8796–8807 *Nucleic Acids Research, 2020, Vol. 48, No. 15* *Published online 11 July 2020*
doi: 10.1093/nar/gkaa589

Impact of 5-formylcytosine on the melting kinetics of DNA by ^{1}H NMR chemical exchange

Romeo C. A. Dubini [1,2], Alexander Schön[1], Markus Müller [1], Thomas Carell [1] and Petra Rovó [1,2,*]

[1]Faculty of Chemistry and Pharmacy, Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstraße 5-13, 81377 Munich, Germany and [2]Center for Nanoscience (CeNS), Faculty of Physics, Ludwig-Maximilians-Universität München, Schellingstraße 4, 80799 Munich, Germany

Received May 17, 2020; Revised June 24, 2020; Editorial Decision June 28, 2020; Accepted July 02, 2020

ABSTRACT

5-Formylcytosine (5fC) is a chemically edited, naturally occurring nucleobase which appears in the context of modified DNA strands. The understanding of the impact of 5fC on dsDNA physical properties is to date limited. In this work, we applied temperature-dependent ^{1}H Chemical Exchange Saturation Transfer (CEST) NMR experiments to non-invasively and site-specifically measure the thermodynamic and kinetic influence of formylated cytosine nucleobase on the melting process involving dsDNA. Incorporation of 5fC within symmetrically positioned CpG sites destabilizes the whole dsDNA structure—as witnessed from the ∼2°C decrease in the melting temperature and 5–10 kJ mol^{-1} decrease in $\Delta G°$—and affects the kinetic rates of association and dissociation. We observed an up to ∼5-fold enhancement of the dsDNA dissociation and an up to ∼3-fold reduction in ssDNA association rate constants, over multiple temperatures and for several proton reporters. Eyring and van't Hoff analysis proved that the destabilization is not localized, instead all base-pairs are affected and the transition states resembles the single-stranded conformation. These results advance our knowledge about the role of 5fC as a semi-permanent epigenetic modification and assist in the understanding of its interactions with reader proteins.

INTRODUCTION

Nucleic acids play a fundamental role within all three domains of life, however their 'alphabet' is intrinsically restricted to only four letters: G, C, A and T (U for RNA). In order to expand this narrow ensemble of constituents, an additional instructional layer can be integrated in the genome through chemically editing the four canonical bases, whose properties on the microscopic scale (base stacking, hydrogen bonding, solvation properties, stabilizing cross-strand terms) lie at the heart of DNA and RNA nature, structure, function and consequent biological role (1–3). Such alterations, broadly termed 'epigenetic modifications', have been extensively reviewed (4–7). Naturally occurring chemically edited bases are found in DNA and RNA of all kingdoms of life and play a crucial regulatory role in both development (8–12) and pathogenesis (13–17) of a given organism. Their presence influences the physicochemical properties of the resulting nucleic acid molecule, thus reshaping the biological role of the unmodified analogs (18). Within the spectrum of epigenetic modifications in DNA, the ones involving cytosine have been studied to the largest extent, primarily due to their (i) remarkable abundance, (ii) ubiquity across all domains of life and (iii) vast range of biological roles. The most frequent chemical modification of cytosine is 5-methylcytosine (5mC), which predominantly occurs within symmetric CpG sites and is associated with stable transcriptional repression, cell differentiation, X-chromosome inactivation and imprinting (19–22). 5mC can also be further modified by Ten-Eleven Translocation (TET) enzymes, which are involved in the context of the active demethylation pathway and subsequent base excision repair (BER) cycle (23,24). The step-wise oxidation of 5mC process leads to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) in a sequential manner. In addition, protein-DNA binding structural studies have revealed that TET preferentially oxidizes 5mC in symmetric CpG sites over non-CpG sites (25). Despite being orders of magnitude less abundant than 5mC or 5hmC, 5fC is equally broadly distributed in genomic DNA (5,26). Since its discovery, 5fC has been confirmed to (i) be a genomically stable and semi-permanent modification (27,28), (ii) vary its concentration depending on the nature, health condition and developmental state of the tissue (29) and (iii) be recognized by several reader proteins with exquisite specificity over an overwhelming excess of background represented by C, 5mC, and 5hmC (30,31).

*To whom correspondence should be addressed. Tel: +49 89 2180 77652; Email: petra.rovo@lmu.de

Taken together, those evidences indicate that 5fC retains an undefined number of uncharacterized epigenetic roles, in addition to being a key intermediate in the active demethylation pathway (32).

In order to fully appreciate the nature and the extent of the various functions that 5fC performs, the properties of formylated CpG sites have been subjected to extensive investigations in the past decade. Whether 5fC is responsible for the alteration of the canonical B-DNA secondary structure is still not a firmly established notion, and contradictory conclusions have been reported (33–36). In contrast, there is reasonable consensus in accounting for the enhanced degree of dsDNA flexibility as a result of 5fC incorporation that has been attributed to an increase in the base pair wobbling around its canonical Watson–Crick, base-paired conformation (18,37,38). It has been proposed that the destabilization is due to the electron withdrawing nature of the formyl substituent, which would in turn weaken the hydrogen-bond strength between the modified base and the counter-stranded paired guanine (38,39). Recent work by Sanstead *et al.* has found that 5fC accounted for a reduction in cooperativity of duplex melting and a lowering of the dissociation barrier, providing a global characterization of heavily modified oligomers via T-jump IR spectroscopy (40). On the other hand, NMR measurement on water-exchangeable protons provided information about the extent of additional base-pair opening caused by 5fC incorporation and other modified cytosine bases in the presence of ammonia acting as a catalyst for the water-exchange process (34).

Despite the above contributions, a number of questions still remain unanswered. In particular, it is entirely unknown how 5fC affects the processes of DNA hybridization. Then, the site-specific energetics of the destabilization are unclear, i.e. whether the activation barrier between ds- to ssDNA is reduced due to a stabilization of the transition state, or due to a destabilization of the ground state. How does the transition state look like? Does it resemble the ds- or ssDNA structure? Lastly, very little is known about the range of the enhanced flexibility, i.e. how locally 5fC destabilizes dsDNA.

Characterizing the subtle differences, whether of structural or dynamical origin, of chemically modified nucleic acids molecules can be a demanding task, especially if site-specific information is needed. Solution-state NMR represents an attractive option: it does not require any bulky physical label which could confound the significance of the small chemical modification, and it allows accessing atomic resolution information on local dynamics on a wide range of time scales. NMR spin relaxation techniques have been shown to allow for such studies on a plethora of biomolecules. Out of the several time domains that can be accessed, the biologically relevant μs-ms frame is amenable to investigation employing $R_{1\rho}$ and CPMG Relaxation Dispersion (RD) techniques (41–48). More recently another experiment, the Chemical Exchange Saturation Transfer (CEST), has been proposed (49), which allows for the elucidation of even slower events, occurring on the ms–s time scale. The essential working principle of the CEST method, whose detailed theoretical framework has been recently outlined (50,51), relies on the presence of a major, 'ground state' (GS) conformer of a target biomolecule that undergoes an exchange process with one or more distinct and sparsely populated 'excited state' (ES) conformer(s) on the ms time scale. Experimentally, in a CEST experiment the intensity of a signal of interest is recorded as a function of the resonance offset frequency of a weakly saturating field, which is stepped through a certain interval across the region of interest. The characterization of such transiently and sparsely populated exchanging states can become invaluable in the attempt of describing the motions in both proteins and nucleic acids, yielding information on the relative populations of the involved GS and ES, the precise time frame of the exchange process, as well as an estimation of the ES chemical shift, which can be indirectly informative on the structure of the sporadic conformation (52–54).

In this study, we apply a recently described ^{1}H-only CEST experiment to gain insights into the exchange kinetics and thermodynamics of several proton reporters simultaneously. In this endeavour, we exploit a subset of proton probes taking advantage of the SELectively Optimized Proton Excitation (SELOPE) scheme (55), which allows for detailed NMR investigations of non-isotopically labeled samples.

To further exploit the wealth of information provided by CEST, we present an atom-specific description of the influence of 5fC base-pair thermodynamics and kinetics of both the melting and the annealing processes, which are mechanisms of primary importance given the involvement of unwound DNA in transcription, repair, and subsequent recombination. With this goal in mind, we introduce a novel methodological framework that permits the extraction of thermodynamic (ΔG°, ΔH°, ΔS°, T_m) and kinetic parameters ($\Delta G^\ddagger$, $\Delta H^\ddagger$, $\Delta S^\ddagger$) from the data provided by the ^{1}H CEST experiments.

Our results corroborate the notion that 5fC does not perturb the canonical B-DNA conformation of dsDNA, but it does curtail the overall stability of the double-stranded structure. More specifically, it impacts the kinetic rates of both the melting and annealing processes. Combining these effects, 5fC shifts the dissociation equilibrium constant up to 9-fold towards the single-stranded conformer. By further dissection of the kinetic rates through Eyring formalism, we identified that the origin of the destabilization can be traced back to the ground state of each nucleobase involved in the dsDNA conformation, evidencing the wide range of the 5fC-induced features. In addition, we find the transition state bridging the single and double-stranded structures closely resembles the ssDNA conformation, consistent with the consensus nucleation mechanism for DNA hybridization (3).

Taken together, the above observations indicate that the repercussions arising from the inclusion of 5fC are only appreciable in a dynamic context and aid in the unravelling of its biological role. In our analysis, we provide a unified view of the processes of dissociation and association, resolving the energetics behind both processes with atomic resolution.

MATERIALS AND METHODS

Sample preparation

The canonical dsDNA sample Can_{12} was purchased after purification via high-performance liquid chromatography

(HPLC) and desalting procedure from Ella Biotech GmbH (Martinsried, Germany).

