

Molecular insight into how the position of an abasic site modifies DNA duplex stability and dynamics

Brennan Ashwood,^{1,†} Michael S. Jones,^{2,†} Yumin Lee,¹ Joseph R. Sachleben,³
Andrew L. Ferguson,^{2,*} Andrei Tokmakoff^{1,*}

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

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

³Biomolecular NMR Core Facility, Biological Sciences Division, The University of Chicago, Chicago, IL 60637, United States

[†]These authors contributed equally

*Correspondence: andrewferguson@uchicago.edu, tokmakoff@uchicago.edu

Abstract

Local perturbations to DNA base-pairing stability from lesions and chemical modifications can alter the stability and dynamics of an entire oligonucleotide. End effects may cause the position of a disruption within a short duplex to influence duplex stability and structural dynamics, yet this aspect of nucleic acid modifications is often overlooked. We investigate how the position of an abasic site (AP site) impacts the stability and dynamics of short DNA duplexes. Using a combination of steady-state and time-resolved spectroscopy and molecular dynamics simulations, we unravel an interplay between AP-site position and nucleobase sequence that controls energetic and dynamic disruption to the duplex. The duplex is disrupted into two segments by an entropic barrier for base pairing on each side of the AP site. The barrier induces fraying of the short segment when an AP site is near the termini. Shifting the AP site inward promotes a transition from short-segment fraying to fully encompassing the barrier into the thermodynamics of hybridization, leading to further destabilization of the duplex. Nucleobase sequence determines the length scale for this transition by tuning the barrier height and base-pair stability of the short segment, and certain sequences enable out-of-register base pairing to minimize the barrier height.

Significance

Whether in biology or technological applications, functions of short nucleic acids rely on tunable base-pair stability and dynamics. The continuous stretch of stable base-pairing and stacking interactions in canonical nucleic acids promotes duplex stability and cooperative hybridization between single strands where base pairs form in a concerted manner. Epigenetic modifications and lesions can locally destabilize base pairing and stacking, which may disrupt the cooperativity in oligonucleotides. The degree of thermodynamic and dynamic disruption also depends on the modification position within the duplex. Using time-resolved spectroscopy and computational modeling, our investigation reveals how the position of an abasic site tunes the disruption of base-pairing stability and kinetics in oligonucleotide duplexes.

Introduction

Design of nucleic acid technologies such as microarrays for sequencing and gene expression profiling, DNA points accumulation for imaging in nanoscale topography (DNA-PAINT), dynamic DNA devices, and CRISPR-Cas systems for gene editing rely on a predictive understanding of duplex hybridization and base pairing in oligonucleotides.(1-4) Most important for these areas is a fundamental understanding of hybridization thermodynamics, kinetics, and dynamics as a function of nucleobase sequence and for non-canonical base-pairing interactions. Although molecular interpretations for sequence-dependent duplex stability are still developing,(5, 6) nearest-neighbor (NN) models provide an empirical yet quantitative prediction of duplex binding free energy,(7, 8) and efforts have been undertaken to predict hybridization kinetics.(9-12) Numerous types of nucleobase chemical modifications and lesions are also known to influence the stability of base pairing,(13-15) and these effects are often quantified as a function of local nucleobase sequence around the modified site in NN models.(16-18) Additionally, end effects are likely to cause the position of a modification within a short oligonucleotide to influence the binding stability. Base-pair mismatches have been reported to be more destabilizing when located progressively inward from the duplex termini,(19-23) and this behavior may be general for any destabilizing modification. Such position dependence likely results from an interplay of position-dependent entropic penalties and disruption of local base-pairing dynamics, but details of the underlying molecular mechanisms at play are not yet resolved.

Formation of an abasic site (apurinic/apyrimidinic or AP site) results from the loss of a nucleobase through spontaneous or enzymatic cleavage of the glycosidic bond and is one of the most naturally abundant DNA lesions.(24, 25) Relative to mismatches, which may exhibit an array of non-canonical base-pairing configurations and structural deformations depending on their sequence context,(26-28) AP sites exhibit minimal perturbation to DNA duplex structure.(29, 30) Therefore, AP sites can be used to probe positional effects on duplex stability and base-pairing dynamics with minimal complexity added from non-canonical base pairing interactions. Previous work showed that duplex destabilization arising from an AP site at the center of an oligonucleotide strongly depends on the identity of the adjacent base pairs.(31-33) We recently demonstrated that the degree of destabilization from a central AP site depends on the full oligonucleotide sequence in addition to the adjacent base pairs, a consequence arising from a sequence-dependent free-energy penalty for nucleating base pairs on each side of the AP site.(34) However, it remains unclear how the position of the AP site will tune this penalty, the overall duplex destabilization, and the dynamics of base pairing in the oligonucleotide. These position-dependent dynamics have implications for the repair of AP sites and their mutagenic properties, such as in the stalling of transcription.(35)

We investigate the impact of shifting the AP-site position on the disruption of base-pairing stability and dynamics in multiple 11-mer template oligonucleotides with variable arrangement of guanine:cytosine (G:C) and adenine:thymine (A:T) base pairs. Temperature-dependent infrared (IR) and ^1H nuclear magnetic resonance (NMR) spectroscopy demonstrate that AP sites increasingly disfavor duplex hybridization as they move inward from the terminus. A cooperative helix-coil model and molecular dynamics (MD) simulations employing the 3-site-per-nucleotide (3SPN.2) coarse-grained model(36, 37) reveal that the position-dependent destabilization stems from a nucleation penalty for base pairing on each side of the AP site, which leads to a transition from multi-base-pair fraying when an AP site is near the termini to metastable half-dehybridization with central AP sites. The presence of these metastable partially-dehybridized duplex segments is directly resolved through temperature-jump IR spectroscopy (T-jump IR). Sequence-specific effects complicate the generality of these observations. For instance, ^1H NMR spectroscopy and all-atom MD simulations indicate that certain segment sequences may circumvent this nucleation penalty by forming out-of-register base pairs. Experimental results can consistently be interpreted

with 3SPN.2 MD simulations to provide detailed insight into the mechanism by which sequence and the position of an AP site alter the free-energy landscape for duplex hybridization.

Materials and Methods

Oligonucleotide preparation

Unmodified and modified DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT) at desalt-grade purity. AP sites were incorporated as a tetrahydrofuran group (dSpacer) rather than a deoxyribose group, removing complications from the chemical reactivity of the 1'-hydroxyl group. The tetrahydrofuran group likely exhibits a distinct distribution of configurations from a deoxyribose group, specifically in the potential for the group to rotate extrahelical.⁽²⁹⁾ We expect these differences to have a minor impact on duplex stability and kinetics as shown previously for other AP-site derivatives.⁽³¹⁾

Oligonucleotide samples were further desalted using 3 kDa centrifugal filters (Amicon) or dialyzed in ~1.5 L ultrapure water for 24 h using Slide-A-Lyzer cassettes (2kDa cutoff, Thermo Scientific). Oligonucleotides were prepared in pH* 6.8 sodium phosphate buffer (SPB, 200 mM dibasic and 200 mM monobasic, $[\text{Na}^+] = 600 \text{ mM}$) for all measurements. Samples were prepared in deuterated SPB, lyophilized to dryness, and re-dissolved in D₂O for Fourier-transform IR spectroscopy (FTIR) and T-jump IR spectroscopy to avoid spectral interference with the H₂O bending vibrational band. Oligonucleotide concentration was verified with UV absorbance using a NanoDrop UV/Vis spectrometer (Thermo Scientific).

Temperature-dependent FTIR spectroscopy

FTIR temperature series were collected using the same method as previously reported.⁽³⁸⁾ FTIR spectra were acquired using a Bruker Tensor FTIR spectrometer at 2 cm⁻¹ resolution. Samples were held between two 1 mm CaF₂ windows with a pathlength set by a 50 μm spacer. All measurements were performed with a 1:1 ratio of complementary strands and a total oligonucleotide concentration of 2 mM. To ensure oligonucleotides start in their minimum energy conformation, the solution of complementary strands was placed in a water bath at 90 °C for 3 min and cooled to room temperature under ambient conditions before each measurement. The temperature was ramped with a ~2.6 °C step size and equilibrated for 3 min at each step. A discrete wavelet transform using the Mallat algorithm and symlet family was applied to the 1490 – 1750

cm⁻¹ region of the FTIR spectra to separate and subtract the D₂O background absorption from the data.(39, 40)

Two-dimensional IR spectroscopy

Two-dimensional infrared spectroscopy (2D IR) temperature series were acquired using a previously described setup employing a pump-probe beam geometry.(41) Spectra were collected using parallel pulse polarization (ZZZZ) at a fixed waiting time (t_2) of 150 fs. The pump pulse pair delay (t_1) was scanned from -160 to 1900 fs in 16 fs steps. To monitor the full range of ring and carbonyl vibrations, the detection frequency axis was measured with approximately 6.2 cm⁻¹ resolution. Samples were prepared identically as for FTIR measurements. Sample temperature was controlled using a recirculating chiller (Ministat 125, Huber). Temperature series were performed with a ~2.8 °C step size and equilibrated for 4 min at each step.

Temperature-jump IR spectroscopy

We have previously described the temperature-jump (T-jump) IR spectrometer used in this work.(42, 43) Measurements were performed using the same sample conditions and sample cell as for FTIR measurements. The T-jump magnitude ($\Delta T = T_f - T_i$) was set to ~15 °C for all measurements as determined from the change in mid-IR solvent transmission following heating by the near-IR T-jump pulse. The T-jump induced changes in mid-IR vibrations of DNA are probed using heterodyned dispersed vibrational echo spectroscopy (HDVE)(44) with parallel (ZZZZ) pulse polarization and fixed at a waiting time (t_2) of 150 fs. The real part of the t-HDVE spectrum is reported and contains similar information to an IR pump-probe spectrum.

¹H NMR spectroscopy

¹H NMR measurements were performed on a Bruker AVANCE III 600 MHz spectrometer equipped with a Bruker TXI probe. Nuclear Overhauser effect spectroscopy (NOESY) experiments were performed with a mixing time (t_{mix}) of 200 ms unless noted otherwise. Measurements were performed in both 95% H₂O/5% D₂O and fully deuterated pH* 6.8 SPB solutions. 2048 and 1900 complex points were acquired in detection (t_2) and evolution (t_1) delays, respectively, over spectral widths of 24 ppm for 95% H₂O samples and 12 ppm for fully deuterated solutions. Solutions contained ~1 mM 3-(trimethylsilyl)propionic-2,2,3,3-d₄ acid (TMSP, Sigma-Aldrich) as a frequency reference.

Total correlation spectroscopy (TOCSY) temperature series measurements were measured with DIPSI II isotropic mixing and $t_{mix} = 20$ ms. The sample was equilibrated for 5 min and auto gradient and lock shimmed with TopShim at each temperature before acquiring spectra. 2048 and 512 complex points were acquired in t_2 and t_1 , respectively, over spectral widths of 12 ppm.

Coarse-grained molecular dynamics simulations

MD simulations were performed using the coarse-grained 3-Site-Per-Nucleotide (3SPN.2) model.(36) AP sites are incorporated by removing a nucleobase site as previously reported.(34) All systems were simulated using the large-scale atomic/molecular massively parallel simulator (LAMMPS) package(45) compiled with the 3SPN.2 model plugin.(36) Simulations were performed in the number, volume, and temperature (NVT) ensemble in a periodic box of side length 8.5 nm, equivalent to an effective 5.3 mM oligomer concentration. Classical mechanical equations of motion were integrated via Langevin dynamics in the scheme developed by Bussi and Parrinello(46) with a 20 fs integration time step. Solvent was treated implicitly with an experimentally informed friction coefficient of 9.94×10^{-11} m²/s.(36, 47) A 600 mM implicit salt concentration was used and electrostatic interactions were treated using the Debye-Hückel with a 5 nm cutoff radius.(48) All simulations were initialized from unmodified (WT) B-DNA duplex configurations generated by the 3SPN.2 software.

Enhanced sampling of hybridization free-energy landscape

We sampled the hybridization free-energy landscape at various temperatures using well-tempered metadynamics (WTMetaD) via the plugin for free-energy calculations with molecular dynamics (PLUMED),(49, 50) as described previously.(34) For each system, we selected a collective variable (CV) that described the average distance between the beads representing native Watson-Crick-Franklin (WCF) base pairs (r_{bp}). The CV was calculated by averaging over all available in-register pair distances – 11 for WT or 10 for AP sequences. Statistics were accumulated over the WTMetaD runs within the quasi-static regime where the applied bias was converged(51) and was analytically reweighted to report unbiased thermodynamic averages and free-energy landscapes in various order parameters beyond those in which sampling was conducted.

Markov state model temperature-jump simulations

Markov state models (MSMs)(52, 53) were constructed from 25 independent and unbiased 10 μ s 3SPN.2 MD simulations of each WT and AP system. MD simulations were performed near the sequence-dependent melting temperature calculated from the WTMetaD simulations ($T_{m,MD}$) and listed in Table S4. Initial velocities were assigned from a Maxwell-Boltzmann distribution at $T_{m,MD}$. Each simulation was conducted for 12 μ s and frames were saved to disc every 100 ps, resulting in 1-2 hybridization and dehybridization events per trajectory. The first 2 μ s of each run was discarded to produce $25 \times 10 \mu\text{s} = 250 \mu\text{s}$ (2.5 million frames) of simulation data for each sequence.

