

Thermodynamics and kinetics of DNA and RNA dinucleotide hybridization to gaps and overhangs

Brennan Ashwood,^{1,2} Michael S. Jones,³ Aleksandar Radakovic,⁴ Smayan Khanna,³ Yumin Lee,^{1,2} Joseph R. Sachleben,⁵ Jack W. Szostak,¹ Andrew L. Ferguson,³ Andrei Tokmakoff^{1,2,*}

¹Department of Chemistry, The University of Chicago, 5735 S. Ellis Avenue, Chicago, IL 60637

²The James Franck Institute and Institute for Biophysical Dynamics, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, United States

³Pritzker School of Molecular Engineering, The University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, United States

⁴Department of Genetics, Harvard Medical School, Boston, MA 02115

⁵Biomolecular NMR Core Facility, Biological Sciences Division, The University of Chicago, Chicago, IL 60637, United States

*Correspondence: tokmakoff@uchicago.edu

Abstract

Hybridization of short nucleic acid segments (<4 nucleotides) to single-strand templates occurs as a critical intermediate in processes such as non-enzymatic nucleic acid replication and toehold-mediated strand displacement. These templates often contain adjacent duplex segments that stabilize base pairing with single-strand gaps or overhangs, but the thermodynamics and kinetics of hybridization in such contexts are poorly understood due to experimental challenges of probing weak binding and rapid structural dynamics. Here we develop an approach to directly measure the thermodynamics and kinetics of DNA and RNA dinucleotide dehybridization using steady-state and temperature-jump infrared spectroscopy. Our results suggest that dinucleotide binding is stabilized through coaxial stacking interactions with the adjacent duplex segments as well as from potential non-canonical base pairing configurations and structural dynamics of gap and overhang templates revealed using molecular dynamics simulations. We measure timescales for dissociation ranging from 0.2 to 40 μ s depending on the template and temperature. Dinucleotide hybridization and dehybridization involves a significant free energy barrier with characteristics resembling that of canonical oligonucleotides. Together, our work provides an initial step for predicting the stability and kinetics of hybridization between short nucleic acid segments and various templates.

Significance

Processes in biology and nanotechnology rely on hybridization of nucleic acid segments as short as mononucleotides. In these contexts, the binding stability and timescale of dehybridization ultimately inhibit or enable biochemical reactions or structural rearrangements to occur after binding. Understanding the energetics and time-dependence of short nucleic acid hybridization, particularly to larger templates, is critical for understanding processes such as non-enzymatic RNA replication and toehold-mediated strand displacement. We demonstrate an approach for extracting temperature-dependent energetics and kinetics of dinucleotide hybridization to gap and overhang templates using steady-state and time-resolved infrared spectroscopy. Our findings indicate that binding between a dinucleotide and template is stabilized beyond standard hybridization, and the methods established here can be applied broadly to various sequences and templates.

Introduction

DNA and RNA duplex hybridization has been investigated for more than sixty years with an aim of developing predictive models for how thermodynamic and kinetic properties will vary with molecular factors such as strand length, sequence, and chemical modifications.(1-3) The development of numerous models, such as the quantitatively accurate nearest-neighbor (NN) models(4, 5) for thermodynamics or the kinetic-zipper model,(6, 7) shape the modern understanding of nucleic acid properties as well as the development of nucleic acid technology.(8, 9) However, most of these models were developed for oligonucleotides, and the hybridization properties of nucleic acid segments shorter than six nucleotides in length have largely been neglected. Hybridization of such short segments is often perceived as irrelevant due to their poor binding stability in aqueous solution, yet there are biological processes where segments as short as mononucleotides must hybridize. Many of these processes, such as the binding of 2'-deoxynucleotide triphosphates to single-strand DNA (ssDNA) during strand replication or binding between three nucleotide anticodons of tRNA and their complement mRNA, utilize additional protein-nucleic acid interactions to promote binding.(10, 11) However, processes such as non-enzymatic replication of nucleic acids and toehold-mediated strand displacement require hybridization of 1-4 base pair patches without the help of proteins.(8, 12-14)

A conserved feature among the examples of short oligonucleotide hybridization is that a short segment binds to a single-strand overhang adjacent to a duplex region or in a gap between

two duplex segments. The adjacent duplexes offer coaxial stacking interactions that stabilize hybridization of the short oligonucleotide. The thermodynamic benefit of coaxial stacking has been measured for DNA and RNA oligonucleotides and is often found to be similar to stacking in B-DNA or A-RNA, respectively.(15-20) The structural constraints of the sugar-phosphate backbone at a nick site are relaxed relative to a covalently-linked base pair step and may enable more stabilizing stacking configurations.(21, 22) Additionally, the single-strand segments in overhang and gap templates may have different physical properties from free single-strands such as reduced configurational flexibility, increased stacking, and different hydration that influence the stability and dynamics of binding with short oligonucleotides. A few studies have shown evidence for increased rigidity of gap and overhang single-stranded regions relative to free single-strands as well as bending of duplexes containing 1 or 2 nucleotide gaps,(23-29) and these properties are likely highly sensitive to nucleobase sequence and cation concentrations.(30-32)

The binding stability of mononucleotides with overhang and gap templates has recently been studied. NMR-monitored titrations of mononucleotides with overhangs revealed dissociation constants (K_d) ranging from ~10 mM for G:C base pairing to ~200 mM for A:T base pairing.(33, 34) Isothermal titration calorimetry (ITC) and FRET-monitored titrations have shown that the K_d for binding onto a gap decreases nearly 3 orders of magnitude to values of 0.2-0.6 mM,(35, 36) which corresponds to a ~12 kJ mol⁻¹ increase in the dissociation free energy (ΔG_d°) and is consistent with forming an additional nearest-neighbor step.(4, 37) These studies give a sense for the stability of short oligonucleotide hybridization to overhangs and gaps and indicate that it depends on sequence and whether DNA or RNA is used, yet, with the exception of ITC, the thermodynamic information is limited to dissociation constants (K_d or ΔG_d°) at the measurement temperature. A detailed understanding of hybridization onto overhangs and gaps requires knowledge of enthalpic and entropic contributions to stability and the role of relevant environmental variables such as temperature and counterion concentration.(38-40) More importantly, no studies have directly monitored the time-dependence of hybridization and dehybridization for short oligonucleotides onto templates, and these timescales determine whether reactions such as non-enzymatic extension or toehold-mediated strand displacement are possible.

Here we report the temperature-dependent thermodynamics and kinetics of short oligonucleotide dehybridization from overhangs and gaps using steady-state and temperature-

jump infrared (T-jump IR) spectroscopy. IR spectroscopy is sensitive to base pairing and stacking interactions and distinctly resolves changes in A:T and G:C base pairing without the need for synthetic labels.(41, 42) Characterization of temperature-dependent binding stability enables accurate determination of enthalpic and entropic contributions and is particularly relevant in non-enzymatic replication where temperature variations may play a key role.(38, 39) We focus on the binding of DNA and RNA dinucleotides and demonstrate our approach with a model system of an adenine-adenine (AA) dinucleotide binding next to pure G:C duplex regions. This model system was chosen to maximize spectral contrast between binding of the dinucleotide and the structural changes and dynamics within the template itself, but our approach is applicable for any nucleobase sequence. We find quantitative agreement between thermodynamic results obtained from IR spectroscopy and those from complementary temperature-dependent ¹H NMR and ITC measurements of binding. From T-jump IR measurements, we extract a time constant ranging from 200 ns to 40 μs for dissociation of AA from the template that depends on the temperature and template, while the time constant for association is a few microseconds and shows a minor dependence on temperature. We also study the binding of 3-mer (AAA) and 4-mer (AAAA) adenine DNA oligonucleotides to assess the length-dependence of association in this short oligonucleotide regime. Our experimental results as well as all-atom molecular dynamics (MD) simulations suggest that differences in the structure and dynamics of the bound complex and template compared to free oligonucleotides lead to significantly different binding thermodynamics and kinetics for overhangs and gaps.

Materials and Methods

Synthesis and purification of RNA oligonucleotides

RNA oligonucleotides were synthesized in house on a K&A H-6 solid-phase synthesizer. Phosphoramidites and reagents for the K&A synthesizer were purchased from Glen Research and Chemgenes. 1.5 mL of a 1:1 mixture of 28 % aqueous ammonium hydroxide and 40 % aqueous methylamine was used to cleave the RNA from the solid support for 15 minutes at room temperature, and was followed by deprotection in the same solution for 15 minutes at 65 °C. After cooling, the solutions were evaporated for 1 hour in a SpeedVac vacuum concentrator and lyophilized to dryness. The dried 2'-O-tertbutyldimethylsilyl (2'-O-TBDMS) protected RNA was dissolved in 115 μL DMSO, 60 μL triethylamine, and 75 μL triethylamine trihydrofluoride and

deprotected for 2.5 hours at 65 °C. After cooling, the RNA was purified by two different methods depending on its length:

Short (≤ 6 nucleotides) oligonucleotides were purified by a modified Glen-Pak protocol (Glen Research). After the RNA was quenched, bound to the resin, and washed to remove failed sequences, it was eluted with 50 % acetonitrile/water v/v. Following evaporation to dryness, the RNA was dissolved in 500 μ L 2 % trifluoroacetic acid to deprotect the 5'-O-dimethoxytrityl (5'-O-DMT) for 5 minutes at room temperature. The deprotection reaction was quenched with 500 μ L 1 M triethylammonium bicarbonate pH 9 and desalted using Sep-Pak C18 cartridges (Waters).

Longer (> 7 nucleotides) oligonucleotides were precipitated with 25 μ L 5 M ammonium acetate and 1 mL 2-propanol at -78 °C for 20 minutes, washed with 1 mL 80 % ethanol in water v/v, air-dried, dissolved in 5 mM EDTA, 99 % formamide in water v/v, and purified by preparative 20 % polyacrylamide gel electrophoresis (19:1 with 7 M urea). After crushing and soaking the desired gel band in 5 mM sodium acetate, 2 mM EDTA for 16 hours at room temperature, the extracted RNA was desalted using Sep-Pak C18 cartridges (Waters).

Purified RNA was characterized by high-resolution mass spectrometry on an Agilent 6540 Q-TOF mass spectrometer.

Preparation of DNA oligonucleotides

DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT) at desalt grade purity. Short (≤ 6 nucleotides) oligonucleotides were further desalted using Sep-Pak C18 cartridges (Waters), evaporated for 1 hour in a SpeedVac vacuum concentrator, and lyophilized to dryness. Longer (> 7 nucleotides) oligonucleotides were dialyzed in ~ 1.5 L ultrapure water for 24 h using Slide-A-Lyzer cassettes (2 kDa cutoff, Thermo Scientific), where the water was replaced every ~ 6 hrs, and then lyophilized to dryness. For all measurements, oligonucleotides were prepared in deuterated pH* 6.8 400 mM sodium phosphate buffer (SPB, $[\text{Na}^+] = 600$ mM). Oligonucleotide concentration was verified with UV absorbance using a NanoDrop UV/Vis spectrometer (Thermo Scientific).

FTIR spectroscopy

FTIR spectra were measured with a Bruker Tensor 27 FTIR spectrometer at 2 cm^{-1} resolution. 30 μ L of sample was placed between two 1 mm thick CaF_2 windows with a pathlength defined by a 50 μ m Teflon spacer. The windows were enclosed in brass jacket that is temperature-

controlled using a recirculating chiller (Ministat 125, Huber). The temperature at the sample was measured with a thermocouple in thermal contact to one CaF₂ window for chiller set points from -3 to 105 °C, which was used to calibrate the reported temperatures in this work.

Prior to titration experiments performed at 1 °C, the gap or overhang template solution was incubated at 85 °C for 2 min and cooled to room temperature under ambient conditions. Samples were prepared with an AA-to-template molar ratio ranging from 0 to 4 with a constant template concentration of 1 mM. Samples were incubated at 1 °C for 10 min prior to each measurement. For FTIR temperature series, all oligonucleotides were prepared at a 1 mM concentration and annealed prior to measurements as for titration measurements. Temperature series were performed in ~2.6 °C steps and the sample was equilibrated for 3 min at each temperature prior to acquiring spectra. 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 each spectrum.(43)

Two-dimensional IR Spectroscopy

Two-dimensional IR (2D IR) measurements were performed using a previously described setup with a pump-probe beam geometry.(44) 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. Samples were prepared identically as for FTIR measurements. Sample temperature was controlled using a recirculating chiller (Ministat 125, Huber). Temperature-dependent drift of the sample out of the pump and probe focus and overlap region was minimized by programmatically maximizing the pump-probe signal at each temperature using a motorized translation stage prior to measuring the 2D IR data.

