PNAS INAUGURAL ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Disruption of energetic and dynamic base pairing cooperativity in DNA duplexes by an abasic site

Brennan Ashwood[a,1], Michael S. Jones[b,1], Andrew L. Ferguson[b,2], and Andrei Tokmakoff[a,2]

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2022.
Contributed by Andrei Tokmakoff; received November 9, 2022; accepted February 27, 2023; reviewed by Hashim M. Al-Hashimi, Andrew H. Marcus, and Thomas E. Ouldridge

DNA duplex stability arises from cooperative interactions between multiple adjacent nucleotides that favor base pairing and stacking when formed as a continuous stretch rather than individually. Lesions and nucleobase modifications alter this stability in complex manners that remain challenging to understand despite their centrality to biology. Here, we investigate how an abasic site destabilizes small DNA duplexes and reshapes base pairing dynamics and hybridization pathways using temperature-jump infrared spectroscopy and coarse-grained molecular dynamics simulations. We show how an abasic site splits the cooperativity in a short duplex into two segments, which destabilizes small duplexes as a whole and enables metastable half-dissociated configurations. Dynamically, it introduces an additional barrier to hybridization by constraining the hybridization mechanism to a step-wise process of nucleating and zipping a stretch on one side of the abasic site and then the other.

DNA | hybridization | cooperativity | molecular dynamics | time-resolved spectroscopy

Significance

Biological functions of nucleic acids rely on the stability and dynamics of base pairing interactions. While hydrogen-bonding contacts describe a specific base pair, base pairing stability within duplexes is primarily shaped by the continuous stretches of stacking interactions between adjoining nucleobases that give rise to highly cooperative base pair breaking and formation. Damage or epigenetic modification of a given nucleobase can disrupt this cooperativity and may change the properties of adjoining base pairs. Our investigation used time-resolved spectroscopy and computational modeling to directly reveal how an abasic site alters the stability, kinetics, and dynamics of base pairing in the duplex.

Numerous modifications to canonical base pairs, including epigenetic chemical modifications, mismatches, and lesions, alter local base pairing interactions in double-stranded DNA which may in turn shift global duplex thermodynamic stability by interrupting cooperative interactions. Cooperativity refers to the nonadditive energetic advantages of forming base pairs in a stretch of multiple adjacent nucleotides concertedly rather than individually (1). This phenomenon underlies the quantitatively accurate nearest neighbor (NN) models of DNA thermodynamic stability (2, 3). While cooperativity in base pairing is understood to be rooted in the stacking interactions between nucleobases, the molecular factors bridging local perturbations and global destabilization are not fully understood. Identifying these factors requires site-specific characterization of the hybridization free energy landscape and duplex structural dynamics over wide time windows. This motivates the combination of temperature-jump infrared (T-jump IR) spectroscopy and coarse-grained molecular dynamics (MD) simulations that are employed in this work to access thermodynamic, kinetic, and dynamic insight into how an abasic site disrupts base pairing and stacking in short oligonucleotides.

The loss of a nucleobase to form an abasic site (apurinic/apyrimidinic or AP-site) is among the most naturally abundant DNA modifications due to spontaneous or enzymatic cleavage of the glycosidic bond (4, 5). Previous work showed that an AP-site destabilized the DNA duplex in a sequence-dependent fashion (6–11) that was correlated with the identity of base pairs adjacent to the modified site; however, an AP-site may also disrupt base pairing well beyond nearest neighbors (12). Since duplexes containing an AP-site retain a B-form conformation on average (13–15), an AP-site influences duplex stability and dynamics through disruption of cooperative interactions rather than a large-scale structural change.

We characterize the disruption of cooperativity in various sequence contexts by investigating the impact of central AP-sites in three 11-mer template oligonucleotides that are expected to exhibit wide variation in dynamic behavior due to different relative placement of G:C and A:T base pairs (16–22). T-jump IR spectroscopy separately monitors changes in A:T and G:C base pairing from nanoseconds to milliseconds following a laser T-jump that induces duplex melting. MD simulations employing the 3-site-per-nucleotide (3SPN.2) coarse-grained model (16, 23) are used to parameterize a kinetic Markov state model (MSM) (24, 25) that allows us to interpret the experiments in terms of the temperature-induced reshaping of the duplex free energy landscape. Our results indicate that duplex destabilization primarily stems from the loss of base pairing and stacking interactions around the AP-site as well as disruption of base pairing cooperativity within the duplex. The AP-site splits the base pairing dynamics of the duplex into two segments that must each overcome

Author contributions: B.A., M.S.J., A.L.F., and A.T. designed research; B.A. and M.S.J. performed research; B.A. and M.S.J. analyzed data; and B.A. and A.T. wrote the paper.

Reviewers: H.M.A., Columbia University; A.H.M., University of Oregon; and T.E.O., Imperial College London.

Competing interest statement: A.L.F. is a cofounder and consultant of Evozyne, Inc., and a coauthor of US Patent Application 16/887,710, US Provisional Patent Applications 62/853,919, 62/900,420, and 63/314,898 and International Patent Applications PCT/US2020/035206 and PCT/US2020/050466.

[1] B.A. and M.S.J. contributed equally to this work.

[2] To whom correspondence may be addressed. Email: andrewferguson@uchicago.edu or tokmakoff@uchicago.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2219124120/-/DCSupplemental.

a significant free energy barrier to hybridize. Although our study is limited to oligonucleotides, length-dependent trends in duplex destabilization behavior enable us to address the impact of an AP-site in biological DNA.

Results and Discussion

AP-Site Destabilizes the DNA Duplex through Disruption of Base Pair Cooperativity. The DNA sequences shown in Fig. 1*A* were chosen to vary the position of three or four stabilizing G:C base pairs within an otherwise A:T sequence. Sequence "CGCcap" places them on one end to create asymmetric base pairing stability, "CCends" places them on both ends to minimize terminal base pair fraying (19), and "GCGcore" centers them to promote frayed A:T terminal base pairs (19, 20, 26). For each oligonucleotide, we compare the hybridization of the wild-type (WT) sequence to its complement with one containing a central AP-site at the 6th base position (AP6).

Experiments showing how a central AP-site destabilizes the duplex are presented in Fig. 1 *B–D* and *SI Appendix*, Figs. S1–S5. Duplex melting curves obtained by Fourier-transform infrared spectroscopy (FTIR) show a 20 to 30 °C reduction in melting temperature (T_m) that equates to a drop of 20 to 30 kJ/mol in the dehybridization free energy between AP6 and WT sequences ($\Delta\Delta G^\circ_{d,37} = \Delta G^\circ_{d,37}(AP) - \Delta G^\circ_{d,37}(WT)$), consistent with previous studies (6, 8, 11). Calculations of $\Delta\Delta G^\circ_{d,37}$ using Santa Lucia's NN model (2, 27) merely removing the two central NN interaction parameters involved with the AP-site only account for a portion of the experimental value, suggesting that duplex destabilization is not simply additive in the NN energies (Fig. 1*D*, Section S1.4). An additional free energy penalty affecting both the dehybridization enthalpy (ΔH^o_d) and entropy (ΔS^o_d) must arise from the AP-site (*SI Appendix*, Fig. S1). We also find that $\Delta\Delta G^\circ_{d,37}$ is ~5 kJ/mol larger for CCends-AP6 than CGCcap-AP6 even though the modified and adjacent base pairs to the AP-site are identical in each sequence, suggesting that the oligonucleotide sequence further from the AP-site contributes to $\Delta\Delta G^\circ_{d,37}$.

MD simulations with the 3SPN.2 model and well-tempered metadynamics (WTMetaD) were used to gain insight into the molecular factors governing the thermodynamic stability of various oligonucleotide structures. Structurally 3SPN.2 predicts that duplexes with the AP-site are virtually unchanged from the B-form conformation of the WT duplex and the nucleobase opposite to the AP-site is predominantly intrahelical, consistent with previous experimental studies and all-atom MD simulations (*SI Appendix*, Figs. S6 and S7) (13, 14). The extrahelical configuration of GCGcore-AP6 is negligible, yet there is minor population for CCends-AP6 (0.3%) and CGCcap-AP6 (0.8%). This sequence dependence presumably arises from the twofold larger crossstacking energy between guanine nucleobases relative to thymines in 3SPN.2 (23). The stability of each duplex was assessed using simulations as a function of temperature to compute duplex–fraction melting curves and the Helmholtz free energy for dehybridization $\Delta\Delta F^o_{d,37}$ (Figs. 1*D* and *SI Appendix*, Figs. S8–S13). The simulations show a similar destabilization by the AP-site as in experiment, suggesting that the coarse-grained model is properly accounting for the duplex destabilization by the AP-site. We note that the simulated sequence dependence of $\Delta\Delta F^o_{d,37}$ differs from the experimental $\Delta\Delta G^\circ_{d,37}$, but any differences lie within the uncertainty of the data. Previous studies have verified the accuracy of 3SPN.2 for predicting T_m in canonical duplexes with near-physiological salt concentrations ($[\mathrm{Na}^+]$ = 120 mM) (23), but we find that they are consistently 5 to 10 °C lower than experimental values, which is likely due to the high salt concentrations ($[\mathrm{Na}^+]$ = 600 mM) used in this work.