An additional 25 μ s MD simulation was performed for each system 15 °C below $T_{m,MD}$, and the data were projected into the microstates generated at $T_{m,MD}$ and used as the initial distribution for relaxation simulation with the MSM built at $T_{m,MD}$. Each microstate was scored in terms of the average number of intact A:T and G:C base pairs. Base pairs were assigned using a radial cutoff of 0.85 nm for G:C and 0.90 nm for A:T, and each nucleobase was restricted to one base pair at a time. Further details of MSM T-jump simulations were reported previously.(34)

All-atom molecular dynamics simulations

CGCcap-AP2 and CCend-AP2 DNA duplex topologies were constructed from an intact template of each sequence using AMBERTools22.(54) Only the first 5 base pairs of each sequence were used for all-atom simulation (CGCAT for CGCcap-AP2 and CCTAT CCend-AP2) to minimize unnecessary computational costs for the MD simulations. As described below, the last three residues of each 5-mer were constrained to their native state to reduce fraying and simulate the presence of a longer stabilizing duplex. The AP2 base atoms were removed and replaced with a hydrogen atom, and a small excess charge was redistributed to the 3' hydroxyl. The parmbsc1 force field was used with TIP3 water.(55, 56) A cubic simulation box with periodic boundaries and a 4.58 nm side length was used to maintain a 1 nm spacing between the DNA and box boundaries. NaCl ions were added to the box to maintain a 600 mM ionic concentration and to balance the negative charge from the 8 phosphate backbone groups.

Energy minimization was performed with the steepest descent algorithm to ensure a maximum force below $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. Consecutive equilibration simulations were performed in the NVT and number, pressure, and temperature (NPT) ensembles for 100 ps each. Production runs were performed for 1 μs in the NVT ensemble. Simulations were propagated with a 2 fs time step using a leap-frog integrator,(57) and frames were saved every 200 ps. The simulation temperature was maintained at 300 K using a velocity-rescaling thermostat.(58) The linear constraint solver (LINCS) algorithm was used to constrain hydrogen bonds,(59) and a Particle Mesh Ewald was used to calculate long-range electrostatic interactions.(60) The E root-mean-square deviation (eRMSD)(61) of the last three base pairs was calculated using PLUMED v2.8.(49, 62) eRMSD compares structures using only the relative distance and orientation of nucleobases, whereas the full structure RMSD often fails to differentiate between nucleic acid conformations with distinct base-pairing configurations. To ensure the last three base pairs retained their native configuration, the eRMSD was restrained by setting an upper wall at 0.6. For each sequence, 20 independent simulations were performed from the same initial state.

Trajectory analysis was performed using the MDTraj and PyEMMA Python libraries.(63, 64) Average hydrogen bond distances were calculated between the three hydrogen bonding atoms on the 5' terminal C1 (N4H2, N3, O2) and corresponding atoms (O6, H1, N2H1) on the 3' terminal G1. For the CCend-AP2 sequence, an out-of-register hydrogen bond distance was calculated between the same set of C1 atoms and the corresponding set on the internal G2. This second pairing represents a C:C mismatch for CGCcap-AP2; therefore, an analogous distance was calculated using the average distance of the available hydrogen bonding atoms on each residue.

Results & Discussion

Position-dependent destabilization of DNA duplex by an abasic site

Duplex destabilization depends on AP-site position

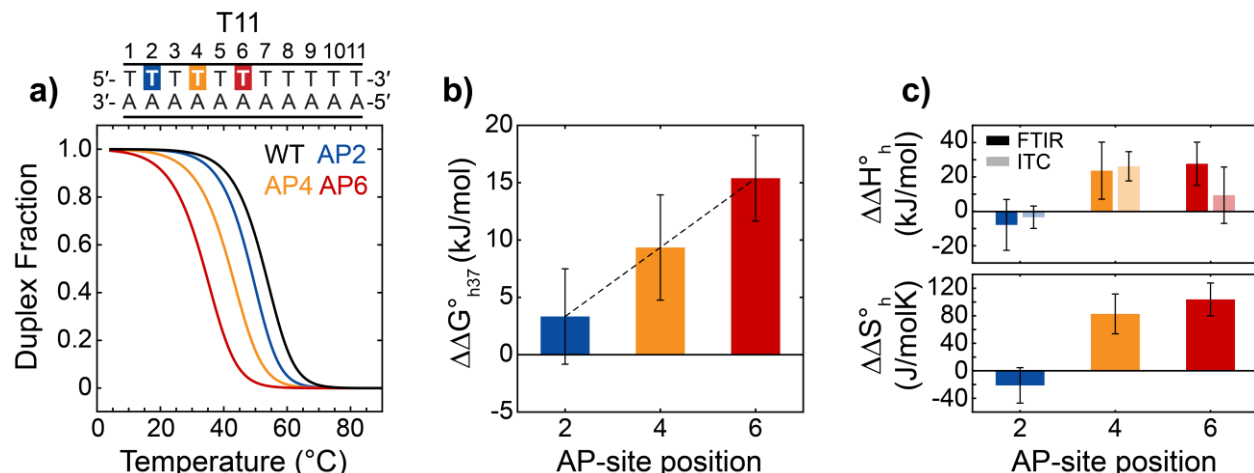


Figure 1. DNA duplex destabilization depends on the AP-site position. (a) Duplex melting curves for a homogeneous A:T sequence that is unmodified (WT, black) or contains an AP site at the 2 (AP2), 4 (AP4), or 6 (AP6) position. Melting curves were extracted from a two-state analysis of FTIR temperature series (Section S1.2). (b) Change in hybridization free energy at 37 °C for each AP sequence with respect to the WT sequence, $\Delta\Delta G_{h37}^\circ = \Delta G_{h37}^\circ - \Delta G_{h37}^\circ(WT)$. The dashed line corresponds to a linear fit with a slope of 3 kJ/mol per base-pair index. (c) Change in hybridization enthalpy ($\Delta\Delta H_h^\circ$, top) and hybridization entropy ($\Delta\Delta S_h^\circ$, top) with respect to the WT sequence. Light $\Delta\Delta H_h^\circ$ bars correspond to values measured with ITC in non-deuterated solution (Fig. S4). FTIR and ITC $\Delta\Delta H_h^\circ$ values were shifted to 25 °C using the change in heat capacity (ΔC_p) measured for T11-WT with ITC (Fig. S5). FTIR and ITC error bars correspond to 95% confidence intervals from two-state fits. An AP site is least destabilizing near the duplex termini and becomes increasingly destabilizing when moved inward.

An AP site destabilizes the duplex, and the magnitude of destabilization depends on the position of the AP site. We compare the thermodynamic impact of incorporating an AP site at the second (AP2), fourth (AP4), and central sixth (AP6) position relative to the unmodified form (WT) of a homogenous sequence 5'-TTTTTTTTTTT-3' + complement (T11). Duplex melting curves extracted from a two-state thermodynamic analysis of FTIR temperature series show that the melting transition midpoint (T_m) decreases by 5 to 20 °C as the AP site is shifted from the second to sixth position (Figs. 1a, S1-S3, and Table S1). The change in the hybridization free energy between WT and AP sequences ($\Delta\Delta G_{h37}^\circ = \Delta G_{h37}^\circ(AP) - \Delta G_{h37}^\circ(WT)$) increases linearly in 6 kJ/mol increments from AP2 to AP4 to AP6 (Fig. 1b). The change in hybridization van't Hoff

enthalpy ($\Delta\Delta H_h^\circ$) and entropy ($\Delta\Delta S_h^\circ$) also increase by ~ 45 kJ/mol and ~ 120 J/molK over this range, respectively. After correcting to an equivalent temperature of 25 °C, isothermal titration calorimetry (ITC, Section S1.3) gives similar $\Delta\Delta H_h^\circ$ values to FTIR for T11-AP2 and T11-AP4 but a smaller value for T11-AP6 (Fig. 1c).

Position-dependent destabilization arises from segment nucleation barrier

The position-dependent destabilization from an AP site may be understood through a statistical treatment of duplex hybridization. Helix-coil (HC) models provide one of the simplest statistical descriptions of cooperative DNA duplex melting thermodynamics.(65, 66) Base-pairing thermodynamics are described by two equilibrium constants: s , which is the Boltzmann factor of the free energy for adding an individual base pair next to an intact base pair (ΔG_{int}°) and σ , which is the Boltzmann factor of the free energy for nucleating a stretch of base pairs (ΔG_{nuc}°).

$$s = e^{-\Delta G_{int}^\circ / RT} \quad (1)$$

$$\sigma = e^{-\Delta G_{nuc}^\circ / RT} \quad (2)$$

Since this is a mean-field model, we apply the average value of s across all possible contacts within the duplex, $\langle s \rangle$, as determined from nearest-neighbor (NN) enthalpy (ΔH_h°) and entropy (ΔS_h°) parameters.(7) The large penalty to nucleation makes $\sigma \ll 1$, which is the source of cooperativity in the model. We assume ΔG_{nuc}° is purely entropic such that σ is independent of temperature.

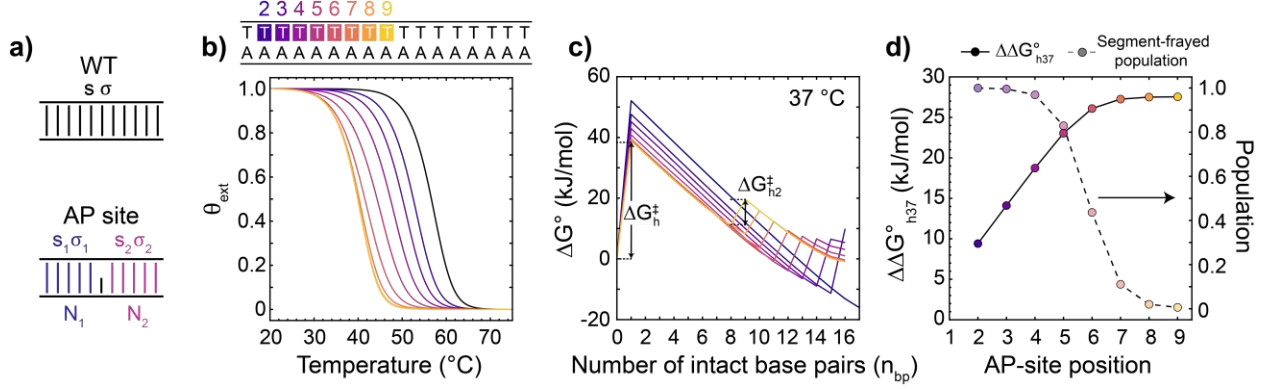


Figure 2. Helix-coil modeling of position-dependent duplex destabilization from an AP site. (a) Schematic of two-stretch helix-coil (HC) model used to model the impact of AP-site position on duplex stability. The WT duplex contains a single nucleation penalty (σ) and base-pair stability constant (s) applied uniformly to each base-pair site.(65, 66) An AP site splits the duplex into two base-pair stretches with their own s and σ values. (b) Fraction of duplexes containing one or more base pairs (θ_{ext}) for an unmodified T17:A17 sequence (black) and those containing an AP site at the 2 – 9 positions (purple to yellow). $\sigma_1 = \sigma_2 = 10^{-4}$ for all sequences. (c) Free-energy profiles at 37 °C for each sequence plotted as a function of the number of intact base pairs (n_{bp}), illustrating barriers for strand association ($\Delta G_h^\ddagger$) and nucleation of the second-segment ($\Delta G_{h2}^\ddagger$). (d) $\Delta \Delta G_{h37}^\circ$ (dark circles and solid line) and the corresponding fraction of duplexes with the short base-pair segment completely frayed (light circles and dashed line). Shifting the AP site inward from the terminus leads to an increase in the number of frayed base pairs and greater destabilization of the duplex. The degree of destabilization levels off once the binding stability of the short segment is large enough to overcome the nucleation penalty.

It is generally unfavorable to have multiple disconnected stretches of base pairs in short duplexes, so we describe the WT system with an internal partition function ($Z_{int,D}$) limited to a single stretch of base pairs.(65-67)

$$Z_{int,D} = 1 + \sigma \sum_{n_{bp}=1}^N (N - n_{bp} + 1) \langle s \rangle^{n_{bp}} \quad (3)$$

N is the maximum number of possible base pairs, and n_{bp} is the number of intact base pairs for a given microstate. As detailed in Section S2, we can extend the HC model to treat an AP site as a defect that splits the duplex into two stretches of N_1 and N_2 base pairs (Fig. 2a). The internal partition function for the duplex can then be written as a product of two partition functions similar to eq. 3 for the two stretches, $Z_{int,D} = Z_{int,D1} Z_{int,D2}$. Each stretch has its own nucleation penalty (σ_1, σ_2) and $\langle s \rangle$ value, which corresponds to the average base-pair equilibrium constant across the

respective stretch. Our usage of a second nucleation penalty is similar to incorporating a ‘defect’ free-energy penalty as in previous statistical models of duplexes containing a mismatch.(19)

To calculate melting curves from the HC model, we determine the fraction of oligonucleotide strands that contain one or more base pairs (θ_{ext}) as a function of temperature. θ_{ext} depends on both external and internal degrees of freedom of the system. We evaluate the external partition functions of the single-strand and duplex states as the number of possible ways of arranging single-strand and duplex molecules as a self-avoiding walk on a 3D cubic lattice,(68) allowing us to compute θ_{ext} in terms of $Z_{int,D}$.

$$\theta_{ext} = 1 + \frac{1 - \sqrt{1 + 2c_{tot}\gamma Z_{int,D}}}{c_{tot}\gamma Z_{int,D}} \quad (4)$$

$$\text{where } \gamma = 6c^\circ N_A V_{ss}$$

c° is the standard-state concentration of 1 M, N_A is Avogadro’s constant, V_{ss} is the volume of a single-strand molecule, and c_{tot} is the total concentration of oligonucleotides. We neglect the temperature-dependence of V_{ss} such that the temperature-dependence of θ_{ext} comes only from $Z_{int,D}$.

