5-Carboxylcytosine and Cytosine Protonation Distinctly Alter the Stability and Dehybridization Dynamics of the DNA Duplex

Brennan Ashwood[1,2,3], Paul J. Sanstead[1,2,3], Qing Dai[1,2], Chuan He[1,2,4,5], Andrei Tokmakoff[1,2,3,*]

[1]Department of Chemistry, [2]Institute for Biophysical Dynamics, [3]James Franck Institute, [4]Department of Biochemistry and Molecular Biology, and [5]Howard Hughes Medical Institute, The University of Chicago, Chicago, Illinois 60637, United States

Abstract

Applications associated with nucleobase protonation events are grounded in their fundamental impact on DNA thermodynamics, structure, and hybridization dynamics. Of the canonical nucleobases, N3 protonation of cytosine (C) is the most widely utilized in both biology and nanotechnology. Naturally occurring C derivatives that shift the N3 pK_a introduce an additional level of tunability. The epigenetic nucleobase 5-carboxylcytosine (caC) presents a particularly interesting example since this derivative forms Watson-Crick base pairs of similar stability and displays pH-dependent behavior over the same range as the canonical nucleobase. However, the titratable group in caC corresponds to the exocyclic carboxyl group rather than N3, and the implications of these divergent protonation events toward DNA hybridization thermodynamics, kinetics, and base pairing dynamics remain poorly understood. Here, we study the pH-dependence of these physical properties using model oligonucleotides containing C and caC with FTIR and temperature-jump IR spectroscopy. We demonstrate that N3 protonation of C completely disrupts duplex stability, leading to large shifts in the duplex/single-strand equilibrium, a reduction in the cooperativity of melting, and an acceleration in the rate of duplex dissociation. In contrast, while increasing 5-carboxyl protonation in caC-containing duplexes induces an increase in base pair fluctuations, the DNA duplex can tolerate substantial protonation without significant perturbation to the duplex/single-strand equilibrium. However, 5-carboxyl protonation has a large impact on hybridization kinetics by reducing the transition state free energy. Our thermodynamic and kinetic analysis provides new insight on the impact of two divergent protonation mechanisms in naturally occurring nucleobases on the biophysical properties of DNA.

Introduction

The physical and biological properties of DNA are highly sensitive to environmental factors such as temperature,[1-4] counterions,[5] and pH.[6-9] In particular, pH effects can greatly impact nucleic acids through direct protonation or de-protonation of specific nucleobase sites, leading to

changes in secondary structure[6, 10-11] and duplex dissociation.[7, 12-13] Proton concentrations across different intracellular compartments are highly regulated and vary substantially from pH 8.0 in mitochondria to pH 7.2 in the nucleus to pH 4.7 in lysosomes.[14] Different local environments regulate DNA properties and can potentially lead to damage. Even before the determination of DNA's double helical structure,[15] it had been shown that polymeric DNA dissociates under alkaline and acidic conditions.[8] Since then, many details related to the identities of protonated species,[6-7, 16-18] conformational changes,[7, 16] and their thermodynamic impact on duplex DNA (dsDNA) have been investigated.[7, 13, 19] It has been demonstrated that the thermal melting temperature of DNA is greatly reduced under acidic conditions and largely stems from perturbations to G:C base pairing.[6, 17-18, 20] As a result, the degree of protonation tolerated in dsDNA and its thermodynamic impact is highly-dependent on the level of G:C content.

In general, the pH-dependent properties of nucleic acids observed in biology are driven by protonation at the N3 position (Fig. 1) of cytosine (C) due to its relatively high and tunable pK_a compared to other protonatable bases.[10] In addition to initiating mismatches and G:C Hoogsteen base pairs,[21] N3 protonation drives the formation of C quadruplex structures known as intercalated motifs (i-motifs).[10] Recently, i-motif structures have been observed *in vivo*, particularly in human promoter regions, suggesting that they may play a significant role in gene regulation.[22]

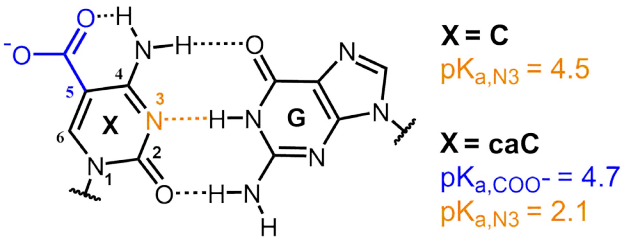


Figure 1: Chemical structure of X:G base pair, where X corresponds to cytosine (C) or 5-carboxylcytosine (caC). Titratable groups for C and caC are shown with their respective pK_a values for the isolated nucleoside in solution.[23]

While N3 protonation of C is correlated to most pH-induced function of canonical nucleic acids, numerous modified nucleobases exist *in vivo* that have unique pH-dependent properties. For example, 5-bromination lowers the pK_a of uracil from ~10 to ~8,[24] allowing for significant N3 deprotonation under physiological conditions that may facilitate damage to nucleic acids.[25] More recently, cytosine derivatives involved in the active DNA demethylation pathway, where

5-methylcytosine (mC) is sequentially oxidized to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and finally 5-carboxylcytosine (caC),[26-27] have been discovered in mammals. In this context fC and caC are selectively excised by thymine DNA glycosylase (TDG) to give an abasic site which is repaired with C by a base excision repair (BER) pathway to complete the demethylation cycle.[27] Neither mC nor hmC are associated with significant shifts in N3 pK_a (p$K_{a,N3}$). In contrast, fC exhibits a reduced p$K_{a,N3}$ of 2.4 but shows no pH-dependence in its excision activity, while caC has two titratable groups with pK_a values of 2.1-2.4 (p$K_{a,N3}$) and 4.2-4.8 (C5-COO^-, pK_{a,COO^-}), [23, 28-29] and shows accelerated excision under mildly acidic conditions.[28, 30] The pH-dependent behavior of the caC nucleobase stands out among the epigenetic cytosine derivatives since it has a pK_a in a similar range to C, but the titratable group associated with this protonation is the exocyclic carboxyl group rather than N3. While addition of a proton at N3 occupies a hydrogen bond acceptor site that blocks the formation of a Watson-Crick base pair (Fig. 1), protonation of the exocyclic carboxyl group situated in the major groove would appear less perturbative, but could nevertheless lead to unique pH-dependent thermodynamic or kinetic effects in modified dsDNA.

In nucleic acid nanotechnology, the sensitivity of DNA to C protonation has motivated the design of numerous pH-driven nanodevices.[31-32] In particular, the formation of i-motif and DNA triple helices in C- and CG-rich sequences, respectively, require C protonation and offer tunability of nucleic acid secondary structure.[10, 32] These secondary structures are highly sensitive to sequence composition and environmental factors, which can shift p$K_{a,N3}$ by 0.5-1 units.[10] Such shifts allow for the construction of uniquely pH-sensitive DNA tools for a variety of applications. For example, pH-driven conformational switches have been used to enhance the location specificity of drug-delivery agents,[33] increase control over DNA strand displacement equilibria,[34] and develop sensors to map the pH of biological environments.[35] While C N3 protonation is the primary mechanism behind these pH-devices, more recent work has demonstrated enhanced tunability through the incorporation of modified nucleobases.[36-38] In particular, substitution of C to caC has been shown to destabilize i-motif formation in many sequences with high sensitivity to the position of modification as well as destabilize DNA triplex formation under physiological conditions.[38-39] Destabilization of each type of structure is most likely due to the reduction in p$K_{a,N3}$ upon 5-carboxylation. However, the dynamic properties, such as folding rate and mechanism, of pH-driven devices are also essential to their function. As has been shown for 5-bromocytidine,[36]

caC may be able to tune specific aspects of i-motif folding or DNA triplex association dynamics important for nanotechnology.

The pH-dependent properties of canonical and modified DNA utilized in biology and nanotechnology are determined by the impact of nucleobase protonation on local and global base pairing properties. Therefore, a detailed understanding of the pH-dependence of hybridization thermodynamics, kinetics, and structural dynamics is warranted. The incorporation of caC may lead to unique pH-dependences in these properties compared to canonical DNA that are nevertheless tunable over the same pH range due to the fact that pK_{a,COO^-} is similar to $pK_{a,N3}$ in C, the 5-position is oriented in the major groove, and protonation of the exocyclic carboxyl group may not directly disturb caC:G base pairing. So far, the biophysical impact of caC has primarily been studied at neutral pH, showing only minor perturbations to double-stranded DNA.[39-42] However, the protonated carboxyl group is much more electron-withdrawing[43] and caC has a reduced $pK_{a,N3}$, which has been proposed to weaken G:caC base pairing.[23] Additionally, simulations have suggested that 5-carboxyl protonation may tune local solvation and increase local base pair fluctuations.[42]

Despite interest in the pH-dependent properties of C and caC, a detailed understanding of nucleobase protonation on duplex dissociation, particularly the kinetics and base pairing dynamics involved in the process, remains incomplete. Here, we characterize the role of 5-carboxyl and N3 protonation on DNA duplex thermodynamics and dehybridization dynamics of a ten-nucleotide sequence using FTIR and temperature-jump infrared (IR) spectroscopy. The mid-IR spectrum of nucleic acids is sensitive to nucleobase-specific changes in base pairing and protonation,[23, 44-48] making IR spectroscopy particularly useful for the study of pH-dependent hybridization of DNA. We demonstrate that N3 protonation of C drastically shifts the duplex/single-strand equilibrium and reduces the barrier to dissociation without significantly impacting the hybridization transition state energy. In contrast, caC-modified dsDNA can handle essentially complete protonation of the 5-carboxyl sites without significantly shifting the duplex-to-single-strand transition. Instead, 5-carboxyl protonation leads to highly sloped melting curve baselines that are interpreted as a reduction in the internal base pairing within duplexes. Protonation of caC still leads to a large reduction in the dissociation barrier, but, in contrast to N3 protonation of C, does so through lowering the transition state free energy rather than destabilizing the duplex state. Our results

demonstrate that caC and C impart distinct pH-dependent properties onto dsDNA, providing insight into the biological function of these protonation events as well as their potential utility in nucleic acid nanotechnology.

Results

pH-Dependent FTIR Melting Experiments on caC- and C-Containing Oligonucleotides

As an initial assessment of the pH dependence of hybridization, Fig. 2 shows FTIR temperature series of the sequence 5′-TAXGXGXGTA-3′, when X = caC (Fig. 2a,b), and when X = C (Fig. 2d,e), prepared at pH* 6.8 and 3.7 and ranging from 3 to 97 °C in ca. 4.5 °C steps. Samples are prepared in deuterated buffer to eliminate the background from the H_2O bend vibration.[49] D_2O has been shown to have a negligible influence on DNA thermodynamics relative to H_2O.[50] The solution pD is 0.44 greater than the measured pH*,[51] so the deuteron concentration at pH* 6.8 corresponds to standard physiological values. The in-plane ring vibrations, carbonyl stretches, and -ND_2 bends of the nucleobases absorb in the 1500-1750 cm^{-1} frequency range and each base contributes a unique vibrational fingerprint to the oligonucleotide spectrum.[44, 46-47]

At pH* 6.8 and 3 °C, both sequences exist overwhelmingly as duplexed DNA. The corresponding spectra of these XG-rich duplexes (Fig. 2a,d) appear similar to one another, although the carboxyl group of caC mixes strongly with the in-plane base vibrations and results in an additional absorption near 1570 cm^{-1}.[23] As the temperature increases, the spectrum in this frequency range is reshaped significantly and several prominent changes are indicative of dehybridization. Below 1600 cm^{-1}, the spectrum is dominated by G ring mode absorptions whose intensity increases as X:G base pairing is disrupted. A similar increase corresponding to A ring mode absorption is observed near 1625 cm^{-1}. At higher frequency, the spectrum is congested with multiple overlapping peaks, with the growth in intensity near 1665 cm^{-1} corresponding to T, X, and G carbonyl absorptions. The increase due to G at this frequency is accompanied by a drop in absorption near 1685 cm^{-1} since this G carbonyl mode is shifted 20 cm^{-1} when engaged in a Watson-Crick base pair. The highest frequency T carbonyl at 1690 cm^{-1} can also contribute to this intensity reduction upon the loss of T:A base pairing.

Reducing the pH* of the oligonucleotide solution significantly alters the FTIR spectrum, particularly at low temperature, as evident through comparison of the pH* 6.8 spectra in Fig. 2a,d against the pH* 3.7 spectra in Fig. 2b,e. FTIR spectra measured at intermediate pH* points at both

low and high temperature are shown in Fig. S1. When X = caC, a loss in absorbance is observed near 1575 and 1625 cm^{-1} as well as a shift of the 1650 cm^{-1} feature to higher frequency and gain at 1700 cm^{-1}. These spectral changes are consistent with the pH*-dependent FTIR spectra of 2′-deoxy-5′-carboxylcytidine (dcaC) (Fig. S1c & S2). Therefore, we can assign the pH*-dependent changes of X = caC to protonation of the caC nucleobase. Adenine (AMP N1 pK_a: 3.9-4.0) and guanine (GMP N7 pK_a: 3.0-3.3) can also be protonated under acidic conditions,[52-53] but comparison of the spectra in Fig. 2b,e against the pH-dependent FTIR spectra of these nucleosides suggests there is little of these protonation events over the studied pH* range (Fig. S4). The FTIR spectra of X = caC additionally reveal a reduction in amplitude and shift of the 1685 cm^{-1} guanine mode to lower frequency as caC is protonated. The loss of this feature at neutral pH reports on DNA duplex dissociation, and is always accompanied by an increase in absorbance of the G peaks centered at 1575 and 1665 cm^{-1}, as discussed above. For X = caC, the absorption at 1575 cm^{-1} instead decreases in amplitude due to shifts in the caC ring vibration upon protonation, suggesting that the loss of the 1685 cm^{-1} feature at low temperature does not result from duplex dissociation. The frequency and amplitude of the 1685 cm^{-1} peak is known to be sensitive to DNA conformation as well as the local base pairing and stacking environment,[44-45, 47, 54-55] and therefore numerous factors (Fig. S6 & S7) could be responsible for its modulation as discussed in the SI.

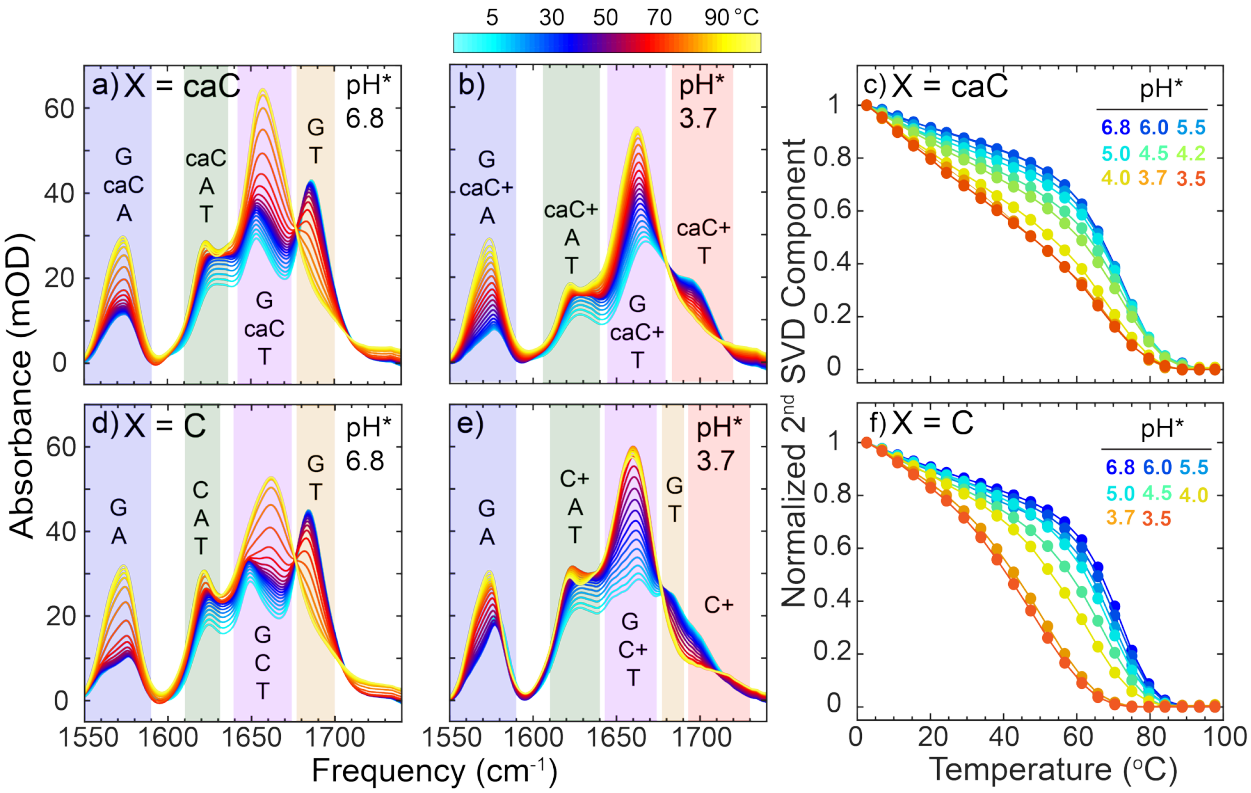


Figure 2: FTIR temperature series of 5′-TAXGXGXGTA-3′ for X = caC prepared at (a) pH* 6.8 and (b) pH* 3.7 from 3 to 97 °C. Colored windows indicate in-plane nucleobase vibration contributions to each frequency region with the degree of contribution decreasing from top to bottom within a region. (c) Normalized 2nd components from singular value decomposition (SVD) across FTIR temperature series from pH* 6.8 to 3.5. (d,e,f) Corresponding figures for X = C at pH* 6.8 and 3.7.

From pH* 6.8 to 3.5, X = C exhibits a shift in the 1650 cm^{-1} peak to higher frequency and growth of a band near 1700 cm^{-1} at low temperature (Fig. 2e). Both of these spectral changes are consistent with pH*-dependent FTIR spectra of 2′-deoxycytidine (dC), suggesting N3 of C is protonated when X = C over this pH* range, in agreement with the reported p$K_{a,N3}$ of 4.2-4.5.[17, 23, 29] The significant loss and gain at 1685 and 1660 cm^{-1}, respectively, as well as the intensity gain at 1575 cm^{-1} also suggest changes to guanine either through protonation or a reduction in base pairing. The pK_a of guanine's N7 position (3-3.3) is much lower than p$K_{a,N3}$ of C,[52] and signatures of guanine protonation (Fig. S4) are not present in either the low or high-temperature oligomer spectra. Instead, the intensity changes in the guanine features must be attributed to the loss of G:C base pairing due to N3 protonation of C.

To further characterize the pH*-dependent spectral changes from N3 protonation, we performed an FTIR titration of the X = C sequence from pH* 6.8 to 1.8 at 3 °C (Fig. 3a). Over this pH* range, DNA dissociates to single strands due to protonation of C, A, G, and the phosphate backbone. The spectral changes as the pH* of the solution decreases in Fig. 3a are dominated by features that indicate the loss of G:C base pairing coincident with N3 protonation of C, as assigned above. Evidence of N1 and N7 protonation of A and G, respectively, is only observed at the lowest pH* points sampled (Fig. S4 & S5). Therefore, the 2nd SVD component of the FTIR spectra across this pH* range should report on the N3 C protonation-driven duplex-to-single-strand transition and can be used to estimate the fraction of intact duplexes at 3 °C as a function of pH*. When fit to a Henderson-Hasselbalch equation, the 2nd SVD component duplex-to-single-strand and N3 protonation trend show an inflection point of 3.9 ± 0.1 that corresponds to the acid denaturation melting point (pH_m) of the duplex/single-strand equilibrium at 3 °C. In analogy to the melting temperature (T_m) in thermal denaturation experiments, pH_m corresponds to the pH where half of the possible duplex species are intact. The value of pH_m for X = C is consistent with those previously measured for polymeric DNA.[7]

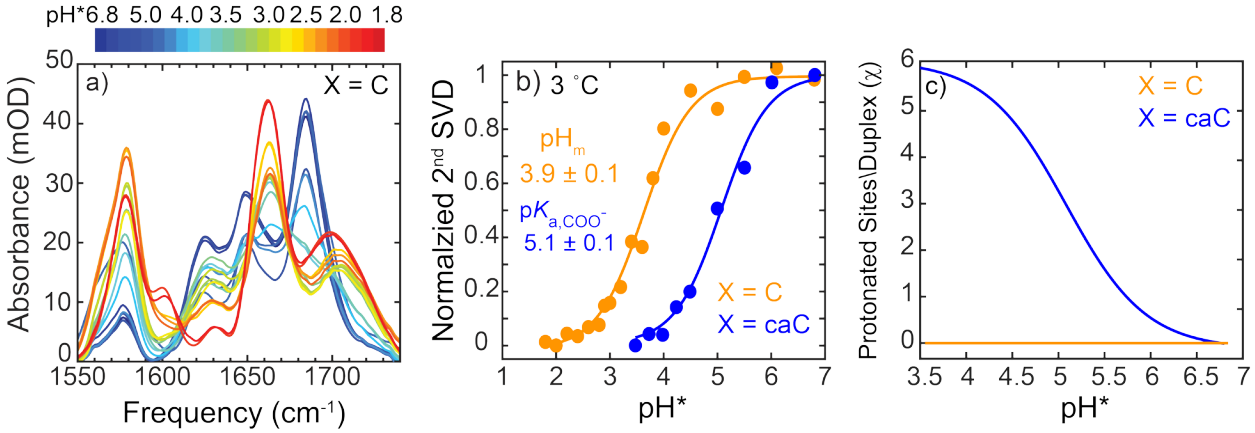


Figure 3: (a) FTIR titration of X = C from pH* 6.8 to 1.8 at 3 °C. (b) 2nd SVD component of FTIR titration on X = C (orange) and X = caC (blue). The 2nd SVD component of X = C and X = caC are fit to eq. S10 and eq. S6, respectively, to extract the inflection point and degree of protonation in the duplex and single-strand. The error in pH_m and pK_{a,COO^-} represent the 95% confidence interval of the fit. (c) Protonated sites per duplex (χ) calculated for X = caC (blue) and X = C (orange) as a function of pH*. Details behind the derivation of eq. S6 and S10 are presented in the SI.

Our pH*-dependent FTIR measurements of X = C and X = caC may be used to estimate the number of protonated C and caC sites, respectively, per duplex (χ) as a function of pH*. For X = caC, an apparent pK_{a,COO^-} of 5.1 is determined from SVD across pH* at 3 °C (Fig. 3b). The apparent pK_{a,COO^-} shows no measurable change from 3 to 97 °C (Fig. S3), suggesting that pK_{a,COO^-} of the X = caC duplex and single-strand are similar. Therefore, we assume that the 5-carboxyl protonation equilibrium for X = caC is identical in the duplex and single-strand and does not strongly depend on temperature. The small change in pK_{a,COO^-} with temperature is consistent with previous temperature-dependent measurements of nucleic acid protonation.[56] Additionally, the X = caC sequence is assumed to be highly duplexed at low temperature across the pH* range studied here based on the spectroscopic signatures consistent with extensive caC:G base pairing observed at pH* 3.7 at 3 °C discussed above. The resulting degree of X = caC duplex protonation as a function of pH* is shown in Fig. 3c. The X = caC duplex is almost completely deprotonated at pH* 6.8 whereas an average of 5+ sites per duplex are protonated below pH* 4.4. Further details regarding our estimation of the degree of protonation are presented in the SI.

