Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Epigenetically modified nucleobases: characterization of structure and dynamics

Romeo Cosimo Arrigo Dubini

aus

Mailand, Italien

2022

CONTENTS

LIST OF FIGURES TO THE INTRODUCTION

ABSTRACT

Deoxyribonucleic acid (DNA) is the most fundamental biomolecule of all three domains of life, acting as the premier storage of genetic information for development, functioning, growth and reproduction of all living things. In order to perform its universal and transcendent biological role, DNA needs to be chemically stable in time, and simultaneously capable of conformational flexibility. The thermodynamics and kinetics involving the dynamics of canonical strands have been central to structural biology and biophysical studies throughout the past two decades. As a result, our understanding of such systems has enjoyed a proportional growth. On the other hand, the precise features of epigenetically modified DNA strands remain to date poorly understood.

This dissertation focuses on the characterization of structural and dynamical features of epigenetically modified cytosine derivatives (5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), in particular) via state-of-the-art solution-state Nuclear Magnetic Resonance (NMR) spectroscopy methodologies, allowing for non-invasive and site-specific measurements coupled with well-established physical chemistry frameworks. When dealing with epigenetically modified bases in the context of a double-stranded DNA strand, our approach featured the design and synthesis of palindromic oligomeric systems carrying a single, centrally positioned naturally occurring modification. In order to comment on potential structural features, an initial assessment was achieved via a comprehensive set of homo- (^{1}H-^{1}H Nuclear Overhouser Effect SpectroscopY (NOESY)) and heteronuclear (^{1}H–^{13}C Heteronuclear Single Quantum Coherence (HSQC), ^{1}H–^{15}N band-Selective Optimized-Flip-Angle Short-Transient (SOFAST) Heteronuclear Multiple Quantum Coherence (HMQC)) 2D spectra for resonance assignment and chemical shift analysis purposes. Subsequent studies focusing on ^{1}H isotropic chemical shift relaxation-based approaches yielded a detailed description of the conformational equilibria involving epigenetically-edited DNA on the slower- and faster-intermediate regimes. Further, an original theoretical framework has been developed, capable of extracting thermodynamic and kinetic parameters using exclusively ^{1}H Chemical Exchange Saturation Transfer (CEST)-derived data. The conservation of the

oligomeric sequence and relevant experimental conditions for the study of different bases allowed for complete comparability between distinct epigenetically modified cytosine nucleosides. Extensive discussions revolving around the biological implications of our observations are presented across the enclosed manuscripts, especially with respect to the active demethylation pathway and 5fC, 5caC's hypothesized additional regulatory roles.

ACKNOWLEDGEMENTS

ACRONYMS

NMR Nuclear Magnetic Resonance

UV Ultraviolet absorbance spectroscopy

CD Circular dichroism spectroscopy

smFRET single-molecule fluorescence resonance energy transfer

FT-IR Fourier-transform infrared spectroscopy

MRI Magnetic resonance imaging

PCR Polymerase chain reaction

$R_{1\rho}$ rotating frame relaxation dispersion

GS Ground state

ES Excited state

NOE Nuclear Overhauser effect

SELOPE SELectively Optimized Proton Experiment

NOESY Nuclear Overhouser Effect SpectroscopY

HSQC Heteronuclear Single Quantum Coherence

HMQC Heteronuclear Multiple Quantum Coherence

SOFAST band-Selective Optimized-Flip-Angle Short-Transient

DOSY Diffusion Ordered SpectroscopY

CEST Chemical Exchange Saturation Transfer

CPMG Carr–Purcell–Meiboom–Gill

CSP Chemical shift perturbation

DSS Sodium trimethylsilylpropanesulfonate

TMS Tetramethylsilane

C Cytosine

A Adenine

T Thymine

G   Guanine

5mC 5-methylcytosine

5hmC 5-hydroxymethylcytosine

5fC 5-formylcytosine

5caC 5-carboxylcytosine

DNA Deoxyribonucleic acid

RNA Ribonucleic acid

WC Watson-Crick base pair

HG Hoogsteen base pair

BER Base-excision repair

DNMT DNA methyltransferase

AID Activation-induced cytidine deaminase

TDG Thymine DNA glycosylase

TET Ten-eleven translocation enzyme

ESC Embryonic stem cells

SAM S-adenosyl-L-methionine

KIE Kinetic isotopic effect

EWG Electro-withdrawing group

PUBLISHED AND PRESENTED WORKS LIST

PEER–REVIEWED JOURNAL ARTICLES

Parts of this thesis have been published in academic peer-reviewed journals and/or presented to attendees of conferences, seminars or webinars.

1. <u>R.C.A. Dubini</u>, A. Schön, M. Müller, T. Carell, P. Rovó*; Impact of 5-formylcytosine on the melting kinetics of DNA by ^{1}H NMR chemical exchange, *Nucleic Acids Res.* **2020**, *48* (15), 8796-8807, DOI: doi.org/10.1093/nar/gkaa589.

2. <u>R.C.A. Dubini</u>, E. Korytiaková, T. Schinkel, P. Heinrichs, T. Carell, P. Rovó*; ^{1}H NMR Chemical Exchange Techniques Reveal Local and Global Effects of Oxidized Cytosine Derivatives, *ACS Phys. Chem Au* **2022**, DOI: doi.org/10.1021/acsphyschemau.1c00050.