Figure 2b shows calculated θ_{ext} melting curves for an extended homogeneous sequence, T17:A17. Moving an AP site from the second to sixth base pair site leads to a gradual reduction in the duplex melting temperature, consistent with the experimental observation for T11. The shift in melting temperature becomes minor and eventually levels off as the AP site moves closer to the duplex center. This trend in the θ_{ext} curves may be understood by examining the free-energy profile (FEP) as a function of the number of intact base pairs (n_{bp}) calculated from the HC model (Fig. 2c). The WT sequence exhibits a single free-energy barrier to hybridization ($\Delta G_h^\ddagger$) between the single-strand ($n_{bp} = 0$) and duplex states that arises from the reduction in translation and configurational (set by σ) entropy upon binding of single strands, and formation of the remaining base pairs is cooperative and downhill in energy. In the HC model, the AP site disrupts this cooperativity and introduces an additional entropic barrier for forming base pairs on each side of the AP site ($\Delta G_{h2}^\ddagger$), and we recently experimentally and computationally verified the presence of $\Delta G_{h2}^\ddagger$ in duplexes with a central AP site.(34) The AP-site position controls the position of $\Delta G_{h2}^\ddagger$

along the hybridization FEP. When an AP site is at the second base-pair position, the adjacent terminal base pair does not have enough binding stability to overcome $\Delta G_{h2}^\ddagger$ and remains highly frayed at 37 °C. As the AP site shifts from the second to sixth base-pair position, the shorter duplex segment increases in length and base-pairing stability but remains highly frayed (Fig. 2c,d). The number of frayed base pairs increases with segment length, leading to the position-dependent decrease in duplex stability observed from the θ_{ext} curves and $\Delta\Delta G_{h37}^\circ$. Once the weak segment reaches a length of seven base pairs, the binding stability overcomes $\Delta G_{h2}^\ddagger$ and duplex destabilization levels off. $\Delta G_{h2}^\ddagger$ is fully encompassed into the duplex state at these AP-site positions, creating a local free-energy minimum along n_{bp} that corresponds to configurations with one segment dehybridized. The leveling off is also facilitated by a reduction in the magnitude of $\Delta G_{h2}^\ddagger$ from 14 to 9 kJ/mol as the AP site moves inward from the termini.

The HC calculations suggest that the position-dependent destabilization from an AP site depends only on the magnitude of the nucleation barrier and the binding stability of the short segment. An additional minor effect is that the number of base-pair arrangements possible over the whole duplex decreases as the AP site moves closer to the duplex, resulting in a reduction of base-pairing combinational entropy and a free-energy penalty of 1 – 2 kJ/mol (Fig. S6). Given the small magnitude of this effect, duplex destabilization is essentially only position-dependent when an AP site is close to the termini. This range is roughly ~1 – 9 base pairs from our model calculations but depends on the molecular properties of $\Delta G_{h2}^\ddagger$ and the short duplex segment.

Interplay of nucleobase sequence and AP-site position in destabilization of the duplex

Beyond the factors that influence a homogeneous duplex, destabilization by an AP site will depend on multiple aspects of the nucleobase sequence. Previous reports showed that $\Delta\Delta G_{h37}^\circ$ is highly sensitive to the identity of the nucleobase being removed as well as the bases adjacent to the AP site.(32, 33) Further, nucleobase sequence will influence the binding stability of the short duplex segment as well as the magnitude of $\Delta G_{h2}^\ddagger$. We use three additional 11-mers with distinct base-pairing properties to investigate sequence-dependent destabilization from an AP site (Fig. 3a). “CGCcap” contains three G:C base pairs on one end to create asymmetry in the duplex base-pairing stability, “CCends” places a pair of G:C base pairs at the termini to minimize base-pair fraying, and “GCGcore” has central G:C base pairs to promote fraying of the A:T termini.(38, 69)

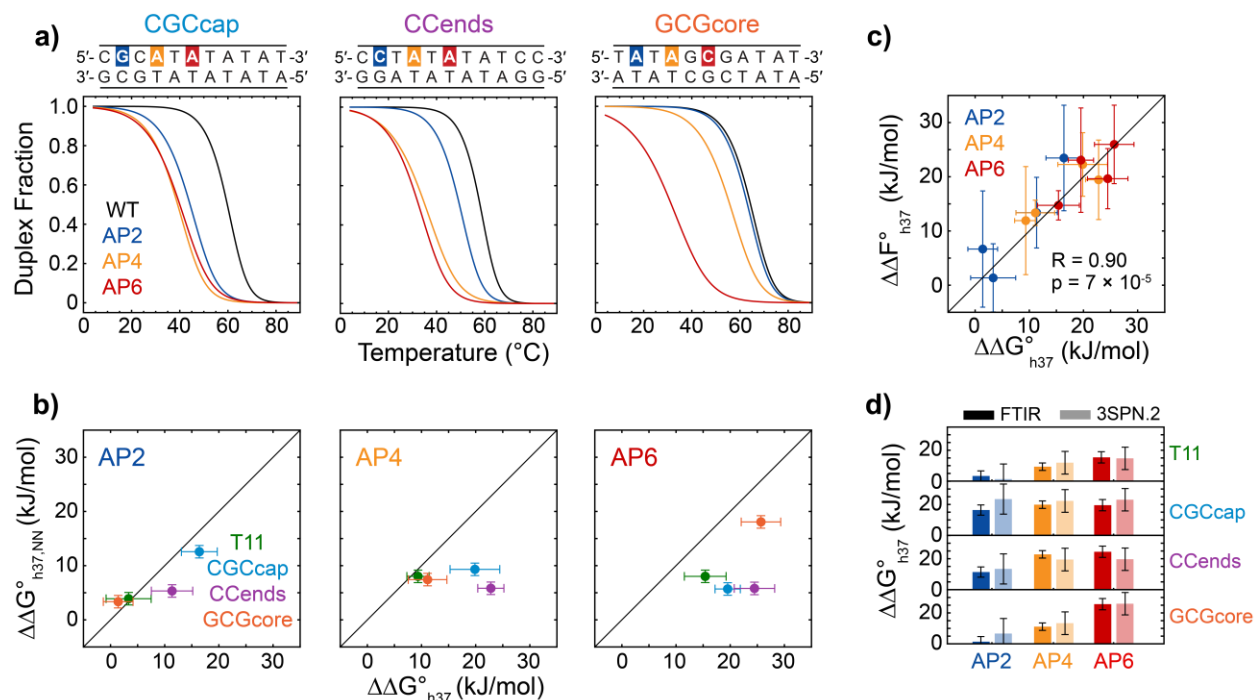


Figure 3. Sequence-dependent duplex destabilization from an AP site. (a) Duplex melting curves for CGCcap, CCends, and GCGcore sequences determined from a two-state analysis of FTIR temperature series. (b) Scatter plots of $\Delta\Delta G_{h37}^{\circ}$ from FTIR vs. nearest-neighbor (NN) model calculations of duplex free-energy change from an AP site ($\Delta\Delta G_{h37,NN}^{\circ}$) for all sequences. Plots are separated by AP-site position. The comparison indicates that the NN model systematically underestimates $\Delta\Delta G_{h37}^{\circ}$ and poorly predicts sequence-dependent destabilization from an AP site. (c) Scatter plot of $\Delta\Delta G_{h37}^{\circ}$ from FTIR vs. $\Delta\Delta F_{h37}^{\circ}$ from 3SPN.2-determined melting curves for all sequences with an AP site. The Pearson correlation coefficient (R) and p-value are listed. (d) Comparison of $\Delta\Delta G_{h37}^{\circ}$ (dark bars) and $\Delta\Delta F_{h37}^{\circ}$ (light bars) for each sequence. Error bars for $\Delta\Delta G_{h37}^{\circ}$ and $\Delta\Delta F_{h37}^{\circ}$ are propagated from two-state fits, and NN error bars correspond to the reported standard deviation from comparison with experimental data.(7)

An AP site is generally more destabilizing when shifted inward from the termini, but Fig. 3a illustrates how the relative placement of G:C and A:T base pairs significantly tunes this position-dependent trend. At identical position, incorporating an AP site at a G:C base pair is more destabilizing than at an A:T base pair, which is most notable for AP2 sites as observed in the comparison of $\Delta\Delta G_{h37}^{\circ}$ values extracted for CGCcap-AP2/CCends-AP2 and T11-AP2/GCGcore-AP2 (Fig. 3b). However, the removal of an interior A:T base pair can be equivalent or even more destabilizing than removing a near-terminal G:C base pair as found in comparing CCends-AP2 and CCends-AP6. For CGCcap, the strength of nearest-neighbor base-pairing and stacking interactions around the AP site decreases from AP2 to AP4 to AP6 modifications. The destabilization from disrupting the G:C-rich region in CGCcap-AP2 almost matches the

combination of destabilization from A:T disruption and incorporation of a nucleation penalty in CGCcap-AP4 and CGCcap-AP6, leading to near cancellation of the position-dependence in $\Delta\Delta G_{h37}^\circ$.

The AP-site position and its local sequence context alone do not fully capture the sequence-dependent trends in $\Delta\Delta G_{h37}^\circ$. This point is illustrated in Fig. 3b by comparing $\Delta\Delta G_{h37}^\circ$ values from FTIR and those calculated with Santa Lucia's NN model ($\Delta\Delta G_{h37,NN}^\circ$).⁽⁷⁾ $\Delta\Delta G_{h37,NN}^\circ$ values for AP4 and AP6 sequences are calculated by removing the two NN parameters associated with the AP site. For AP2 sequences, the 3' NN parameter is removed and the 5' parameter is halved to account for having just a single base-pair and stacking interaction at the terminal base. The NN model systematically underestimates the magnitude of $\Delta\Delta G_{h37}^\circ$ due to the lack of a free-energy penalty for base pairing on each side of the AP site. Further, NN effects do not predict the sequence-dependence of $\Delta\Delta G_{h37}^\circ$ at fixed AP-site position. For example, $\Delta\Delta G_{h37}^\circ$ spans 14 kJ/mol for the AP4 sequences yet $\Delta\Delta G_{h37,NN}^\circ$ values for these sequences fall within a ~ 3 kJ/mol window. In another example, CCends-AP6 and GCGcore-AP6 exhibit $\Delta\Delta G_{h37}^\circ$ values within 1 kJ/mol and their $\Delta\Delta G_{h37,NN}^\circ$ predictions are separated by 12 kJ/mol. These observations suggest the presence of additional sequence-dependent contributions to the free-energy penalty from an AP site.

Sequence-dependent nucleation barrier and stability of the weak segment

We next aim to evaluate how the properties of the segment-nucleation barrier and binding stability of the weak segment depend on nucleobase sequence. The HC model qualitatively captures the effect of sequence on duplex destabilization (Figs. S6 – S7), but it neglects numerous molecular factors that are necessary for an accurate description of the underlying base-pairing free-energy landscape. Therefore, we characterized the underlying hybridization free-energy landscape as a function of AP-site position using coarse-grained MD simulations. In particular, the threshold segment length for stable hybridization, its sequence-dependence, and the molecular behavior of few-base-pair segments were examined.