Temperature-jump IR Spectroscopy

The T-jump IR spectrometer used in this work has been described previously.(45, 46) Optical heating increases the sample temperature within 7 ns and the ensuing changes in base pairing and stacking of DNA and RNA are monitored with heterodyned dispersed vibrational echo spectroscopy (HDVE)(47) from nanosecond to millisecond time delays. The HDVE spectrum is collected with parallel pulse polarization (ZZZZ) and at a fixed waiting time (t_2) of 150 fs and reports on the nucleobase ring and carbonyl vibrational bands similarly to an IR pump-probe spectrum. The T-jump laser power was attenuated with a polarizer to set a T-jump magnitude

($\Delta T = T_f - T_i$) of ~ 13 °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.

Given the complex form of the T-jump relaxation kinetics spanning multiple time scales, we chose a model-free analysis method to extract experimental relaxation rates. The time-domain t-HDVE response at different IR frequencies was converted into a rate spectrum using an inverse Laplace transform that employs a maximum entropy regularization method (MEM-iLT) which is described elsewhere.^(48, 49) We calculate the observed relaxation rate, k_{obs} , for a given process by calculating the amplitude-weighted mean value of the rate for a corresponding peak in the rate spectrum.

Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) measurements were performed using a MicroCal iTC200 (Malvern Panalytical). Template oligonucleotides were placed in the cell with a concentration that varied from 60 μ M to 3 mM depending on the sequence and the complement short oligonucleotide was placed in the syringe with a concentration of 100-300 μ M. Sample conditions for each measurement are listed in Table S2. All samples were prepared in deuterated pH 6.8 400 mM SPB and degassed under vacuum for >20 min at the experimental temperature prior to each measurement. ITC measurements started with an initial 0.4 μ L injection followed by 19 injections of 2 μ L aliquots of the titrant oligonucleotide solution into the cell. The syringe needle constantly stirred the cell solution with a spin rate of 1000 rpm, and the injection duration and time interval between injections were adjusted to avoid signal saturation in high concentration samples. Injection settings for each sample are listed in Table S2.

NMR Spectroscopy

NMR spectra were acquired on a Bruker AVANCE III 500 MHz spectrometer equipped with a Bruker TXI probe. Temperature series were performed in 2.5 °C steps and the sample was equilibrated for 5 min and auto gradient and lock shimmed with TopShim at each temperature. prior to acquiring spectra. Samples were prepared in fully deuterated 400 mM SPB for all measurements and contained ~ 1 mM 3-(Trimethylsilyl)propionic-2,2,3,3-d₄ acid (Sigma-Aldrich) as a frequency reference.

Total correlation spectroscopy (TOCSY) measurements were measured with DIPSI II isotropic mixing and a mixing time of 60 ms. 2048 and 256 complex points were acquired in t_2 and t_1 , respectively, over sweep widths of 24 ppm.

Diffusion-ordered spectroscopy (DOSY) was performed on a Bruker AVANCE IIIHD 600 MHz spectrometer. Measurements were performed using 2D stimulated echo pulse sequence with bipolar gradients and WATERGATE solvent suppression (Bruker TopSpin, steppgp1s19). Spectra were acquired from 0 to 95% of the maximum gradient field strength in 5 % intervals where the maximum field strength was 5.35 T/cm. The magnetic field gradient was calibrated with a 3D printed phantom with known spacing between the water samples.

Molecular dynamics simulations

DNA and RNA duplex topologies were constructed using AMBERTools22(50) in canonical B and A forms, respectively. AA-gap complexes were generated from the canonical topologies by removing phosphate atoms after the 6th (5'-GA-3') and 8th position (5'-AG-3') nucleotides and adding terminal 5' and 3' hydrogens using the GROMACS molecular modeling suite.(51) Gap templates were generated by entirely removing the A nucleotides from the canonical topologies. MD simulations were performed with the force fields that are currently most accurate for modelling canonical DNA and RNA duplex structure: The bsc1-AMBER force field(52) with TIP3P(53) water model for DNA systems, and the DES-AMBER RNA force field(54) with TIP4P water model(55) for RNA systems. To enable more direct comparisons between DNA and RNA simulation results, we performed additional simulations of DNA using the DES-AMBER force field for DNA(56) with TIP4P water model and simulations of RNA using the bsc1-AMBER force field with TIP3P water model (Section S5). Cubic simulation boxes with periodic boundaries were constructed to ensure a minimum of 1 nm of spacing between any nucleic acid atoms and the boundaries. For 14-mer DNA this corresponded to a box with 7.16 nm side lengths with 11770 water molecules, and for 14-mer RNA a box with 6.77 nm side lengths with 9845 water molecules. NaCl ion pairs were added to the simulation box to maintain an ionic concentration of 600 mM, and additional Na⁺ ions were added to balance the negatively charged phosphate backbone.

Energy minimization was performed on each topology using the steepest descent algorithm until the maximum force was below 1000 kJ mol⁻¹ nm⁻¹. Equilibration was performed for 100 ps in the NVT ensemble and then 100 ps in the NPT ensemble. Production runs were performed for

up to 4 μ s in the NVT ensemble. Simulations were propagated with a 2 fs time step using a leap-frog integrator, and frames were saved every 200 ps. Simulation temperature was set to 310 K and maintained by a velocity-rescaling thermostat.(57) In the NPT equilibration runs, pressure was set to 1 bar and maintained using a Parinello-Rahmen barostat. Particle Mesh Ewald(58) was used to calculate long-range electrostatic interactions, employing a 1.0 nm real space cutoff and a 0.16 Fourier grid spacing that was optimized during runtime, and constraints were placed on hydrogen bonds using the LINCS algorithm.(59)

Cambridge conventional helical parameters(60) and backbone torsions were calculated from MD trajectories using a Biobb BioExcel Building Blocks workflow(61, 62), which leverages the Curves+ program and its associated Canal tool(63, 64). Intra-molecular residue distances and solvent distributions were calculated using the MDTraj Python library. Additional statistical analyses and visualizations were generated using the PyEMMA and Seaborn Python libraries.(65, 66)

Results and Discussion

Low-temperature hybridization of dinucleotides to overhangs and gaps

Figure 1a illustrates the model sequences that we designed to study the hybridization of a 2'-deoxyadenosine dinucleotide onto G:C rich DNA templates incorporating a single-stranded thymine overhang or thymine gap. Overhang templates (TTo) are composed of a 6-mer primer strand bound to the 3' end of a 14-mer template strand. Gap templates also have a 6-mer helper strand bound to the 5'-end of the template. Additionally, we studied the binding of adenosine dinucleotide onto a single-stranded uracil RNA gap template (UUg). The abbreviation AA will be used to refer to both the 2'-deoxyadenosine dinucleotide for DNA and adenosine dinucleotide for RNA samples throughout. The primer and helper segments consist of six guanine:cytosine (G:C) base pairs in order to maximize their binding stability while also keeping their size relatively small to minimize IR absorption from the template itself. The hybridized product between the template and AA is termed the AA-overhang (TTo:AA) or AA-gap complex (TTg:AA, UUg:AA). Although formation of a fully formed gap or overhang complex can potentially involve numerous binding equilibria, the choice of a weakly bound AA to a far more stable GC-rich template means that the

dissociation of the complex is well described through two equilibrium constants: K_d for the dissociation of AA from the template and $K_{d,Temp}$ for dissociation of the template.

We first test for evidence that AA binds to gaps and overhangs at low temperature (1 °C) using FTIR spectroscopy. Figure 1 compares the FTIR spectra in the 1540 – 1720 cm^{-1} region of isolated TTg, TTo, and UUG templates with samples containing an equimolar mixture of template and AA. This spectral region contains in-plane carbonyl and ring stretching vibrations of the nucleobases that are sensitive to base pairing and stacking interactions and have previously been studied in detail.(41, 42, 67, 68) The templates are fully intact at 1 °C such that spectral changes from adding AA report on binding of AA to the template. Each mixed sample exhibits an increase in absorbance near 1575 and 1625 cm^{-1} resulting from the A ring vibrations of AA. The 1650 – 1700 cm^{-1} region primarily contains carbonyl vibrations. The shifting of intensity from 1665 to 1685 cm^{-1} indicates increased stacking and base pairing of guanines adjacent to the gap, and the overlapping gain in intensity of the thymine carbonyl band near 1695 cm^{-1} indicates the formation of A:T base pairing.(41) The intensity of the 1625 cm^{-1} band directly reports on the stacking and base pairing interactions experienced by AA, and its suppression in TTg:AA and UUG:AA relative to TTo:AA indicates greater binding of AA in the gap at 1°C. Further, FTIR-monitored titrations of AA with each template indicate an increase in fraction of TTg bound to AA at equimolar conditions ($\theta_{1:1}$) progressing from TTo:AA to UUG:AA to TTg:AA (Figs. 1e & S1). K_d decreases by over an order of magnitude in switching the template from an overhang to a gap as previously observed for binding of guanosine monophosphate (GMP) to RNA.(35)

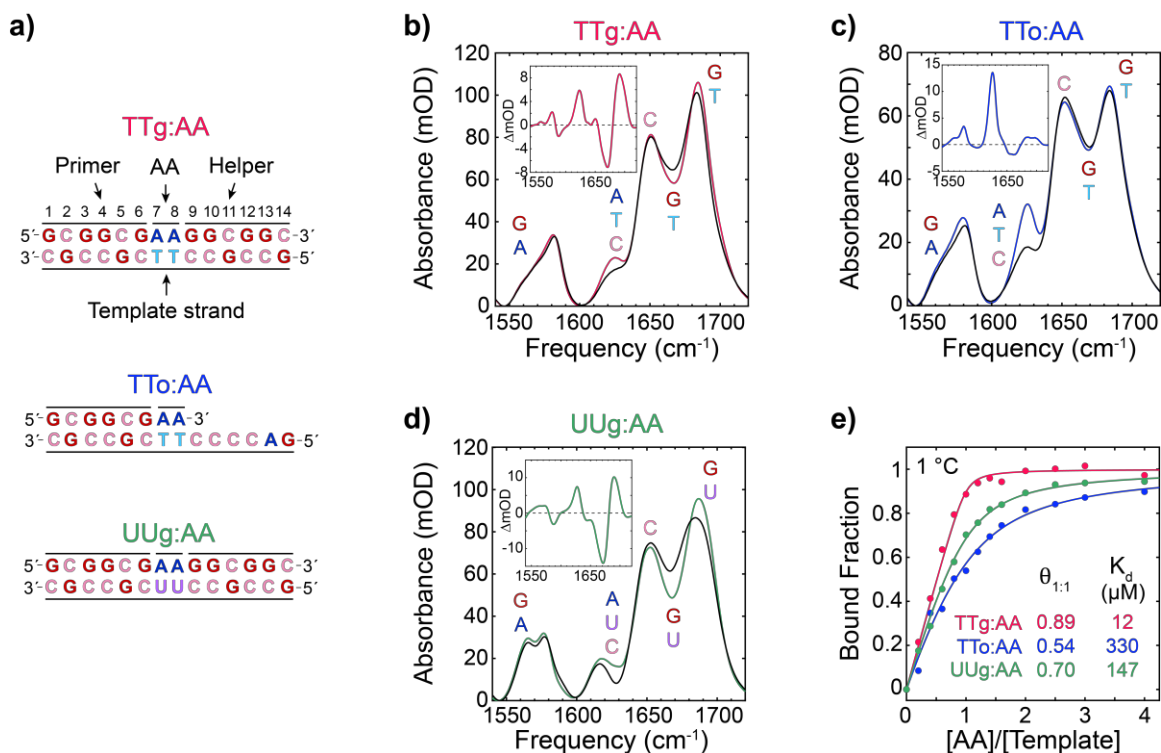


Figure 1. Low temperature hybridization of AA dinucleotide to gap and overhang. (a) Model sequences studied in this work. Gap templates are comprised of a 14-mer template strand bound to 6-mer primer and 6-mer helper strands whereas the overhang templates are only bound to the primer strand. (b, c, d) FTIR spectra at 1 °C of 1 mM gap and overhang templates in pH* 6.8 400 mM sodium phosphate buffer (SPB) without AA (black lines) and after adding 1 mM AA (color lines). Difference spectra upon addition of AA are shown as insets. Labels indicate the nucleobases contributing to each frequency region of the FTIR spectra. (e) Titrations of AA against 1 mM TTg, TTo, and UUg templates at 1 °C monitored with FTIR. The fraction of template bound to AA is determined from fitting the second singular value decomposition (SVD) component of FTIR data between 1650 – 1720 cm^{-1} to a two-state binding model (solid lines, eq. S2). The fraction of AA bound to template at a 1:1 molar AA:template ratio ($\theta_{1:1}$) and the dissociation constant (K_d) are listed for each complex. Additional spectra and fitting details are provided in Section S1.1.

