Supporting Information

Disruption of Energetic and Dynamic Base Pairing Cooperativity in DNA Duplexes by an Abasic Site

Brennan Ashwood,[1,‡] Michael S. Jones,[2,‡] Andrew L. Ferguson,[2,*] Andrei Tokmakoff[1,*]

[1]Department of Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, United States

[2]Pritzker School of Molecular Engineering, The University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, United States

[‡]These authors contributed equally

*Corresponding authors: andrewferguson@uchicago.edu, tokmakoff@uchicago.edu

Table of Contents

S1. Duplex dehybridization thermodynamics

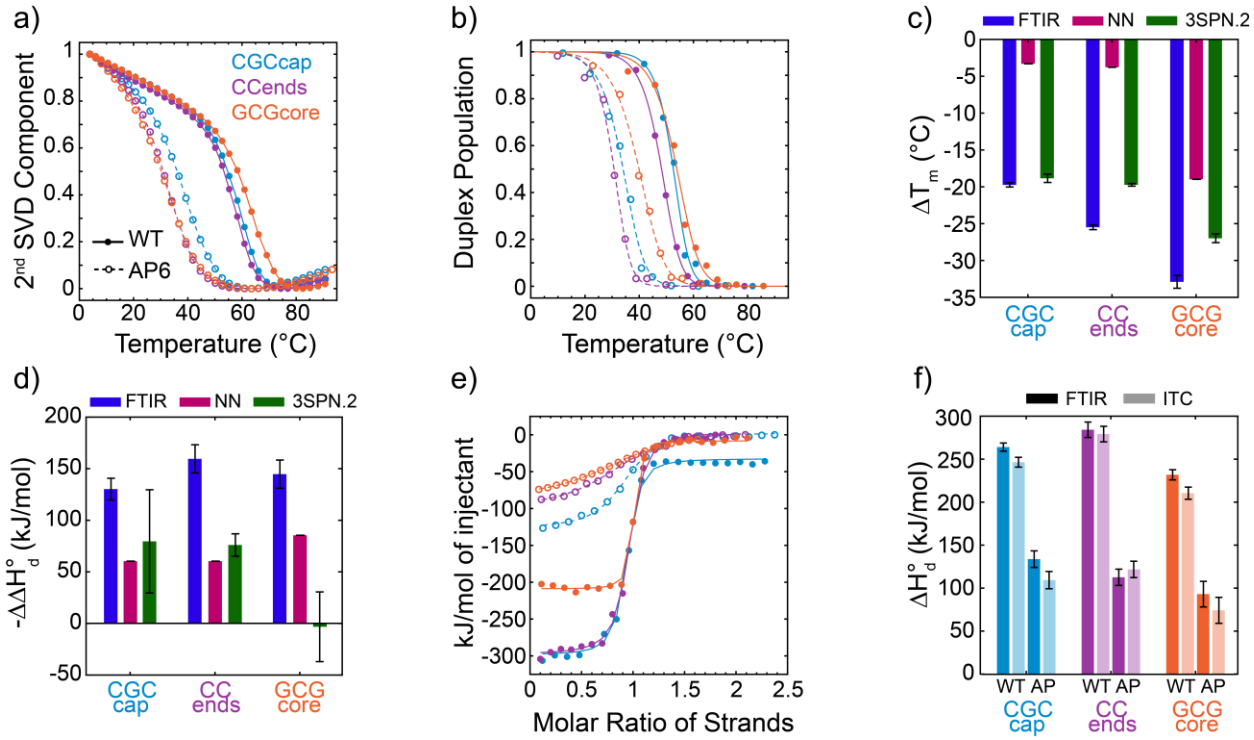

Figure S1 – Influence of AP-site on duplex dehybridization thermodynamics. (a) Normalized 2nd SVD components from FTIR temperature series (Fig. S2) for each WT and AP6 sequence. (**b)** Temperature-dependent fraction of intact duplexes from 3SPN.2 MD simulations employing WTmetaD (Section S2.3). Data are fit to two-state thermodynamic model (solid and dashed lines, Eq. S6). (**c)** Difference in T_m between AP6 and WT sequences (ΔT_m) from FTIR melting curves, NN model, and 3SPN.2 melting curves. (**d)** Difference in dehybridization enthalpy (ΔH_d°) between AP6 and WT sequences ($\Delta\Delta H_d^\circ$) from FTIR melting curves and NN model. Values from 3SPN.2 melting curves correspond to the change in dehybridization internal energy $\Delta\Delta U_d^\circ$. (**e)** isothermal titration calorimetry (ITC) thermograms plotted as molar ratio between complement DNA strands. Data are fit to a two-state model (solid and dashed lines, Eq. S1). ITC sample conditions are listed in Table S1. (**f)** Comparison of ΔH_d° determined from FTIR melting curves (dark) and ITC (light). The measured change in heat capacity between dissociated and hybridized states (ΔC_p) of WT sequences (Fig. S3) is applied to the FTIR-determined ΔH_d° values (Eq. S8) to account for differences between T_m and the ITC temperature. Error bars for all FTIR, ITC, and 3SPN.2 parameters indicate 95% confidence intervals from two-state model fits. Error bars for NN values correspond to the previously reported error in the NN model.(1)

S1.1 Temperature-dependent FTIR spectra

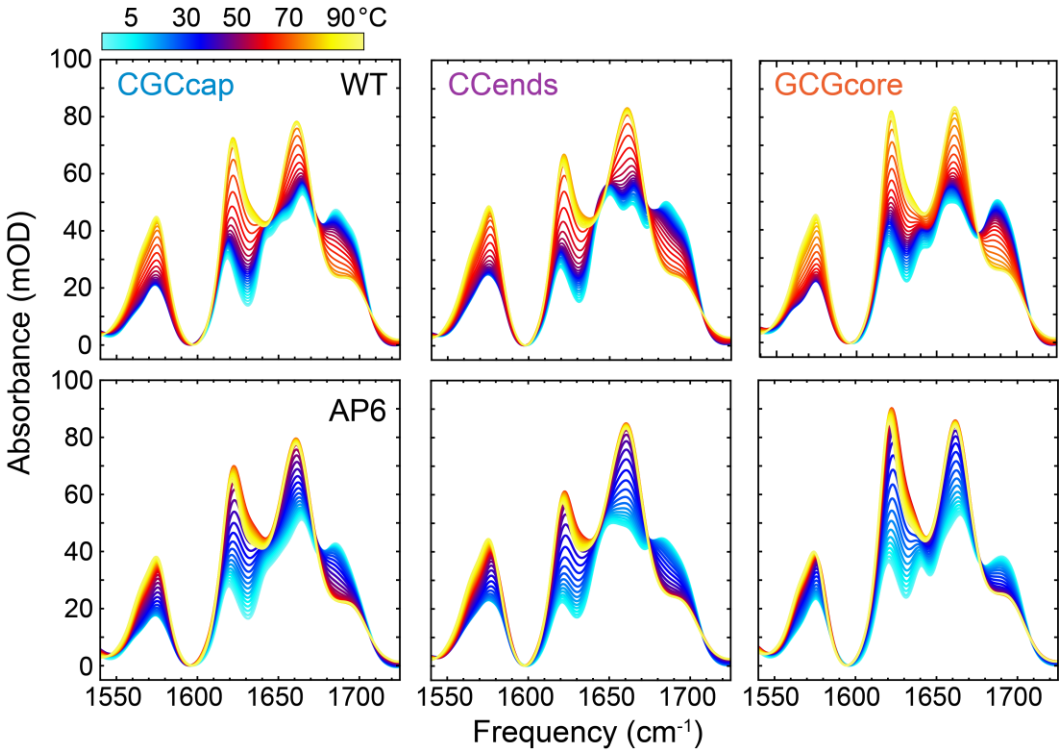

Figure S2 – Temperature-dependent FTIR spectra of WT and AP6 sequences. FTIR temperature series for CGCcap, CCends, and GCGcore WT (top row) and AP6 (bottom row) sequences measured from 3.8 to 93.3 °C. All samples were prepared in deuterated pH* 6.8 400 mM SPB with a total oligonucleotide concentration of 2 mM.

S1.2 ITC measurements conditions and determination of ΔC_p

ITC measurements started with an initial 0.4 µL injection followed by 19 2 µL aliquot injections of the titrant oligonucleotide solution into the cell. Injections of 4 s duration were performed every 180 s, and the syringe needle constantly stirred the cell solution with a spin rate of 1000 rpm. The power required to re-equilibrate the sample and reference cells was monitored as a function of time. The power profile of each injection was integrated over time to obtain the change in heat from the injection (ΔQ). The heat of dilution was measured for the titrant at each experimental temperature and subtracted from ΔQ after integration. Curves of ΔQ vs molar ratio of titrant and cell oligonucleotide were fit to a single site binding model(2) with free parameters for binding enthalpy (ΔH_d°), binding constant (K_a), and reaction stoichiometry (n).

$$\Delta Q_i = Q_i - Q_{i-1} + \frac{V_{inj}}{V_o}\left[\frac{Q_i + Q_{i-1}}{2}\right] \quad \text{(S1)}$$

$$Q_i = \frac{nM_{c,i}V_o\Delta H^\circ}{2}\left[1 + \frac{T_{c,i}}{nM_{c,i}} - \frac{1}{nK_aM_{c,i}} - \sqrt{\left(1 + \frac{T_{c,i}}{nM_{c,i}} + \frac{1}{nK_aM_{c,i}}\right)^2 - \frac{4T_{c,i}}{nM_{c,i}}}\right] \quad \text{(S2)}$$

Q_i is the total heat exchanged after injection i of titrant. $M_{c,i}$ and $T_{c,i}$ are the concentration of cell oligonucleotide and titrant oligonucleotide, respectively, in the cell after injection i. V_o is the initial volume of sample in the cell, which was 200 µL for all measurements, and V_{inj} is the volume of titrant used in each injection (2 µL).

Table S1. Experimental conditions used for ITC measurements shown in Fig. S1. All samples were prepared in 400 mM pH* 6.8 sodium phosphate buffer (SPB).

Sequence	T (°C)	Syringe Strand	[Syringe] (µM)	Cell Strand	[Cell] (µM)
CGCcap					
WT	25	ATATATATGCG	120	CGCATATATAT	10
AP6	10	ATATATATGCG	245	CGCAT_TATAT	20
CCends					
WT	25	GGATATATAGG	100	CCTATATATCC	10
AP6	10	GGATATATAGG	220	CCTAT_TATCC	20
GCGcore					
WT	25	ATATCGCTATA	110	TATAGCGATAT	10
AP6	10	ATATCGCTATA	2000	TATAG_GATAT	200

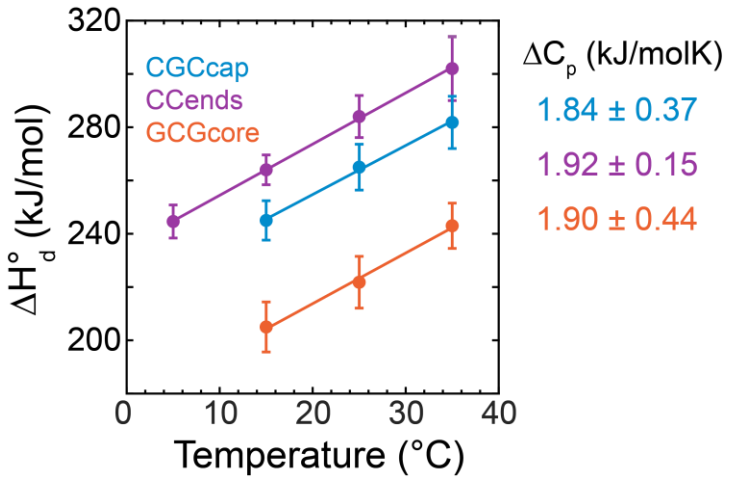

Figure S3 – Temperature-dependent isothermal titration calorimetry of WT sequences. Enthalpy of dissociation (ΔH_d°) determined with ITC as a function of temperature for CGCcap,

CCends, and GCGcore WT sequences. Vertical error bars indicate 95% confidence intervals from nonlinear least squares fitting of ITC thermograms. Temperature-dependent ΔH_d° data are fit to linear functions (solid lines) and the slope (ΔC_p) is listed for each sequence.

S1.3 Two-state model of duplex-to-single-strand transition

The FTIR temperature series of each DNA sequence studied here exhibit a single sigmoidal melting transition and therefore can be most simply described using a two-state model of the duplex (*D*)-to-single-strand (*S*) transition for non-self-complementary oligonucleotides. This is a standard treatment of optical melting curves,(3, 4) but we provide details here to clarify our fitting and data analysis.

$$D \rightleftharpoons S_1 + S_2 \tag{S3}$$

In the two-state description, the D state represents an average over the ensemble of all base pairing configurations (frayed, shifted, etc.). The equilibrium constant for dissociation (K_d) is related to the fraction of intact duplexes (θ_D),

$$K_d = \frac{[S_1][S_2]}{[D]} = \frac{(1-\theta_D)^2 c_{tot}}{2\theta_D} \tag{S4}$$

$$\theta_D = \frac{2[D]}{c_{tot}} \tag{S5}$$

where c_{tot} is the total oligonucleotide concentration. Eq. S4 assumes a 1:1 molar ratio between complementary DNA strands ($[S_1] = [S_2]$), and all FTIR measurements in this work were performed under this condition. Eq. S4 can be arranged to express θ_D in terms of K_d.

$$\theta_D = 1 + \frac{K_d - \sqrt{K_d^2 + 2c_{tot}K_d}}{c_{tot}} \tag{S6}$$

The temperature dependence of θ_D and K_d are determined by the enthalpy (ΔH_d°) and entropy (ΔS_d°) difference between the D and S states.

$$K_d(T) = \exp\left[-\frac{\Delta H_d^\circ(T)}{RT} + \frac{\Delta S_d^\circ(T)}{R}\right] \tag{S7}$$

In reality, both ΔH_d° and ΔS_d° are temperature-dependent due to a change in heat capacity (ΔC_p) between the D and S states, which is thought to arise from temperature-dependent stacking interactions in each of the single-strands and a change in solvent accessible surface area upon dehybridization.(5-8) For dissociation of short oligonucleotides, ΔC_p follows a length-scaling of 150-250 J/molK per base-pair(9, 10) and depends on the sequence and counterion concentration and can also be temperature-dependent.(5, 6) Assuming ΔC_p is constant over the temperature range of interest, we can express ΔH_d° and ΔS_d° as a function of temperature.

$$\Delta H_d^\circ(T) = \Delta H_d^\circ(T_m) + \Delta C_p(T - T_m) \tag{S8}$$

$$\Delta S_d^\circ(T) = \Delta S_d^\circ(T_m) + \Delta C_p \ln\left(\frac{T}{T_m}\right) \tag{S9}$$

ΔH_d° and ΔS_d° are expressed relative to their values at the melting temperature (T_m), which are the van't Hoff enthalpy and entropy. T_m, which here is defined as the temperature where $\theta_D = 0.5$, can be re-cast in terms of ΔH_d° and ΔS_d°.

$$T_m = \frac{\Delta H_d^\circ(T_m)}{\Delta S_d^\circ(T_m) - R\ln(c_{tot}/4)} \tag{S10}$$

ΔH_d° and T_m also determine the slope of the melting transition at T_m:

$$\left(\frac{\partial \theta_D}{\partial T}\right)_{T_m} = \frac{\Delta H_d^\circ(T_m)}{6RT_m^2} \tag{S11}$$

To fit the FTIR melting data, ΔH_d° and T_m were used as free parameters and ΔC_p was set to 0 initially. The 2nd SVD components from FTIR data were then fit to θ_D with baselines described by additional slope (m_D, m_S) and intercept (b_D, b_S) terms.

$$V_{Fit}^{(2)} = \theta_D(L_D - L_S) + L_S \tag{S12}$$

$$L_D = m_D T + b_D \tag{S13a}$$

$$L_S = m_S T + b_S \tag{S13b}$$

Fig. S4a shows fits of the FTIR 2nd SVD components to Eq. S12 with $\Delta C_p = 0$. We can also fit the melting data using ΔC_p determined from ITC (Fig. S3), which is shown in Fig. S4b. Compared to profiles of θ_D determined from fits with $\Delta C_p = 0$ (Fig. S5a), those including ΔC_p exhibit reduced hybridization at lower temperature due to higher K_d values. Overall, K_d values determined from ITC agree better with the melting curve fits that include ΔC_p, which is expected since the ΔC_p values were determined ITC. The largest discrepencies are observed for the lowest K_d measurements (ITC $K_d < 10^{-7}$), but these ITC determined values are expected to have the highest error due to having very sharp thermogram transitions. $\Delta\Delta G^{\circ}_{d,37}$ values are nearly the same with or without inclusion of ΔC_p in the thermodynamic fitting (Fig. S5) but this comparison does not account for minor differences in ΔC_p between WT and AP6 sequences.

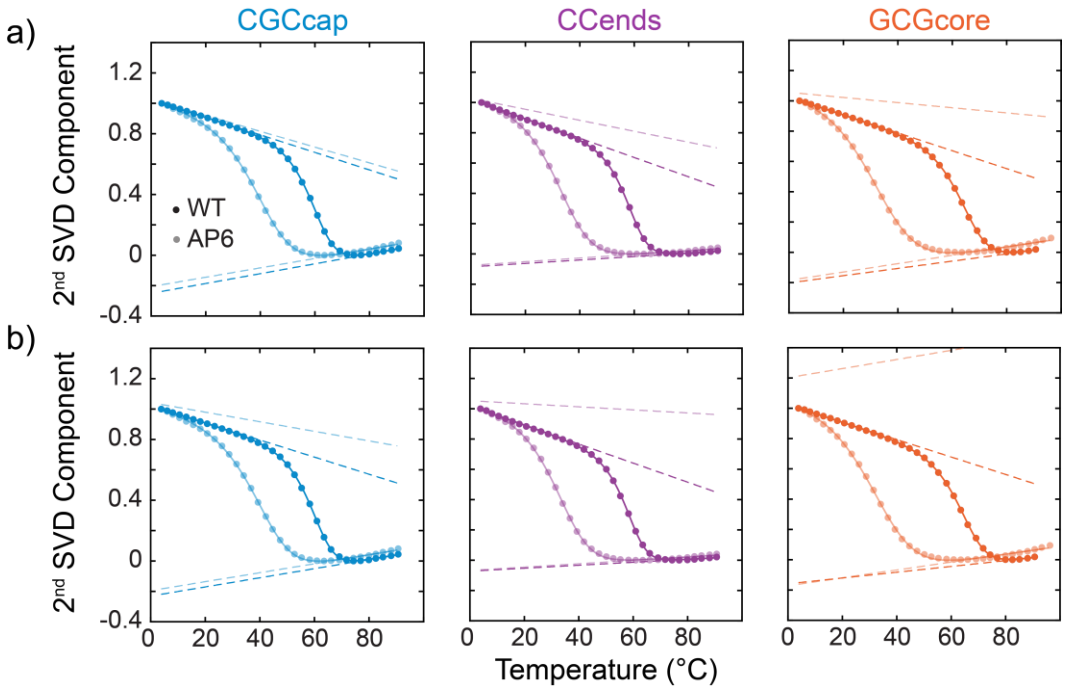

Figure S4 – Two-state fitting to determine duplex melting curves. 2nd SVD components from FTIR temperature series fit (solid lines) to 2-state thermodynamic model (Eq. S12) **(a)** with $\Delta C_p = 0$ and **(b)** with ΔC_p fixed at the value determined from ITC (Fig. S2). Dashed lines indicate high (L_D) and low (L_S) baselines from the fits.