The computed hybridization free energy profile (FEP) as a function of the number of intact base pairs (n_{bp}) is compared for WT and AP6 sequences in Fig. 2*A*. These profiles overlap for the early steps in hybridization (n_{bp} = 0 to 5 for CGCcap and CCends and n_{bp} = 0 to 2 for GCGcore) and span the barrier to nucleating the

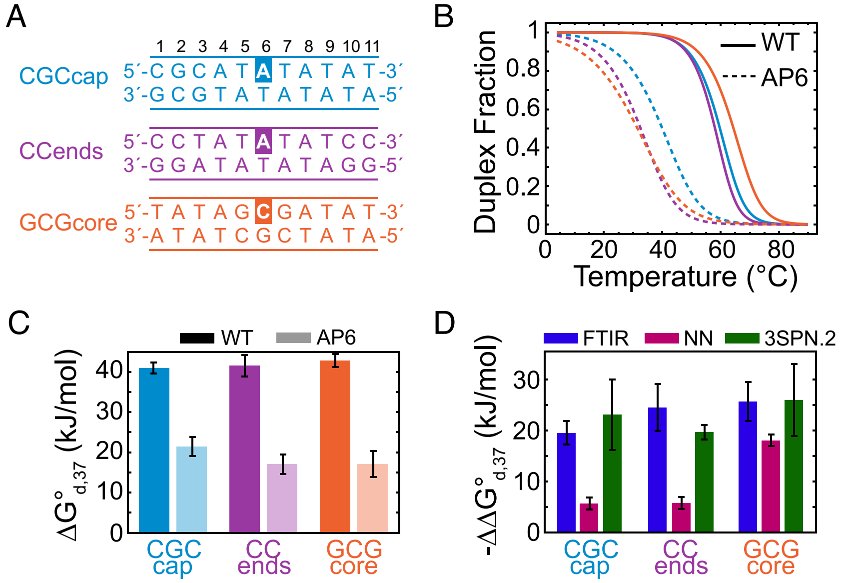


Fig. 1. DNA duplex destabilization induced by an AP-site. (*A*) CGCcap, CCends, and GCGcore sequences studied in this work, identifying the six-position of the AP-site along the top strand (shaded). (*B*) Duplex melting curves determined from two-state analysis of an FTIR temperature series for WT (solid lines) and AP6 (dashed lines) sequences (Section S1.3). (*C*) Duplex dehybridization free energy at 37 °C ($\Delta G^\circ_{d,37}$) determined from FTIR data for WT (dark) and AP6 (light) sequences. (*D*) Drop in $\Delta G^\circ_{d,37}$ for each AP6 sequence with respect to WT sequence ($\Delta\Delta G^\circ_{d,37} = \Delta G^\circ_{d,37}(AP) - \Delta G^\circ_{d,37}(WT)$) compared to those determined from the NN model (Section S1.4). Also shown are the change in dehybridization Helmholtz free energy ($\Delta\Delta F^\circ_{d,37}$) obtained from melting curves generated from 3SPN.2 MD simulations. FTIR and 3SPN.2 error bars correspond to 95% CIs from fits to a two-state model (*SI Appendix*, Fig. S4) while NN error bars correspond to the reported error in the model (2).

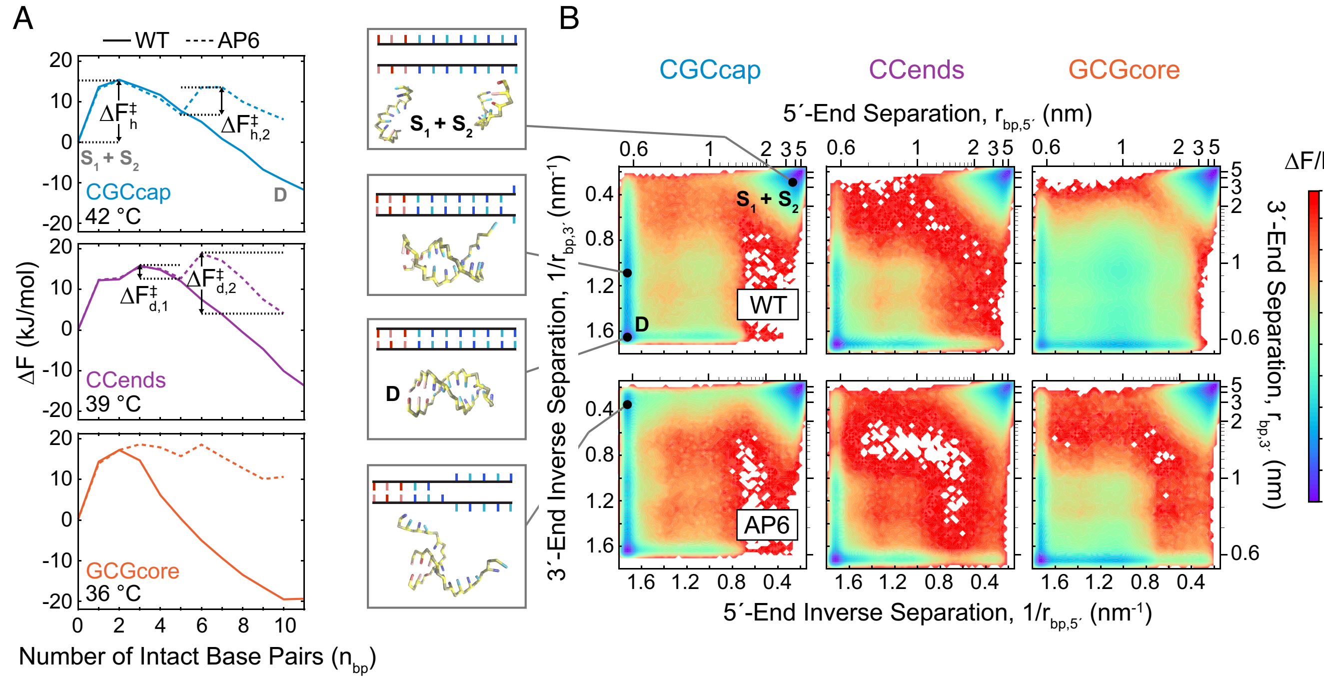


Fig. 2. Influence of AP-site on base pairing free energy landscape. (*A*) Free energy profiles as a function of the number of intact base pairs, $\Delta F(n_{bp}) = -RT\ln P(n_{bp})/P(0)$ for WT (solid lines) and AP6 (dashed lines) sequences extracted from 3SPN.2 MD simulations with WTMetaD at temperatures 10-20°C below $T_{m,MD}$ of each WT sequence. Base pairs are determined using a 0.7-nm radial cutoff between the center of mass of a nucleobase site and that of its complement. Hybridization barriers in n_{bp} for forming the first ($\Delta F_h^\ddagger$) and second ($\Delta F_{h,2}^\ddagger$) base pair segments of the AP6 sequence are indicated for CGCcap. Barriers along the reverse dehybridization direction for dissociating the second ($\Delta F_{d,2}^\ddagger$) and first ($\Delta F_{d,1}^\ddagger$) base pair segments are shown for CCends-AP6. (*B*) Free energy surfaces for all sequences plotted as a function of the average separation between base pairs 1, 3, and 5 on the 5′-end ($r_{bp,5'}$) and 7, 9, and 11 on the 3′-end ($r_{bp,3'}$). 5′- and 3′-ends refer to the top strand in Fig. 1*A*. Free energy (ΔF) is plotted relative to the minimum of the single-strand well ($S_1 + S_2$). Surfaces were generated using 2.5 million frames from 25 unbiased 10 µs simulation trajectories near the 3SPN.2-determined melting temperature of each sequence and are binned and plotted on a r_{bp}^{-1} scale to better resolve different base pairing configurations. Addition of an AP6-site creates local minima (*Top Left* and *Bottom* corners) in the free energy landscape corresponding to configurations with only the 5′- or 3′-end hybridized that makes up the $n_{bp} = 5$ minimum in panel b (*SI Appendix*Fig. S12).

duplex ($\Delta F_h^\ddagger$) at n_{bp} = 2 to 3. The free energy of duplex formation continues downhill with increasing n_{bp} in the WT sequences, as expected for a two-state nucleation and zipping process (28–30), but the AP6 sequences display a minimum at n_{bp} = 5 and must pass through an additional free energy barrier to fully hybridize ($\Delta F_{h,2}^\ddagger$). The dominant base pairing configurations of AP6 sequences at n_{bp} = 5 are those with one side of the AP-site fully intact and the other side completely dissociated, indicating a step-wise process of hybridizing one segment and then the other (*SI Appendix*, Fig. S8). The nucleobase opposite to the AP-site is predominantly intrahelical at n_{bp} = 5. Therefore, an additional free energy penalty for base pair nucleation on the second side of the AP-site contributes to $\Delta\Delta G_d^o$, which is absent in the NN model.

We also observe that $\Delta F_{h,2}^\ddagger < \Delta F_h^\ddagger$, reflecting the difference in hybridization mechanism for the first and second segments. The temperature dependence of $\Delta F_h^\ddagger$ and $\Delta F_{h,2}^\ddagger$ indicates that each arises from an entropic penalty that is partially balanced by a favorable internal energy change for base pairing (*SI Appendix*, Fig. S11). The reduction in translational and orientational entropy for bimolecular association contributes to $\Delta F_h^\ddagger$ but not $\Delta F_{h,2}^\ddagger$, where conformational entropy of the partially bound chains dominates. For a $n_{bp} = 5$ configuration with one fully hybridized segment, the separation of the unbound segment base pairs adjacent to the AP-site is more constrained than the terminal base pairs. As a result, nucleation of the second segment usually occurs by forming a base pair adjacent to the AP-site, as shown below.

