoogsteen.

S1.3 Directionality

On binding to the duplex, the TFO can be in one of two orientations with respect to the duplex strand to which it is bound: parallel or antiparallel. This is illustrated in Fig. S3. For a 5'–3' purine duplex strand, the antiparallel triplex forms when the TFO runs 3'–5' with respect to this strand, and the parallel when it runs 5'–3'.

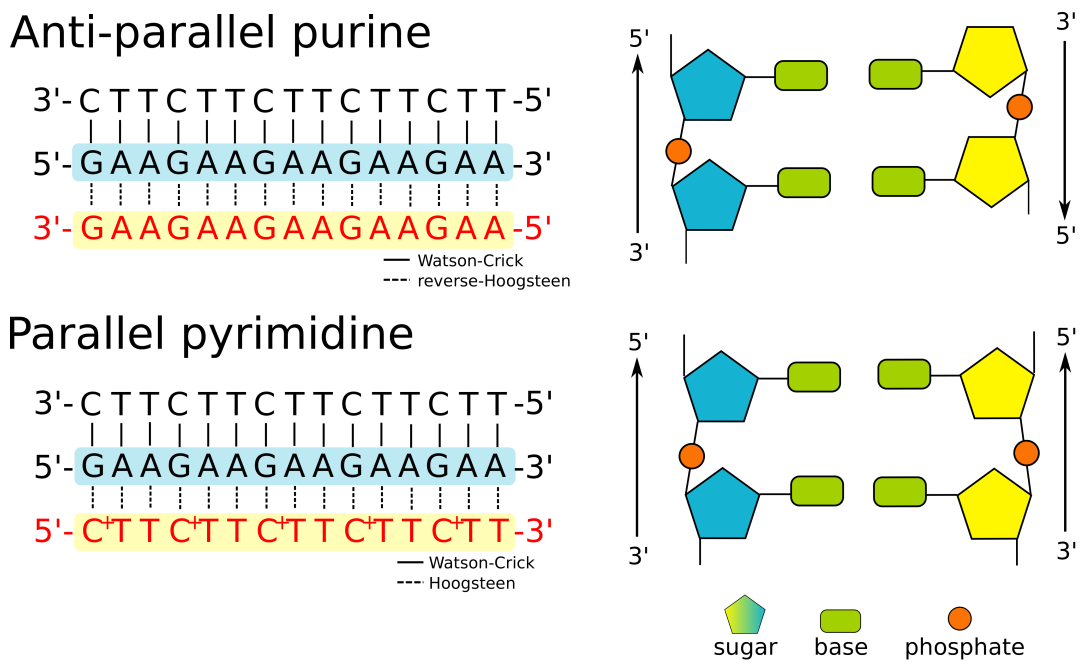


Fig. S3 (Left) 15-base antiparallel-purine and parallel-pyrimidine triplexes studied in this work. The TFO is shown in red. (Right) The directionality of the strands (3'–5' or 5'–3') refers to the orientation of the sugars of each strand relative to each other. With reference to the triplex structure on the left, the two strands shown are the central one, and TFO. In an antiparallel structure (top) the 3' terminus of strand 1 is at the opposite end of the molecule to the 3' terminus of strand 2. In the parallel structure (bottom) the 3' termini of each strand are at the same end of the molecule.

S2 Structures of netropsin derivatives

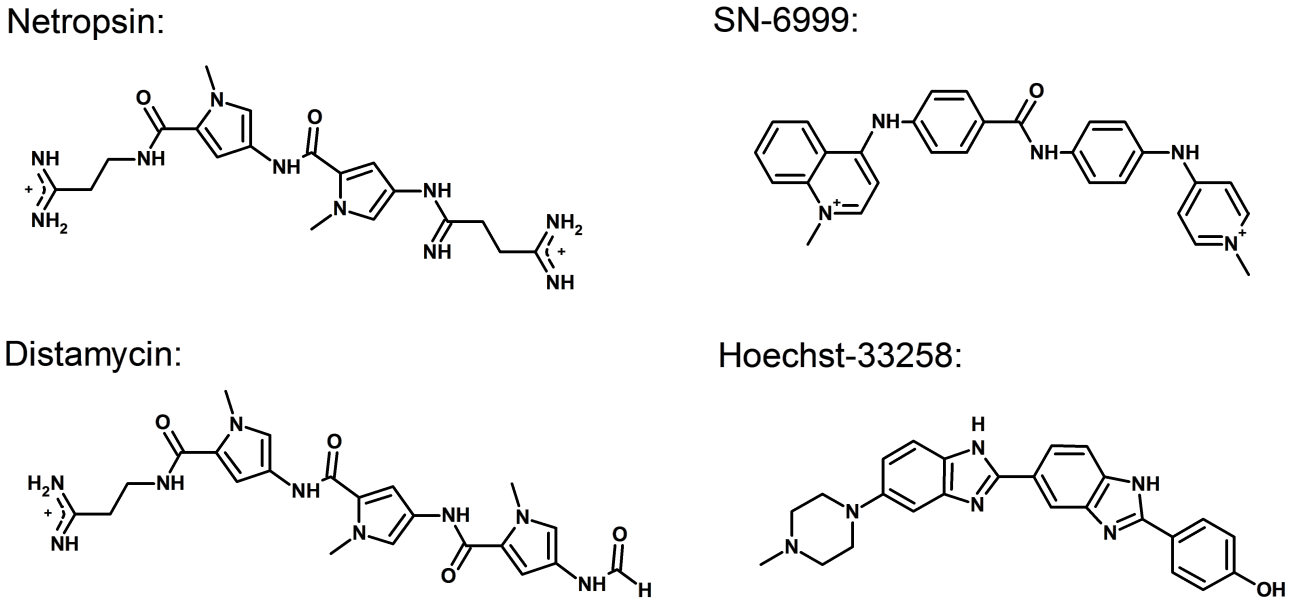
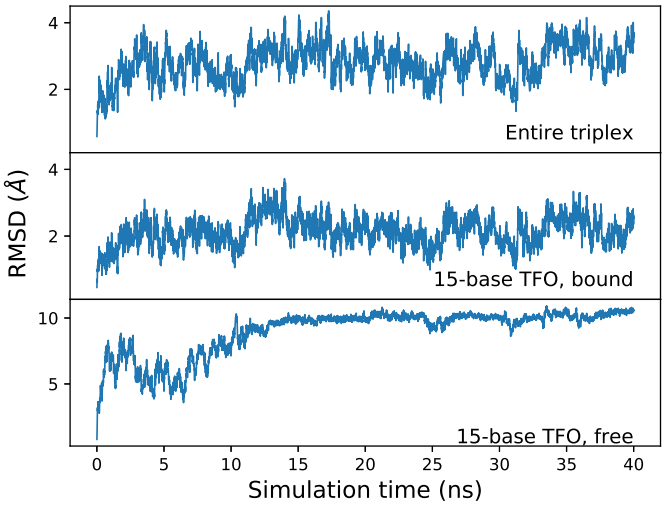