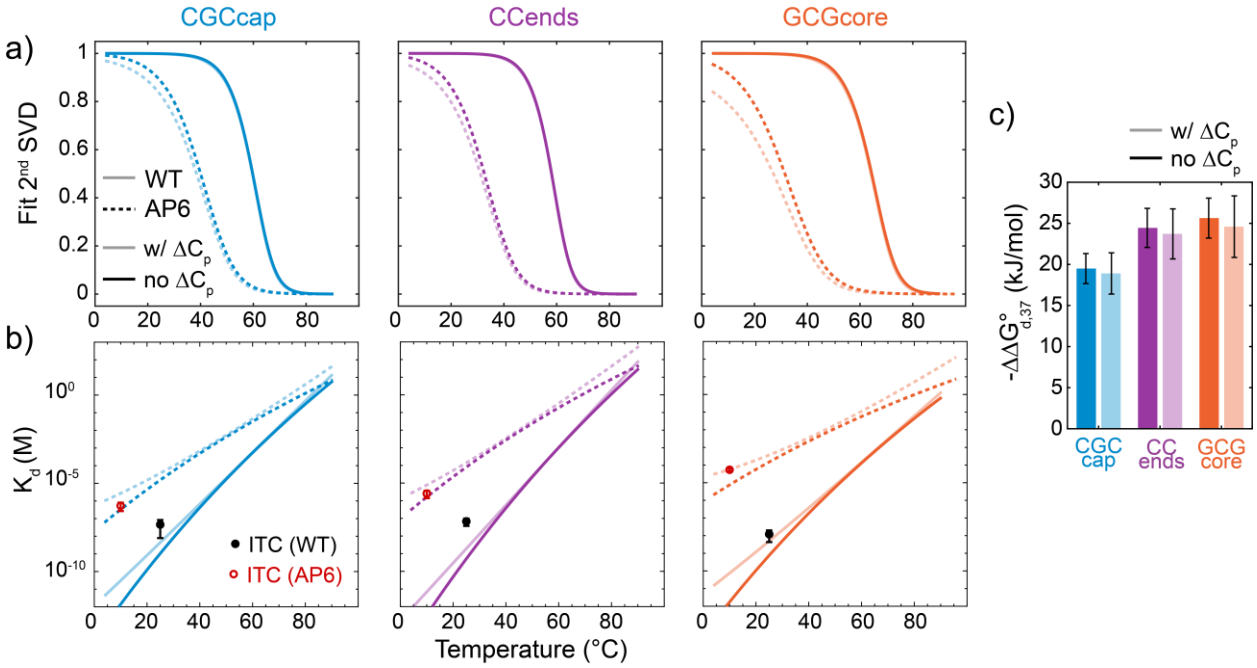

Figure S5 – Impact of temperature-dependent ΔH°_{d} and ΔS°_{d} on melting curves. θ_D determined from fitting FTIR 2nd SVD components to two-state model (Eq. S12) **(a)** with $\Delta C_p = 0$ and **(b)** ΔC_p fixed at the value determined from ITC (Fig. S3). **(c)** Temperature-dependent equilibrium constant for duplex dissociation (K_d) determined from two-state fits to FTIR 2nd SVD components with $\Delta C_p = 0$ (solid lines) and with ΔC_p set to value from ITC (transparent lines). K_d values determined at select temperatures with ITC are shown as circles. Error bars indicate 95% confidence interval from fitting ITC thermograms to a single-site binding model.

Table S2. Thermodynamics parameters for duplex melting curves determined from two-state fits to FTIR temperature series 2nd SVD components (Fig. S5a) and 3SPN.2-determined duplex fractions (Fig. S1b). The internal dehybridization energy (ΔU°_{d}) and Helmholtz dehybridization free energy ($\Delta F^{\circ}_{d,37}$) determined from 3SPN.2 MD are compared with ΔH°_{d} and $\Delta G^{\circ}_{d,37}$ from experiment, respectively.

Sequence		ΔH°_{d} (kJ/mol)	ΔU°_{d} (kJ/mol)	ΔS°_{d} (J/molK)		$\Delta G^{\circ}_{d,37}$ (kJ/mol)	$\Delta F^{\circ}_{d,37}$ (kJ/mol)	T_m (°C)	
		FTIR	3SPN.2	FTIR	3SPN.2	FTIR	3SPN.2	FTIR	3SPN.2
CGCcap	WT	310 ± 6	322 ± 17	869 ± 16	921 ± 47	40.9 ± 1.4	35.8 ± 2.5	59.8 ± 0.1	53.5 ± 0.3
	AP6	180 ± 10	329 ± 31	512 ± 30	1010 ± 98	21.4 ± 2.4	16.1 ± 5.0	40.0 ± 0.6	33.7 ± 0.4
CCends	WT	342 ± 10	345 ± 8	970 ± 32	1010 ± 24	41.5 ± 2.8	31.7 ± 1.7	58.1 ± 0.2	48.3 ± 0.1
	AP6	183 ± 9	269 ± 8	535 ± 31	829 ± 25	17.0 ± 2.4	12.0 ± 1.5	32.6 ± 0.6	28.5 ± 0.2
GCGcore	WT	284 ± 6	303 ± 32	779 ± 20	864 ± 91	42.8 ± 1.7	35.3 ± 5.0	64.4 ± 0.2	53.9 ± 0.6
	AP6	140 ± 13	307 ± 11	396 ± 42	958 ± 36	17.1 ± 3.4	9.3 ± 2.8	31.5 ± 1.7	26.9 ± 0.2

Table S3. Thermodynamics difference parameters between AP6 and WT duplexes determined from two-state fits to FTIR temperature series 2nd SVD components (Fig. S5a) and 3SPN.2-determined duplex fractions (Fig. S1b).

Sequence	$\Delta\Delta H^\circ_d$ (kJ/mol)	$\Delta\Delta U^\circ_d$ (kJ/mol)	$\Delta\Delta S^\circ_d$ (J/molK)		$\Delta\Delta G^\circ_{d,37}$ (kJ/mol)	$\Delta\Delta F^\circ_{d,37}$ (kJ/mol)	ΔT_m (°C)	
	FTIR	3SPN.2	FTIR	3SPN.2	FTIR	3SPN.2	FTIR	3SPN.2
CGCcap	-130 ± 11	8 ± 43	-357 ± 34	89 ± 114	-19.5 ± 2.8	-23.1 ± 7.0	-19.7 ± 0.6	-19.7 ± 0.5
CCends	-159 ± 14	-76 ± 13	-435 ± 44	-181 ± 35	-24.5 ± 3.6	-19.7 ± 1.7	-25.5 ± 0.6	-19.7 ± 0.2
GCGcore	-145 ± 14	3 ± 40	-384 ± 46	94 ± 104	-25.7 ± 3.6	-26.0 ± 7.0	-32.9 ± 1.8	-27.0 ± 0.7

S1.4 Nearest-neighbor (NN) model predictions

We compare the two-state thermodynamic parameters obtained from FTIR temperature series with estimates from Santa Lucia's unified NN parameters(1) corrected to a sodium concentration of 600 mM to match experimental conditions.(11) The NN model contains both propagation, which is the energy of extended a duplex by a given dinucleotide step, and initiation parameters. The initiation parameters contain both sequence-independent and sequence-dependent terminal effects that depend on whether the terminal base pair is G:C or A:T. We treat sequences containing an AP-site as two separate duplexes of length N_1 and N_2 where i is the dinucleotide index across the duplex. The propagation enthalpy ($\Delta H^\circ_{p,NN}$) and entropy ($\Delta S^\circ_{p,NN}$) are the sum of dinucleotide contributions across each duplex segment.

$$\Delta H^o_{p,NN} = \sum_{i=1}^{N_1-1} \Delta H^o_{NN,i} + \sum_{i=N_1}^{N_1+N_2-1} \Delta H^o_{NN,i} \quad \text{(S14a)}$$

$$\Delta S^o_{p,NN} = \sum_{i=1}^{N_1-1} \Delta S^o_{NN,i} + \sum_{i=N_1}^{N_1+N_2-1} \Delta S^o_{NN,i} \quad \text{(S14b)}$$

Initiation parameters are included just for base pairs 1 and 11.

$$\Delta H^o_{NN} = \Delta H^o_{p,NN} + \Delta H^o_{init,1} + \Delta H^o_{init,11} \quad \text{(S15a)}$$

$$\Delta S^o_{NN} = \Delta S^o_{p,NN} + \Delta S^o_{init,1} + \Delta H^o_{init,11} \quad \text{(S15b)}$$

Table S4. Stability of AP6 duplex halves from NN calculations. Values are corrected to a sodium ion concentration of 600 mM.(11)

Sequence	Stretch	ΔH_d° (kJ/mol)	ΔS_d° (J/molK)	$\Delta G_{d,37}^\circ$ (kJ/mol)	$\Delta\Delta G_{d,37}^\circ$ (5′-3′) (kJ/mol)
CGCcap-AP6	5’ (5’-CGCAT-3’)	150.6	409.4	23.7	16.5
	3’ (5’-TATAT-3’)	110.9	334.3	7.2	
CCends-AP6	5’ (5’-CCTAT-3’)	125.9	359.1	14.6	-0.1
	3’ (5’-TATCC-3’)	127.6	364.1	14.7	
GCGcore-AP6	5’ (5’-TATAG-3’)	113.4	336.4	9.1	-1.2
	3’ (5’-GATAT-3’)	115.0	337.7	10.3	

S2. Free energy surfaces and thermodynamics from 3SPN.2 simulations

S2.1 Impact of AP-site on DNA duplex structure

The 3SPN.2 model has not previously been used to study oligonucleotides containing an AP-site, and it is important to assess whether the model can incorporate an AP-site without structural artifacts. Both all-atom MD simulations and NMR experiments indicate that an AP-site does not disturb the global B-form conformation of the DNA duplex.(12-14) To assess the duplex structure of the sequences studied in this work, we ran 3SPN.2 MD simulations at 20 °C and compared observed structural parameters between sequences WT and AP6 sequences (Fig. S6). The average RMSD relative to B-form DNA in sequences containing an AP-site is similar to their respective WT sequence, regardless of whether the AP-site-containing strand or unmodified strand is considered. The average base pair and intermolecular phosphate separation distances are also similar between WT and AP sequences. The most significant differences arise from sampling half-dehybridized configurations in CGCcap-AP6. Overall, duplexes containing an AP-site still adopt a global B-form structure in 3SPN.2 MD simulations.

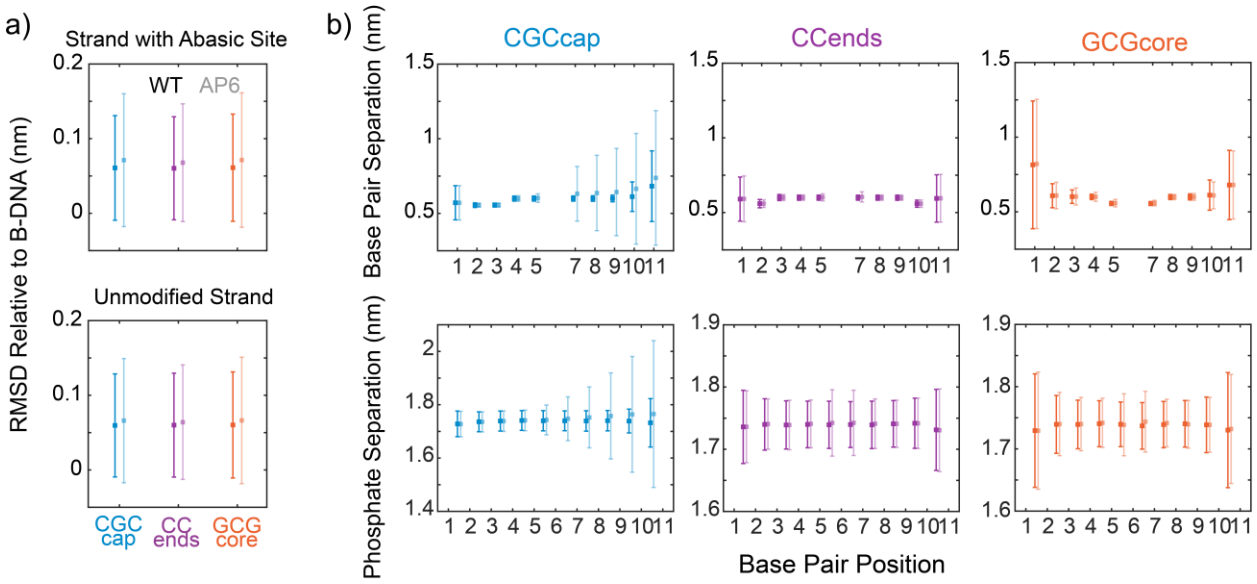

Figure S6 – Impact of AP-site on B-DNA structure observed with 3SPN.2. (a) RMSD relative to a single strand reference B-DNA structure for the strand containing the AP-site and the unmodified complement strand. Points correspond to average RMSD over 25 × 1.2 μs trajectories run at 20 °C and error bars indicate standard deviations. **(b)** Average separation for base pairs shared among all sequences. **(c)** Average intermolecular separation between backbone phosphates for all nucleotide sites.

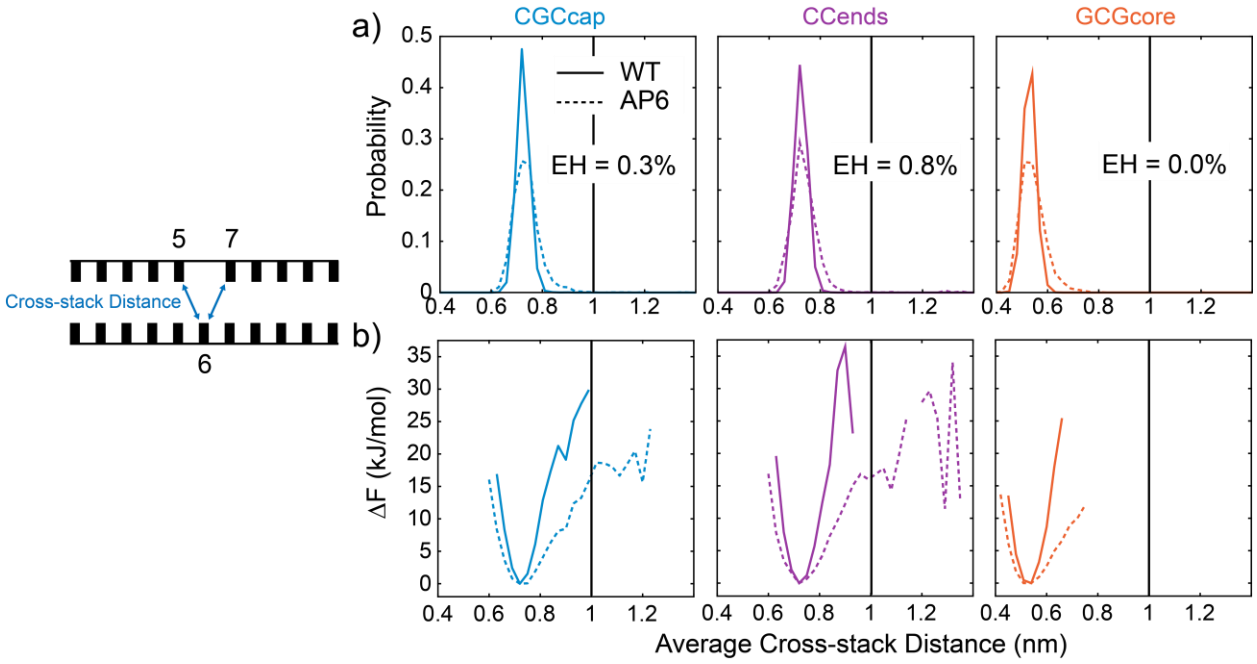

Figure S7 – Conformation of free nucleotide in DNA duplex from 3SPN.2. (a) Probability distributions for the average of each cross-stacking distance between the nucleobase opposite the AP-site and nucleobases adjacent to the AP-site (schematic shown on the left). The same distance is plotted for WT sequences (solid lines). Populations include all configurations from WTMetaD

simulation with $n_{bp} \geq 9$. **(b)** Corresponding FEPs plotted in ΔF relative to the intrahelical free energy minimum. Configurations with an average cross-stacking distance greater than 1 nm are found to be extrahelical and make of 0.3 and 0.8% of the population (EH) for CGCcap-AP6 and CCends-AP6. Negligible extrahelical population is observed for GCGcore-AP6.

S2.2 Enhanced sampling with well-tempered metadynamics

Although equilibrium simulations of coarse-grained DNA systems can provide reasonable estimates for duplex free energy surfaces and stationary distributions, the computational load required for adequate sampling becomes intractable for the combinatorics of all sequences, AP-sites, and temperatures investigated in this study. To efficiently sample the full duplex-to-single-strand transition, we employed well-tempered metadynamics (WTMetaD) to enhance sampling of hybridized, dissociated, and intermediate states.(15) Metadynamics is often described as performing on-the-fly deposition of "computational sand" onto a system's underlying free-energy landscape. As the simulation progresses, free-energy minima are "filled" with an artificial biasing potential and the system is forced to explore new regions of phase space.(16, 17) Eq. S16 shows how a bias potential V is constructed as function of some collective variable *s* over time *t*. The Gaussian height *W*, and width σ, and deposition time τ are hyperparameters that determine how aggressively the system is biased.

$$V(s,t) = \sum_{t'=\tau..}^{t'<t} W \exp\left[\frac{-\left(s - s(t')\right)^2}{2\sigma^2}\right] \tag{S16}$$

In WTMetaD, the bias height *W* is tempered as function of the potential already deposited at a given collective variable (CV) as shown in Eq. S17. The rate of tempering is determined by the bias factor γ, which acts as an additional hyperparameter during optimization.

$$W = W_o \exp\left[-\frac{V(s,t)}{k_B T}\right] \quad \text{where, } \gamma = \frac{T + \Delta T}{T} \tag{S17}$$

Once the WTMetaD simulation has converged, the biasing potential is reweighted according to Eq. S18 to reveal the unbiased free energy landscape. The first 10% of simulation data is excluded during reweighting to ensure that the bias has reached a quasi-static regime. Note that the reweighting coordinate s need not be the same CV used for biasing.

$$P_{target}(s) = P_{bias}(s) \exp\left[-\frac{V_{bias}(s)}{k_B T}\right] \tag{S18}$$

We then report the Helmholtz free energy profile (FEP) as the difference relative to the probability maximum of the fully dissociated state (S_1+S_2).

$$\Delta F(s) = -RT \ln\left[\frac{P_{target}(s)}{P_{target}\left(s_{S_1+S_2}\right)}\right] \quad \text{(S19)}$$

For each system, we selected a nearly identical CV that described the average distance between native Watson-Crick-Franklin base pairs (r_{bp}). The CV was calculated by averaging over all available native pair distances – 11 for WT or 10 for AP sequences. Due to the 0.05 nm difference in A:T vs G:C base pair equilibrium distance, this CV changes slightly across the studied sequences but remains nearly identical for comparative purposes. We evaluated convergence and optimized metadynamics hyperparameters by comparing biased FEPs against a set of several long (500 μs) unbiased free-energy surfaces projected on the same CV. With a Gaussian height of 0.6 kJ/mol, a Gaussian width of 0.01 nm, and bias factor of 4, we were able to achieve convergence – defined as agreement between the free energy landscapes in r_{bp} between the long unbiased and biased simulations to within 2.0 kJ/mol – in less than 10 μs or ~12 CPU-hours simulation time for a single sequence/AP combination on an Intel E5-2680v4 CPU core. In order to obtain better statistics, we ran each system for twice as long (20 μs or ~24 CPU-hours) and averaged over the last 5 μs to account for fluctuations in the bias.(16) This also enabled error estimation of the underlying FEP. In addition to r_{bp}, we tried supplementing our analysis with additional CVs such as individual base pair distances, center-of-mass between strands, and angles between strands as have been used in previous work(18) via 2D WTMetaD and parallel-bias metadynamics.(19) When evaluating convergence against the long equilibrium FEPs, we found no significant improvement with the addition of multiple CVs and often observed slower convergence than the 1D case. We performed biased simulations for all sequences at 9 temperatures ranging from $T_{m,MD}$-20°C to $T_{m,MD}$+20°C. The PLUMED v2.7 libraries(20) compiled with LAMMPS were used to calculate collective variable, apply WTMetaD bias, and reweight our observables to eliminate the effect of bias and return estimates of the unbiased free-energy surfaces.