Additional sequence-dependent details can be observed in the free energy surfaces (FESs) in Fig. 2*B*, which are computed at the 3SPN.2-determined melting temperature ($T_{m,MD}$) and projected into the average base pair separation (r_{bp}) on either end of the duplex. FESs are binned and plotted on a r_{bp}^{-1} scale to better resolve different basins within the duplex state. Here, 5′- and 3′-sides of the duplex refer to the top strand in Fig. 1*A*. Fully intact sides have $r_{bp} < 0.7$ nm, whereas well-separated strand segments have $r_{bp} = 3$ to 5 nm. Addition of an AP6 site creates metastable configurations with only the 5′- or 3′-segment hybridized. This provides a structural rationalization for the emergence of the $\Delta F_{h,2}^\ddagger$ barriers observed in Fig. 2*A*. The relative stability of 5′-hybridized-to-3′-hybridized configurations qualitatively follows the expected stability of those segments predicted by the NN model (*SI Appendix*, Table S4).

Direct Observation of Internal Nucleation Barrier with T-Jump Spectroscopy. We experimentally test for the presence of $\Delta F_{h,2}^\ddagger$ and a half-hybridized metastable state by characterizing duplex dehybridization with time-resolved T-jump IR spectroscopy (19, 31, 32). We equilibrate the sample at a temperature 15 °C below T_m, where most of the DNA is duplexed and use a 15 °C optical T-jump to heat the solution within 7 ns and promote duplex dehybridization. Changes in base pairing interactions are probed at varying time delays after the T-jump using heterodyned dispersed vibrational echo spectroscopy (HDVE) (31), which provides a transient IR spectrum reporting on changes in nucleobase ring and carbonyl vibrational bands (Fig. 3*A*). The transient spectrum that we report is the change of the spectrum at a given T-jump delay relative to the initial temperature, T_i. Each nucleobase contains a distinct IR fingerprint of ring and carbonyl vibrations (33, 34) that are reshaped upon base pair formation (20, 35), providing separate probes for G:C and A:T

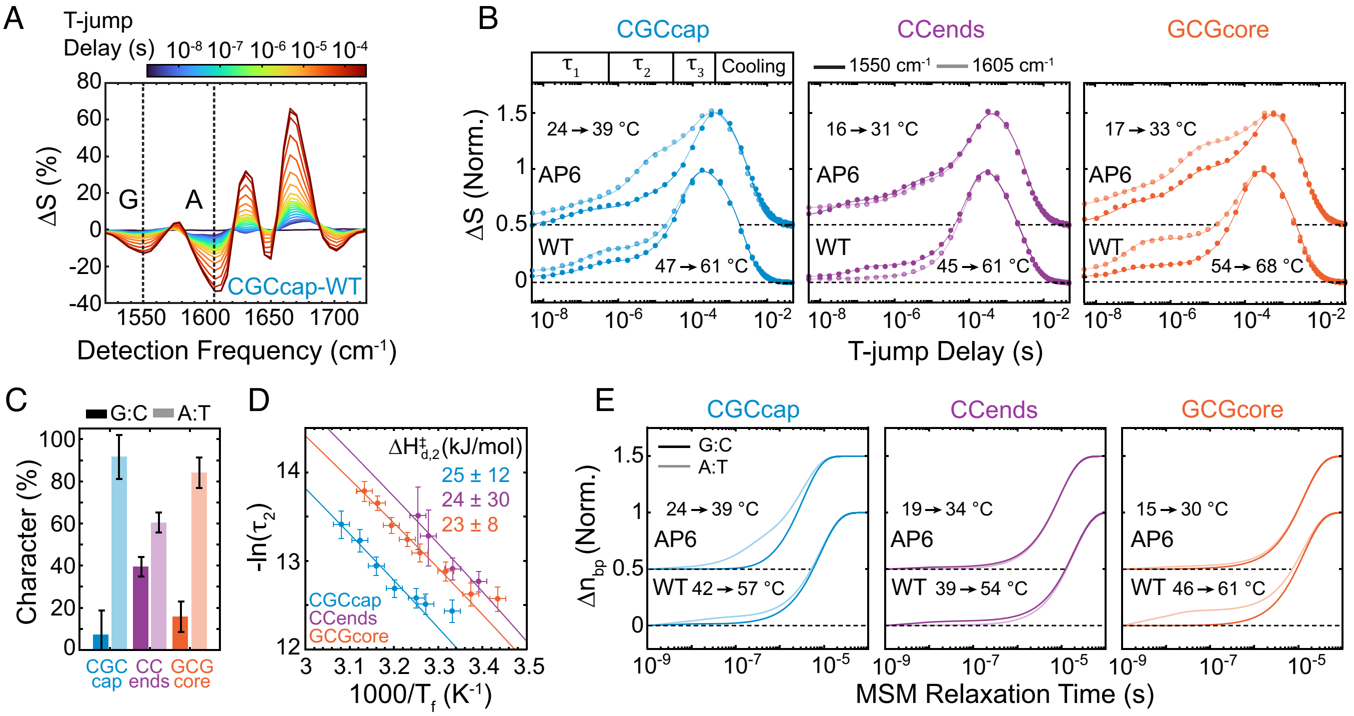


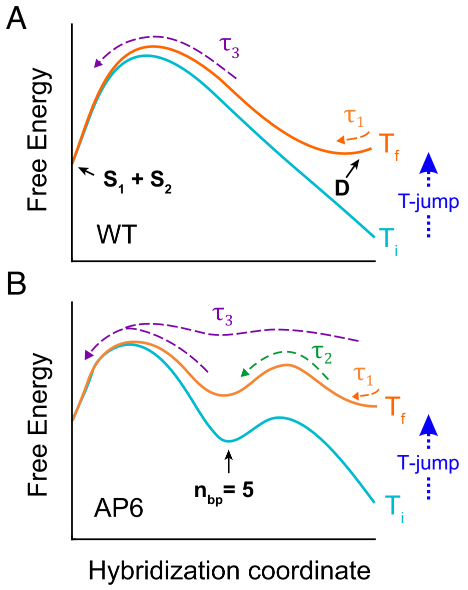
Fig. 3. Direct observation of segment nucleation barrier from T-jump IR spectroscopy. (*A*) T-jump IR difference (t-HDVE) spectra of CGCcap-WT from 5 ns to 1 ms delays after the T-jump, plotted as the change relative to the maximum of the spectrum before the T-jump: $\Delta S(t) = S(t)/\max(S(T_i))$. (*B*) Normalized $\Delta S(t)$ time traces at 1,550 cm^{-1} (dark, G:C) and 1,605 cm^{-1} (light, A:T) cm^{-1} plotted for the CGCcap, CCends, and GCGcore sequences, showing the initial and final temperatures for each sequence ($T_i \rightarrow T_f$). T-jumps are performed from approximately T_m −15 °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. (*C*) Percentage of G:C (dark) and A:T (light) base pair loss character observed in half-dehybridization t-HDVE responses. Percentages are derived from t-HDVE amplitudes of the 1,550 and 1,605 cm^{-1} bands, respectively, determined from global lifetime fitting. Error bars indicate 95% CIs propagated from global fits. (*D*) Temperature-dependent observed rates for half-dehybridization responses (τ_2^{-1}) with fit to a Kramers-like equation in the high friction limit (solid lines, Eq. S27). Vertical error bars correspond to 95% CIs from global lifetime fitting and horizontal error bars are the SD in T-jump magnitude. The enthalpic barrier for half-dehybridization ($\Delta H^{\ddagger}_{d,2}$) is reported. (*E*) Markov state model (MSM) T-jump simulations of the normalized change in intact G:C (dark) and A:T (light) base pairs, $\Delta n_{bp}(t) = [n_{bp}(t) - n_{bp}(T_i)]/[n_{bp}(T_f) - n_{bp}(T_i)]$. Traces are shifted vertically with respect to one another, and dashed lines indicate respective baselines. T-jumps were initialized from a population distribution determined at $T_i = T_{m,MD}$ −15 °C, projected into the microstates of an MSM constructed at $T_f \approx T_{m,MD}$ and allowed to relax to the ensemble determined at T_f. The presence of a base pair is based on a $r_{bp,eq}$ + 0.3 nm radial cutoff between a nucleobase and its complement, where $r_{bp,eq}$ is the equilibrium base pair separation for G:C (0.55 nm) and A:T (0.6 nm) base pairs. The relaxation kinetics are insensitive to small changes in the radial cutoff value (*SI Appendix*, Fig. S28). The MSM relaxation profiles contain two to three kinetic components spanning from 1 ns to 100 μs, and CGCcap-AP6 is the only sequence that exhibits a significant amplitude of the half-dehybridization component that appears from 10 ns to 1 μs.

base pairing using the guanine band at 1,550 cm^{-1} and adenine band at 1,605 cm^{-1} (*SI Appendix*, Figs. S14–S16).