As discussed above, the pH*-induced duplex-to-single-strand transition in the X = C sequence appears to be driven by N3 protonation of C. The convolution of these two processes indicates that N3 protonation of C occurs over the same pH*-range as duplex denaturation, but it is not clear how protonation is distributed among duplex and single-strand species. Assuming the effective p$K_{a,N3}$ and pH_m at 3 °C are equivalent for X = C, the fraction of intact duplex species (θ_{ext}) and degree of N3 protonation in the duplex and single-strand can be related as a function of pH* (See SI for details). As shown in Fig. 3c, X = C is essentially only protonated in the single-strand while the duplex cannot tolerate N3 protonation at even one site. This result may also be interpreted as a significant reduction of p$K_{a,N3}$ in the duplex relative to the single-strand,[57] leading to a sharp decline in duplex fraction at pH* values where single-strand protonation becomes favorable.

Having identified the spectroscopic signatures of dehybridization and nucleobase protonation, we can assess the pH* dependent melting behavior of the two model sequences. To take into account the global changes to the spectrum upon melting, a SVD is performed on the FTIR temperature series at each pH* to determine a melting curve.[45, 58] The melting curves measured from pH* 6.8 to 3.5 are shown in Fig. 2c and 2f for the X = caC and C sequences,

respectively. While the two sequences have nearly identical melting profiles at pH* 6.8, they exhibit starkly different trends with decreasing pH*. The X = caC melting profile shows only minor shifts in the inflection point of the sigmoidal melting transition with descending pH*, but the low-temperature baseline slope steepens significantly under acidic conditions. In contrast, the melting transition of X = C shifts drastically to lower temperature and broadens as pH* is reduced. Baseline slopes are often observed in thermal melting curves and are attributed to factors such as evaporation, temperature-dependent changes in sample path length, or DNA base solvation.[59-60] However, a comparison of spectral changes along the low-temperature baseline when X = caC (Fig. S8) reveals gains in intensity at 1575 and 1660 cm^{-1}, indicating the loss of G:C base pairing, at pH* 3.7 that are not observed under neutral conditions. This observation as well as the abnormally large changes in baseline slope with decreasing pH* when X = caC suggest that the low-temperature baseline is related to changes in base pairing and base stacking within the duplex state. Based on the pH* and temperature dependence of the mid-IR spectra, the pH* dependence of the melting curves, and the assignment of the extent of protonation in the duplex state, we conclude that there is distinct pH-dependent melting behavior due to divergent protonation events in C- and caC-containing duplexes. Whereas the X = C sequence protonates at N3 C over this pH* range and appears unable to tolerate protonation in the duplex state, the X = caC sequence protonates at the exocyclic 5-carboxyl group, resulting in a duplex much more robust to reductions in pH.

pH-Dependent Relaxation Kinetics Monitored with Temperature-Jump Spectroscopy

In addition to equilibrium melting experiments, we employed transient temperature jump (T-jump) measurements to assess the kinetic and dynamic impact of nucleobase protonation in the X = C and caC oligonucleotides. The difference heterodyned dispersed vibrational echo (t-HDVE) spectrum, which can be interpreted like a pump-probe difference spectrum,[61] is used to track changes to the DNA ensemble at delays following the T-jump. Illustrative time traces tracked at the most intense difference feature of 1670 cm^{-1}, which contains contributions from G:C and A:T base pairing, are shown for the X = C and caC sequences at pH* 6.8 and 3.7 in Fig. 4. For both sequences and at each pH*, three distinct timescales are observed: (1) a small amplitude rise within ~200 ns (λ_{ns}), (2) a larger rise near ~100 μs ($\lambda_{\mu s}$), and (3) decay of the difference signal due to thermal relaxation and re-hybridization in ~ 2 ms. As shown in the SI using temperature-jump two-dimensional IR spectroscopy (t-2DIR, Fig. S9), λ_{ns} is primarily assigned to fraying of the A:T

termini while $\lambda_{\mu s}$ corresponds to the duplex-to-single-strand transition. Previous T-jump measurements of short DNA oligomers containing A:T termini have revealed a 10-100 ns AT response that was assigned to fraying of the termini.[45, 62] Spectral changes associated with $\lambda_{\mu s}$ are similar to an equilibrium thermal difference spectrum (Fig. S9) between the initial (T_i) and final (T_f) temperatures, indicating the response corresponds to the duplex-to-single-strand transition.

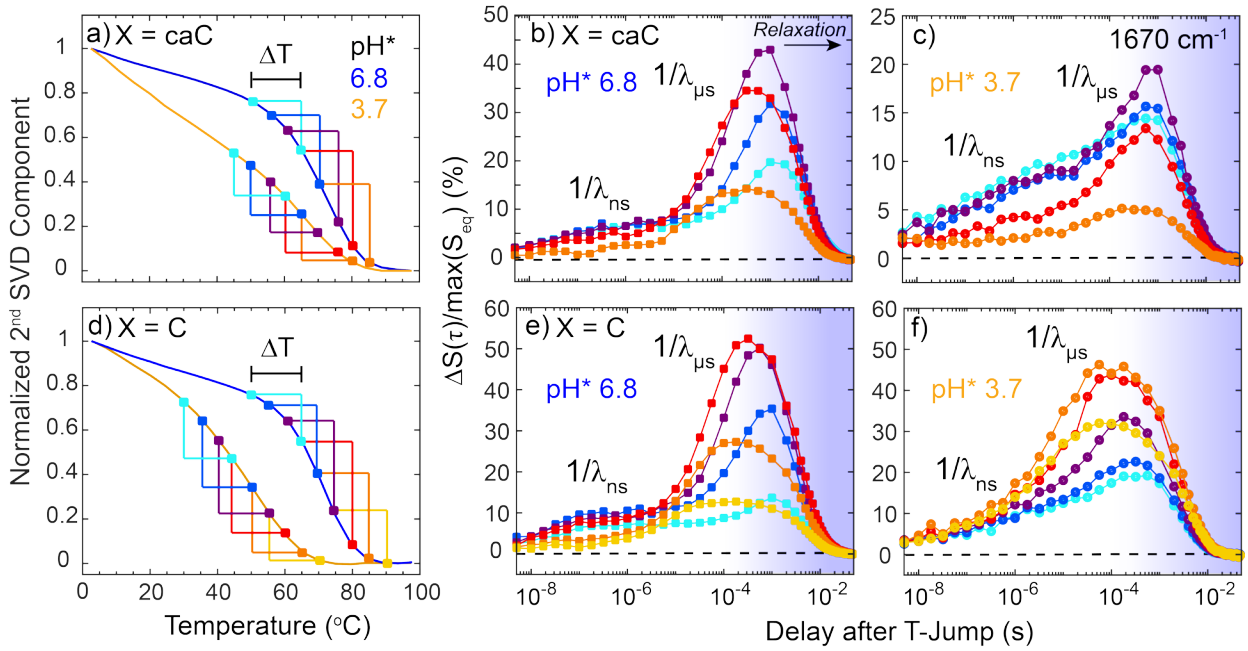


Figure 4: (a) Temperature ranges for T-jump experiments of X = caC along the 2nd SVD components from FTIR temperature series at pH* 6.8 (blue) and 3.7 (orange). Kinetic traces probed at 1670 cm^{-1} from the t-HDVE spectra (Figs. S10 & S11) for each T-jump range at (b) pH* 6.8 and (c) pH* 3.7. (d-f) Corresponding plots for X = C at pH* 6.8 and 3.7. Kinetic traces at intermediate pH* values are shown in Figs. S12 & S13.

For each sequence and pH* condition, T-jump experiments were performed with a fixed jump magnitude ($\Delta T \approx 15$ °C) and varying T_i along the duplex-to-single-strand transition region (Fig. 4a,d). Under neutral conditions, X = C and X = caC show similar temperature-dependent relaxation kinetics. In each case, $\lambda_{\mu s}$ increases exponentially with temperature, as observed in previous T-jump experiments of DNA dissociation.[1-2, 58, 62-65] Additionally, the magnitude of the $\lambda_{\mu s}$ response varies with the expected change in equilibrium base pairing (Fig. 4a,d) between T_i and T_f. In contrast to $\lambda_{\mu s}$, the temperature-dependence of λ_{ns} is negligible within our experimental resolution. The magnitude of the λ_{ns} response remains unchanged across the low T_i sampled, but

sharply reduces at high temperature. Since λ_{ns} primarily corresponds to fraying of the A:T termini as well as other changes in base pairing, the amplitude of its response is expected to decrease as T_i approaches a condition where the remaining duplexes in solution are largely frayed at equilibrium. More details relating the signal change of λ_{ns} to duplex thermodynamics are discussed in the SI.

As the solution pH* is reduced, both X = C and X = caC exhibit drastically different relaxation kinetics compared to those at neutral conditions. The $\lambda_{\mu s}$ response becomes increasingly non-exponential (Fig. S14) and faster. Additionally, the variation of $\lambda_{\mu s}$ and its associated signal amplitude with temperature are greatly reduced, signifying a reduction in the barrier to dissociation and cooperativity of the duplex-to-single-strand transition, respectively. For X = caC, the signal change of the duplex/single-strand response decreases at low pH*, consistent with the apparent change in the melting curve shown in Fig. 4d and adds further support that the highly sloped low-temperature baseline observed upon 5-carbxoyl protonation arises from changes in internal duplex base pairing. Overall, our results demonstrate N3 and 5-carboxyl protonation have distinct and significant impacts on DNA duplex stability and dehybridization kinetics.

Self-Consistently Modeling pH-Dependent DNA Hybridization Thermodynamics and Kinetics

To quantify the impact of protonation on the hybridization of our model sequences and to provide a consistent framework with which to discuss all of the experimental results, we propose a model that self-consistently describes the thermodynamics and kinetics of duplex formation. The melting of short oligonucleotides is typically assumed to occur in an all-or-nothing fashion, where DNA strands are fully base paired or separated.[60, 66] However, the apparent loss of A:T and G:C base pairing at low temperature and asymmetric melting curves exhibited by X = caC suggests a degree of pH-dependent pre-melting within the DNA duplex. To account for both the sigmoidal duplex-to-single-strand transition and the loss of base pairing at lower temperatures that gives rise to a sloping baseline, we interpret the melting curve as a total base pairing fraction, $\theta(T)$ that can be separated into external (θ_{ext}) and internal (θ_{int}) contributions (see SI for additional details):[67-68]

$$\theta(T) = \theta_{\text{int}}(T)\theta_{ext}(T) \quad (1)$$

Here, θ_{ext} refers to the fraction of intact duplexes containing at least one base pair and is related to the duplex/single-strand ($D \rightleftharpoons 2S$) equilibrium constant, K_d, that can be described by an external

enthalpy ($\Delta H°_{ext}$) and entropy ($\Delta S°_{ext}$). The average fraction of intact base pairs within the duplex is given by θ_{int}. For polymeric DNA, reductions in θ_{int} are the dominant factor in duplex dissociation and typically exhibit a more gradual and less cooperative dependence on temperature relative to θ_{ext}.[67] The low temperature changes in base pairing that manifest as asymmetry in the melting curve also appear to accumulate non-cooperatively with temperature, consistent with the expected profile of θ_{int}. For the purpose of modelling θ_{int}, we derive an expression for the average fraction of intact base pairs with respect to the equilibrium constant for forming or breaking a single base pair (see SI for details) that is described by an internal enthalpy ($\Delta H°_{\text{int}}$) and entropy ($\Delta S°_{int}$).However, it is an oversimplification to assign low-temperature loss of base pairing to the discrete loss of base pair contacts for such a short duplex. Instead, it is more likely that the steepening duplex baseline of X = caC upon 5-carboxyl protonation corresponds to loosening of caC:G base pairing along continuous structural coordinates or an increase in base pair structural fluctuations as suggested previously.[42]

The model can be extended to incorporate the T-jump results by assuming a two-state equilibrium of self-complementary oligomers, where the observed relaxation rate $\lambda_{\mu s}$ is related to the association (k_a) and dissociation (k_d) rates:[69]

$$\lambda_{\mu s} = k_d + 4\left[S\right]_{T_f} k_a \tag{2}$$

The single-strand concentration ([*S*]) at the final temperature (T_f) is obtained from the melting curve. In practice, we use a maximum entropy inverse-Laplace transform approach (MEM-iLT)[70] to extract λ_{ns} and $\lambda_{\mu s}$, the observed rates of each process (Fig. S15). To determine both the activation enthalpy ($\Delta H^{\ddagger}$) and entropy ($\Delta S^{\ddagger}$) of hybridization we fit k_a and k_d to a Kramers model in the high friction limit:[71]

$$k_d = \frac{C_d^o \lambda_{ns}}{\eta(T)} \exp\left(\frac{\Delta S_d^{\ddagger}}{R}\right) \exp\left(\frac{-\Delta H_d^{\ddagger}}{RT}\right) \tag{3}$$

$$k_a = \frac{C_a^o \lambda_{ns}}{\eta(T)} \exp\left(\frac{\Delta S_a^{\ddagger}}{R}\right) \exp\left(\frac{-\Delta H_a^{\ddagger}}{RT}\right) \tag{4}$$

where λ_{ns} is the observed rate of the 10-100 ns response measured in the t-HDVE kinetic traces, and $\eta(T)$ is the temperature-dependent viscosity of D_2O.[72] Here, λ_{ns} is taken as an estimate for the rate of diffusive hybridization in analogy to similar treatments in protein folding studies.[73-75] This stems from our assignment of λ_{ns} as largely corresponding to terminal A:T fraying dynamics, which have been shown to be diffusive in nature.[62] The temperature-dependence of λ_{ns} is negligible within the error of our measurement over the surveyed temperature range, therefore the mean over the lowest three T_i is used in the Kramers analysis. An additional unit parameter is included in the dissociation ($C°_d$) and association ($C°_a$) pre-exponential factor fixed at 1 Pa·s and 1 Pa·s·M^{-1}, respectively, which amounts to assuming that the pre-exponential factor is equivalent for association and dissociation, effectively placing any additional contributions into $\Delta S^{\ddagger}$.

For a two-state process on a 1D free energy surface, the standard free energy difference between the duplex and single-strand states ($\Delta G°_{ext}$) is equal to the difference between the dissociation ($\Delta G^{\ddagger}_d$) and association ($\Delta G^{\ddagger}_a$) activation free energies. Therefore, the duplex/single-strand equilibrium constant can be expressed in terms of enthalpic and entropic barriers:

$$K_d(T) = \exp\left(\frac{(\Delta S_d^{\ddagger} - \Delta S_a^{\ddagger})}{R}\right)\exp\left(\frac{(\Delta H_a^{\ddagger} - \Delta H_d^{\ddagger})}{RT}\right) \quad (5)$$

where the numerators in the arguments of the exponentials are equal to $\Delta S°_{ext}$ and $\Delta H°_{ext}$, respectively. Using eqs. 1-5, the thermodynamic and kinetic data can be globally fit to self-consistently describe the pH*-dependent hybridization thermodynamics and kinetics of X = C and X = caC. In total, the thermodynamic and kinetic data is described by six parameters: $\Delta H^{\ddagger}_d$, $\Delta H^{\ddagger}_a$, $\Delta S^{\ddagger}_d$, $\Delta S^{\ddagger}_a$, ΔH_{int}, ΔS_{int}. An additional parameter (A) that scales the melting 2nd SVD components by a value between 0 and 1 at the lowest temperature measured is also needed (See SI). In an all-or-nothing melting model, thermal melting curves are normalized to 1 along the duplex baseline, reflecting that all duplexes are intact at the lowest temperature. However, sequences with substantial changes in internal base pairing prior to duplex dissociation may not be fully intact at the lowest temperature measured, and the normalization amplitude (A) used in fitting the thermal melting curves is proposed to report on the degree of internal base pairing at the lowest temperature sampled (3 °C).

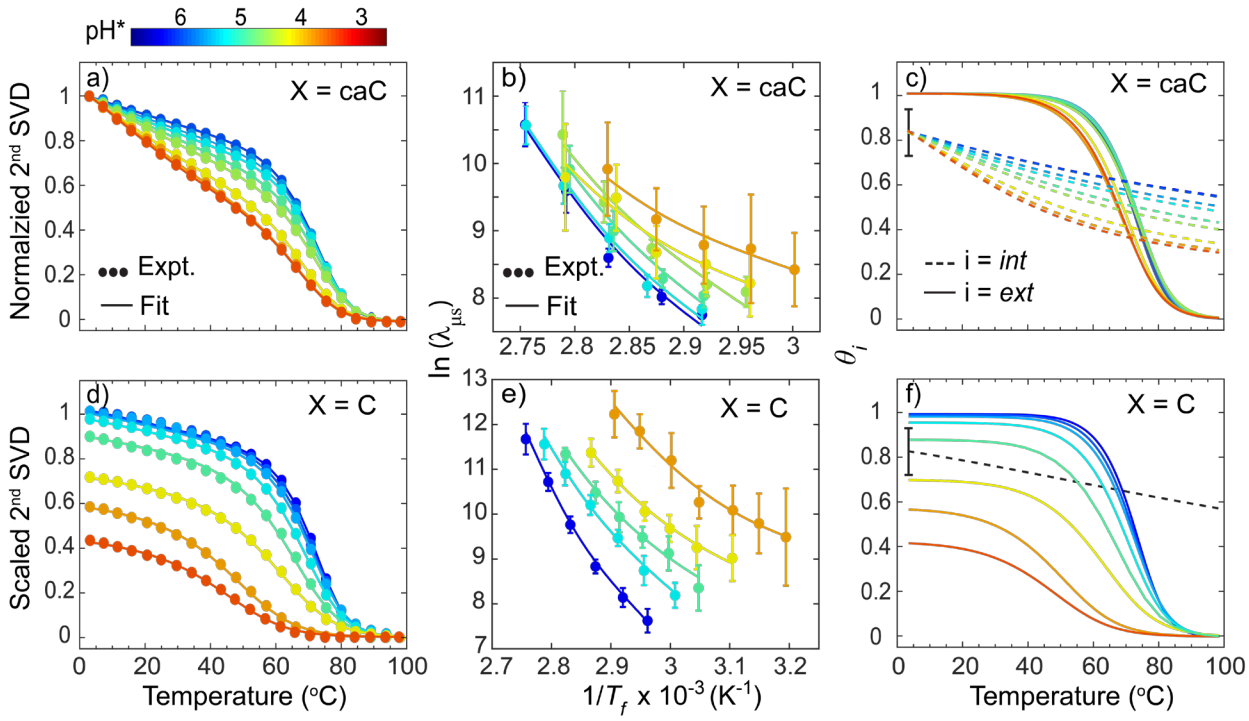


Figure 5: (a,b) Fits (solid lines) to the 2nd SVD components from FTIR temperature series and observed relaxation rates ($\lambda_{\mu s}$) from T-jump measurements of X = caC using eqs. 1-5. 2nd SVD components at pH* 6.0, 5.5, and 3.5 were not globally fit with $\lambda_{\mu s}$ because T-jump measurements were not performed at those pH* conditions. Error bars in (b) represent the amplitude-weighted standard deviation in $\lambda_{\mu s}$ across probe frequency. (c) Extracted profiles of θ_{ext} (solid line) and θ_{int} (dashed line) from pH* 6.8 to 3.5. θ_{int} is normalized to 0.825, which is the average value of $1/A$ across pH* for X = caC (Fig. S16). The error bar in the value of θ_{int} at 3 °C is the mean error in $1/A$ across pH*. (d-f) Corresponding plots for X = C.

Fig. 5 shows fits using eqs. 1-5 to the 2nd SVD component melting curves and $\lambda_{\mu s}$ across temperature for X = caC and X = C. T-jump experiments were performed at select pH* points among the measured equilibrium melting experiments. Therefore, melting 2nd SVD components acquired at pH* 6.0, 5.5, and 3.5 were fit without $\lambda_{\mu s}$ to eq. 1 using $\Delta H°_{ext}$, $\Delta S°_{ext}$, $\Delta H°_{int}$, and $\Delta S°_{int}$. As shown in Fig. S16 & S17, consistent thermodynamic values are determined when fitting the 2nd SVD components with or without kinetic data. The profiles of θ_{int} and θ_{ext} determined from the fits for X = caC are shown in Fig. 5c. Upon a reduction in pH*, the profile of θ_{ext} exhibits a small shift toward lower temperature and broadens slightly while θ_{int} becomes more sharply decreasing along temperature. The value of A for X = caC is assumed to account for deviations in

θ_{int} from 1 at 3 °C and was found to be insensitive to pH* within the accuracy of our measurement and model (Fig. S16). Therefore, the average value of $1/A$ across pH* (0.825) was applied to each 2nd SVD component and the average fit error across pH* is used to estimate the error in A.

A comparison of the FTIR temperature series from pH* 6.8 to 3.5 for X = C (Fig. 2f) shows substantial broadening and shifting of the melting transition as pH* is reduced, indicating a loss of melting cooperativity and overall destabilization of the duplex, respectively. This provides further evidence that the low pH* samples do not begin from a duplex fraction of unity at 3 °C. The thermal 2nd SVD components measured for the X = C sequence can be modeled using eq. 1, where both θ_{int} and θ_{ext} contribute to the melting profile. However, our FTIR titration analysis suggests that the X = C sequence cannot tolerate N3 protonation in the duplex state, and it is a reasonable approximation that θ_{int} is independent of pH*. Therefore, we fit the pH* 6.8 2nd SVD component to eq. 1, determining θ_{int}, θ_{ext}, and a 3 °C normalization offset of 0.825. Under neutral conditions, the slope of θ_{int} is expected to arise from fraying of the A:T termini (Fig. S8). The normalized 2nd SVD components at all pH* points were scaled at 3 °C (Fig. 5d) by the value of θ_{ext} determined from the FTIR titration of X = C (Fig. 3a) and the form of θ_{int} at pH* 6.8 (dashed line in Fig. 5f) was assumed constant across all values of pH*. As shown in Fig. 5, both the thermal 2nd SVD components and temperature trend in $\lambda_{\mu s}$ are well fit within this treatment.