Nucleation barrier and stability of the weak base-pair segment

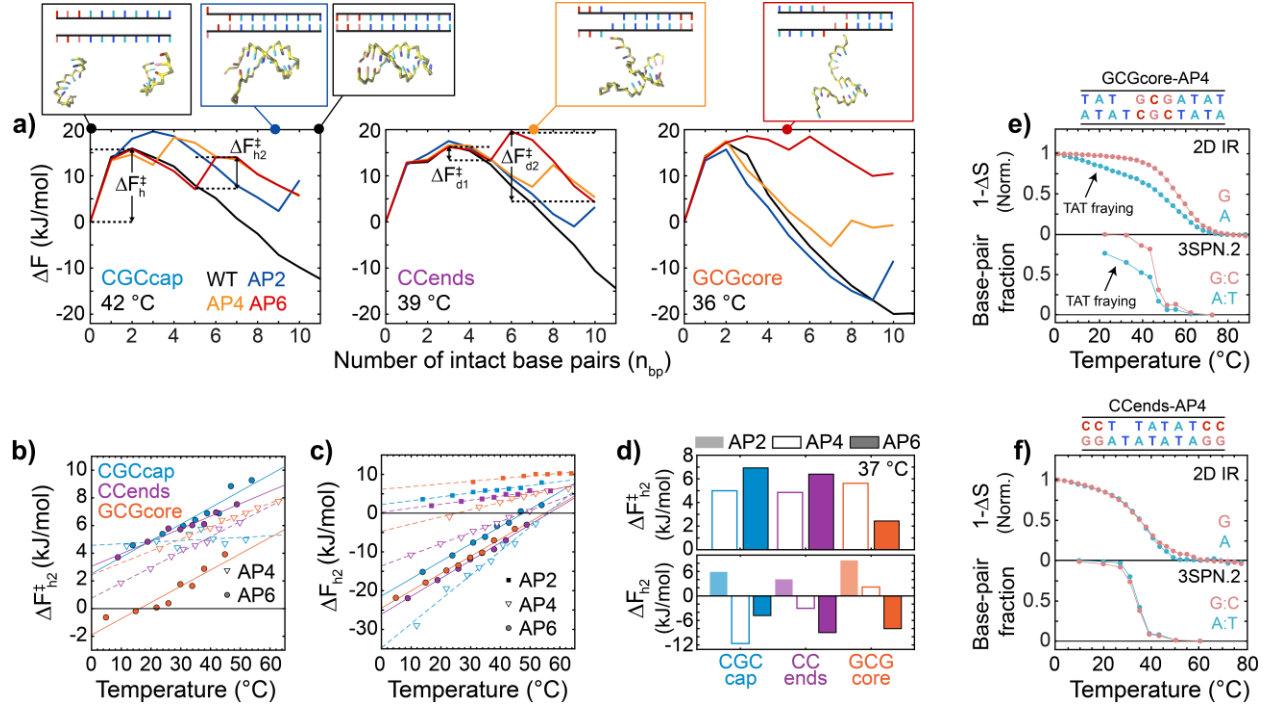


Figure 4. Binding stability of the weak segment. (a) Free-energy profiles as a function of the number of intact base pairs, $\Delta F(n_{bp}) = -RT \ln P(n_{bp})/P(0)$, for each sequence from 3SPN.2 MD simulations employing WTMetaD. Base pairs were assigned using a 0.7 nm radial separation cutoff. Simulations were carried out at temperatures 10-20 °C below the 3SPN.2-determined T_m ($T_{m,MD}$). AP2, AP4, and AP6 sequences show an additional hybridization barrier for $n_{bp} > 2$ for base pairing on both sides of the AP site ($\Delta F_{h2}^{\ddagger}$). (b) $\Delta F_{h2}^{\ddagger}$ for AP4 (triangles) and AP6 (circles) sequences as a function of temperature. (c) Free-energy difference between states with weak-segment frayed and intact, $\Delta F_{h2} = -RT \ln P(n > n_{sd})/P(n_{sd})$, for all sequences where n_{sd} is the number of base pairs in the segment-dehybridized state. Dark-solid, dashed, and light-solid lines correspond to linear fits for AP6, AP4, and AP2 sequences, respectively. Extracted internal-energy and entropy changes are provided in Fig. S16. (d) (top) $\Delta F_{h2}^{\ddagger}$ and (bottom) ΔF_{h2} at 37 °C for each sequence. 3SPN.2 MD simulations suggest that two to five base pairs, depending on the G:C content, provide enough stability for a segment to hybridize. (e) (top) Normalized temperature-dependent change in two-dimensional infrared (2D IR) signals relative to the initial temperature, $1 - \Delta S(T) = [S(T) - S(1^\circ\text{C})]/S(1^\circ\text{C})$, at guanine (G) and adenine (A) ring vibrational bands for GCGcore-AP4. Corresponding 2D IR spectra are shown in Fig. S20. (bottom) Fraction of intact G:C and A:T base pairs from 3SPN.2 MD simulations carried out at 8 – 9 temperatures across the duplex melting transition. Base pairs were assigned using a radial cutoff of 0.7 nm. (f) Same plot for CCends-AP4. Fraying of the TAT segment is experimentally observed in GCGcore-AP4 but not for the CCT segment in CCends-AP4, consistent with the sequence dependence of weak-segment stability predicted by the simulations.

We conducted MD simulations with the 3SPN.2 coarse-grained model employing well-tempered metadynamics (WTMetaD) to sample the hybridization free-energy landscape at 7 – 8

temperatures from $T_{m,MD} - 20\text{ }^{\circ}\text{C}$ to $T_{m,MD} + 20\text{ }^{\circ}\text{C}$, where $T_{m,MD}$ is the melting temperature determined with 3SPN.2. Incorporation of an AP site in 3SPN.2 has negligible impact on the B-DNA duplex structure, and the free nucleobase remains predominantly intrahelical (Fig. S8),(34) consistent with previous structural characterization of duplexes containing AP sites.(29) Two-state analysis of 3SPN.2-determined melting curves produce hybridization Helmholtz free energies (ΔF_{h37}° and $\Delta\Delta F_{h37}^{\circ}$) that are well-correlated ($R = 0.90$) with experimental values (Figs. 3cd & S9), indicating that 3SPN.2 reasonably predicts the sequence-dependent free-energy penalty from an AP site. In contrast, ΔH_h° and ΔS_h° are poorly captured by 3SPN.2, particularly for sequences containing an AP site (Figs. S9).

Figure 4 shows FEPs computed from the probability distribution along n_{bp} for CGCcap, CCends, and GCGcore sequences that qualitatively agree with those from the HC model (Figs. 2c & S7). WT sequences show a single free-energy barrier to hybridization ($\Delta F_h^{\ddagger}$) peaked at $n_{bp} = 2$ or 3 and hybridization is energetically downhill for $n_{bp} > 3$. AP sequences must overcome a second barrier ($\Delta F_{h2}^{\ddagger}$) to form a fully intact duplex. Just as for $\Delta G_{h2}^{\ddagger}$ from the HC model, $\Delta F_{h2}^{\ddagger}$ arises from the energy penalty for nucleating the second base-pair segment and creates a local free-energy minimum that corresponds to duplex configurations containing intact base pairs on only one side of the AP site (Figs. S10 and S11). $\Delta F_{h2}^{\ddagger}$ lies between $n_{bp} = 9$ and $n_{bp} = 10$ in AP2 sequences and is not observable along the n_{bp} coordinate, yet a barrier for forming the terminal base pair is observed as a function of average base-pair separation (r_{bp} , Figs. S12-S15). Both $\Delta F_h^{\ddagger}$ and $\Delta F_{h2}^{\ddagger}$ arise from entropic penalties for nucleation and are partially balanced by favorable enthalpy changes for base pairing (Figs. 4b & S16). $\Delta F_h^{\ddagger}$ is predominantly a reflection of the reduction in translational and orientational entropy upon bimolecular association, which are absent from $\Delta F_{h2}^{\ddagger}$. Instead, $\Delta F_{h2}^{\ddagger}$ is dominated by a reduction in conformational entropy of the unhybridized segment and is therefore 2-to-5-fold smaller than $\Delta F_h^{\ddagger}$ (Fig. S16). $\Delta F_{h2}^{\ddagger}$ is also sensitive to the sequence of nucleating base pairs adjacent to the AP site (Fig. 4d). $\Delta F_{h2}^{\ddagger}$ is smallest for GCGcore-AP6, particularly at low temperature, and involves formation of a G:C base pair adjacent to the AP site whereas $\Delta F_{h2}^{\ddagger}$ requires formation of one or two A:T base pairs on the second segment in the other sequences, leading to a larger barrier height.

The AP-site position and nucleobase sequence determine the position and magnitude of the $\Delta F_{h2}^\ddagger$ barrier along n_{bp} and the base-pairing stability of the second segment. The combination of these factors determines the free energy for hybridization of the second segment (ΔF_{h2}). The terminal base pair adjacent to the AP site in AP2 sequences has an insufficiently stabilizing potential energy for base pairing to overcome the unfavorable entropy, leading to a positive free-energy change for forming the last base pair ($\Delta F_{h2} > 0$, Fig. 4c). The threshold length of the weak segment to overcome the nucleation penalty and reach $\Delta F_{h2} < 0$ depends on the sequence of the segment. If the weak segment contains multiple G:C base pairs like in CCends-AP4, then three base pairs are sufficient to form a stable segment. In contrast, the pure A:T segment of GCGcore-AP4 is highly frayed at 37 °C and four to six base pairs are needed to form a stable segment as evidenced by the results for CGCcap-AP4 and CGCcap-AP6. These sequence-dependent length thresholds for forming a stably-bound segment are also observed experimentally from FTIR and 2D IR temperature series that monitor the dissociation of A:T and G:C base pairs as reported by changes in adenine and guanine ring modes, respectively (Figs. 4e,f and S17-S23). GCGcore-AP4 exhibits fraying of the short segment at temperatures below full dissociation of the duplex whereas such fraying is negligible for CCends-AP4 and GCGcore-AP6. (Fig. 4e,f) The ~4 °C shift of the A:T melting profile below the G:C melting profiles in CGCcap-AP6 suggest that the 5'-TATAT-3' segment does fray in a minor population of duplexes near the dissociation transition (Fig. S21). The resulting experimental melting profiles qualitatively agree with the temperature-dependent fraction of intact G:C and A:T base pairs from simulation. Further analysis of the total spectral change from FTIR and 2D IR melting experiments suggests that the short segments of GCGcore-AP2 and GCGcore-AP4 have a significant frayed population even below 4 °C (Figs. S17 and S18). These observations support the sequence-dependent trends in ΔF_{h2} from 3SPN.2 MD simulations and confirm that the length threshold for stable binding ($\Delta G_{h2} < 0$) of pure A:T segments is in the range of 4 – 7 base pairs at physiological temperatures while that for G:C segments is as low as 2 – 3 base pairs. The nucleation free-energy penalty for the second duplex segment is significantly lower than for the initial duplex nucleation, therefore it is important to note that these values underestimate thresholds for stable hybridization of free duplexes and are most applicable for forming segments that are covalently linked to another bound segment.

The sequence-dependence in the length threshold for forming a stably-bound segment contributes to the mismatch between $\Delta\Delta G_{h37}^\circ$ and $\Delta\Delta G_{h37,NN}^\circ$ observed in Fig. 3b. For instance, the spread in $\Delta\Delta G_{h37}^\circ$ for AP4 sequences may primarily occur because the A:T-rich weak segments of T11-AP4 and GCGcore-AP4 are more likely to fray than the weak segments of CGCcap-AP4 and CCends-AP4.

Kinetics of segment-dehybridization with T-jump IR spectroscopy

Only some sequences exhibit a large enough population of configurations with fraying of the weak segment to be observed in IR melting experiments, and the frayed configurations in other sequences are expected to be metastable and located between free-energy barriers for hybridization on each side of the AP site. To test for the presence of metastable configurations and a second barrier to hybridization, we turned to T-jump IR spectroscopy, which can be used to directly measure the kinetics of weak-segment dehybridization.(34) In these experiments, the sample is equilibrated at a temperature below T_m and optically heated by $\sim 15^\circ\text{C}$ ($\Delta T = T_f - T_i$) within 7 ns. The ensuing changes in base pairing of the duplex are monitored with heterodyned dispersed vibrational echo spectroscopy (HDVE)(44) from nanosecond-to-millisecond time delays. HDVE spectroscopy reports on changes in ring and carbonyl vibrational bands similarly to an IR pump-probe spectrum, and we report the spectra as the difference in signal at a given time delay after the T-jump relative to the maximum of the initial-temperature spectrum, $\Delta S(t) = [S(t) - S(T_i)]/\max[S(T_i)]$. We use the change in excited-state absorptions of the guanine ring mode at 1550 cm^{-1} and of the adenine ring mode at 1605 cm^{-1} as reporters of G:C and A:T base pairing, respectively, as in previous studies.(34, 38, 69)