AA binding equilibrium monitored with temperature-dependent IR and NMR spectroscopy

To accurately measure temperature-dependent AA binding stability, IR and NMR spectroscopic probes are employed to monitor changes in base pairing as a function of temperature that are then modelled to extract melting curves for AA dissociation. FTIR temperature series were performed to track AA binding stability from 1 to 96 °C (Fig. 2a,b). Both AA-gap complexes show two melting transitions that correspond to dissociation of AA from the template followed by dissociation of the primer and helper from the template strand. The AA dissociation transition in TTg:AA and UUg:AA is observed between 1 and 60 °C as an increase in intensity of the adenine ring mode at 1625 cm^{-1} and the change of carbonyl bands at 1685-1695 and 1665 cm^{-1} . The sharper

dissociation transition of the primer and helper – characterized exclusively by changes in guanine ring and carbonyl bands – occurs at higher temperature as a single step melting transition since the binding stability of the primer and helper to the template strand are nearly identical. The midpoint for primer and helper dissociation is shifted 13 °C higher in UUg:AA due to the more stabilizing nature of GC base pair steps in A-RNA relative to B-DNA.(4, 5) AA unbinding in TTo:AA is already halfway complete at 1 °C as indicated in Fig. 1e and is completed by ~40 °C (Fig. S2).

The sigmoidal character of the AA thermal dissociation transition is superimposed on other slowly varying temperature-dependent changes, primarily a result of solvation. We account for additional independent spectral changes by measuring FTIR temperature series of TTg, TTo, and UUg and fully complementary duplexes (TTgd, UUgd) for reference, as described in Section S1.2. In contrast, we find that these linear background changes are only minor contributions in temperature series monitored by two-dimensional IR (2D IR) spectroscopy (Section S1.3). The 2D IR temperature series show profiles for AA dissociation and primer and helper dissociation transitions that correlate closely the transitions observed in FTIR data but with reduced linear baseline contribution to the sigmoidal transitions. We also obtain similar AA dissociation and primer and helper dissociation transitions from ^1H NMR of TTg:AA (Figs. 2b and Section S1.4), lending confidence to the profiles of the IR-monitored dissociation transitions.

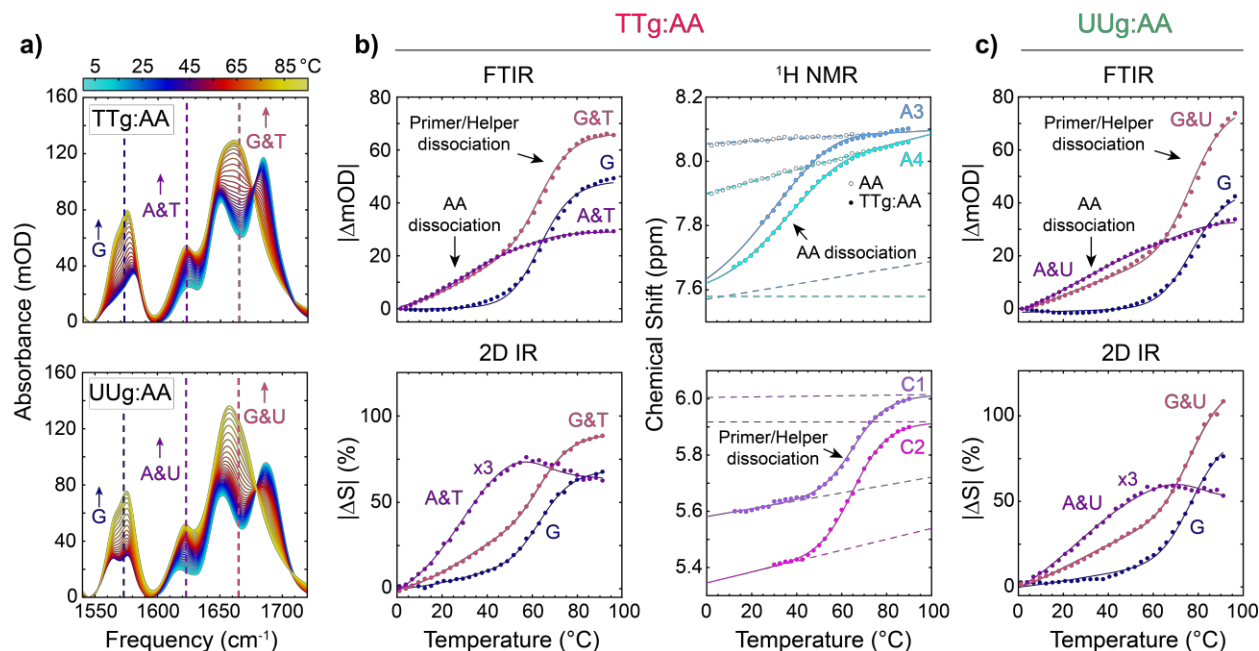


Figure 2. Temperature-induced dissociation of AA. (a) FTIR spectra of TTg:AA and UUg:AA from 1 to 96 °C in ~2.6 °C steps. Solutions contain 1 mM each of AA, primer, helper, and template strands. (b)

Thermal AA and template dissociation of the TTg:AA complex monitored with FTIR, 2D IR, and ^1H NMR. FTIR traces are plotted as the absolute value change in absorption relative to 1°C for select frequencies marked in (a) that primarily report on intensity changes in G (1573 cm^{-1} , blue), A & T (1623 cm^{-1} , purple), and G & T (1665 cm^{-1} , pink). 2D IR traces at select frequencies similar to the FTIR data are plotted as the absolute value percent change in signal relative to 1°C ($|\Delta S|$). From ^1H NMR, the AA dissociation transition is monitored using the chemical shift of two aromatic protons from adenine (A3 and A4), and H5 aromatic protons of cytosine (C1 and C2). The temperature-dependent chemical shifts of A3 and A4 in free AA are shown as open circles. The corresponding ^1H NMR spectra are shown in Fig. S9. (c) Thermal AA and template dissociation transitions of the UUg:AA complex monitored with (top) FTIR and (bottom) 2D IR. Solid lines for each dataset correspond to fits to a three-state sequential model. FTIR and 2D IR data are globally fit together. The four ^1H NMR peaks are globally fit to the same three-state model and include upper and lower baselines (dashed lines). Temperature-dependent FTIR and 2D IR data for TTo:AA and TTg, TTo, and UUg are shown in Sections S1.2 and S1.3.

Temperature-jump IR spectroscopy of AA dissociation kinetics

To directly monitor the time-dependence of AA dehybridization from gaps and overhangs, we employed T-jump IR spectroscopy(45, 46) using T-jumps of $\Delta T \approx 13^\circ\text{C}$ across the center of the AA melting transition (Figs. 3 and S11). The absorption change at 1605 cm^{-1} by the adenine ring mode and at 1545 cm^{-1} by the guanine ring mode provide selective reporters for changes in A:T and G:C base pairing, whereas the 1672 cm^{-1} band tracks both A:T and G:C base pair disruption through the thymine and guanine carbonyl vibrations.(49, 69) For all complexes, time traces at these three frequencies exhibit what appears to be a single kinetic component from 500 ns to $10\text{ }\mu\text{s}$ prior to decay of the signal due to thermal re-equilibration of the sample from 1 to 10 ms (Fig. 3b). The large amplitude of the A:T response at 1605 cm^{-1} and its microsecond timescale support assignment of the process to AA unbinding.

Without AA, the gap and overhang templates also exhibit a response over a similar time window but with lower amplitude, particularly at the adenine ring mode, relative to the AA-template complexes. The dominant G:C character of the response suggests that it may arise from base pair disruption adjacent to the gap or at the primer and helper termini; however, the same response is also observed in the fully formed duplexes (Figs. S12-S13), which indicates it arises from fraying of the primer and helper G:C termini. This observation is consistent with previous NMR studies that found G:C base pairs adjacent to a gap to be more stable than those at the duplex termini,(28) and T-jump experiments indicating G:C fraying in DNA oligonucleotides on microsecond timescales.(70) G:C fraying occurs faster than AA dissociation yet with enough temporal overlap that only a single kinetic component is resolved at each frequency. AA dissociation dominates the observed kinetic component at most frequencies, but the signal change

at 1672 cm^{-1} contains enough amplitude from G:C fraying that the extracted observed rate is faster than at 1605 cm^{-1} for TTg:AA at low temperature (Figs. S14-S15).

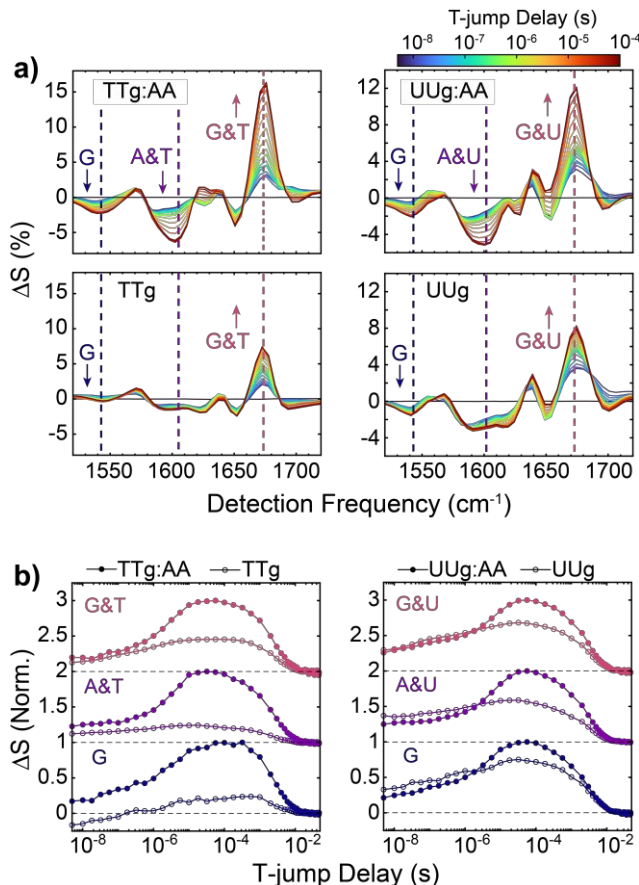


Figure 3. Time-dependence of AA dissociation from DNA and RNA gaps directly observed with T-jump IR spectroscopy. (a) Heterodyned dispersed vibrational echo difference spectra (t-HDVE) at T-jump delays ranging from 5 ns to 180 μs for TTg:AA, TTg, UUg:AA, and UUg following a T-jump from 19 to 33 $^{\circ}\text{C}$. Spectra are plotted as the change relative to the maximum signal value of the initial temperature spectrum ($\Delta S(t) = S(t)/\max(S(T_i))$). (b) Normalized $\Delta S(t)$ time traces at 1545 cm^{-1} (blue, guanine ring mode), 1605 cm^{-1} (purple, adenine ring mode), and 1672 cm^{-1} (red, guanine and thymine carbonyl modes) for TTg:AA and UUg:AA (filled circles) as well as TTg and UUg (open circles). The probe frequencies are marked as vertical dashed lines in the t-HDVE spectra. Traces at different frequencies are shifted vertically with respect to one another and dashed lines indicate respective baselines. AA dissociation is observed as a single kinetic component from 1 to 10 μs , and additional base pairing dynamics from the template itself occur on an overlapping timescale.