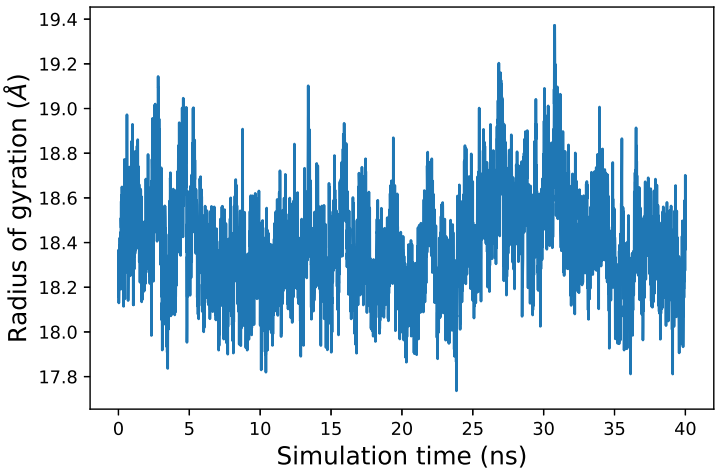
Fig. S4 The structures of netropsin and several structurally similar minor-groove binders. All feature a crescent shape, complementary to that of the minor groove, and a series of heterocyclic and aromatic hydrocarbon rings.

S3 Changes in structure and energetics during equilibration simulations

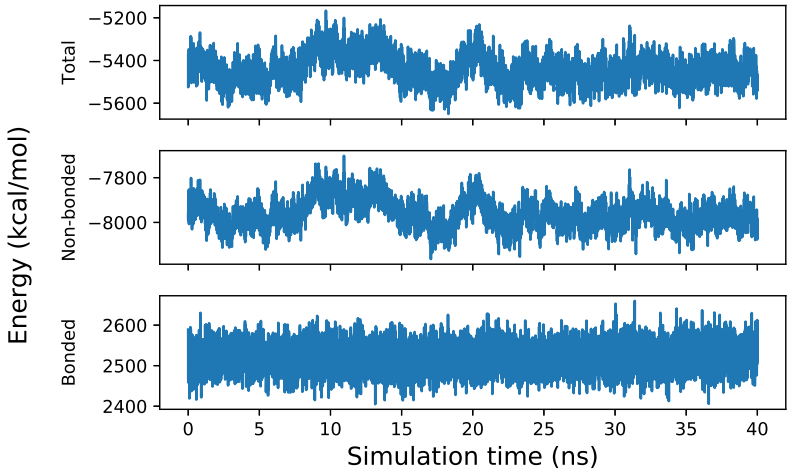
S3.1 Explicit solvent (40 ns)



(a) Root-mean-squared deviation (RMSD) of atom positions in the triplex or the free TFO.



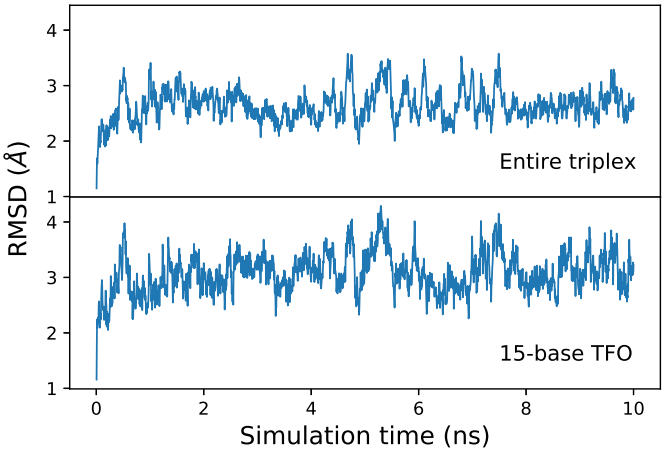
(b) Radius of gyration of the triplex.