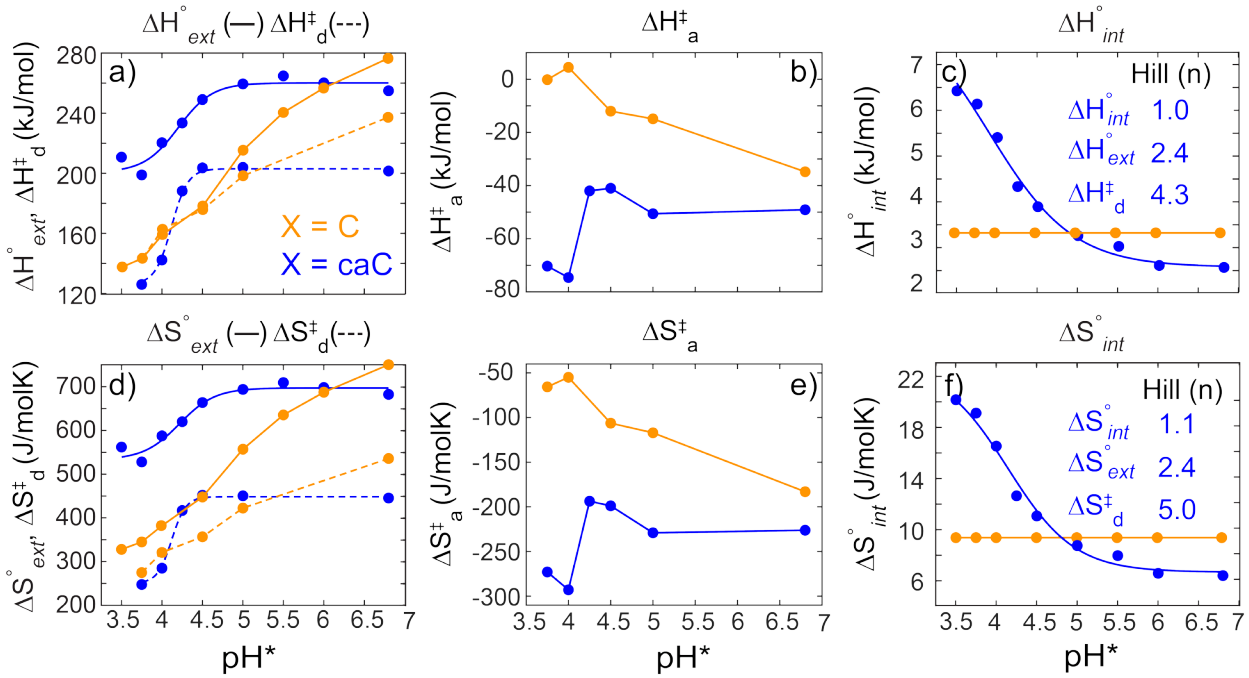


Figure 6: Thermodynamic and kinetic parameters determined from fits to melting 2nd SVD components and temperature trends in $\lambda_{\mu s}$. (a) External enthalpy ($\Delta H°_{ext}$, solid lines) and enthalpic dissociation barrier ($\Delta H^{\ddagger}_{d}$, dashed lines) for X = C (orange) and X = caC (blue). Parameters for X = caC are fit to the Hill equation. Corresponding plots for (b) enthalpic association barrier ($\Delta H^{\ddagger}_{a}$), (c) internal enthalpy ($\Delta H°_{int}$), (d) external entropy ($\Delta S°_{ext}$, solid lines) and entropic dissociation barrier ($\Delta S^{\ddagger}_{d}$, dashed lines), (e) entropic association barrier ($\Delta S^{\ddagger}_{a}$), and (f) internal entropy ($\Delta S°_{int}$). Hill coefficients (n) from fits to X = caC parameters are shown in (c) and (f).

Fig. 6 presents the trends in thermodynamic and kinetic parameters determined from the fits to thermal 2nd SVD components and $\lambda_{\mu s}$ across pH* for X = caC and X = C. Over the studied pH* range, X = caC exhibits reductions in $\Delta H°_{ext}$ and $\Delta S°_{ext}$ of ca. 40 kJ/mol and 150 J/molK, respectively. X = C shows substantially greater reductions in $\Delta H°_{ext}$ and $\Delta S°_{ext}$ of ca. 140 kJ/mol and 400 kJ/mol, further demonstrating the large disruption of duplex formation upon N3 protonation shown here and by others.[7, 12-13] 5-carboxyl protonation also leads to a reduction in θ_{int} as well as increase in $\Delta H°_{int}$ and $\Delta S°_{int}$. The dissociation barriers $\Delta H^{\ddagger}_{d}$ and $\Delta S^{\ddagger}_{d}$ both decrease upon 5-carboxyl and N3 protonation while opposite trends are observed in $\Delta H^{\ddagger}_{a}$ and $\Delta S^{\ddagger}_{a}$ between the two types of protonation. At pH* 6.8, each sequence exhibits negative association barriers as observed for DNA hybridization in many previous works.[1, 62-64, 76] However, $\Delta H^{\ddagger}_{a}$ and $\Delta S^{\ddagger}_{a}$ each become more negative upon 5-carboxyl protonation and less negative upon N3 protonation. Overall, the activation parameters resemble those in activation energy (E) and pre-exponential factor (A) determined from an Arrhenius analysis of the kinetic data (Fig. S19-S21). Further discussion of the hybridization thermodynamic and kinetic parameter trends with pH* follows.

Examining the trends in thermodynamic and kinetic parameters in Fig. 6 across pH* reveals distinct behavior due to 5-carboxyl and N3 protonation. For X = caC, each parameter remains almost unchanged between pH* 6.8 and 5.0 and then changes sharply across lower pH* values with titration-like behavior. In contrast, X = C exhibits large, but more gradual changes across the pH* window studied. The trends for X = caC can each be well fit by the Hill equation[77] with consistent apparent pK_a values of ca. 4.5 and variable Hill coefficients (n). For fitting of binding curves, n > 1 indicates positive cooperativity between binding sites that leads to sharpening of the binding transition. The trends in external thermodynamics as well as enthalpic and entropic barriers reveal sharp transitions that fit to values of n > 1.

Discussion

Impact of N3 and 5-Carboxyl Protonation on the Thermodynamic Stability of DNA

Similar to previous studies,[6-7, 12-13] the X = C duplex is greatly destabilized as a result of N3 protonation, and our analysis suggests that this specific sequence essentially cannot tolerate any protonation in the duplex state. However, the degree of duplex destabilization from N3 protonation and maximum number of protonated C bases is expected to vary with sequence composition. Previous work with polymeric DNA and short RNA sequences has shown that duplex stability is dependent on the GC content in the sequence.[7, 12-13, 18] UV CD measurements suggested that up to 50% of C bases in polymeric DNA could be protonated prior to signs of complete duplex dissociation at low temperature.[6] Instead, C protonation leads to disruption of many GC base pairs prior to full dissociation of the duplex.[18] It is clear that N3 protonation generally destabilizes the duplex state, but the degree and nature of the destabilization will depend on the context of the GC base pairs. Therefore, while X = C completely dissociates upon a N3 protonation event, other sequences may be able to tolerate some degree of protonation depending on the number and location of GC base pairs present. Regardless, it is well-established that the thermodynamics of nucleic acid hybridization in both polymers and oligonucleotides is additive and can be decomposed into the nearest-neighbor contributions of discrete dinucleotide steps.[78-79] It is thus reasonable to expect the present results to apply generally for these base pairs and their local sequence context whether in short duplexes or longer polymeric DNA.

In contrast to the X = C sequence, protonation of the exocyclic carboxyl group of caC is much less perturbative to the duplex/single-strand equilibrium, shifting the melting inflection point down by ~6 °C in going from pH* 6.8 to 3.5. As a result, the X = caC duplex can tolerate near complete protonation of caC sites (Fig. 3), and this leads to measurable changes in internal base pairing in the duplex. As caC is protonated, clear signatures of the loss of G:C base pairing appear along the low-temperature baseline of X = caC (Fig. S8). For short oligonucleotides like those studied here, it is unlikely that the low-temperature G:C changes correspond to complete loss of discrete base pair contacts. Instead, these signatures suggest loosening of hydrogen bonding or base stacking along continuous structural coordinates. Recent MD simulations of caC-containing duplexes demonstrated that 5-carboxyl protonation increases the degree of structural fluctuations within the modified X:G base pair.[42] Motions along these structural coordinates alter the

intramolecular base pair distance and orientation, which are expected to alter guanine and C vibrational modes. Therefore such structural fluctuations may account for the observed melting curve baselines.[47] The physical origin for this base pair loosening may stem from weakened caC:G base pairing due to the electron-withdrawing nature of the protonated carboxylic acid group, consistent with similar base pair loosening observed for DNA containing 5-formylcytosine.[68]

Assignment of the two pK_a values observed for caC (2.1-2.4 and 4.2-4.8) has been debated in recent reports.[23, 80-81] Using FTIR titrations and simulated IR spectra of 2'-deoxy-5-carboxylcytosine, we previously concluded that the pK_a values at 4.2-4.8 and 2.1-2.4 correspond to protonation of the 5-carboxyl and N3 positions, respectively.[23] The pH-dependent behavior of X = C and X = caC observed in the current study further support these assignments. Overall, protonation of X = C is shown disrupt the DNA duplex to a much greater degree than protonation of X = caC from pH* 6.8 – 3.5. If X = caC was protonated at the N3 position in the studied pH* range, similar pH-dependent behavior to X = C would be expected. Therefore, we confidently assign the pK_a value at 4.2-4.8 in the caC nucleobase and within duplex DNA to protonation of the 5-carboxyl position.

N3 and 5-Carboxyl Protonation Distinctly alter DNA Hybridization Kinetics

To date, the impact of nucleobase protonation on dissociation and hybridization has remained elusive from a dynamical perspective, despite the importance of these processes in biology and nanotechnology applications. Not only do our results add insight to how N3 protonation impacts the stability of short oligonucleotides as well as the tunability of base pairing stabilities through caC protonation, but we have also characterized the impact on the kinetics of hybridization.

Just as N3 protonation greatly destabilizes the DNA duplex, it also increases the rate of duplex dissociation. For example, k_d increases by nearly a factor of 500 at T_i 55 °C between pH* 6.8 and 3.7 (Fig. S19). 5-carboxyl protonation leads to a far more modest increase in the dissociation rate at T_i 55 °C by a factor of 7-10 (Fig. S20) between pH* 6.8 and 3.7. Global fitting of the kinetics and thermodynamics indicate that both N3 and 5-carboxyl protonated oligonucleotides are still well described by a two-state duplex/single-strand model. Over the pH* range studied, each type of protonation results in a similar reduction of 80-100 kJ/mol in $\Delta H^{\ddagger}_d$ and 200-250 J/mol·K in $\Delta S^{\ddagger}_d$. The trend upon N3 protonation suggests that $\Delta H^{\ddagger}_d$ and $\Delta S^{\ddagger}_d$ would

continue to decrease at pH* <3.5. However, due to significant protonation of adenine and eventually guanine at pH* <3.5, we have restricted our study to the pH* window between 6.8 and 3.7. In contrast to the dissociation barriers, both $\Delta H^{\ddagger}_{a}$ and $\Delta S^{\ddagger}_{a}$ become less negative for X = C and more negative for X = caC as pH* decreases.

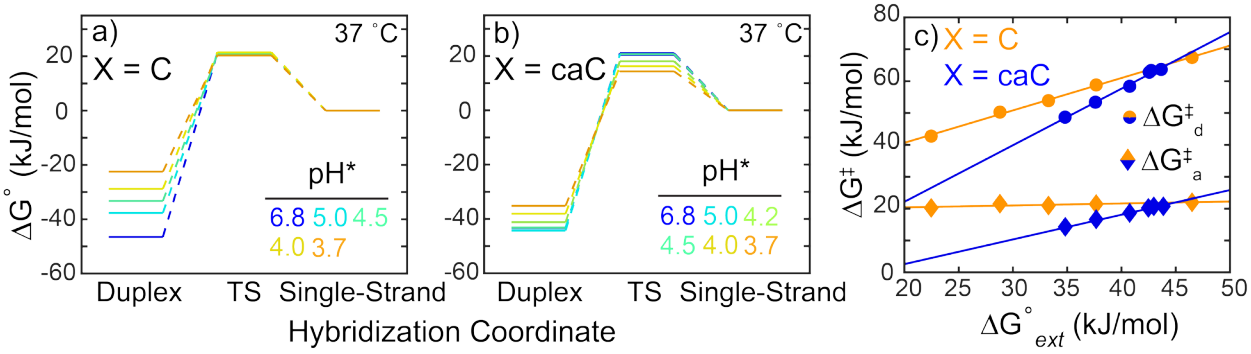


Figure 7: Representative 1D free energy plots for (a) X = C and (b) X = caC at 37 °C constructed using the association ($\Delta G^{\ddagger}_{a}$) and dissociation ($\Delta G^{\ddagger}_{d}$) free energy barriers determined from global fitting between T-jump and FTIR temperature series results. (c) Relationship between $\Delta G^{\ddagger}_{d}$ (circles) and $\Delta G^{\ddagger}_{a}$ (diamonds) and $\Delta G°_{ext}$ at 37 °C for X = C (orange) and X = caC (blue). Solid lines indicate linear fits.

For both dissociation and association, the enthalpy and entropy are highly pH*-dependent, and therefore simplistic free energy diagrams may be informative for understanding the impact from each type of protonation. Fig. 7 presents free energy plots along a 1D hybridization coordinate for X = C and X = caC at 37 °C. These three-level surfaces are composed of the duplex/single-strand free energy change (left), hybridization transition state (TS, center), and single-strand (right). The single-strand free energy is assumed to be independent of pH* and is set as the reference state. At 37 °C, X = C and X = caC exhibit reductions in $\Delta G^{\ddagger}_{d}$ of 25 and 20 kJ/mol, respectively, in going from pH* 6.8 to 3.7. $\Delta G^{\ddagger}_{a}$ is essentially independent of N3 protonation at this temperature. Across the studied pH* range, linear relationships are observed between $\Delta G^{\ddagger}_{d}/\Delta G^{\ddagger}_{a}$ and $\Delta G°_{ext}$ for each sequence (Fig 7c). For X = C, a slope of 1.04 is observed when plotting $\Delta G^{\ddagger}_{d}$ against $\Delta G°_{ext}$ and a slope of 0.06 is observed when plotting $\Delta G^{\ddagger}_{a}$ against $\Delta G°_{ext}$, indicating that reduction in $\Delta G^{\ddagger}_{d}$ at low pH* is due to destabilization of the duplex relative to the single-strand rather than a change in the transition state energy. In contrast, X = caC exhibits a

slope of 1.77 for $\Delta G^{\ddagger}_{d}$ vs. $\Delta G°_{ext}$ and 0.77 for $\Delta G^{\ddagger}_{a}$ vs. $\Delta G°_{ext}$, indicating that both destabilization of the duplex relative to the single-strand and a reduction in transition state free energy occur upon 5-carboxyl protonation.

Proposed Mechanistic Impact of N3 and 5-Carboxyl Protonation on DNA Hybridization

One can gain further insight into the mechanistic implications of the thermodynamic and kinetic results if we assume that the hybridization transition state involves the formation of some small subset of native contacts followed by rapid downhill formation of the remaining base pairs, in analogy to the classic nucleation-zipper mechanism of helix-to-coil transitions in biomolecules.[2-3] At neutral pH*, both sequences exhibit negative values of $\Delta H^{\ddagger}_{a}$ (Fig. 6b). This result has been observed routinely in DNA hybridization,[1-2, 4, 58, 62-64, 76] and is assumed to account for base pair formation in the transition state. A large entropic penalty due to loss of translational and configuration freedom accompanies this initial base pair formation, causing the rate of association to decrease as temperature increases.

Interestingly, $\Delta H^{\ddagger}_{a}$ approaches 0 kJ/mol and $\Delta S^{\ddagger}_{a}$ becomes less negative with increasing N3 protonation of the X = C oligonucleotide. Multiple possibilities could account for the reduction in magnitude of $\Delta H^{\ddagger}_{a}$. The X = C single-strand is expected to become highly protonated as pH* decreases, as evidenced by the pH-driven denaturation observed in Fig. 3. For GC-rich sequences such as the one studied here, hybridization is assumed to initiate at C:G base pairs, but protonation at N3 interferes with the ability to form C:G contacts. Therefore, the number of contacts in the transition state may remain unchanged but offer less enthalpic stabilization due to the competition between base pair hydrogen bonding and protonation at N3. In addition, the reduction in magnitude of $\Delta S^{\ddagger}_{a}$ may result from changes in solvation and sodium cation environment around the protonated single-strand as noted previously.[82] As more N3 sites are protonated, the negative charge density of the single-strand is reduced, and the number of bound sodium cations and excluded chloride and phosphate anions decrease. Analysis of the FTIR titration (Fig. 3) on X = C suggests that N3 protonation can only occur in the single-strand. As a result, the charge density and counterion environment around the protonated single-strand and deprotonated duplex are quite different, and hybridization of the former must involve the loss of N3 protons, gain of associated sodium cations, and exclusion of additional anions. These changes likely contribute to the thermodynamic and kinetic effects observed for N3 protonation in the X = C sequence.

The decrease in the magnitude of $\Delta H^{\ddagger}_{a}$ and $\Delta S^{\ddagger}_{a}$ at low pH* could also reflect a reduction in the number of base pairs formed in the hybridization transition state. $\Delta H^{\ddagger}_{a}$ ranges from -35 to 0 kJ/mol over the pH* range studied here. The average enthalpic benefit associated with forming a base pair in the CG region of X = C is estimated to be -21 kJ/mol based on the unified nearest-neighbor parameters for DNA hybridization.[78] Therefore, the change in $\Delta H^{\ddagger}_{a}$ suggests that the number of base pair contacts in the transition state shifts from ~2 to 0 upon protonation. For the X = C sequence, the reduction of $\Delta H^{\ddagger}_{a}$ to 0 kJ/mol would suggest that no base pairs are formed in the transition-state at low pH*. Even in this case, k_a is 100 times slower (~10^7 $M^{-1}s^{-1}$) than the estimated diffusion-limited association rate constant determined from fluorescence correlation spectroscopy (FCS) measurements of short oligonucleotides,[76] indicating that an entropic barrier to duplex formation dictates DNA hybridization kinetics even without the formation of base pair contacts in the transition state.

While protonation of the X = caC sequence also increases the rate of duplex association, $\Delta H^{\ddagger}_{a}$ and $\Delta S^{\ddagger}_{a}$ are both observed to increase in magnitude as pH* is lowered. Again, this could be interpreted as changes to the energetics of initial contact formation or reflect a change in the structure of the hybridization transition state. 5-carboxyl protonation has the potential to impact intra- and inter-strand electrostatics, caC:G base pairing, base stacking, and major groove solvation. Our pH*-dependent FTIR results indicate that caC sites within X = caC are essentially completely deprotonated at pH* 6.8, adding substantial negative charge to the duplex and single-strand states. Protonation of caC should reduce electrostatic repulsion between the exocyclic carboxyl group and the phosphate backbone. However, the absence of negative charge may alter the degree of hydration in the major groove. Our model for internal base pairing suggests that 5-carboxyl protonation increases the enthalpic stabilization of base pair formation (Fig. 6c) even though contacts become loosened overall. Such an increase in $\Delta H°_{int}$ may explain the increase in magnitude of $\Delta H^{\ddagger}_{a}$ as pH* decreases, but not the increase in entropic penalty. The change in $\Delta S^{\ddagger}_{a}$ is more complex as 5-carboxyl protonation can alter factors such as base hydration and single-strand flexibility.[83] Both the increase in magnitude of $\Delta H^{\ddagger}_{a}$ and $\Delta S^{\ddagger}_{a}$ and decrease in $\Delta H^{\ddagger}_{d}$ and $\Delta S^{\ddagger}_{d}$ could also be explained by an increase in the number of base pair contacts associated with the hybridization transition state, which could in turn result from the increased structural fluctuation of caC:G base pairs upon 5-carboxyl protonation.[42]

Our results demonstrate that N3 and 5-carboxyl protonation significantly perturb DNA hybridization and dissociation kinetics, but in a nucleobase-specific manner that can be rationalized in terms of duplex destabilization and changes to the transition state for a two-state process. However, it is important to note that while the nature of the transition state may change, so too may the distribution of transition states. As shown in Fig. 4 & S14, both sequences exhibit non-exponential relaxation kinetics at low pH* that may stem from inhomogeneity among the duplex ensemble or the transition state. This spread of rates may be modulated by base pairing fluctuations, solvent and ion interactions, as well as distributions of protonated species. A more detailed discussion regarding the origin of the non-exponential kinetics is presented in the SI.

Protonated caC Sites Cooperatively Impact Duplex-to-Single-Strand Thermodynamics and Kinetics

5-carboxyl protonation reduces the overall magnitude of the thermodynamic parameters for duplex dissociation ($\Delta H°_{ext}$, $\Delta S°_{ext}$, $\Delta G°_{ext}$) in an apparently cooperative manner. As shown in Fig. 6, trends in $\Delta H°_{ext}$, $\Delta S°_{ext}$, and $\Delta G°_{ext}$ for X = caC follow a Hill profile with Hill coefficient >1, signifying positive cooperativity. Typically, Hill equations are applied to identify cooperativity in binding problems,[84] but here the protonation of X = caC is noncooperative (n = 1, Fig. 3). Instead, 5-carboxyl protonation seems to cooperatively shift the thermodynamics of the duplex-to-single-strand transition. Our analysis indicates that nearly all caC sites in the duplex state become protonated from pH* 6.8 to 3.5, and therefore it is possible that multiple protonated caC:G base pairs interact. It seems most likely that such cooperativity would arise from adjacent protonated base pairs, where increased structural fluctuations of each may amplify the weakening of hydrogen bonding and base stacking interactions. Similar effects were predicted for C methylation in which simulations of adjacent methylated CpG dinucleotides influenced base pair fluctuations in a more than additive manner.[85]

Similar to the shifts in duplex/single-strand equilibrium, 5-carboxyl protonation appears to alter DNA hybridization kinetics in a cooperative manner as both dissociation and association barriers show sharp transitions as a function of pH* and exhibit a Hill coefficient greater than 1 (Fig. 6). Additionally, the pH* trends in dissociation barrier show a sharper transition (larger Hill coefficient) than observed for the external thermodynamic terms. As shown in the 1D free energy surfaces (Fig. 7b), the pH*-dependence of $\Delta G^{\ddagger}_{d}$ at 37 °C has contributions from duplex

destabilization and a reduction in free energy of the transition state as pH* is lowered. The former corresponds to $\Delta G°_{ext}$ while the latter is only captured by $\Delta G^{‡}_{d}$ and $\Delta G^{‡}_{a}$. Therefore, the more cooperative manner in which $\Delta G^{‡}_{d}$ varies as a function of pH* relative to $\Delta G°_{ext}$ suggests that protonation of multiple 5-carboxyl sites leads to a cooperative change in the free energy of the hybridization transition state in addition to the duplex state. A comparison of $\Delta H^{‡}_{d}$ with $\Delta H°_{ext}$ and $\Delta S^{‡}_{d}$ with $\Delta S°_{ext}$ reveal the same trends in apparent cooperativity with 5-carboxyl protonation.

N3 protonation leads to a larger reduction in DNA duplex stability and dissociation barrier than 5-carboxyl protonation, but it does so almost linearly as a function of pH* (Fig. 6a) over the range studied. This more gradual reduction reflects a non- or anti-cooperative impact of N3 protonation on the duplex-to-single-strand transition and is consistent with our FTIR titration results that estimate the X = C duplex cannot tolerate N3 protonation, making cooperative interactions between multiple protonated sites in the duplex very unlikely.