Thermodynamic analysis of AA dissociation

The results demonstrate steady-state and time-resolved thermal dissociation of AA, which we self-consistently model to extract quantitative thermodynamic and kinetic information. In general, complete dissociation of the AA-template complex may proceed through multiple

pathways differentiated by the order in which AA, primer, or helper segments dissociate. Exact treatment of these coupled equilibria leads to a fourth or higher-order dependence of each binding fraction on the concentration of a given species, which we find does not have a stable solution over the full temperature range of each transition. However, the much greater binding stability of the primer and helper allows us to neglect species with AA bound to the template strand without the primer and helper. We also model primer and helper dissociation from the template strand as a unimolecular equilibrium between the template and “dissociated” template that is equivalent to treating binding of the primer and helper as folding of hairpins.⁽¹⁶⁾ This results in a three-state sequential model described by equilibrium constants for AA dissociation (K_d) and dissociation of the primer and helper ($K_{d,Temp}$). Using TTg:AA as an example:



where dTTg is the “dissociated” template. The assumption in eq. 1b sharpens the primer and helper melting profile relative to its true bimolecular form as illustrated in Fig. S23, but this is not a serious concern because this only affects our analysis in the temperature range where the transitions overlap.

To extract melting curves for AA dissociation, FTIR and 2D IR temperature series for a given sequence are globally fit to eqs. 1a-b using the spectral decomposition described in Section S3 and constraints from the titration at 1 °C. The resulting melting curves are shown in Fig. 4 and thermodynamic parameters are shown in Fig. 4 and Table 1, and the corresponding spectral components are shown in Figs. S20-S21. Overall, the global fit describes both data sets well (Fig. 2). The first spectral component corresponds to AA dissociation and contains signatures of A:T base pair breaking whereas the second component contains only G:C base pair breaking features. The AA dissociation curves resemble the raw temperature-dependent change in absorbance and start from a bound fraction less than 1 at 1 °C. We also find quantitative agreement between the TTg:AA dissociation curves determined from FTIR and 2D IR and from global fitting of temperature-dependent ¹H NMR chemical shifts (Fig. S9), lending support to the accuracy of the IR dissociation curves.

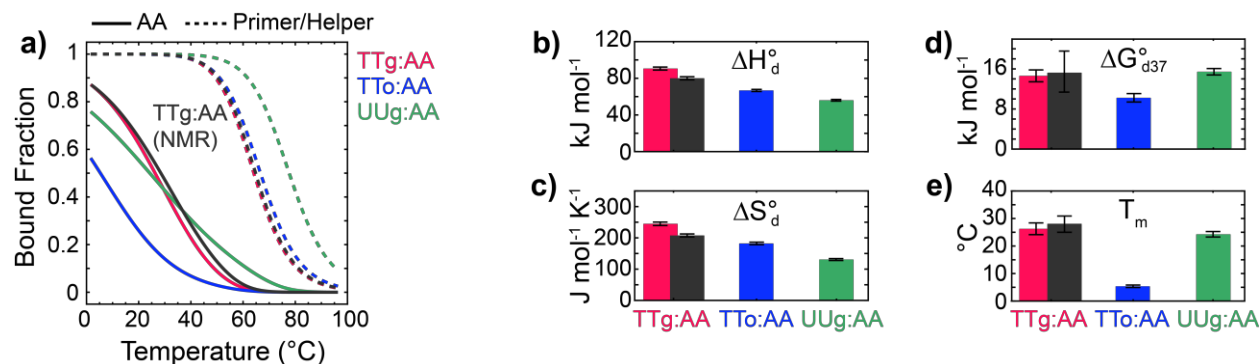


Figure 4. Thermodynamics of AA dissociation from gap and overhang templates extracted with three-state global fitting of FTIR and 2D IR temperature series. (a) AA (solid) and primer and helper (dashed) dissociation curves for TTg:AA, TTo:AA, and UUG:AA extracted from global fitting of FTIR and 2D IR temperature series to a three-state sequential model (eqs. 1a-b, Section S3). Dark gray lines correspond to dissociation components of TTg:AA obtained from ^1H NMR temperature series (Fig. S9). AA dissociation (b) enthalpy (ΔH_d°), (c) entropy (ΔS_d°), (d) free energy at 37 °C (ΔG_{d37}°), and (e) melting temperature (T_m) determined for each sequence. Dark gray bars indicate values from ^1H NMR. Error bars are derived from 95% confidence intervals in fit parameters ΔH_d° and ΔS_d° .

Table 1. Thermodynamics for dissociation of AA from templates determined with global fitting to three-state sequential model (eqs. 1a-b). Error are derived from 95% confidence intervals in fit parameters ΔH_d° and ΔS_d° . ΔH_d° and the dissociation free energy at 8 °C (ΔG_{d8}°) from ITC measurements are also listed where errors correspond to 95% confidence intervals from two-state fits (Section S4.2).

Sequence	Method	T_m^c (°C)	ΔH_d° (kJ mol $^{-1}$)	ΔS_d° (J mol $^{-1}$ K $^{-1}$)	ΔG_{d8}° ^d (kJ mol $^{-1}$)
TTg:AA	IR ^a	26 ± 2	91 ± 2	245 ± 6	21.7 ± 1.1
	NMR	28 ± 3	79 ± 5	203 ± 20	21.5 ± 4.5
	ITC ^b	-	83 ± 5	-	23.1 ± 1.3
UUG:AA	IR ^a	24 ± 1	56 ± 1	131 ± 3	19.2 ± 0.6
	ITC ^b	-	51 ± 1	-	23.5 ± 0.4
TTo:AA	IR ^a	5 ± 1	67 ± 1	182 ± 4	16.1 ± 0.8

^aIR refers to global fits of FTIR and 2D IR temperature series. ^bIsothermal titration calorimetry (ITC) was performed at 8 °C. ^c T_m is defined as the temperature where half of the AA is bound to the template. ^dFree energy change for AA dissociation at 8 °C.

The shape of AA dissociation curves varies between DNA and RNA as well as the type of template. TTg:AA and UUG:AA exhibit similar transition midpoints (T_m), defined as the temperature where half of the total AA concentration is bound to the template, yet UUG:AA shows a broader transition with a dissociation curve slope at T_m of 0.013 K $^{-1}$ relative to 0.021 K $^{-1}$ in TTg:AA. The broader transition width in UUG:AA arises from a 35 kJ mol $^{-1}$ lower dissociation

enthalpy (ΔH_d°) and 115 J mol⁻¹ K⁻¹ lower dissociation entropy (ΔS_d°) relative to TTg:AA. This difference in ΔH_d° is also directly observed through ITC measurements (Table 1 and Section S4.2). Although TTg:AA and UUG:AA exhibit a similar dissociation free energy at 37 °C (ΔG_{d37}°), the larger ΔH_d° and ΔS_d° values lead to greater binding stability for TTg:AA at lower temperatures and UUG:AA at higher temperatures, but we note that ITC gives similar ΔG_d° for each sequence at 8 °C. Thymine contributes greater π - π stacking interactions than uracil,(71, 72) which is why AA is the only dinucleotide step with nearest-neighbor ΔH_d° and ΔS_d° parameters that are larger for DNA (33 kJ mol⁻¹; 93 J mol⁻¹ K⁻¹) than RNA (28 kJ mol⁻¹; 79 J mol⁻¹ K⁻¹). (4, 5) This effect contributes to the larger values of ΔH_d° and ΔS_d° in TTg:AA relative to UUG:AA but cannot solely account for the observed magnitude of this difference. The free energy of AA dissociation at 37 °C is roughly 4 kJ mol⁻¹ higher in TTg:AA than TTo:AA due to the added stacking stabilization from the 3'-G, and this value is similar to previous measurements of coaxial stacking in 5'-AG-3' steps of DNA oligonucleotides.(15, 17, 18)

Kinetic analysis of AA dissociation

The observed relaxation rates for AA dissociation kinetics, k_{obs} , are extracted from T-jump measurements with varying initial temperature across the AA melting transition of each sequence (Figs. 5 & S11-S13). Only AA dissociation and G:C fraying are observed below final temperatures $T_f = 42$ °C for DNA and 52 °C for RNA. At higher temperatures, primer and helper dissociation is observed on 100 μ s timescales and is more than 100-fold slower than AA dissociation (Fig. S13). The timescale separation between AA dissociation and primer and helper dissociation is sufficient to treat k_{obs} with two-state kinetics.(73) Using TTg:AA as an example,

$$k_{obs} = k_d + k_a ([TTg] + [AA]) \quad (2)$$

where k_a and k_d are the rate constants for AA association and dissociation. Since $K_d = k_d/k_a$, these rate constants are extracted using the concentrations of template, [TTg], and AA, [AA], determined from the AA dissociation curves in Fig. 4. Due to the broad observable temperature range of AA-gap dissociation transitions, T-jump measurements can be performed at multiple T_f values below T_m where k_{obs} is dominated by k_a . Over this temperature range, the dissociation rate follows Arrhenius behavior and increases from 3×10^3 to 2×10^6 s⁻¹ while k_a shows just a minor 2-to-5-fold increase. Such contrasting temperature-dependence of k_a and k_d has consistently been

observed in studies of short DNA and RNA oligonucleotides (5 – 20 bp),(6, 7, 49, 74, 75) and has been used to characterize the transition-state between hybridized and dissociated states. We fit the temperature trends of k_a and k_d to a Kramers-like model for diffusive barrier crossing where the rate is inversely proportional to the solvent viscosity, η . Written for the dissociation rate:

$$k_d(T) = (c^\circ)^{1-m} \frac{k_B T}{h} \frac{\eta_{37}}{\eta(T)} \exp\left(-\frac{\Delta H_d^\ddagger - T\Delta S_d^\ddagger}{RT}\right) \quad (3)$$

and the same expression applies for association. Here the free energy barrier ($\Delta G_d^\ddagger$) is written in terms of the enthalpic ($\Delta H_d^\ddagger$) and entropic ($\Delta S_d^\ddagger$) barriers. The temperature-dependent viscosity of D₂O is taken relative to its value at 37 °C (η_{37}/η).⁽⁷⁶⁾ Also, c° is the standard state concentration of 1 M, and m is the reaction molecularity ($m = 1$ for k_d and $m = 2$ for k_a). We use a simple form of the Eyring pre-factor ($k_B T/h$). This approximation is known to overestimate the frequency of diffusive motion at the barrier for biomolecular folding and binding.^(77, 78) While this impacts the absolute magnitudes of $\Delta S_d^\ddagger$ and $\Delta G_d^\ddagger$, we assume that the true pre-factor is the same across our sequences such that relative values of $\Delta S_d^\ddagger$ and $\Delta G_d^\ddagger$ are meaningful. Fitting of our data indicates that the free energy barrier for AA dissociation arises from an enthalpic penalty ($\Delta H_d^\ddagger > 50$ kJ mol⁻¹) that is partially compensated by a positive $\Delta S_d^\ddagger$, whereas the free energy barrier for AA association ($\Delta G_a^\ddagger$) is dominated by an entropic barrier ($\Delta S_a^\ddagger < 0$).

The kinetics of AA association and dissociation vary substantially depending on the template. $\Delta G_d^\ddagger$ is ~ 4.5 kJ mol⁻¹ smaller for TTo:AA than TTg:AA at 37 °C, which matches the difference in ΔG_{a37}° , suggesting that the 5-fold faster unbinding of AA from an overhang compared to a gap primarily comes from reduced binding stability. The rate of association or $\Delta G_{a37}^\ddagger$ are essentially the same for TTo:AA and TTg:AA but with small differences in its temperature-dependence as quantified by $\Delta H_a^\ddagger$. There are also significant differences between the kinetics of AA dissociation in UUG:AA and TTg:AA. k_d increases more sharply with temperature in TTg:AA, as quantified by a ~ 25 kJ mol⁻¹ larger $\Delta H_d^\ddagger$ relative to UUG:AA, which likely comes from enhanced stacking interactions in the hybridized state of TTg:AA and is consistent with the thermodynamic results presented above. On the other hand, k_a is similar for each sequence across the measured temperature range.