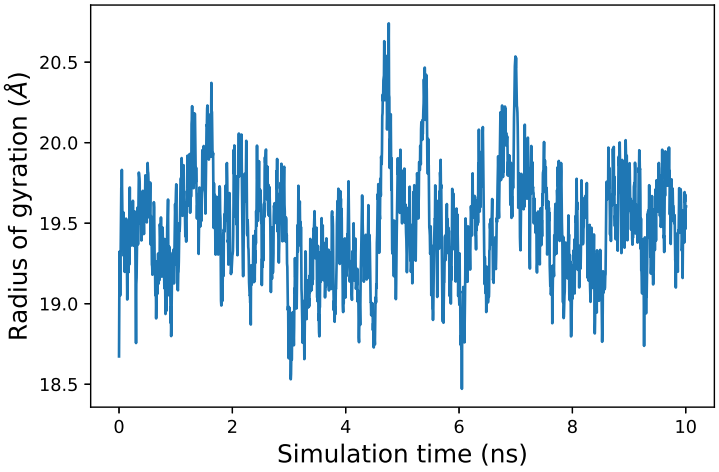
(c) Fluctuations in total, bonded, and non-bonded energies of the triplex.

Fig. S5 RMSD, radius of gyration, and energy over the course of a typical 40 ns equilibration simulation of the 15-base homopurine-TFO triplex with no netropsin in explicit solvent. RMSD is also shown for the free 15-base TFO in solution.

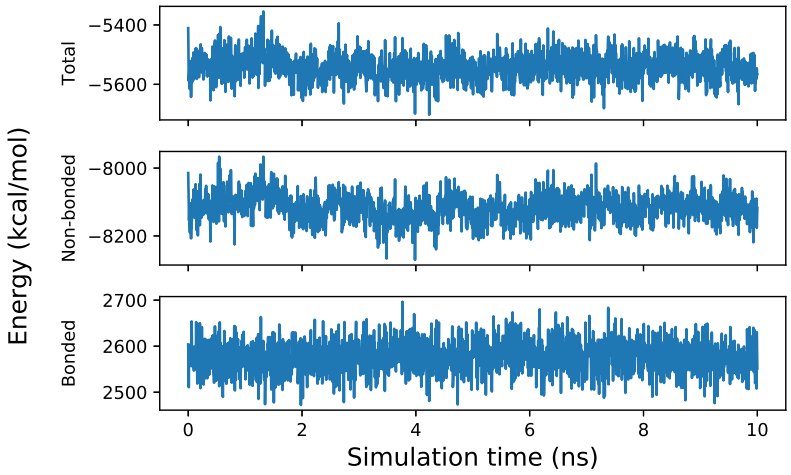
S3.2 Implicit solvent (10 ns)



(a) Root-mean-squared deviation (RMSD) of atom positions in the triplex.



(b) Radius of gyration of the triplex.



(c) Fluctuations in total, bonded, and non-bonded energies of the triplex.

Fig. S6 RMSD, radius of gyration, and energy over the course of a typical 10 ns equilibration simulation of the 15-base homopurine-TFO triplex with no netropsin in implicit solvent.

S4 Starting and equilibrium structures of the bound TFO

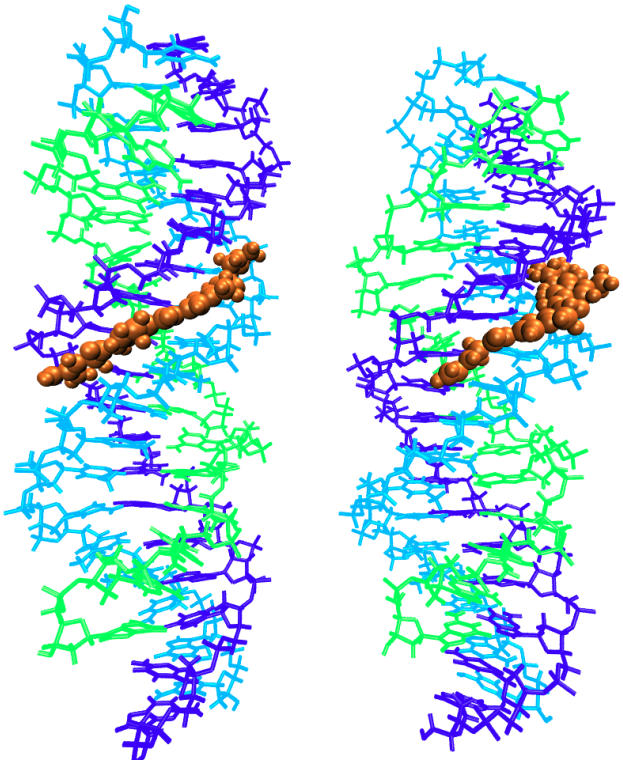


Fig. S7 Structure prior to equilibration (left) and the starting structure for FEP (right) for the 15-base homopurine TFO with netropsin (TFO green, duplex blue, netropsin orange).

S5 Convergence of free energy calculations

For each annihilation used to complete the thermodynamic cycle, 192 ns of simulation was required for the 15-base TFO. This relatively long time was required as the system equilibrated slowly, requiring almost 5 ns to equilibrate within each window (Fig. S8). Furthermore, structural fluctuations for the free TFO in solution appear to occur on a time scale on the order of 1–2 ns (Fig. S8), indicating that to effectively sample all conformations within each window, long simulation times are needed.