The fdC-phosphoramidite (fdC-PA) and subsequently the modified dsDNA sample Mod_{12} was prepared via phosphoramidite chemistry as previously reported (56). Solid phase synthesis of oligonucleotides containing fdC was performed on an ABI 394 DNA/RNA synthesizer (Applied Biosystems) using standard DNA synthesis conditions with a cartridge scale of 1 μmol. The phosphoramidites Bz-dA, Ac-dC, iBu-dG and dT as well as the PS carriers were purchased from LinkTechnologies. For the reaction of the fdC-PA a coupling time of 180 s was applied. The terminal DMT protecting group was cleaved after DNA synthesis on the synthesizer. Basic and acidic deprotection of all oligonucleotides was performed according to literature (56). Purification of the oligonucleotides was achieved with a HPLC system (Agilent 1260 Infinity II 400 bar pump and a Agilent 1260 Infinity II VWD detecting at 260 nm) applying a buffer system of 0.1 M triethylammonium acetate in water (buffer A) and 0.1 M triethylammonium acetate in 80 % aqueous MeCN (buffer B), a gradient of 0–30% buffer B in 45 min and a flow rate of 5.0 ml/min. As stationary phase Nucleodur columns (250/10 mm, C18ec, 5 μm) from Macherey-Nagel were used. Purified oligonucleotides were analyzed by MALDI-TOF (Bruker Autoflex II). Quantification of oligonucleotides was performed via UV/Vis spectroscopy with a NanoDrop ND-1000 spectrophotometer at 260 nm.

Both samples were dissolved in an aqueous buffer consisting of 15 mM Na_2HPO_4/NaH_2PO_4 (pH 7.0), 25 mM NaCl in H_2O. Annealing was performed by heating the dsDNA-containing buffer solution to 90°C for 5 min and slowly cooling it to 5°C in ∼90 min, after which it was allowed to return to room temperature. Then, the NMR sample was prepared with the addition of 0.02% NaN_3, 25 μM DSS and 5% D_2O, resulting in final sample concentrations of 1.00 and 1.58 mM for Can_{12} and Mod_{12}, respectively, as determined via UV spectrophotometric measurements at 260 nm using the extinction coefficient calculated via the nearest neighbor approximation.

UV/Vis spectroscopy

UV/Vis melting profiles of the oligonucleotides were measured at 260 nm with a JASCO V-650 UV/Vis spectrophotometer between 20 and 85°C (scanning rate of 1°C/min), and each sample was measured four times. Samples were placed into 100 μl cuvettes and diluted with the same Na_2HPO_4/NaH_2PO_4, NaCl aqueous buffer as used in the NMR experiment. Before each measurement, a layer of mineral oil was placed on the surface of the sample in order to prevent water evaporation. Mod_{12} was measured at six concentrations (1.23, 1.53, 2.45, 3.72, 6.14 and 12.27 μM), Can_{12} at four (1.17, 1.76, 3.52 and 7.04 μM). All concentration values yielded absorption values within the linear range of the spectrometer.

NMR spectroscopy

All experiments were performed on Bruker Avance III spectrometer operating at a ^{1}H Larmor frequency of 800 MHz (corresponding to a magnetic field of 18.8 T) equipped with a 5 mm triple-resonance cryogenically cooled probe. Standard 2D NOESY (mixing time of 150–250 ms, shown in Supplementary Figure S1D and E) spectra were recorded at 37°C for resonance assignment. Site-selective spin relaxation measurements were performed following the SELOPE scheme; these included ^{1}H CEST, on-resonance ^{1}H $R_{1\rho}$, recorded with a spin-lock strength of 4 kHz, and ^{1}H R_1 experiments at temperatures between 37 and 62°C. The employed $R_{1\rho}$ and ^{1}H CEST pulse sequences have been modified from Schlagnitweit *et al.* (55) in order to allow for data acquisition in an pseudo-3D interleaved fashion, while an in-house pseudo-2D ^{1}H R_1 pulse sequence (displayed in Supplementary Figure S5) with selective excitation and water suppression was used to record ^{1}H inversion recovery. ^{1}H $R_{1\rho}$ relaxation decays were collected with time delays of 20, 30, 40, 60, 80, 100 and 120 ms, while the inversion recovery profile of the ^{1}H R_1 measurement was collected with time delays of 0.5, 1, 5, 10, 25, 50, 75, 100, 200, 500, 750 and 1000 ms. All ^{1}H CEST profiles were recorded with a spin lock B_1 field of 25 Hz using a T_{EX} of 0.4 s, additional measurement conditions for each temperature are summarized in Supplementary Table S1. All NMR experiments were recorded with a relaxation delay (d_1) of 1 s, with the exception of R_1 measurements for which d_1 was set to 10 s. When not at the water resonance to allow for water suppression, the ^{1}H carrier was positioned in the middle of the H8/H6/H2 region allowing for selective irradiation of the desired aromatic sites (Figure 1). The CEST measurements were performed at 37, 45, 53, 57, 58, 59, 60, 61, 62°C for Can_{12} and at 37, 45, 53, 55, 56, 57, 58, 59, 60°C for Mod_{12}. In total, eleven sites for Can_{12} and twelve for Mod_{12} (namely for C2H6, G3H8, A4H2, A4H8, T5H6, C6H6, A8H2, A8H8, T9H6, C10H6, and C12H6 for Can_{12} and G1H8, C2H6, A4H2, A4H8, T5H6, C6H6, C6H7, A8H2, A8H8, T9H6, C10H6, and C12H6 for Mod_{12}) could be uniquely picked and assigned in the CEST spectra. Reconstruction of pseudo-3D interleaved spectra was achieved through an in-house Perl script. Both assignment and spin relaxation data was processed and inspected using TopSpin 3 or 4. Calibration of the DSS signal at 0 ppm allowed for exact spectral referencing at all temperatures. Subsequent ^{1}H resonance assignments and raw relaxation data extraction (including peak height and signal to noise) were carried out in Sparky (57).

RESULTS

We studied the following 12mer, homo-formylated palindromic sequence: 5′-GCGAT-X-GATCGC-3′, where X = 5fC in the modified sample, Mod_{12} and unedited C in the canonical sample, Can_{12} (Figure 1A). The selection was made taking into consideration that (i) it contains a centrally positioned CpG domain, which is the naturally occurring settings of 5fC and (ii) being a self-complementary sequence, the resulting ^{1}H-based NMR spectra are less crowded, and thus less prone to be affected by overlapping signals. The latter feature can be ascertained by inspecting the 2D SELOPE spectrum of Mod_{12} (Figure 1B), where each signal arising from the H6 of C and T, and H8, H2 protons of A in Mod_{12} is resolved. Guanine H8 protons

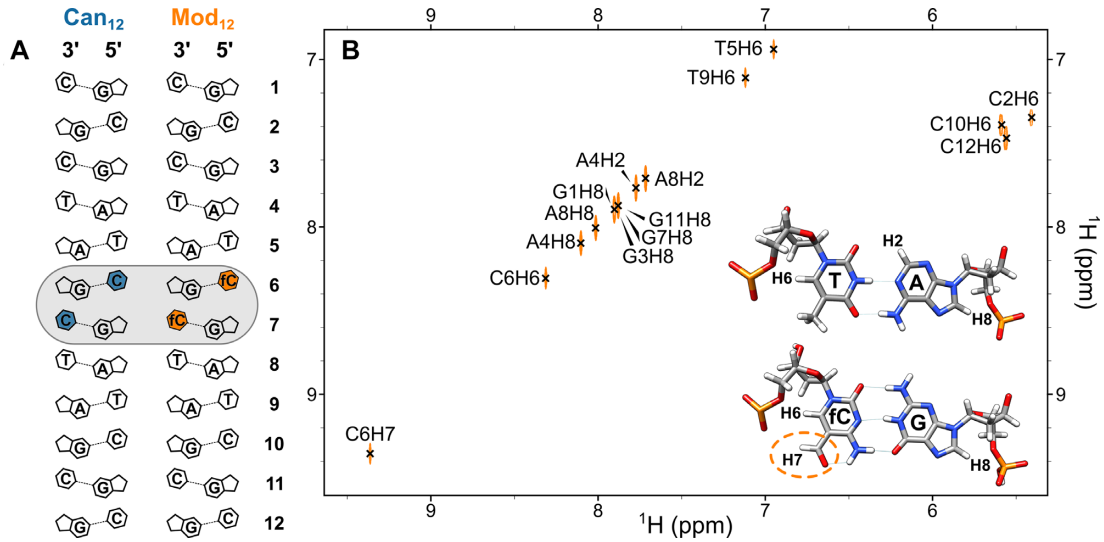

Figure 1. (**A**) DNA constructs of Can_{12} and Mod_{12}, gray box highlights the site of the 5fC incorporation. (**B**) SELOPE 2D ^{1}H–^{1}H correlation spectrum of the Mod_{12} aromatic region. H6 protons of canonical cytosine bases appear as crosspeaks due to their ^{3}J coupling to H5, while the H6 proton of the formylated cytosine C6, lacking the coupling partner H5, appears at the diagonal. The formyl proton C6H7 appears at 9.4 ppm.

represent an exception as they almost completely overlap for G3, G7 and G11 H8 in Mod_{12} in CEST spectra, thus no site specific analysis was possible for these sites.

Structural impact

A set of NOESY and natural abundance ^{1}H–^{13}C HSQC spectra were recorded to weigh on the influence of the 5fC modification on the ground-state conformation of DNA. The very weak intensity of the H2'–H4' cross-peaks in NOESY spectra (indicative of the C2'-endo sugar puckering) and dihedral angles obtained from the $^{3}J^{1}H$–^{1}H couplings of the sugar moieties confirmed the predominant presence of B-DNA conformation in both samples (58).

The chemical shift values for all ^{1}H and ^{13}C nuclei are highly comparable between the two samples, as expected, the only differences occur for sites nearby the modified base. Most notably, the resonance frequency of both amino protons of the formylated base shift markedly downfield from 6.5 and 8.2 to 7.9 and 9.1 ppm, respectively indicating the formation of a strong hydrogen bond between the formyl group and the amino protons, as per the predicted 5fC:G base-pair in Figure 1 B). This intramolecular hydrogen bond is stable even at high temperatures, only marginally weakening (0.08 ppm upfield shift) when the sample approaches the melting temperature. Besides, the 0.4 ppm upfield shift the G7 imino proton resonance of the modified sample with respect to the same proton of Can_{12} (Supplementary Figure S1A–C) suggests a reduced H-bonding and thus weakened base-pairing capacity in the 5fC:G with respect to C:G pair as observed by others (34,38).