Conclusions

We have investigated the impact of cytosine and 5-carboxylcytosine protonation on the thermodynamics and kinetics of DNA dissociation and hybridization using FTIR and optically-induced temperature-jump IR spectroscopy. Our results demonstrate that between pH* 6.8 and 3.5, DNA containing C and caC are predominantly protonated at the N3 and 5-carboxyl positions, respectively. Each protonation is shown to have a distinct impact on the thermodynamics and kinetics of dissociation and association between model DNA oligonucleotides. Protonation at N3 completely disrupts the ability of the DNA to duplex, shifts the melting inflection point by >20 °C, and reduces the cooperativity of the duplex-to-single-strand transition. These thermodynamic properties are accompanied by a speedup in the rate of duplex dissociation and increasingly stretched exponential kinetics at low pH*, with the reduction in dissociation barrier primarily achieved through destabilization of the duplex state. 5-carboxyl protonation leads to highly sloped melting baselines that reflect an accumulation of disrupted base pairing contacts in the duplex state, but perturbations to the duplex/single-strand equilibrium are comparatively minor. Regardless of the identity of X, the dissociation free energy barrier is reduced to a similar degree over the pH* range studied. However, N3 protonation does reduces the barrier through destabilization of the DNA duplex while 5-carboxyl protonation reduces the dehybridization transition state energy. Unlike with N3 protonation, X = caC duplexes can be highly protonated,

and protonation of multiple sites is observed to alter duplex/single-strand thermodynamics and kinetics in a cooperative manner.

Protonation of N3 in cytosine plays critical roles in DNA damage, triplex association, and i-motif formation and is utilized to create pH-driven nanodevices. Additionally, the potential importance of the protonation equilibrium of 5-carboxylcytosine in cytosine demethylation has recently come to light. Each of these functions and applications rely on protonation-driven perturbations to double helical DNA, and this work demonstrates the unique ability of N3 and 5-carboxyl protonation to alter the stability, hybridization kinetics, and base pairing dynamics of nucleic acids. The loosening of caC:G base pairing and reduction in dissociation barrier initiated by caC protonation may assist in its selective recognition within the cytosine demethylation pathway.[27-28, 30, 86] Such pH-dependent base pair loosening as well as the overall reduction of p$K_{a,N3}$ upon substitution of caC for C may also prove useful in tuning the physical properties of DNA triplex and i-motif formation. Recent work has shown that caC can effectively fine tune the thermodynamic stability of both structures,[38-39] and the results and analysis in the present work provide physical insight into the observed thermodynamic response of the DNA triplex and i-motif to incorporation of caC as well as the prediction of pH-dependent dynamical behavior upon protonation of the 5-carboxyl group.

Materials and Methods

Synthesis and Purification of 5′-TA(caC)G(caC)G(caC)GTA-3′

Unmodified and 5caC phosphoramidites were purchased from Glen Research and DNA oligomers were synthesized at 1 μmol scale in several batches. After oligomer synthesis, the beads were treated with 0.1 M K_2CO_3 in 1:1 MeOH/water at 42 °C for 16 hours, and then acetic acid was added to neutralize the pH to 7.0. Oligomers were subsequently purified with dialysis in ultrapure water at 4 °C for 48 hours and lyophilized to a powder form.

Oligonucleotide Sample Preparation

The sequence 5′-TACGCGCGTA-3′ was purchased from Integrated DNA Technologies (IDT) at desalt grade purity. Samples were purified further with 3 kD cutoff centrifugal filters (Amicon). For IR spectroscopic measurements, all labile protons were exchanged in deuterium oxide (D_2O, Cambridge Isotopes, 99.9%). Samples were prepared at a concentration of 1 mM in

deuterated 20 mM sodium phosphate buffer at pH* 6.8 with 16 mM NaCl. Units of pH* indicate the measured pH of a deuterated solution using a standard glass electrode pH meter. The pK_a values reported here were first converted to represent the value that would be determined in a non-deuterated solution.[51] Solutions were prepared at reported pH* values through the addition of concentrated DCl. To minimize HOD content, samples were lyophilized after addition of DCl and re-dissolved in the appropriate volume of pure D_2O.

Equilibrium FTIR and 2D IR Measurements

FTIR spectra were measured with a Bruker Tensor FTIR spectrometer at 1 cm^{-1} resolution. Samples were placed between two 1 mm thick CaF_2 windows separated by a 50 μm Teflon spacer enclosed within a home-built brass jacket. The jacket temperature is controlled with a recirculating chiller (Ministat 125, Huber). The sample temperature for a given chiller bath temperature was determined using a thermocouple attached to the center of the CaF_2 window. The FTIR titration of X = C was performed using a home-built flow cell with a 50 μm path length and 1 mm CaF_2 windows. Samples prepared at each pH* condition were flowed into the sample cell using a syringe pump (Harvard Apparatus).

Two-dimensional (2D IR) measurements were performed on a previously described setup with a BOXCAR geometry.[87-88] 2D IR spectra were collected with ZZZZ polarization and at a fixed waiting time (τ_2) of 150 fs. The coherence time (τ_1) was scanned from -60 to 2500 fs and -60 to 2000 fs for rephasing and non-rephasing surfaces, respectively, with a 4 fs step size.

Transient T-jump IR Spectroscopy

The details of our temperature-jump (T-jump) spectrometer have previously been described in detail.[87-88] In brief, the output of a frequency-doubled Nd:YAG (YG 980, Quantel) was sent through an optical parametric oscillator (OPO) to generate 2 μm pulses (5 ns, 20 mJ, 20 Hz) used to pump the O-D stretch overtone of D_2O. Time-dependent changes to the DNA structure are monitored between 5 ns and 50 ms with nonlinear infrared spectroscopy. Transient heterodyne detected vibrational echo (t-HDVE) infrared spectra were acquired in ZZZZ polarization and the local oscillator was stepped in 5 fs intervals between -10 and 10 fs relative to maximum interference with the DVE signal. t-HDVE spectra were processed with Fourier Transform Spectral Interferometry,[61, 89] and the recovered dispersed pump-probe (t-DPP) is used throughout

the study. Transient 2D IR (t-2DIR) measurements were acquired with undersampling along τ_1 at a fixed τ_2 of 150 fs. A τ_1 step size of 16 fs was used. Rephasing and non-rephasing FIDs were scanned from -60 to 1750 fs and -60 to 1250 fs, respectively. Initial temperatures (T_i) were set using a recirculating chiller connected to a brass sample jacket as for FTIR measurements. The temperature-jump magnitude (ΔT) was determined through monitoring the change in mid-IR D_2O transmission and was set by adjusting the voltages applied to the Nd:YAG flashlamps or by using a polarizer to attenuate the 2 μm output. In this study, ΔT was set to ~15 °C for each measurement.

Associated Content

Supporting Information

Detailed description and interpretation of pH-dependent FTIR spectra; Assignment of observed T-jump timescales using T-jump 2D IR spectroscopy (t-2DIR); T-jump data of intermediate pH* samples for X = C and X = caC; Description of T-jump rate extraction using maximum entropy inverse Laplace transform (MEM-iLT) approach; Description and interpretation of stretched-exponential T-jump relaxation kinetics; Correlation between θ_{int} and λ_{ns}; Arrhenius analysis of T-jump kinetic data. This information is available free of charge via the Internet at http://pubs.acs.org.

Author Information

Corresponding Author

*tokmakoff@uchicago.edu

Notes

The authors declare no competing financial interest.

Acknowledgements

B.A., P.J.S., and A.T. acknowledge the National Science Foundation (Grant No. CHE-1856684) and the National Institute of General Medical Sciences of the National Institutes of Health (Award No. R01GM118774) for support of this research. B.A. acknowledges support from the NSF GRFP. Q.D. was supported by the National Institutes of Health (Grant No. 5K01HG006699). C.H. acknowledges the National Institutes of Health (Award No. R01HG006827). The authors

also thank Ram Itani for design and construction of the flowcell used for FTIR titration measurements.

References

1. Chen, C.; Wang, W.; Wang, Z.; Wei, F.; Zhao, X. S., Influence of secondary structure on kinetics and reaction mechanism of DNA hybridization. *Nucleic Acids Res.* **2007,** *35*, 2875-2884.
2. Craig, M. E.; Crothers, D. M.; Doty, P., Relaxation kinetics of dimer formation by self complementary oligonucleotides. *J. Mol. Biol.* **1971,** *62*, 383-401.
3. Pörschke, D.; Uhlenbeck, O.; Martin, F., Thermodynamics and kinetics of the helix-coil transition of oligomers containing GC base pairs. *Biopolymers* **1973,** *12*, 1313-1335.
4. Wetmur, J. G.; Davidson, N., Kinetics of renaturation of DNA. *J. Mol. Biol.* **1968,** *31*, 349-370.
5. Anderson, C. F.; Record Jr, M. T., Salt-nucleic acid interactions. *Annu. Rev. Phys. Chem.* **1995,** *46*, 657-700.
6. Zimmer, C.; Luck, G.; Venner, H.; Frič, J., Studies on the conformation of protonated DNA. *Biopolymers* **1968,** *6*, 563-574.
7. Zimmer, C.; Venner, H., Protonation of cytosine in DNA. *Biopolymers* **1966,** *4*, 1073-1079.
8. Gulland, J. M.; Jordan, D.; Taylor, H., 213. Deoxypentose nucleic acids. Part II. Electrometric titration of the acidic and the basic groups of the deoxypentose nucleic acid of calf thymus. *J. Chem. Soc.* **1947**, 1131-1141.
9. Mergny, J.-L.; Lacroix, L.; Han, X.; Leroy, J.-L.; Helene, C., Intramolecular folding of pyrimidine oligodeoxynucleotides into an i-DNA motif. *J. Am. Chem. Soc.* **1995,** *117*, 8887-8898.
10. Abou Assi, H.; Garavís, M.; González, C.; Damha, M. J., i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.* **2018,** *46*, 8038-8056.
11. Frank-Kamenetskii, M. D.; Mirkin, S. M., Triplex DNA structures. *Annu. Rev. Biochem.* **1995,** *64*, 65-95.
12. Mariani, A.; Bonfio, C.; Johnson, C. M.; Sutherland, J. D., pH-Driven RNA strand separation under prebiotically plausible conditions. *Biochemistry* **2018,** *57*, 6382-6386.
13. Woodbury, C.; Record Jr, M., A range of G+ C-independent denaturation solvents for DNA. *Biopolymers* **1975,** *14*, 2417-2420.
14. Casey, J. R.; Grinstein, S.; Orlowski, J., Sensors and regulators of intracellular pH. *Nat. Rev. Mol. Cell Biol.* **2010,** *11*, 50-61.
15. Watson, J. D.; Crick, F. H., Molecular structure of nucleic acids. *Nature* **1953,** *171*, 737-738.
16. Courtois, Y.; Fromageot, P.; Guschlbauer, W., Protonated polynucleotide structures: 3. An optical rotatory dispersion study of the protonation of DNA. *Eur. J. Biochem.* **1968,** *6*, 493-501.
17. González-Olvera, J. C.; Durec, M.; Marek, R.; Fiala, R.; Morales-García, M. d. R. J.; González-Jasso, E.; Pless, R. C., Protonation of Nucleobases in Single-and Double-Stranded DNA. *ChemBioChem* **2018,** *19*, 2088-2098.
18. Smol'janinova, T.; Zhidkov, V.; Sokolov, G., Analysis of difference spectra of protonated DNA: determination of degree of protonation of nitrogen bases and the fractions of disordered nucleotide pairs. *Nucleic Acids Res.* **1982,** *10*, 2121-2134.

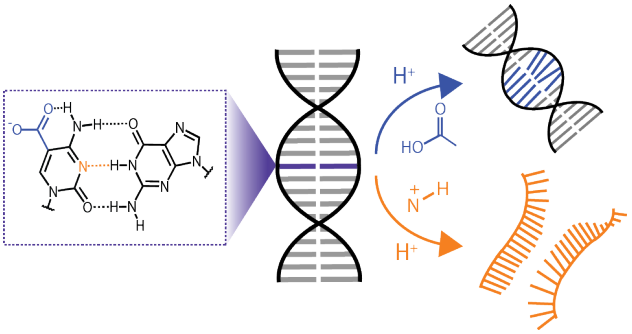
19. Sorokin, V. A.; Gladchenko, G. O.; Valeev, V. A., DNA protonation at low ionic strength of solution. *Macromol. Chem. Phys.* **1986,** *187*, 1053-1063.
20. Tajmir-Riahi, H.; Ahmad, R.; Naoui, M.; Diamantoglou, S., The effect of HCl on the solution structure of calf thymus DNA: a comparative study of DNA denaturation by proton and metal cations using Fourier transform IR difference spectroscopy. *Biopolymers* **1995,** *35*, 493-501.
21. Nikolova, E. N.; Goh, G. B.; Brooks III, C. L.; Al-Hashimi, H. M., Characterizing the protonation state of cytosine in transient G· C Hoogsteen base pairs in duplex DNA. *J. Am. Chem. Soc.* **2013,** *135*, 6766-6769.
22. Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D., I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* **2018,** *10*, 631-637.
23. Dai, Q.; Sanstead, P. J.; Peng, C. S.; Han, D.; He, C.; Tokmakoff, A., Weakened N3 hydrogen bonding by 5-formylcytosine and 5-carboxylcytosine reduces their base-pairing stability. *ACS Chem. Biol.* **2015,** *11*, 470-477.
24. Privat, E. J.; Sowers, L. C., A proposed mechanism for the mutagenicity of 5-formyluracil. *Mutat. Res.* **1996,** *354*, 151-156.
25. Driggers, P.; Beattie, K., Effect of pH on the base-mispairing properties of 5-bromouracil during DNA synthesis. *Biochemistry* **1988,** *27*, 1729-1735.
26. Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y., Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **2011,** *333*, 1300-1303.
27. Zhang, L.; Lu, X.; Lu, J.; Liang, H.; Dai, Q.; Xu, G.-L.; Luo, C.; Jiang, H.; He, C., Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.* **2012,** *8*, 328-330.
28. Maiti, A.; Michelson, A. Z.; Armwood, C. J.; Lee, J. K.; Drohat, A. C., Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA. *J. Am. Chem. Soc.* **2013,** *135*, 15813-15822.
29. La Francois, C. J.; Jang, Y. H.; Cagin, T.; Goddard, W. A.; Sowers, L. C., Conformation and proton configuration of pyrimidine deoxynucleoside oxidation damage products in water. *Chem. Res. Toxicol.* **2000,** *13*, 462-470.
30. Hashimoto, H.; Hong, S.; Bhagwat, A. S.; Zhang, X.; Cheng, X., Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucleic Acids Res.* **2012,** *40*, 10203-10214.
31. Dong, Y.; Yang, Z.; Liu, D., DNA nanotechnology based on i-motif structures. *Acc. Chem Res.* **2014,** *47*, 1853-1860.
32. Idili, A.; Vallée-Bélisle, A.; Ricci, F., Programmable pH-triggered DNA nanoswitches. *J. Am. Chem. Soc.* **2014,** *136*, 5836-5839.
33. Porchetta, A.; Idili, A.; Vallée-Bélisle, A.; Ricci, F., General strategy to introduce pH-induced allostery in DNA-based receptors to achieve controlled release of ligands. *Nano Lett.* **2015,** *15*, 4467-4471.
34. Amodio, A.; Zhao, B.; Porchetta, A.; Idili, A.; Castronovo, M.; Fan, C.; Ricci, F., Rational design of pH-controlled DNA strand displacement. *J. Am. Chem. Soc.* **2014,** *136*, 16469-16472.

35. Modi, S.; Nizak, C.; Surana, S.; Halder, S.; Krishnan, Y., Two DNA nanomachines map pH changes along intersecting endocytic pathways inside the same cell. *Nat. Nanotechnol.* **2013,** *8*, 459-467.
36. Lannes, L.; Halder, S.; Krishnan, Y.; Schwalbe, H., Tuning the pH Response of i-Motif DNA Oligonucleotides. *ChemBioChem* **2015,** *16*, 1647-1656.
37. Zamiri, B.; Mirceta, M.; Bomsztyk, K.; Macgregor Jr, R. B.; Pearson, C. E., Quadruplex formation by both G-rich and C-rich DNA strands of the C9orf72 (GGGGCC) 8•(GGCCCC) 8 repeat: effect of CpG methylation. *Nucleic Acids Res.* **2015,** *43*, 10055-10064.
38. Wright, E. P.; Abdelhamid, M. A. S.; Ehiabor, M. O.; Grigg, M. C.; Irving, K.; Smith, N. M.; Waller, Z. A. E., Epigenetic modification of cytosines fine tunes the stability of i-motif DNA. *Nucleic Acids Res.* **2019**, DOI: 10.1093/nar/gkz1082.
39. Sumino, M.; Ohkubo, A.; Taguchi, H.; Seio, K.; Sekine, M., Synthesis and properties of oligodeoxynucleotides containing 5-carboxy-2′-deoxycytidines. *Bioorg. Med. Chem. Lett.* **2008,** *18*, 274-277.
40. Münzel, M.; Lischke, U.; Stathis, D.; Pfaffeneder, T.; Gnerlich, F. A.; Deiml, C. A.; Koch, S. C.; Karaghiosoff, K.; Carell, T., Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine. *Chem. Eur. J.* **2011,** *17*, 13782-13788.
41. Szulik, M. W.; Pallan, P. S.; Nocek, B.; Voehler, M.; Banerjee, S.; Brooks, S.; Joachimiak, A.; Egli, M.; Eichman, B. F.; Stone, M. P., Differential stabilities and sequence-dependent base pair opening dynamics of Watson–Crick base pairs with 5-hydroxymethylcytosine, 5-formylcytosine, or 5-carboxylcytosine. *Biochemistry* **2015,** *54*, 1294-1305.
42. Ngo, T. T.; Yoo, J.; Dai, Q.; Zhang, Q.; He, C.; Aksimentiev, A.; Ha, T., Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* **2016,** *7*, 10813.
43. Hansch, C.; Leo, A.; Taft, R., A survey of Hammett substituent constants and resonance and field parameters. *Chem. Rev.* **1991,** *91*, 165-195.
44. Banyay, M.; Sarkar, M.; Gräslund, A., A library of IR bands of nucleic acids in solution. *Biophys. Chem.* **2003,** *104*, 477-488.
45. Sanstead, P. J.; Stevenson, P.; Tokmakoff, A., Sequence-dependent mechanism of DNA oligonucleotide dehybridization resolved through infrared spectroscopy. *J. Am. Chem. Soc.* **2016,** *138*, 11792-11801.
46. Peng, C. S.; Jones, K. C.; Tokmakoff, A., Anharmonic vibrational modes of nucleic acid bases revealed by 2D IR spectroscopy. *J. Am. Chem. Soc.* **2011,** *133*, 15650-15660.
47. Lee, C.; Cho, M., Vibrational dynamics of DNA. II. Deuterium exchange effects and simulated IR absorption spectra. *J. Chem. Phys.* **2006,** *125*, 114509.
48. Stelling, A. L.; Xu, Y.; Zhou, H.; Choi, S. H.; Clay, M. C.; Merriman, D. K.; Al-Hashimi, H. M., Robust IR-based detection of stable and fractionally populated G-C+ and A-T Hoogsteen base pairs in duplex DNA. *FEBS Lett.* **2017,** *591*, 1770-1784.
49. Falk, M.; Ford, T., Infrared spectrum and structure of liquid water. *Can. J. Chem.* **1966,** *44*, 1699-1707.
50. Albergo, D. D.; Marky, L. A.; Breslauer, K. J.; Turner, D. H., Thermodynamics of (dG-dC) 3 double-helix formation in water and deuterium oxide. *Biochemistry* **1981,** *20*, 1409-1413.
51. Krężel, A.; Bal, W., A formula for correlating pKa values determined in D2O and H2O. *J. Inorg. Biochem.* **2004,** *98*, 161-166.

52. Benoit, R. L.; Fréchette, M., Protonation of hypoxanthine, guanine, xanthine, and caffeine. *Can. J. Chem.* **1985,** *63*, 3053-3056.
53. Major, D. T.; Laxer, A.; Fischer, B., Protonation studies of modified adenine and adenine nucleotides by theoretical calculations and 15N NMR. *J. Org. Chem.* **2002,** *67*, 790-802.
54. Krummel, A. T.; Zanni, M. T., DNA vibrational coupling revealed with two-dimensional infrared spectroscopy: insight into why vibrational spectroscopy is sensitive to DNA structure. *J. Phys. Chem. B* **2006,** *110*, 13991-14000.
55. Peng, C. S. Two-dimensional infrared spectroscopy of nucleic acids: application to tautomerism and DNA aptamer unfolding dynamics. Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2014.
56. Siegfried, N. A.; O'Hare, B.; Bevilacqua, P. C., Driving forces for nucleic acid pKa shifting in an A+· C wobble: effects of helix position, temperature, and ionic strength. *Biochemistry* **2010,** *49*, 3225-3236.
57. Crothers, D. M., Statistical thermodynamics of nucleic acid melting transitions with coupled binding equilibria. *Biopolymers* **1971,** *10*, 2147-2160.
58. Menssen, R. J.; Tokmakoff, A., Length-dependent melting kinetics of short DNA oligonucleotides using temperature-jump IR spectroscopy. *J. Phys. Chem. B* **2019,** *123*, 756-767.
59. Amunson, K. E.; Anderson, B. A.; Kubelka, J., Temperature effects on the optical path length of infrared liquid transmission cells. *Appl. Spec.* **2011,** *65*, 1307-1313.
60. Mergny, J.-L.; Lacroix, L., Analysis of thermal melting curves. *Oligonucleotides* **2003,** *13*, 515-537.
61. Jones, K. C.; Ganim, Z.; Tokmakoff, A., Heterodyne-detected dispersed vibrational echo spectroscopy. *J. Phys. Chem. A* **2009,** *113*, 14060-14066.
62. Sanstead, P. J.; Tokmakoff, A., Direct observation of activated kinetics and downhill dynamics in dna dehybridization. *J. Phys. Chem. B* **2018,** *122*, 3088-3100.
63. Pörschke, D.; Eigen, M., 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.* **1971,** *62*, 361-381.
64. Williams, A. P.; Longfellow, C. E.; Freier, S. M.; Kierzek, R.; Turner, D. H., Laser temperature-jump, spectroscopic, and thermodynamic study of salt effects on duplex formation by dGCATGC. *Biochemistry* **1989,** *28*, 4283-4291.
65. Holmstrom, E. D.; Dupuis, N. F.; Nesbitt, D. J., Pulsed IR heating studies of single-molecule DNA duplex dissociation kinetics and thermodynamics. *Biophys. J.* **2014,** *106*, 220-231.
66. Owczarzy, R., Melting temperatures of nucleic acids: discrepancies in analysis. *Biophys. Chem.* **2005,** *117*, 207-215.
67. Wartell, R. M.; Benight, A. S., Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Phys. Rep.* **1985,** *126*, 67-107.
68. Sanstead, P. J.; Ashwood, B.; Dai, Q.; He, C.; Tokmakoff, A., Oxidized Derivatives of 5-Methylcytosine Alter the Stability and Dehybridization Dynamics of Duplex DNA. *Submitted*.
69. Bernasconi, C., *Relaxation kinetics*. Academic Press: New York 1976.
70. Kumar, A. T.; Zhu, L.; Christian, J.; Demidov, A. A.; Champion, P. M., 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* **2001,** *105*, 7847-7856.