T-jump IR measurements of WT sequences in Fig. 3*B* show two distinct kinetic components on ~100 ns (τ_1) and ~100 μs (τ_3) timescales, followed by thermal relaxation of the sample back to T_i at ~1 ms. The transient IR spectra are well described by global lifetime fitting with exponentially damped spectral components, and similar results are obtained from inverse-Laplace-transform rate distribution spectra (*SI Appendix*, Figs. S17–S19). These processes have previously been characterized in similar sequences and they are illustrated in Scheme 1*A* through changes to the hybridization free energy profile induced by the T-jump (18, 19, 36, 37). The T-jump destabilizes the duplex state, resulting in barrierless terminal base pair fraying (τ_1). The signal change observed during τ_1 increases in order from CCends-WT to CGCcap-WT to GCGcore-WT due to an increased propensity for fraying in A:T base pairs relative to G:C (*SI Appendix*, Figs. S20 and S21). This fast step is followed by complete dissociation into single strands by an activated barrier crossing (τ_3). The temperature dependence of dehybridization rates obtained from τ_3^{-1} follows Arrhenius behavior with a large enthalpic barrier ($\Delta H^{\ddagger}_{d} > 100$ kJ/mol, *SI Appendix*, Fig. S22) that is consistent with full-strand dissociation (30, 38).

The T-jump IR responses of CGCcap-AP6, CCends-AP6, and GCGcore-AP6 reveal an additional kinetic component that occurs on a ~1 μs timescale (τ_2) and corresponds to dehybridization of the full stretch of base pairs on one side of the AP-site (half-dehybridization). Using the contribution of G (1,550 cm^{-1}) and A (1,605 cm^{-1}) to the spectral response during τ_2 relative to the full time window, we determine the fraction of A:T and G:C base pair loss character during τ_2 (Fig. 3*C* and *SI Appendix*, Figs. S23 and S24). The observed A:T character of 60% for CCends-AP6 and 85% for GCGcore-AP6 matches the expectation for dissociating either the 5′- or 3′-half of the duplex based on the fractional A:T content of these palindromic sequences. On the contrary, the 91% A:T character for CGCcap-AP6 indicates a clear preference for dissociating the pure A:T segment. We find that τ_2^{-1} increases exponentially with temperature (Fig. 3*D*) and can also be described by a Kramers-like equation (Eq. S27) as expected for an activated process. Together, the relative A:T/G:C character and activated nature of the τ_2 response directly reveal the presence of a free energy barrier in AP6 sequences for half-dehybridization ($\Delta G^{\ddagger}_{d,2}$), which in the reverse direction corresponds to the barrier for nucleating the remaining base pairs across the AP-site ($\Delta G^{\ddagger}_{h,2}$). Scheme 1b illustrates how the AP-site introduces the metastable half-hybridized state ($n_{bp} = 5$) into the free energy profile, and the resulting kinetics following the T-jump.

To provide a molecular-level interpretation of the T-jump IR responses, we constructed Markov state models (MSMs) from



Scheme 1. Free energy profile reshaping by a T-jump for WT and AP6 sequences. (*A*) An increase in temperature promotes broadening of the duplex free energy minimum (D) by terminal base pair fraying that occurs on a ~100 ns timescale (τ_1, orange arrows). This is followed by cooperative dissociation of the duplex to single strands ($S_1 + S_2$) on a ~100 µs timescale (τ_3, purple arrows). (*B*) In AP6 sequences, cooperative dissociation is split into multiple steps. Following minor terminal fraying (τ_1), base pairs on one side of the AP-site dehybridize on a ~1 µs timescale (τ_2, green arrows) prior to full dissociation of the duplex (τ_3).

250 µs of unbiased 3SPN.2 simulations near $T_{m,MD}$ to furnish kinetic models of the T-jump relaxation behavior (*SI Appendix*, Section S4). Using base pairing assignments for the 200 microstates, we computed the fractional change in the number of intact A:T and G:C base pairs (Δn_{bp}) as a function of relaxation time at $T_{m,MD}$ after initiating the MSM kinetics with a state occupancy obtained at $T_{m,MD} - 15$ °C (Fig. 3*E*). The 3SPN.2 coarse-grained model is known to accelerate computational timescales by ~1 order of magnitude relative to experiment (18), but maintains a clear temporal separation between terminal fraying (~20 ns), half-dehybridization (~500 ns), and complete dehybridization (~10 µs). Consistent with the FEPs in Fig. 2*A*, we observe significant amplitude of τ_2 response for CGCcap-AP6, the only sequence for which $\Delta F^{\ddagger}_{d,2} < \Delta F^{\ddagger}_{d,1}$ at the final temperature used for the simulation. In contrast, $\Delta F^{\ddagger}_{d,2} > \Delta F^{\ddagger}_{d,1}$ for GCGcore-AP6 and CCends-AP6 sequences, leading to overlap of half-dehybridization and complete dissociation timescales and negligible τ_2 response amplitude. Experimentally, half-dehybridization responses are observed for all the three AP6 sequences, so although our computational model predicts a free energy minimum at $n_{bp} = 5$ and accurately captures $\Delta\Delta F^{o}_{d,37}$, we surmise that small errors in the ratio of $\Delta F^{\ddagger}_{d,1}$ to $\Delta F^{\ddagger}_{d,2}$ – potentially due to inaccurate modeling of the local interactions around the AP-site for which the 3SPN.2 model was not directly parameterized – may lead to substantial changes in the kinetic predictions of the computational model. Nevertheless, the internal consistency of the model in resolving significant τ_2 response as a result of $\Delta F^{\ddagger}_{d,2} < \Delta F^{\ddagger}_{d,1}$ provides a mechanistic understanding of the origin of this observed kinetic behavior.

Impact of AP-Site in Longer Oligonucleotides. The global impact of the AP-site across these 11-mer oligonucleotides raises the question of how an AP-site may disrupt cooperativity in longer oligonucleotides. We compared duplexes of A:T-rich sequences similar to GCGcore containing 11, 15, 19, and 27 possible base pairs (*N*) with and without a central AP-site using the same destabilization metrics as above (Fig. 4 and *SI Appendix*, Figs. S29–S33). The observed duplex destabilization, measured with the difference in T_m between AP and WT sequences (ΔT_m), is greatest for the shortest sequences and drops with increasing *N*, but still remains substantial for $N = 27$. In comparison, the NN model yields values of ΔT_m that are roughly half, but qualitatively follows the same trend with *N*. The NN model applies only a

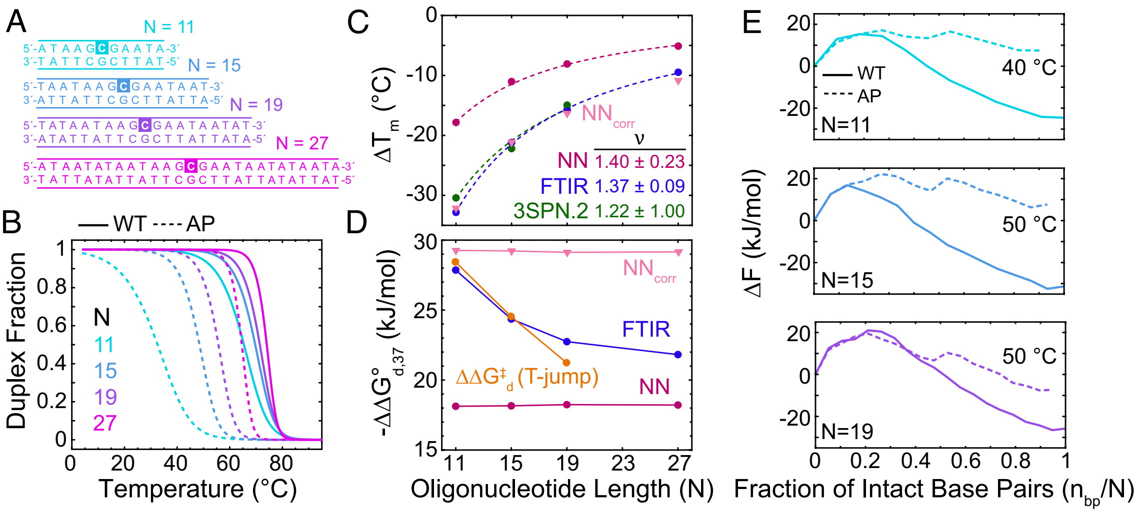


Fig. 4. Influence of oligonucleotide length on duplex destabilization induced by an AP-site. (*A*) GCGcore-like sequences of variable length in terms of number of possible base pairs (*N*), identifying the central AP-site in the top strand (shaded). Sequences contain a GCG motif with a variable number of A:T base pairs on each side leading to $N = 11$, 15, 19, and 27. (*B*) FTIR melting curves for GCGcore-like sequences where solid and dashes lines correspond to WT and AP sequences, respectively. All sequences were measured with a 1:1 strand ratio, total strand concentration of 2 mM, and in pH 6.8* 400 mM SPB. (*C*) $\Delta T_m = T_m(AP) - T_m(WT)$ and (*D*) $-\Delta\Delta G^{o}_{d,37}$ between AP and WT sequences. Values from FTIR melting curves are compared with those from NN calculations. Dashed lines indicate fits to a power law $\Delta T_m(N) = AN^{-\nu}$, and ν is listed where error corresponds to 95% CI from the fit. The NN ΔT_m trend was corrected to the experimental data (NN_{corr}) by fitting with an entropic factor. (*E*) FEPs for WT (solid lines) and AP (dashed lines) N = 11, 15, and 19 sequences extracted from 3SPN.2 MD WTMetaD simulations at the indicated temperatures that are approximately 20 °C below $T_{m,MD}$ of each WT sequence. FEPs are plotted as a function of the fraction of intact base pairs (n_{bp}/N) within a base pair separation cutoff of 0.7 nm.