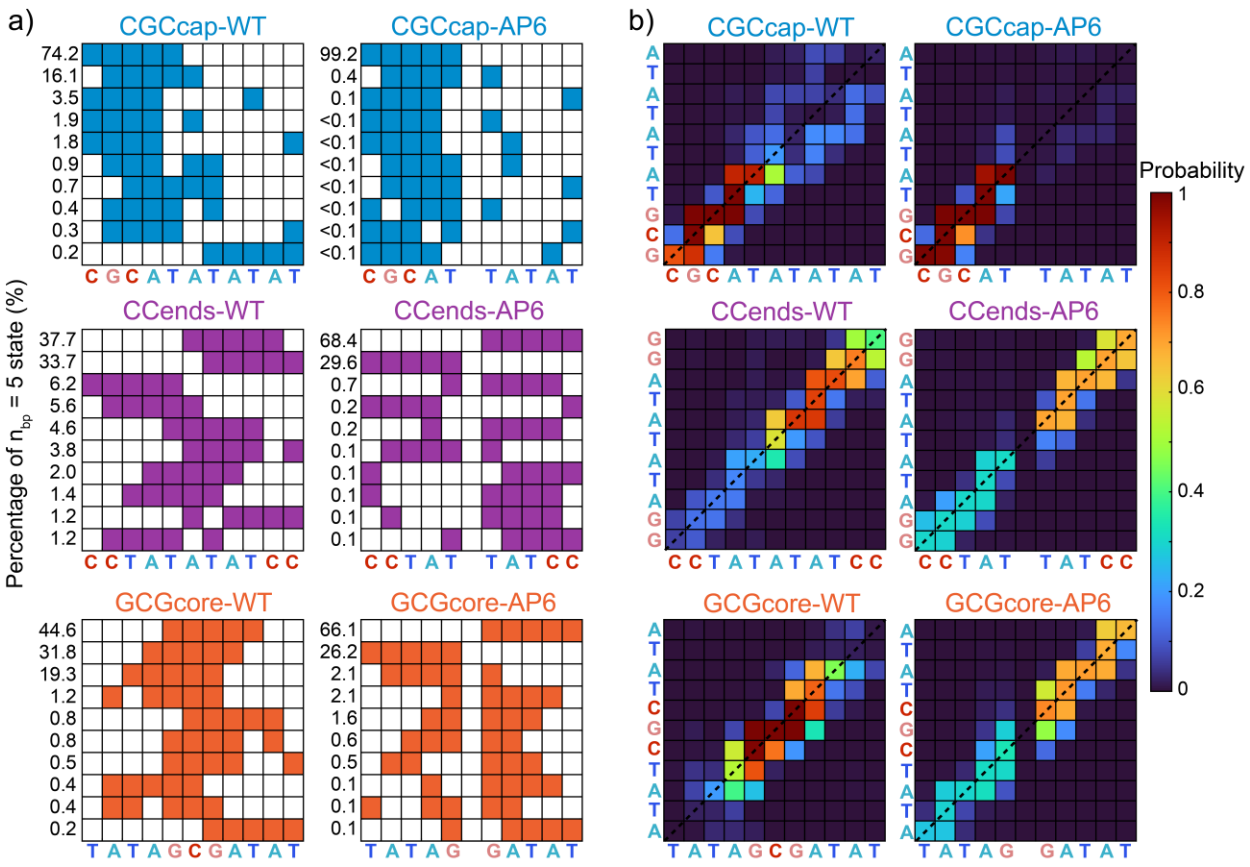

Figure S8 – Distribution of duplex configurations containing five intact base-pairs from 3SPN.2 MD simulations. (a) Ten most probable base pairing configurations contributing to the n_{bp} = 5 state from the FEPs in Fig. 2a for each WT and AP6 sequence. The probability of a given configuration (%) is listed on the left. Filled squares correspond to an intact base-pair and white squares indicate a broken base-pair. A base-pair separation cutoff of 0.7 nm was used. (**b)** Contact maps indicating the probability for each nucleobase intermolecular distance to be below 0.7 nm in the n_{bp} = 5 state. Squares along the diagonal (dashed line) correspond to probabilities for in-register distances. Contact maps indicate that the n_{bp} = 5 state is dominated by configurations with in-register base-pairs.

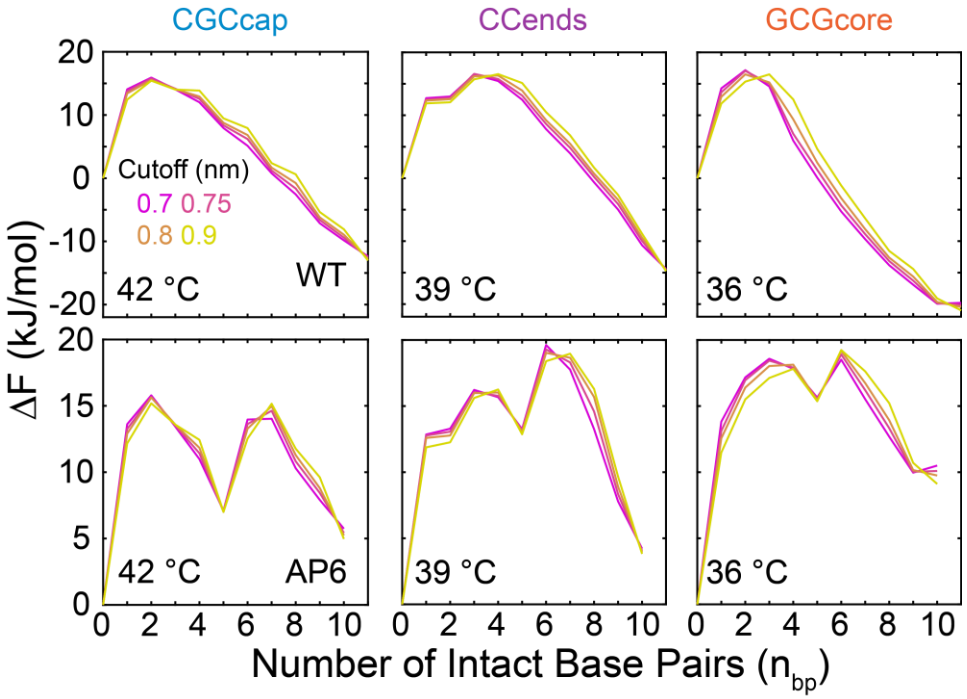

Figure S9 – Sensitivity of 3SPN.2 FEPs to choice of base-pair separation cutoff. FEPs are shown for WT (top row) and AP6 (bottom row) sequences extracted from WTMetaD simulations. Data were re-weighted as a function of the number of intact base-pairs (n_{bp}). Base pairs are determined using a radial cutoff between a nucleobase and its complement that ranges from 0.7 to 0.9 nm.

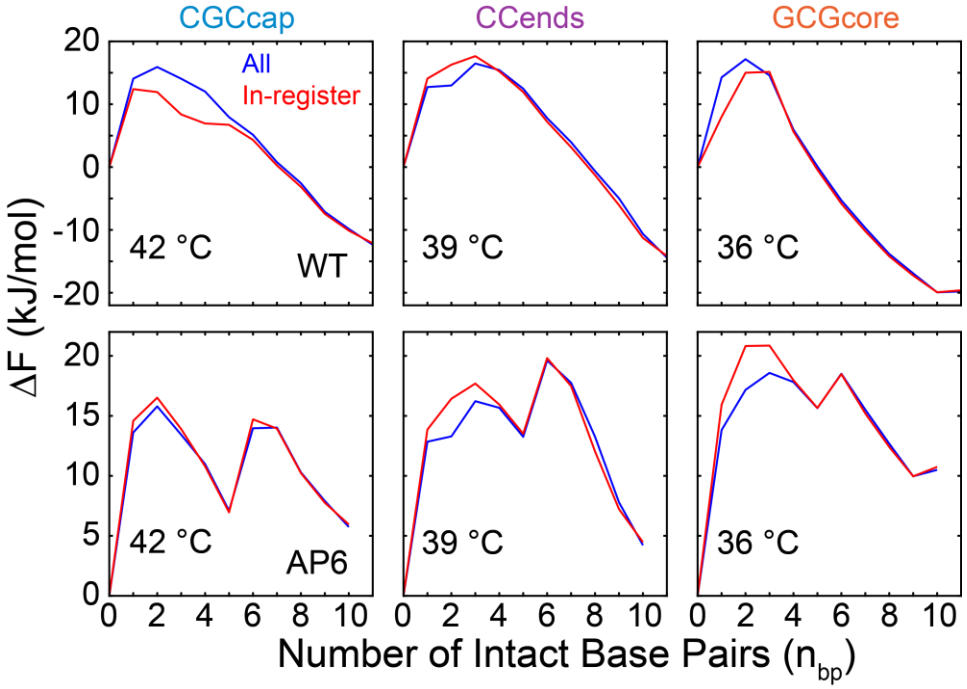

Figure S10 – Comparison of FEPs as a function of all base pairs or only in-register base pairs. FEPs are shown for WT (top row) and AP6 (bottom row) sequences extracted from

WTMetaD simulations. FEPs are shown as a function of all intact base-pairs (n_{bp}, blue) and including just configurations with in-register base pairs (red).

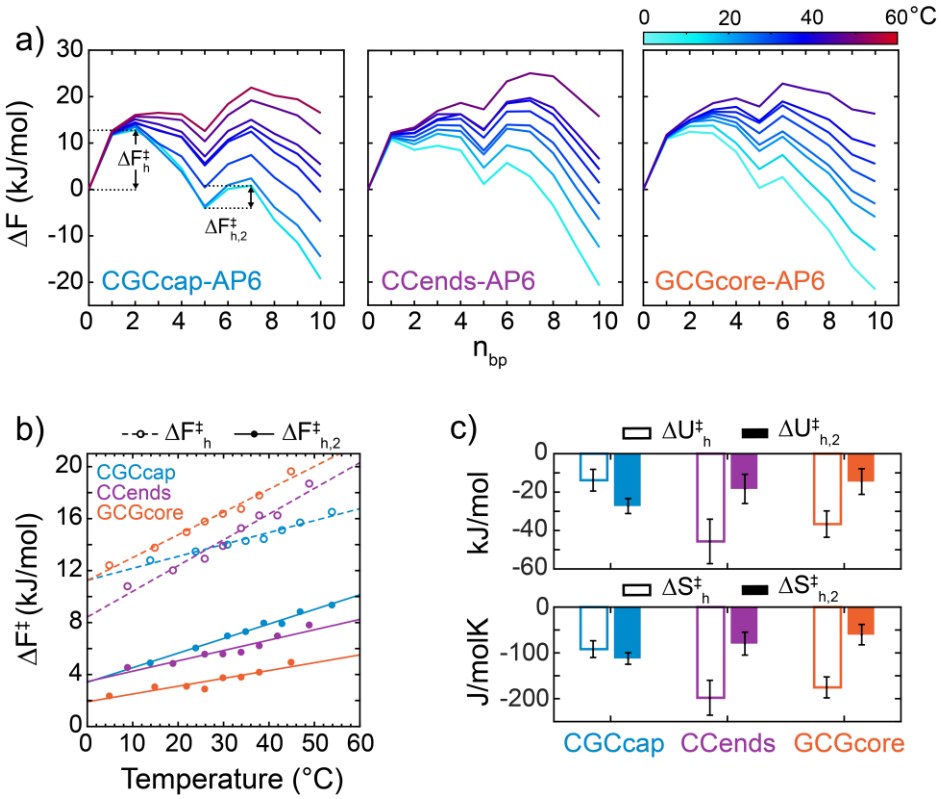

Figure S11 – Temperature-dependence of nucleation barriers determine from 3SPN.2. (a) FEPs of AP6 sequences determined from 3SPN.2 MD simulations with WTMetaD at temperatures from $T_{m,MD}$-20°C to $T_{m,MD}$+20°C. **(b)** Temperature-dependent values of $\Delta F^{\ddagger}_{h}$ (open circles) and $\Delta F^{\ddagger}_{h,2}$ (filled circles) determined from **(a)**. Dashed and solid lines indicate linear fits to $\Delta F^{\ddagger} = \Delta U^{\ddagger} - T\Delta S^{\ddagger}$. **(c)** Internal energy ($\Delta U^{\ddagger}_{h}$ and $\Delta U^{\ddagger}_{h,2}$) and entropic ($\Delta S^{\ddagger}_{h}$ and $\Delta S^{\ddagger}_{h,2}$) contributions to $\Delta F^{\ddagger}_{h}$ and $\Delta F^{\ddagger}_{h,2}$ from fits in in **(b)**. Error bars correspond to 95% confidence intervals from fits.

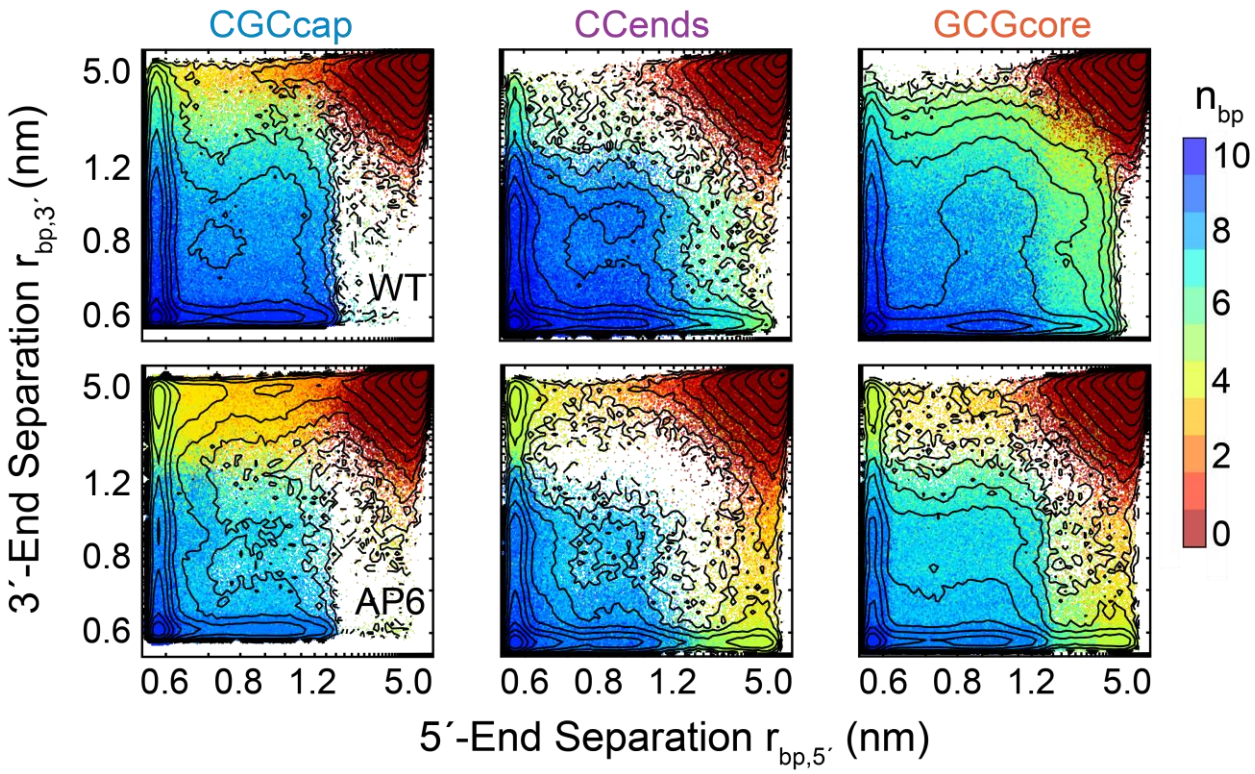

Figure S12 – Relationship between n_{bp} and r_{bp} coordinates. Unbiased simulation data from Fig. 2b plotted as a function of $r_{bp,5'}$ and $r_{bp,3'}$ in units of the number of intact base-pairs (n_{bp}) determined using a 0.7 nm radial cutoff between the center-of-mass of a nucleobase site and that of its complement. Contours correspond to the free energy (-ΔF) and are plotted with uniform 0.66 $\Delta F/k_BT$ spacing. Local free energy minima observed for AP6 sequences in the top-left and bottom-right corners of the free energy landscapes correspond to n_{bp}=5.

S2.3 Determining duplex melting curves from 3SPN.2 simulations

Temperature-dependent simulations for each sequence were used to determine duplex melting curves. FEPs along the biasing CV were calculated using the relation shown in Eq. S19. In order to reduce the effective oligonucleotide concentration from 5.3 mM to the experimental condition of 2 mM, the FEP was analytically extended beyond $r_{bp} = r_{ext} = 4.0$ nm as described by Hinkley et al.(18) To perform this extension, we assumed i) strands were effectively non-interacting beyond r_{ext} and ii) that the average base pair distance was approximately equal to the strand center of mass distance beyond r_{ext}. With these assumptions we used Eq. S20 to extend the Helmholtz FEP (ΔF) to an arbitrary length.

$$\Delta F(r_{bp}) = \Delta F\left(r_{ext}\right) + 2RT\ln\left(r_{ext}\right) - 2RT\ln(r_{bp}) \quad \text{(S20)}$$

The probability density P along r_{bp} was determined from ΔF using Eq. S21, where Z is a partition function used to normalize the distribution to unity.

$$P\left(r_{bp}\right)=\frac{1}{Z}\exp\left(-\frac{F\left(r_{bp}\right)}{RT}\right) \tag{S21}$$

Relative hybridized to dissociated states were determined by integrating across the probability density, where the location of the maximum free-energy barrier between the two states along r_{bp} was used to partition the hybridized probability p_h^{pbc} and dissociated probability p_d^{pbc}. In practice, the selection of the exact partition location had little effect on the relative state populations so long as it was located between r_{bp} = 1-2. The dissociated probability was cutoff at r_{bp} = 7.35 nm in order to replicate the experimental concentration. An example of the extended FEP and state partitioning is shown in Fig. S13. In order to account for local volume fluctuations in the bulk solution that are not present under periodic boundary conditions, we use Eq. S22 from Ouldridge et al(21) to determine an experimentally comparable duplex fraction p_h^{bulk}.

$$p_h^{bulk}=\left(1+\frac{1}{2\phi}\right)-\sqrt{\left(1+\frac{1}{2\phi}\right)^2-1} \quad \text{where, } \phi=\frac{p_h^{pbc}}{p_d^{pbc}} \tag{S22}$$

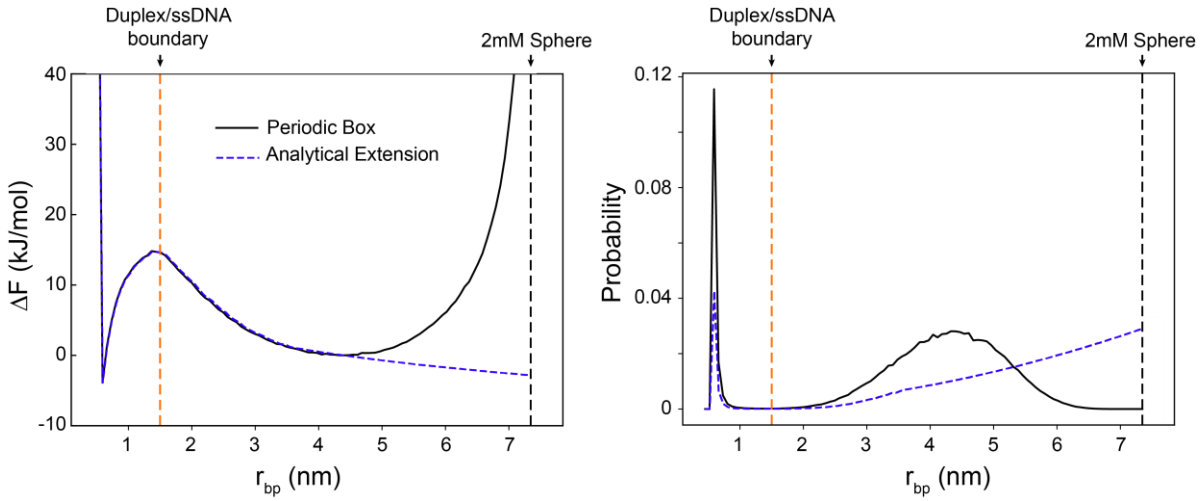

Figure S13 – Extension of FEP from WTMetaD simulation to match experimental oligonucleotide concentration. (a) Reweighted FEP along average base pair separation (r_{bp}) for CCends-WT at 331 K (solid black line, Eq. S19) containing periodic box effects with an effective oligonucleotide concentration of 5.3 mM. Analytical extension(21) of the FEP (blue dashed line) is performed to determine the duplex melting thermodynamics in larger volume sphere that corresponds to an effective oligonucleotide concentration of 2 mM. An extension of the FEP to r_{bp} =7.35 nm is used to ensure a 2 mM effective oligonucleotide concentration. The orange dashed line indicates the boundary between duplex and single-strand states. **(b)** Corresponding plot for probability density along r_{bp}.