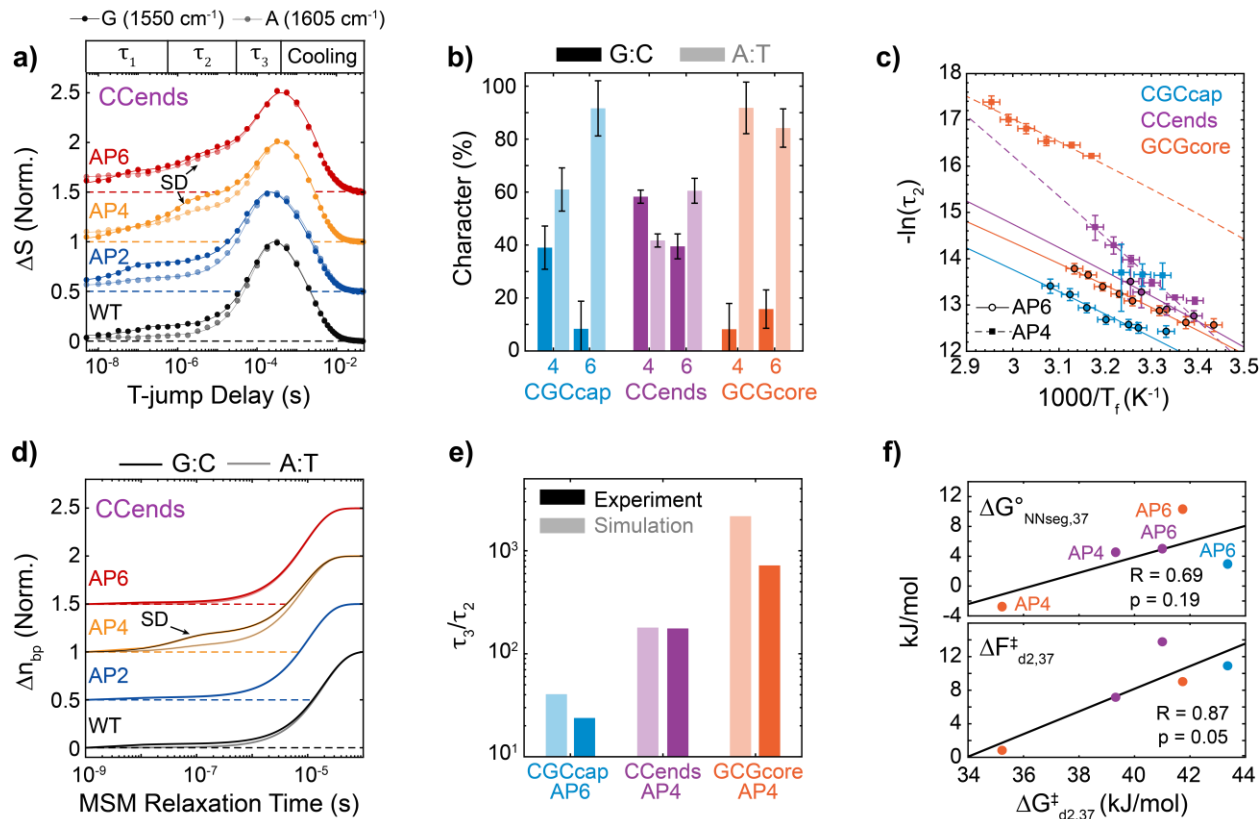


Figure 5. Comparison of segment-dehybridization kinetics observed in experiment and simulation. (a) T-jump IR (t-HDVE) time traces plotted for CCends sequences as the normalized signal change relative to the maximum of the initial-temperature spectrum, $\Delta S(t) = [S(t) - S(T_i)]/\max[S(T_i)]$. ΔS time traces probed at 1550 cm^{-1} (dark) report on changes in G:C base pairing and those at 1605 cm^{-1} (light) report on changes in A:T base pairing. T-jumps are performed from approximately $T_m - 15^\circ\text{C}$ to T_m for each sequence. Traces are shifted vertically with respect to one another, and dashed lines indicate respective baselines. Solid lines correspond to three- or four-component fits from global lifetime analysis (Section S5.2). A microsecond timescale segment-dehybridization response (SD) is observed for CCends-AP4 and CCends-AP6. (b) Percentage of G:C (dark, 1550 cm^{-1}) and A:T (light, 1605 cm^{-1}) base-pair loss character observed in segment-dehybridization t-HDVE responses. Percentages are derived from amplitudes determined from global lifetime fitting (eqs. S38a,b). Error bars indicate 95% confidence intervals propagated from global fits. (c) Temperature-dependent observed rate constants for segment-dehybridization ($1/\tau_2$) determined from T-jump IR. Data are fit to a Kramers-like equation in the high-friction limit (eq. S36 and Fig. S35). The rate trend for CGCcap-AP4 is not fit due to insufficient data. Vertical error bars correspond to 95% confidence intervals from global lifetime fitting and horizontal error bars are the standard deviation in T-jump magnitude. (d) Markov state model (MSM) T-jump simulations of the normalized change in intact G:C (dark) and A:T (light) base pairs, $\Delta n_{bp}(t) = [n_{bp}(t) - n_{bp}(T_i)]/[n_{bp}(T_f) - n_{bp}(T_i)]$, from $T_{m,MD} - 15^\circ\text{C}$ to $T_{m,MD}$. Traces are shifted vertically with respect to one another, and dashed lines indicate respective baselines. Black/gray solid lines for CCends-AP4 correspond to fits to a sum of two exponential components. (e) Ratio of full-strand dissociation time constant (τ_3) to τ_2 from T-jump IR (dark) and MSM T-jump simulations (light) indicating qualitative agreement between simulation and experiment. (f) Scatter plots of the segment-dehybridization free-energy barrier determined from T-jump IR ($\Delta G_{d2}^\ddagger$) vs. (top) the dehybridization free energy for the respective segment calculated from the NN model (ΔG_{NNseg}° , Table S3) and (bottom) $\Delta F_{d2}^\ddagger$ determined from 3SPN.2 MD simulations at 37 $^\circ\text{C}$. Linear fits (solid black line) are

shown for each plot with the respective Pearson correlation coefficient (R) and p-value. $\Delta G_{d2}^\ddagger$ shows better correlation with $\Delta F_{d2}^\ddagger$ than ΔG_{NNseg}° .

T-jump measurements reveal distinct kinetic behavior between WT and AP sequences (Figs. 5a and S24-S28). WT sequences show two dehybridization kinetic components on ~ 100 ns (τ_1) and ~ 100 μ s (τ_3) timescales followed by thermal relaxation back to T_i at ~ 1 ms. The T-jump IR data are well described by global lifetime fitting with exponentially damped spectral components, and the same number of components and similar kinetics are obtained from inverse-Laplace-transform rate distribution spectra (Figs. S29-S30). The τ_1 and τ_3 processes were previously assigned to T-jump induced terminal fraying of the duplex and full-strand dehybridization, respectively,(34) consistent with reports for similar sequences.(69) The temperature-dependent trends in timescale and amplitude of τ_1 and τ_3 support their assignment to terminal fraying and complete strand dissociation (Figs. S31-S33). Many of the AP4 and AP6 sequences show an additional kinetic component (labeled τ_2) between 500 ns to 10 μ s with sequence-dependent amplitude and timescale. The τ_2 response corresponds to dehybridization of a complete stretch of base pairs on one side of the AP site (segment dehybridization) and reveals the presence of a free-energy barrier from the fully intact duplex to segment-dehybridized configurations ($\Delta G_{d2}^\ddagger$). The observation of $\Delta G_{d2}^\ddagger$ verifies the presence of $\Delta G_{h2}^\ddagger$ in the reverse direction, as illustrated from the FEPs in Fig. 4a and previously reported for the AP6 sequences.(34)

The spectral properties and kinetics of the τ_2 process support its assignment to segment-dehybridization and provide insight into the properties of $\Delta G_{h2}^\ddagger$. The fraction of A:T and G:C base-pair loss character during τ_2 determined from global lifetime fitting (Figs. 5b and S34) match the expected signal change for dehybridizing either half of CCends-AP6 and GCGcore-AP6, the 5'-TATAT-3' segment in CGCcap-AP6, and the 5'-CCT-3' segment in CCends-AP4. Additionally, the segment dehybridization rate, τ_2^{-1} , increases exponentially with temperature as expected for an active process, illustrated by the Arrhenius plot in Fig. 5c, which we analyze using a Kramers-like equation (eq. S36). The τ_2 response of CGCcap-AP4 is more difficult to interpret. It has low amplitude and mixed G:C and A:T character even though each segment is purely composed of G:C or A:T base pairs. Due to the similar binding stability of 5'-CGC-3' and 5'-TATATAT-3' segments, τ_2 may contain similar amplitude from both segment-hybridization

pathways. Lastly, we observe that GCGcore-AP4 does not have a τ_2 response; however, unlike all other sequences, the rate τ_1^{-1} increases exponentially with temperature in an activated manner. We therefore assign τ_1 in GCGcore-AP4 to dehybridization of the 5'-TAT-3' segment and note that it may overlap in time with terminal fraying dynamics from the other end of the duplex.

We find large variations in segment-dehybridization kinetics across sequence and segment length from T-jump IR measurements. To more directly connect these measurements to simulation, we built kinetic models of T-jump relaxation using Markov state models (MSMs) constructed from 250 μ s of unbiased 3SPN.2 simulations near $T_{m,MD}$ for each sequence (Figs. S36-S38) as previously reported for WT and AP6 sequences.(34) The change in the fraction of intact A:T and G:C base pairs (Δn_{bp}) is computed as a function of relaxation time at $T_{m,MD}$ after initiating the MSM kinetics with a population distribution obtained at $T_{m,MD}-15^\circ\text{C}$ to replicate the initial experimental T-jump ensemble. Relaxation traces in Figs. 5d and S24 show two to three kinetic components that correspond to terminal fraying (1 – 30 ns, τ_1), segment-dehybridization (50 – 500 ns, τ_2), and complete strand dissociation (1 – 50 μ s, τ_3). The MSM relaxation timescales are accelerated by ~ 1 order of magnitude relative to T-jump IR measurements, as previously reported for the 3SPN.2 model,(70) and therefore we use the ratio between τ_3 and τ_2 (or τ_1 for GCGcore-AP4) to compare simulation and experimental timescales. Experimental and simulated τ_3/τ_2 values show qualitative agreement and an increase from CGCcap-AP6 to CCends-AP4 to GCGcore-AP4, suggesting that 3SPN.2 accurately captures sequence-dependent variation in $\Delta G_{d2}^\ddagger$ (Fig. 5e). We also find a correlation ($R = 0.87$) between the $\Delta G_{d2}^\ddagger$ and $\Delta F_{d2}^\ddagger$ from the FEPs in Fig. 4a at 37 $^\circ\text{C}$ (Fig. 5f). The correlation between $\Delta G_{d2}^\ddagger$ and the calculated NN dehybridization free energy of the respective unbinding segment (ΔG_{NNseg}°) is substantially poorer ($R = 0.69$), suggesting that both the dehybridization energy of the segment (ΔG_{d2}) and its nucleation barrier ($\Delta G_{h2}^\ddagger$) significantly contribute to the sequence-dependence of $\Delta G_{d2}^\ddagger$.

MSM relaxation profiles generally exhibit a lower segment-dehybridization response amplitude than in experiment, and negligible response is observed for CGCcap-AP4, CCends-AP6, and GCGcore-AP6 even though each sequence possesses a metastable segment-dehybridized state (Figs. 4a, 5d, and S39). For each of these three sequences, we observe $\Delta F_{d2}^\ddagger$ to be larger than the free-energy barrier for dehybridization of the final base-pair segment ($\Delta F_{d1}^\ddagger$) at

$T_{m,MD}$ (Fig. S16). When $\Delta F_{d2}^\ddagger > \Delta F_{d1}^\ddagger$, the waiting time to surpass $\Delta F_{d2}^\ddagger$ is the rate-limiting step for dehybridization, and we do not expect to observe a segment-dehybridization response in the T-jump kinetics. In contrast, we observe $\Delta F_{d2}^\ddagger \leq \Delta F_{d1}^\ddagger$ for sequences where there is a large amplitude segment-dehybridization response (CCends-AP4, CGCcap-AP6, GCGcore-AP4). Segment-dehybridization is only observed experimentally for CGCcap-AP4 and CCends-AP6 at the few lowest measured temperatures, which may be explained by $\Delta F_{d2}^\ddagger$ increasing relative to $\Delta F_{d1}^\ddagger$ and surpassing it at higher temperatures (Fig. S16). The overall lower amplitude of segment-dehybridization response in simulations therefore suggests that the ratio of $\Delta F_{d2}^\ddagger$ to $\Delta F_{d1}^\ddagger$ is generally overestimated, and as $\Delta F_{d2}^\ddagger$ is fairly correlated with $\Delta G_{d2}^\ddagger$, we suggest that sequence-dependent error in this ratio primarily arises from $\Delta F_{d1}^\ddagger$. Error in $\Delta F_{d1}^\ddagger$ may stem from inaccurate modeling of base-pair and stacking interactions around the AP site for which 3SPN.2 is not parameterized. Further, 3SPN.2 simulations tend to poorly predict sequence-dependent variation in the dehybridization barrier among canonical oligonucleotides.(70)