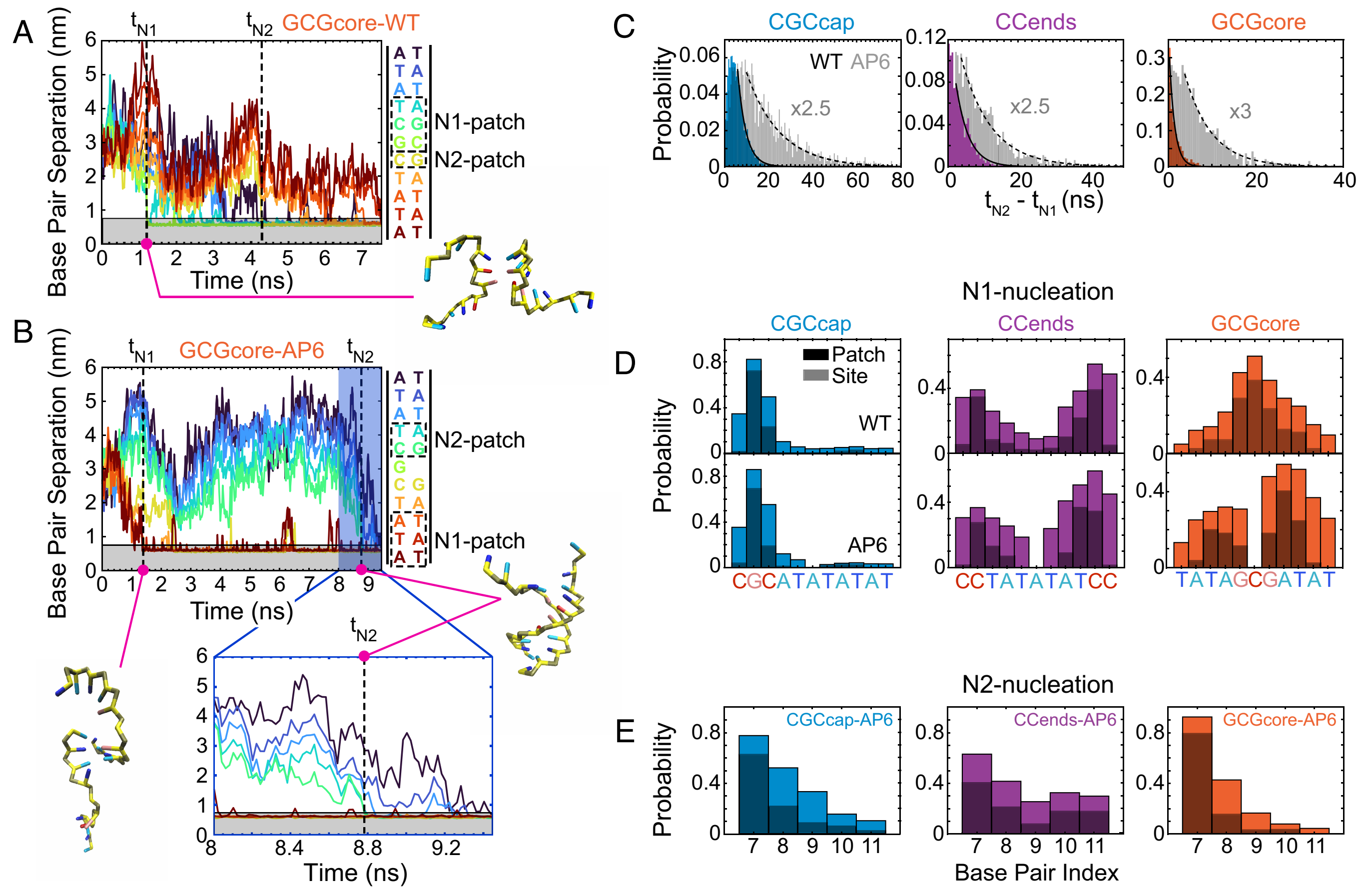


Fig. 5. Splitting of hybridization dynamics into two nucleation events by an AP-site revealed from 3SPN.2 MD simulations. (*A*) Example hybridization trajectory at 20 °C of GCGcore-WT plotting the separation of each in-register base pair contact (color coded). The horizontal solid line indicates the base pair separation cutoff of 0.7 nm, and vertical dashed lines indicate the time points of nucleation (t_{N1} and t_{N2}) in the example trajectories. N1-nucleation is defined as the first permanent formation of any base pair during the hybridization event and N2-nucleation is the first permanent formation of a base pair on the other half of the duplex. N1- and N2-patches correspond to all base pairs with a separation below 0.7 nm at t_{N1} and t_{N2}, respectively. Dashed boxes denote the base pairs within the N1- and N2-patches for the example trajectories. (*B*) Example hybridization trajectory of GCGcore-AP6 at 20 °C. The bottom panel magnifies the portion of the trajectory near t_{N2} for GCGcore-AP6. The hybridization event is declared complete when the average base pair separation reaches 0.6 nm. The GCGcore-WT trajectory exhibits a single nucleation event where t_{N2} is defined just for comparison with GCGcore-AP6, which undergoes a distinct nucleation event for each half of the duplex. (*C*) Probability distribution of time difference between N2- and N1-nucleation (t_{N2}-t_{N1}). Gray distributions correspond to AP6 sequences scaled up vertically by a factor of 2.5 to 3 for clarity. Distribution tails are fit to an exponential decay for WT (solid) and AP6 (dashed) sequences. AP6 sequences exhibit wider t_{N2}-t_{N1} distributions than WT sequences due to the added free energy barrier for N2-nucleation. (*D*) Probability that a given base pair site is part of the N1-patch (colored bars) or the N1-nucleation site (gray bars) determined from >1,000 hybridization trajectories for each sequence. The AP-site reshapes the N1-patch in a sequence-dependent fashion. (*E*) Probability that a base pair site is part of the N2-patch (colored bars) or the N2-nucleation site (gray bars). Base pair index refers to the distance in sequence from the terminal base pair on the duplex half where nucleation occurs (*SI Appendix*, Fig. S34*B*). Regardless of sequence, the most probable site for N2-nucleation is the base pair adjacent to the AP-site. Nucleation properties calculated using a base pair separation cutoff of 1.0 nm are shown in *SI Appendix*, Fig. S35.

constant free energy penalty independent of N to account for the AP-site. We find that this quantity can be corrected to nearly reproduce the experimental data by fitting the NN ΔT_m trend to that from FTIR using an entropic factor of 34.5 J/molK. This correction corresponds to a free energy penalty of −10.7 kJ/mol at 37 °C that approximately matches the difference between NN and experimental $\Delta\Delta G^{\circ}_{d,37}$ values for sequences with the GCG motif. This observation suggests that the AP-site defect introduces a single thermodynamic penalty that is "diluted out" with increasing N. In contrast to the trends in ΔT_m, the experimental $\Delta\Delta G^{\circ}_{d,37}$ decreases with N (Fig. 4*D*) by an amount dependent on temperature (*SI Appendix*, Fig. S29*d*). This could be a nonadditive contribution to the AP-site destabilization, which may reflect that the fraction of $\Delta G^{\circ}_{d,37}$ lost by incorporating an AP-site ($\Delta\Delta G^{\circ}_{d,37}/\Delta G^{\circ}_{d,37}(WT)$) is much higher for short sequences than long ones.

The sizeable influence of a single lesion on the stability of a 27-mer duplex led us to investigate the predictions of the NN model for much larger oligonucleotides (*SI Appendix*, Fig. S29 *E* and *F*). These calculations predict that ΔT_m does not drop to an offset, but appears to follow power law behavior, $\Delta T_m \sim N^{-\nu}$, with $\nu \sim 1.1$ in A:T-rich sequences. The same power law behavior with $\nu \sim 1.4$ fits the experimental ΔT_m trend (Fig. 4*C*), and similar behavior is calculated for $\Delta\Delta G^{\circ}_{d,37}/\Delta G^{\circ}_{d,37}(WT)$. Of course, this NN calculation does not account for the fact that long stretches of A:T base pairs (>20 bp) may form bubbles (39), or that an AP-site could significantly impact bubble behavior as observed for base pairing mismatches (40). The ΔT_m trend is expected to be sensitive to sequence and A:T/G:C content. For example, NN calculations predict ΔT_m to approach zero more sharply with an exponent of $\nu \sim 3.5$ when extending a CCends template with GC steps due to the greater $\Delta G^{\circ}_{d,37}$ associated with GC steps.

For comparison, 3SPN.2 simulations employing WTMetaD were performed for the GCGcore-like sequences of length $N = 11$, 15, and 19. We find that one central AP-site has a negligible effect on the end-to-end distance of duplex DNA, regardless of N, but does have a small lengthening effect on single strands that becomes negligible as N increases (*SI Appendix*, Fig. S32). The computed

FEPs as a function of the fraction of intact base pairs have a similar free energy barrier $\Delta F^{\ddagger}_{h,2}$ at $n_{bp}/N = 0.5$ with a barrier height of ~4 kJ/mol for all lengths (Fig. 4*E*), indicating that the disruption of hybridization cooperativity and second nucleation penalty arising from $\Delta F^{\ddagger}_{h,2}$ persists even in the long duplex limit.