Aldehydes in aqueous solution can undergo an acid/base catalyzed hydration reactions (59). Contrary to previous reports on different dsDNA sequences featuring 5fC in the context of X:C mismatches (where X is either 5mC, 5hmC, 5fC or 5caC) (60), we do not detect resonances compatible with such a hydrate form neither in ^{1}H nor in ^{13}C based experiments.

In accordance with previous studies, we did not identify any difference in the ground-state conformation of Can_{12} and Mod_{12} (34–36). The two samples displayed localized and discrete chemical shifts in both the ^{1}H and ^{13}C dimensions due to the presence of the formyl substituent on the edited cytosine (C6) ring and the nearest neighbouring bases T5 and G7 (Supplementary Figure S1). Considering the affinities between the spectral features of the two samples, we conclude that 5fC is well tolerated in the context of B-DNA conformation.

Concentration dependent UV/Vis melting studies

As an initial assessment of the effect of 5fC on the stability of the dsDNA conformer, we determined the concentration-dependent melting temperature (T_m) of the two samples via UV/Vis spectroscopy. Since DNA annealing is a bimolecular process, the melting temperature of the double-stranded conformation is highly concentration dependent, causing its T_m to steadily increase with the molarity of the sample (61). Such relationship can be exploited to obtain an overall estimation about the melting thermodynamics (ΔG°, ΔH°, ΔS°) and to extrapolate to the expected melting temperature to higher concentrations; in our case to the concentration range of the NMR samples which is approximately two- to three-orders-of-magnitude higher than that of the UV/Vis samples.

Hence, we recorded the melting profiles of both samples at multiple concentrations in the 1–10 μM range, and extracted the thermodynamic parameters assuming a linear relationship between $\ln(C_t)$ and T_m, where C_t is the total monomer concentration (Supplementary Figure 2, S2 and S3). This method assumes that ΔH° is temperature independent i.e. the specific heat capacity (C_p) is constant within the temperature range under consideration. In Table 1, we compare the fitted bulk thermodynamic parameters with the site-specific thermodynamic parameters obtained from CEST NMR measurements (*vide infra*). Systematic devia-

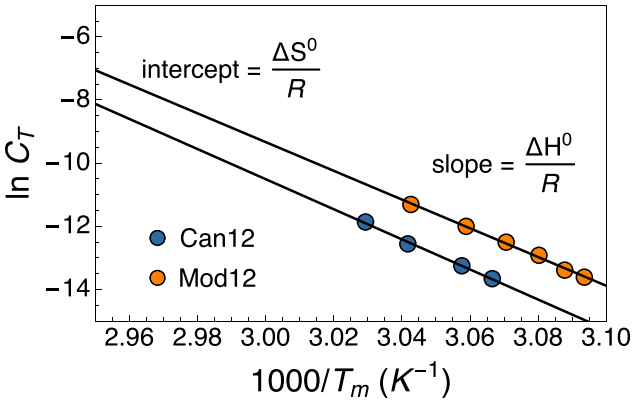

Figure 2. Concentration-dependent UV/Vis melting temperatures analysis of Can_{12} (blue) and Mod_{12} (orange). The reciprocal of the observed T_m is plotted as a function of the logarithm of the total DNA concentration (C_t). The slope and the intercept of the fitted linear corresponds to $\Delta H^\circ/R$, and $\Delta S^\circ/R$, respectively. The obtained thermodynamic parameters and the expected T_m at 1 mM concentration are listed in Table 1.

tion between the derived thermodynamic parameters of the two methods indicate a non-zero ΔC_p, which is expected but rarely taken into account explicitly (62).

The fitted standard enthalpy and enthropy changes, and derived Gibbs free energy differences indicate that inclusion of a 5fC site in a dsDNA oligomer leads to a consistent decrease in stability. The lower extent of cohesion between the two strands in Mod_{12} is apparent from the decrease of both ΔH° and ΔS° with respect to ΔH° and ΔS° of Can_{12}. The enthalpy–entropy compensation is only partial, leading to a decrease of ΔG° by 5 kJ mol^{-1} at 37°C and to a lowered melting temperature ($\Delta T_m = 2$°C) at 1 mM total monomer concentration.

Temperature-dependent CEST profiles

While UV/Vis melting studies provided a coarse glimpse into the impact of 5fC modification on the overall DNA stability, they are not sufficient to account for the kinetics related to the hybridization and melting processes. Conversely, CEST measurements recorded in a temperature-dependent manner can reveal both the kinetics and thermodynamics of ssDNA association and dsDNA dissociation in a site-specific manner. Therefore, we directed our efforts to study the melting profiles of Can_{12} and Mod_{12} using ^{1}H CEST experiments.

Overall, we recorded CEST profiles in a SELOPE pseudo-3D interleaved fashion for both samples at multiple temperatures between 37 and 62°C. We focused on the spectral region between 6.5 and 10 ppm, populated by resonances belonging to purines' H8 and H2 and pyrimidines' H6 in addition to the formyl proton of the formylated cytosine site (C6H7) (Figure 1B).

Due to reliance of the CEST technique on the non-equal chemical shifts between the interconverting states (in this case between dsDNA and ssDNA), only those sites could be analysed where this condition was met. The population of the ssDNA species fell below the detection limit at the three lowest measured temperatures (37, 45 and 53°C for both samples), and hence, those data sets were not employed in the parameter fitting procedure (see Supplementary Data text for details about the fitting and error estimation). High-temperature CEST profiles together with their fits to a two-state exchange model for three exemplary proton reporters (T5H6, C6H6 and A8H8) are displayed in Figure 3, and the rest is shown in Supplementary Figure S4.

We observed a single excited state species across all conditions for both samples. Coincidentally, the chemical shift of the single, uniquely detectable ES closely resembles the predicted value for a random coil shifts, whose values were calculated considering the nearest and next-nearest neighbor effect (63,64). A comparison between the CEST-derived chemical shifts and the predicted values is offered in Supplementary Table S2. This outcome suggests that there is a single predominant exchange process which involves melting of the dsDNA and annealing of ssDNA, and no detectable bulge or kink formation is present on the slow ms timescale. A comparison between the proton chemical shift values of the ds- and ssDNA strands is shown in Figure 4, which displays the simulated 1D spectra featuring the sites that could be identified and characterized via CEST. The simulations used the chemical shifts and R_1, R_2 relaxation rate constants that were obtained from the CEST fitting procedure.

We proceeded by fitting the CEST profiles assuming an all-or-none two-state exchange model (dsDNA $\rightleftharpoons$ 2 ssDNA), which we consider to be a reasonable assumption given the relatively short length of the selected sequence (65). As fitted parameters we obtained populations (p_D for dsDNA, $1 - p_D$ for ssDNA), exchange kinetics (k_{ex}), and chemical shifts of the exchanging states at each of the five highest temperatures, while we used the the sixth (lowest temperature of the ensemble for each sample) CEST profile to compare the predicted back-calculated value from the fits to the experimental value. To ease the fitting procedure, we measured ^{1}H longitudinal (R_1) and transverse (R_2) relaxation rates separately at multiple temperatures and used those rates as inputs for the CEST fits assuming that the rates of the dsDNA and ssDNA states are the same. The details of the fitting procedure is discussed in the Supplementary Data Text, and the fitted values are listed in Supplementary Tables S3 and S4.

Site specific quantification of thermodynamic and kinetic parameters

To obtain a rigorous, site-specific assessment of the DNA melting and annealing processes and the associated energetics, we analysed the CEST-derived site specific populations and exchange rates in the context of Eyring and van't Hoff theories. Within the proposed framework, concentrations are explicitly taken into account and extracted parameters are thus normalized and amenable to a scrupulous comparison.

The observed exchange rate in the CEST experiment (k_{ex}) is the sum of the forward (k_1) and backward (k_{-1}) reactions, where the forward rate is equivalent to the rate of duplex dissociation ($k_1 = k_d$), and the backward rate is related to the association rate, k_a

$$k_{-1} = 2[\text{ssDNA}]_{eq} k_a, \quad (1)$$

Table 1. Thermodynamic and kinetic parameters of the dsDNA melting process obtained from the van't Hoff and Eyring analysis of the CEST-derived exchange parameters, and from UV/Vis melting studies. Errors are given as one standard deviation