Contributions from non-canonical base pairing

Out-of-register base pairing at near-terminal AP sites

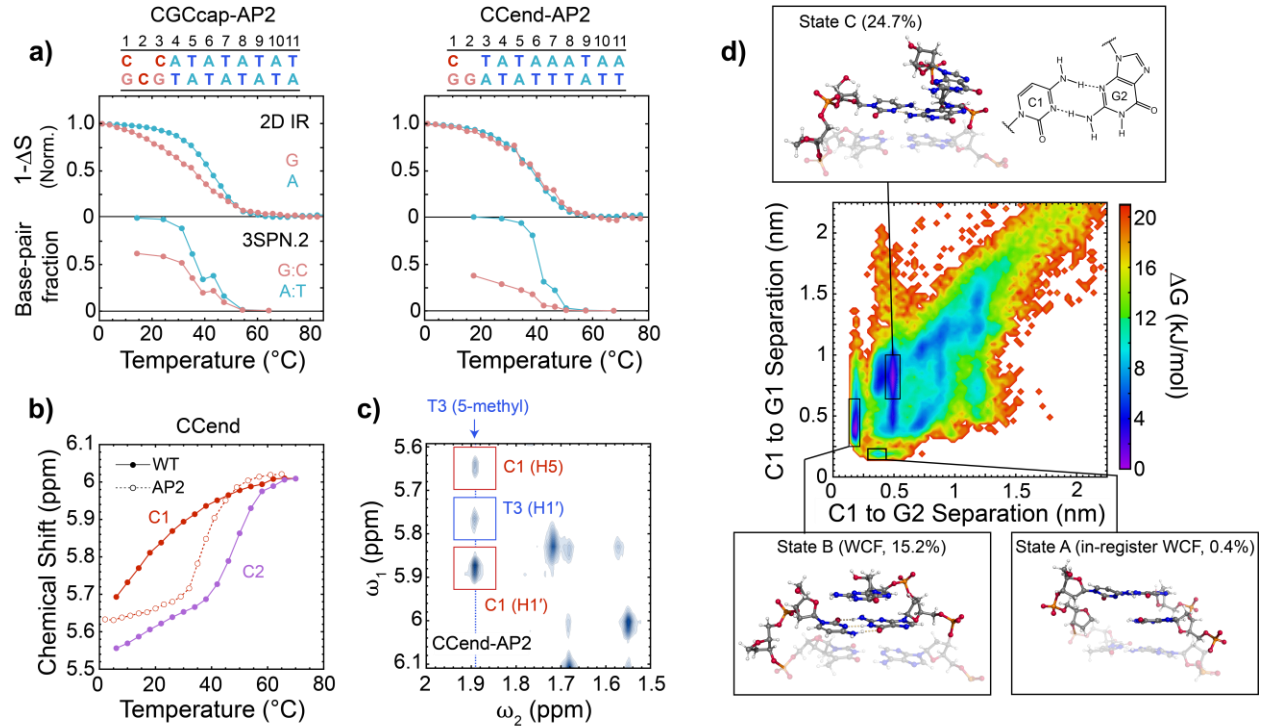


Figure 6. Out-of-register base pairing in CCends-AP2. (a) (top) Normalized temperature-dependent change in 2D IR signals relative to the initial temperature, $1 - \Delta S(T) = [S(T) - S(1^\circ\text{C})]/S(1^\circ\text{C})$, at guanine (G) and adenine (A) ring vibrational bands for CGCcap-AP2 and CCend-AP2. Corresponding 2D IR spectra are shown in Fig. S20. (bottom) Fraction of intact G:C and A:T base pairs from 3SPN.2 MD simulations carried out at 8 – 9 temperatures across the duplex melting transition. Low-temperature fraying of the weak segment is observed in experiment and simulation for CGCcap-AP2, yet not observed experimentally for CCend-AP2. (b) NMR chemical shifts of cytosine H5 nuclei for CCend-WT (filled circles, solid lines) and CCend-AP2 (open circles, dashed lines) determined as a function of temperature from TOCSY measurements. The terminal cytosine (C1) shows low-temperature fraying in CCend-WT whereas the second cytosine (C2) signal primarily changes during duplex dissociation. C1 does not exhibit fraying CCend-AP2 and instead tracks duplex dissociation. (c) NOESY cross peaks between H5/H1' nuclei and 5-methyl protons of thymine for CCend-AP2. Cross-peaks are observed between C1 and T3 (cyan boxes). (d) Free-energy surface from $20 \times 1 \mu\text{s}$ all-atom MD simulations of CCend-AP2 using the AMBER-bsc1 force field at 30°C . The x-axis is the average separation between hydrogen-bonding atoms between C1 and G2 (out-of-register), and the y-axis is the average separation between those atoms in C1 and G1 (in-register). Snapshots of base pairs 1-3 from select free-energy minima are shown, and the population from a given minimum is determined by integrating over the marked rectangles. Additional states are highlighted in Fig. S48. In agreement with experiment, C1 primarily base pairs out-of-register yet contains multiple free-energy minima with distinct base-pairing and stacking configurations.

The simulations presented in Fig. 4 suggest that the terminal base pair next to an AP2 site is highly frayed at physiological temperatures regardless of sequence, but our experiments indicate that the terminal base pair can stably bind in certain sequence contexts. As introduced in Fig. 4e, we monitor the temperature-dependence of G:C and A:T features in the IR spectrum to check for the presence of segment-dehybridization at temperatures well below T_m in CGCcap-AP2 and CCend-AP2, a modified form of CCends with only a single CC terminus. Relative to FTIR, 2D IR temperature series are found to provide greater contrast between spectral changes arising from disruption of base pairs and those giving rise to melting curve baselines (Figs. S19 and S21). CGCcap-AP2 shows a clear deviation from all-or-nothing melting behavior where G:C base-pair character is lost at temperatures 20-30 $^\circ\text{C}$ below A:T melting, likely corresponding to disruption of the terminal G:C base pair as predicted from 3SPN.2 MD simulations (Fig. 6a). In contrast, the temperature profiles of G:C and A:T signatures in CCend-AP2 are the same, suggesting that the terminal G:C base pair is stabilized relative to CGCcap-AP2.

The lack of partial dehybridization in CCend-AP2 suggests its terminal base pair is stabilized relative to the other sequences, and we used ^1H NMR spectroscopy to further examine the dominant duplex configurations of CCend and CGCcap sequences. Melting profiles for

individual G:C base pairs were extracted from the temperature-dependent chemical shift of cytosine H5 nuclei isolated using ^1H - ^1H TOCSY temperature series (Figs. 6b and S40 – S47). Interior cytosines (C2, C3) of CCend-WT and CGCcap-WT show sigmoidal transitions over the temperature range of the duplex melting transition measured with IR spectroscopy, but most of the frequency change in the terminal C1 H5 nuclei occurs at temperatures well below the melting transition. We assign the low-temperature behavior at C1 to fraying of the terminal base pair, which is further supported by the lack of such changes in GCGcore-WT where all cytosines are at the center of the duplex (Fig. S47). The temperature-dependence of the C1 H5 frequency in CCend-AP2 follows the duplex melting transition, suggesting that the terminal G:C base pair is intact and dissociates together with the rest of the duplex as indicated by 2D IR temperature series. The ^1H - ^1H NOESY spectrum reveals intense cross-peaks between C1 and T3 nuclei, indicating that C1 is stacked with T3 in the duplex state and presumably base paired with G2 rather than G1 (Figs. 6c, S42 – S43). This out-of-register base pairing creates a duplex with one stretch of base pairs, circumventing the nucleation penalty for forming a new segment. Similar out-of-register configurations have been proposed previously around AP sites.(29, 71) In contrast, the temperature-dependent cytosine H5 frequencies in CGCcap-AP2 indicate partial dehybridization of the G:C cap more than 10 °C below the duplex melting transition. The TOCSY spectra of CGCcap-AP2 also show additional cytosine peaks from 2 to 26 °C that suggest multiple slowly interconverting duplex configurations may be present at low temperature (Figs. S45 – S46), yet the assignment of these structures are still unclear.

All-atom MD simulations provide insight into the possible base-pairing and stacking configurations adopted by CCend-AP2 and CGCcap-AP2. Free-energy surfaces (FESs) constructed from 20 independent 1 μs simulations indicate that each sequence adopts a variety of non-canonical configurations that interconvert on sub-microsecond timescales (Figs. S48 – S49). Interpretation of the simulations results must be done with care since the AMBER-bsc1 force is not parameterized for AP sites. While we cannot anticipate accurate predictions of the relative stabilities of the various states, the predictions are nonetheless useful for identifying putative configurational states and their relative propensities to help interpret experimental measurements. Only 1% of CGCcap-AP2 configurations exhibit in-register base pairing between C1 and G1, and C1 instead prefers to stack with C3 and interact with C2. Similar for CCend-AP2, configurations with in-register base pairing between C1 and G1 only make up 0.4% of the population (Fig. 6d).

The most populated states are those with C1 paired to G2 and the free deoxyribose group fully extrahelical (states B and C in Fig. 6d). State B (15.2%) is characterized by a Watson-Crick-Franklin (WCF) base pair between C1 and G2 with stacking of G1 over the base pair as often observed for duplexes with dangling ends.⁽⁷²⁾ The C1 nucleotide rotates approximately 180° in state C (24.7%) and forms a base pair composed of two amino-imino hydrogen bonds, N3(C) to N2H(G) and N4H(C) to N3(G), and G1 stacks only with G2. This type of base pair has previously been observed for mononucleotides binding to RNA templates and in other folded RNA structures.^(73, 74) Another common configuration involves C1 stacked between bases on the other strand (state D). State B is most consistent with the NOESY spectrum for CCend-AP2 because it is the only configuration where the H5 nuclei of C1 is close enough to the T3 methyl group (2.7 Å vs. 8.8 Å in state C and 8-10 Å in state D) to enable significant dipolar coupling to produce an intense cross-peak. While the NOESY data suggests a configuration like state B has the largest population experimentally, we cannot rule out the presence of states C and D as they interconvert with state B too quickly to be resolved as separate sets of NMR peaks.

Potential role of out-of-register duplex configurations in repetitive sequences

The stability of the duplex for the repetitive T11 sequence can be affected by out-of-register base pairing, and these configurations are observed in 3SPN.2 MD simulations. Figure 7 shows duplex free-energy minima at $n_{bp} = 10$ for T11-WT, $n_{bp} = 9$ for T11-AP2, $n_{bp} = 7$ for T11-AP4, and $n_{bp} = 5$ for T11-AP6 corresponding to duplex configurations shifted by one, two, four, and six base-pair indices, respectively. The shifted configurations for T11-AP2 and T11-AP4 are the dominant duplex species even at $T_{m,MD} - 20$ °C, but the population of those for T11-WT and T11-AP6 grow with increasing temperature until the duplex fully dissociates (Figs. 7c and S12). Relative to unbound segments of in-register segment-dehybridized configurations found in CGCcap, CCends, and GCGcore sequences, the single-strand overhangs of out-of-register configurations have greater conformational entropy and a lower electrostatic penalty that lead to enhanced stability of the duplex region.⁽⁷⁵⁾

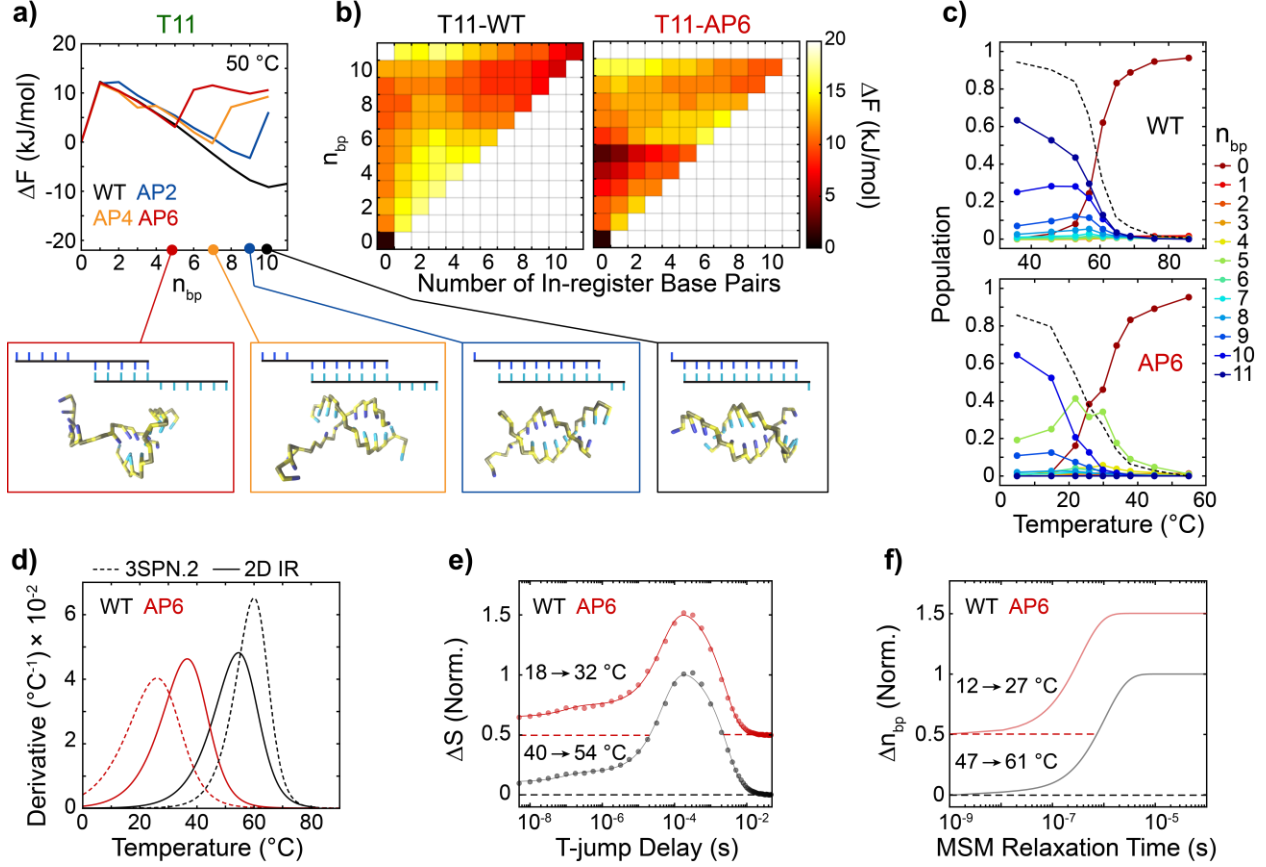


Figure 7. Assessment of out-of-register base pairing induced by AP site in T11. (a) FEPs as a function of n_{bp} for T11 sequences at 50 °C from 3SPN.2 MD simulations with WTMetaD. Representative structures show out-of-register shifted configurations at duplex free-energy minima. (b) 2D free-energy surfaces for (left) T11-WT and (right) T11-AP6 of the number of in-register base pairs vs. n_{bp} . The free-energy minimum at $n_{bp} = 5$ for T11-AP6 corresponds to a completely out-of-register configuration. (c) Temperature-dependent population profiles of duplexes with n_{bp} ranging from 0 to 11 for (top) T11-WT and (bottom) T11-AP6. The black dashed line corresponds to the total fraction of intact base pairs. (d) Derivative of the temperature-dependent fraction of intact base pairs from 3SPN.2 (dashed lines) and 2D IR melting curves (solid lines). Derivatives are taken on melting curves determined from two-state fits to the temperature-dependent data (Figs. S22 – S23). (e) Normalized $\Delta S(t)$ time traces at 1605 cm^{-1} (A:T) plotted for the T11-WT and T11-AP6. T-jumps are performed from approximately $T_m - 15^\circ\text{C}$ to T_m for each sequence. Traces are shifted vertically with respect to one another, and dashed lines indicate respective baselines. Solid lines correspond to three-component fits from global lifetime analysis. (f) MSM T-jump simulations of the normalized change in A:T base pairs, $\Delta n_{bp}(t)$, using the same method as in Fig. 5d. Due to overlapping timescales for partial dehybridization to the out-of-register $n_{bp} = 5$ state and full-strand dissociation, only a single dehybridization kinetic component is observed in the relaxation profiles.