While measurable in the N = 11 AP sequence, the half-dehybridization T-jump response decreases in magnitude in the N = 15 AP sequence and is undetectable for N = 19 (*SI Appendix*, Fig. S33*A*), indicating that $\Delta G^{\ddagger}_{d,2}$ increases and ultimately exceeds $\Delta G^{\ddagger}_{d,1}$ for large N. Similar to $\Delta\Delta G^{\circ}_{d,37}$, a decrease with N is observed in $\Delta\Delta H^{\circ}_{d}$ (*SI Appendix*, Fig. S29*A*) as well as $\Delta\Delta G^{\ddagger}_{d,37}$ from T-jump measurements (Fig. 4*D*). As indicated by the FEPs in Fig. 4*E*, these observations indicate that the role of half-hybridized states in duplex stability decreases as N increases, yet remains important even for oligonucleotides far longer than those investigated here.

Disruption of Cooperativity in Hybridization Dynamics. Both T-jump measurements and computed FEPs reveal the presence of free energy barriers for hybridizing each half of AP6 sequences, but they do not directly describe if or how an AP-site affects the dynamics of recognition and binding during hybridization. Even among studies of unmodified DNA, many of the early events during hybridization are poorly understood, and it is clear that oligonucleotides may hybridize via multiple mechanisms (16–18, 41). We examined the dynamics of base pairing from thousands of hybridization events generated by 3SPN.2 MD simulations at 20 °C for each 11-mer sequence. Trajectories of example hybridization events for GCGcore-WT and -AP6 sequences are shown in Fig. 5 *A* and *B*, plotting the separation between all in-register base pairs as a function of time. Trajectories are trimmed to the time interval between the first base pair separation distance falling below 2.5 nm and when all distances fall below 0.7 nm. From each trajectory, we isolate the time point for first permanent base pair formation, the N1-nucleation point t_{N1}, and define a nucleation patch (N1-patch) as any intact base pairs present at that point, whether they are transient or permanent.

As illustrated in Fig. 5 *A* and *B*, WT sequences exhibit a single nucleation event—here through a N1-patch of three adjacent bases—followed by concerted formation of the remaining base pairs on either side of the patch within a few nanoseconds in a classic nucleation and zipping process (30, 38). In contrast, AP6 sequences tend to hybridize one half of the duplex and remain in the half-hybridized state with frayed termini for many nanoseconds before nucleating the second half (N2-nucleation) at t_{N2}. This is illustrated through the changes in the distribution of $t_{N1} - t_{N2}$ time intervals between WT and AP6 sequences in Fig. 5*C* where t_{N2} is defined for WT sequences as the time point for nucleating a base pair on the half opposite from where t_{N1} occurs. These dynamics are consistent with the presence of two barriers to hybridization, $\Delta F^{\ddagger}_{h}$ and $\Delta F^{\ddagger}_{h,2}$, and suggest that n_{bp} is a good coordinate to describe hybridization.

N1- and N2-nucleation events have qualitatively different properties that strongly depend on nucleobase sequence. Quantifying the N1-nucleation position and N1-patch size shows that initial nucleation most likely occurs near G:C base pairs with a N1-patch spread over 1 to 5 contacts (Fig. 5*D*). A central AP-site localizes the N1-patch into one half of the duplex (*SI Appendix*, Fig. S34), which is most disruptive to GCGcore and reshapes the nucleation free energy barrier as shown in Fig. 2*A*. The N1-patch of CCends and CGCcap is less perturbed by a central AP-site, and its corresponding WT and AP6 FEPs have nearly identical profiles from n_{bp} = 0 to 5.

N2-nucleation typically occurs with all base pairs in the first half intact, which constrains the dissociated segment structure relative to single-stranded DNA (42, 43) and reduces the barrier to N2-nucleation. As a result, N2-nucleation is most probable adjacent to the AP-site but with characteristics that depend on the identity of the base pair adjacent to the AP-site as well as the sequence of the entire segment (Fig. 5*E*). GCGcore-AP6 demonstrates that an adjacent G:C base pair promotes N2-nucleation next to the AP-site relative to sequences with an adjacent A:T base pair and does not require as many base pairs in the N2-patch (*SI Appendix*, Fig. S34*C*). Although each have an adjacent A:T base pair, the terminal G:C base pairs of CCends-AP6 promote a higher probability of N2-nucleation away from the AP-site than in CGCcap-AP6 and exemplify that N2-nucleation dynamics depend on the sequence of the entire segment, not just the nucleobases adjacent to the AP-site.

Conclusions

Our study combines T-jump IR spectroscopy and coarse-grained MD simulations to reveal a hierarchy of consequences arising from a single AP-site in duplex DNA, spanning thermodynamic destabilization, hybridization pathways, and melting kinetics. An AP-site reduces the global cooperativity by effectively splitting the base pairing energetics and dynamics of the duplex into two segments. The level of impact is sequence dependent where both the loss of base pairing interactions and height of the barrier for hybridizing across the AP-site depend on the identity of surrounding base pairs. This disruption of base pairing energetics and dynamics may be important for the recognition of AP-sites by AP endonuclease (5, 44, 45) and subsequent repair that plays a critical role in maintaining genome integrity (46). Additionally, damage-induced AP-sites may form in clusters (47, 48) that will likely amplify the magnitude of the disruption observed from a single AP-site in this work.

The hallmark of DNA duplex stability is the nonlinearity of cooperative interactions that depend on the concerted formation of multiple continuous contacts, but we see that even one perturbation in this cooperativity can unravel duplex stability and base pairing dynamics. We have focused on an AP-site, but there are noncanonical base pairs and epigenetic and damage-induced nucleobase modifications that impose an energetic penalty similar to that of an AP-site ($-\Delta\Delta G^{\circ}_{d,37}$ = 5 to 20 kJ/mol) and likely disrupt base pairing cooperativity within the duplex (3, 49–53). Base pairing stability and dynamics themselves tune the interactions between DNA and other molecules yet also give rise to mechanical dynamics and distortions that facilitate protein recognition (54, 55).

The contribution of cooperativity to stability is recognized not only among nucleic acids but for any binding partners that rely on a continuous stretch of interactions between subunits such as for protein–protein association and self-assembly of material and biological subunits. Even though the nature of these interactions is completely different, a disruption through mutation or defect may reshape stability and dynamics, for better or for worse.

Materials and Methods

Oligonucleotide Preparation. DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT) at desalt-grade purity. AP-sites were incorporated as a tetrahydrofuran group (dSpacer). Samples were purified with 3 kDa centrifugal filters (Amicon). All oligonucleotides were prepared in pH* 6.8 400 mM sodium phosphate buffer (SPB, [Na^{+}] = 600 mM). This high salt concentration was chosen to assure that the complete melting transition for all duplexes could be observed. For FTIR spectroscopy, samples were prepared in deuterated SPB solution to avoid spectral interference with the H_2O bending vibrational band. Oligonucleotide concentration was verified with UV absorbance using a NanoDrop UV/Vis spectrometer (Thermo Scientific).

Temperature-Dependent FTIR Spectroscopy. All samples used for FTIR spectroscopy measurements were prepared at a 1:1 ratio of complementary strands and a total strand concentration of 2 mM. This concentration was required in order to reach duplex absorbance values of 20 to 40 mOD from 1,550 to 1,720 cm^{-1}, which is ~6% of the D_2O solvent background absorbance. 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 prior to each measurement to ensure oligonucleotides start in their lowest energy conformation. A discrete wavelet transform using the Mallat algorithm and symlet family was applied to the 1,490 to 1,750 cm^{-1} region of the FTIR spectra to separate and subtract the D_2O background absorption from the data (56, 57).

Isothermal Titration Calorimetry. Isothermal titration calorimetry (ITC) measurements were performed using a MicroCal iTC200 (Malvern Panalytical). Single-stranded oligonucleotides were prepared in the cell with a concentration of 10 to 30 μM and the complement strand was placed in the syringe with a concentration of 100 to 300 μM. All samples were prepared in nondeuterated pH 6.8 400 mM SPB and degassed under vacuum for >20 min at the experimental temperature prior to each measurement. Sample conditions for each measurement are listed in *SI Appendix*, Table S1 and additional experimental details are provided in *SI Appendix, Section S1.2*. The heat of dilution was measured for the titrant at each experimental temperature and subtracted from integrated thermogram of each measurement.

Temperature-Jump IR Spectroscopy. The temperature-jump (T-jump) IR spectrometer used in this work has been described previously (32, 58). The T-jump magnitude ($\Delta T = T_f$ to T_i) was set to ~15 °C for all measurements. The initial temperatures (T_i) were set using a recirculating chiller connected to a brass sample jacket, and ΔT was determined from the change in mid-IR transmission following the T-jump pulse. The T-jump response of the DNA samples is probed using transient heterodyned dispersed vibrational echo spectroscopy (t-HDVE) (31) with parallel (ZZZZ) pulse polarization and fixed at a constant waiting time (t_2) of 150 fs. We report the real part of the t-HDVE spectrum, which contains similar information to an IR pump-probe spectrum.