S3. Temperature-jump IR Spectroscopy

S3.1 Characterization of t-HDVE spectra

We use the in-plane carbonyl and ring stretching vibrations of the DNA nucleobases as probes for base-pair formation and breakage.(22, 23) Each nucleobase exhibits a distinct mid-IR spectrum from 1500 to 1720 cm^{-1} that provide separate probes of G:C and A:T base pairing. HDVE and t-HDVE spectra are complicated by overlapping changes in fundamental and excited-state absorption (ESA) transitions, and CGCcap, CCends, and GCGcore sequences each contain both A:T and G:C loss features (Fig. S14). The window from 1620 to 1720 cm^{-1} of the t-HDVE contains significant overlap of A:T and G:C features. The 1660 cm^{-1} difference band contains overlapping changes in fundamental and ESA T carbonyl transitions as well as a gain in G carbonyl fundamental intensity. The 1625 cm^{-1} band contains overlap between a positive gain in the A ring fundamental and negative gain in the G carbonyl ESA. However, the window from 1520 to 1615 cm^{-1} is less congested. The difference feature centered at 1605 cm^{-1} is dominated by changes in A and T ring ESA bands and that at 1550 cm^{-1} is dominated by changes in G ring ESA bands, providing select probes to monitor changes in A:T and G:C base pairing, respectively.

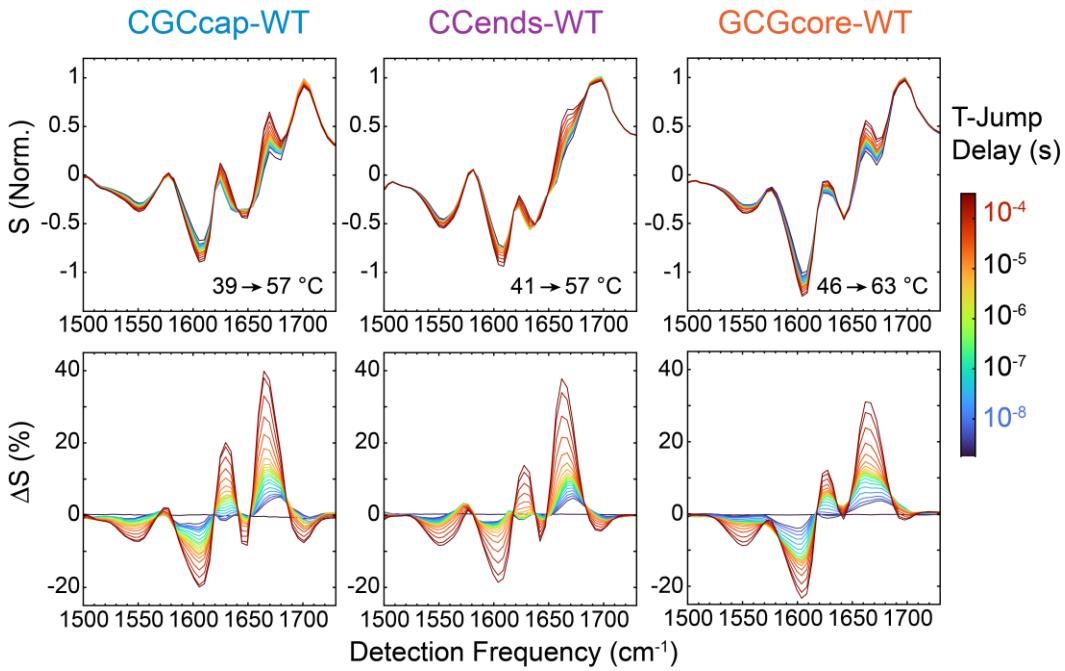

Figure S14 – Heterodyned dispersed vibrational echo (HDVE) spectra. HDVE spectra (top row) are shown as a function of T-jump delay for each WT sequence. The T-jump starting and final temperature are indicated for each sequence. Difference spectra relative to the initial temperature spectrum (t-HDVE) are shown for each respective measurement in the bottom row. t-HDVE spectra are plotted in units of percent signal change relative to the initial temperature spectrum ($\Delta S(t) = S(t)/\max(S(T_i))$).

T-jump two-dimensional IR (t-2D IR) measurements were performed to provide additional clarification into t-HDVE spectral changes. The absorptive 2D IR spectrum spreads the content of the HDVE spectrum over an additional (excitation) frequency axis. Fig. S15 shows 2D IR spectra of CGCcap-WT taken at different temperatures with parallel (ZZZZ) pump pulse polarization and at a waiting time (t_2) of 150 fs. Red-yellow features along the diagonal correspond to fundamental vibrational transitions (GSB, 0-1). Blue features just below the diagonal are excited-state absorption (ESA, 1-2) transitions where the shift along the detection frequency relative to the diagonal reflects the anharmonicity of the vibrational potential. Off-diagonal features report on intra- and inter-base coupling between nucleobase vibrations.(23-27)

Due to the extra acquisition time compared to t-HDVE measurements, t-2D IR spectra were only collected at a few select T-jump delays for each sequence. Similar to t-HDVE spectra, t-2D IR spectra are plotted as the percent change difference relative to the initial temperature (T_i) spectrum (ΔS). Red indicates the gain of a positive or loss of a negative feature while blue corresponds to a loss of positive amplitude or gain of negative amplitude. Equilibrium and transient thermal difference spectra at a T-jump delay of 320 μs are shown for CGCcap-WT in Fig. S15b-c. The equilibrium and 320 μs transient difference spectra are nearly identical, indicating that the DNA ensemble fully relaxes to the equilibrium duplex and single-strand populations at the T-jump final temperature (T_f) within hundreds of microseconds.

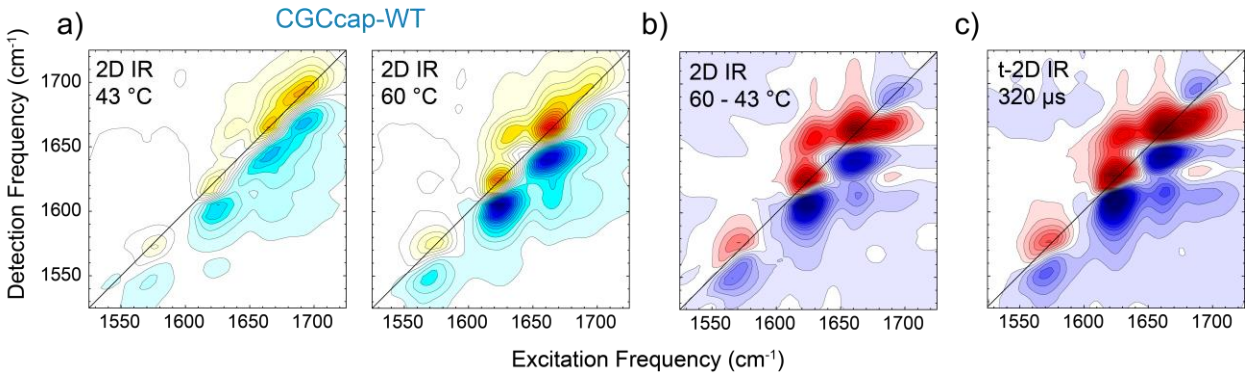

Figure S15 - Comparison of equilibrium thermal difference 2D IR and late-time t-2D IR spectra. **(a)** Absorptive 2D IR spectra of CGCcap-WT at 43 and 60 °C collected with ZZZZ pulse polarization at t_2 = 150 fs. **(b)** Equilibrium thermal difference spectrum between 60 and 43 °C and **(c)** t-2D IR spectrum at a delay of 320 μs for a T-jump from 43 to 58 °C.

In theory, the projection of the absorptive t-2D IR spectra onto the detection frequency axis should be equivalent to the real part of the t-HDVE spectra.(28, 29) Therefore, the t-2D IR spectra provide insight into the overlapping contributions to the t-HDVE spectra. Fig. S16 compares the t-HDVE spectra and t-2D IR spectra projected onto the detection frequency axis of each WT sequence at a T-jump delay of 320 μs, and good agreement is observed in each case. From 1610 to 1690 cm^{-1}, it is clear that t-HDVE features arise from combinations of features that report on A:T and G:C base pairing for CGCcap, CCends, and GCGcore sequences as shown previously for other DNA oligonucleotides.(30, 31) The largest feature in the t-HDVE spectra near 1660 cm^{-1} has positive contributions from loss of the T carbonyl ESA, gain of the G carbonyl GSB, and gain in the cross-peak between T carbonyl/ring and ring vibrations. The positive t-HDVE peak at 1625 cm^{-1} contains interference between gain in A ring GSB intensity and G carbonyl ESA intensity. In contrast, the t-HDVE spectra are less congested from 1520 to 1610 cm^{-1}. The negative feature near 1605 cm^{-1} primarily comes from a gain in the A ring ESA, but also overlaps with gains in T-T cross-peak ESA features. Therefore, this region of the t-HDVE spectra primarily reports on changes in A:T base pairing. The low-frequency tail of the A ring ESA gain interferes with a gain in G ring mode intensity near 1575 cm^{-1}. However, the gain in the ESA of these G ring modes does not significantly interfere with other features and provides a probe for G:C base pairing.

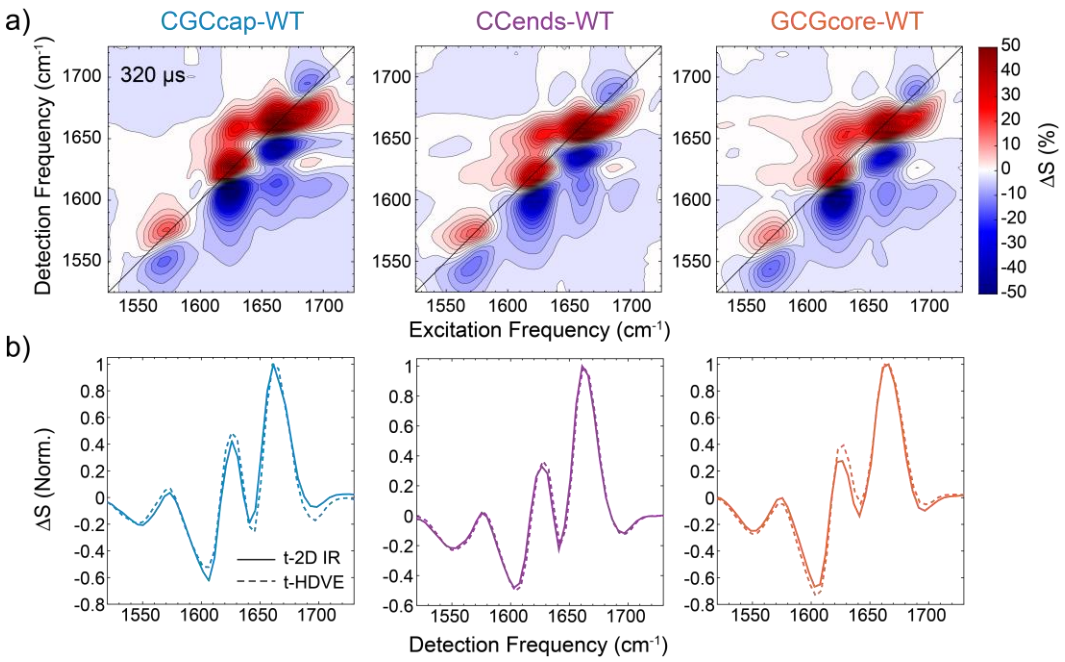

Figure S16 - Insight into t-HDVE spectral features from t-2D IR. **(a)** t-2D IR difference spectra for each WT sequence at a T-jump delay of 320 μs. 14 contours with uniform ΔS = 3.5% spacing are plotted for each sequence. **(b)** Projections of t-2D IR spectra (solid lines) in **(a)** onto detection frequency axis compared with t-HDVE spectra at a T-jump delay of 320 μs.

S3.2 Global lifetime analysis of t-HDVE data

The rate spectra representation of the t-HDVE data in Fig. S19 provides a method to assess the number of distinct exponential relaxation components in a dataset. However, the width of a component in the rate domain depends on both rate heterogeneity as well as experimental noise, and this can complicate quantitative comparisons between measurements. An alternative analysis method is global lifetime analysis, which globally fits the time-domain traces at all probe wavelengths to a user-defined number of kinetic components.(32) We applied global lifetime analysis to the t-HDVE data ($\psi_{t\text{-}HDVE}$) using three or four exponential components (N_{comp}) based an examination of the rate- and time-domain data.

$$\psi_{t-HDVE}\left(\omega,\tau_{TJ}\right)=\sum_{n=1}^{N_{comp}}A_n\left(\omega\right)\exp\left(-\frac{\tau_{TJ}}{\tau_n}\right) \tag{S23}$$

The first two or three components account for the T-jump response while the final component accounts for thermal relaxation and re-hybridization of the sample. Global fitting was performed by minimizing the objective function (R) in Eq. S24 using a nonlinear least-squares solver where

I is an identity matrix, C is the matrix of time-dependent concentration profiles, and C^+ denotes the pseudoinverse of C.(33)

$$R = \left\| \left(I - CC^{+} \right) \psi_{t-HDVE} \right\|^2 \tag{S24}$$

Fig. S17 shows an example of the global lifetime fitting applied to CGCcap-AP6. CCends-AP4, CCends-AP6, and GCGcore-AP6 where each dataset is fit to a sum of four exponential decays. The frequency-dependent amplitudes associated with each exponential component make up the decay-associated spectra (DAS) shown in Fig. S17c. The final DAS corresponds to the thermal relaxation response and has the opposite amplitude sign relative to the T-jump components.

Fig. S18 shows the DAS for each sequence determined from three or four component global lifetime fitting. As expected, the DAS and associated time constants reveal similar information to the rate distributions in Fig. S19. All sequences exhibit a first component with time constant from tens to hundreds of nanoseconds. Previous studies of short oligonucleotide dissociation with T-jump IR demonstrated that the structural dynamics on this timescale can primarily be assigned to terminal base pair fraying.(24, 30, 34-36) The first component spectra in Fig. S18 (DAS 1) are consistent with this assignment and more details are discussed in Section S3.5. A second component (DAS 2) with time constant from hundreds of ns to many μs is only needed for CGCcap-AP6, CCends-AP6, and GCGcore-AP6. The final T-jump component (DAS 3) occurs from a few to hundreds of μs depending on the temperature and sequence and corresponds to full strand dissociation and association.

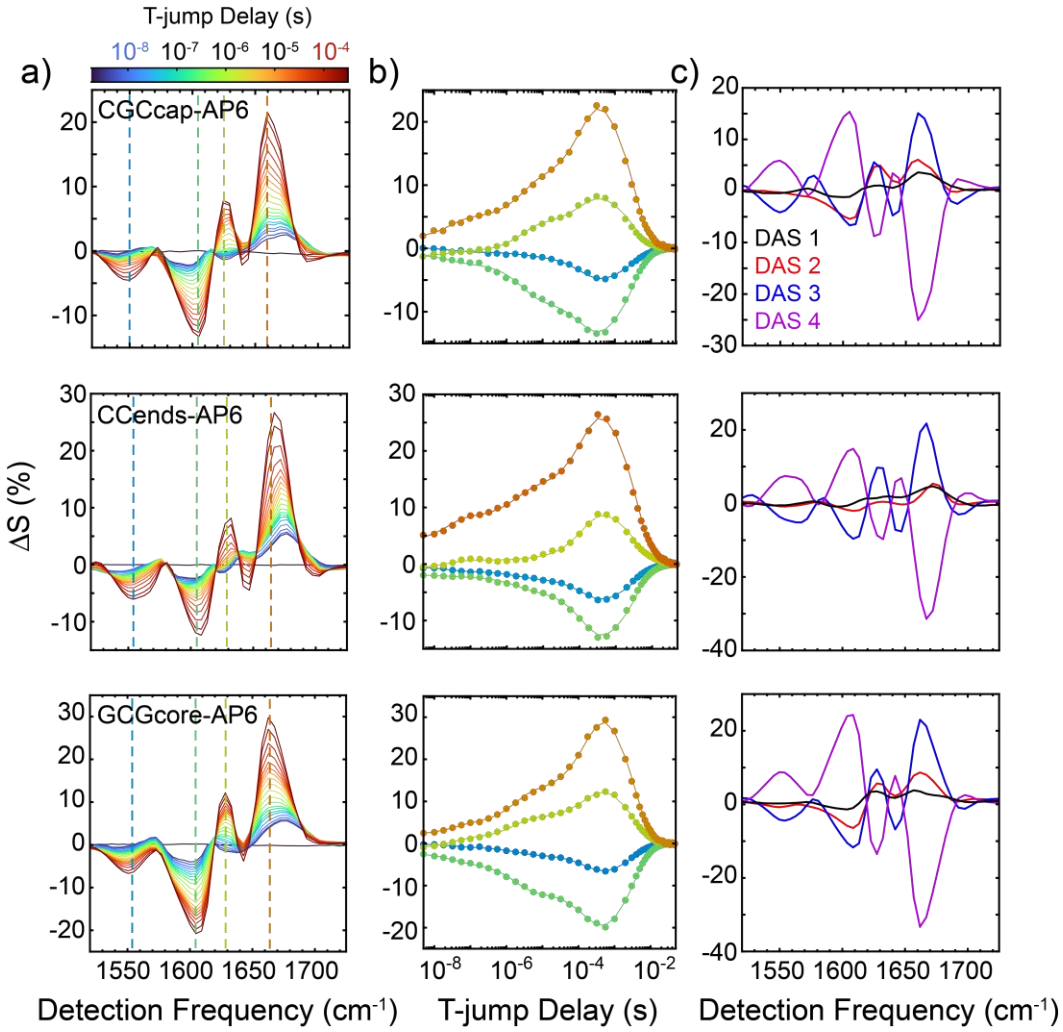

Figure S17 - Example of global lifetime fitting applied to CGCcap-AP6, CCends-AP6, and GCGcore-AP6 t-HDVE data. **(a)** t-HDVE spectra from 5 ns to 560 μs. **(b)** Representative time traces probed at 1554, 1605, 1625, and 1665 cm^{-1} with fits (solid lines) from global fitting to a series of four exponentials. **(c)** Decay-associated spectra (DAS) from four-component global fitting. The final component of opposite sign amplitude corresponds to thermal relaxation and rehybridization of the duplex.

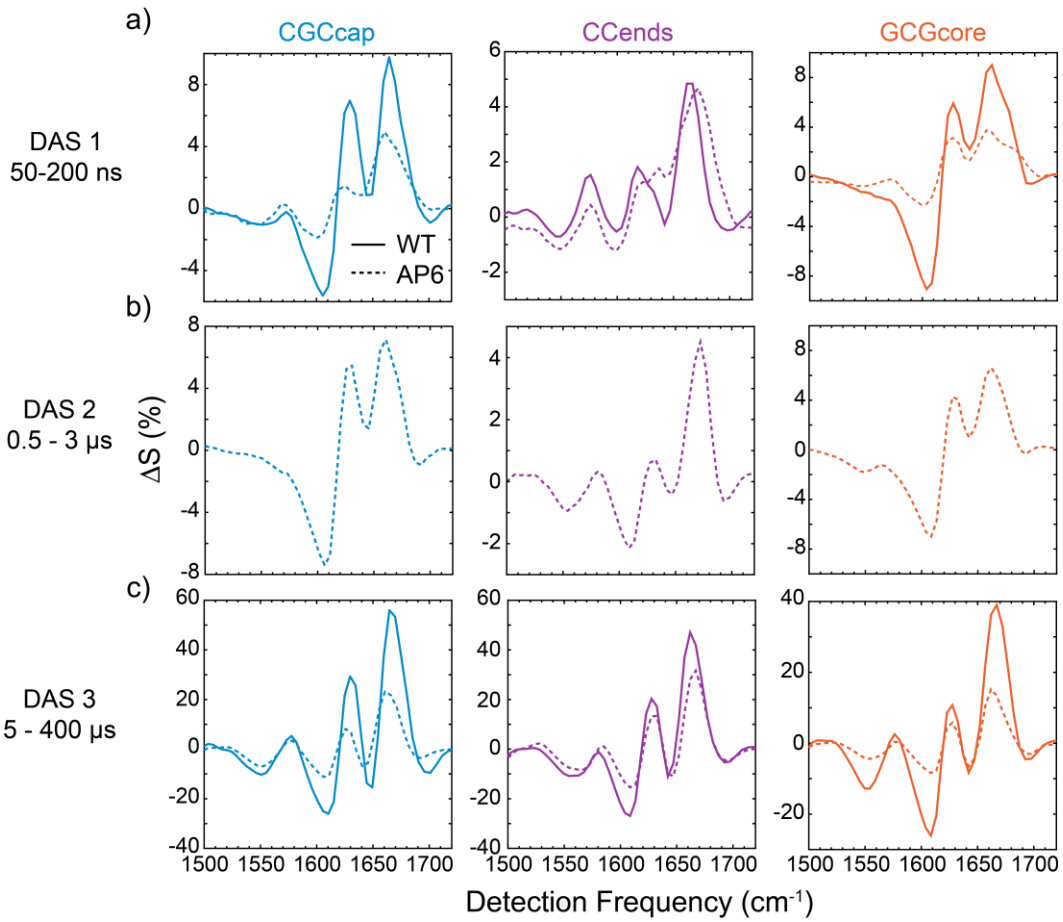

Figure S18 - Decay-associated spectra (DAS) determined from global lifetime analysis of t-HDVE data for all sequences. DAS corresponding to thermal relaxation are not shown. DAS are shown for a T-jump with T_f near the respective T_m of each sequence. **(a)** The first spectral component (DAS 1) has a time constant ranging from tens to hundreds of ns and is observed for all sequences. **(b)** An intermediate component (DAS 2) is only observed for CGCcap-AP6, CCends-AP6, and GCGcore-AP6 with a time constant ranging from hundreds of nanoseconds to many microseconds. **(c)** DAS 3 corresponds to the duplex-to-single-strand transition observed for all sequences on a timescale ranging from a few to hundreds of microseconds.