71. Kramers, H. A., Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940,** *7*, 284-304.
72. Cho, C.; Urquidi, J.; Singh, S.; Robinson, G. W., Thermal offset viscosities of liquid H2O, D2O, and T2O. *J. Phys. Chem. B* **1999,** *103*, 1991-1994.
73. Chung, H. S.; Tokmakoff, A., Temperature-dependent downhill unfolding of ubiquitin. I. Nanosecond-to-millisecond resolved nonlinear infrared spectroscopy. *Proteins* **2008,** *72*, 474-487.
74. Jäger, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M., The folding mechanism of a β-sheet: The WW domain. *J. Mol. Biol.* **2001,** *311*, 373-393.
75. Crane, J. C.; Koepf, E. K.; Kelly, J. W.; Gruebele, M., Mapping the transition state of the WW domain β-sheet. *J. Mol. Biol.* **2000,** *298*, 283-292.
76. Dupuis, N. F.; Holmstrom, E. D.; Nesbitt, D. J., Single-molecule kinetics reveal cation-promoted DNA duplex formation through ordering of single-stranded helices. *Biophys. J.* **2013,** *105*, 756-766.
77. Hill, A. V., The combinations of haemoglobin with oxygen and with carbon monoxide. I. *Biochem. J.* **1913,** *7*, 471-480.
78. SantaLucia, J., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci.* **1998,** *95*, 1460-1465.
79. Spasic, A.; Berger, K. D.; Chen, J. L.; Seetin, M. G.; Turner, D. H.; Mathews, D. H., Improving RNA nearest neighbor parameters for helices by going beyond the two-state model. *Nucleic Acids Res.* **2018,** *46*, 4883-4892.
80. Hardwick, J. S.; Lane, A. N.; Brown, T., Epigenetic modifications of cytosine: biophysical properties, regulation, and function in mammalian DNA. *BioEssays* **2018,** *40,* 1700199.
81. Drohat, A. C.; Coey, C. T., Role of base excision "repair" enzymes in erasing epigenetic marks from DNA. *Chem. Rev.* **2016,** *116*, 12711-12729.
82. Mingot, F., Quantitative test of Record's theory for proton-induced lowering of DNA melting temperature. *Biopolymers* **1981,** *20*, 2121-2136.
83. Chen, H.; Meisburger, S. P.; Pabit, S. A.; Sutton, J. L.; Webb, W. W.; Pollack, L., Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci.* **2012,** *109*, 799-804.
84. Weiss, J. N., The Hill equation revisited: uses and misuses. *FASEB J.* **1997,** *11*, 835-841.
85. Teng, X.; Hwang, W., Effect of methylation on local mechanics and hydration structure of DNA. *Biophys. J.* **2018,** *114*, 1791-1803.
86. Pidugu, L. S.; Dai, Q.; Malik, S. S.; Pozharski, E.; Drohat, A. C., Excision of 5-carboxylcytosine by Thymine DNA Glycosylase. *J. Am. Chem. Soc.* **2019,** *141*, 18851-18861.
87. Jones, K. C.; Ganim, Z.; Peng, C. S.; Tokmakoff, A., Transient two-dimensional spectroscopy with linear absorption corrections applied to temperature-jump two-dimensional infrared. *J. Opt. Soc. Am. B* **2012,** *29*, 118-129.
88. Chung, H. S.; Khalil, M.; Smith, A. W.; Tokmakoff, A., Transient two-dimensional IR spectrometer for probing nanosecond temperature-jump kinetics. *Rev. Sci. Instrum.* **2007,** *78*, 063101.
89. Lepetit, L.; Chériaux, G.; Joffre, M., Linear techniques of phase measurement by femtosecond spectral interferometry for applications in spectroscopy. *J. Opt. Soc. Am. B* **1995,** *12*, 2467-2474.

TOC Graphic:



5-Carboxylcytosine and Cytosine Protonation Distinctly Alter the Stability and Dehybridization Dynamics of the DNA Duplex

Brennan Ashwood[1,2,3], Paul J. Sanstead[1,2,3], Qing Dai[1,2], Chuan He[1,2,4,5], Andrei Tokmakoff[1,2,3,*]

[1]Department of Chemistry, [2]Institute for Biophysical Dynamics, [3]James Franck Institute, [4]Department of Biochemistry and Molecular Biology, and [5]Howard Hughes Medical Institute, The University of Chicago, Chicago, Illinois 60637, United States

Supporting Information

1. pH-Dependent FTIR Spectra of X = C and X = caC

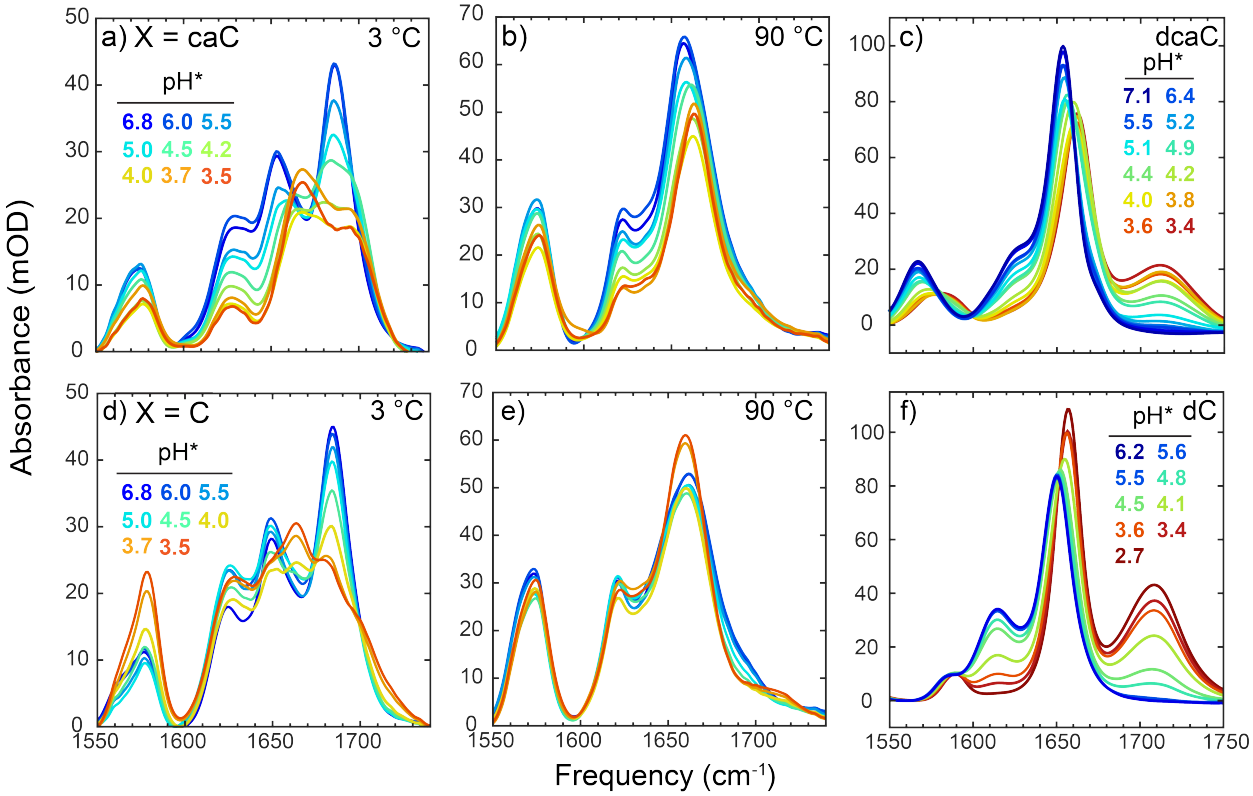


Figure S1: FTIR spectra of X = caC at (a) 3 °C and (b) 90 °C from pH* 6.8 to 3.5. (c) FTIR spectra of 2'-deoxy-5-carboxylcytidine (dcaC) from pH* 7.1 to 3.4 at room temperature in D_2O. FTIR spectra of X = C at (d) 3 °C and (e) 90 °C from pH* 6.8 to 3.5. (f) FTIR spectra of 2'-deoxycytidine (dC) from pH* 6.2 to 2.7 at room temperature in D_2O.

Fig. S1 presents FTIR spectra of X = C and X = caC at 3 °C and 90 °C from pH* 6.8 to 3.5. The 3 °C spectrum of X = caC corresponds to the duplex state and its changes with pH* reflect protonation of X = caC. As pH* is reduced, the 1575 and 1625 cm^{-1} features decrease in intensity, the 1650 cm^{-1} band shifts to 1665 cm^{-1}, and the 1685 cm^{-1} peak drastically drops in amplitude and shifts to lower frequency while a new peak arises at 1700 cm^{-1}. Similar pH*-dependent changes are observed in the 90 °C spectrum corresponding to the single-strand. The spectral changes exhibited by X = caC with pH* are consistent with 2'-deoxy-5-carboxylcytidine (dcaC), suggesting the duplex and single-strand spectral changes are due to caC protonation. Interestingly, the protonated caC peak at 1700 cm^{-1} decays at high temperature. The loss of the 1700 cm^{-1} peak is also observed the FTIR temperature series of dcaC at pH* 4.0 (Fig. S2), suggesting it may come

from changes in nucleobase solvation with temperature. Another possibility is that the pK_a of caC decreases as temperature is increased. However, 2nd SVD components along pH* for X = caC at 3, 50, and 90 °C (Fig. S3a) each show similar titration behavior, suggesting that the pK_a for caC does not change significantly with temperature.

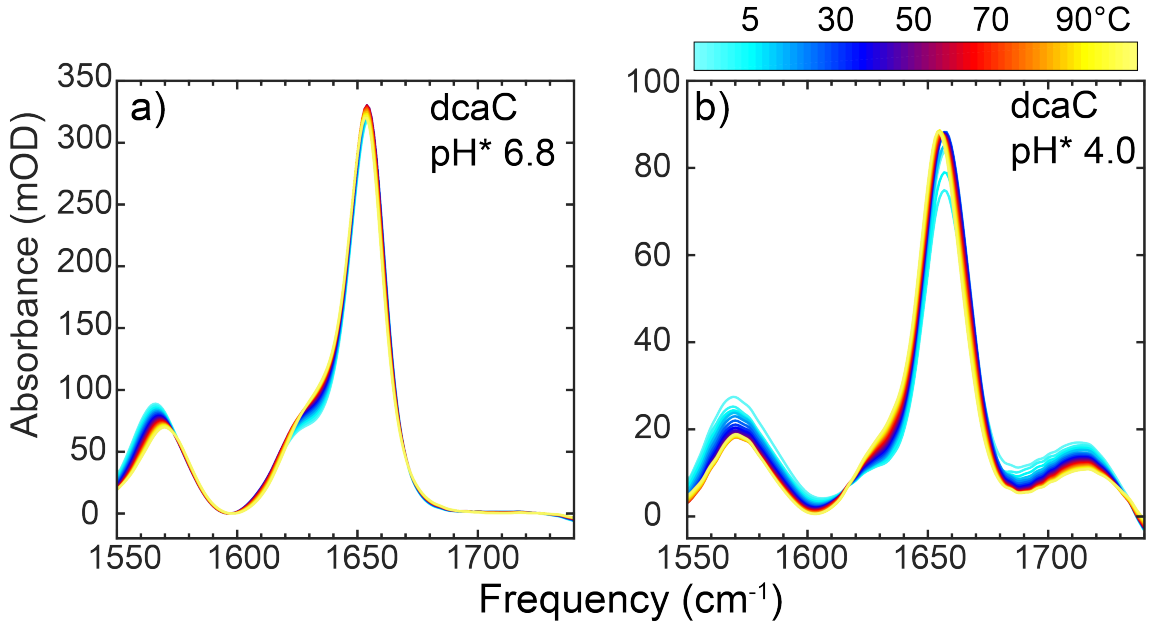


Figure S2: FTIR temperature series from 3 – 97 °C for 2'-deoxy-5-carboxylcytidine (dcaC) prepared at pH* (a) 6.8 and (b) 4.0.

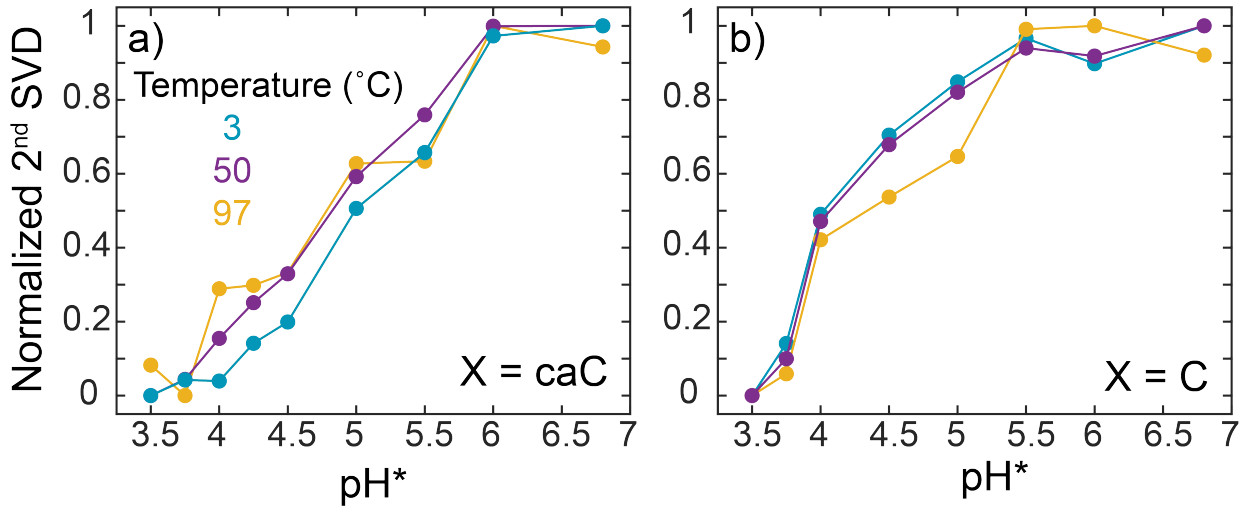


Figure S3: 2nd SVD components of pH*-dependent FTIR spectra of (a) X = caC and (b) X = C from pH* 6.8 to 3.5 at 3, 50, and 97 °C.

As pH* is reduced, the spectral changes of X = C resembles those of 2'-deoxycytidine. In particular, the C feature at 1650 cm^{-1} shifts to 1660 cm^{-1} and a new band is formed near 1700 cm^{-1}. Additionally, the 1575 and 1685 cm^{-1} rise and decay at 3 °C, respectively, as pH* is reduced, indicating duplex dissociation. FTIR spectra of X = C from pH* 6.8 to 1.8 at 3 °C are shown in Fig. 3.

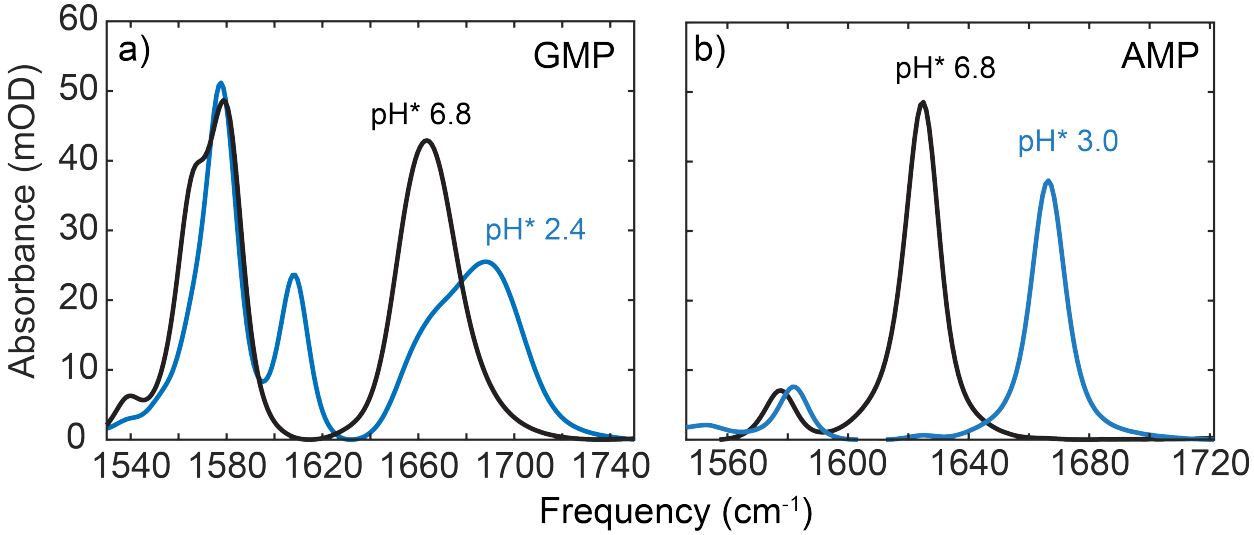


Figure S4: FTIR spectra of (a) 2'-deoxyguanosine 5'-monophosphate (GMP) at pH* 6.8 and 2.4 and (b) 2'-deoxyadenosine 5'-monophosphate (AMP) at pH* 6.8 and 3.0. Samples were prepared at 3 mg/ml in deuterated 20 mM sodium phosphate buffer with 16 mM NaCl.

Among the acidic titratable groups of the canonical DNA nucleobases, C possesses the highest pK_a. However, protonation of adenine (AMP pK_a: 3.9-4.0) and guanine (GMP pK_a: 3.0-3.3) at the N1 and N7 positions, respectively, become possible at low pH* as well. Neutral and low pH* FTIR spectra for AMP and GMP are shown in Fig. S4. N1 protonation of adenine leads to a shift of the 1625 cm^{-1} ring mode to 1665 cm^{-1}. Such a spectral shift is not apparent in the pH*-dependent FTIR spectra of X = C and X = caC. However, the spectral changes associated with caC protonation are largely overlapped with those of adenine, making it difficult to assess the degree of adenine protonation in X = caC. In X = C, signatures of adenine protonation become evident around pH* 3.0 (Fig. 3). An SVD of X = C along pH* in the 1600-1640 cm^{-1} frequency window (Fig. S5), which is dominated by changes in adenine, reveals titration behavior shifted much lower in pH* relative to C protonation. The extracted pK_a of 2.7 is ~1 pH unit lower than that for AMP

in solution and suggests that the pK_a of adenine is shifted in the presence of the other DNA nucleobases. The degree of adenine protonation per duplex is calculated using eq. S10 and indicates that adenine sites remain almost completely deprotonated over the pH* range used for thermodynamic and kinetic measurements of X = C in this work. Signs of guanine protonation are only observed below pH* 2.5 in Fig. 3, suggesting that the pK_a of guanine is also shifted lower in X = C.

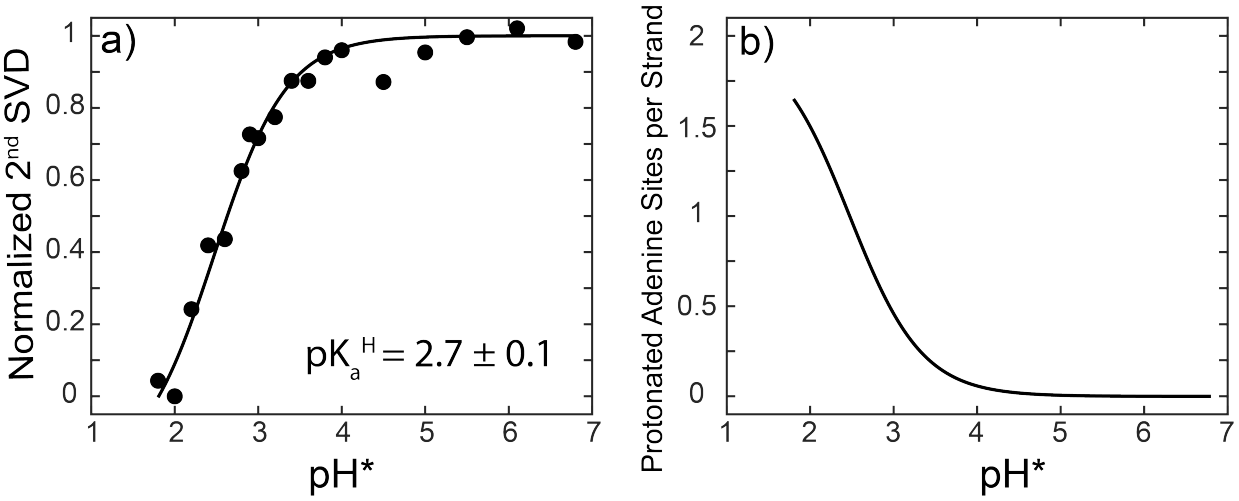


Figure S5: (a) 2nd SVD component for FTIR titration of X = C over the frequency window 1600-1640 cm^{-1} fit to the Henderson-Hasselbalch equation (solid line). (b) Number of protonated adenine sites per X = C strand as a function of pH* determined from the pK_a in (a).

2. Explanations for Loss of 1685 cm^{-1} Peak Upon Protonation of X = caC

As shown in Fig. S1a, the pH*-dependent FTIR spectra of X = caC at low temperature show clear signs of caC nucleobase protonation. However, the origin behind the additional loss and shift of the 1685 cm^{-1} guanine carbonyl peak may be related to additional changes in the DNA duplex and requires further explanation. DNA duplex dissociation has always been shown to contain a gain in absorbance of the 1575 cm^{-1} guanine ring mode in addition to changes of the 1685 cm^{-1}. Due to the lack of a gain in the 1575 cm^{-1} at 3 °C as a function of pH*, the duplex fraction (θ_{ext}) is assumed to be unperturbed by 5-carboxyl protonation at low temperature. In general, the 1685 cm^{-1} feature has been shown to be very sensitive to DNA structural changes.[1-3] In particular, the formation of hydrogen bonds with the N7 atom of an adjacent guanine, which

occurs in G-quadruplex and G:C Hoogsteen base pair formation, have been shown to significantly reduce the 1685 cm^{-1} absorbance.

The frequency window from 1300-1500 cm^{-1} primarily reports on nucleobase ring modes and deoxyribose-nucleobase coupled vibrations in DNA. Some of these features report on deoxyribose conformation and can be used to detect shifts from anti- to syn-orientations. This is particularly important to consider for X = caC as major groove substituents have been known to promote such changes in sugar conformation and may impact local base pairing in a such a way that alters the carbonyl region of the nucleobase spectrum.[2, 4-5] Fig. S6a shows the FTIR spectra of X = caC at pH* 6.8 and 3.5 from 1300-1500 cm^{-1}. The region is crowded with vibrations from each of the nucleobases. The region from 1400-1500 cm^{-1} is dominated by bands from adenine and guanine that remain fairly constant with pH*. In contrast the lower frequency half of the spectrum changes significantly in going to low pH*. The feature at 1380 cm^{-1} decreases in intensity and shifts to slightly lower frequency while the band at 1340 cm^{-1} is essentially lost. Similar changes are observed in the FTIR spectrum of dcaC (Fig. S6b), suggesting that the pH*-dependence to this frequency region result from the caC nucleobase itself. Density functional theory (DFT) harmonic frequency calculations were performed on dcaC to try to assign the vibrations. While distinct spectra are observed at each pH*, the discrepancy with the experimental spectra makes it difficult to separately assign the features at 1340 and 1380 cm^{-1} in the X = caC spectrum.