Molecular Dynamics Simulations with 3SPN.2. Molecular dynamics (MD) simulations were performed using the coarse-grained 3-Site-Per-Nucleotide (3SPN.2) model (23). Although not parameterized against sequences containing an AP-site, 3SPN.2 is directly amenable to modification for AP-site representation and simulation. Since backbone dihedral potentials are determined solely by sugar and phosphate sites, removal of one base bead does not intrinsically disrupt the backbone structure. Nonbonded interactions are constructed anisotropically as a function of the angles formed between proximate intra- and inter-strand bases and their bonded sugars. The targeted removal of one base eliminates nonbonded interactions with that base without disrupting nonbonded interactions between all other intact bases. In this way, base pairing and stacking interactions in the vicinity of the AP-site are minimally perturbed.

All systems were simulated using the LAMMPS package (59) compiled with the 3SPN.2 model plugin (23). Simulations were performed in the 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 (60) with a 20 fs integration time step. Solvent was treated implicitly with an experimentally informed friction coefficient of 9.94×10^{-11} m^2/s (23, 61). A 600 mM implicit salt concentration was used, and electrostatic interactions were treated using the Debye–Hückel equation with a 5-nm cutoff radius (62). All simulations were initialized from WT B-DNA duplex states generated by the 3SPN.2 software. Free energies obtained from probability densities are reported as the Helmholtz free energy ΔF.

Enhanced Sampling of Hybridization Free Energy Landscape. We employed different sampling methods depending on the hybridization property of interest. To efficiently sample the duplex-to-single-strand thermodynamic free energy landscape at various temperatures, we employed well-tempered metadynamics (WTMetaD) via the PLUMED plugin (63) to enhance sampling of hybridized, dissociated, and intermediate states by accelerating the dynamics in a predefined collective variable (CV) (64). For each system, we selected a CV that described the average distance between the beads representing native Watson–Crick–Franklin 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 and was analytically reweighted to report unbiased thermodynamic averages and free energy landscapes in various order parameters beyond those in which sampling was conducted. Full details of the WTMetaD simulation protocol is provided in *SI Appendix, Section S2.2*.

Markov State Models. Markov state models (MSMs) are long-time kinetic models that provide quantitative and interpretable descriptions of the thermodynamic stability and dynamical transitions between various metastable states of a molecular system (24, 25, 65–67). We constructed MSMs from 25 independent and unbiased 10 μs molecular dynamics simulations for each of the WT and AP6 DNA systems. Simulations were performed near the sequence-dependent T_m calculated from the WTMetaD simulations ($T_{m,MD}$). Initial velocities were assigned from a Maxwell–Boltzmann distribution at $T_{m,MD}$. Each simulation was conducted for 12 μs and frames saved to disc every 100 ps, resulting in 1-2 hybridization and dehybridization events per trajectory. The first 2 μs of each run was discarded to produce 25 × 10 μs = 250 μs (2.5 million frames) of simulation data for each sequence. An additional 25 × 1 μs of simulation data was generated for each system 15 °C below $T_{m,MD}$ in order to furnish a low-temperature initial distribution for MSM relaxation experiments. The 2D FESs shown in Fig. 2*B* were produced by reweighting unbiased trajectories according to the MSM microstate transition matrix. Full details of MSM construction and validation as well as T-jump relaxation simulations can be found in *SI Appendix, Section S4*.

Hybridization Trajectories. To efficiently sample hybridization events to characterize nucleation behavior (Fig. 5), 1,000 to 2,500 simulations of each sequence were initiated from a dissociated configuration, generated by shifting equilibrium duplexes by 2 nm in the *x*, *y*, and *z* directions, and were declared complete when reaching a defined duplex state cutoff of $r_{bp} = 0.6$ nm, which corresponds to a fully hybridized duplex. An umbrella potential with a force constant k = 100 kJ/nm^2mol was applied for $r_{bp} > 3$ nm to prevent diffusive separation of the oligonucleotide strands. We verified that this cutoff was sufficiently large that it did not influence the binding and association events by confirming negligible differences in the hybridization mechanism when comparing trajectories harvested under restraining potentials with $r_{bp} > 3$ nm, k = 100 kJ/nm^2mol and $r_{bp} > 4$ nm, k = 10 kJ/ nm^2mol. Trajectories were trimmed to analyze the time window between when at least one in-register base pair separation distance falls below 2.5 nm to when all base pair separation distances fall below 0.7 nm.

Supporting Information. Temperature-dependent FTIR data; description of melting curve determination from two-state model; details of enhanced WTMetaD 3SPN.2 simulation and melting curve determination; t-HDVE spectral assignments; description of global lifetime fitting and determination of A:T/G:C character in τ_2 response; description of two-state kinetic analysis of τ_3 response; description of MSM construction and T-jump relaxation simulations; FTIR and t-HDVE kinetic data for oligonucleotide length series.

Data, Materials, and Software 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/ (68). Scripts for running equilibrium simulations and building Markov State models are available at https://github.com/mrjoness/abasic-kinetics/ (69). All PLUMED input scripts were submitted to the PLUMED-NEST public repository and are available at https://www.plumed-nest.org/eggs/22/037/ (70). All unbiased and biased MD trajectories (65 GB), time- and rate-domain t-HDVE data, FTIR temperature series, and experimental and 3SPN.2-determined melting curves are uploaded to Zenodo and available at 10.5281/zenodo.7199303 (71).

ACKNOWLEDGMENTS. This work was supported by the National Institute of General Medical Sciences of the NIH (Award No. R01-GM118774) and the NSF under Grant No. CHE-2152521. B.A. acknowledges support from the NSF Graduate Research Fellowship Program. B.A. thanks Yumin Lee and Lukas Whaley-Mayda for helpful discussions and providing feedback on the manuscript. B.A. thanks Elena Solomaha of the University of Chicago BioPhysics Core Facility for training on isothermal titration calorimetry instrumentation. 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 graphics processing unit (GPU)-based cyberinfrastructure supported by the NSF under Grant No. DMR-1828629.

Author affiliations: [a]Department of Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, Chicago, IL 60637; and [b]Pritzker School of Molecular Engineering, The University of Chicago, Chicago, IL 60637