S3.3 Rate distributions from inverse Laplace transform

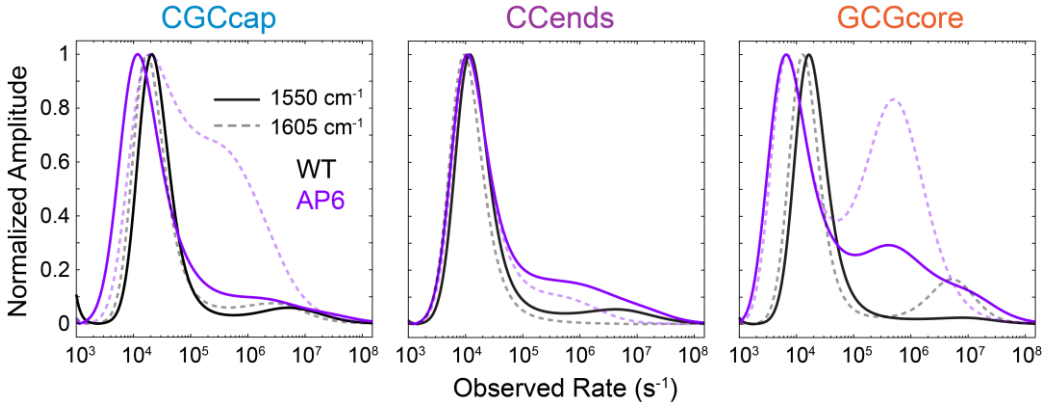

Figure S19 – t-HDVE rate distributions. Rate-domain T-jump traces are shown at 1550 (solid lines) and 1605 cm^{-1} (dashed lines) for WT (black) and AP6 (blue) sequences. Rate distributions were transformed from the time-domain t-HDVE data using an inverse Laplace transform maximum entropy method (MaxEnt-iLT) as described previously.(30, 37)

S3.4 Temperature-dependent t-HDVE data

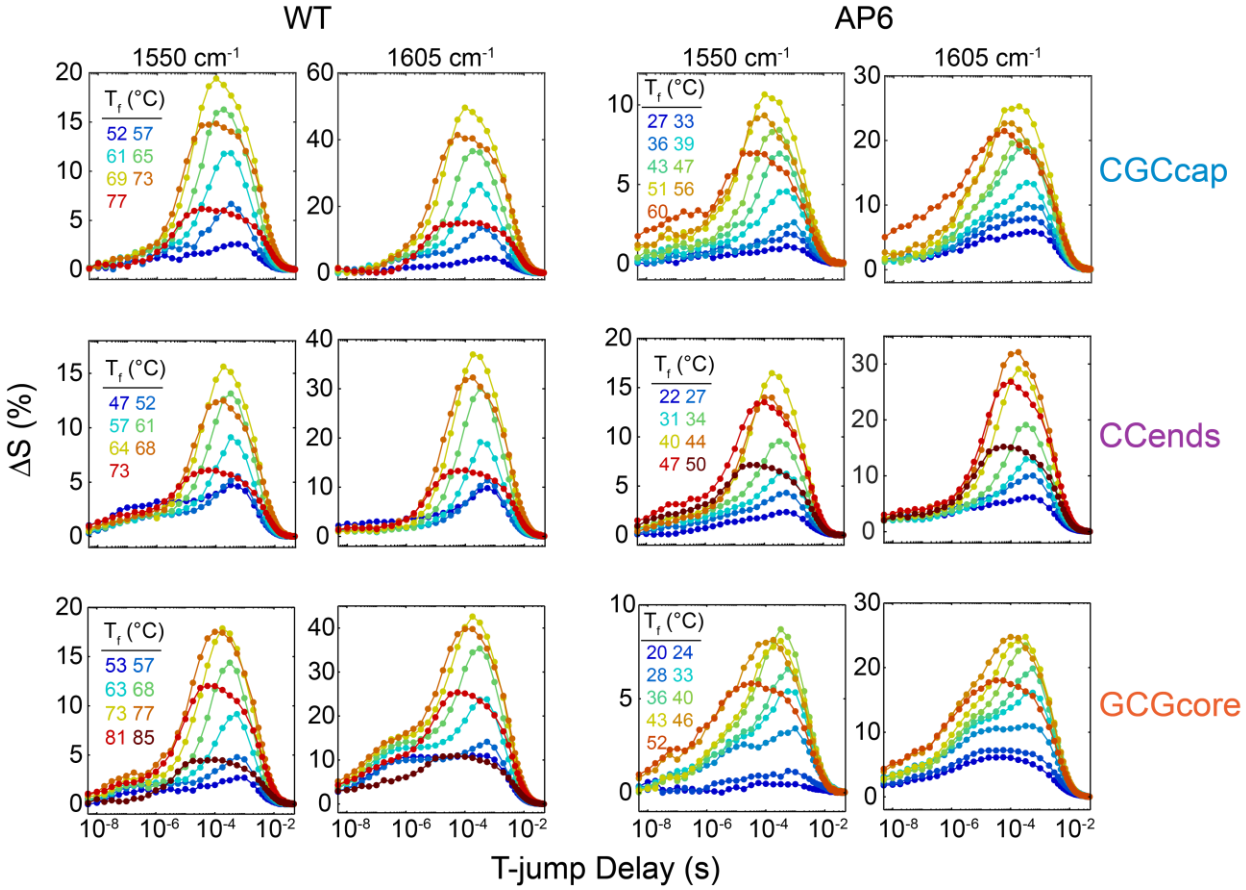

Figure S20 – Temperature-dependent t-HDVE data. t-HDVE time traces at various temperatures for CGCcap, CCends, and GCGcore sequences at probe frequencies 1550 and 1605 cm^{-1}.

S3.5 Terminal fraying in t-HDVE response

Previous studies have shown that the tens of nanosecond T-jump response (τ_1 process) observed in short DNA duplexes primarily corresponds to reshaping of the duplex free energy minimum due to terminal base pair fraying.(24, 30, 34-36, 38) Fig. S21 shows the rates and amplitudes of the τ_1 response extracted from global lifetime fitting. The observed amplitudes vary across WT sequences due to different arrangement of G:C and A:T base pairs. A stable G:C center leads to substantial A:T terminal fraying in GCGcore-WT, and CGCcap-WT has the next greatest fraying amplitude that primarily comes from the A:T end. CCends-WT has a relatively small fraying responses due to the terminal G:C base pairs. The relative sequence-dependent fraying response amplitude is similar to that observed in simulated MSM T-jump responses (Fig. 3e) Incorporation of an AP-site is shown to generally decrease the amplitude of the τ_1 process and make it undetectable in some cases. However, the reduction in fraying amplitude may also by due to the lower experimental temperatures used for AP6 sequences.

The τ_1 responses have much less signal change than the slower responses, which makes it difficult to compare the temperature-dependent observed rate constants between sequences. For most sequences, there is a minor increase in observed rate as T_f increases (Fig. S21a), and such has been observed previously for A:T terminal fraying responses.(30, 36)

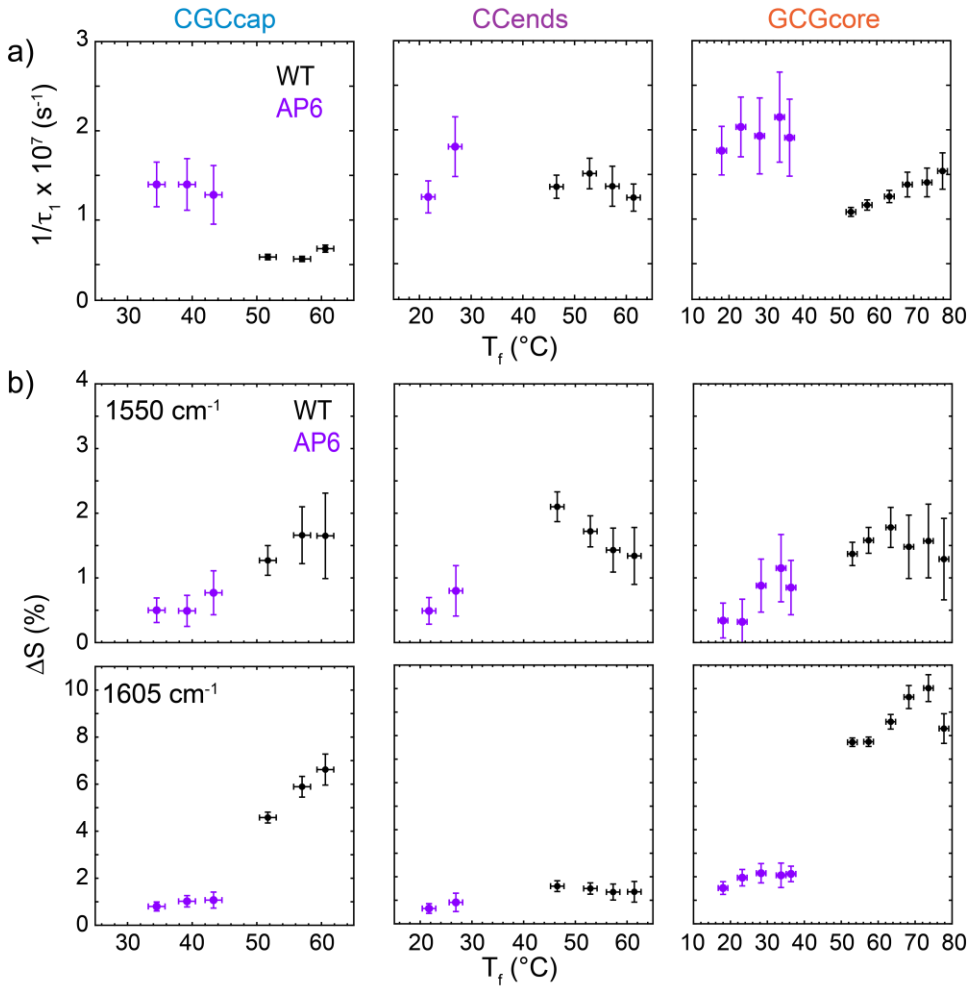

Figure S21 - Kinetic trends of τ_1 T-jump response across sequence and temperature. (a) Observed rate ($1/\tau_1$) and **(b)** amplitudes at 1550 and 1605 cm^{-1} of first component from global lifetime fitting for all sequences as a function of T_f. Vertical error bars indicate 95% confidence intervals from global fits and horizontal error bars correspond to the measured standard deviation in T-jump magnitude.

S3.6 Determination of dehybridization and hybridization rate constants

The temperature-dependence of the T-jump response during τ_3 (Figs. S20 and S22) provides insight into how an AP-site alters the kinetics of duplex dissociation and association. Dissociation (k_d) and hybridization (k_h) rate constants were derived by applying a two-state relaxation kinetics model to $1/\tau_3$ (Eq. S25) in Fig. S22a. FTIR melting curves (Fig. 1) were used to determine single-strand concentrations ([S_1] and [S_2]) and the duplex-to-single-strand equilibrium constant (K_D) at each T_f.(39)

$$1/\tau_3 = k_d + k_h\left(\left[S_1\right]_{T_f} + \left[S_2\right]_{T_f}\right) \quad \text{(S25)}$$

$$1/\tau_3 = \frac{k_d K_{D,T_f}}{K_{D,T_f} + \left[S_1\right]_{T_f} + \left[S_2\right]_{T_f}}, \text{ where } K_{D,T_f} = \frac{k_{d,T_f}}{k_{h,T_f}} \quad \text{(S26)}$$

Many of the studied sequences undergo partial dissociation during τ_1 and τ_2 prior to complete dehybridization. As a result, a two-state model reports on the kinetics between the duplex ensemble after τ_2 and the single-strand ensemble. The temperature-dependence of k_d and k_h can be described by a Kramers-like expression to extract enthalpic ($\Delta H^{\ddagger}$) and entropic ($\Delta S^{\ddagger}$) barriers where the temperature-dependence of the D_2O solvent viscosity (η) is taken into account.(40) c° is the standard state concentration of 1 M raised to the reaction molecularity (m = 1 for k_d and m = 2 for k_h)

$$k_{d/h}(T) = \left(c^{\circ}\right)^{1-m} \frac{k_b T}{\hbar} \frac{\eta_{37}}{\eta(T)} \exp\left(-\frac{\Delta H^{\ddagger}_{d/h} - T\Delta S^{\ddagger}_{d/h}}{RT}\right) \quad \text{(S27)}$$

In Eq. S27, we assume a transmission coefficient of 1, yet the true value is likely much smaller.(41) Overestimating the transmission coefficient leads to a reduction in $\Delta S^{\ddagger}$ and therefore an increase in free energy barrier ($\Delta G^{\ddagger}$) for both dehybridization and hybridization. However, assuming the transmission coefficient or speed limit of (de)hybridization is independent of sequence, the relative trends of $\Delta S^{\ddagger}$ and $\Delta G^{\ddagger}$ across sequence and AP-site position are still informative.

All sequences exhibit a large enthalpic dehybridization barrier ($\Delta H_d^{\ddagger}$ >150 kJ/mol, Fig. S22) primarily due to the breaking of base pairing interactions, consistent with previous studies of short oligonucleotides.(30, 42-45) Incorporation of an AP-site reduces $\Delta H_d^{\ddagger}$ in all sequences. The positive $\Delta H_d^{\ddagger}$ is accompanied by an increase in entropy ($\Delta S_d^{\ddagger}$, Fig. S21) that partially compensates for $\Delta H_d^{\ddagger}$ and leads to more minor variation in the dissociation free energy barrier at 37 °C ($\Delta G_{d,37}^{\ddagger}$, Fig. S22) compared to $\Delta H^{\ddagger}_d$. Each sequence shows a reduction in $\Delta G_{d,37}^{\ddagger}$ when an AP-sites is incorporated, and this change in $\Delta G_{d,37}^{\ddagger}$ roughly matches the shift in free energy difference between duplex and single-strand states ($\Delta G_{d,37}^{o}$, Fig. 1c).

T-jump measurements are inherently more sensitive to dehybridization kinetics since $k_d >> k_h([S_1]+[S_2])$ over much of the experimental temperature range. Typically, only the few lowest temperature conditions measured are sensitive to k_h, and the value of $1/\tau_3$ under these conditions is often near the slower end of our experimental time window. Previous work has shown that k_h exhibits a non-Arrhenius temperature dependence when measured over a wide enough temperature range,(44) so additional measurements of k_h should be performed below the melting transition for proper assessment of hybridization kinetics. With these cautions in mind, we examine the trends in k_h determined across the sequences in this work (Fig. S22). Overall, the values of k_h show minor variation with sequence and temperature that are typically within the estimated experimental error. The minor temperature-dependence suggests the free energy barrier ($\Delta G_h^{\ddagger}$) is dominated by a loss of entropy over the measured temperature ranges. However, low temperature mixing experiments have revealed clear sequence-dependent kinetics in hybridization rate.(46) As shown in Fig. S22f, the value of $\Delta G_h^{\ddagger}$ at 37 °C is almost same for all sequences, and this would suggest that AP-sites do not significantly impact the rate of association. However, more direct measurements of k_h are required to assess the impact of AP-sites on association kinetics.

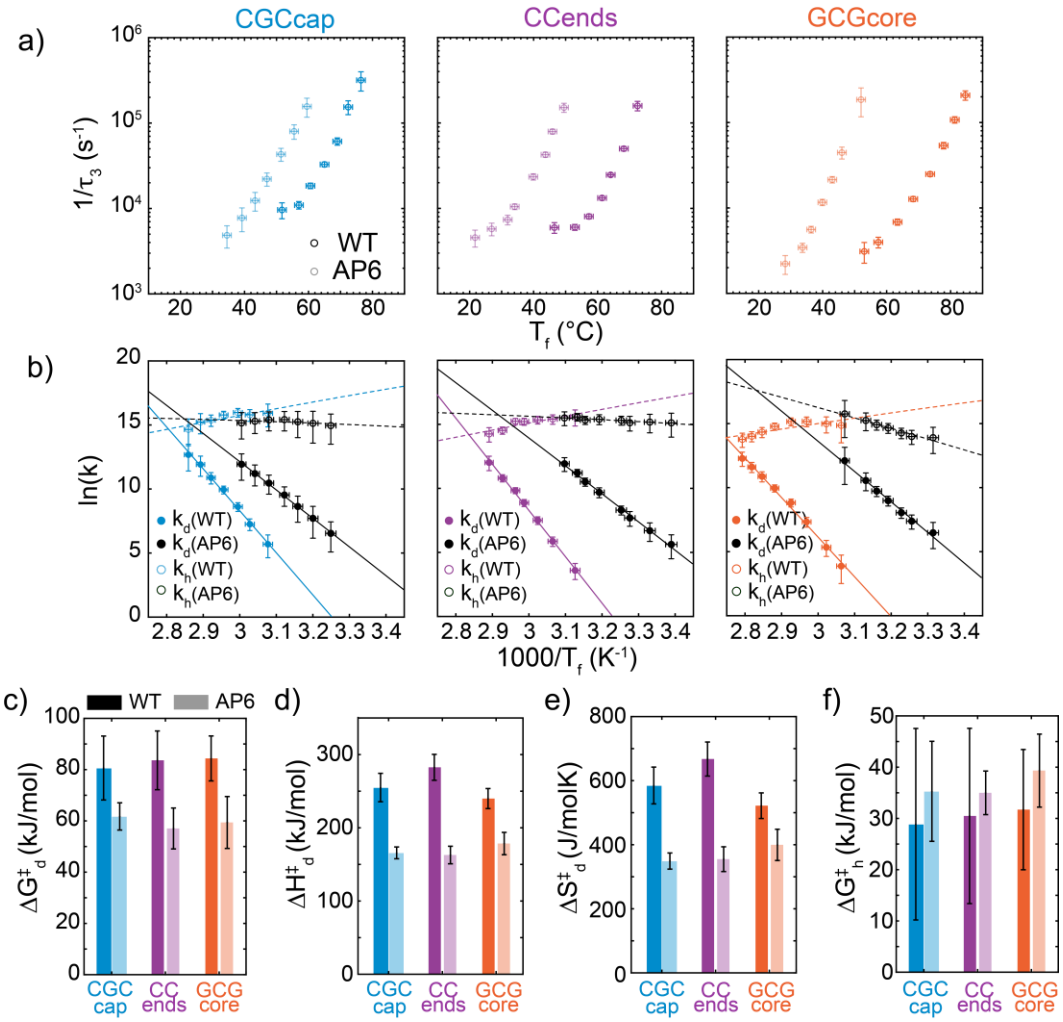

Figure S22 – Temperature-dependent dehybridization and hybridization kinetics measured with T-jump IR spectroscopy. (a) Slowest observed relaxation rates ($1/\tau_3$) for WT (dark) and AP6 (light) sequences obtained from global lifetime fitting of t-HDVE data (Section S3.2). Vertical error bars indicate 95% confidence intervals from global fits and horizontal error bars correspond to the measured standard deviation in T-jump magnitude. (**b)** Dehybridization (k_d) and hybridization (k_h) rate constants extracted from two-state modelling of $1/\tau_3$. Vertical error bars indicate 95% confidence intervals propagated from global fits and horizontal error bars correspond to the measured standard deviation in T-jump magnitude. Temperature-trends are fit to a Kramers-like equation (Eq. S27, solid lines for WT, dashed lines for AP6). (**c)** Dehybridization free energy barrier at 37 °C ($\Delta G^{\ddagger}_{d,37}$) (**d)** Dehybridization enthalpy barrier ($\Delta H^{\ddagger}_{d}$) **(e)** Dehybridization entropic barrier ($\Delta S^{\ddagger}_{d}$) (**f)** Hybridization free energy barrier ($\Delta G^{\ddagger}_{h,37}$) determined from Kramers fit in (**b)** for WT (dark) and AP6 (light) sequences. Error bars correspond to 95% confidence intervals from Kramers fit.