For identification of anti- to syn- rotation in guanine, a signature of Z-DNA or Hoogsteen base pairing, we can look for low intensity features at 1350 and 1325 cm^{-1} in deuterated solutions, which report on anti- and syn-purine orientation, respectively.[2] At pH* 6.8 and 3.5, a shoulder and peak are observed at 1350 cm^{-1}, respectively, but no feature is observed at 1325 cm^{-1}. This suggests that anti deoxyribose orientation dominates across the pH* range studied here and that Hoogsteen base pairs or Z-DNA are not induced upon 5-carboxyl protonation of X = caC.

One possibility to consider is that the carboxylic acid of caC can donate a hydrogen bond to adjacent guanine bases in the duplex. In B-form DNA, both the carboxylic acid and N7 position of guanine orient toward the major groove and are nearly stacked on top of one another. Therefore, the carboxylic acid could potentially donate a hydrogen bond to the N7 position of guanine (Fig. S7), and this interaction may be responsible for the suppression and shift of the 1685 cm^{-1}

peak to lower frequency. Another possibility is that caC protonation alters local base pairing, base-stacking, or solvation interactions with guanine bases. The exocyclic COOH group is much more electron withdrawing than COO^-, potentially weakening G:C base pairing and altering stacking interactions with adjacent guanine bases.[6] Additionally, protonation of caC changes the charge and potentially solvation environment in the major groove of DNA.[7] Each of these effects may directly affect interactions with guanine, particularly at the N7 position. Overall, the exact nature for loss and shift of the 1685 cm^{-1} mode to lower frequency at low temperature and pH* is still unclear, but the absence of an intensity gain at 1575 cm^{-1} or large shift in duplex-to-single-strand melting inflection point suggest that it is not due to a reduction in duplex concentration.

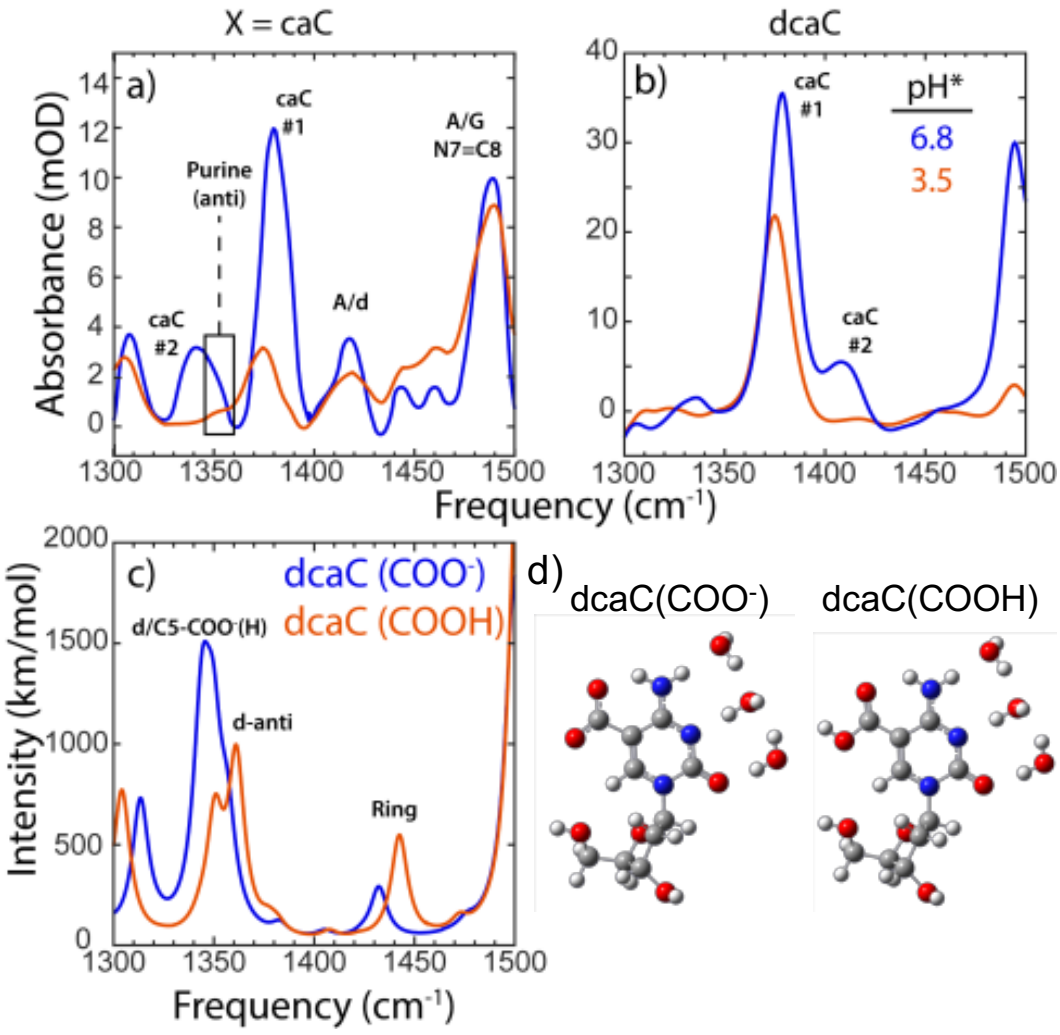


Figure S6: (a) FTIR spectrum of (a) X = caC and (b) 2'-deoxy-5-carboxylcytidine (dcaC) at pH* 6.8 and 3.5 in the frequency window 1300-1500 cm^{-1}. (c) Calculated IR spectra of 5-carboxyl

protonated and deprotonated dcaC from harmonic frequency calculations of optimized structures with solvating D_2O molecules (d) performed at the B3LYP/6-311++G(d,p)/IEFPCM level of theory. Each vibrational mode is shown with a Lorentzian lineshape and 5 cm^{-1} FWHM. Here, d indicates vibrations with significant deoxyribose character.

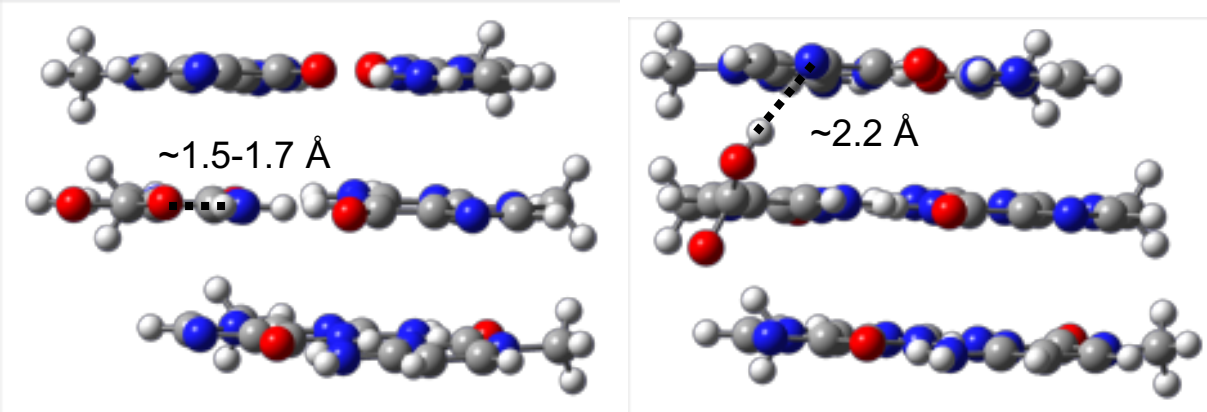


Figure S7: Structure of B-form 5'-GCG-3' trinucleotide containing a protonated 5-carboxyl substituent on the middle cytosine. The left indicates the planar structure of the carboxylic acid with the carbonyl forming a hydrogen bond to the ND_2 group at the 4 position. The right demonstrates that the carboxylic acid hydrogen can donate a hydrogen-bond to the N7 position of adjacently stacked guanine bases.

3. Signatures of Pre-Melting Along Low-Temperature Baseline of FTIR Melting Curves

Fig. S8 shows the FTIR spectra of X = caC at pH* 6.8 and 3.7 taken along the low-temperature melting curve baseline (3 – 38 °C). At pH* 6.8, the main spectral changes within this temperature range are increases in absorbance from 1620 to 1680 cm^{-1} and a small reduction at 1690 cm^{-1}. These peaks contain overlapping contributions from each nucleobase, and therefore it is difficult to be certain whether the temperature-dependent changes in Fig. S8a correspond to changes in A:T and/or G:C base pairing. However, each of the observed spectral trends are consistent with a reduction in A:T base pairing, and may reflect fraying of the A:T termini.

At pH* 3.7, X = caC exhibits an increase in absorbance at 1575 and 1660 cm^{-1} in addition to the spectral changes observed at pH* 6.8. These extra changes are consistent with the loss of G:C base pairing and provide evidence for a pre-melting of G:C contacts, in addition to fraying of the A:T termini, prior to the cooperative duplex-to-single-strand transition.

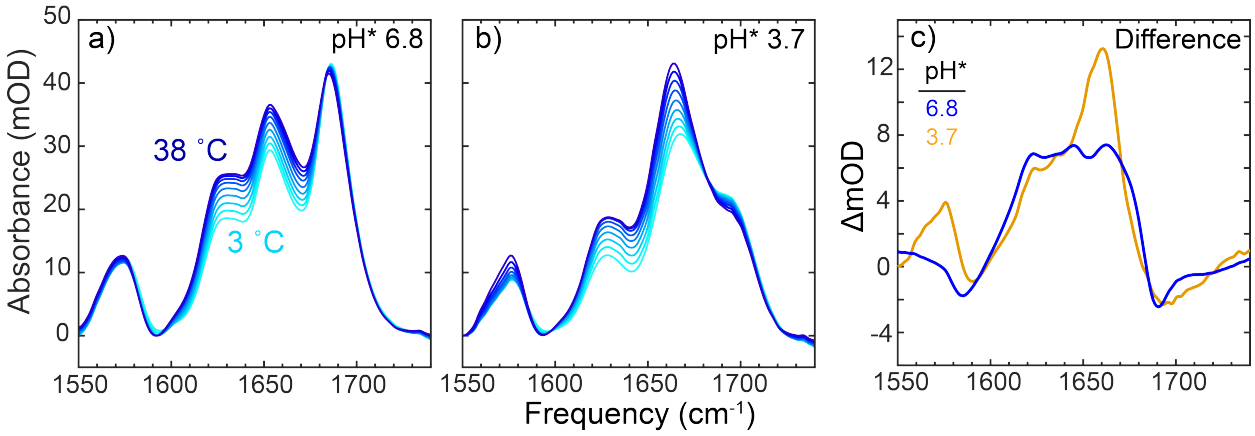


Figure S8: FTIR temperature series of X = caC from 3 to 38 °C at (a) pH* 6.8 and (b) pH* 3.7. (c) Difference spectra between 38 and 3 °C.

4. Determining Degree of Protonation in X = C and X = caC Duplex

The degree of 5caC protonation in X = caC can be directly determined from the 2nd SVD component across pH* shown in Fig. 3 of the main text. The 2nd SVD component is well fit to the Henderson-Hasselbach equation. At 3 °C, the duplex fraction (θ_{ext}) is assumed to be essentially independent of pH*, and therefore the fit 2nd SVD component describes the degree of protonation in the duplex state. The number of protonated 5caC sites per duplex (χ) is given by Equation S1:

$$\chi = \frac{[HA]\theta_{ext}}{[D]} \quad \text{(S1)}$$

where [HA] and [D] are the concentration of protonated 5caC sites and duplex, respectively. There are three 5caC bases per DNA strand, so the total concentration of sites is three times greater than the DNA concentration (c_{tot}):

$$[HA]+[A^-] = 3c_{tot} \quad \text{(S2)}$$

Rearranging the Henderson-Hasselbach equation gives an expression for [HA]:

(S3) $$[HA] = \frac{[A^-]}{10^{pH-pK_a}}$$

which can be expressed in terms of c_{tot} using Equation S4:

(S4) $$[HA] = \frac{3c_{tot}}{1+10^{pH-pK_a}}$$

The DNA duplex concentration ([D]) can be cast in terms of c_{tot} and θ_{ext}:

(S5) $$[D] = \frac{\theta_{ext} c_{tot}}{2}$$

Then, Equations S2, S4, and S5 can be used to determine χ for X = caC:

(S6) $$\chi = \frac{6}{1+10^{pH-pK_a}}$$

The plot of χ for X = caC is shown in Fig. 3. Over the pH* range studied here, X = caC varies from fully deprotonated at pH* 6.8 to near complete protonation of all six sites at pH* 3.5.

Unlike X = caC, θ_{ext} of X = C is highly sensitive to pH* and requires a modified treatment to determine the degree of duplex protonation. An FTIR titration of X = C (Fig. 3) was performed, where the 2nd SVD component is assumed to primarily report on the pH*-induced duplex-to-single-strand transition. The duplex-to-single-strand transition and cytosine N3 protonation appear to occur over the same pH* range, indicating that pH_m, the pH at which θ_{ext} = 0.5 at 3 °C, is equivalent to the apparent pK_a of N3 protonation. Therefore, the duplex fraction can be explicitly related to degree of N3 protonation in the single-strand and duplex at a specific pH*. The percentage of protonated N3 sites is given by Equation S7:

(S7a) $$\theta_p = \frac{[HA]}{[HA]+[A^-]}$$

(S7b) $$= \frac{1}{1+10^{pH-pK_a}}$$

Here we define [HA] in terms of single-strand ([S]) and duplex ([D]) concentrations:

(S8) $$[HA] = c_S[S] + c_D[D]$$

where c_S and c_D represent the average number of protonated N3 sites in the single-strand and duplex states, respectively. By defining [S] and [D] in terms of c_{tot} and $\theta_{ext,}$ and using Equation S7a and S2:

(S9) $$\theta_p = \frac{1}{3}\left[c_S(1-\theta_{ext}) + \frac{c_D\theta_{ext}}{2} \right]$$

By re-arranging this expression for θ_p and substituting in Equation S7b, an expression relating θ_{ext} to single-strand and duplex protonation is given:

(S10) $$\theta_{ext} = \frac{\frac{3}{10^{pH-pK_a}+1} - c_S}{0.5c_D - c_S}$$

Assuming that the pH*-induced duplex-to-single-strand transition follows the N3 titration curve, Equation S10 describes the relation between θ_{ext} and N3 protonation at low temperature. Equation S10 can be used to fit the 2nd SVD component from the FTIR titration of X = C, where pK_a, $c_{S,}$ and c_D are used as fit parameters. The 2nd SVD component is well fit (Fig. 3) only when c_S = 2.92 and c_D = 0.02, indicating that the X = C duplex cannot tolerate any degree of N3 protonation. This suggests that the thermodynamic and kinetic consequences of N3 protonation purely arise from shifting in the duplex-to-single-strand transition equilibrium at low pH*.

5. Assignment of Temperature-Jump Processes with Equilibrium and Transient 2DIR

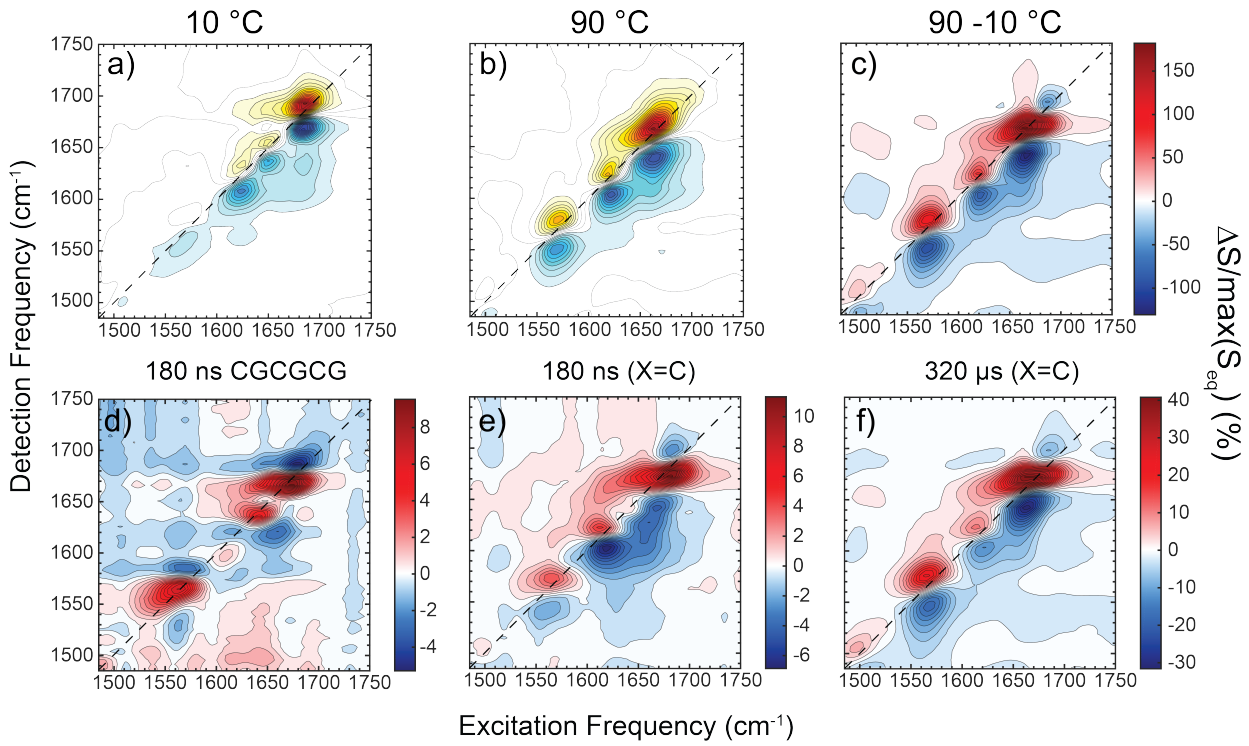


Figure S9: (a) Equilibrium 2DIR spectrum of X = C at pH* 6.8 and 10 °C with a ZZZZ polarization scheme and at a waiting time (τ_2) of 150 fs. (b) Corresponding 2DIR spectrum at 90 °C. (c) Difference 2DIR spectrum between panels (b) and (a) plotted in percent change of signal. (d) Temperature-jump 2DIR (t-2DIR) difference spectrum of 5'-CGCGCG-3' taken at a T-Jump delay of 180 ns and in the ZZZZ polarization scheme. (e) Corresponding t-2DIR spectrum for X = C at pH* 6.8. (f) t-2DIR spectrum of X = C at a T-Jump delay of 320 µs.

Transient heterodyned dispersed vibrational echo (t-HDVE) spectra are acquired at many T-Jump delays in order to determine λ_{ns} and $\lambda_{\mu s}$ in a time-efficient manner. However, the HDVE spectrum is equivalent to the projection of the full two-dimensional IR (2D IR) spectrum onto the detection axis, and therefore features in the HDVE spectrum may be congested with multiple features from the 2D IR spectrum. Therefore, examination of the temperature-jump 2D IR (t-2DIR) spectrum at select T-Jump delays is used to make more informed assignments of λ_{ns} and $\lambda_{\mu s}$. Fig. S9 shows the equilibrium 2D IR spectra of X = C at 10 and 90 °C. Features along the diagonal (dashed line) exist as positive (red)/negative (blue) doublets. Positive peaks along the diagonal contain to ground-state bleach and stimulated emission pathways that correspond to transitions between the ground and first vibrational excited-state. Negative peaks correspond to excited-state absorption

between the first and second vibrational excited-states and are shifted to lower frequency along the detection axis, reporting on the anharmonic shape of the vibrational potentials. Off-diagonal features correspond to cross-peaks between separate vibrational modes and may report on vibrational coupling or dynamics. The 3 °C spectrum of X = C contains doublets centered at 1575, 1620, 1650, and 1685 cm^{-1} along the excitation axis, consistent with the FTIR spectra shown in Fig. S1. As the temperature is increased to 90 °C and the duplex dissociates, the 1685 cm^{-1} guanine peak decays with a correspond gain at 1660 cm^{-1} and gain of the 1575 cm^{-1} guanine ring mode and 1620 cm^{-1} adenine ring mode. The difference spectrum between these two temperatures (Fig. S9c) reflects these intensity changes.

The t-2DIR spectra at T-Jump delays of 180 ns and 320 μs for X = C are shown in Fig. S9e and f, respectively. The 320 μs surface shows strikingly similarity to the equilibrium difference 2D IR surface between 90 and 10 °C, suggesting that $\lambda_{\mu s}$ corresponds to the duplex-to-single-strand transition. On the other hand, the 180 ns spectrum is distinct from the 320 μs and equilibrium difference spectra. In particular, the response at 1620 cm^{-1}, which corresponds to A and T absorptions, is larger relative to the other features than at 320 μs and contains nearly 50% of the response observed at 320 μs. In contrast, the gain in intensity of the G ring mode at 1575 cm^{-1} at 180 ns contains only 10% of the total response observed at 320 μs. Additionally, the carbonyl region of the 180 ns surface is dominated by a loss at 1690 cm^{-1} with a relatively small gain at 1660 cm^{-1}. Both the changes in the ring and carbonyl mode regions suggest that the 180 ns response primarily corresponds to the loss of A:T base pairing rather than G:C base pairing. The response at 1620 cm^{-1} could also have contributions from C. To account for this, we measured the t-2DIR spectrum of 5'-CGCGCG-3' at 180 ns and observe a very weak difference signal from C at 1620 cm^{-1}. However, the difference signal is negative in the ground-state bleach and has opposite sign from the AT response in this region. Therefore, the 1620 cm^{-1} difference feature for X = C is confidently assigned to AT bases.

6. t-HDVE Spectra and Additional Kinetic Traces

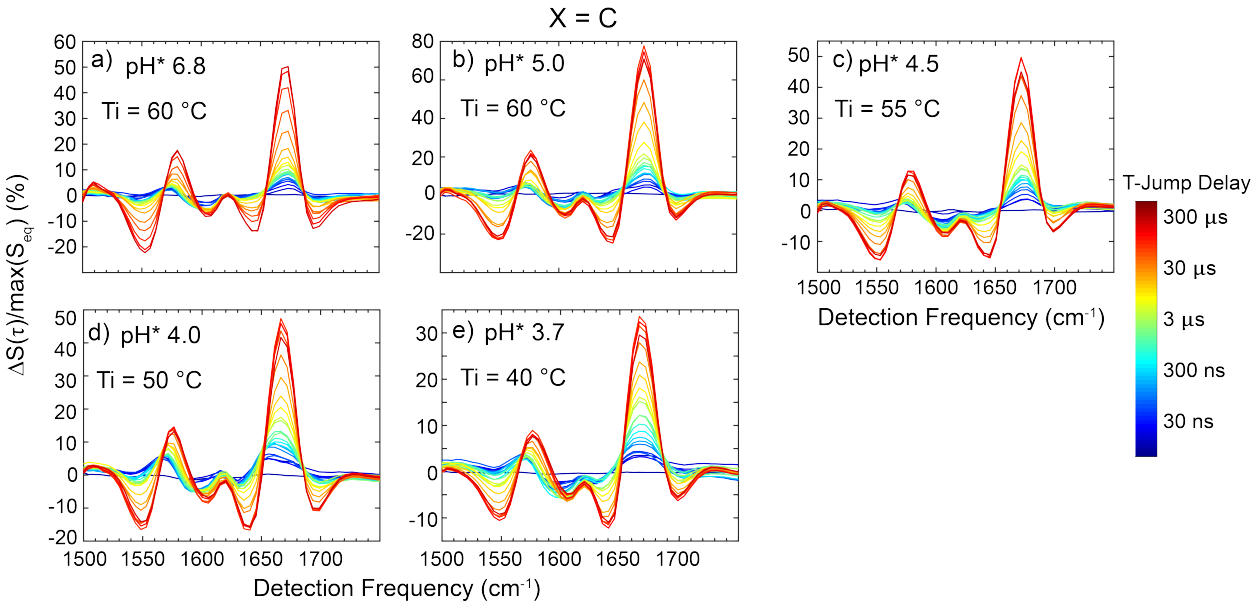


Figure S10: t-HDVE spectra of X = C at select initial temperature (T_i) at each pH*. Spectra at T-jump delays up to 500 µs are shown.