Out-of-register base pairing is commonly observed in coarse-grained MD simulations and statistical modeling of repetitive DNA oligonucleotides,(37, 70, 76-78) yet it is challenging to

verify the population of such configurations experimentally. The IR spectral features of in- and out-of-register A:T base pairing are expected to be essentially identical, therefore our experimental data does not directly report on the presence of out-of-register base pairing. Additionally, the experimental temperature-dependent melting behavior of T11 sequences does not indicate a significant population of out-of-register configurations and suggests that 3SPN.2 simulations overestimate the thermodynamic stability of out-of-register configurations for T11 AP sequences. The probability of out-of-register hybridization pathways may be overestimated due to smoothening of the free-energy landscape generally induced by coarse-graining.⁽⁷⁹⁾ The total change in adenine FTIR and 2D IR signal upon thermal dissociation is nearly constant in T11-WT and each AP sequence, suggesting each duplex contains a similar degree of A:T base pairing at the lowest measured temperature in the duplex state (Fig. S34). Further, 3SPN.2 predicts that partial dehybridization of T11-AP4 and T11-AP6 occurs at temperatures below $T_{m,MD}$, which broadens the overall melting curve and reduces the slope at the inflection point relative to T11-WT for the fraction of intact base pairs by 35% and for the fraction of intact duplexes by 18% (Fig. 7d and S23). 2D IR experiments show only a 5% reduction in the melting curve slope of T11-AP6 and no indication of partial hybridization in the melting profile (Fig. S22), suggesting that the out-of-register configurations are less populated than in the simulations. In contrast, T11-AP4 exhibits a steeper low-temperature baseline and broader melting transition than T11-AP6 that may result from temperature-dependent fraying of the three-base-pair segment or shifting out-of-register (Fig. S23a).

We also find that MSM and T-jump IR relaxation are insensitive to the kinetics of out-of-register shifting (Figs. 7e,f and S17). Although a thermally-induced increase of the $n_{bp} = 5$ state is predicted for T11-AP6 and $\Delta F_{d2}^{\ddagger} < \Delta F_{d1}^{\ddagger}$, the MSM relaxation timescale for shifting overlaps with full-strand dissociation (Fig. 7f), which was observed previously in similar canonical oligonucleotides.⁽⁷⁰⁾ Shifting from the $n_{bp} = 10$ to $n_{bp} = 5$ state first entails breaking of all in-register contacts followed by formation of five out-of-register base pairs. The barrier height for shifting is likely much larger than indicated by $\Delta F_{d2}^{\ddagger}$ in Fig. 7a, and n_{bp} is a poor coordinate to describe the shifting transition. As a result, no partial-dehybridization response is observed in MSM relaxation profiles. Similarly, T-jump IR measurements only show responses for terminal

fraying and full-strand dissociation and cannot indicate nor rule out a role of out-of-register base pairing in these sequences (Fig. 7e).

Conclusions

Our study reveals molecular details underpinning the position-dependent duplex destabilization from an AP site within DNA oligonucleotides by combining temperature-dependent IR and NMR spectroscopy, T-jump IR kinetics, and coarse-grained MD simulations. An AP site destabilizes the duplex through a loss of stacking and base-pairing interactions that disrupt cooperative base pairing and introduce a free-energy barrier for nucleating base-pair segments on each side of the AP site. This nucleation barrier leads to a position-dependent duplex destabilization from the AP site. Experiments and simulations show that the nucleation barrier promotes fraying of the weak segment when the AP site is near the termini. As the AP site moves inward by three to six base-pair sites, the short segment has enough stability to bind, and the frayed configuration becomes metastable as observed in T-jump measurements. At this position, the nucleation barrier is fully encompassed into the duplex and the AP site is maximally destabilizing. The AP-site positions over which fraying becomes metastable depend on the nucleobase sequence in the weak segment and adjacent to the AP site, leading to poor prediction of duplex destabilization by the NN model even at a fixed AP-site position. While the transition occurs at weak-segment lengths of two to five base pairs depending on the G:C content of the segment, certain short-segment sequences may be able to circumvent the nucleation barrier by forming base pairs out-of-register as found experimentally for CCend-AP2.

This work has focused on AP sites, yet there are numerous non-canonical base pairs, nucleobase modifications, and damaged-induced lesions that impart a comparable energetic penalty to duplex formation ($\Delta\Delta G_{h37}^{\circ} = 5\text{-}20$ kJ/mol).^(8, 13, 80, 81) Although they alter local base-pairing and stacking interactions in different ways, such modifications and non-canonical base pairs may disrupt base-pairing cooperativity and follow a position-dependence similar to an AP site. For example, base-pair mismatches impose a sequence-dependent free-energy penalty to duplex formation that follows a similar position-dependence as for AP sites.^(8, 19, 20) Greater reduction in duplex stability occurs as the mismatch moves inward 3 – 4 base pairs from the termini. While the base-pairing geometry of mismatches is highly sequence-dependent, the ‘defect penalty’ used to model their behavior is essentially the same as the nucleation penalty described

in our work and supports that the position-dependent thermodynamic and dynamic consequences from destabilizing modifications may be similar to what we find for AP sites.

Data Availability

Python scripts for generating abasic configurations from intact 3SPN.2 files, performing metadynamics simulations, and reweighting free-energy surfaces are available at <https://github.com/mrjoness/abasic-thermo/>. Scripts for running equilibrium simulations and building Markov state models are available at <https://github.com/mrjoness/abasic-kinetics/>. All PLUMED input scripts were submitted to the PLUMED-NEST public repository and are available at <https://www.plumed-nest.org/eggs/22/037/>.

Initialization files and trajectory data for unbiased coarse-grained simulations used to construct Markov state models, biased coarse-grained simulations over multiple temperatures, and all-atoms simulations of AP2 sequences are uploaded to Zenodo and available at [10.5281/zenodo.8169462](https://doi.org/10.5281/zenodo.8169462). Free-energy profiles along n_{bp} and r_{bp} , Markov state model relaxation profiles, time- and rate-domain t-HDVE data, FTIR temperature series, 2D IR temperature series, and NMR temperature series are available at [10.5281/zenodo.8174843](https://doi.org/10.5281/zenodo.8174843).

Supporting Material

Supporting material can be found online at xxx.

Author Contributions

B.A., M.S.J., A.L.F, and A.T designed the research. B.A., Y.L. and J.R.S. performed experiments. B.A. performed helix-coil modeling. M.S.J. performed molecular dynamics simulations. B.A. and M.S.J. analyzed the data. B.A. and A.T. wrote the manuscript, and M.S.J. wrote the computational methods section. All authors provided feedback on the manuscript.

Declaration of Interests

A.L.F. is a co-founder and consultant of Evozyne, Inc. and a co-author of US Patent Applications 16/887,710 and 17/642,582, US Provisional Patent Applications 62/853,919, 62/900,420,

63/314,898, 63/479,378, and 63/521,617 and International Patent Applications PCT/US2020/035206 and PCT/US2020/050466.

Acknowledgments

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health (Award No. R01-GM118774) and the National Science Foundation under Grant No. CHE-2152521 and CHE-2155027. B.A. acknowledges support from the NSF Graduate Research Fellowship Program. This work was completed in part with resources provided by the University of Chicago Research Computing Center. We gratefully acknowledge computing time on the University of Chicago high-performance GPU-based cyberinfrastructure supported by the National Science Foundation under Grant No. DMR-1828629.

Supporting References

References (82 – 99) appear in the supporting material.

References

1. Bumgarner R. 2013. Overview of DNA microarrays: types, applications, and their future. *Curr. Protoc. Mol. Biol.* 101:22.21. 21-22.21. 11.
2. Strauss S. and Jungmann R. 2020. Up to 100-fold speed-up and multiplexing in optimized DNA-PAINT. *Nat. Methods* 17:789-791.
3. Song J., Li Z., Wang P., Meyer T., Mao C., and Ke Y. 2017. Reconfiguration of DNA molecular arrays driven by information relay. *Science* 357:eaan3377.
4. Klein M., Eslami-Mossallam B., Arroyo D. G., and Depken M. 2018. Hybridization kinetics explains CRISPR-Cas off-targeting rules. *Cell Rep.* 22:1413-1423.
5. Altun A., Garcia-Ratés M., Neese F., and Bistoni G. 2021. Unveiling the complex pattern of intermolecular interactions responsible for the stability of the DNA duplex. *Chem. Sci.* 12:12785-12793.
6. Yakovchuk P., Protozanova E., and Frank-Kamenetskii M. D. 2006. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* 34:564-574.
7. SantaLucia J. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Nat. Acad. Sci.* 95:1460-1465.
8. SantaLucia Jr J. and Hicks D. 2004. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* 33:415-440.
9. Hertel S., Spinney R., Xu S., Ouldrige T. E., Morris R., and Lee L. 2021. The stability and number of nucleating interactions determine DNA hybridization rates in the absence of secondary structure. *Nucleic Acids Res.* 50:7829-7841.

10. Zhang J. X., Fang J. Z., Duan W., Wu L. R., Zhang A. W., Dalchau N., Yordanov B., Petersen R., Phillips A., and Zhang D. Y. 2018. Predicting DNA hybridization kinetics from sequence. *Nat. Chem.* 10:91-98.
11. Todisco M. and Szostak J. W. 2022. Hybridization kinetics of out-of-equilibrium mixtures of short RNA oligonucleotides. *Nucleic Acids Res.* 50:9647-9662.
12. Phan T. T., Phan T. M., and Schmit J. D. 2023. Beneficial and detrimental effects of non-specific binding during DNA hybridization. *Biophys. J.* 122:835-848.
13. Dubini R. C. A., Korytiaková E., Schinkel T., Heinrichs P., Carell T., and Rovó P. 2022. ¹H NMR chemical exchange techniques reveal local and global effects of oxidized cytosine derivatives. *ACS Phys. Chem. Au* 2:237-246.
14. Sanstead P. J., Ashwood B., Dai Q., He C., and Tokmakoff A. 2020. Oxidized derivatives of 5-methylcytosine alter the stability and dehybridization dynamics of duplex DNA. *J. Phys. Chem. B* 124:1160-1174.
15. Liu B., Shi H., Rangadurai A., Nussbaumer F., Chu C.-C., Erharter K. A., Case D. A., Kreutz C., and Al-Hashimi H. M. 2021. A quantitative model predicts how m6A reshapes the kinetic landscape of nucleic acid hybridization and conformational transitions. *Nat. Commun.* 12:1-17.
16. Watkins Jr N. E. and SantaLucia Jr J. 2005. Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. *Nucleic Acids Res.* 33:6258-6267.
17. Szabat M., Prochota M., Kierzek R., Kierzek E., and Mathews D. H. 2022. A Test and Refinement of Folding Free Energy Nearest Neighbor Parameters for RNA Including N6-Methyladenosine. *J. Mol. Biol.* :167632.
18. Hopfinger M. C., Kirkpatrick C. C., and Znosko B. M. 2020. Predictions and analyses of RNA nearest neighbor parameters for modified nucleotides. *Nucleic Acids Res.* 48:8901-8913.
19. Naiser T., Kayser J., Mai T., Michel W., and Ott A. 2008. Position dependent mismatch discrimination on DNA microarrays—experiments and model. *BMC Bioinform.* 9:1-12.
20. Naiser T., Ehler O., Kayser J., Mai T., Michel W., and Ott A. 2008. Impact of point-mutations on the hybridization affinity of surface-bound DNA/DNA and RNA/DNA oligonucleotide-duplexes: comparison of single base mismatches and base bulges. *BMC Biotechnol.* 8:1-23.
21. Pozhitkov A., Noble P. A., Domazet-Lošo T., Nolte A. W., Sonnenberg R., Staehler P., Beier M., and Tautz D. 2006. Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Res.* 34:e66-e66.
22. Ding F., Cocco S., Raj S., Manosas M., Nguyen T. T. T., Spiering M. M., Bensimon D., Allemand J.-F., and Croquette V. 2022. Displacement and dissociation of oligonucleotides during DNA hairpin closure under strain. *Nucleic Acids Res.* 50:12082-12093.
23. Todisco M., Ding D., and Szostak J. W. 2023. Transient States During Oligonucleotides Hybridization: Insights from Annealing Kinetics of Mismatches and Bulges. *bioRxiv*, doi:10.1101/559754 (preprint posted Sept. 28, 2023)
24. Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature* 362:709-715.
25. Thompson P. S. and Cortez D. 2020. New insights into abasic site repair and tolerance. *DNA Repair* 90:102866.