		Thermodynamics			Dissociation kinetics				Association kinetics		
		ΔG^0_{37C} kJ mol^{-1}	ΔH^0 kJ mol^{-1}	ΔS^0 J K^{-1} mol^{-1}	T_m (1 mM) °C	$\Delta G^{\ddagger}_{d,37C}$ kJ mol^{-1}	$\Delta H^{\ddagger}_d$ kJ mol^{-1}	$\Delta S^{\ddagger}_d$ J K^{-1} mol^{-1}	$\Delta G^{\ddagger}_{a,37C}$ kJ mol^{-1}	$\Delta H^{\ddagger}_a$ kJ mol^{-1}	$\Delta S^{\ddagger}_a$ J K^{-1} mol^{-1}
Mod_{12}	G1H8	64.6 ± 7.4	470 ± 128	1306 ± 389	75 ± 14.1	86.8 ± 4.6	122 ± 79	115 ± 241	22.2 ± 4.7	− 347 ± 80	− 1191 ± 244
	C2H6	74.7 ± 10.4	769 ± 158	2238 ± 475	62 ± 0.5	99.7 ± 5.6	413 ± 84	1011 ± 253	25. ± 5.5	− 356 ± 83	− 1227 ± 250
	A4H2	73.1 ± 1.	712 ± 16	2060 ± 49	63 ± 0.1	101.6 ± 0.6	432 ± 9	1064 ± 28	28.5 ± 0.6	− 281 ± 9	− 996 ± 28
	A4H8	71.9 ± 1.2	716 ± 19	2078 ± 58	62 ± 0.1	101.7 ± 1.2	444 ± 18	1102 ± 55	29.8 ± 1.1	− 273 ± 18	− 975 ± 53
	T5H6	64.3 ± 1.1	613 ± 18	1768 ± 53	62 ± 0.1	93.9 ± 0.8	340 ± 12	792 ± 38	29.6 ± 0.8	− 273 ± 12	− 976 ± 37
	C6H6	67.4 ± 1.3	652 ± 20	1886 ± 60	63 ± 0.1	99.9 ± 1.6	422 ± 24	1038 ± 73	32.5 ± 1.2	− 230 ± 19	− 848 ± 58
	C6H7	48.6 ± 1.6	344 ± 26	953 ± 78	68 ± 0.8	118.8 ± 1.8	640 ± 28	1680 ± 86	70.2 ± 1.9	296 ± 30	727 ± 90
	A8H2	71.2 ± 0.6	678 ± 9	1955 ± 28	64 ± 0.1	99.6 ± 0.4	398 ± 6	962 ± 18	28.4 ± 0.4	− 280 ± 6	− 994 ± 18
	A8H8	65.3 ± 0.9	604 ± 14	1737 ± 42	63 ± 0.1	98.2 ± 0.7	383 ± 11	920 ± 34	32.9 ± 0.6	− 221 ± 10	− 817 ± 31
	T9H6	64.6 ± 1.2	611 ± 18	1761 ± 55	63 ± 0.1	97.1 ± 1.	383 ± 16	920 ± 47	32.5 ± 1.	− 228 ± 15	− 841 ± 45
	C10H6	66.4 ± 1.2	636 ± 19	1835 ± 57	63 ± 0.2	97.5 ± 1.	384 ± 16	924 ± 49	31.1 ± 1.1	− 252 ± 17	− 911 ± 50
	C12H6	69.7 ± 1.	679 ± 15	1964 ± 47	63 ± 0.1	98.7 ± 0.9	402 ± 14	977 ± 43	29. ± 1.	− 277 ± 15	− 987 ± 45
	UV/Vis	50.4 ± 1.2	338 ± 25	1056 ± 77	66 ± 1.0						
Can_{12}	C2H6	71.1 ± 1.7	641 ± 26	1838 ± 77	65 ± 0.2	99.2 ± 2.2	372 ± 31	880 ± 94	28.1 ± 2.1	− 269 ± 30	− 958 ± 91
	G3H8	69. ± 1.3	608 ± 20	1738 ± 59	66 ± 0.2	106.4 ± 1.9	459 ± 28	1138 ± 84	37.4 ± 1.6	− 149 ± 23	− 600 ± 71
	A4H2	72.5 ± 0.9	626 ± 14	1784 ± 41	67 ± 0.2	100.1 ± 0.7	363 ± 11	847 ± 32	27.7 ± 0.7	− 263 ± 11	− 936 ± 32
	A4H8	77.6 ± 1.1	732 ± 16	2110 ± 48	65 ± 0.1	108.3 ± 1.8	498 ± 25	1258 ± 76	30.7 ± 1.8	− 234 ± 25	− 852 ± 75
	T5H6	71. ± 1.6	633 ± 23	1812 ± 70	65 ± 0.2	102.7 ± 1.4	415 ± 20	1006 ± 60	31.7 ± 1.3	− 218 ± 19	− 806 ± 57
	C6H6	70.7 ± 1.4	638 ± 21	1829 ± 63	65 ± 0.2	99.8 ± 1.2	379 ± 17	900 ± 51	29.1 ± 1.2	− 259 ± 17	− 929 ± 51
	A8H2	73.8 ± 1.1	643 ± 17	1834 ± 50	67 ± 0.2	103.6 ± 1.	409 ± 15	986 ± 45	29.8 ± 1.	− 233 ± 14	− 849 ± 42
	A8H8	73.1 ± 1.	666 ± 15	1910 ± 45	65 ± 0.1	104.3 ± 1.	439 ± 14	1078 ± 43	31.2 ± 0.9	− 227 ± 13	− 833 ± 40
	T9H6	70.2 ± 1.8	626 ± 27	1791 ± 80	65 ± 0.3	102.2 ± 1.5	410 ± 22	993 ± 66	31.9 ± 1.5	− 216 ± 22	− 798 ± 65
	C10H6	71.3 ± 1.4	642 ± 21	1841 ± 64	65 ± 0.2	100.3 ± 1.3	382 ± 19	908 ± 56	29. ± 1.3	− 260 ± 20	− 933 ± 59
	C12H6	77.6 ± 1.1	723 ± 17	2082 ± 50	65 ± 0.1	104.2 ± 1.3	434 ± 18	1064 ± 55	26.6 ± 1.3	− 289 ± 19	− 1018 ± 58
	UV/Vis	55.2 ± 2.5	408 ± 43	1130 ± 131	69 ± 1.5						

where $[\text{ssDNA}]_{eq} = C_t(1 - p_D)$ is the equilibrium concentration of the ssDNA, and C_t is the total ssDNA concentration. The expression of equation (1) was derived assuming a weak perturbation as described in the Supplementary Data (66). k_a and k_d kinetic rates can be inferred from fitted k_{ex} and p_D parameters as follows:

$$k_a = \frac{k_{ex} p_D}{2C_t(1 - p_D)} \tag{2}$$

$$k_d = k_{ex}(1 - p_D) \tag{3}$$

The temperature dependence of k_d and k_a can be used to derive the activation barriers for the dissociation ($\Delta G^{\ddagger}_d(T)$) and association ($\Delta G^{\ddagger}_a(T)$) processes. Subsequently, the corresponding activation enthalpy ($\Delta H^{\ddagger}_d$, $\Delta H^{\ddagger}_a$) and entropy ($\Delta S^{\ddagger}_d$, $\Delta S^{\ddagger}_a$) changes can be obtained. Further, the ratio between $k_d(T)$ and $k_a(T)$ rates, which gives the equilibrium constant ($K_d(T)$) for dissociation, is relevant for the calculation of the site-specific melting thermodynamics parameters (ΔG°, ΔH°, ΔS° and T_m).

In order to infer on the thermodynamics of the process, we assumed ΔS° and ΔH° to be temperature independent (so-called van't Hoff assumption) and fitted the variation of K_d with temperature accordingly

$$K_d(T) = \exp\left(\frac{-\Delta H^{\circ}}{RT} + \frac{\Delta S^{\circ}}{R}\right) \tag{4}$$

leading to a linear dependence between $\ln K_d$ and $1/T$ (Figure 6). See Supplementary Data text for the details of the derivation of the thermodynamic parameters.

Consistently with published experimental and computational DNA melting kinetic studies, we find that the temperature-dependence of the k_d (k_a) rates follow a Eyring-Polanyi relationship and they increase (decrease) with increasing temperature (3,65). The dissociation is a five orders-of-magnitude slower process than the bimolecular association, the former being in the range of 0.5–10 s^{-1}, and the latter in 0.5–2 × 10^5 M^{-1} s^{-1}. When comparing the two samples at the same temperatures (Figure 5), we find that the k_d rates are consistently and systematically 2- to 5-fold higher for Mod_{12} than for Can_{12}.

In analogy to the dissociation rates, also k_a values show a regular pattern. We detect the hybridization process to be 1.5- to 3.5-fold faster in the canonical sample. These results indicate that 5fC simultaneously facilitates the melting process and hinders the annealing process.

Figure 6 shows the logarithm of the obtained kinetic rates and equilibrium constants for T5H6, C6H6 and A8H8 protons as a function of $1000(1/T - 1/T_{hm})$, where $T_{hm} = n(\sum_i^n T_i^{-1})^{-1}$ is the harmonic mean of the temperatures used in the CEST studies, which is 58 and 60°C for Mod_{12} and Can_{12}, respectively. The linear change of the kinetic rates with T^{-1} suggests a single, well-defined transition state. According to a modified version of the Eyring–Polanyi equation, which takes into account the statistical bias induced by the different temperatures used, the dissociation and association rate can be given as

$$k_i(T) = \frac{\kappa k_B T}{h} \times \exp\left(-\frac{\Delta G^{\ddagger}_i(T_{hm})}{RT_{hm}} - \frac{\Delta H^{\ddagger}_i}{R}\left(\frac{1}{T} - \frac{1}{T_{hm}}\right)\right) \tag{5}$$

where i stands for association (a) or dissociation (d), k_B is the Boltzmann constant, R is the universal gas constant,

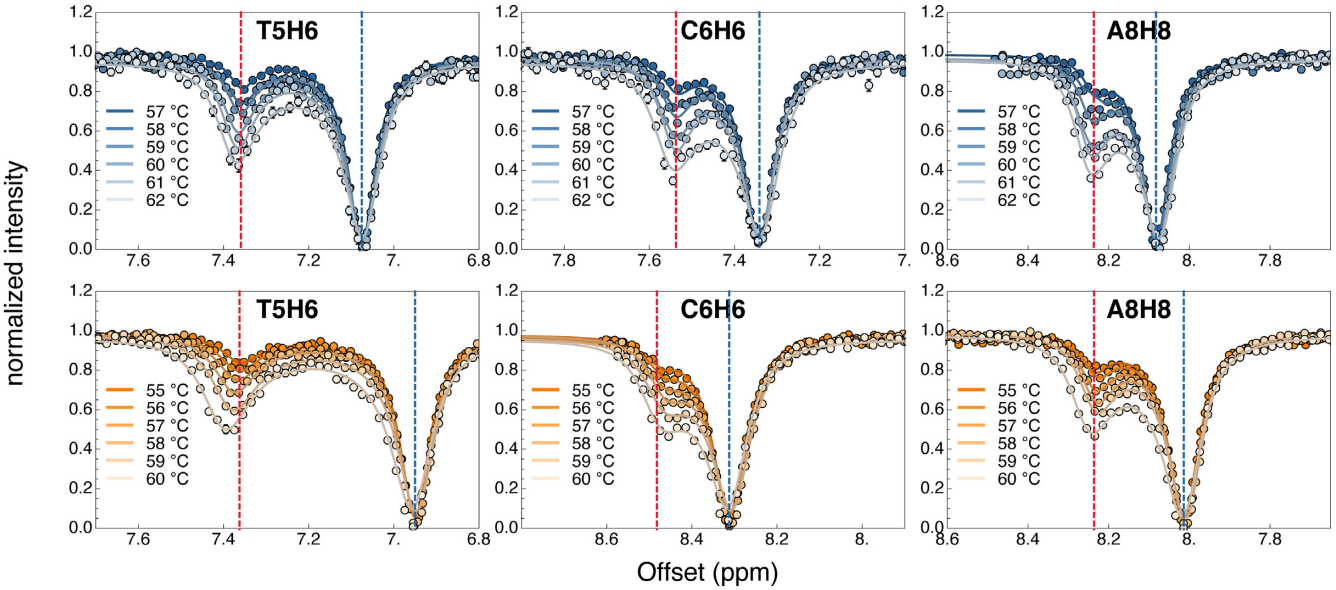

Figure 3. Temperature-dependent ^{1}H CEST profiles and fits of T5H6, C6H6 and A8H8 protons of Can_{12} (blue, upper row) and Mod_{12} (orange, bottom row). Dashed blue and red vertical lines indicate the chemical shifts of the double-stranded and single-stranded conformations, respectively. As expected, the population of ssDNA increases gradually with increasing temperature.