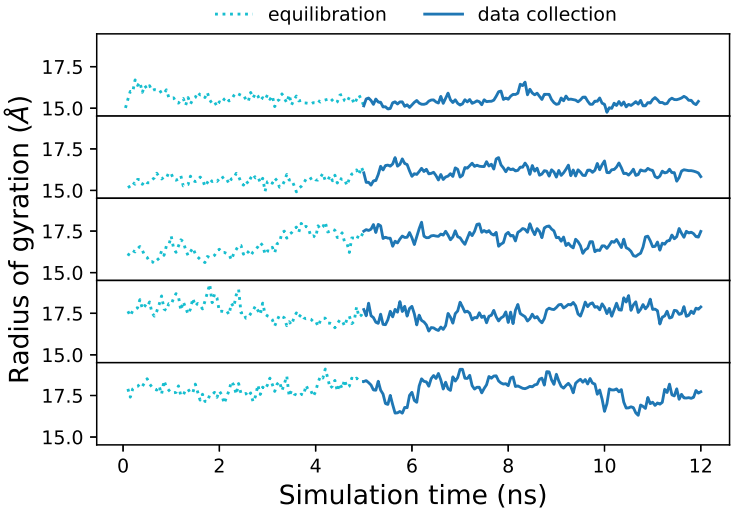


Fig. S8 Radius of gyration over the first 5 windows (top to bottom) of the annihilation of the free 15-base purine TFO in solution: equilibration period (light-blue dotted line); data collection period (dark-blue solid line).

Examining the convergence of the calculated free energies, for both the 3- and 15-base homopurine TFOs, the simulation time at each window appears to be sufficient that further increasing it is unlikely to have a significant effect on the calculated free energy (Figs. S9, S10).

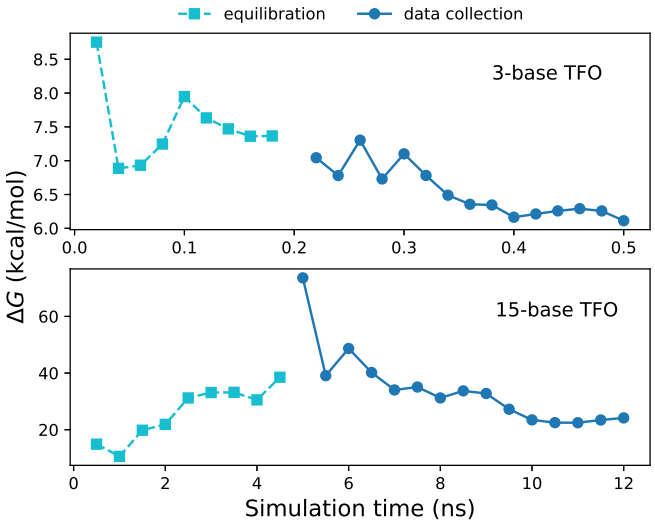
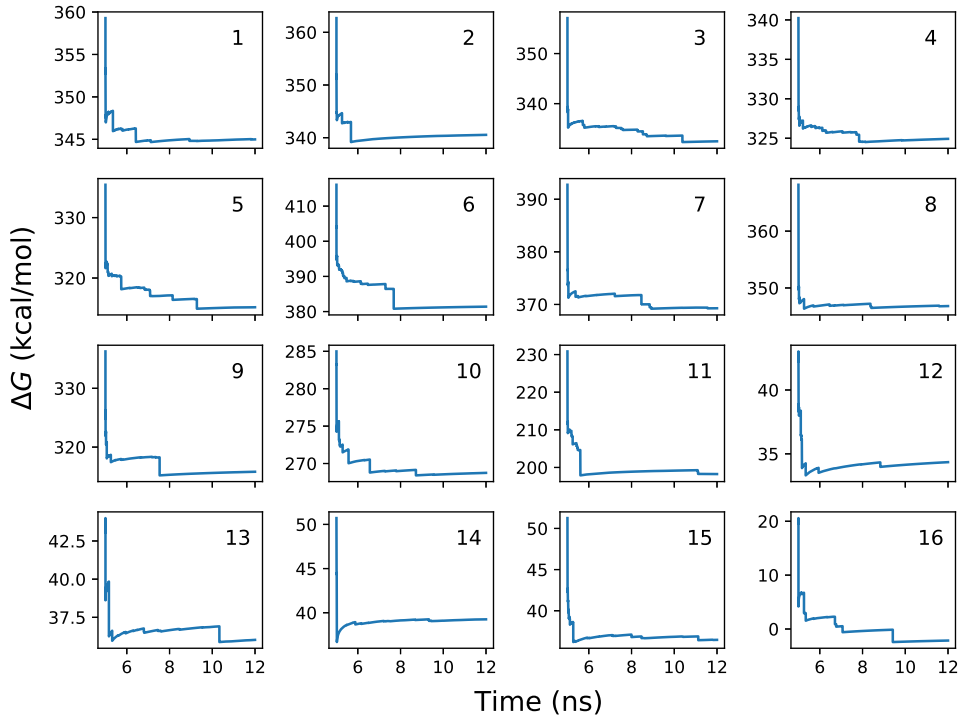
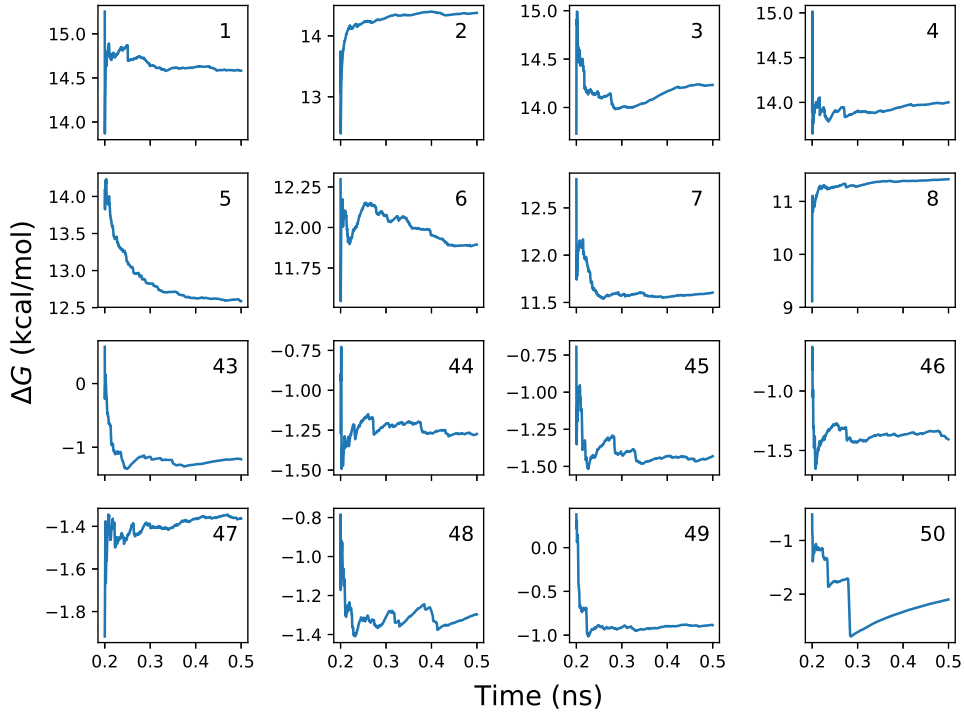