S3.7 Determination of A:T and G:C character in τ_2 process

Fig. 3c shows the fraction of A:T and G:C character in the τ_2 response observed for CGCcap-AP6, CCends-AP6, and GCGcore-AP6. For these calculations, we assume that the t-HDVE signals at 1550 and 1605 cm^{-1} report exclusively on changes of the guanine ring and A/T ring ESA bands, respectively. In reality, there are also small overlapping contributions from changes in the ESA band of the guanine carbonyl mode to the A/T region and the A/T ESA band to the G region as discussed in Section S3.1. The τ_2 T-jump amplitudes are shown in Figure S22a for values of T_f where a response was measurable. The percentage (P) of G:C and A:T dehybridization during the intermediate response is determined from the ratio of total dissociation amplitude ($A_2 + A_3$) and τ_2 response amplitude (A_2) at 1550 and 1605 cm^{-1}.

$$P_{G/A} = \frac{A_{2,G/A}}{A_{2,G/A} + A_{3,G/A}} \tag{S28}$$

The values of P at 1550 and 1605 cm^{-1} are shown in Fig. S23a. P generally decreases in going to higher temperatures due to a greater loss of duplex population (i.e. larger A_3). Then, the A:T and G:C character (C) are determined by the ratio of $P_{A/G}$ to the sum of P_A and P_G weighted by the number of A:T (N_A) and G:C (N-N_A) base pairs in the sequence. N is the total number of starting base pairs in the sequence, which is assumed to be 10 for sequences containing an AP-site.

$$C_{A:T} = \frac{N_A P_A}{(N - N_A) P_G + N_A P_A} \tag{S29a}$$

$$C_{G:C} = \frac{(N - N_A) P_G}{(N - N_A) P_G + N_A P_A} = 1 - C_{A:T} \tag{S29b}$$

The values of $C_{A:T}$ are shown in Fig. S23c and are nearly the same, within experimental error, across the temperature ranges measured. The mean value of $C_{A:T}$ and $C_{G:C}$ across measured temperature points are plotted in Fig. 3c.

We also determined $C_{A:T}$ and $C_{G:C}$ for CCends-AP6 and GCGcore-AP6 using t-2D IR spectroscopy in order to test the validity of the G:C and A:T spectral markers in the t-HDVE data. Spectra were acquired at T-jump delays of 180 ns, 10 μs, and 320 μs to reflect the fast solvation and terminal fraying dynamics, half-dehybridization response, and duplex-to-single-strand transition, respectively (Fig. S24). We used the difference between 10 μs and 180 ns spectra as the

τ_2 response and the difference between 320 μs and 180 ns surfaces for the total dissociation. The integrated signal change over certain regions of the absolute value difference spectra were used to determine G:C and A:T response. We used the G ring mode region (cyan) for the isolated G:C response and the A/T ring mode region (blue) for A:T. We also compare with the signal change of the T intrabase cross-peak (purple) between the carbonyl/ring vibration near 1660 cm^{-1} and ring vibration at 1630 cm^{-1} as well as the overlapping signal changes from G carbonyl and T carbonyl/ring vibrations (pink). The ratio between G:C (cyan) and A:T (blue) amplitudes determined on the 10 μs – 180 ns surface and 320 μs – 180 ns surface should be nearly equivalent to the $P_{G/A}$ values determined from t-HDVE data. Then, $C_{A:T}$ and $C_{G:C}$ can be determined from the t-2D IR data using Eqs. S29a,b. The C values determined with t-2D IR and t-HDVE data are nearly identical (Figs. S24c and 3c), which indicates that the minor interference between A/T and G/C cross-peak and ESA amplitude changes is negligible at 1550 and 1605 cm^{-1} in the t-HDVE spectra.

In addition to calculating the relative A:T and G:C loss contributions to the intermediate T-jump response, the fraction of total base pair dissociation during the intermediate response (D_2) can be estimated. We simply calculate D_2 as the weighted mean of G:C and A:T intermediate dissociation percentages.

$$D_2 = \frac{\left(N - N_A\right) P_{G:C} + N_A P_{A:T}}{N} \tag{S30}$$

In contrast to $C_{G:C}$ and $C_{A:T}$, D_2 strongly depends on temperature and typically decreases sharply towards higher temperatures. The temperature trends of D_2 are shaped by the temperature-dependence of the duplex-to-single-strand response (A_3), which is determined by the change in duplex and single-strand concentration during the T-jump.

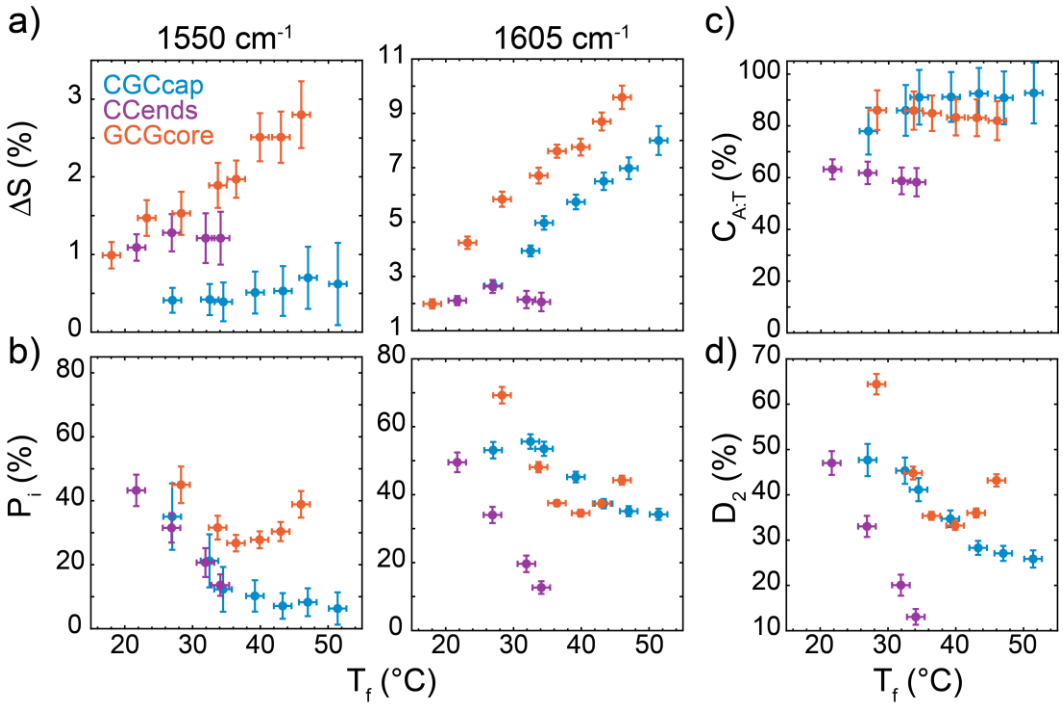

Figure S23 – Calculation of A:T/G:C character in τ_2 response. (a) Amplitudes of second component (DAS 2) probed at 1550 and 1605 cm^{-1} from global lifetime fitting of t-HDVE data for CGCcap-AP6, CCends-AP6, and GCGcore-AP6 as a function of T_f. **(b)** Amplitudes in **(a)** divided by the sum of DAS 2 and DAS 3 amplitudes (Eq. S28). **(c)** A:T character of τ_2 response at each temperature from Eq. S29. **(d)** Fraction of total dehybridization response occurring during τ_2 determined using Eq. S30.

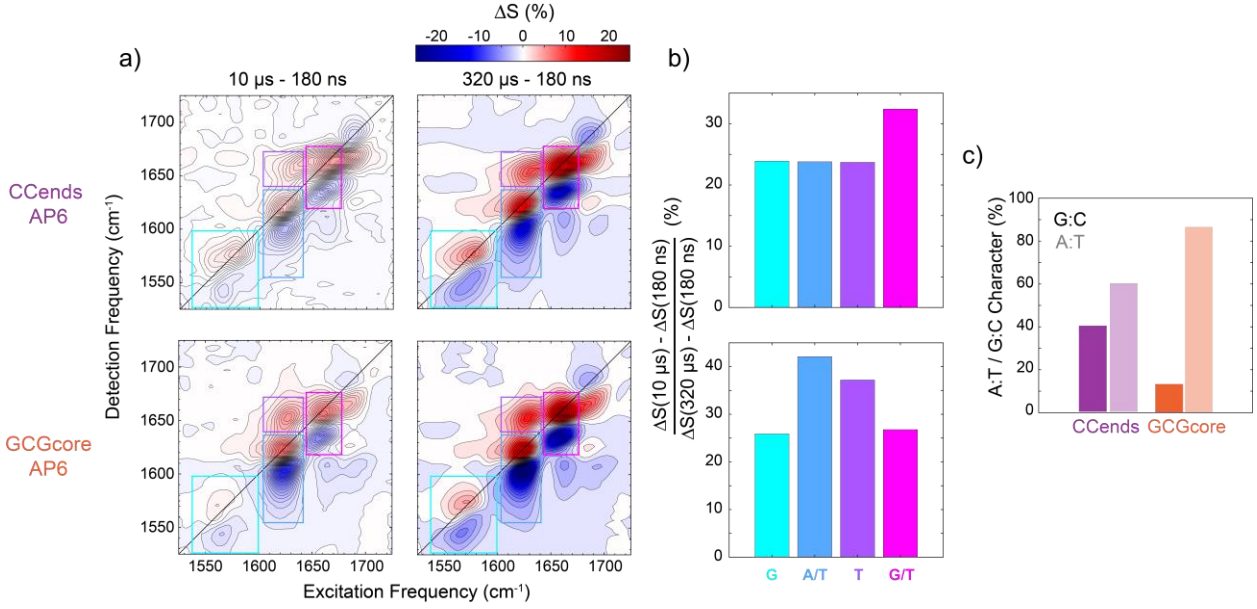

Figure S24 - Determination of relative A:T and G:C spectral content in τ_2 response from t-2D IR. (a) t-2D IR spectra taken as the difference between (left) 10 μs and 180 ns and (right)

320 μs and 180 ns for CCends-AP6 and GCGcore-AP6 sequences. Spectral change amplitude is plotted as the percent change relative to the T_i spectrum. 320 μs – 180 ns spectra are plotted with 14 contours using uniform ΔS = 1.4 % spacing for CCends-AP6 and GCGcore-AP6. 10 μs – 180 ns spectra are plotted with 14 contours using uniform 0.4 and 0.7 % spacing for CCends-AP6 and GCGcore-AP6, respectively. **(b)** Ratio of integrated spectral change between 10 μs – 180 ns and 320 μs – 180 ns t-2D IR spectra in the spectral regions denoted by boxes in **(a)**. Integrations were performed over the marked regions on the absolute value t-2D IR spectra to avoid cancellation between positive and negative amplitudes. **(c)** Percentage of G:C (dark) and A:T (light) base pair loss character observed in τ_2 T-jump responses for CCends-AP6 and GCGcore-AP6 determined from t-2D IR measurements.

S4. Construction and validation of Markov state models from 3SPN.2 MD simulations

S4.1 Construction and validation of state-reversible VAMPnets (SRV) MSMs

We employed MSMs to infer quantitative and interpretable long-time kinetic models, perform experimental relaxation comparisons, and elucidate key transition states and pathways from our aggregated simulation data. MSMs have a robust theoretical foundation(47-49) and details of our MSM construction pipeline for CG DNA can be found in previous work.(36, 50) In brief, we followed three fundamental steps to generate optimized MSMs across each system: (i) trajectory featurization, (ii) dimensionality reduction, (iii) microstate clustering and microstate transition matrix inference. Instead of additional hierarchical clustering into coarse macrostates, we calculated all observables from microstates directly and performed relaxation experiments on the microstate transition matrix. Calculations were performed using the PyEMMA software package.(51)

Featurization. Trajectories consisting of Cartesian coordinates of the DNA strands over time were featurized using the MDtraj Python libraries.(52) An optimal feature set should contain all the required information to represent the essential system dynamics while eliminating unwanted translation and rotational invariances. As in our previous work,(36) we adopt intermolecular pairwise distances between the centers of mass of all available CG nucleobase sites, resulting in 11 x 11 = 121 features for WT systems and 10 x 11 = 110 features for systems containing an AP-site. Due to 5’ vs. 3’ asymmetries and that each sequence is non-self-complementary, each feature is unique and permutationally invariant. We used VAMP-2 scoring(53-55) – the sum of the squared estimated eigenvalues of the transfer operator – to evaluate the kinetic variance of several

competing features sets and found (i) reciprocal distances perform better than non-reciprocal distances and (ii) the addition of more features such as intramolecular distances between nucleobases on the same strand or phosphate and sugar distances provide very little improvement to the VAMP-2 score. Based on these findings, we selected reciprocal nucleobase intermolecular distances as a consistent feature set for each system.

Dimensionality Reduction. The featurized trajectories were embedded into a kinetically informed low-dimensional space in preparation for microstate clustering. In order to find an optimal kinetic representation, we built independent State-free Reversible VAMPnets (SRVs) for each system. SRVs function as a nonlinear analog to time-lagged independent components analysis (tICA) and are used to retain a maximum amount of a system's kinetic variance.(56) The twin-lobed neural network structure constructs an optimized low-dimensional space as a function of its input features, and produces higher resolution MSMs.(36, 50) Due to the large number of systems studies, optimized hyperparameter tuning across each sequence/AP combination would have entailed burdensome computational costs and also limited comparability between systems. As such we used the same number of SRV embedding dimensions, discrete microstates, and training epochs, etc. for each system. For each of these hyperparameters, we ensured that the embedding space was rich enough to capture all relevant kinetic states but not so expansive as to diminish clustering efficiency or overfit the model. To verify the optimized embedding dimensionality, we performed five-fold cross-validation of our SRV-MSM pipeline and found a plateau in the VAMP-2 score for both training and test sets when using a five-dimensional SRV embedding (Fig. S25). Although some systems show a small increase in VAMP-2 score beyond five dimensions, most test scores either remained the same or dropped slightly, indicating a loss in generality with higher dimensionality. SRVs were trained using a previously developed package (https://github.com/hsidky/srv) employing the default network architecture of two hidden layers each comprising 100 neurons and tanh activation functions, a learning rate of 0.001, a batch size of 50,000, and a lag time of 1 ns.

SRV Microstate clustering. For each system, the SRV embeddings of 2.5 million frames were clustered into microstates using k-means clustering. We found little improvement in VAMP-2 score when the number of clusters exceeded 200, motivating our choice of 200 microstate clusters to discretize the SRV coordinates for each system. For all system, we found

that a lag time of 1 ns was long enough to converge leading slow timescales while resolving higher order faster dynamics (Fig. S26). Chapman-Kolmogorov (CK, Fig. S27) tests show that predictions from our MSMs constructed with a lag time of 1 ns are in approximate agreement with MSMs constructed at longer lag times.

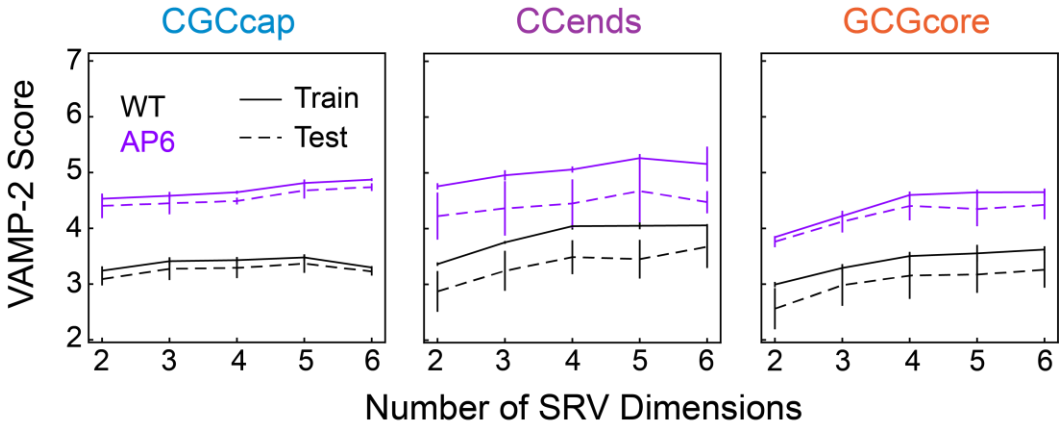

Figure S25 - VAMP-2 scoring with 5-fold cross-validation across all systems as a function of embedding dimensions. A plateau in the training and testing scores at or before 5 dimensions indicates a near optimal embedding space across systems. A decrease in VAMP-2 beyond 5 dimensions for some sequences may be indicative of overfitting. Error bars correspond to the standard deviation computed over the cross-validation partitions.

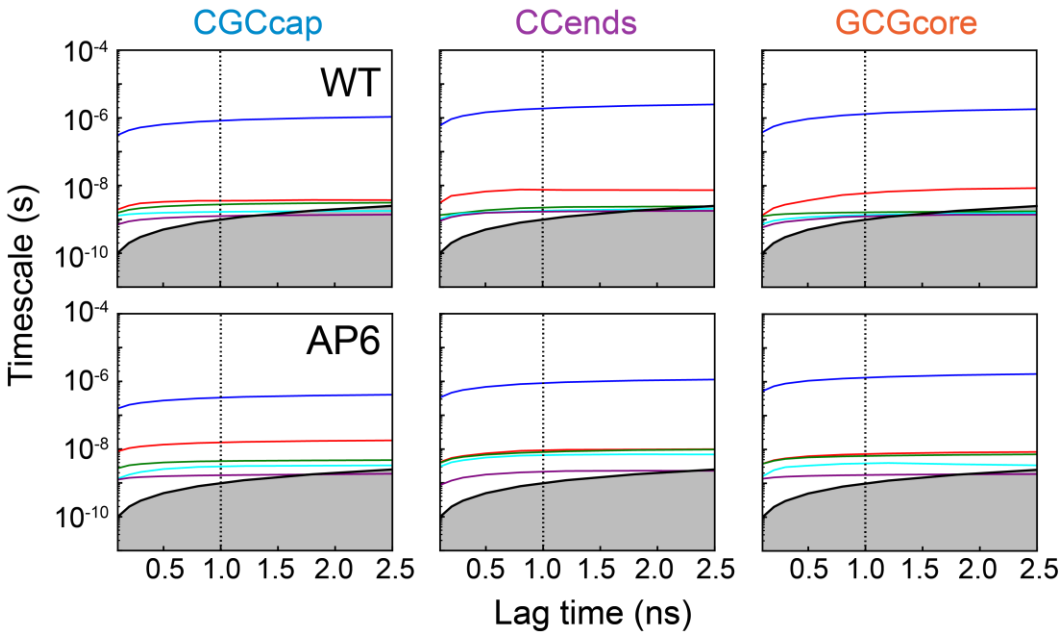

Figure S26 - Implied timescale convergence plots as a function of lag time for each system after SRV dimensionality reduction and k-means clustering. At a lag time of 1 ns (vertical dashed line), we note that most slow processes have converged and that many fast processes are still resolvable. As noted previously(36) the leading slow mode corresponds to the overall hybridization/dissociation and converges much slower than higher order modes.