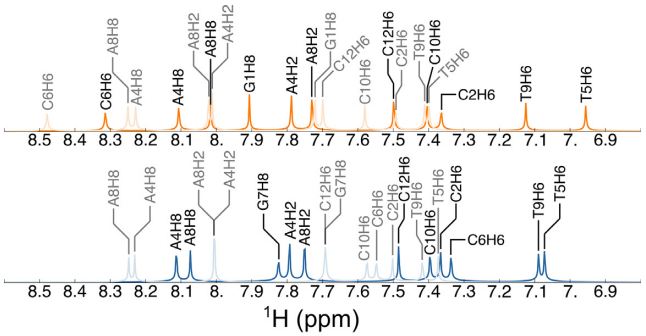

Figure 4. Simulated ^{1}H 1D spectra. Darker and lighter shades represent the ground-state (dsDNA) and excited-state (ssDNA) conformations. Only those sites are included that showed a second dip in the CEST profiles.

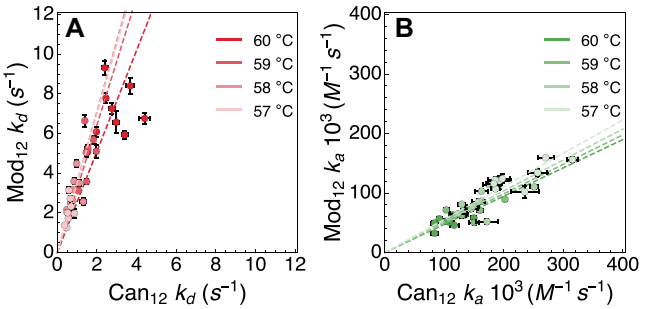

Figure 5. Correlation between the site-specific k_d (**A**) and k_a (**B**) kinetic rates between the observable sites of Mod_{12} and Can_{12} at 57, 58, 59 and 60°C. Dashed lines represents the linear fits at each temperature with a slope of 2.5–3.5 for k_d and 0.47–0.55 for k_a.

and κ is the transmission coefficient, assumed to be 1. The slope and intercept of the linear fit to $\ln k_i/T$ versus $1000(1/T - 1/T_{hm})$ gives the activation enthalpy change ($\Delta H_i^{\ddagger}$) and the Gibbs free energy change at T_{hm} ($\Delta G_i^{\ddagger}(T_{hm})$) respectively, from which the activation entropy ($\Delta S_i^{\ddagger}$) can be derived as

$$\Delta S_i^{\ddagger} = \frac{\Delta H_i^{\ddagger} - \Delta G_i^{\ddagger}(T_{hm})}{T_{hm}} \quad (6)$$

From Figure 6 and Supplementary Figure S4 it is apparent that, for all sites, our results indicate thermally activated kinetics with a positive dissociation (Arrhenius behavior) barrier and a negative association barrier (anti-Arrhenius), with the notable exception of the formyl C6H7 proton of the modified base (Supplementary Figure S4) where both processes follow an Arrhenius behavior. The robustness of the CEST-based thermodynamics and kinetics analysis is reflected in the exceptionally high consistency between the temperature-dependence of the kinetic rates and equilibrium parameters when the same proton of the two samples are compared.

Site-specific activation kinetics and thermodynamics parameters for both samples are summarized in Table 1, and visualized in the insets of Figure 6 and in Figure 7. Consistently with the UV/Vis melting study, we found that the 5fC modification destabilizes the 12-mer DNA by ~4-8 kJ mol^{-1} at 37°C, and decreases the melting temperature by ~2°C. As a support for the two-state exchange model, we found that the destabilization effect is not localized, but rather it involves the whole molecule. The ΔS and ΔH parameters are scattered over a large range and seemingly there is little or no correlation between their absolute value and their location in the sequence or chemical type (i.e. if they are purine or pirimidine H2, H6, or H8 protons). The only main outlier is the formyl proton of the modified cytosine base, C6H7, whose ΔH° and ΔS° are half the size, $\Delta H_d^{\ddagger}$ and $\Delta S_d^{\ddagger}$ are ~70% larger, and $\Delta H_a^{\ddagger}$ and $\Delta S_a^{\ddagger}$ have the oppo-

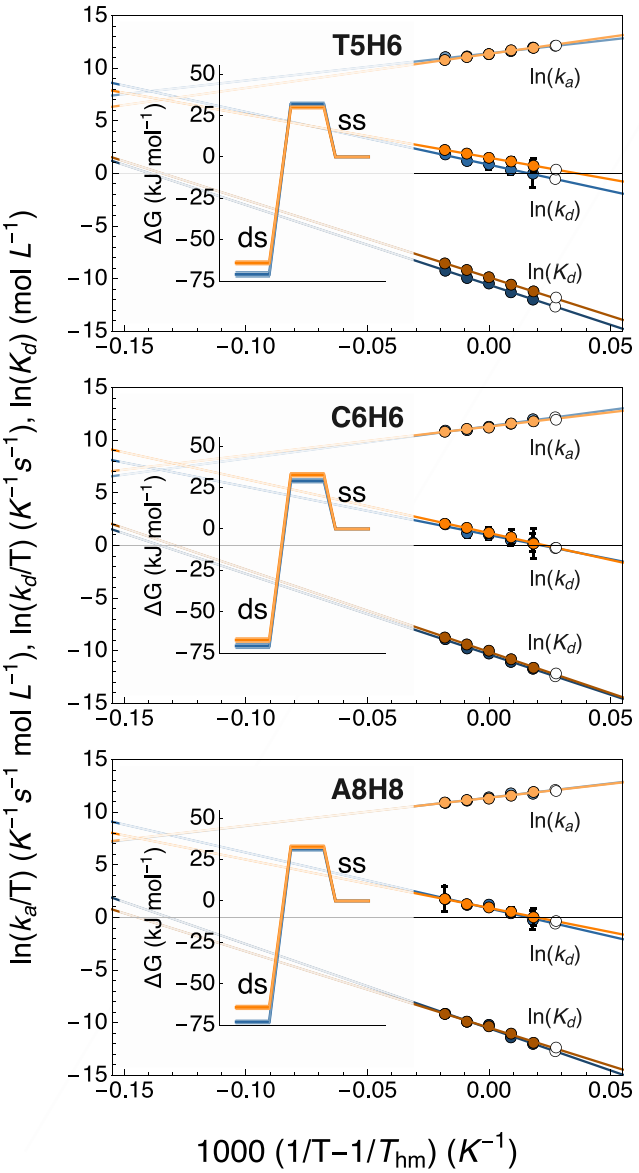

Figure 6. van't Hoff plots are displayed for three selected sites. Shades of orange and blue indicate data entries and linear fits for Mod_{12} and Can_{12}, respectively. White data points represent back-calculated values. Fitted values for the association and dissociation rate constants (k_a and k_d) and the equilibrium constant of dissociation K_d are listed in Supplementary Tables S5 (Mod_{12}) and S6 (Can_{12}). Insets present the relevant Gibbs free energy plots at each site at 37°C.

site sign as those of the other H2, H6 and H8 protons in the molecule. To assess whether this large difference in the thermodynamics and kinetics values of C6H7 reflects a divergence from a two-state equilibrium we plotted ΔH as a function of ΔS (such a plot reflects the genuine entropy-enthalpy compensation). Since $\Delta G = \Delta H - T\Delta S$, therefore the plot of ΔH versus ΔS gives a linear with a slope of T_c which corresponds to the characteristic temperature of the measurement, and the intercept reflects the ΔG at T_c. Figure 7A–C displays the entropy-enthalpy compensation plots for both thermodynamic and kinetic processes with a fitted linear including all observable protons of both samples. Most sites are clustered together, while C6H7, G1H8 and C2H6 protons of Mod_{12} have distinctly different ΔH and ΔS values, however their ΔH and ΔS compensate each other leading to a ΔG, which agrees well with the overall ΔG of the system. The fitted slopes (T_c) and intercepts ($\Delta G(T_c)$) are close to the experimental T_{hm} and obtained average $\Delta G(T_{hm})$ values.

Since ΔG is the difference of two large quantities, the ΔH versus ΔS plot is ill-suited to indicate any site-specific differences in the Gibbs free energy differences. The subtle destabilization caused by the epigenetically relevant 5fC nucleotide can be better appreciated in Figure 7 D where $\Delta G_d^{\ddagger}$ is plotted as a function of ΔG° for both samples. Here again, C6H7 seems to be an outlier with much higher activation free energies and lower equilibrium energies than the rest of the molecule. Generally, both the ΔG° and $\Delta G^{\ddagger}$ values are decreased by ~10 and ~20 kJ mol^{-1} in Mod_{12} with respect to the same values in Can_{12}, respectively. Otherwise, we see a linear correlation between the activation and equilibrium free energies (C6H7 was excluded from this analysis). In terms of linear Gibbs free energy relation, the slope $d\Delta G^{\ddagger}/d\Delta G^{\circ}$, with a value between 0 and 1, measures how much the transition state resembles the GS or ES conformations (67). In our case, the slope of the Gibbs energy changes associated with the DNA melting process is 0.83 which indicates that the transition state is structurally related to the single-stranded conformation.