Fig. S9 Convergence of FEP simulations for annihilation of the TFO in the 3- and 15-base homopurine-TFO triplexes: equilibration period, which is not included in the calculation of ΔG_{dissoc} (light-blue dotted line); data collection period (dark-blue solid line). Each point is the value of the free energy calculated using data from each window up to the specified time excluding the equilibration period. Approaching the end of the simulation, the calculated ΔG does not vary significantly.



(a) The variation in free energy over 16 windows of the free energy calculation for the annihilation of the bound 15-base homopurine TFO in the presence of netropsin. Data was collected after a 5 ns equilibration period at each of the 16 windows.



(b) The variation in free energy over the first and final 8 windows of the free energy calculation for the annihilation of the bound 3-base homopurine TFO in the presence of netropsin. Data was collected after a 0.25 ns equilibration period at each window of 50 windows.

Fig. S10 Measures of the convergence of the FEP simulations of the annihilation of 3- and 15-base homopurine TFOs. It generally appears that the time scale chosen for simulation at each window is sufficient for convergence to be reached.

S6 Concentration dependence of binding of netropsin to the triplex

The concentration of netropsin required in solution for total dissociation of the triplex can be analysed as follows. The binding of netropsin to DNA can be modelled approximately using the Langmuir adsorption model, which considers non-cooperative binding of an adsorbate (here netropsin) to a substrate (here DNA). Using this model, the fractional coverage of the surface (fraction of occupied bases sites in the minor groove), θ, in terms of the solution concentration of netropsin, [net], can be expressed as

$$\theta = \frac{K[\text{net}]}{1+K[\text{net}]}, \tag{S1}$$

giving

$$K[\text{net}] = \frac{\theta}{1-\theta} \tag{S2}$$

for the binding constant K, which is related to the binding free energy ΔG of netropsin to DNA by

$$K = \exp\left(\frac{-\Delta\text{G}}{k_B T}\right). \tag{S3}$$

Netropsin was found in the simulations presented in this work to have a binding site of approximately 7 bases. From the free energy calculations, it can estimated that two netropsin molecules are required to completely unbind a 15-base TFO. Therefore, 14 out of 15 base sites will be occupied, giving $\theta_{\text{min}} = 14/15$ as the minimum fractional coverage to completely unbind the TFO. Thus, the minimum concentration of netropsin, $[\text{net}]_{\text{min}}$, required for full triplex destabilisation is

$$[\text{net}]_{\text{min}} = \frac{1}{K}\frac{\theta_{\text{min}}}{1-\theta_{\text{min}}} \tag{S4}$$

$$= \exp\left(\frac{\Delta\text{G}}{k_B T}\right)\frac{\theta_{\text{min}}}{1-\theta_{\text{min}}} \tag{S5}$$

$$= \exp\left(\frac{\Delta\text{G}}{k_B T}\right)\frac{14/15}{1-14/15} \tag{S6}$$

$$= 14\exp\left(\frac{\Delta\text{G}}{k_B T}\right) \tag{S7}$$

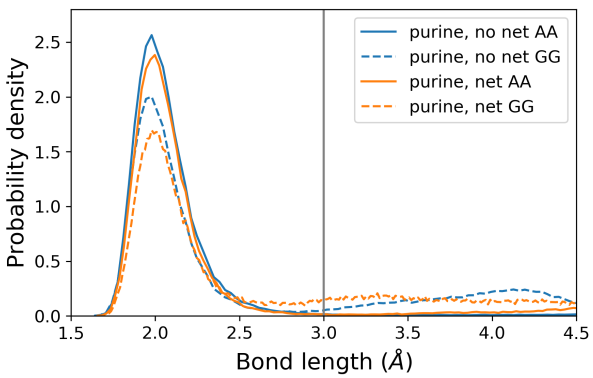
Thus

$$\Delta\text{G} = k_B T \ln\left(\frac{[\text{net}]_{\text{min}}}{14}\right). \tag{S8}$$