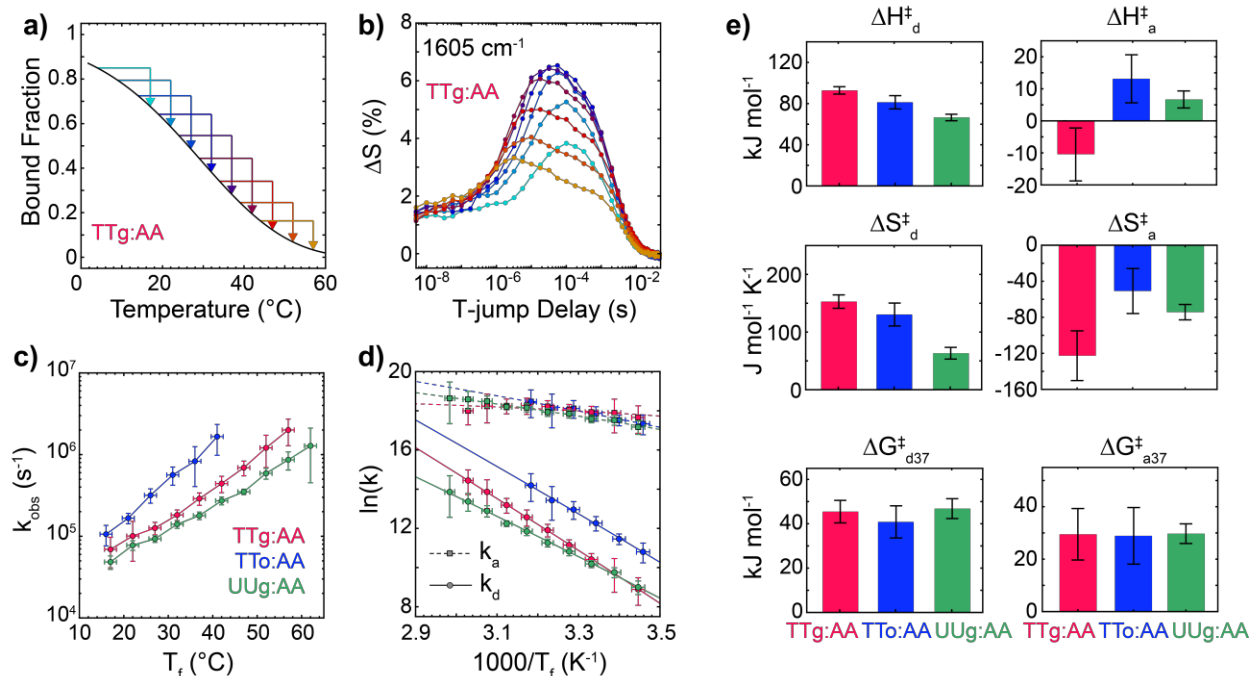


Figure 5. Kinetics of AA association and dissociation from gaps and overhangs. (a) Illustration of a series of T-jumps performed from various initial to final temperatures ($T_i \rightarrow T_f$) with $\Delta T \sim 13$ °C across the AA dissociation transition of TTg:AA. (b) t-HDVE time traces of TTg:AA probed at the adenine ring mode (1605 cm^{-1}) for each temperature condition shown in (a). (c) Observed relaxation rate (k_{obs}) for AA unbinding as a function of T_f for all sequences. Rates correspond to the amplitude-weighted mean across the adenine ring mode response (1585 to 1610 cm^{-1}) in the rate-domain spectra determined from MEM-iLT.(48, 49) Vertical error bars indicate standard deviation over the rate spectra and horizontal error bars correspond to the measured standard deviation in T-jump magnitude. (d) Rate constants for AA association (k_a) in $\text{M}^{-1} \text{ s}^{-1}$ and dissociation (k_d) in s^{-1} as a function of T_f determined from a two-state analysis of k_{obs} (eq. 2). Data for all sequences is shown in Fig. S27. (e) Enthalpic ($\Delta H_d^\ddagger$, $\Delta H_a^\ddagger$), entropic ($\Delta S_d^\ddagger$, $\Delta S_a^\ddagger$), and free energy (37°C , $\Delta G_{d37}^\ddagger$, $\Delta G_{a37}^\ddagger$) barriers determined from fitting k_d and k_a temperature-trends to a Kramers-like model (eq. 3). Error bars indicate 95% confidence intervals propagated from the fits.

Structural properties of AA-gap complexes from molecular dynamics simulations

The energetics of AA hybridization onto overhangs and gaps differs from free strand hybridization due to a combination of differences in the template and AA-gap complex structural configurations relative to canonical duplexes and single-strands. We characterized the differences in conformation between these species using all-atom MD simulations, finding that most configurational differences are localized near the central AA binding site. As observed previously, gap templates adopt bent configurations in high population which reduces hydration of the gap relative to single strands (Figs. S29-S30).(27, 28, 79) AA-gap complexes exhibit greater configurational freedom than canonical duplexes (Figs. 6 & S31-S33), with the largest differences

found in backbone and base-stacking structure rather than base pair geometry. TTg:AA exhibits broadened distributions in shift, slide, and twist parameters at the 5'-GA-3' and 5'-AG-3' nicks, whereas the stacking geometry of the between adenine bases is nearly identical to B-DNA (ΔD_x , $\Delta D_y < 0.4 \text{ \AA}$ and $\Delta \Omega < 4^\circ$). A broader spread of stacking orientations at the nick sites is unsurprising given the greater flexibility in the backbone, particularly for glycosidic bond angles (χ) of the A and G bases on the 3'-side of each nick.⁽²²⁾ The nick sites enable greater local flexibility in the backbone, and in particular for the glycosidic bond angles (χ) of the A and G bases on the 3'-side of each nick. The greater freedom of the backbone near the nick sites also reduces the effective population of BII backbone configurations across the 5'-CGAAGG-3' motif (Fig. S34). The fluctuations along slide, twist, shift, and χ coordinates are partially correlated in a given nick site (Fig. 6b). In contrast, the structural fluctuations in the 5'-GA-3' and 5'-AG-3' nicks appear largely independent of one another (Fig. 6c).

The structural differences between UUG:AA and an A-RNA duplex are also minor yet distinct from DNA in a few regards. Overall, UUG:AA exhibits larger changes in stacking orientation (twist, shift) and less change in the backbone torsional angles (χ , δ , γ) relative to DNA. Further, the base pairing and stacking geometry of the AA differs with broader distributions of shear, stretch, and buckle values, an increase in mean rise from 3.0 to 3.3 $\text{\AA}$, and sub-populations with larger shift and twist values. In contrast to DNA, distinct states are observed along twist and shift coordinates for the 5'-GA-3' and 5'-AG-3' and the mean values differ from A-RNA by as much as 20° and 1.5 $\text{\AA}$, respectively.

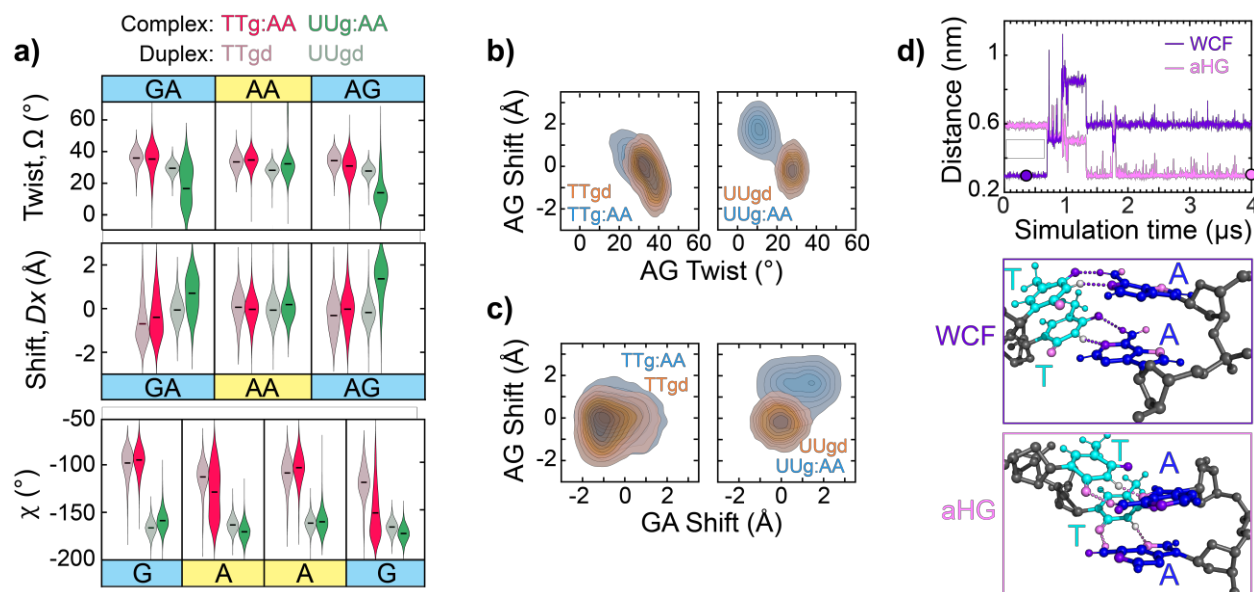


Figure 6. Structural properties of AA-gap complexes from all-atom MD simulations. (a) Comparison of structural variation between AA-gap complex and canonical duplex. Violin plots show structural distributions for base pair twist (Ω) and shift (D_x), and glycosidic bond dihedral angle (χ) for the three central base pair steps of TTgd (red, light), TTg:AA (red, dark), UUgd (green, light), and UUg:AA (green, dark). Black horizontal bars indicate the distribution's median value obtained from a Gaussian kernel estimation using Scott's rule.⁽⁸⁰⁾ Distributions were generated from 1.5 μs of simulation for each sequence using the bsc1-AMBER force field for DNA and DES-AMBER force field for RNA. Only Watson-Crick-Franklin (WCF) base pair configurations are included in the distributions. Other structural parameters using both force fields are shown in Figs. S31-S33. (b) Contour plots showing the correlation of 5'-AG-3' twist vs. shift. Nine contours with uniform 10% spacing are plotted for each sequence. Diagonally elongated distributions indicate that fluctuations along different base pair coordinates within the same nick site are partially correlated with Pearson correlation coefficients of -0.61 for DNA and -0.56 for RNA. (c) Contour plots showing the correlation of 5'-GA-3' shift vs. 5'-AG-3' shift. Symmetrical distributions indicate that fluctuations at each nick site are uncorrelated with Pearson correlation coefficients of -0.04 for DNA and 0.05 for RNA. (d) 4 μs trajectory of TTg:AA showing the A:T base pairs switching from WCF to an anti-Hoogsteen (aHG) geometry. Trajectories are plotted in terms of the mean WCF (purple) and aHG (pink) distances across each A:T base pair. The WCF distance is defined as the mean of adenine-N6 to thymine-O4 and adenine-N1 to thymine-N3 distances and the aHG distance is the average of adenine-N7 to thymine-N3 and adenine-N6 to thymine-O2 distances. Visualizations of each base pair geometry at time points marked with filled circles indicate WCF distances as purple dashed lines and aHG distances as pink dashed lines.

During one of two 4 μs trajectories for TTg:AA, we observed AA switch from canonical Watson-Crick-Franklin (WCF) to anti-Hoogsteen (aHG) base pairing geometry. As shown in Fig. 6d, the 5'-A flips at 700 ns followed by the 3'-A 200 ns afterward, however, both adenines remain poorly aligned with the gap until reaching the final aHG geometry at 1.3 μs . Our 8 μs of total simulation time for this sequence is insufficient to determine the relative stability of WCF and HG

configurations, therefore we applied the On-the-fly Probability Enhanced Sampling (OPES) method to calculate the free energy differences for AA dissociation from a WCF or aHG geometry ($\Delta\Delta F_{WCF-aHG} = \Delta F_{d37,aHG} - \Delta F_{d37,WCF}$; See Section S5.3). WCF and aHG geometries have nearly the same population and stability at 37 °C with $\Delta\Delta F_{WCF-aHG} \approx 0$ kJ mol⁻¹. Although not previously tested for dinucleotide-gap complexes or aHG base pairing, simulations with the bsc1-AMBER force field typically capture accurate relative stabilities between WCF and syn-Hoogsteen base pairing geometries.(81, 82) Simulations with the DES-AMBER force field instead suggest that the WCF geometry is more stable ($\Delta\Delta F_{WCF-aHG} \sim -8$ kJ mol⁻¹). For UUG:AA, WCF and aHG geometries have similar stabilities regardless of the employed force field (Fig. S35). The high stability of an aHG geometry may appear unusual, but this has been reported as the dominant base pairing geometry between AA and TT DNA dinucleotides mediated through self-assembling tripyridyl-triazine capped hydrophobic cages in aqueous solution, so far the only structural characterization of dinucleotide duplexes in aqueous solution.(83)