1. K. Dill, S. Bromberg, *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience* (Garland Science, 2010).
2. J. SantaLucia, A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Nat. Acad. Sci. U.S.A.* **95**, 1460–1465 (1998).
3. J. SantaLucia Jr., D. Hicks, The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 415–440 (2004).
4. T. Lindahl, Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
5. P. S. Thompson, D. Cortez, New insights into abasic site repair and tolerance. *DNA Repair* **90**, 102866 (2020).
6. C. A. Gelfand, G. E. Plum, A. P. Grollman, F. Johnson, K. J. Breslauer, Thermodynamic consequences of an abasic lesion in duplex DNA are strongly dependent on base sequence. *Biochem.* **37**, 7321–7327 (1998).
7. C. A. Minetti *et al.*, Impact of bistrand abasic sites and proximate orientation on DNA global structure and duplex energetics. *Biopolymers* **109**, e23098 (2018).
8. J. Sági, A. B. Guliaev, B. Singer, 15-mer DNA duplexes containing an abasic site are thermodynamically more stable with adjacent purines than with pyrimidines. *Biochem.* **40**, 3859–3868 (2001).
9. H. Ide *et al.*, Influence of alpha-deoxyadenosine on the stability and structure of DNA. Thermodynamic and molecular mechanics studies. *Biochem* **34**, 6947–6955 (1995).
10. D. V. Pyshnyi, A. A. Lomzov, I. A. Pyshnaya, E. M. Ivanova, Hybridization of the bridged oligonucleotides with DNA: Thermodynamic and kinetic studies. *J. Biomol. Struct. Dyn.* **23**, 567–579 (2006).
11. G. Vesnaver, C.-N. Chang, M. Eisenberg, A. P. Grollman, K. J. Breslauer, 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. U.S.A.* **86**, 3614–3618 (1989).
12. M. J. Morten, S. G. Lopez, I. E. Steinmark, A. Rafferty, S. W. Magennis, Stacking-induced fluorescence increase reveals allosteric interactions through DNA. *Nucleic Acids Res.* **46**, 11618–11626 (2018).
13. J. Chen, F.-Y. Dupradeau, D. A. Case, C. J. Turner, J. Stubbe, DNA oligonucleotides with A, T, G or C opposite an abasic site: Structure and dynamics. *Nucleic Acids Res.* **36**, 253–262 (2008).
14. J. Zálešák, M. Lourdin, L. Krejčí, J.-F. Constant, M. Jourdan, Structure and dynamics of DNA duplexes containing a cluster of mutagenic 8-oxoguanine and abasic site lesions. *J. Mol. Biol.* **426**, 1524–1538 (2014).
15. M. Lukin, C. de Los Santos, NMR structures of damaged DNA. *Chem. Rev.* **106**, 607–686 (2006).
16. D. M. Hinckley, J. P. Lequieu, J. J. de Pablo, Coarse-grained modeling of DNA oligomer hybridization: Length, sequence, and salt effects. *J. Chem. Phys.* **141**, 035102 (2014).
17. T. E. Ouldridge, P. Šulc, F. Romano, J. P. Doye, A. A. Louis, DNA hybridization kinetics: Zippering, internal displacement and sequence dependence. *Nucleic Acids Res.* **41**, 8886–8895 (2013).
18. M. S. Jones, B. Ashwood, A. Tokmakoff, A. L. Ferguson, Determining sequence-dependent DNA oligonucleotide hybridization and dehybridization mechanisms using coarse-grained molecular simulation, Markov state models, and infrared spectroscopy. *J. Am. Chem. Soc.* **143**, 17395–17411 (2021).
19. P. J. Sanstead, A. Tokmakoff, Direct observation of activated kinetics and downhill dynamics in DNA dehybridization. *J. Phys. Chem. B* **122**, 3088–3100 (2018).
20. P. J. Sanstead, P. Stevenson, A. Tokmakoff, Sequence-dependent mechanism of DNA oligonucleotide dehybridization resolved through infrared spectroscopy. *J. Am. Chem. Soc.* **138**, 11792–11801 (2016).
21. M. Schickinger, M. Zacharias, H. Dietz, Tethered multifluorophore motion reveals equilibrium transition kinetics of single DNA double helices. *Proc. Nat. Acad. Sci. U.S.A.* **115**, E7512–E7521 (2018).
22. S. Hertel *et al.*, The stability and number of nucleating interactions determine DNA hybridization rates in the absence of secondary structure. *Nucleic Acids Res.* **50**, 7829–7841 (2021).
23. D. M. Hinckley, G. S. Freeman, J. K. Whitmer, J. J. De Pablo, An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **139**, 144903 (2013).
24. V. S. Pande, K. Beauchamp, G. R. Bowman, Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **52**, 99–105 (2010).
25. C. Wehmeyer *et al.*, Introduction to Markov state modeling with the PyEMMA software. *Living J. Comput. Mol. Sci.* **1**, 5965 (2019).
26. J. Araque, M. Robert, Lattice model of oligonucleotide hybridization in solution. II. Specificity and cooperativity. *J. Chem. Phys.* **144**, 125101 (2016).
27. R. Owczarzy, B. G. Moreira, Y. You, M. A. Behlke, J. A. Walder, Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochem.* **47**, 5336–5353 (2008).
28. D. Poland, H. A. Scheraga, *Theory of helix-coil transitions in biopolymers* (Academic Press Inc, 1970).
29. P. A. Thompson, W. A. Eaton, J. Hofrichter, Laser temperature jump study of the helix⇌ coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model. *Biochem.* **36**, 9200–9210 (1997).
30. M. E. Craig, D. M. Crothers, P. Doty, Relaxation kinetics of dimer formation by self complementary oligonucleotides. *J. Mol. Biol.* **62**, 383–401 (1971).
31. K. C. Jones, Z. Ganim, A. Tokmakoff, Heterodyne-detected dispersed vibrational echo spectroscopy. *J. Phys. Chem. A* **113**, 14060–14066 (2009).
32. H. S. Chung, M. Khalil, A. W. Smith, A. Tokmakoff, Transient two-dimensional IR spectrometer for probing nanosecond temperature-jump kinetics. *Rev. Sci. Instrum.* **78**, 063101 (2007).
33. C. S. Peng, K. C. Jones, A. Tokmakoff, Anharmonic vibrational modes of nucleic acid bases revealed by 2D IR spectroscopy. *J. Am. Chem. Soc.* **133**, 15650–15660 (2011).
34. M. Banyay, M. Sarkar, A. Gräslund, A library of IR bands of nucleic acids in solution. *Biophys. Chem.* **104**, 477–488 (2003).
35. C. Lee, M. Cho, Vibrational dynamics of DNA. II. Deuterium exchange effects and simulated IR absorption spectra. *J. Chem. Phys.* **125**, 114509 (2006).
36. R. Fritzsch *et al.*, Monitoring base-specific dynamics during melting of DNA–ligand complexes using temperature-jump time-resolved infrared spectroscopy. *J. Phys. Chem. B* **123**, 6188–6199 (2019).
37. J. Dale *et al.*, Combining steady state and temperature jump IR spectroscopy to investigate the allosteric effects of ligand binding to dsDNA. *Phys. Chem. Chem. Phys.* **23**, 15352–15363 (2021).
38. D. Pörschke, M. Eigen, Co-operative non-enzymatic base recognition III. Kinetics of the helix–coil transition of the oligoribouridylic· oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *J. Mol. Biol.* **62**, 361–381 (1971).
39. Y. Zeng, A. Montrichok, G. Zocchi, Bubble nucleation and cooperativity in DNA melting. *J. Mol. Biol.* **339**, 67–75 (2004).
40. Y. Zeng, G. Zocchi, Mismatches and bubbles in DNA. *Biophys. J.* **90**, 4522–4529 (2006).
41. Y. Yin, X. S. Zhao, Kinetics and dynamics of DNA hybridization. *Acc. Chem. Res.* **44**, 1172–1181 (2011).
42. B. Israels *et al.*, Submillisecond conformational transitions of short single-stranded DNA lattices by photon correlation single-molecule förster resonance energy transfer. *J. Phys. Chem. B* **125**, 9426–9440 (2021).
43. P. Qu, X. Yang, X. Li, X. Zhou, X. S. Zhao, Direct measurement of the rates and barriers on forward and reverse diffusions of intramolecular collision in overhang oligonucleotides. *J. Phys. Chem. B* **114**, 8235–8243 (2010).
44. C. D. Mol, T. Izumi, S. Mitra, J. A. Tainer, DNA-bound structures and mutants reveal abasic DNA binding by APE1 DNA repair and coordination. *Nature* **403**, 451–456 (2000).
45. A. A. Kuznetsova *et al.*, Substrate specificity of human apurinic/apyrimidinic endonuclease APE1 in the nucleotide incision repair pathway. *Nucleic Acids Res.* **46**, 11454–11465 (2018).
46. A. Huskova, B. Landova, E. Boura, J. Silhan, The rate of formation and stability of abasic site interstrand crosslinks in the DNA duplex. *DNA Repair* **113**, 103300 (2022).
47. E. Bignon *et al.*, Correlation of bistranded clustered abasic DNA lesion processing with structural and dynamic DNA helix distortion. *Nucleic Acids Res.* **44**, 8588–8599 (2016).
48. A. G. Georgakilas, P. O'Neill, R. D. Stewart, Induction and repair of clustered DNA lesions: What do we know so far? *Radiat. Res.* **180**, 100–109 (2013).
49. S. K. Singh *et al.*, Characterization of DNA with an 8-oxoguanine modification. *Nucleic Acids Res.* **39**, 6789–6801 (2011).
50. R. C. A. Dubini *et al.*, 1H NMR chemical exchange techniques reveal local and global effects of oxidized cytosine derivatives. *ACS Phys. Chem. Au* **2**, 237–246 (2022).
51. M. T. Woodside *et al.*, Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* **314**, 1001–1004 (2006).
52. I. I. Cisse, H. Kim, T. Ha, A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nat. Struct. Mol. Biol.* **19**, 623–627 (2012).
53. A. Rangadurai *et al.*, Measuring thermodynamic preferences to form non-native conformations in nucleic acids using ultraviolet melting. *Proc. Nat. Acad. Sci. U.S.A.* **119**, e2112496119 (2022).
54. A. Afek *et al.*, DNA mismatches reveal conformational penalties in protein–DNA recognition. *Nature* **587**, 291–296 (2020).
55. T. Ngo *et al.*, Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* **7**, 1–9 (2016).
56. P. M. Ramos, I. Ruisánchez, Noise and background removal in Raman spectra of ancient pigments using wavelet transform. *J. Raman Spectrosc.* **36**, 848–856 (2005).
57. S. Mallat, A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 674–693 (1989).
58. K. C. Jones, Z. Ganim, C. S. Peng, A. Tokmakoff, Transient two-dimensional spectroscopy with linear absorption corrections applied to temperature-jump two-dimensional infrared. *J. Opt. Soc. Am. B* **29**, 118–129 (2012).
59. S. Plimpton, Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
60. G. Bussi, M. Parrinello, Accurate sampling using Langevin dynamics. *Phys. Rev. E* **75**, 056707 (2007).
61. A. E. Nkodo *et al.*, Diffusion coefficient of DNA molecules during free solution electrophoresis. *Electrophoresis* **22**, 2424–2432 (2001).
62. P. Debye, E. Hückel, De la theorie des electrolytes. I. abaissement du point de congelation et phenomenes associes. *Physikalische Zeitschrift* **24**, 185–206 (1923).
63. G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
64. A. Barducci, G. Bussi, M. Parrinello, Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 020603 (2008).
65. U. Sengupta, M. Carballo-Pacheco, B. Strodel, Automated Markov state models for molecular dynamics simulations of aggregation and self-assembly. *J. Chem. Phys.* **150**, 115101 (2019).
66. J. D. Chodera, F. Noé, Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **25**, 135–144 (2014).

67. B. E. Husic, V. S. Pande, Markov state models: From an art to a science. *J. Am. Chem. Soc.* **140**, 2386–2396 (2018).
68. M. S. Jones, Abasic-thermo. GitHub. https://github.com/mrjoness/abasic-thermo/. Deposited 26 October 2021.
69. M. S. Jones, Abasic-kinetics. GitHub. https://github.com/mrjoness/abasic-kinetics/. Deposited 1 March 2022.
70. M. S. Jones, plumID:22.037, PLUMED-NEST. https://www.plumed-nest.org/eggs/22/037/. Deposited 16 October 2022.
71. B. Ashwood, M. S. Jones, A. L. Ferguson, A. Tokmakoff, Supporting data for disruption of energetic and dynamic base pairing cooperativity in DNA duplexes by an abasic site. Zenodo. https://zenodo.org/record/7199304#.ZBJNIXbMJaQ. Deposited 14 October 2022.