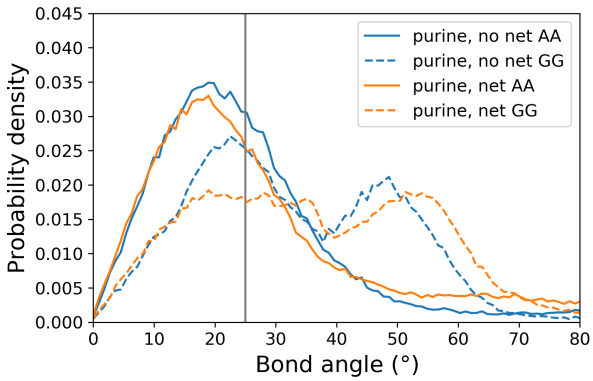
Given a typical LD50 for drugs of 10-100 μM, a conservative estimate of the maximum concentration of netropsin tolerable in a biological system would be $[\text{net}]_{\text{min}} = 1$ μM. Using the above equation, this would require a binding free energy of netropsin to triplex of approximately -7 kcal/mol.

S7 Hydrogen bond length and angle distributions

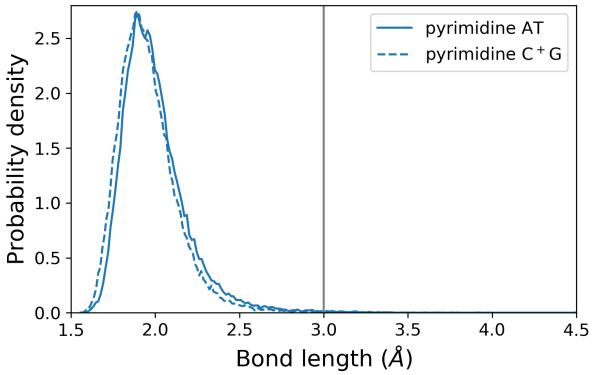
The average length and angle of each hydrogen bond between nucleotide bases were calculated using the bond length and angle definitions in Fig. 3 of the main paper. Additionally, the distributions of the bond lengths/angles for each type of hydrogen bond (AA-rH, AT-H, GG-rH, C^+G-H) were calculated with the results shown in Figs. S11 and S12. Angles are presented as deviation from linearity, ie. the difference between 180° and the given angle. Comparing the one-dimensional distributions of either bond length or angle for the purine-TFO triplexes with and without netropsin (Fig. S11a, b), both AA and GG peaks broaden and shift to slightly longer lengths on adding netropsin, indicating a weakening of hydrogen bonds. Examining the two-dimensional distributions (Fig. S12), while all structures are predominantly localised in the low length/angle region, both purine triplexes show a number of hydrogen bonds which are at higher length/angles. This is enhanced further for the structure in which netropsin is bound, indicating again that its binding likely reduces the strength of the interactions between hydrogen-bonding atoms.