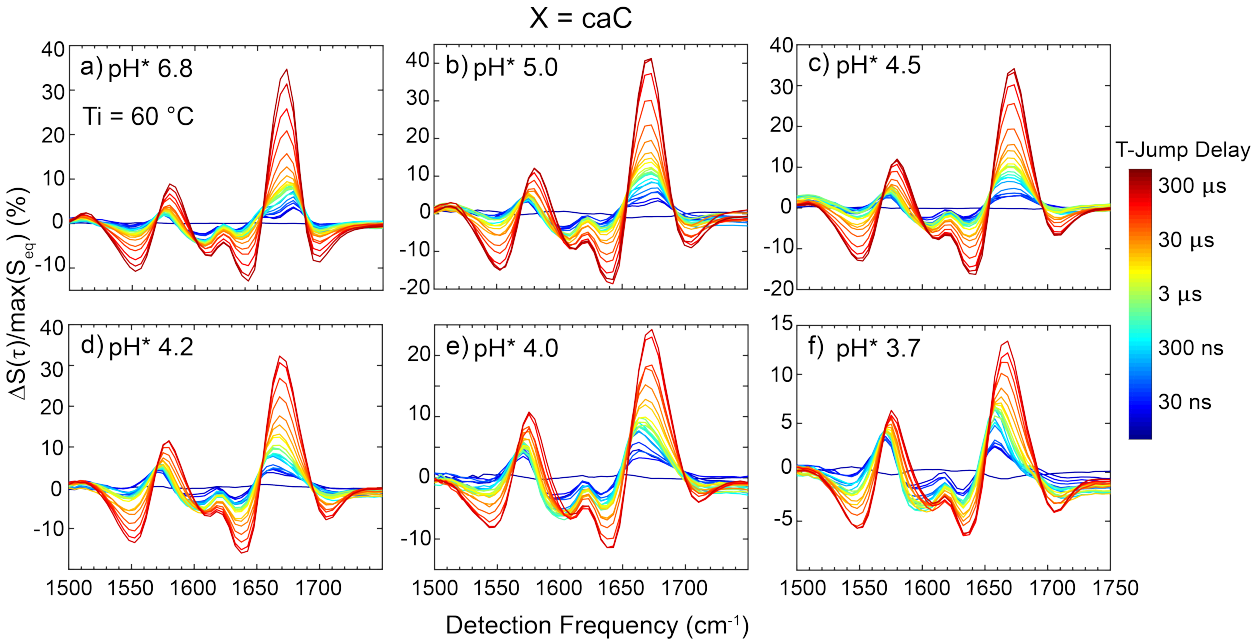


Figure S11: t-HDVE spectra of X = caC at an initial temperature (T_i) of 60 °C at each pH*. Spectra at T-jump delays up to 500 µs are shown.

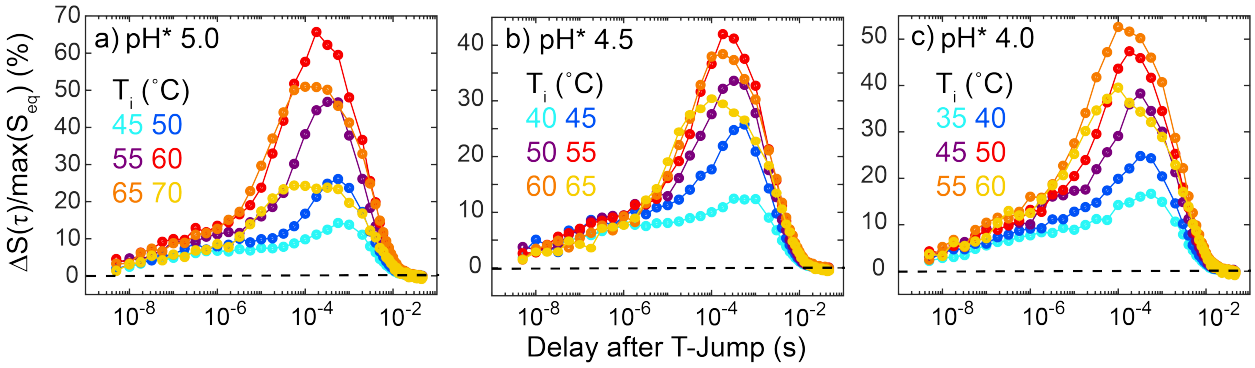


Figure S12: t-HDVE kinetic traces probed at 1670 cm^{-1} for X = C at pH* (a) 5.0, (b) 4.5, and (c) 4.0 and each initial temperature (T_i). Kinetic traces at pH* 6.8 and 3.7 are shown in Fig. 4 of the main text.

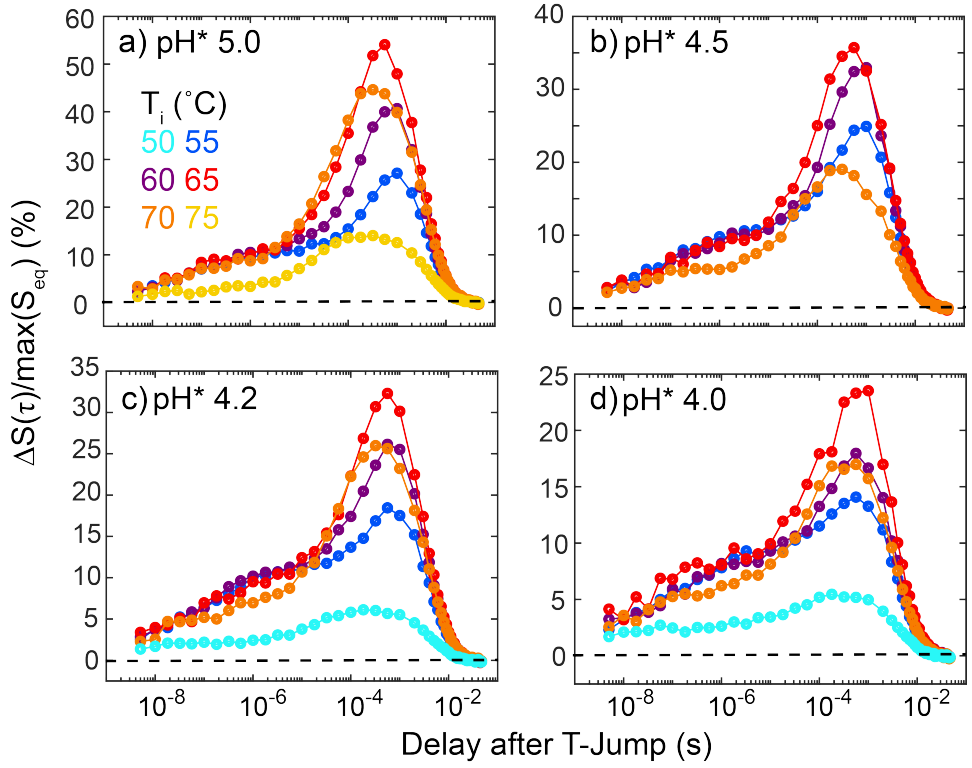


Figure S13: t-HDVE kinetic traces probed at 1670 cm^{-1} for X = caC at pH* (a) 5.0, (b) 4.5, (c) 4.2, and (d) 4.0 and each initial temperature (T_i). Kinetic traces at pH* 6.8 and 3.7 are shown in Fig. 4 of the main text.

7. pH-Dependent Stretched Exponential Relaxation Kinetics

It is apparent in Fig. 4c and f that the 10-100 μs T-jump response from both X = C and X = caC becomes increasingly non-exponential as pH* is reduced. Fig. S14 shows fits of the $\lambda_{\mu s}$ response at 1670 cm^{-1} to a stretched exponential function and the resulting stretch parameter (β), where a decrease in β represents increasingly non-exponential behavior. Fits were performed with the time constant fixed to that determined via the MEM-iLT representation (Section 8 of SI), such that only the amplitude and β were fit. Additionally, the time-domain data was truncated to omit the 10-100 ns response during fitting. For both X = caC and X = C, β drops substantially below pH* 4.5, with values reaching ~0.8 and 0.6, respectively, at pH* 3.7. Stretched exponential kinetics have previously been observed in T-jump studies of biomolecular dissociation and may result from two distinct origins:[8-9] (1) an inhomogeneous starting ensemble that dissociates with the same mechanism, but with a distribution of rates, and (2) a homogeneous ensemble that dissociates with a variety of mechanisms, leading to distribution in rates. These are two limiting cases, but it is possible that both factors may contribute here.

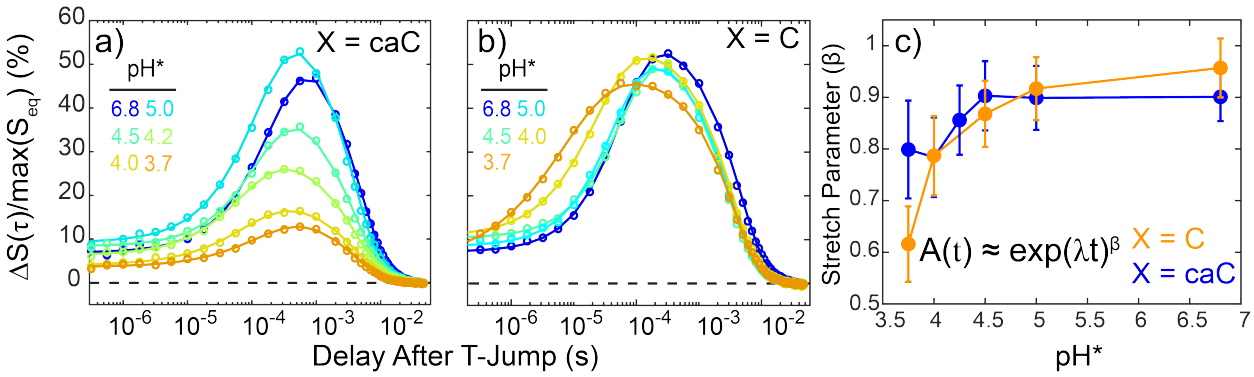


Figure S14: Time-domain traces of t-HDVE data from (a) X = caC and (b) X = C probed at 1670 cm^{-1} from pH* 6.8 to 3.7. Data shown corresponds to T-jump across the center of the duplex-to-single-strand transition. Traces were fit to a stretched exponential function (solid lines), where the stretch parameter (β) is shown in (c) as a function of pH*. β values shown were averaged across experiments for the middle three T_i conditions. Error bars indicate the 95% confidence interval from the fits and averaged over the three temperatures.

The pH*-dependent duplex/single-strand thermodynamics and kinetics may be used to interpret the observed stretched exponential kinetics. The X = caC sequence exhibits a reduction in θ_{int} and duplex-to-single-strand cooperativity (Fig. 5) as pH* is reduced. The pH*-dependence

of θ_{int} is primarily related to changes in G:C base pairing, and previous MD simulations have suggested that 5-carboxyl protonation increase local fluctuations in base pairing degrees of freedom.[10] The increased inhomogeneity in the duplex state at low pH* for caC may lead to a sum of exponential relaxation rates that appear as a single stretched response. On the other hand, the reduction in melting cooperativity (Fig. 2 and 5) suggests that dehybridization may occur in a less concerted manner, leading to a greater distribution of potential pathways, which could contribute to the observed stretched kinetics. As discussed above, the X = C cannot essentially tolerate any C protonation without dissociating, and therefore a distribution of protonated duplexes is unlikely to produce the highly non-exponential dissociation kinetics at low pH*. Instead, the stretch factor appears correlated with the large reduction in duplex-to-single-strand transition cooperativity observed at low pH*, suggesting that the less concerted transition gives rise to a distribution of transition states and timescales for dissociation.

The non-exponential relaxation kinetics suggest inhomogeneity in the duplex ensemble, transition state, or both, we have globally fit the kinetic and thermodynamic data for X = C and X = caC to a Kramers model consisting of a single duplex and transition state free energy. Incorporation of the stretched exponential relaxation results would require significant assumptions regarding its origin in the duplex state or transition state as well as additional parameters that may lead to overfitting of the thermodynamic and kinetic data. Therefore, even for the conditions that exhibit stretched kinetics, we have taken the maximum amplitude rate from the rate-domain representation of the kinetic data (Section 8 of SI) to be used as a representative λ_{obs}. This is identical to using the time constant obtained from fitting time-domain traces to a stretched exponential function (Figure S14). In general, the physical interpretation of this time constant is complex but can be roughly related to the median timescale of the process over the values of β observed for X = caC and X = C.[9] Therefore, the rates and barriers shown in Figs. 5 & 6 most closely represent their respective median value among a distribution.

8. Extraction of Observed Relaxation Rate from Temperature-Jump Data with MEM-iLT

As mentioned in the main text, we determine the observed rate(s) from t-HDVE data using a rate representation in which the data is inverse Laplace transformed into the rate domain using a

maximum entropy approach (MEM-iLT).[11] This method allows for extraction of observed rates without fitting the data to specific kinetic models such as a sum of exponentials. Rate-domain representations, or rate maps, are shown for X = caC at pH* 6.8 and 3.7 in Fig. S15a and b, respectively. Orange and purple indicate positive gain and loss in the t-HDVE data. The observed rates (λ_{obs}) shown in Fig. 6 are determined through finding the highest amplitude rate (solid line, Fig. S15c) at each frequency shown as the black dashed line in Figs. S15d and e. Then, an amplitude-weighted mean is taken across frequency as shown by the black solid lines in Figs. S15d and e, and these are the values of λ_{obs} shown in Fig. 6.

At low pH*, both X = caC and X = C exhibit increases stretched exponential behavior in the duplex-to-single-strand transition response (Fig. 8). In this case, the rate distribution becomes more asymmetric with a tail extending to larger rates (Fig. S15b and c) and the maximum amplitude rate (Max) may not best represent the rate of the process.[12] Another approach is to use the rate at which the integrated area under the duplex-to-single-strand transition is half of its total value (dashed lines Fig. S15c, Half Area). For highly symmetric rate distributions, this value should be nearly identical to the maximum amplitude value but will differ for skewed distributions. Observed rates using the Half Area method are shown in blue in Figs. S15d and e. At each pH*, the Half Area method leads to a faster observed rate, as the duplex-to-single-strand transition rate distributions exhibit a positive skew. In general, the Half Area method shifts the observed rate to higher values (Fig. S15f), but does not change the association and dissociation Arrhenius barriers determined via a two-state assumption. However, the overall shift in observed rate will slightly impact $\Delta S^{\ddagger}_{d}$ and $\Delta S^{\ddagger}_{a}$ determined via fits to a Kramers model (Fig. 6). While this approach may better capture the characteristic rate of a significantly stretched process, there are a couple concerns for application to these specific sequences. First, there is significant overlap in the rate domain between the duplex-to-single-strand transition and fast response (10^{6}-10^{8} s^{-1}), making it difficult to correctly determine the area under the duplex-to-single-strand response. Secondly, noise in the kinetic data leads to symmetric broadening of the rate distribution.[11-12] If broadening due to noise is comparable to or larger than the width of the rate distribution without noise, then the skew of the distribution will be significantly reduced. For X = caC and X = C, the signal-to-noise in the t-HDVE data is reduced at low pH*, and therefore noise broadening may make large contributions to the rate representation. Overall, these complications have led us to use the Max method for determining all λ_{obs} in this work.

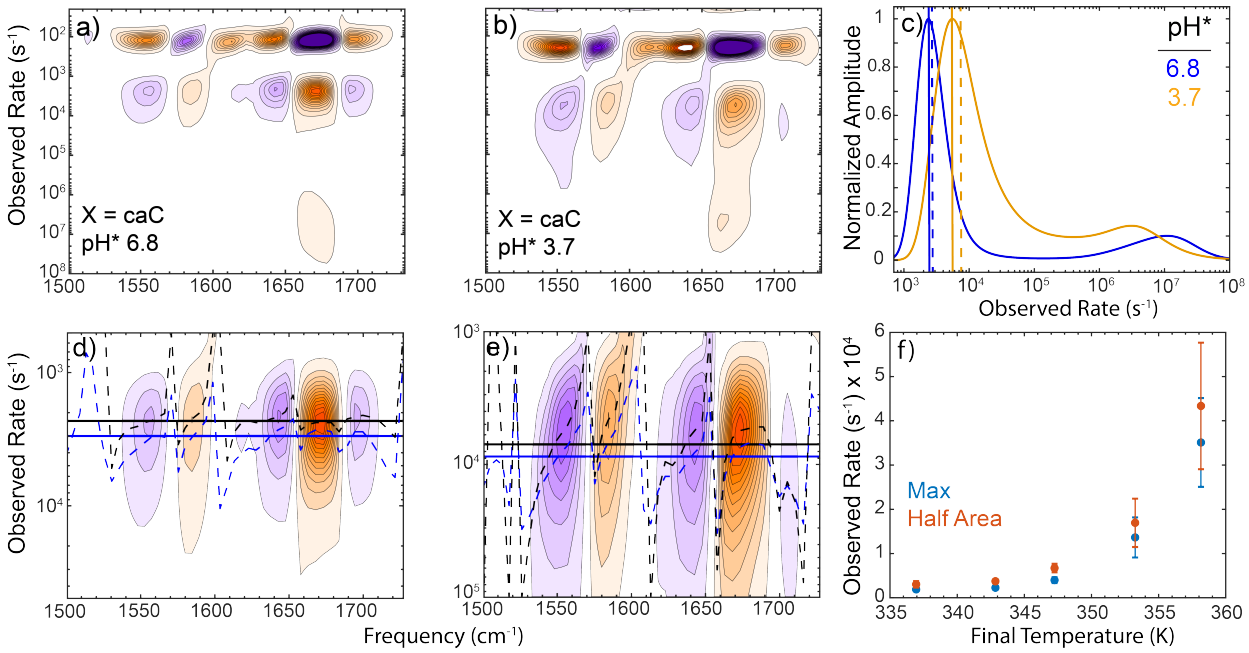


Figure S15: Rate-domain representations of t-HDVE data shown for X=caC at (a) pH* 6.8 and (b) pH* 3.7 where orange and purple indicate gain and loss of signal, respectively. (c) Slices through the rate maps at 1673 cm^{-1} for each pH* condition. Solid vertical lines represent maximum observed rate (Max) while dashed lines indicate the rate at which the integrated area for the 10^3-10^4 s^{-1} process (duplex-to-single-strand transition) is half of its full value (Half Area). Rate maps are zoomed in on the duplex-to-single-strand process for (d) pH* 6.8 and (e) pH* 3.7. Dashed lines indicate the observed rate determined at each frequency using the Max (black) and Half Area (blue) methods. Solid lines represent the amplitude-weighted mean values across frequency for each method. (f) Observed duplex-to-single-strand rate for X = caC at pH* 6.8 as a function of temperature using the Max and Half Area methods.

9. Determination of Melting Curve Normalization Amplitudes

The melting of short oligonucleotides is typically assumed to occur in an all-or-nothing fashion, where DNA strands are fully duplexed or separated.[13-14] However, X=caC exhibits the apparent loss of G:C base pairing along the low temperature baseline of the 2nd SVD component, suggesting a degree of pre-melting within the DNA duplex. To account for both the sigmoidal duplex-to-single-strand transition and the loss of base pairing at lower temperatures, we have modeled the 2nd SVD component as a total base pairing fraction ($\theta(T)$) that can be separated into internal (θ_{int}) and external (θ_{ext}) contributions.[15] The details of this model have been described

elsewhere.[16] However, it should be mentioned here that the model requires a normalization parameter (A):

$$\theta(T) = A\theta_{\text{int}}(T)\theta_{ext}(T) \quad \text{(S11)}$$

The value of A reports on the degree of base pairing character at the lowest temperature measured. In an all-or-nothing model, the melting curve would be normalized to 1 at the lowest measured temperature as the duplex state is assumed to be fully intact. However, sequences with substantial changes in internal base-pairing prior to full dissociation may not be fully base-paired at the lowest temperature sampled. Therefore, A is assumed to account for deviations in θ_{int} below 1 at 3 °C. For X = caC, the values of $1/A$ determined at each pH* were the same within error (Fig. S16f), so the mean value of 0.825 is used for fits shown in the main text. As shown in Fig. S16, almost identical thermodynamic trends are observed whether $1/A$ is fixed at 0.825 or allowed to float at each pH*. For X = C, N3 protonation is modeled to completely dissociate the duplex, and therefore θ_{int} is expected to be independent of pH*. The 2nd SVD component at pH* 6.8 was fit to determine θ_{int} and A, and these values were applied for all pH* points (Fig. S17).

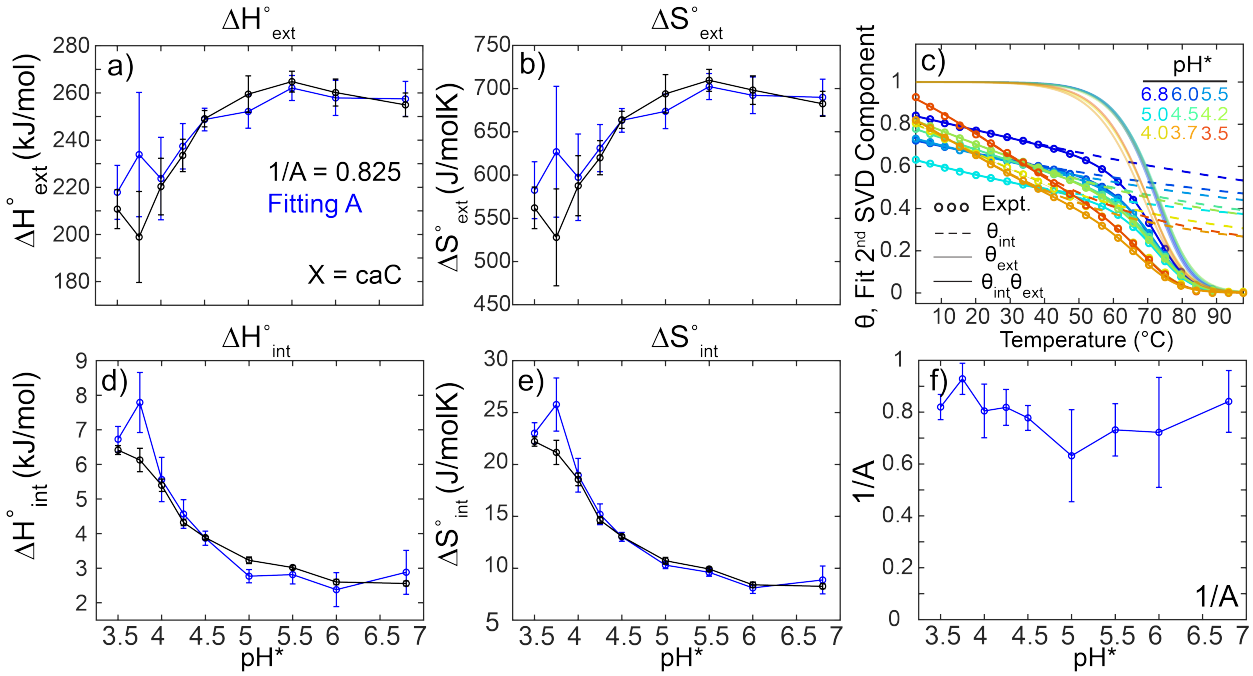


Figure S16: (a) Values of $\Delta H°_{ext}$ determined from fits of thermal 2nd SVD components of X = caC to the thermodynamic model described above. Values are shown when fitting with a fixed normalization amplitude (A) of 0.825 (black) and including A as an additional fit parameter for each pH*. (b) Corresponding plot for $\Delta S°_{ext}$. (c) Fit 2nd SVD components scaled by their respective

value of 1/*A*. Extracted temperature profiles of θ_{int} (dashed line) and θ_{ext} (transparent solid) for each pH* are shown. (d) Fit values of $\Delta H°_{int}$. (e) Fit values of $\Delta S°_{int}$. (f) Fit values of 1/*A*. Error bars in each plot represent the 95% confidence interval from the fit.