DISCUSSION

The naturally occurring 5fC DNA modification is an epigenetic marker whose biological roles have not been yet precisely defined (4–7). In order to account for the impact of 5fC incorporation into dsDNA, we studied how such a modification reshapes the thermodynamics and kinetics of the process of DNA melting and annealing.

We propose that our findings can be interpreted within the framework of two separate, but complementary aspects: the role of 5fC as an independent, semi-permanent modification and its unique interactions with TDG protein within the framework of the active demethylation cycle.

5fC as a semi-permanent modification

In agreement with previous studies, we identified no indications of 5fC-induced wobble base-pairing arising from an imino-amino tautomerization processes, nor any hint pointing towards a deviation from the B-DNA structure (34–36). If there is no obvious modification on the ground-state structure of dsDNA due to 5fC incorporation then the questions arise, whether it impacts the properties of a rarely visited excited state or it alters the potential energy surface around the accessible conformations.

To address these questions, we performed a thorough, site-specific thermodynamics and kinetics study on the melting process involving two 12mer dsDNA samples. From the preliminary UV/Vis melting study, we could confirm that 5fC destabilizes the DNA structure when compared to canonical C by increasing the equilibrium constant of the dissociation, but this method was unable to clarify

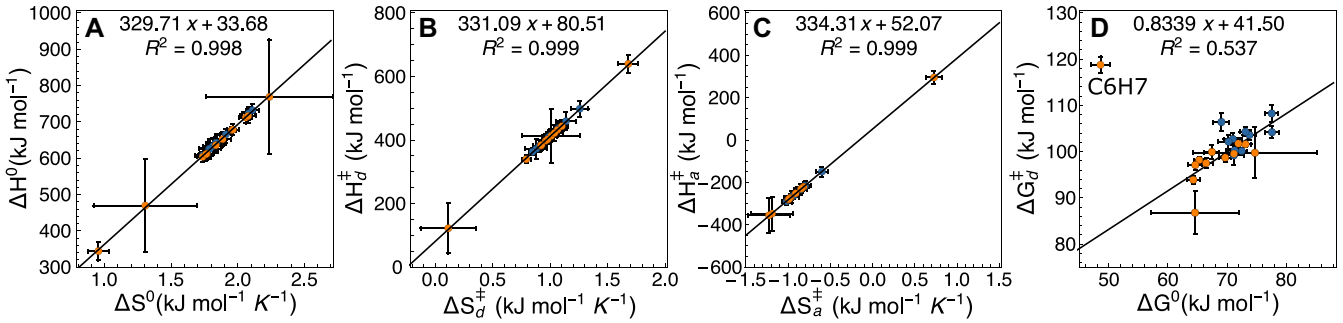

Figure 7. (**A–C**) Correlation between the enthalpy and entropy changes of the thermodynamic and kinetic processes for Can_{12} (blue) and Mod_{12} (orange). The solid lines represent a linear fit in the form of $\Delta H = T_c\Delta S + \Delta G$. (**D**) Correlation between the dissociation and equilibrium free energy changes of Can_{12} (blue) and Mod_{12} (orange) obtained at 37°C.

if this increase was a consequence of the decreased association rate, or the increased dissociation rate. Hence, we turned our attention to CEST measurements which provide site-specific information about both the kinetics (exchange rates) and the thermodynamics (populations) of simple equilibrium processes. With CEST, we could establish that the decreased stability of 5fC modified DNA was a consequence of the ∼2–5-fold increase of the dissociation rate and the ∼1.5–3.5-fold decrease of the association rate, leading to a ∼4–9 times higher equilibrium dissociation constant.

We propose that these results are relevant for our understanding of 5fC as an independent, semi-permanent modification and can be explained considering the molecular grounds of the destabilization. Experimental and computational studies have described the electron-withdrawing (EW) properties of the formyl group on position 5 (38,39), and it is conceivable that the very same EW effects that modulate the dissociation process have a similar impact on the annealing kinetics by curtailing the extent to which the critical amount of bases needed to trigger the nucleation process, a key factor in double-stranded nucleic acid hybridization, come in contact (3).

By performing the CEST measurements in a temperature-dependent fashion we ascertained that the kinetic rates and the equilibrium constants follow an Eyring and van't Hoff relationship, respectively (cf. Figure 6 and Supplementary Figure S4). Indeed, the overall similarity of the profiles between equivalent sites of the two samples suggests that inclusion of 5fC leaves the transition states and potential energy surfaces unaltered. Consistently with previous investigations of the melting kinetics of short dsDNA sequences, the dissociation kinetics were found to be compatible with an Arrhenius process, in opposition to the annealing mechanism, which appears to progressively slow down at higher temperatures (3,65). The validity of the above observations holds for both samples, indicating that 5fC does not drastically perturb dsDNA and does not fundamentally change its hybridization and melting dynamics.

In order to understand the subtle differences arising from the incorporation of formylated cytosine in DNA, we resorted to a global analysis of the thermodynamics and kinetic parameters determined via our CEST and UV/Vis studies. The correlation of the activation free energy change ($\Delta G^{\ddagger}_{d}$) with the standard Gibbs free energy change (ΔG^{0}) assists in rationalizing the elusive 5fC-induced destabilization, which stems from the higher energy associated with the ground state of the dsDNA conformer. Indeed, values relative to Mod_{12} tend to cluster at the lower energy side of the correlation plot, meaning that the lower energy barrier separating ds- and ssDNA is mostly, if not entirely due to the lower cohesion between the two strands. Furthermore, the impact of the destabilization generated by the centrally placed modification is not localized as previously suggested (38,68,69). Rather, it appears to be broad and all base pairs are consistently affected when compared to the canonical sample. This evidence implies that 5fC curtails the extent to which bases in dsDNA can cooperatively stabilize each other, causing a global weakening of the double stranded structure, albeit without disrupting the original B-DNA geometry. By examining the ^{1}H 1D spectra of Mod_{12} and Can_{12}, it can be appreciated how the sole proton in the imino region which is shifted upon 5fC incorporation is G7H1, indicating that the hydrogen bond strength of the centrally positioned C6:G7 base pair is significantly weakened upon cytosine formylation. In sharp contrast, other base pairs, and most notably even the immediate neighbour T5:A8, are entirely unaffected in terms of H-bonding capacity. The above observation is deceptively simple when compared to the numerical results obtained via our CEST analysis. How can then such a localized effect impact global processes such as melting and reannealing? We propose that the weakened H-bond of 5fC6:G7 diminishes the extent to which the entire ensemble of bases can cooperatively stabilize each other via mechanisms of π–π stacking, cross-strands interactions etc. as previously reported (1–3). Lower H-bond propensity translates into a higher propensity for base wiggling on very fast (ns–μs) time scales, which in turn weakens the extent to which neighboring bases stabilize each other. A schematic representation of this process is depicted in Supplementary Figures S6 and S7.

Previous research has established the notion that the rate limiting step in the process of ssDNA association dynamics is the nucleation (3,70). During the formation of a DNA double strand, the transition state structure only features a few preformed base pairs. Then, by progressively 'zipping-up' the remaining sites until it reaches the termini, and achieves the formation of the double strand. The nega-

tive $\Delta H_a^{\ddagger}$ that we observed for all but one sites corroborates this idea and indicates that the transition state towards a double-stranded structure is enthalpically stabilized by newly formed hydrogen bonds. Remarkably, the only proton which follows an opposite trend is the formyl proton of 5fC. Here, the positive $\Delta H_a^{\ddagger}$ suggests that the formyl group forms enthalpically favorable interactions (e.g. H-bonds) even in the single-stranded state, and these interactions should break and newly form during the hybridization process. The anomalous behavior of C6H7 with respect to other proton reporters could stem from the rotation of the formyl group along the axis of the carbonyl carbon and the covalently bonded ring carbon, which would normally occur on the nanosecond timescale. We speculate that the strong hydrogen bond between the carbonyl oxygen and the nearby amino protons slows this rotation down to the ms time scale and makes this exchange process interfere with the melting kinetics. Thus, the observed kinetics and thermodynamics of C6H7 is a mixture of both bond rotation and DNA hybridization.

The correlation between standard and activation Gibbs free energy changes has been long used to comment on the nature of the transition state (71). The slope of the fitted linear normally ranges between 0 and 1, meaning that the transition state structurally resembles the reactant or the product, respectively. Consistently with this view, we find a slope of 0.83, providing experimental support to the nucleation theory where the otherwise floppy strands are anchored only at few locations and hence resembling the single-stranded form. This finding is in contrast to the conclusion of the studies related to individual base-pair openings, which found that the transition state of that process resembles the dsDNA architecture more than the ssDNA random coil (67).

In view of the foregoing, we conclude that formylation of cytosine within its naturally occurring setting distinctly enhances dsDNA opening. It has been long established that the presence of the simplest cytosine epigenetic modification, 5mC, is linked to 'genetic silencing', i.e. repression of gene expression (19–21). 5mC has also been confirmed to stabilize the dsDNA conformer, both kinetically and thermodynamically (34,38,40). Our study suggests that, in sharp contrast, by enhancing dsDNA double strand opening 5fC might facilitate gene expression. The fact that formylated cytosine is abundant (relative to its presence in different types of genomes) in the context of cancerous and stem cells, where cell activity is enhanced, is perhaps more than coincidental.