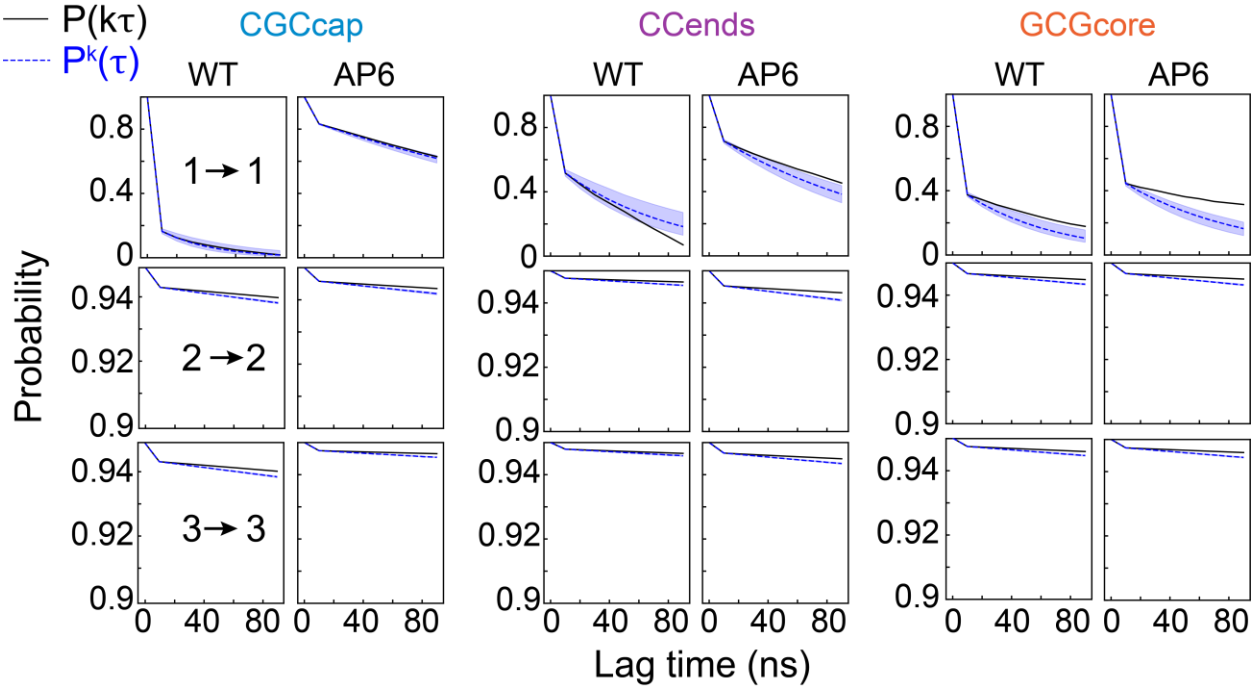

Figure S27 - Chapman-Kolmogorov tests for SRV-MSMs constructed for all sequences using a lag time of 1 ns. For each system, the self-transition probabilities for 3 PCCA+ sets are shown. Strong agreement between probabilities predicted from k applications of an MSM constructed at a lag time of 1 ns ($P^k(\tau)$, dashed blue line) and those computed from an MSM built at a lag time of k ns ($P(k\tau)$, black solid line) indicate robust Markovian behavior out to at least 100 ns. The blue shaded region indicates the estimated standard deviation in $P^k(\tau)$.

S4.2 MSM T-jump relaxation simulations

MSM relaxation simulations are a powerful way to probe dynamical signature of an out-of-equilibrium distribution as it equilibrates over time.(57) This technique is therefore amenable to T-jump experiments where a distribution favoring duplex configurations transitions to an equilibrium distribution favoring single-stranded states. Eq. S31 shows how an experimentally relevant observable, *y*, can be monitored over time (E_y) as some initial distribution, p_0, is propagated *k* times by a transition matrix *P*.

$$E_y(k\tau) = p_0^T P(\tau)^k \, y \tag{S31}$$

For each system, we defined P as our optimized MSM transition matrix, built on extensive sampling near $T_{m,MD.}$ When determining an initial distribution for each system, we found that initiating simulations in the fully hybridized microstate at $T_{m,MD}$ was not representative of the low-temperature distribution. Instead, we ran a series of short (25 x 1 μs) simulations at 15 °C below $T_{m,MD}$ for each sequence, projected these data into pre-defined $T_{m,MD}$ microstates, and used the low-temperature microstate distribution as the initial distribution for the MSM built at $T_{m,MD}$. Although these low-temperature simulations do not fully sample the dissociated or metastable states, they provide a good estimate for the ensemble of partially dissociated duplex configurations 15 °C below $T_{m,MD}$.

In addition to an initial distribution and transition matrix, a scoring function is needed to draw comparison with experiment. Given that a quantitative mapping between DNA base pair separation and IR spectrum is unknown, we developed two approximate scoring functions that quantify the number of intact A:T and G:C base pairs. The score assigned to each microstate is the average number of intact A:T or G:C base pairs within that microstate. Each base was restricted to form one intermolecular pair, resulting in a scoring range between zero and the number of A:T or G:C pairs in a sequence. In Eqs. S32a,b scores for a given microstate a_i are calculated as a mean over all x frames in set of microstate frames S_i. All available base pairs x_p are evaluated against the cutoff c, which was defined as 0.3 nm greater than the equilibrium separation between nucleobase sites (center-of-mass) involved the base pair (0.85 nm for G:C and 0.90 nm for A:T). In practice, the exact definition of this cutoff had little effect the resulting relaxation profiles (Figure S28).

$$y_i^{AT} = \frac{1}{N_i} \sum_{x \in S_i} \sum_{p=0}^{N_{\max}^{AT}} \begin{cases} 1 & x_p^{AT} < c^{AT} \\ 0 & x_p^{AT} \geq c^{AT} \end{cases} \tag{S32a}$$

$$y_i^{GC} = \frac{1}{N_i} \sum_{x \in S_i} \sum_{p=0}^{N_{\max}^{GC}} \begin{cases} 1 & x_p^{GC} < c^{GC} \\ 0 & x_p^{GC} \geq c^{GC} \end{cases} \tag{S32b}$$

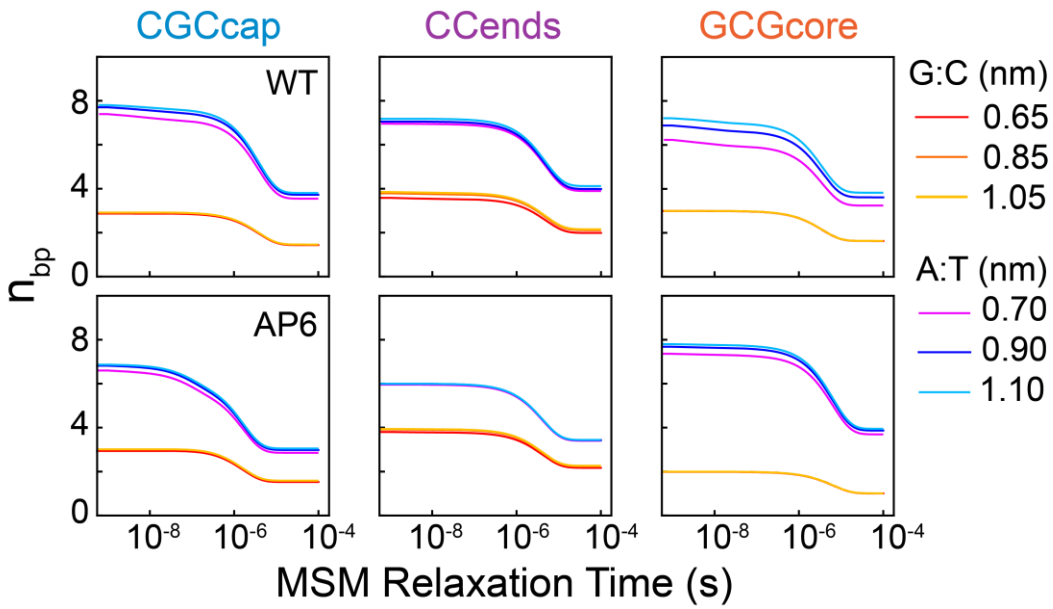

Figure S28 – MSM simulated T-jump responses with various base pair cutoff distances. Simulated T-jump relaxation profiles are shown in terms of the number of intact A:T (blue) and G:C (red) base pairs (n_{bp}). Relaxation profiles are shown using base pair separation cutoff values ranging from 0.65 to 1.05 nm for G:C base pairs and 0.70 to 1.10 nm for A:T base pairs. All profiles are qualitatively consistent across cutoff values where lower cutoff values produce systematically lower amplitude responses across all sequences.

S5. Oligonucleotide length series

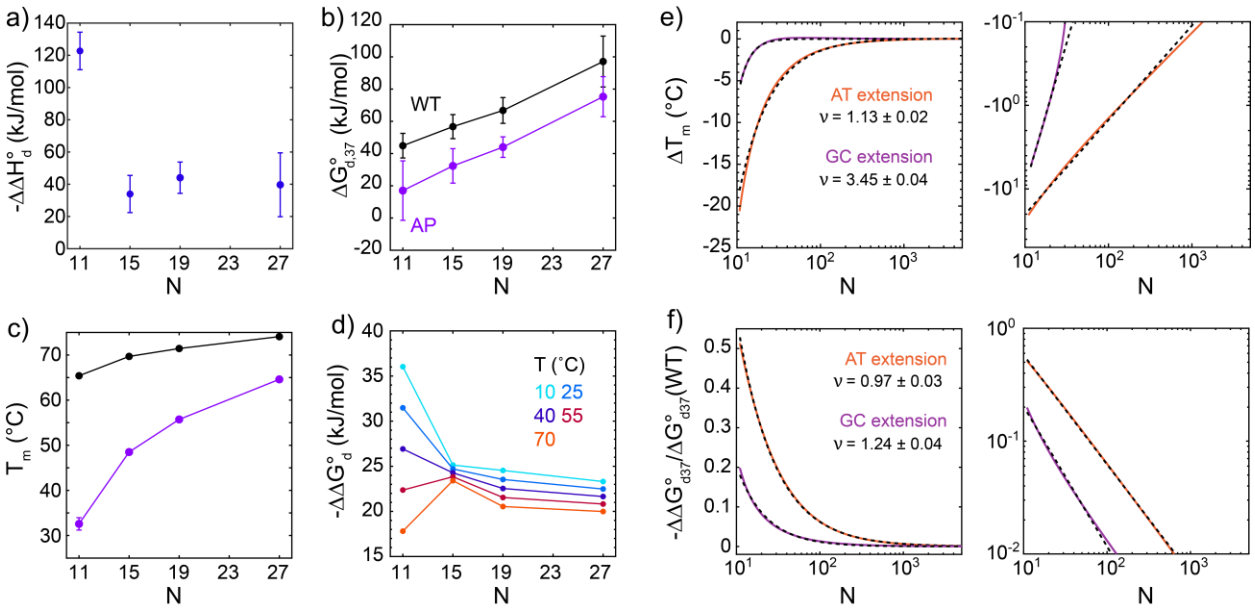

Figure S29 – Oligonucleotide length effects on dehybridization thermodynamics and T-jump IR relaxation. (a) $\Delta\Delta H^\circ_d$ **(b)** $\Delta G^\circ_{d,37}$ and **(c)** T_m from FTIR melting curves and NN model predictions for length-dependent GCGcore-like oligonucleotides shown in Fig. 4a. FTIR error bars correspond to 95% confidence intervals from two-state fits. **(d)** $\Delta\Delta G^\circ_{d,37}$ from FTIR melting curves at various temperatures. **(e)** NN model predictions of ΔT_m for GCGcore and CCends template sequences extended from N = 11 to 5,000 shown on (left) log-linear and (right) log-log scales. The GCGcore and CCends sequences were symmetrically extended with AT and GC dinucleotide repeats, respectively. The NN predictions approximately follow power-law behavior as illustrated by fits to $\Delta T_m(N) = AN^{-\nu}$ (dashed black lines), and ν is listed for each sequence with an error that corresponds to the 95% confidence intervals of the fit. The log-log plots indicate minor deviations from power-law behavior for ΔT_m>0.1˚C. **(f)** NN model predictions of $\Delta\Delta G^\circ_{d,37} / \Delta G^\circ_{d,37}(\mathrm{WT})$ on (left) log-linear and (right) log-log scales.

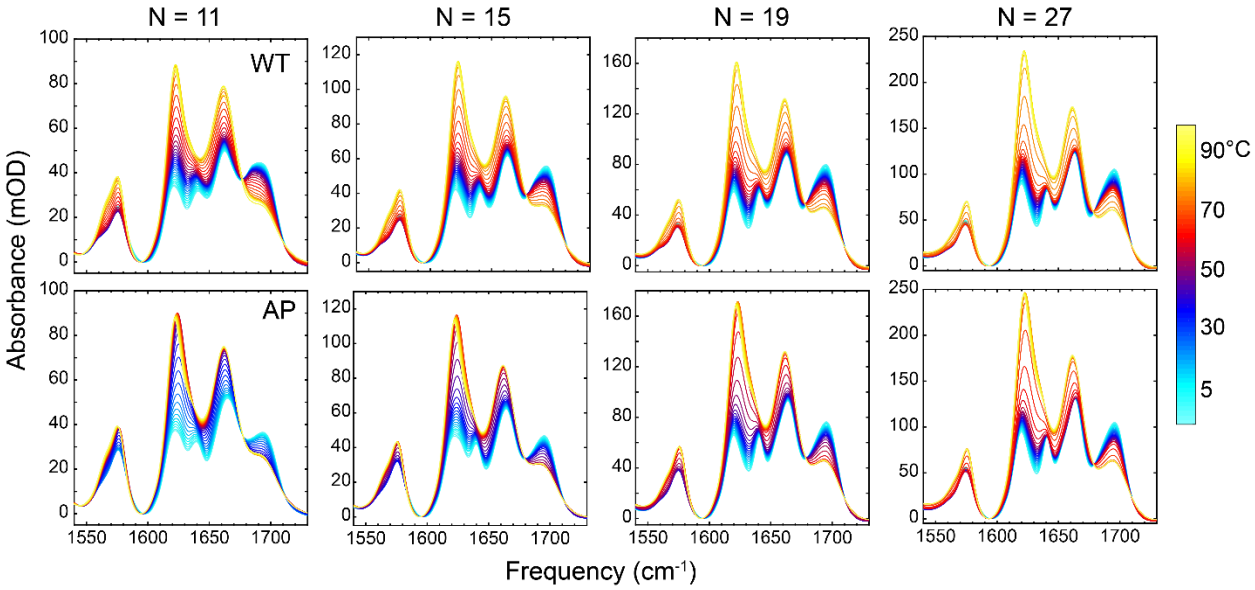

Figure S30 – Temperature-dependent FTIR spectra of different length oligonucleotides. FTIR temperature series for N = 11, 15, 19, and 27 sequences shown in Fig. 4a measured from 3.8 to 96.3 °C. All samples were prepared in deuterated pH* 6.8 400 mM SPB with a total oligonucleotide concentration of 2 mM.

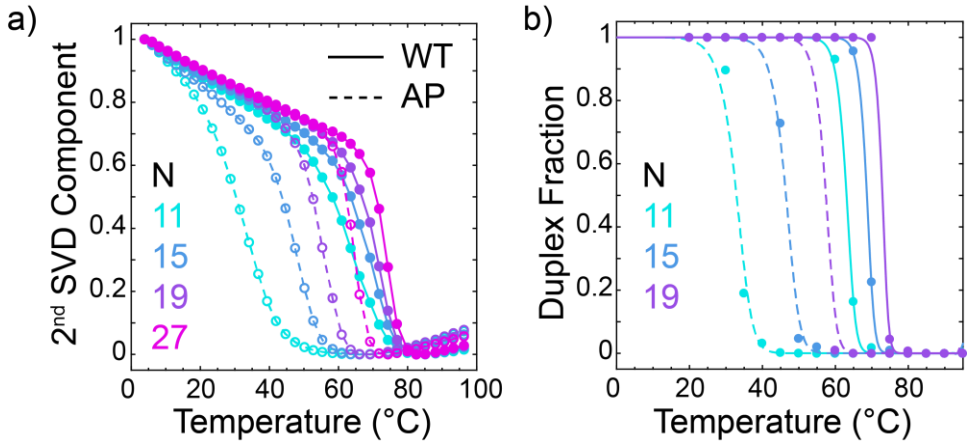

Figure S31 – Length-dependent melting curves from FTIR measurements and 3SPN.2 simulations. (a) Normalized 2nd SVD components from FTIR temperature series (Fig. S30) for N = 11, 15, 19, and 27 sequences shown in Fig. 4a. **(b)** Temperature-dependent fraction of intact duplexes from 3SPN.2 simulations employing WTMetaD for N = 11, 15, and 19 sequences. Determination of duplex fraction is presented in Section S2.3. Solid lines correspond to fits to two-state thermodynamic model (Section S1.3).

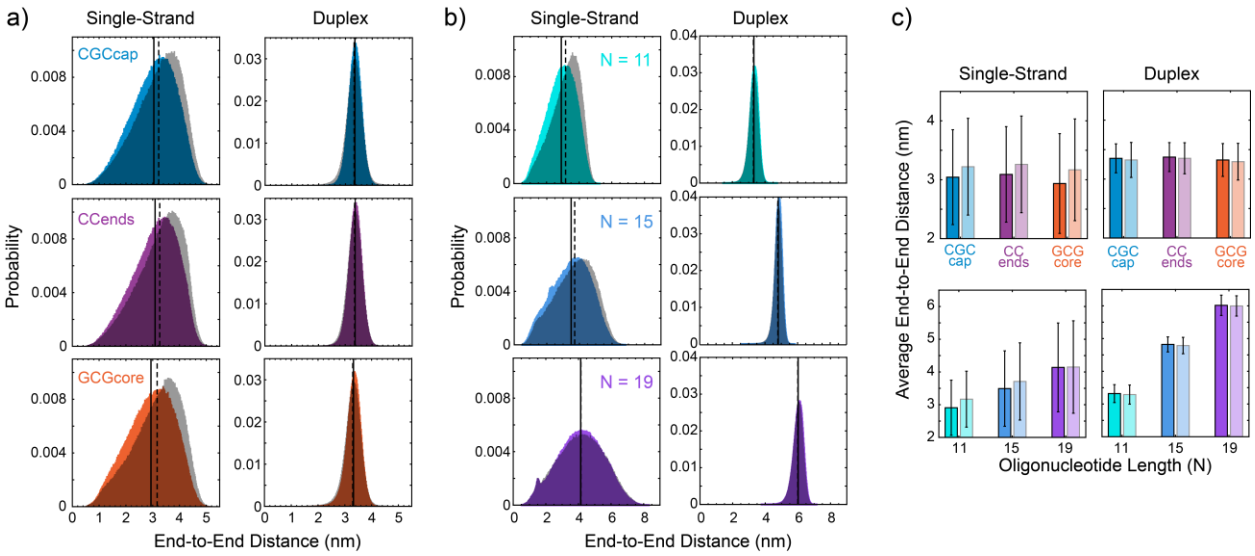

Figure S32 – End-to-end distance of single-strand and duplex DNA from 3SPN.2 simulations. **(a)** Distribution of (left) single-strand and (right) duplex end-to-end distances from 8×100 ns of unbiased 3SPN.2 MD simulations of each state at 300 K for CGCcap, CCends, and GCGcore sequences. Single-strand simulations were performed with only the AP-site-containing oligonucleotide. WT distributions are shown as solid colors and the overlaid black transparent distribution is that for the corresponding AP6 sequence. The solid and dashed black lines indicate the mean of the WT and AP6 distributions, respectively. **(b)** End-to-end distance distributions for variable length GCGcore-like sequences (N = 11, 15, 19) shown in Fig. 4. **(c)** Average end-to-end distance from distributions in **(a)** and **(b)**. Error bars correspond to standard deviation from distributions.