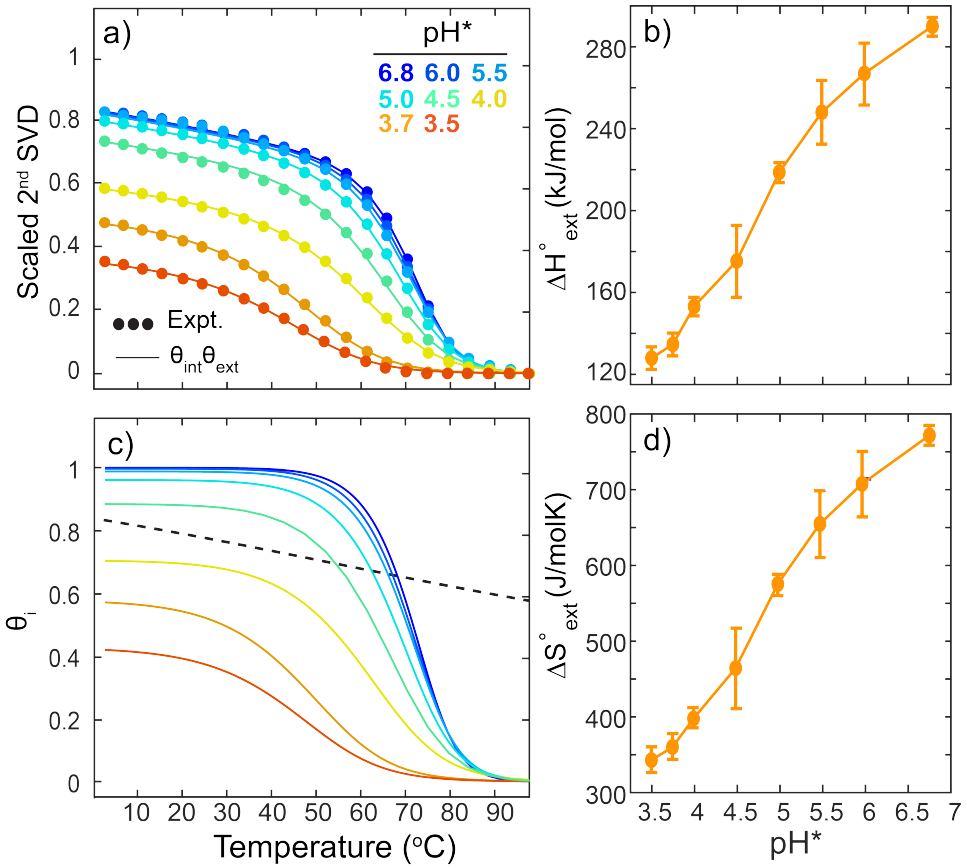


Figure S17: (a) Scaled melting 2nd SVD components of X = C from pH* 6.8 and 3.5 fit to the thermodynamic model described above. (b) Values of $\Delta H°_{ext}$ extracted from fit. (c) Extracted profile of θ_{ext} (solid transparent) for each pH*. The black dashed line is the pH*-independent θ_{int} profile for X = C applied in fitting. (d) Values of $\Delta S°_{ext}$ extracted from fit. Error bars in each plot represent the 95% confidence interval from the fit.

10. Experimental Correlation Between θ_{int} and λ_{ns}

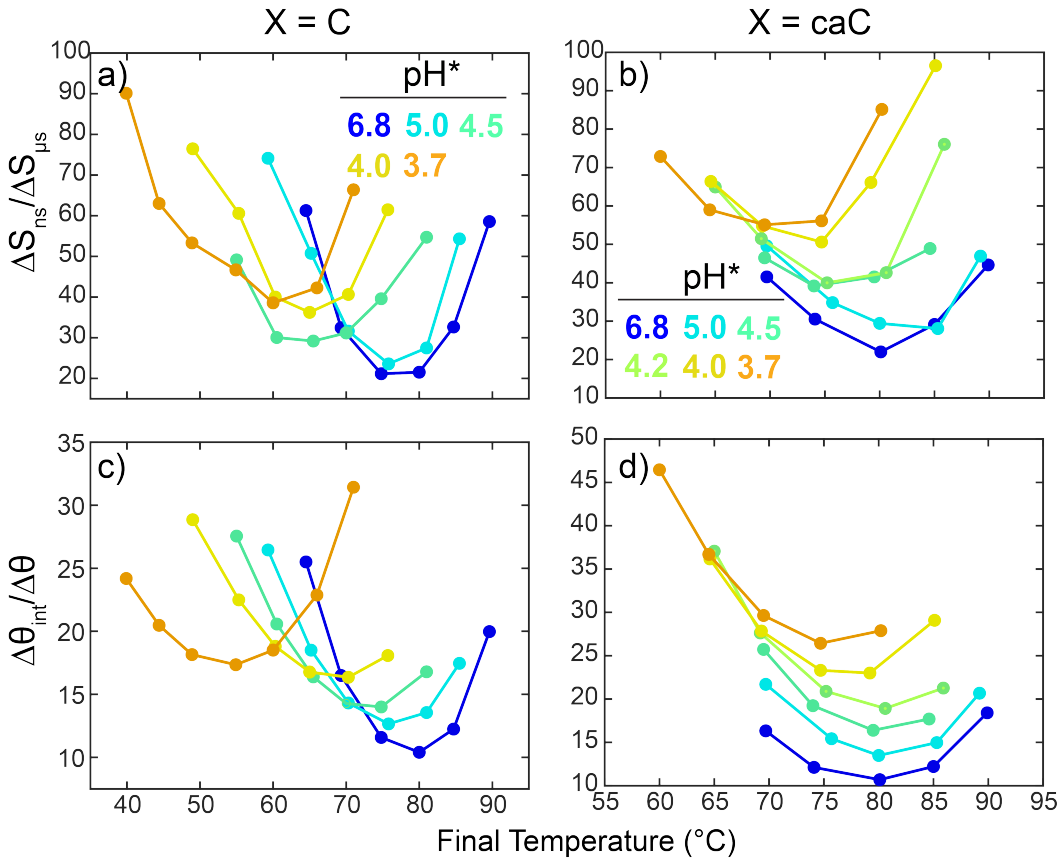


Figure S18: (a) Ratio of the integrated t-HDVE spectrum at a T-Jump delay of 1 μs and that of the maximum signal for X = C as a function of pH*. (c) Ratio of the change in θ_{int} over the total melting curve change ($\Delta\theta$) for the temperature interval in (a) for X = C. (b,d) Corresponding plots for X = caC from pH* 6.8 to 3.7.

As discussed in SI section 5, λ_{ns} is assigned to the loss of internal duplex base pairing, which for the sequences studied here primarily corresponds to fraying of the A:T termini and loosening of G:C base pairs upon 5-carboxyl protonation. Since λ_{ns} is related to changes in internal duplex base pairing, it should be related to θ_{int} determined from thermodynamic fits of the FTIR melting curves (Fig. 5). Fig. S18a,b shows the ratio ($\Delta S_{ns}/\Delta S_{\mu s}$) of the integrated t-HDVE spectrum at a T-Jump delay of 1 μs and that at the delay corresponding to maximum response for X = C and X = caC. $\Delta S_{ns}/\Delta S_{\mu s}$ reports on the percentage of the total duplex-to-single-strand response that comes from λ_{ns}. For both X = C and X = caC, $\Delta S_{ns}/\Delta S_{\mu s}$ follows a parabolic-like form as temperature increase across the melting transition. The initial decrease and then increase in $\Delta S_{ns}/\Delta S_{\mu s}$ arise from both changes in ΔS_{ns} and $\Delta S_{\mu s}$ as a function of temperature. For both X = C

and X = caC, the minimum of the parabola-like form of $\Delta S_{ns}/\Delta S_{\mu s}$ becomes larger in amplitude and broader across temperature as pH* is reduced. A comparison with θ_{int} from the thermal melting curves reveals similar trends. Fig. S18c,d shows the ratio ($\Delta\theta_{int}/\Delta\theta$) of the change in internal base pairing fraction relative to the change in overall base pairing fraction over the temperature intervals used for T-Jump measurements. $\Delta\theta_{int}/\Delta\theta$ reports on the percentage of overall base pairing reduction in a given temperature interval due to loss of internal base pairing. Interestingly, $\Delta\theta_{int}/\Delta\theta$ shows similar trends to $\Delta S_{ns}/\Delta S_{\mu s}$ across temperature and pH* for X = C and X = caC. For X = C, $\Delta\theta_{int}$ remains constant with pH*, but $\Delta\theta$ exhibits a broader melting transition that is reflected in its temperature-dependence as well as that of $\Delta S_{\mu s}$. In contrast, X = caC exhibits pH*-dependent changes from internal and external base pairing contributions. These combined changes lead to a larger shift in the minimum of $\Delta\theta_{int}/\Delta\theta$ across temperature from pH* 6.8 to 3.7 as compared to X = C, which is consistent with shift in $\Delta S_{ns}/\Delta S_{\mu s}$. Overall, the similarity between $\Delta\theta_{int}/\Delta\theta$ and $\Delta S_{ns}/\Delta S_{\mu s}$ for X = C and X = caC further supports our assignment of λ_{ns} and its connection to melting thermodynamics.

11. Application of Arrhenius Analysis to Temperature-Jump Data

In addition to the Kramers model in eq. 3 and 4, the temperature dependence of $\lambda_{\mu s}$ may also be described by a standard Arrhenius equation.

(S12) $$k = A\exp\left(-\frac{E}{RT}\right)$$

First, $\lambda_{\mu s}$ is decomposed into dissociation (k_d) and association (k_a) rate constants using eq. 2, and then both k_d and k_a can be fit to eq. S12 to determine the apparent activation energy (E) and Arrhenius pre-factor (A). For both X = C and X = caC, the pH*-dependence of dissociation barrier (E_d) and association barrier (E_a) (Fig. S19, S20, S21) resemble $\Delta H^{\ddagger}_d$ and $\Delta H^{\ddagger}_a$ in Fig. 6, respectively, while the dissociation pre-factor (A_d) and association pre-factor (A_a) trend similarly to $\Delta S^{\ddagger}_d$ and $\Delta S^{\ddagger}_a$.

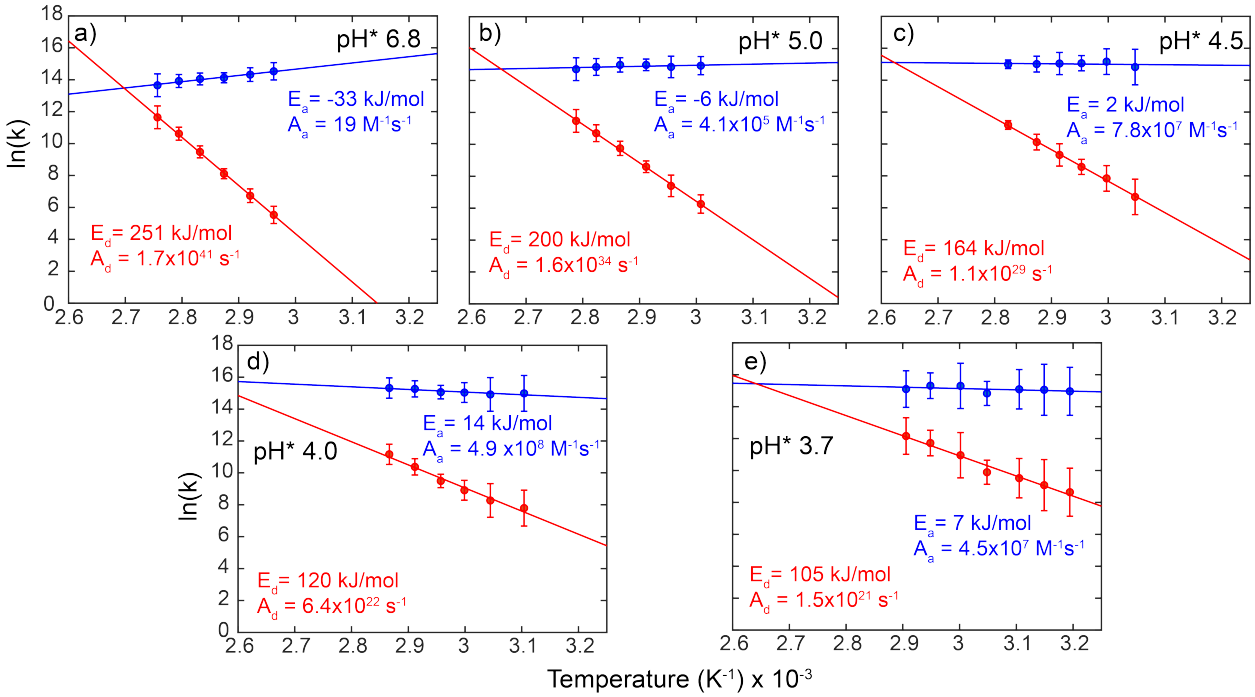


Figure S19: Arrhenius plots of dissociation (red) and association (blue) rates determined from T-Jump experiments for X = C from pH* 6.8 to 3.7. Error bars indicate standard deviation in maximum rate taken from rate maps across frequency.

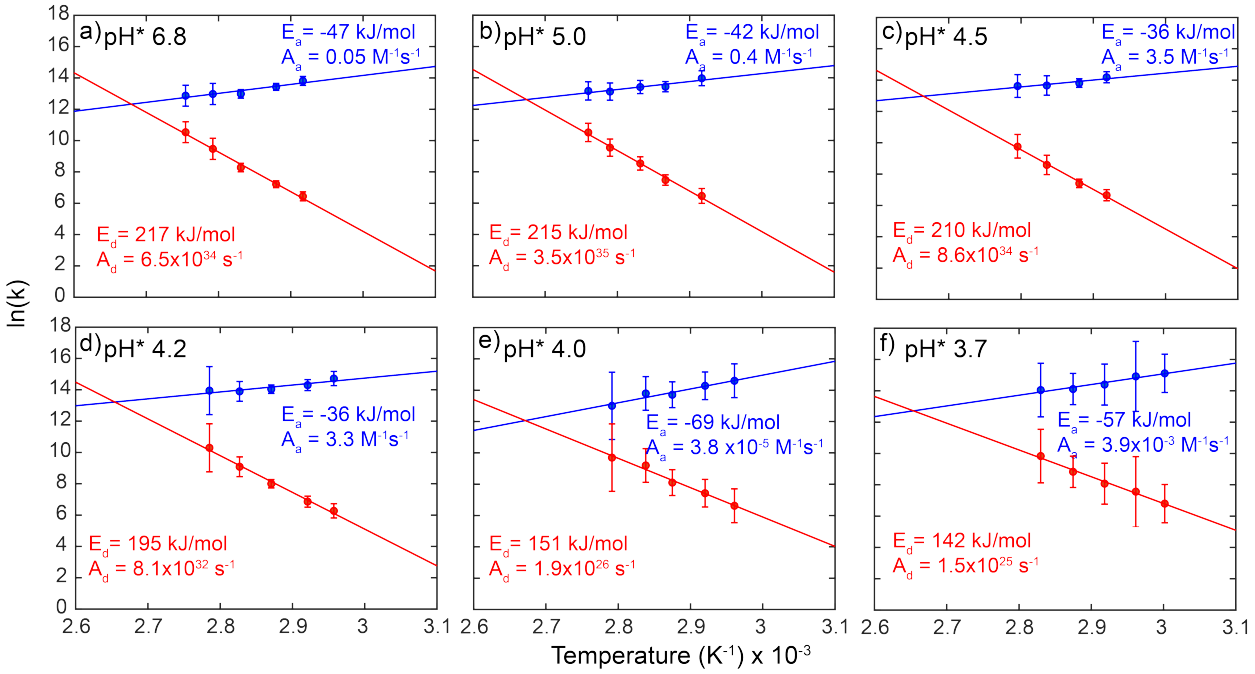


Figure S20: Arrhenius plots of dissociation (red) and association (blue) rates determined from T-Jump experiments for X = caC from pH* 6.8 to 3.7. Error bars indicate standard deviation in maximum rate taken from rate maps across frequency.

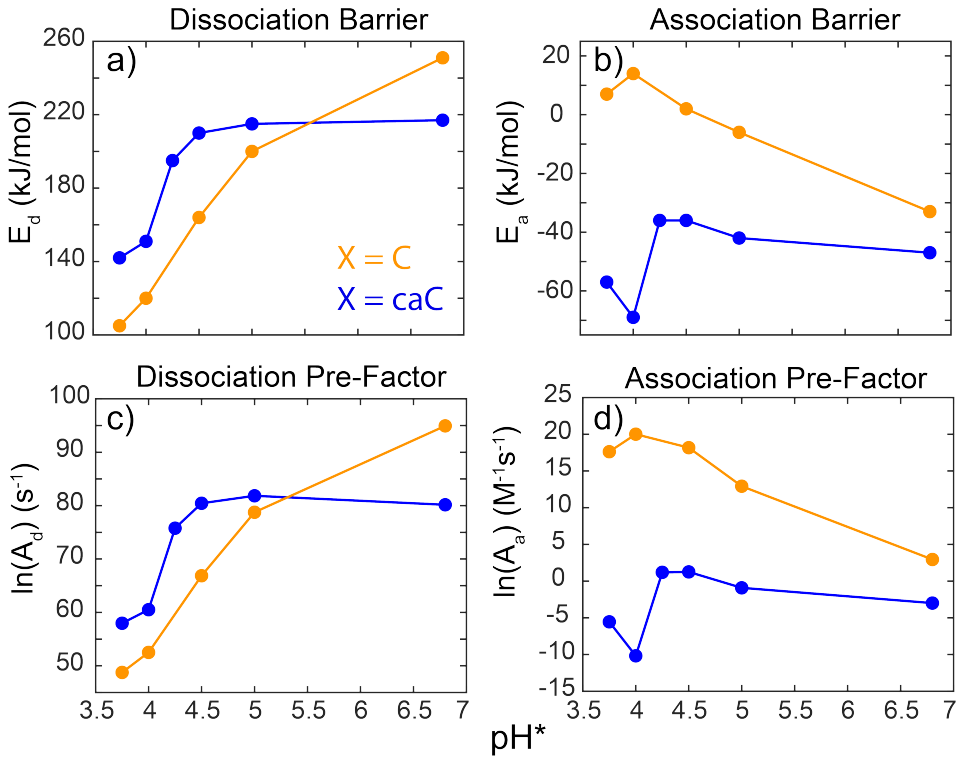


Figure S21: Trends in (a) dissociation barrier (E_d), (b) association barrier (E_a), (c) dissociation pre-factor (A_d), and (d) association pre-factor (A_a) as a function of pH* for X = C (black) and X = caC (blue).

12. Supporting References

1. Krummel, A. T.; Zanni, M. T., DNA vibrational coupling revealed with two-dimensional infrared spectroscopy: insight into why vibrational spectroscopy is sensitive to DNA structure. *J. Phys. Chem. B* **2006,** *110*, 13991-14000.
2. Peng, C. S. Two-dimensional infrared spectroscopy of nucleic acids: application to tautomerism and DNA aptamer unfolding dynamics. Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2014.
3. Lee, C.; Cho, M., Vibrational dynamics of DNA. II. Deuterium exchange effects and simulated IR absorption spectra. *J. Chem. Phys.* **2006,** *125*, 114509.
4. Lee, C.; Cho, M., Vibrational dynamics of DNA: IV. Vibrational spectroscopic characteristics of A-, B-, and Z-form DNA's. *J. Chem. Phys.* **2007,** *126*, 04B607.
5. Jovin, T. M.; Soumpasis, D. M.; McIntosh, L. P., The transition between b-DNA and z-DNA. *Annu. Rev. Phys. Chem.* **1987,** *38*, 521-558.
6. Dai, Q.; Sanstead, P. J.; Peng, C. S.; Han, D.; He, C.; Tokmakoff, A., Weakened N3 hydrogen bonding by 5-formylcytosine and 5-carboxylcytosine reduces their base-pairing stability. *ACS Chem. Biol.* **2015,** *11*, 470-477.
7. Teng, X.; Hwang, W., Effect of methylation on local mechanics and hydration structure of DNA. *Biophys. J.* **2018,** *114*, 1791-1803.

8. Sabelko, J.; Ervin, J.; Gruebele, M., Observation of strange kinetics in protein folding. *Proc. Natl. Acad. Sci.* **1999,** *96*, 6031-6036.
9. Johnston, D., Stretched exponential relaxation arising from a continuous sum of exponential decays. *Phys. Rev. B* **2006,** *74*, 184430.
10. Ngo, T. T.; Yoo, J.; Dai, Q.; Zhang, Q.; He, C.; Aksimentiev, A.; Ha, T., Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* **2016,** *7*, 10813.
11. Kumar, A. T.; Zhu, L.; Christian, J.; Demidov, A. A.; Champion, P. M., 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* **2001,** *105*, 7847-7856.
12. Sanstead, P. J. C., Investigation of DNA Dehybridization through Steady-State and Transient Temperature-Jump Nonlinear Infrared Spectroscopy. Ph. D. Dissertation, University of Chicago, Chicago, IL, **2018**.
13. Owczarzy, R., Melting temperatures of nucleic acids: discrepancies in analysis. *Biophys. Chem.* **2005,** *117*, 207-215.
14. Mergny, J.-L.; Lacroix, L., Analysis of thermal melting curves. *Oligonucleotides* **2003,** *13*, 515-537.
15. Wartell, R. M.; Benight, A. S., Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Phys. Rep.* **1985,** *126*, 67-107.
16. Sanstead, P. J.; Ashwood, B.; Dai, Q.; He, C.; Tokmakoff, A., Oxidzed Derivatives of 5-Methylcytosine Alter the Stability and Dehybridization Dynamics of Duplex DNA. *Submitted*.