5fC as a substrate

A central conundrum of epigenetics is how reader proteins recognize their substrates with such high specificity over the overwhelming background represented by canonical nucleic acids. This puzzle is especially intriguing for TDG which can differentiate between carbonyl-bearing epigenetically modified cytosines such as 5fC and 5caC versus canonical, methylated and hydroxymethylated (C, 5mC, 5hmC) cytosines. Several studies have proposed the so-called 'pinch-push-pull' mechanism to explain the outstanding recognition capabilities of TDG towards dsDNA mismatches (e.g. the G–T mismatched base-pair), 5fC and 5caC (although in a highly pH-dependent manner) (72–75). According to the consensus enzymatic model for TDG substrate recognition, base extrusion is proposed to be initiated by residue Arg275, which interferes with Watson-Crick base-pairing geometry and triggers the process of substrate flipping into the active site of TDG. An array of subsequent recognition and binding processes precedes the enzymatic event of the glycosidic bond cleavage. Although our work does not involve studies of DNA-protein interactions, and thus is ill-suited to comment on the whole enzymatic process, we propose that our results fit nicely within the above mentioned 'pinch-push-pull' mechanism. Indeed, we suggest that the lower activation barrier at 37°C that 5fC-bearing DNA encounters in the course of a melting process facilitates the base extrusion enacted by TDG. Upon interrogation of Arg275, the weakened (by the inclusion of 5fC) Watson-Crick base pair geometries facilitate base flipping into the active site of the enzyme. In this context, canonical bases (or other epigenetic modifications lacking the carbonyl moiety which are not recognized by TDG, such as 5mC and 5hmC) are then released and resume their Watson-Crick hydrogen bonding. Otherwise, 5fC is recognized and retained in the active site, allowing TDG to perform base-excision.

CONCLUSION

In this work, we investigated the effect of the inclusion of 5fC within its naturally occurring settings and conditions in the context of a double-stranded DNA 12mer, focusing our efforts on the slow time-scale dynamics close to the melting temperature.

We achieved a quantitative assessment of the degree of destabilization degree by 5fC via ^{1}H CEST techniques, which yielded the populations of both the single and double stranded DNA conformers, their chemical shifts, and the overall exchange kinetic rates k_{ex} for each temperature. Further decomposition of the site-specific k_{ex} into k_d and k_a allowed for the determination of kinetic rates for both the melting and annealing processes. We find that both k_d and k_a are consistently affected by the presence of 5fC and their combined impact on the dissociation equilibrium constant near the melting temperature can be almost as large as an order of magnitude. To gain further insights into the energetics of double helix to single helix transition, we developed a methodological framework for the extraction of thermodynamic (ΔG°, ΔH°, ΔS°, T_m) and kinetic ($\Delta G^{\ddagger}$, $\Delta H^{\ddagger}$, $\Delta S^{\ddagger}$) parameters from the CEST derived k_{ex} and population values. Our findings indicate that inclusion of 5fC generates a tenuous degree of destabilization in the duplex state without affecting the transition state architecture, which remains structurally indistinguishable from its canonical counterpart. Said destabilization affects the conformational dsDNA-ssDNA conformational equilibrium. We speculate the identified features induced by 5fC onto dsDNA might account for the both its role as an independent, semi-permanent epigenetic marker and the selective recognition and activity of the TDG protein towards sites bearing the modification.

In this study, we confined our analysis to slow ms timescale motions, focusing on the site-specific quantification of

5fC-related energetics. The noticeable and localized weakening of the 5fC:G base pair in contrast to its canonical C:G counterpart implies that the origin of the induced destabilization effects has its roots in a faster time-scale motion, which has as of yet gone undetected. Further research will be needed to unravel and characterize possible excited states on faster time scales that might be impacting in a localized fashion the dsDNA flexibility at (or near) the 5fC:G base pair.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Anthony Crisp for insightful discussion and help in the measurements of UV/Vis melting profiles.

FUNDING

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [SFB 1309-325871075, in part]; EU-ITN LightDyNAmics [765266]; ERC-AG EpiR [741912]; Center for NanoScience, the Excellence Clusters CIPSM; Fonds der Chemischen Industrie. Funding for open access charge: The SFB1309 framework (https://www.sfb1309.de/about-sfb-1309/).
Conflict of interest statement. None declared.

REFERENCES

1. Poater,J., Swart,M. and Bickelhaupt,F.M. (2014) B-DNA structure and stability: the role of hydrogen bonding, π−π stacking interactions, twist-angle, and solvation. *Org. Biomol. Chem.*, **12**, 4691–4700.
2. Hobza,P. and Šponer,J. (1999) Structure, energetics, and dynamics of the nucleic acid base pairs: nonempirical ab initio calculations. *Chem. Rev.*, **99**, 3247–3276.
3. Ouldridge,T.E., Šulc,P., Romano,F., Doye,J.P.K. and Louis,A.A. (2013) DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic Acids Res.*, **41**, 8886–8895.
4. Bilyard,M.K., Becker,S. and Balasubramanian,S. (2020) Natural, modified DNA bases. *Curr. Opin. Chem. Biol.*, **57**, 1–7.
5. Carell,T., Kurz,M.Q., Müller,M., Rossa,M. and Spada,F. (2018) Non-canonical bases in the genome: the regulatory information layer in DNA. *Angew. Chem. Int. Ed.*, **57**, 4296–4312.
6. Chen,Y., Hong,T., Wang,S., Mo,J., Tian,T. and Zhou,X. (2017) Epigenetic modification of nucleic acids: from basic studies to medical applications. *Chem. Soc. Rev.*, **46**, 2844–2872.
7. Hardwick,J.S., Lane,A.N. and Brown,T. (2018) Epigenetic modifications of cytosine: biophysical properties, regulation, and function in mammalian DNA. *BioEssays*, **40**, doi:10.1002/bies.201700199.
8. Sasaki,H. and Matsui,Y. (2008) Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat. Rev. Genet.*, **9**, 129–140.
9. Yamaguchi,S., Hong,K., Liu,R., Inoue,A., Shen,L., Zhang,K. and Zhang,Y. (2013) Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell Res.*, **23**, 329–339.
10. Neri,F., Incarnato,D., Krepelova,A., Rapelli,S., Anselmi,F., Parlato,C., Medana,C., Dal Bello,F. and Oliviero,S. (2018) Single-base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter DNA methylation dynamics. *Cell Rep.*, **10**, 674–683.
11. Basanta-Sanchez,M., Wang,R., Liu,Z., Ye,X., Li,M., Shi,X., Agris,P.F., Zhou,Y., Huang,Y. and Sheng,J. (2017) TET1-mediated oxidation of 5-formylcytosine (5fC) to 5-carboxycytosine (5caC) in RNA. *ChemBioChem*, **18**, 72–76.
12. Van Haute,L., Powell,C.A. and Minczuk,M. (2017) Dealing with an unconventional genetic code in mitochondria: the biogenesis and pathogenic defects of the 5-formylcytosine modification in mitochondrial $tRNA^{(Met)}$. *Biomolecules*, **7**, 24.
13. Dawson,M.A. and Kouzarides,T. (2012) Cancer epigenetics: from mechanism to therapy. *Cell*, **150**, 12–27.
14. Cavalli,G. and Heard,E. (2019) Advances in epigenetics link genetics to the environment and disease. *Nature*, **517**, 489–499.
15. Baylin,S. and Jones,P.A. (2011) A decade of exploring the cancer epigenome – biological and translational implications. *Nat. Rev. Cancer*, **11**, 726–734.
16. Wu,H. and Zhang,Y. (2011) Mechanisms and functions of Tet protein- mediated 5-methylcytosine oxidation. *Genes Dev.*, **25**, 2436–2452.
17. Storebjerg,T.M., Strand,S.H., Høyer,S., Lynnerup,A., Borre,M., Ørntoft,T.F. and Sørensen,K.D. (2018) Dysregulation and prognostic potential of 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) levels in prostate cancer. *Clin. Epigenet.*, **10**, 105.
18. Derreumaux,S., Chaoui,M., Tevanian,G. and Fermandjian,S. (2001) Impact of CpG methylation on structure, dynamics and solvation of cAMP DNA responsive element. *Nucleic Acids Res.*, **29**, 2314–2326.
19. Bird,A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
20. Moarii,M., Boeva,V., Vert,J. and Reyal,F. (2015) Changes in correlation between promoter methylation and gene expression in cancer. *BMC Genomics*, **16**, 873.
21. Jones,P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.*, **13**, 484–492.
22. Smith,Z.D. and Meissner,A. (2013) DNA methylation: roles in mammalian development. *Nat. Rev. Genet.*, **14**, 204–220.
23. Koivunen,P. and Laukka,T. (2018) The TET enzymes. *Cell. Mol. Life Sci.*, **75**, 1339–1348.
24. Ito,S., Shen,L., Dai,Q., Wu,S.C., Collins,L.B., Swenberg,J.A., He,C. and Zhang,Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1304.
25. Hu,L., Li,Z., Cheng,J., Rao,Q., Gong,W., Liu,M., Shi,Y.G., Zhu,J., Wang,P. and Xu,Y. (2013) Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell*, **155**, 1545–1555
26. Pfaffeneder,T., Hackner,B., Truß,M., Münzel,M., Müller,M., Deiml,C.A., Hagemeier,C. and Carell,T. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed.*, **50**, 7008–7012.
27. Su,M., Kirchner,A., Stazzoni,S., Müller,M., Wagner,M., Schröder,A. and Carell,T. (2016) 5-Formylcytosine could be a semipermanent base in specific genome sites. *Angew. Chem. Int. Ed.*, **55**, 11797–11800.
28. Bachman,M., Uribe-Lewis,S., Yang,X., Burgess,H.E., Iurlaro,M., Reik,W., Murrell,A. and Balasubramanian,S. (2015) 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.*, **11**, 555–557.
29. Zhang,Y. and Zhou,C. (2019) Formation and biological consequences of 5-Formylcytosine in genomic. *DNA Repair*, **81**, 102649.
30. Pfeifer,G.P., Szabó,P.E. and Song,J. (2020) Protein interactions at oxidized 5-methylcytosine bases. *J. Mol. Biol.*, **432**, 1718–1730.
31. Rausch,C., Hastert,F.D. and Cardoso,M.C. (2019) DNA modification readers and writers and their interplay. *J. Mol. Biol.*, **432**, 1731–1746.
32. Wu,X. and Zhang,Y. (2017) TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.*, **18**, 517–534.
33. Raiber,E.A., Murat,P., Chirgadze,D.Y., Beraldi,D., Luisi,B.F. and Balasubramanian,S. (2015) 5-formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.*, **22**, 44–49.
34. Szulik,M.W., Pallan,P.S., Nocek,B., Voehler,M., Banerjee,S., Brooks,S., Joachimiak,A., Egli,M., Eichman,B.F. and Stone,M.P. (2015) Differential stabilities and sequence-dependent base pair opening dynamics of Watson Crick base pairs with 5-hydroxymethylcytosine, 5-formylcytosine, or 5-carboxylcytosine. *Biochemistry*, **134**, 8148–8161.
35. Hardwick,J.S., Ptchelkine,D., El-Sagheer,A.H., Tear,I., Singleton,D., Phillips,S.E.V., Lane,A.N. and Brown,T. (2017) 5-Formylcytosine does not change the global structure of DNA. *Nat. Struct. Mol. Biol.*, **24**, 544–552.