26. Kimsey I. J., Szymanski E. S., Zahurancik W. J., Shakya A., Xue Y., Chu C.-C., Sathyamoorthy B., Suo Z., and Al-Hashimi H. M. 2018. Dynamic basis for dG• dT misincorporation via tautomerization and ionization. *Nature* 554:195-201.
27. Afek A., Shi H., Rangadurai A., Sahay H., Senitzki A., Xhani S., Fang M., Salinas R., Mielko Z., Pufall M. A., et al. 2020. DNA mismatches reveal conformational penalties in protein–DNA recognition. *Nature* 587:291-296.
28. Rossetti G., Dans P. D., Gomez-Pinto I., Ivani I., Gonzalez C., and Orozco M. 2015. The structural impact of DNA mismatches. *Nucleic Acids Res.* 43:4309-4321.
29. Chen J., Dupradeau F.-Y., Case D. A., Turner C. J., and Stubbe J. 2008. DNA oligonucleotides with A, T, G or C opposite an abasic site: structure and dynamics. *Nucleic Acids Res.* 36:253-262.
30. Lukin M. and de Los Santos C. 2006. NMR structures of damaged DNA. *Chem. Rev.* 106:607-686.
31. Vesnaver G., Chang C.-N., Eisenberg M., Grollman A. P., and Breslauer K. J. 1989. 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.
32. Gelfand C. A., Plum G. E., Grollman A. P., Johnson F., and Breslauer K. J. 1998. Thermodynamic consequences of an abasic lesion in duplex DNA are strongly dependent on base sequence. *Biochem.* 37:7321-7327.
33. Sági J., Guliaev A. B., and Singer B. 2001. 15-mer DNA duplexes containing an abasic site are thermodynamically more stable with adjacent purines than with pyrimidines. *Biochem.* 40:3859-3868.
34. Ashwood B., Jones M. S., Ferguson A. L., and Tokmakoff A. 2023. Disruption of Energetic and Base Pairing Cooperativity in DNA Duplexes by an Abasic Site. *Proc. Nat. Acad. Sci.* 120:e2219124120.
35. Wang W., Walmacq C., Chong J., Kashlev M., and Wang D. 2018. Structural basis of transcriptional stalling and bypass of abasic DNA lesion by RNA polymerase II. *Proc. Nat. Acad. Sci.* 115:E2538-E2545.
36. Hinckley D. M., Freeman G. S., Whitmer J. K., and De Pablo J. J. 2013. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* 139:10B604_601.
37. Hinckley D. M., Lequieu J. P., and de Pablo J. J. 2014. Coarse-grained modeling of DNA oligomer hybridization: length, sequence, and salt effects. *J. Chem. Phys.* 141:07B613_611.
38. Sanstead P. J., Stevenson P., and Tokmakoff A. 2016. Sequence-dependent mechanism of DNA oligonucleotide dehybridization resolved through infrared spectroscopy. *J. Am. Chem. Soc.* 138:11792-11801.
39. Ramos P. M. and Ruisánchez I. 2005. Noise and background removal in Raman spectra of ancient pigments using wavelet transform. *J. Raman Spectrosc.* 36:848-856.
40. Mallat S. 1989. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 11:674-693.
41. Ashwood B., Lewis N. H., Sanstead P. J., and Tokmakoff A. 2020. Temperature-Jump 2D IR Spectroscopy with Intensity-Modulated CW Optical Heating. *J. Phys. Chem. B* 124:8665-8677.

42. Chung H. S., Khalil M., Smith A. W., and Tokmakoff A. 2007. Transient two-dimensional IR spectrometer for probing nanosecond temperature-jump kinetics. *Rev. Sci. Instrum.* 78:063101.
43. Jones K. C., Ganim Z., Peng C. S., and Tokmakoff A. 2012. Transient two-dimensional spectroscopy with linear absorption corrections applied to temperature-jump two-dimensional infrared. *J. Opt. Soc. Am. B* 29:118-129.
44. Jones K. C., Ganim Z., and Tokmakoff A. 2009. Heterodyne-detected dispersed vibrational echo spectroscopy. *J. Phys. Chem. A* 113:14060-14066.
45. Plimpton S. 1995. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* 117:1-19.
46. Bussi G. and Parrinello M. 2007. Accurate sampling using Langevin dynamics. *Phys. Rev. E* 75:056707.
47. Nkodo A. E., Garnier J. M., Tinland B., Ren H., Desruisseaux C., McCormick L. C., Drouin G., and Slater G. W. 2001. Diffusion coefficient of DNA molecules during free solution electrophoresis. *Electrophoresis* 22:2424-2432.
48. Debye P. and Hückel E. 1923. De la theorie des electrolytes. I. abaissement du point de congelation et phenomenes associes. *Phys. Z.* 24:185-206.
49. Tribello G. A., Bonomi M., Branduardi D., Camilloni C., and Bussi G. 2014. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* 185:604-613.
50. Barducci A., Bussi G., and Parrinello M. 2008. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* 100:020603.
51. Bussi G. and Laio A. 2020. Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.* 2:200-212.
52. Wehmeyer C., Scherer M. K., Hempel T., Husic B. E., Olsson S., and Noé F. 2019. Introduction to Markov state modeling with the PyEMMA software. *Living J. Comput. Mol. Sci.* 1:5965.
53. Pande V. S., Beauchamp K., and Bowman G. R. 2010. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* 52:99-105.
54. Case D. A., Aktulga H. M., Belfon K., Ben-Shalom I. Y., Berryman J. T., Brozell S. R., Cerutti D. S., Cheatum III T. E., Cisneros G. A., Cruzeiro V. W. D., et al. 2022. AMBER 2022. University of California, San Francisco.
55. Ivani I., Dans P. D., Noy A., Pérez A., Faustino I., Hospital A., Walther J., Andrio P., Goñi R., and Balaceanu A. 2016. Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* 13:55-58.
56. Mark P. and Nilsson L. 2001. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* 105:9954-9960.
57. Hockney R. W. and Eastwood J. W. (1988) *Computer simulation using particles* (CRC Press).
58. Bussi G., Donadio D., and Parrinello M. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.
59. Hess B., Bekker H., Berendsen H. J., and Fraaije J. G. 1997. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463-1472.
60. Darden T., York D., and Pedersen L. 1993. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089-10092.
61. Bottaro S., Di Palma F., and Bussi G. 2014. The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res.* 42:13306-13314.

62. Bonomi M. 2019. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* 16:670-673.
63. McGibbon R. T., Beauchamp K. A., Harrigan M. P., Klein C., Swails J. M., Hernández C. X., Schwantes C. R., Wang L.-P., Lane T. J., and Pande V. S. 2015. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* 109:1528-1532.
64. Scherer M. K., Trendelkamp-Schroer B., Paul F., Pérez-Hernández G., Hoffmann M., Plattner N., Wehmeyer C., Prinz J.-H., and Noé F. 2015. PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *J. Chem. Theory Comput.* 11:5525-5542.
65. Poland D. and Scheraga H. A. (1970) *Theory of helix-coil transitions in biopolymers* (Academic Press, Inc., New York, New York, 10003).
66. Applequist J. and Damle V. 1965. Thermodynamics of the helix-coil equilibrium in oligoadenylic acid from hypochromicity studies. *J. Am. Chem. Soc.* 87:1450-1458.
67. Wartell R. M. and Benight A. S. 1985. Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Phys. Rep.* 126:67-107.
68. Sanstead P. J. and Tokmakoff A. 2019. A lattice model for the interpretation of oligonucleotide hybridization experiments. *J. Chem. Phys.* 150:185104.
69. Sanstead P. J. and Tokmakoff A. 2018. Direct observation of activated kinetics and downhill dynamics in DNA dehybridization. *J. Phys. Chem. B* 122:3088-3100.
70. Jones M. S., Ashwood B., Tokmakoff A., and Ferguson A. L. 2021. 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.
71. Au R. Y., Ng K. S., Chi L. M., and Lam S. L. 2012. Effect of an abasic site on strand slippage in DNA primer-templates. *J. Phys. Chem. B* 116:14781-14787.
72. Isaksson J. and Chattopadhyaya J. 2005. A uniform mechanism correlating dangling-end stabilization and stacking geometry. *Biochem.* 44:5390-5401.
73. Zhang W., Tam C. P., Wang J., and Szostak J. W. 2016. Unusual base-pairing interactions in monomer-template complexes. *ACS Cent. Sci.* 2:916-926.
74. Lemieux S. and Major F. 2002. RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Res.* 30:4250-4263.
75. Di Michele L., Mognetti B. M., Yanagishima T., Varilly P., Ruff Z., Frenkel D., and Eiser E. 2014. Effect of inert tails on the thermodynamics of DNA hybridization. *J. Am. Chem. Soc.* 136:6538-6541.
76. Ouldrige T. E., Šulc P., Romano F., Doye J. P., and Louis A. A. 2013. DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic Acids Res.* 41:8886-8895.
77. Araque J. and Robert M. 2016. Lattice model of oligonucleotide hybridization in solution. II. Specificity and cooperativity. *J. Chem. Phys.* 144:125101.
78. Maciejczyk M., Spasic A., Liwo A., and Scheraga H. A. 2014. DNA duplex formation with a coarse-grained model. *J. Chem. Theory Comput.* 10:5020-5035.
79. Xiao S., Sharpe D. J., Chakraborty D., and Wales D. J. 2019. Energy landscapes and hybridization pathways for DNA hexamer duplexes. *J. Phys. Chem. Lett* 10:6771-6779.

80. Roost C., Lynch S. R., Batista P. J., Qu K., Chang H. Y., and Kool E. T. 2015. Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. *J. Am. Chem. Soc.* 137:2107-2115.
81. Singh S. K., Szulik M. W., Ganguly M., Khutsishvili I., Stone M. P., Marky L. A., and Gold B. 2011. Characterization of DNA with an 8-oxoguanine modification. *Nucleic Acids Res.* 39:6789-6801.
82. Freire E., Schön A., and Velazquez-Campoy A. 2009. Isothermal titration calorimetry: general formalism using binding polynomials. *Methods Enzymol.* 455:127-155.
83. Jost D. and Everaers R. 2009. A unified Poland-Scheraga model of oligo- and polynucleotide DNA melting: salt effects and predictive power. *Biophys. J.* 96:1056-1067.
84. Owczarzy R., Moreira B. G., You Y., Behlke M. A., and Walder J. A. 2008. Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochem.* 47:5336-5353.
85. Zeng Y., Montrichok A., and Zocchi G. 2004. Bubble nucleation and cooperativity in DNA melting. *J. Mol. Biol.* 339:67-75.
86. Applequist J. and Damle V. 1963. Theory of the Effects of Concentration and Chain Length on Helix—Coil Equilibria in Two-Stranded Nucleic Acids. *J. Chem. Phys.* 39:2719-2721.
87. Sim A. Y., Lipfert J., Herschlag D., and Doniach S. 2012. Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle x-ray scattering. *Phys. Rev. E* 86:021901.
88. Ashwood B., Jones M. S., Radakovic A., Khanna S., Lee Y., Sachleben J. R., Szostak J. W., Ferguson A. L., and Tokmakoff A. 2023. Thermodynamics and kinetics of DNA and RNA dinucleotide dehybridization from gaps and overhangs. *Biophys. J.* 122:3323-3339.
89. Mullen K. M., Vengris M., and van Stokkum I. H. 2007. Algorithms for separable nonlinear least squares with application to modelling time-resolved spectra. *J. Glob. Optim.* 38:201-213.
90. Dale J., Howe C. P., Toncrova H., Frittsch R., Greetham G. M., Clark I. P., Towrie M., Parker A. W., McLeish T. C., and Hunt N. T. 2021. 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.
91. Frittsch R., Greetham G. M., Clark I. P., Minnes L., Towrie M., Parker A. W., and Hunt N. T. 2019. Monitoring base-specific dynamics during melting of DNA–ligand complexes using temperature-jump time-resolved infrared spectroscopy. *J. Phys. Chem. B* 123:6188-6199.
92. Kumar A. T., Zhu L., Christian J., Demidov A. A., and Champion P. M. 2001. 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.
93. Bernasconi C. (1976) *Relaxation kinetics* (Academic Press).
94. Cho C., Urquidi J., Singh S., and Robinson G. W. 1999. Thermal offset viscosities of liquid H₂O, D₂O, and T₂O. *J. Phys. Chem. B* 103:1991-1994.
95. Neupane K., Hoffer N. Q., and Woodside M. 2018. Measuring the local velocity along transition paths during the folding of single biological molecules. *Phys. Rev. Lett.* 121:018102.

96. Rauzan B., McMichael E., Cave R., Sevcik L. R., Ostrosky K., Whitman E., Stegemann R., Sinclair A. L., Serra M. J., and Deckert A. A. 2013. Kinetics and thermodynamics of DNA, RNA, and hybrid duplex formation. *Biochem.* 52:765-772.
97. Ashwood B., Sanstead P. J., Dai Q., He C., and Tokmakoff A. 2019. 5-Carboxylcytosine and Cytosine Protonation Distinctly Alter the Stability and Dehybridization Dynamics of the DNA Duplex. *J. Phys. Chem. B* 124:627-640.
98. Dupuis N. F., Holmstrom E. D., and Nesbitt D. J. 2013. Single-molecule kinetics reveal cation-promoted DNA duplex formation through ordering of single-stranded helices. *Biophys. J.* 105:756-766.
99. Pyshnyi D. V., Lomzov A. A., Pyshnaya I. A., and Ivanova E. M. 2006. Hybridization of the bridged oligonucleotides with DNA: Thermodynamic and kinetic studies. *J. Biomol. Struct. Dyn.* 23:567-579.