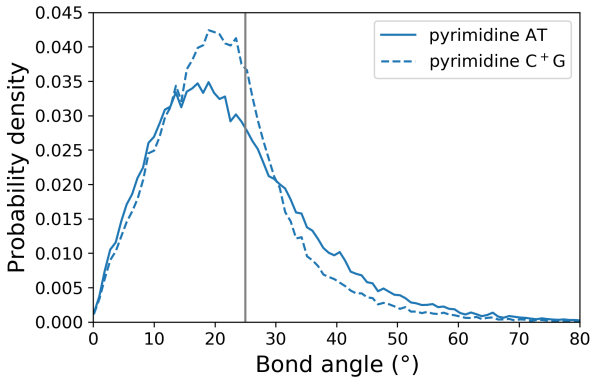
(a) Purine bond length distribution



(b) Purine bond angle distribution

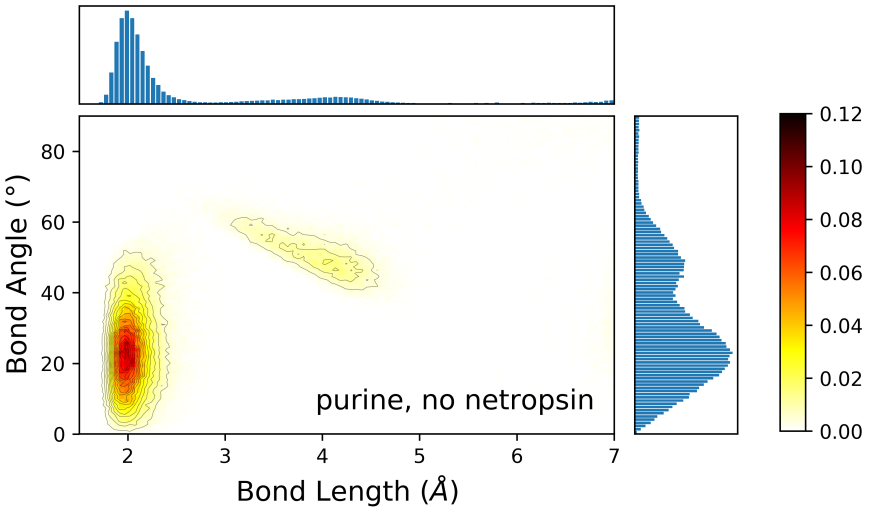


(c) Pyrimidine bond length distribution

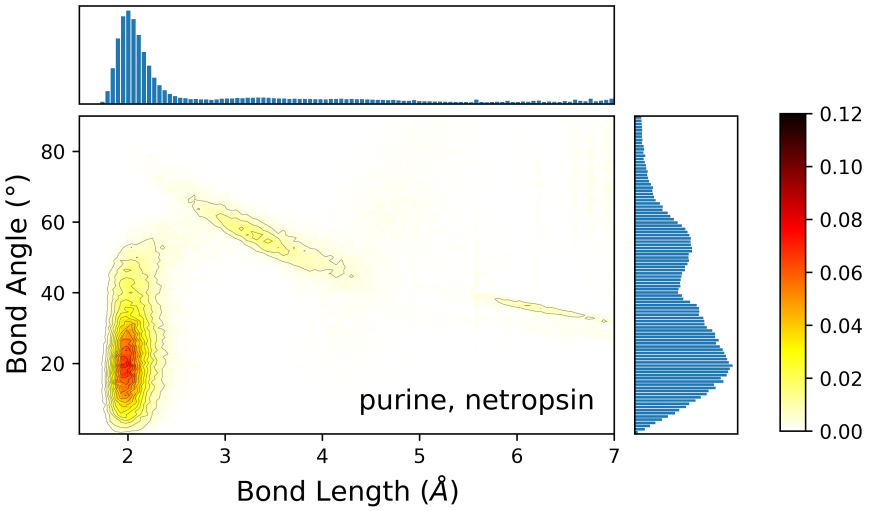


(d) Pyrimidine bond angle distribution

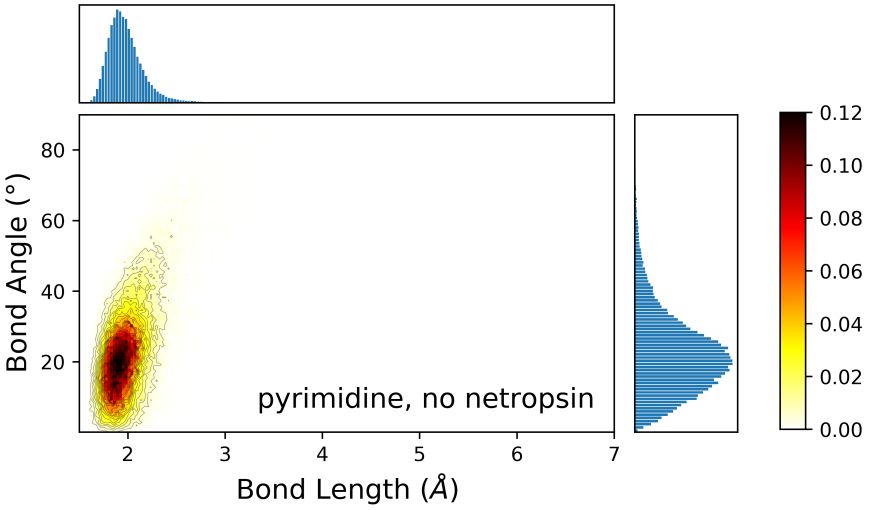
Fig. S11 Normalised bond and angle distributions for each type of hydrogen bond (purine: AA, GG, pyrimidine: C^+G, AT) between the TFO and duplex of the homopurine-TFO triplex with and without netropsin, and the homopyrimidine-TFO triplex in the absence of netropsin, calculated from a 20-ns equilibrium simulation trajectory. The bond angles are measured as deviations from collinearity (180°). Grey vertical lines indicate the hydrogen bond length and angle cutoffs used to calculate the number of hydrogen bonds in the system.



(a) Purine-TFO triplex, no netrospin



(b) Purine-TFO triplex, netrospin



(c) Pyrimidine-TFO triplex, no netrospin

Fig. S12 Joint probability distribution of hydrogen bond lengths and angles for the three triplex structures studied in this work. Individual probability distributions for bond and angle data are also shown for all three structures. Contours are at intervals of 0.01.