36. Fu,T., Liu,L., Yang,Q., Wang,Y., Xu,P., Zhang,L., Liu,S., Dai,Q., Ji,Q., Xu,G. *et al.* (2019) Thymine DNA glycosylase recognizes the geometry alteration of minor grooves induced by 5-formylcytosine and 5-carboxylcytosine. *Chem. Sci.*, **10**, 7407–7417.
37. Ngo,T.T.M., Yoo,J., Dai,Q., Zhang,Q., He,C., Aksimentiev,A. and Ha,T. (2016) Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.*, **7**, 10813.
38. Dai,Q., Sanstead,P.J., Peng,C.S., Han,D. and He,C. (2016) Weakened N3 hydrogen bonding by 5-formylcytosine and 5-carboxylcytosine reduces their base-pairing stability. *ACS Chem. Biol.*, **11**, 470–477.
39. Nunes,F.B., Bettega,M.H.F. and Almeida Sanchez,S. (2017) Substituents' effect in electron attachment to epigenetic modifications of cytosine. *J. Chem. Phys.*, **146**, 244314.
40. Sanstead,P.J., Ashwood,B., Dai,Q., He,C. and Tokmakoff,A. (2020) Oxidized derivatives of 5-methylcytosine alter the stability and dehybridization dynamics of duplex DNA. *J. Phys. Chem. B.*, **124**, 1160–1174.
41. Mulder,F.A.A., Mittermaier,A., Hon,B., Dahlquist,F.W. and Kay,L.E. (2001) Studying excited states of proteins by NMR. *Nat. Struct. Mol. Biol.*, **11**, 932–935.
42. Nikolova,E.N., Kim,E., Wise,A.A., Brien,P.J.O., Andricioaei,I. and Al-Hashimi,H.M. (2011) Transient Hoogsteen base pairs in canonical duplex DNA. *Nature*, **470**, 498–502.
43. Tollinger,M., Skrynnikov,N.R., Mulder,F.A.A., Forman-Kay,J.D. and Kay,L.E. (2001) Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.*, **123**, 11341–11352.
44. Marušič,M., Schlagnitweit,J. and Petzold,K. (2019) RNA dynamics by NMR spectroscopy. *ChemBioChem*, **20**, 2685–2710.
45. Zhou,H., Kimsey,I.J., Nikolova,E.N., Sathyamoorthy,B., Grazioli,G., McSally,J., Bai,T., Wunderlich,C.H., Kreutz,C., Andricioaei,I. *et al.* (2016) $m^1$1A and m^1G disrupt A-RNA structure through the intrinsic instability of Hoogsteen base pairs. *Nat. Struct. Mol. Biol.*, **23**, 803–810.
46. Juen,M.A., Wunderlich,C.H., Nußbaumer,F., Tollinger,M., Kontaxis,G., Konrat,R., Hansen,D.F. and Kreutz,C. (2016) Excited states of nucleic acids probed by proton relaxation dispersion NMR spectroscopy. *Angew. Chem. Int. Ed.*, **55**, 12008–12012.
47. Nikolova,E.N., Gottardo,F.L. and Al-Hashimi,H.M. (2012) Probing transient Hoogsteen hydrogen bonds in canonical duplex DNA using NMR relaxation dispersion and single-atom substitution. *J. Am. Chem. Soc.*, **8**, 3667–3670.
48. Dethoff,E.A., Petzold,K., Chugh,J., Casiano-Negroni,A. and Al-Hashimi,H.M. (2012) Visualizing transient low-populated structures of RNA. *Nature*, **491**, 724–728.
49. Vallurupalli,P., Bouvignies,G. and Kay,L.E. (2012) Studying 'Invisible' excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.*, **134**, 8148–8161.
50. Vinogradov,E., Sherry,A.D. and Lenkinski,R.E. (2013) CEST: from basic principles to applications, challenges and opportunities. *J. Magn. Reson.*, **229**, 155–172.
51. Palmer III,A.G. and Koss,H. (2019) In: *Methods in Enzymology*. Springer, Berlin, pp. 177–227.
52. Zhao,B., Baisden,J.T. and Zhang,Q. (2020) Probing excited conformational states of nucleic acids by nitrogen CEST NMR spectroscopy. *J. Magn. Reson.*, **310**, 106642.
53. Delaforge,E., Milles,S., Bouvignies,G., Bouvier,D., Boivin,S., Salvi,N., Maurin,D., Martel,A., Round,A., Lemke,E.A. *et al.* (2015) Large-scale conformational dynamics control H5N1 influenza polymerase PB2 binding to Importin α. *J. Am. Chem. Soc.*, **137**, 15122–15134.
54. Shi,H., Liu,B., Nussbaumer,F., Rangadurai,A., Kreutz,C. and Al-Hashimi,H.M. (2019) NMR chemical exchange measurements reveal that N^6-methyladenosine slows RNA annealing. *J. Am. Chem. Soc.*, **51**, 19988–19993.
55. Schlagnitweit,J., Steiner,E., Karlsson,H. and Petzold,K. (2018) Efficient detection of structure and dynamics in unlabeled RNAs: the SELOPE approach. *Chem. Eur.J.*, **24**, 6067–6070.
56. Schröder,A.S., Steinbacher,J., Steigenberger,B., Gnerlich,F.A., Schiesser,S., Pfaffeneder,T. and Carell,T. (2014) Synthesis of a DNA promoter segment containing all four epigenetic nucleosides: 5-methyl-, 5-hydroxymethyl-, 5-formyl-, and 5-carboxy-2′-deoxycytidine. *Angew. Chem. Int. Ed.*, **53**, 315–318.
57. Lee,W., Tonelli,M. and Markley,J.L. (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, **31**, 1325–1327.
58. Kim,S., Lin,L. and Reid,B.R. (1992) Determination of nucleic acid backbone conformation by proton NMR. *Biochemistry*, **31**, 782–791.
59. Hilal,S.H., Bornander,L.L. and Carreira,L.A. (2005) Hydration equilibrium constants of aldehydes , ketones and quinazolines. *QSAR Comb. Sci.*, **24**, 631–638.
60. Johnson,R.P., Fleming,A.M., Perera,R.T., Burrows,C.J. and White,H.S. (2017) Dynamics of a DNA mismatch site held in confinement discriminate epigenetic modifications of cytosine. *J. Am. Chem. Soc.*, **139**, 2750–2756.
61. Mergny,J. and Lacroix,L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515–537.
62. Dragan,A., Privalov,P. and Crane-Robinson,C. (2019) Thermodynamics of DNA: heat capacity changes on duplex unfolding. *Eur. Biophys.J.*, **48**, 773–779.
63. Lam,S.L., Ip,L.N., Cui,X. and Ho,C.N. (2002) Random coil proton chemical shifts of deoxyribonucleic acids. *J. Biomol. NMR*, **24**, 329–337.
64. Lam,S.L. (2007) DSHIFT: a web server for predicting DNA chemical shifts. *Nucleic Acid Res.*, **35**, W713–W717.
65. Menssen,R.J. and Tokmakoff,A. (2019) Length-dependent melting kinetics of short DNA oligonucleotides using temperature-jump IR spectroscopy. *J. Phys. Chem. B*, **123**, 756–767.
66. Nöltig,B. (2006) In: *Protein Folding Kinetics, Biophysical Methods*. Springer, Berlin.
67. Coman,D. and Russu,I.M. (2005) A nuclear magnetic resonance investigation of the energetics of basepair opening pathways in DNA. *Biophys.J.*, **89**, 3285–3292.
68. Hashimoto,H., Hong,S., Bhagwat,A.S., Zhang,X. and Cheng,X. (2012) Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucleic Acid Res.*, **40**, 10203–10214.
69. Maiti,A. and Drohat,A.C. (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine. *J. Biol. Chem.*, **286**, 35334–35338.
70. Sicard,F., Destainville,N. and Manghi,M. (2015) DNA denaturation bubbles: free-energy landscape and nucleation/closure rates. *J. Chem. Phys.*, **142**, 034903.
71. Leffler,J.E. (1953) Parameters for the description of transition states. *Science*, **117**, 340–341.
72. Coey,C.T., Malik,S.S., Pidugu,L.S., Varney,K.M., Pozharski,E. and Drohat,A.C. (2016) Structural basis of damage recognition by thymine DNA glycosylase: key roles for N-terminal residues. *Nucleic Acids Res.*, **44**, 10248–10258.
73. Maiti,A., Morgan,M.T., Pozharski,E. and Drohat,A.C. (2008) Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. *Proc. Natl. Acad. Sci. U.S.A.*, **26**, 8890–8895.
74. Dodd,T., Yan,C., Kossmann,B.R., Martin,K. and Ivanov,I. (2018) Uncovering universal rules governing the selectivity of the archetypal DNA glycosylase TDG. *Proc. Natl. Acad. Sci.U.S.A.*, **115**, 5974–5979.
75. Drohat,A.C. and Coey,C.T. (2016) Role of base excision 'repair' enzymes in erasing epigenetic marks from DNA. *Chem. Rev.*, **116**, 12711–12729.