Length scaling of A_n dissociation from templates

To assess the scaling of dissociation thermodynamics and kinetics for 2-to-4 nucleotides sequences, we experimentally examined the dehybridization of AAA (A3) and AAAA (A4) from DNA gaps and overhangs using the extended templates shown in Fig. 7a. As expected, longer A_n oligonucleotides bind with greater stability and show a decrease in both k_d and k_a (Figs. 7b-d & S27). Both gaps and overhangs exhibit an average increase in $\Delta G_{d,8}^\circ$ of 2.7-3.0 kJ mol⁻¹ per base pair (bp), yet ΔH_d° and ΔS_d° increase more sharply with length for overhangs and become equivalent to the gap for A4 dissociation. The overhang values for $\Delta\Delta H_d^\circ$, $\Delta\Delta S_d^\circ$, and $\Delta\Delta G_{d,8}^\circ$ are still well below the 33 kJ mol⁻¹ bp⁻¹, 94 J mol⁻¹ K⁻¹ bp⁻¹, and 6.7 kJ mol⁻¹ bp⁻¹ values predicted for appending AA dinucleotide steps by Santa Lucia's DNA nearest-neighbor (NN) model (Fig. 7e).(4)

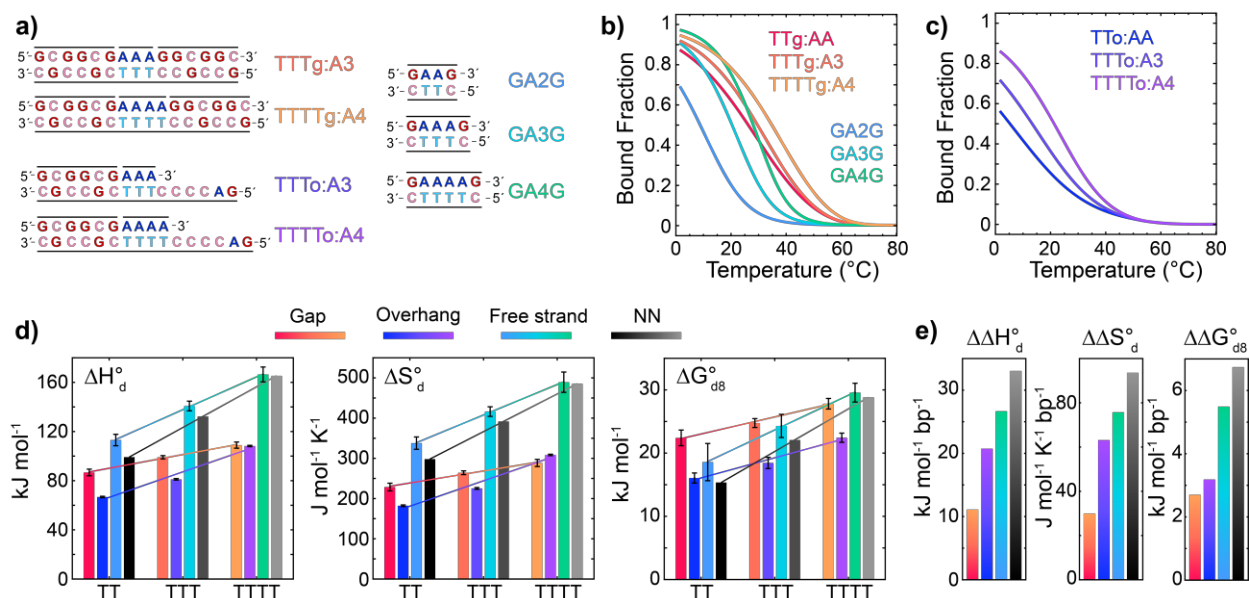


Figure 7. Comparing length-dependent dissociation thermodynamics in gaps, overhangs, and free strands. (a) Extended templates for 3 and 4 nucleotide gap (TTTg:A3, TTTTg:A4) and overhang (TTTo:A3, TTTTo:A4) sequences bound to A3 and A4 oligonucleotides. Dissociation of A_n ($n = 2, 3, 4$) segments from gaps and overhangs are compared with dissociation of GA_nG single-strands with their complement strands. (b, c) A_n melting curves for gaps and overhangs extracted from global fitting of FTIR and 2D IR temperature series to three-state sequential model (eq. 1). GA_nG melting curves are shown in (b) and were extracted from a two-state model fit to FTIR temperature series (Section S4.1). GA_nG melting curves were measured at 10 mM oligonucleotide concentration and corrected to 2 mM. (d) ΔH_d° , ΔS_d° , and dissociation free energy at 8 °C (ΔG_{d8}°) for all sequences. Red-orange bars correspond to DNA A_n -gap complexes, blue-purple bars correspond to DNA A_n -overhang complexes, light blue-green bars are for GA_nG strands, and black-gray indicate GA_nG dissociation parameters calculated from Santa Lucia's nearest-neighbor (NN) model and corrected to $[Na^+] = 600$ mM.(4, 84) Values for gap and GA_nG systems are an average over those determined from IR spectroscopy and ITC (Section S4.2) while overhang values are only from FTIR and 2D IR global fits. Error bars indicate 95% confidence intervals from fits to thermodynamic models. (e) Slopes from linear fits of each thermodynamic parameter across A_n length (solid lines in (d)). Data are presented as the change per appended A:T base pair (bp) in the dissociation enthalpy ($\Delta\Delta H_d^\circ$), entropy ($\Delta\Delta S_d^\circ$), and free energy at 8 °C ($\Delta\Delta G_{d8}^\circ$). Although AA-gap complex binding is more stable than GA2G, $\Delta\Delta H_d^\circ$, $\Delta\Delta S_d^\circ$, and $\Delta\Delta G_{d8}^\circ$ are each more than 2-fold greater in GA_nG strands than for A_n dissociation from gaps.

To test for the origin of different length-scaling dissociation thermodynamics between overhangs, gaps, and the NN model, we compare to dissociation of oligonucleotides 5'-GAAG-3' (GA2G), 5'-GAAAG-3' (GA3G), and 5'-GAAAAG-3' (GA4G) with their respective complement strands (Figs. 7 & S24-S26). ΔH_d° and ΔS_d° are both greater for GA2G than AA-gap dissociation while ΔG_{d8}° is ~ 4 kJ mol $^{-1}$ lower for GA2G. As a result, the GA2G dissociation transition is much sharper than for AA-gap dissociation yet with T_m shifted ~ 15 °C lower in temperature. Values of

$\Delta\Delta H_d^\circ$, $\Delta\Delta S_d^\circ$, and $\Delta\Delta G_{d8}^\circ$ for GA_nG are more than 2-fold greater than for gap sequences and are much closer the NN model values (Fig. 7e). This leads to nearly equivalent ΔG_{d8}° for TTTTg:A4 and GA4G, and binding between free single-strands may progressively become more stable than for binding of longer oligonucleotides onto gaps and overhangs. Additionally, we find that length-scaling of the dissociation free energy barrier ($\Delta\Delta G_d^\ddagger$) is ~ 2.5 -fold greater than $\Delta\Delta G_d^\circ$ for gaps (Fig. S28), which is a much larger difference than previously reported for dehybridization of free single-strands.(74, 85) This sharp dependence of $\Delta G_d^\ddagger$ (and k_d) on length is important for predicting dissociation kinetics in template systems and also indicates a length-dependence in the free energies of the association transition state and/or the unbound state.(86, 87)

Based on previous studies of gap and overhang structure,(23-28) it is possible that the single-stranded binding region in gaps and overhangs becomes less ordered as the length increases from 2 to 4 nucleotides. However, our all-atom MD simulations indicate essentially no difference in the structural distribution of the canonical unbent T_n configurations in gaps, but instead highly bent configurations and alternative stacking geometries are observed for certain gap lengths (Fig. S29). AMBER force fields, including bsc1-AMBER, tend to overestimate base-stacking stability in single-strand DNA as they are parameterized for duplex DNA,(56, 88) and therefore these results should be interpreted with caution.

Conclusions

We demonstrate an approach for measuring temperature-dependent binding thermodynamics and kinetics of nucleic acids as short as dinucleotides onto gaps and overhangs using IR spectroscopy. By determining accurate melting curves, our method extracts enthalpic and entropic contributions to binding that are typically missing in previous titration studies. Further, to our knowledge we provide the first measurements of dinucleotide association and dissociation timescales (0.5 – 20 μs) through the application of T-jump IR spectroscopy. Although our study is limited to pure A_n segments, we identify striking energetic differences between cases of association onto gaps or overhangs with free-strand association that have not been observed previously.

Most notable is that association of AA onto a gap with G:C base pairs on each side is energetically more favorable than association between 5'-GAAG-3' and 5'-CTTC-3' ($\Delta\Delta G_{d8}^\circ = 4$

kJ mol⁻¹, Fig. 7d). This distinction may arise from energetic differences in either the bound or dissociated states or a combination of both through factors such as single-strand stacking, optimized coaxial stacking, or non-canonical base pairing. Most association measurements, including the approach in this work, directly probe energetic differences between the bound and dissociated states but not energetic differences between the respective bound or dissociated states of different samples. As a result, we cannot directly dissect the enhanced stability of AA-template complexes but rather infer how the energy of the associated and dissociated states may change between gap, overhang, and free single-strand scenarios.

A clear distinction between gap and free single-strand binding scenarios is present in both the enthalpic and entropic contributions to association. ΔH_d° and ΔS_d° are lower for the gap, and therefore the extra stability of TTg:AA relative to GA2G arises from an entropic benefit. Enhanced binding stability of dinucleotides onto gaps and overhangs may arise from greater base pair ordering and constraints in water dynamics within the single-strand binding region and/or increased configurational flexibility in the AA-template complexes relative to canonical B-DNA, which we observe in MD simulations near the nick sites (Figs. 6a & S30-S31). The former effect decreases the entropy of the dissociated state while the latter increases that of the hybridized state, each of which will decrease ΔS_d° relative to free stand hybridization. The same effect that may increase rigidity and stacking interactions in the gap single-strand will also reduce the enthalpy of the dissociated state. Another reduction in enthalpy may arise from bending about the single-stranded segment in gap templates as observed in our MD simulations (Fig. S29) and reported in previous studies.(26, 28, 29, 89, 90) Lastly, our simulations of TTg:AA suggest that AA can bind in WCF or aHG geometries with significant population, although the relative stability depends on the choice of force field (Fig. S35). aHG configurations are likely only relevant to binding of pure A:T or A:U dinucleotides or G:G mismatch pairing where the aHG configuration is a stable option. To date, the only crystallographic studies of dinucleotide binding to a gap or overhang template are for G:C base pairing of modified GG RNA dinucleotides,(91-93) a scenario where aHG configurations are highly improbable. Unfortunately, we cannot indicate nor disprove the presence of A:T aHG base pairing using the IR and NMR measurements presented in this work, and further structural characterization is necessary to clarify the role of aHG base pairing.

The energetics of dinucleotide association onto gaps and overhangs differs from free-strand hybridization, and these effects stem from underlying dynamics of association and dissociation. Our T-jump IR measurements of DNA and RNA AA association and dissociation kinetics provide a first step toward understanding the structural dynamics. We determine a k_a value of $\sim 5 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$ that corresponds to a time constant of 20 μs at 1 mM oligonucleotide concentration. This k_a value is nearly 100-fold slower than an estimated diffusion-limited rate constant (k_{diff}) of $5 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$ calculated using the translational diffusion coefficients of AA and TTg determined from DOSY NMR measurements and estimated radii of gyration from experimental length-scaling of ssDNA and worm-like chain model of duplex DNA (Section S6).(94, 95) Although faster than association between oligonucleotides of 5 – 20 nucleotide lengths, the 100-fold difference between AA association and diffusion-limited rates suggest that numerous unsuccessful collisions and/or significant structural rearrangement occur prior to successful binding. Association rate constants for longer oligonucleotides are also consistently reported to be 100-1000-fold smaller than k_{diff} ,⁽⁷⁴⁾ suggesting that each system's hybridization transition-state and encounter complex may share certain characteristics. Even for longer oligonucleotides these properties are still poorly understood but are thought to involve some degree of single-strand ordering, a small number of base pair contacts, and potentially rearrangement of water molecules and counterions. In the case of dinucleotide association to gaps and overhangs, water and counterion rearrangement is still necessary, and the ordering of the binding region of the gap or overhang contains a large penalty. The time ordering of these events are unclear for each system and must be investigated further to develop a molecular picture for the dynamics of short oligonucleotide hybridization.