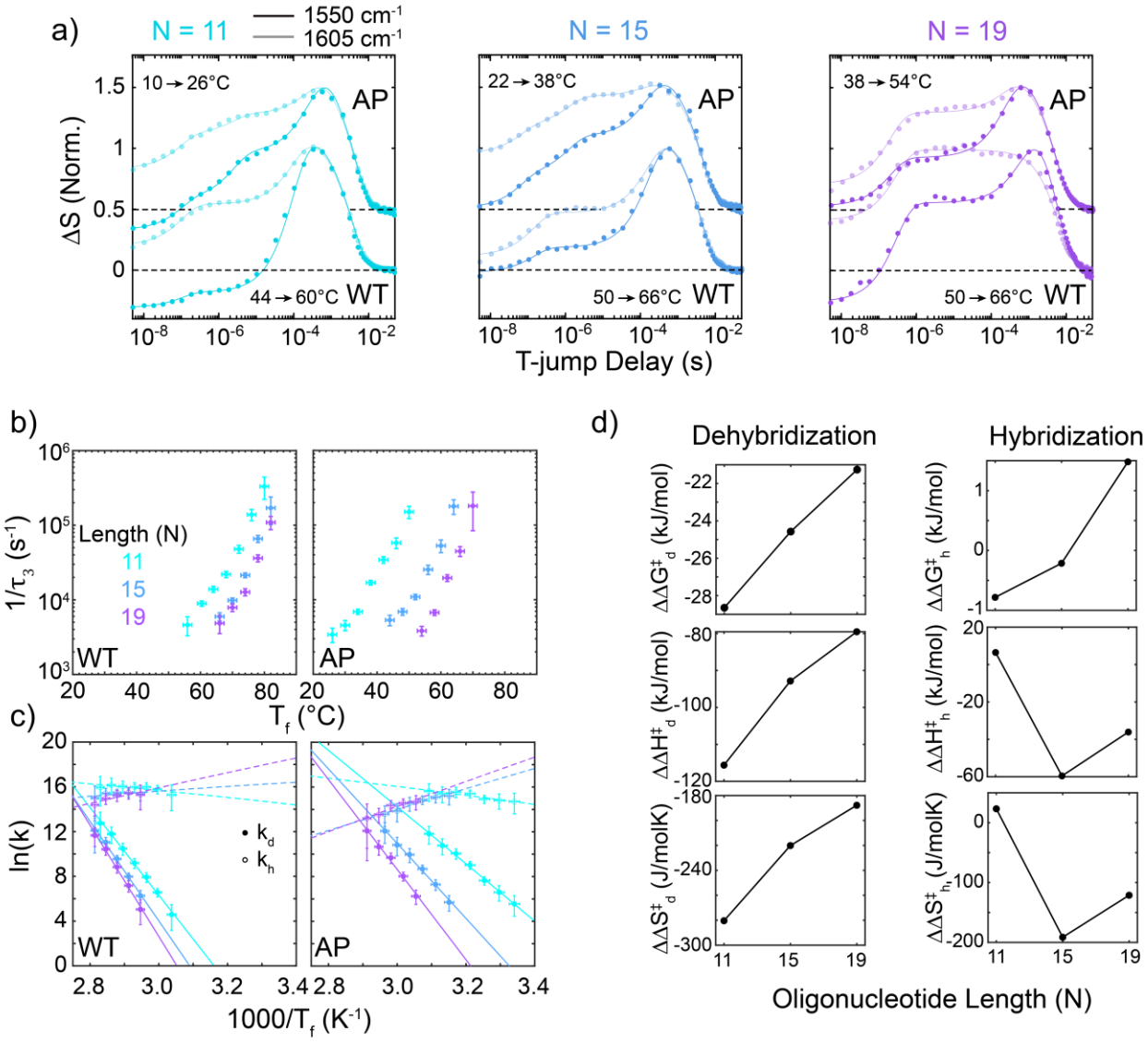

Figure S33 – Length-dependent impact of AP-site on dehybridization and hybridization kinetics. (a) Normalized low-temperature (indicated on plot) T-jump IR responses at 1550 (dark) and 1605 cm^{-1} (light) for GCGcore-like N = 11, 15, and 19 (left) WT and (right) AP sequences. **(b)** Slowest observed relaxation rates ($1/\tau_3$) for WT (opaque) and AP6 (transparent) sequences obtained from global lifetime fitting of t-HDVE data (Section S3.2). Vertical error bars indicate 95% confidence intervals from global fits and horizontal error bars correspond to the measured standard deviation in T-jump magnitude. **(c)** Dehybridization (k_d, filled circles) and hybridization (k_h, open circles) rate constants extracted from two-state modelling of observed rates (Section S3.5). Temperature-trends are fit to a Kramers-like equation in the high-friction limit (Eq. S27), solid lines for WT, dashed lines for AP). **(d)** Dehybridization free energy barrier ($\Delta G^{\ddagger}_{d,37}$), dehybridization enthalpy barrier ($\Delta H^{\ddagger}_{d}$), dehybridization entropic barrier ($\Delta S^{\ddagger}_{d}$), hybridization free energy barrier ($\Delta G^{\ddagger}_{h,37}$), hybridization enthalpy barrier ($\Delta H^{\ddagger}_{h}$), and hybridization entropy barrier ($\Delta S^{\ddagger}_{h}$) determined from Kramers fit in **(c)** for WT (black) and AP (black) sequences. Differences between AP and WT values are also shown.

S6. Hybridization nucleation behavior from 3SPN.2 simulations

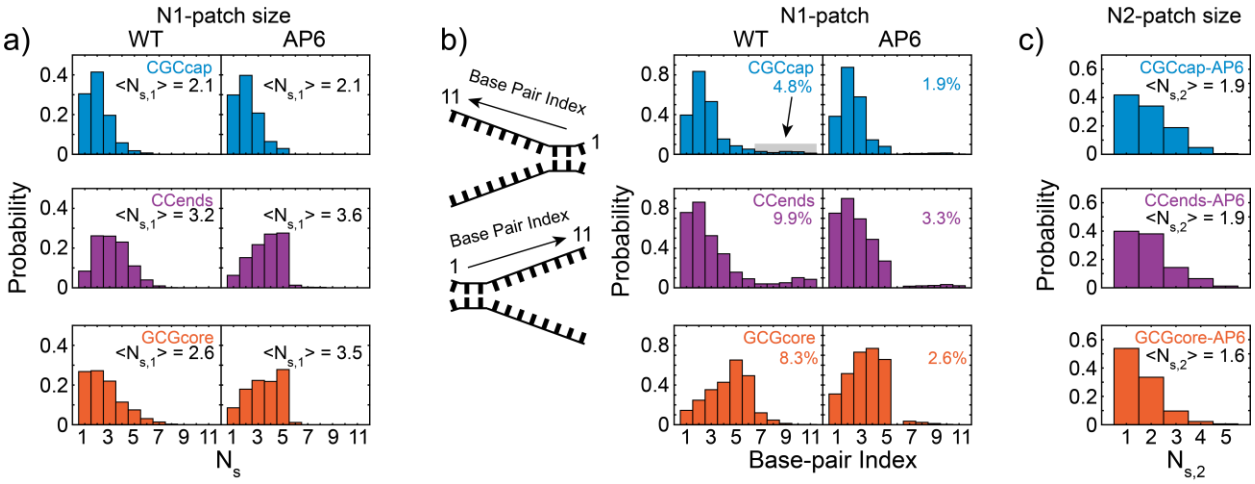

Figure S34 – Nucleation site and patch properties from 3SPN.2 MD hybridization trajectories. (a) Probability distribution for N1-patch size ($N_{s,1}$), which corresponds to the number of intact base-pairs at t_{N1}, determined with a base-pair separation cutoff of 0.7 nm. The average value of $N_{s,1}$, $\langle N_{s,1} \rangle$, is listed for each sequence and tends to increase in the presence of the AP-site depending on the sequence. **(b)** Probability that a base-pair is involved in the N1-patch determined from 1000-25000 hybridization trajectories for each sequence. Base-pair index refers to the distance (in number of base-pairs) from the terminal base-pair on the duplex half where N1-nucleation occurs (illustrated on the left). The sum over base-pair index 7-11 is shown for each sequence. **(c)** Probability distribution for N2-patch size ($N_{s,2}$) in AP6 sequences, which corresponds to the number of intact base-pairs at t_{N2} on the duplex half unbound at t_{N1} (base-pair index 7-11) The average value of $N_{s,2}$, $\langle N_{s,2} \rangle$, is listed for each sequence and is smaller for GCGcore-AP6 than CGCcap-AP6 and CCends-AP6.

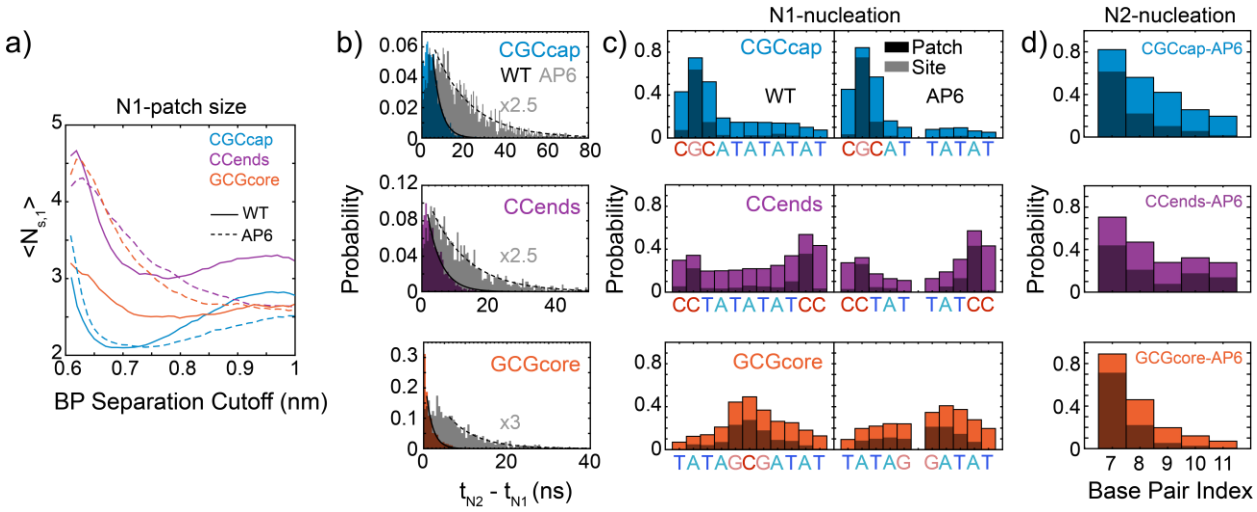

Figure S35 – Impact of base pair separation cutoff on hybridization nucleation behavior observed in 3SPN.2 simulations. (a) The average number of base pairs in the N1-patch $\langle N_{s,1} \rangle$ as a function of base pair separation cutoff length for WT (solid lines) and AP6 (dashed lines) sequences. **(b)** Probability distribution of time difference between N2- and N1-nucleation (t_{N2} - t_{N1}) determined using a base-pair separation cutoff of 1.0 nm. Shaded distributions correspond to AP6 sequences and their probability is multiplied by a factor of 2.5-3 for clarity. Distribution tails are fit to an exponential decay for WT (solid) and AP6 (dashed) sequences. **(c)** Probability that a base-pair site is part of the N1-patch determined using a base pair separation cutoff of 1.0 nm. **(d)** Probability for a base-pair being part of the N2-patch. Base pair index refers to the distance (in number of base pairs) from the terminal base pair on the duplex half where N1-nucleation occurs (Fig. S34).

Supplementary References

1. SantaLucia J, A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Nat. Acad. Sci.* **95**,1460-1465 (1998).
2. Freire E, Schön A, & Velazquez-Campoy A, Isothermal titration calorimetry: general formalism using binding polynomials. *Meth. Enzymol.* **455**,127-155 (2009).
3. Mergny J-L & Lacroix L, Analysis of thermal melting curves. *Oligonucleotides* **13**,515-537 (2003).
4. Marky LA & Breslauer KJ, Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* **26**,1601-1620 (1987).
5. Mikulecky PJ & Feig AL, Heat capacity changes associated with DNA duplex formation: salt-and sequence-dependent effects. *Biochem.* **45**,604-616 (2006).
6. Lang BE & Schwarz FP, Thermodynamic dependence of DNA/DNA and DNA/RNA hybridization reactions on temperature and ionic strength. *Biophys. Chem.* **131**,96-104 (2007).
7. Holbrook JA, Capp MW, Saecker RM, & Record MT, Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: interpretation in terms of coupled processes of formation and association of single-stranded helices. *Biochem.* **38**,8409-8422 (1999).
8. Hadži S & Lah J, Origin of heat capacity increment in DNA folding: The hydration effect. *Biochim. Biophys. Acta* **1865**,129774 (2021).
9. Hughesman CB, Turner RF, & Haynes CA, Role of the heat capacity change in understanding and modeling melting thermodynamics of complementary duplexes containing standard and nucleobase-modified LNA. *Biochemistry* **50**,5354-5368 (2011).
10. Dragan A, Privalov P, & Crane-Robinson C, Thermodynamics of DNA: heat capacity changes on duplex unfolding. *Eur. Biophys. J.* **48**,773-779 (2019).
11. Owczarzy R, Moreira BG, You Y, Behlke MA, & Walder JA, Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochem.* **47**,5336-5353 (2008).
12. Vesnaver G, Chang C-N, Eisenberg M, Grollman AP, & Breslauer KJ, Influence of abasic and anucleosidic sites on the stability, conformation, and melting behavior of a DNA duplex: correlations of thermodynamic and structural data. *Proc. Nat. Acad. Sci.* **86**,3614-3618 (1989).
13. Barsky D, Foloppe N, Ahmadia S, Wilson III DM, & MacKerell Jr AD, New insights into the structure of abasic DNA from molecular dynamics simulations. *Nucleic Acids Res.* **28**,2613-2626 (2000).
14. Chen J, Dupradeau F-Y, Case DA, Turner CJ, & Stubbe J, DNA oligonucleotides with A, T, G or C opposite an abasic site: structure and dynamics. *Nucleic Acids Res.* **36**,253-262 (2008).
15. Barducci A, Bussi G, & Parrinello M, Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**,020603 (2008).
16. Pfaendtner J (2019) Metadynamics to Enhance Sampling in Biomolecular Simulations. *Biomolecular Simulations*, (Springer), pp 179-200.
17. Bussi G & Laio A, Using metadynamics to explore complex free-energy landscapes. *Nature Reviews Physics* **2**,200-212 (2020).
18. Hinckley DM, Freeman GS, Whitmer JK, & De Pablo JJ, An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **139**,10B604_601 (2013).
19. Pfaendtner J & Bonomi M, Efficient sampling of high-dimensional free-energy landscapes with parallel bias metadynamics. *J. Chem. Theory Comput.* **11**,5062-5067 (2015).
20. Tribello GA, Bonomi M, Branduardi D, Camilloni C, & Bussi G, PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185**,604-613 (2014).
21. Ouldridge TE, Louis AA, & Doye JP, Extracting bulk properties of self-assembling systems from small simulations. *J. Phys. Condens.* **22**,104102 (2010).

22. Banyay M, Sarkar M, & Gräslund A, A library of IR bands of nucleic acids in solution. *Biophys. Chem.* **104**,477-488 (2003).
23. Peng CS, Jones KC, & Tokmakoff A, Anharmonic vibrational modes of nucleic acid bases revealed by 2D IR spectroscopy. *J. Am. Chem. Soc.* **133**,15650-15660 (2011).
24. Sanstead PJ, Stevenson P, & Tokmakoff A, Sequence-dependent mechanism of DNA oligonucleotide dehybridization resolved through infrared spectroscopy. *J. Am. Chem. Soc.* **138**,11792-11801 (2016).
25. Hithell G, *et al.*, Ultrafast 2D-IR and optical Kerr effect spectroscopy reveal the impact of duplex melting on the structural dynamics of DNA. *Phys. Chem. Chem. Phys.* **19**,10333-10342 (2017).
26. Krummel AT & Zanni MT, DNA vibrational coupling revealed with two-dimensional infrared spectroscopy: Insight into why vibrational spectroscopy is sensitive to DNA structure. *J. Phys. Chem. B* **110**,13991-14000 (2006).
27. Lee C, Park K-H, Kim J-A, Hahn S, & Cho M, Vibrational dynamics of DNA. III. Molecular dynamics simulations of DNA in water and theoretical calculations of the two-dimensional vibrational spectra. *J. Chem. Phys.* **125**,114510 (2006).
28. Jones KC, Ganim Z, & Tokmakoff A, Heterodyne-detected dispersed vibrational echo spectroscopy. *J. Phys. Chem. A* **113**,14060-14066 (2009).
29. Jones KC, Ganim Z, Peng CS, & Tokmakoff A, Transient two-dimensional spectroscopy with linear absorption corrections applied to temperature-jump two-dimensional infrared. *J. Opt. Soc. Am. B* **29**,118-129 (2012).
30. Sanstead PJ & Tokmakoff A, Direct observation of activated kinetics and downhill dynamics in DNA dehybridization. *J. Phys. Chem. B* **122**,3088-3100 (2018).
31. Sanstead PJ, Ashwood B, Dai Q, He C, & Tokmakoff A, Oxidized derivatives of 5-methylcytosine alter the stability and dehybridization dynamics of duplex DNA. *J. Phys. Chem. B* **124**,1160-1174 (2020).
32. van Stokkum IH, Larsen DS, & Van Grondelle R, Global and target analysis of time-resolved spectra. *Biochim. Biophys. Acta* **1657**,82-104 (2004).
33. Mullen KM, Vengris M, & van Stokkum IH, Algorithms for separable nonlinear least squares with application to modelling time-resolved spectra. *J. Glob. Optim.* **38**,201-213 (2007).
34. Dale J, *et al.*, Combining steady state and temperature jump IR spectroscopy to investigate the allosteric effects of ligand binding to dsDNA. *Phys. Chem. Chem. Phys.* **23**,15352-15363 (2021).
35. Fritzsch R, *et al.*, Monitoring base-specific dynamics during melting of DNA–ligand complexes using temperature-jump time-resolved infrared spectroscopy. *J. Phys. Chem. B* **123**,6188-6199 (2019).
36. Jones MS, Ashwood B, Tokmakoff A, & Ferguson AL, Determining sequence-dependent DNA oligonucleotide hybridization and dehybridization mechanisms using coarse-grained molecular simulation, Markov state models, and infrared spectroscopy. *J. Am. Chem. Soc.* **143**,17395-17411 (2021).
37. Kumar AT, Zhu L, Christian J, Demidov AA, & Champion PM, On the rate distribution analysis of kinetic data using the maximum entropy method: Applications to myoglobin relaxation on the nanosecond and femtosecond timescales. *J. Phys. Chem. B* **105**,7847-7856 (2001).
38. Araque J & Robert M, Lattice model of oligonucleotide hybridization in solution. II. Specificity and cooperativity. *J. Chem. Phys.* **144**,125101 (2016).
39. Bernasconi C *Relaxation kinetics* (Academic Press, Inc., 1976).
40. Cho C, Urquidi J, Singh S, & Robinson GW, Thermal offset viscosities of liquid H2O, D2O, and T2O. *J. Phys. Chem. B* **103**,1991-1994 (1999).
41. Neupane K, Hoffer NQ, & Woodside M, Measuring the local velocity along transition paths during the folding of single biological molecules. *Phys. Rev. Lett.* **121**,018102 (2018).
42. Dupuis NF, Holmstrom ED, & Nesbitt DJ, Single-molecule kinetics reveal cation-promoted DNA duplex formation through ordering of single-stranded helices. *Biophys. J.* **105**,756-766 (2013).

43. Williams AP, Longfellow CE, Freier SM, Kierzek R, & Turner DH, Laser temperature-jump, spectroscopic, and thermodynamic study of salt effects on duplex formation by dGCATGC. *Biochem.* **28**,4283-4291 (1989).
44. Chen C, Wang W, Wang Z, Wei F, & Zhao XS, Influence of secondary structure on kinetics and reaction mechanism of DNA hybridization. *Nucleic Acids Res.* **35**,2875-2884 (2007).
45. Craig ME, Crothers DM, & Doty P, Relaxation kinetics of dimer formation by self complementary oligonucleotides. *J. Mol. Biol.* **62**,383-401 (1971).
46. Carrillo-Nava E, Mejía-Radillo Y, & Hinz H-Jr, Dodecamer DNA duplex formation is characterized by second-order kinetics, positive activation energies, and a dependence on sequence and Mg2+ ion concentration. *Biochem.* **47**,13153-13157 (2008).
47. Pande VS, Beauchamp K, & Bowman GR, Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **52**,99-105 (2010).
48. Husic BE & Pande VS, Markov state models: From an art to a science. *J. Am. Chem. Soc.* **140**,2386-2396 (2018).
49. Wehmeyer C, *et al.*, Introduction to Markov state modeling with the PyEMMA software. *Living Journal of Computational Molecular Science* **1**,5965 (2019).
50. Sidky H, Chen W, & Ferguson AL, High-resolution Markov state models for the dynamics of Trp-cage miniprotein constructed over slow folding modes identified by state-free reversible VAMPnets. *J. Phys. Chem. B* **123**,7999-8009 (2019).
51. Scherer MK, *et al.*, PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *J. Chem. Theory Comput.* **11**,5525-5542 (2015).
52. McGibbon RT, *et al.*, MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **109**,1528-1532 (2015).
53. Noé F & Nuske F, A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Model. Simul.* **11**,635-655 (2013).
54. Scherer MK, *et al.*, Variational selection of features for molecular kinetics. *J. Chem. Phys.* **150**,194108 (2019).
55. Wu H & Noé F, Variational approach for learning Markov processes from time series data. *J. Nonlinear Sci.* **30**,23-66 (2020).
56. Chen W, Sidky H, & Ferguson AL, Nonlinear discovery of slow molecular modes using state-free reversible VAMPnets. *J. Chem. Phys.* **150**,214114 (2019).
57. Noé F, *et al.*, Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments. *Proc. Nat. Acad. Sci.* **108**,4822-4827 (2011).