S8 Hydrogen bond strength

The change in the total hydrogen-bonding interaction energy between the TFO and duplex on netropsin binding was estimated by assuming a dependence of the hydrogen bond strength on bond length, l, identical to that calculated for an AT Hoogsteen pair using ab initio quantum chemistry.[1] The potential energy curve for the AT Hoogsteen pair was fit with an equation of the form

$$U_{\mathrm{hb}}(l) = A\left[\left(\frac{B}{l-d}\right)^{9} - \left(\frac{C}{l-e}\right)^{7}\right] \tag{S9}$$

for parameters A, B, C, d, and e. The form of this equation was chosen to fit the potential energy curve as closely as possible. The fit of eqn (S9) to the AT potential is shown in Fig. S13. Potential energy curves have not previously been calculated for the AA and GG reverse-Hoogsteen pairs found in the purine-TFO triplex or for the C^+G Hoogsten pairs found in the pyrimidine-TFO triplex studied in this work, but the overall dimer interaction energies for these pairs and that of the AT Hoogsten pair have been calculated by quantum-chemical methods.[2] To approximate the potential energy curves of the the AT and C^+G Hoogsten pairs and AA and GG reverse-Hoogsteen pairs, the fitted potential energy curve for the AT Hoogsteen pair from reference 1 was scaled by changing the value of A so that its minimum matched the dimerisation energies of these base pairs in reference 2. Although different methods were used for the calculations in references 1 and 2, the AT Hoogsteen dimerisation energies are similar, as indicated by the similarity between the original and scaled curves for the AT Hoogsteen pair in Fig. S13. The reported dimerisation energies[2] of these base pairs and the values of A parameter used are shown in Table S1.

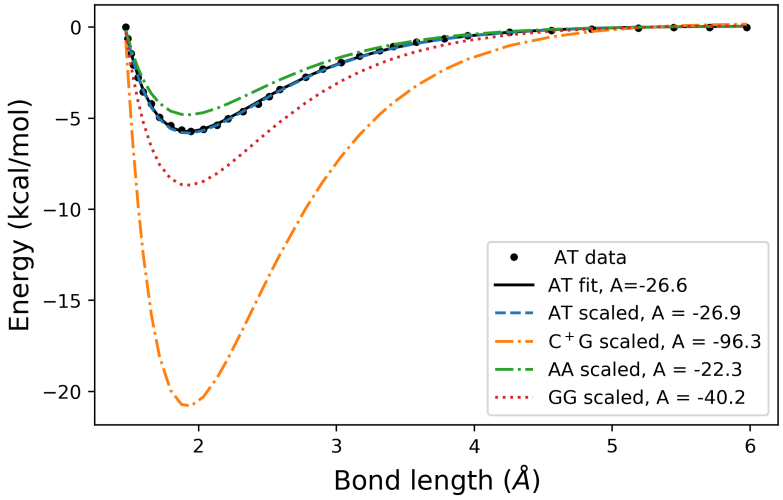


Fig. S13 Fits of the form of eq. (S9) to the hydrogen bond potential for an AT Hoogsteen pair (points). The values of A (in kcal/mol), used to scale the fits for each of the four base pairs, are given in Table S1. B, C, d, and e remained constant for all base pairs: $B = 6.6$ Å, $C = 5.1$ Å, $d = -3.7$ Å, and $e = -2.3$ Å.

Table S1 Dimer interaction energies for relevant base pairs.[2] H=Hoogsteen, rH = reverse-Hoogsteen

Pair	Overall interaction energy (kcal/mol)	Per hydrogen bond (kcal/mol)	A (kcal/mol)
AT (H)	-11.53	-5.77	-26.9
C^+G (H)	-41.25	-20.62	-96.3
AA (rH)	-9.54	-4.77	-22.3
GG (rH)	-17.21	-8.61	-40.2

The average total TFO–duplex hydrogen-bonding energy can then be estimated from

$$\langle U_{\mathrm{total}} \rangle = N_{\mathrm{hb}} \langle U_{\mathrm{hb}} \rangle , \tag{S10}$$

where N_{hb} is the total number of hydrogen bonds that can form between the TFO and duplex and $\langle U_{\mathrm{hb}} \rangle$ is the average strength of a single hydrogen bond, obtained from the calculated potential energy curve $U_{\mathrm{hb}}(l)$ and probability distribution $P(l)$ of hydrogen bond lengths (Fig. S11) using

$$\langle U_{\mathrm{hb}} \rangle = \int P(l) U_{\mathrm{hb}}(l)\, \mathrm{d}l. \tag{S11}$$

The 15-base sequences simulated had a total of 30 potential hydrogen bonds, 20 of which were AA reverse-Hoogsteen or AT Hoogsteen for the purine-TFO or pyrimidine-TFO triplex, respectively, and the remaining 10 GG reverse-Hoogsteen or C^+G Hoogsteen, respectively. The results of these calculations for the three triplexes are found in Table S2.

Table S2 Total hydrogen bond energy for each of the three triplexes studied, calculated as described above based on the distribution of hydrogen bond lengths. H = Hoogsteen, rH = reverse-Hoogsteen.

Triplex type	Pair	Number of bonds, N_{hb}	$\langle U_{\mathrm{hb}} \rangle$ (kcal/mol)	$\langle U_{\mathrm{total}} \rangle$ (kcal/mol)
Purine no netropsin	AA (rH)	20	-4.0	-80
	GG (rH)	10	-5.9	-59
	total	30		**-139**
Purine netropsin	AA (rH)	20	-3.7	-74
	GG (rH)	10	-5.4	-54
	total	30		**-128**
Pyrimidine no netropsin	AT (H)	20	-5.5	-55
	C^+G (H)	10	-19.6	-196
	total	30		**-251**

References

[1] K. Brameld, S. Dasgupta and W. A. Goddard, *J. Phys. Chem. B*, 1997, **101**, 4851–4859.

[2] M. Peters, I. Rozas, I. Alkorta and J. Elguero, *J. Phys. Chem. B*, 2003, **107**, 323–330.