We have primarily focused on short DNA oligonucleotide association, and it is important to assess whether similar characteristics are found for short RNA oligonucleotides, which are more relevant in the context of non-enzymatic replication.⁽⁹⁶⁾ Our results show large differences in energetics and kinetics of AA dissociation from DNA and RNA gaps. ΔH_d° and ΔS_d° values for UUG:AA are approximately half of those predicted by NN models and obtained for TTg:AA (Fig. S26). ΔH_d° and ΔS_d° are likely higher in DNA for multiple reasons. The AA dinucleotide in particular contains enhanced stacking arising from the C5-methyl groups of thymine.^(71, 72, 97) Another contribution may come from the wider breadth of stacking configurations observed in UUG:AA relative to UUGd. However, RNA duplexes are generally more thermodynamically

stable than an equivalent sequence DNA duplex due to multiple effects from the C2'-OH groups. Differences in base stacking, hydrogen bonding between C2'-OH groups and water molecules, and greater uptake and ordering of cations and water molecules are all effects that contribute to larger ΔH_d° , ΔS_d° , and ΔG_d° in A-RNA than B-DNA.(98, 99) Therefore, we predict that binding of other RNA dinucleotide sequences will generally be more stable than the equivalent DNA sequence.

Gaps and overhangs are expected to have a qualitatively similar thermodynamic and kinetic impact on DNA and RNA hybridization, but the relative magnitude may generally differ. For instance, single- and double-strand RNA are shown to have a longer persistence length than DNA and the change in strand rigidity upon introducing an overhang or gap may differ as well as the propensity for bending of gapped RNA.(30, 100) MD simulations of UUG:AA and UUGd indicate that the distribution of stacking geometries at nick sites between primer and helper segments and AA differ from A-RNA while the backbone geometry is relatively unchanged, yet the opposite is found for DNA (Fig. 6a). Such differences in geometry may influence the enthalpy of the AA-gap complex while the enhanced configurational flexibility increases its entropy. Together, the combination of differences between DNA and RNA templates and dinucleotide-template complexes may promote significant contrast in their association energetics and kinetics across variations in dinucleotide sequence and template construction.

Our experimental and computational results indicate multiple structural, energetic, and kinetic factors that differentiate hybridization between short nucleic acids and templates with hybridization between free single-strands, and these differences may play an important role in processes such as non-enzymatic replication. In addition to various structural and chemical factors, the efficiency of this process depends upon the binding probability, dissociation rate of the bound complex, and structural dynamics of the complex.(93, 101) The structural dynamics of the gap template and bound complex suggested by our experimental and computational results are likely to influence the efficiency of non-enzymatic extension. Further, we measure an AA dissociation rate from UUG:AA that is orders of magnitude higher than the catalytic rate (k_{cat}) of non-enzymatic primer extension for a 2-aminoimidazole-bridged (2AI) dinucleotide in an equivalent sequence context,(101) indicating that on average the 2AI dinucleotide binds and dissociates from the gap numerous times before extension occurs. Although measured under different solution conditions, the dissociation constant of UUG:AA at 25 °C ($K_d = 1.05$ mM) is similar to the

Michaelis constant of the 2AI system ($K_m = 0.64$ mM) extracted from a Michaelis-Menten analysis of the extension rate.(101) These observations support the use of a Michaelis-Menten kinetic model for non-enzymatic extension of this sequence and template, and our study demonstrates a method for direct measurement of energetics and kinetics for dinucleotide binding that is applicable to a wide range of oligonucleotide sequences and template designs.

Supporting Material

Supporting material can be found online at XXX.

FTIR titration curve analysis; characterization of FTIR and 2D IR temperature series of templates and short segment dissociation; ^1H NMR temperature series spectra for TTg:AA; temperature-dependent T-jump IR data of templates and short segment dissociation; description of thermodynamic and kinetic models and global fitting; Isothermal titration calorimetry data and analysis; molecular dynamics simulations of DNA gap templates and free single-strands; description of molecular dynamics simulations of WCF to aHG transition with OPES enhanced sampling; diffusion-ordered NMR of gap template and AA.

Author Contributions

B.A., M.S.J., J.W.S., A.L.F., and A.T designed the research. B.A., A.R., Y.L. and J.R.S. performed experiments. M.S.J. and S.K. performed simulations. B.A. and M.S.J. analyzed the data. B.A. and A.T. wrote the manuscript. All authors provided feedback on the manuscript.

Acknowledgements

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health (Award No. R01-GM118774), the National Science Foundation under Grant No. CHE-2152521 and CHE-2155027, and the Simons Foundation (Grant No. 290363). J.W.S. is an Investigator of the Howard Hughes Medical Institute. B.A. acknowledges support from the NSF Graduate Research Fellowship Program. We thank Marco Todisco and Jakob Schauss for helpful discussions and feedback on the manuscript. 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.

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, and 63/479,378 and International Patent Applications PCT/US2020/035206 and PCT/US2020/050466.

Supporting References

References 102 - 127 appear in the Supporting Material.

References

1. Yin Y. and Zhao X. S. 2011. Kinetics and dynamics of DNA hybridization. *Acc. Chem. Res.* 44:1172-1181.
2. Wetmur J. G. 1976. Hybridization and renaturation kinetics of nucleic acids. *Annu. Rev. Biophys.* 5:337-361.
3. Wartell R. M. and Benight A. S. 1985. Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Phys. Rep.* 126:67-107.
4. SantaLucia Jr J. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci.* 95:1460-1465.
5. Xia T., SantaLucia Jr J., Burkard M. E., Kierzek R., Schroeder S. J., Jiao X., Cox C., and Turner D. H. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochem.* 37:14719-14735.
6. Craig M. E., Crothers D. M., and Doty P. 1971. Relaxation kinetics of dimer formation by self complementary oligonucleotides. *J. Mol. Biol.* 62:383-401.
7. Pörschke D. and Eigen M. 1971. 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.
8. Simmel F. C., Yurke B., and Singh H. R. 2019. Principles and applications of nucleic acid strand displacement reactions. *Chem. Rev.* 119:6326-6369.
9. Wei X., Nangreave J., and Liu Y. 2014. Uncovering the self-assembly of DNA nanostructures by thermodynamics and kinetics. *Acc. Chem. Res.* 47:1861-1870.
10. Kiefer J. R., Mao C., Braman J. C., and Beese L. S. 1998. Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature* 391:304-307.
11. Steitz T. A. 2008. A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.* 9:242-253.
12. Zhou L., Ding D., and Szostak J. W. 2021. The virtual circular genome model for primordial RNA replication. *RNA* 27:1-11.
13. Duzdevich D., Carr C. E., Ding D., Zhang S. J., Walton T. S., and Szostak J. W. 2021. Competition between bridged dinucleotides and activated mononucleotides determines the error frequency of nonenzymatic RNA primer extension. *Nucleic Acids Res.* 49:3681-3691.
14. Leveau G., Pfeffer D., Altaner B., Kervio E., Welsch F., Gerland U., and Richert C. 2022. Enzyme-Free Copying of 12 Bases of RNA with Dinucleotides. *Angew. Chem. Int. Ed.*:e202203067.
15. Pyshnyi D. V., Goldberg E. L., and Ivanova E. M. 2003. Efficiency of coaxial stacking depends on the DNA duplex structure. *J. Biomol. Struct. Dyn.* 21:459-467.

16. Lane M. J., Paner T., Kashin I., Faldasz B. D., Li B., Gallo F. J., and Benight A. S. 1997. The thermodynamic advantage of DNA oligonucleotide ‘stacking hybridization’ reactions: energetics of a DNA nick. *Nucleic Acids Res.* 25:611-616.
17. Vasiliskov V. A., Prokopenko D. V., and Mirzabekov A. D. 2001. Parallel multiplex thermodynamic analysis of coaxial base stacking in DNA duplexes by oligodeoxyribonucleotide microchips. *Nucleic Acids Res.* 29:2303-2313.
18. 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.
19. Walter A. E., Turner D. H., Kim J., Lyttle M. H., Müller P., Mathews D. H., and Zuker M. 1994. Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl. Acad. Sci.* 91:9218-9222.
20. Walter A. E. and Turner D. H. 1994. Sequence dependence of stability for coaxial stacking of RNA helices with Watson-Crick base paired interfaces. *Biochem.* 33:12715-12719.
21. Parker T. M., Hohenstein E. G., Parrish R. M., Hud N. V., and Sherrill C. D. 2013. Quantum-mechanical analysis of the energetic contributions to π stacking in nucleic acids versus rise, twist, and slide. *J. Am. Chem. Soc.* 135:1306-1316.
22. Häse F. and Zacharias M. 2016. Free energy analysis and mechanism of base pair stacking in nicked DNA. *Nucleic Acids Res.* 44:7100-7108.
23. Israels B., Albrecht C. S., Dang A., Barney M., von Hippel P. H., and Marcus A. H. 2021. 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.
24. Qu P., Yang X., Li X., Zhou X., and Zhao X. S. 2010. 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.
25. Rivetti C., Walker C., and Bustamante C. 1998. Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. *J. Mol. Biol.* 280:41-59.
26. Lin S., Horning D. P., Szostak J. W., and Chaput J. C. 2009. Conformational analysis of DNA repair intermediates by time-resolved fluorescence spectroscopy. *J. Phys. Chem. A* 113:9585-9587.
27. Kashida H., Kurihara A., Kawai H., and Asanuma H. 2017. Orientation-dependent FRET system reveals differences in structures and flexibilities of nicked and gapped DNA duplexes. *Nucleic Acids Res.* 45:e105-e105.
28. Roll C., Ketterlé C., Faibis V., Fazakerley G. V., and Boulard Y. 1998. Conformations of nicked and gapped DNA structures by NMR and molecular dynamic simulations in water. *Biochem.* 37:4059-4070.
29. Craggs T. D., Sustarsic M., Plochowitz A., Mosayebi M., Kaju H., Cuthbert A., Hohlbein J., Domicieva L., Biggin P. C., and Doye J. P. K. 2019. Substrate conformational dynamics facilitate structure-specific recognition of gapped DNA by DNA polymerase. *Nucleic Acids Res.* 47:10788-10800.
30. Chen H., Meisburger S. P., Pabit S. A., Sutton J. L., Webb W. W., and Pollack L. 2012. Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci.* 109:799-804.
31. Dans P. D., Balaceanu A., Pasi M., Patelli A. S., Petkevičiūtė D., Walther J., Hospital A., Bayarri G., Lavery R., and Maddocks J. H. 2019. The static and dynamic structural heterogeneities of B-DNA: extending Calladine–Dickerson rules. *Nucleic Acids Res.* 47:11090-11102.
32. Harris P. D., Hamdan S. M., and Habuchi S. 2020. Relative contributions of base stacking and electrostatic repulsion on DNA nicks and gaps. *J. Phys. Chem. B* 124:10663-10672.
33. Kervio E., Claasen B., Steiner U. E., and Richert C. 2014. The strength of the template effect attracting nucleotides to naked DNA. *Nucleic Acids Res.* 42:7409-7420.

34. Izgu E. C., Fahrenbach A. C., Zhang N., Li L., Zhang W., Larsen A. T., Blain J. C., and Szostak J. W. 2015. Uncovering the thermodynamics of monomer binding for RNA replication. *J. Am. Chem. Soc.* 137:6373-6382.
35. Tam C. P., Fahrenbach A. C., Björkbohm A., Prywes N., Izgu E. C., and Szostak J. W. 2017. Downstream oligonucleotides strongly enhance the affinity of GMP to RNA primer-template complexes. *J. Am. Chem. Soc.* 139:571-574.
36. Kashida H., Kokubo Y., Makino K., and Asanuma H. 2019. Selective binding of nucleosides to gapped DNA duplex revealed by orientation and distance dependence of FRET. *Org. Biomol. Chem.* 17:6786-6789.
37. Spasic A., Berger K. D., Chen J. L., Seetin M. G., Turner D. H., and Mathews D. H. 2018. Improving RNA nearest neighbor parameters for helices by going beyond the two-state model. *Nucleic Acids Res.* 46:4883-4892.
38. Salditt A., Keil L. M., Horning D. P., Mast C. B., Joyce G. F., and Braun D. 2020. Thermal habitat for RNA amplification and accumulation. *Phys. Rev. Lett.* 125:048104.
39. Mast C. B., Schink S., Gerland U., and Braun D. 2013. Escalation of polymerization in a thermal gradient. *Proc. Natl. Acad. Sci.* 110:8030-8035.
40. Ianeselli A., Mast C. B., and Braun D. 2019. Periodic melting of oligonucleotides by oscillating salt concentrations triggered by microscale water cycles inside heated rock pores. *Angew. Chem. Int. Ed.* 131:13289-13294.
41. 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.
42. Banyay M., Sarkar M., and Gräslund A. 2003. A library of IR bands of nucleic acids in solution. *Biophys. Chem.* 104:477-488.
43. Mallat S. G. 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 11:674-693.
44. 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.
45. 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.
46. 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.
47. Jones K. C., Ganim Z., and Tokmakoff A. 2009. Heterodyne-detected dispersed vibrational echo spectroscopy. *J. Phys. Chem. A* 113:14060-14066.
48. 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.
49. 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.
50. 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.
51. Abraham M. J., Murtola T., Schulz R., Páll S., Smith J. C., Hess B., and Lindahl E. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1:19-25.
52. 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.
53. 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.

54. Tan D., Piana S., Dirks R. M., and Shaw D. E. 2018. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl. Acad. Sci.* 115:E1346-E1355.
55. Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W., and Klein M. L. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926-935.
56. Tucker M. R., Piana S., Tan D., LeVine M. V., and Shaw D. E. 2022. Development of Force Field Parameters for the Simulation of Single-and Double-Stranded DNA Molecules and DNA-Protein Complexes. *J. Phys. Chem. B* 126:4442-4457.
57. Bussi G., Donadio D., and Parrinello M. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.
58. 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.
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. Diekmann S. 1989. Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.* 205:787-791.
61. Bayarri G., Andrio P., Hospital A., Orozco M., and Gelpí J. L. 2022. BioExcel Building Blocks Workflows (BioBB-Wfs), an integrated web-based platform for biomolecular simulations. *Nucleic Acids Res.* 50:W99-W107.
62. Andrio P., Hospital A., Conejero J., Jordá L., Del Pino M., Codo L., Soiland-Reyes S., Goble C., Lezzi D., and Badia R. M. 2019. BioExcel Building Blocks, a software library for interoperable biomolecular simulation workflows. *Sci. Data* 6:1-8.
63. Lavery R., Moakher M., Maddocks J. H., Petkeviciute D., and Zakrzewska K. 2009. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* 37:5917-5929.
64. Blanchet C., Pasi M., Zakrzewska K., and Lavery R. 2011. CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res.* 39:W68-W73.
65. 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.
66. Waskom M. L. 2021. Seaborn: statistical data visualization. *J. Open Source Softw.* 6:3021.
67. Peng C. S., Jones K. C., and Tokmakoff A. 2011. Anharmonic vibrational modes of nucleic acid bases revealed by 2D IR spectroscopy. *J. Am. Chem. Soc.* 133:15650-15660.
68. Lee C. and Cho M. 2006. Vibrational dynamics of DNA. II. Deuterium exchange effects and simulated IR absorption spectra. *J. Chem. Phys.* 125:114509.
69. Ashwood B., Jones M. S., Ferguson A. L., and Tokmakoff A. 2022. Disruption of Energetic and Dynamic Base Pairing Cooperativity in DNA Duplexes by an Abasic Site. *Proc. Natl. Acad. Sci.* 10.1073/pnas.2219124120.
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. Wang S. and Kool E. T. 1995. Origins of the large differences in stability of DNA and RNA helices: C-5 methyl and 2'-hydroxyl effects. *Biochem.* 34:4125-4132.
72. Huang Y., Weng X., and Russu I. M. 2011. Enhanced base-pair opening in the adenine tract of a RNA double helix. *Biochem.* 50:1857-1863.
73. Bernasconi C. (1976) *Relaxation kinetics* (Academic Press).
74. 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.
75. Chen C., Wang W., Wang Z., Wei F., and Zhao X. S. 2007. Influence of secondary structure on kinetics and reaction mechanism of DNA hybridization. *Nucleic Acids Res.* 35:2875-2884.

76. 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.
77. Hori N., Denesyuk N. A., and Thirumalai D. 2018. Frictional effects on RNA folding: Speed limit and Kramers turnover. *J. Phys. Chem. B* 122:11279-11288.
78. 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.
79. Fontana F., Bellini T., and Todisco M. 2022. Liquid Crystal Ordering in DNA Double Helices with Backbone Discontinuities. *Macromolecules* 55:5946-5953.
80. Scott D. W. 1979. On optimal and data-based histograms. *Biometrika* 66:605-610.
81. Stone S. E., Ray D., and Andricioaei I. 2022. Force Field Dependent DNA Breathing Dynamics: A Case Study of Hoogsteen Base Pairing in A6-DNA. *J. Chem. Inf. Model.* 62:6749-6761.
82. Pérez de Alba Ortiz A., Vreede J., and Ensing B. 2022. Sequence dependence of transient Hoogsteen base pairing in DNA. *PLOS Comput. Biol.* 18:e1010113.
83. Sawada T., Yoshizawa M., Sato S., and Fujita M. 2009. Minimal nucleotide duplex formation in water through enclathration in self-assembled hosts. *Nat. Chem.* 1:53-56.
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. Rejali N. A., Ye F. D., Zuijter A. M., Keller C. C., and Wittwer C. T. 2021. Nearest-neighbour transition-state analysis for nucleic acid kinetics. *Nucleic Acids Res.* 49:4574-4585.
86. Szabo A. 1978. Kinetics of hemoglobin and transition state theory. *Proc. Natl. Acad. Sci.* 75:2108-2111.
87. Eaton W. A., Henry E. R., and Hofrichter J. 1991. Application of linear free energy relations to protein conformational changes: the quaternary structural change of hemoglobin. *Proc. Natl. Acad. Sci.* 88:4472-4475.
88. Oweida T. J., Kim H. S., Donald J. M., Singh A., and Yingling Y. G. 2021. Assessment of AMBER force fields for simulations of ssDNA. *J. Chem. Theory Comput.* 17:1208-1217.
89. Roll C., Ketterlé C., Fazakerley G. V., and Boulard Y. 1999. Solution structures of a duplex containing an adenine opposite a gap (absence of one nucleotide) An NMR study and molecular dynamic simulations with explicit water molecules. *Eur. Biophys. J.* 264:120-131.
90. Guo H. and Tullius T. D. 2003. Gapped DNA is anisotropically bent. *Proc. Natl. Acad. Sci.* 100:3743-3747.
91. Zhang W., Tam C. P., Walton T., Fahrenbach A. C., Birrane G., and Szostak J. W. 2017. Insight into the mechanism of nonenzymatic RNA primer extension from the structure of an RNA-GpppG complex. *Proc. Natl. Acad. Sci.* 114:7659-7664.
92. Zhang W., Tam C. P., Zhou L., Oh S. S., Wang J., and Szostak J. W. 2018. Structural rationale for the enhanced catalysis of nonenzymatic RNA primer extension by a downstream oligonucleotide. *J. Am. Chem. Soc.* 140:2829-2840.
93. Zhang W., Walton T., Li L., and Szostak J. W. 2018. Crystallographic observation of nonenzymatic RNA primer extension. *Elife* 7:e36422.
94. 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.
95. Chen Z. Y. and Noolandi J. 1992. Renormalization-group scaling theory for flexible and wormlike polymer chains. *J. Chem. Phys.* 96:1540-1548.
96. Giurgiu C., Fang Z., Aitken H. R., Kim S. C., Pazienza L., Mittal S., and Szostak J. W. 2021. Structure–Activity Relationships in Nonenzymatic Template-Directed RNA Synthesis. *Angew. Chem. Int. Ed.* 133:23107-23114.
97. Conte M. R., Conn G. L., Brown T., and Lane A. N. 1997. Conformational properties and thermodynamics of the RNA duplex r (CGCAAUUUGCG) 2: comparison with the DNA analogue d (CGCAATTGCG) 2. *Nucleic Acids Res.* 25:2627-2634.

98. Cruz-León S. and Schwierz N. 2022. RNA Captures More Cations than DNA: Insights from Molecular Dynamics Simulations. *J. Phys. Chem. B* 126:8646-8654.
99. Auffinger P. and Westhof E. 2001. Water and ion binding around r (UpA) 12 and d (TpA) 12 Oligomers-comparison with RNA and DNA (CpG) 12 duplexes. *J. Mol. Biol.* 305:1057-1072.
100. Abels J., Moreno-Herrero F., Van der Heijden T., Dekker C., and Dekker N. H. 2005. Single-molecule measurements of the persistence length of double-stranded RNA. *Biophys. J.* 88:2737-2744.
101. Ding D., Zhou L., Giurgiu C., and Szostak J. W. 2022. Kinetic explanations for the sequence biases observed in the nonenzymatic copying of RNA templates. *Nucleic Acids Res.* 50:35-45.
102. Stootman F. H., Fisher D. M., Rodger A., and Aldrich-Wright J. R. 2006. Improved curve fitting procedures to determine equilibrium binding constants. *Analyst* 131:1145-1151.
103. Movileanu L., Benevides J. M., and Thomas Jr G. J. 1999. Temperature dependence of the raman spectrum of DNA. Part I—Raman signatures of premelting and melting transitions of poly (dA–dT)· poly (dA–dT). *J. Raman Spectrosc.* 30:637-649.
104. Owczarzy R. 2005. Melting temperatures of nucleic acids: discrepancies in analysis. *Biophys. Chem.* 117:207-215.
105. Mergny J.-L. and Lacroix L. 2003. Analysis of thermal melting curves. *Oligonucleotides* 13:515-537.
106. Krummel A. T. and Zanni M. T. 2006. DNA vibrational coupling revealed with two-dimensional infrared spectroscopy: Insight into why vibrational spectroscopy is sensitive to DNA structure. *J. Phys. Chem. B* 110:13991-14000.
107. Lee C., Park K.-H., Kim J.-A., Hahn S., and Cho M. 2006. Vibrational dynamics of DNA. III. Molecular dynamics simulations of DNA in water and theoretical calculations of the two-dimensional vibrational spectra. *J. Chem. Phys.* 125:114510.
108. Braunlin W. H. and Bloomfield V. A. 1991. Proton NMR study of the base-pairing reactions of d (GGAATTCC): salt effects on the equilibria and kinetics of strand association. *Biochem.* 30:754-758.
109. Mikulecky P. J. and Feig A. L. 2006. Heat capacity changes associated with DNA duplex formation: salt-and sequence-dependent effects. *Biochem.* 45:604-616.
110. Marky L. A. and Breslauer K. J. 1987. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* 26:1601-1620.
111. Böttcher A., Kowerko D., and Sigel R. K. 2015. Explicit analytic equations for multimolecular thermal melting curves. *Biophys. Chem.* 202:32-39.
112. van Stokkum I. H., Larsen D. S., and Van Grondelle R. 2004. Global and target analysis of time-resolved spectra. *Biochim. Biophys. Acta* 1657:82-104.
113. 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.
114. Lee Y., Gorka M., Golbeck J. H., and Anna J. M. 2018. Ultrafast energy transfer involving the red chlorophylls of cyanobacterial photosystem I probed through two-dimensional electronic spectroscopy. *J. Am. Chem. Soc.* 140:11631-11638.
115. Freire E., Schön A., and Velazquez-Campoy A. 2009. Isothermal titration calorimetry: general formalism using binding polynomials. *Meth. Enzymol.* 455:127-155.
116. Todisco M. and Szostak J. W. 2022. Hybridization kinetics of out-of-equilibrium mixtures of short RNA oligonucleotides. *Nucleic Acids Res.* 50:9647-9662.
117. Chen Z. and Znosko B. M. 2013. Effect of sodium ions on RNA duplex stability. *Biochem.* 52:7477-7485.
118. Invernizzi M. 2021. OPES: On-the-fly Probability Enhanced Sampling Method. *arXiv arXiv:2101.06991*. <https://arxiv.org/abs/2101.06991> (preprint posted Jan. 18 2021).
119. Invernizzi M. and Parrinello M. 2020. Rethinking metadynamics: from bias potentials to probability distributions. *J. Phys. Chem. Lett.* 11:2731-2736.

120. Laio A. and Parrinello M. 2002. Escaping free-energy minima. *Proc. Natl. Acad. Sci.* 99:12562-12566.
121. Ensing B., De Vivo M., Liu Z., Moore P., and Klein M. L. 2006. Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Acc. Chem. Res.* 39:73-81.
122. Best R. B., Hummer G., and Eaton W. A. 2013. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci.* 110:17874-17879.
123. 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.
124. 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.
125. 2019. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* 16:670-673.
126. Berg O. G. and von Hippel P. H. 1985. Diffusion-controlled macromolecular interactions. *Annu. Rev. Biophys.* 14:131-158.
127. Stejskal E. O. and Tanner J. E. 1965. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *J. Chem. Phys.* 42:288-292.