CONFERENCES AND INVITED PRESENTATIONS

1. **SFB1309 Nikolaus Symposium 2018**, 30th November 2018
   Presented poster with title *"Unravelling the structural transitions of chemically modified DNA and RNA molecules"*

2. **IRTG1309 Summer School 2019**, 15th–19th September 2019
   Presented poster with title *"5fC in action: an NMR investigation"*

3. **Center for NanoScience (CeNS)/CRC235 Workshop "Evolving Nanosciences"**, 23rd–27th September 2019
   Presented poster with title *"5fC in motion: uncharted free-energy landscapes"*

4. **SFB1309 Nikolaus Symposium 2019**, 6th December 2019
   Presented poster with title *"5fC in motion: uncharted free-energy landscapes"*

5. **SFB1309 Nikolaus Symposium 2020**, 7th December 2020
   Presented poster with title *"Epigenetic DNA and RNA modifications by ^1H NMR"*

6. **IRTG 1309 Science Meeting (AK Rovó edition)**, 31st March 2021
   Presented keynote with title *"Epigenetic DNA and RNA modifications by ^1H NMR"*

7. **SFB 1309 Virtual Summer Symposium 2021**, 1st July 2021

8. **ICMRBS Early Career Researcher Webinar Series**, 14th July 2021
   Presented invited talk with title *"Dynamics of epigenetic DNA modifications by ^1H NMR"*, available on YouTube

9. **IRTG 1309 Wrap-Up event**, 16th–17th February 2022
   Presented invited talk with title *"Epigenetic DNA modifications by ^1H NMR"*

OTHER ACHIEVEMENTS

The following papers have been published in academic peer-reviewed journals, but fall outside the scope of this work.

1. A. Menke, R.C.A. Dubini, P. Mayer, P. Rovó, L.J. Daumann*; Formation of Cisplatin Adducts with the Epigenetically Relevant Nucleobase 5-Methylcytosine, *Eur. J. Inorg. Chem.* **2021**, *1*, 30-36, DOI: doi.org/10.1002/ejic.202000898

2. S.J. Cox, D.C. Rodriguez Camargo, Y.-H. Lee, R.C.A. Dubini, P. Rovó, M.I. Ivanova, V. Padmini, B. Reif, A. Ramamoorthy*; Small molecule induced toxic human-IAPP species characterized by NMR, *Chem. Commun.* **2020**, *56*, 86, 13129-13132, DOI: doi.org/10.1039/D0CC04803H

3. R.C.A. Dubini, H. Jung, C.H. Skidmore, M.C. Demirel*, P. Rovó*; Hydration-induced structural transitions in biomimetic tandem repeat proteins, *J. Phys. Chem. B* **2021**, *125*, 8, 2134-2145, DOI: doi.org/10.1021/acs.jpcb.0c11505

1 | INTRODUCTION

1.1 PREFACE

In this chapter, a concise yet comprehensive summary of structural biology, epigenetic regulation mechanisms and bioNMR spectroscopy concepts and applications is offered.

Section 1.2 focuses on the history and the main features of DNA. The spotlight is on dynamics rather than structure. Intermediate time scale motions and their relevance with respect to biologically relevant phenomena are discussed. Particular attention is devoted to melting, annealing and base-pair dynamics, which are central concepts to this work.

Section 1.3 dwells on the mechanisms and actors of the epigenetic regulation of the genome. Here, the emphasis is placed on Cytosine (C) chemically-modified DNA bases including their origin and their interactions with key enzymes. The active demethylation pathway is then discussed, introducing 5-methylcytosine (5mC)'s oxidized derivatives, specifically 5fC and 5caC. Lastly, a compendium of relevant work in the fields of structural biology and biophysical characterization of epigenetically-modified DNA is presented.

Section 1.4 aims at introducing essential NMR concepts and applied state-of-the-art methodologies for nucleic acids' chemical shift assignment and thereafter characterization of potential conformational exchange phenomena. Special attention is placed on experiments suitable for the study of natural-abundance samples.

1.2 DNA STRUCTURE AND DYNAMICS

The following sections contain considerations on conformational aspects and enzymatic processes involving the most iconic molecule throughout the field of biology. Section 1.2.1 summarizes DNA history and fundamental structural features. Section 1.2.2 introduces the concepts of dynamical landscapes and high-level overview of higher-energy conformational states in DNA. Special care is devoted to describing the annealing and melting processes and examples of base-pair dynamics.

1.2.1 Brief history and general structural features

DNA is the fundamental genetic information carrier, as it dictates the instructions for the development, functioning, growth and reproduction of all known organisms and many viruses. Identified during the 19th century by Swiss physician Friedrich Miescher, it was not isolated until 1878. [1] Throughout the 20th century, DNA research has been one of the central themes across science, mainly in chemistry, medicine, biology and physics. Between 1909 and 1933, several advances narrowed down its composition to nitrogenous bases, deoxyribose sugar and phosphate linkages, in addition to identifying its essential role as a genetic information carrier. [2] It was not until 1937, although, that DNA was reported to present a defined average structure. [3] This revelation spurred intense research efforts in order to more accurately characterize such architecture, resulting in the iconic double helix discovery by Wilkins, Franklin, Watson and Crick. [4] This development, regarded as a cornerstone moment in the history of scientific progress and effectively marking the dawn of molecular biology, ultimately led to Francis Crick's lecture where he first formulated its "central dogma" concept, correlating DNA to Ribonucleic acid (RNA) and finally proteins.

dsDNA consists of two biopolymeric strands which are in turn composed of four distinct nucleotides: Guanine (G) and Adenine (A) (the purines), and C and Thymine (T) (the pyrimidines). DNA nucleotides are made up of a deoxyribose ring, a phosphate backbone (linking two sugar molecules together and forming a chain) and a heterocycle (which is the characteristic portion of a nucleoside), also called the nucleobase or the nitrogenous base.

The X-ray diffraction data reported by Crick *et al.* in ref. 4 first revealed the archetypal double-helix structure, which was later on identified as the most fre-
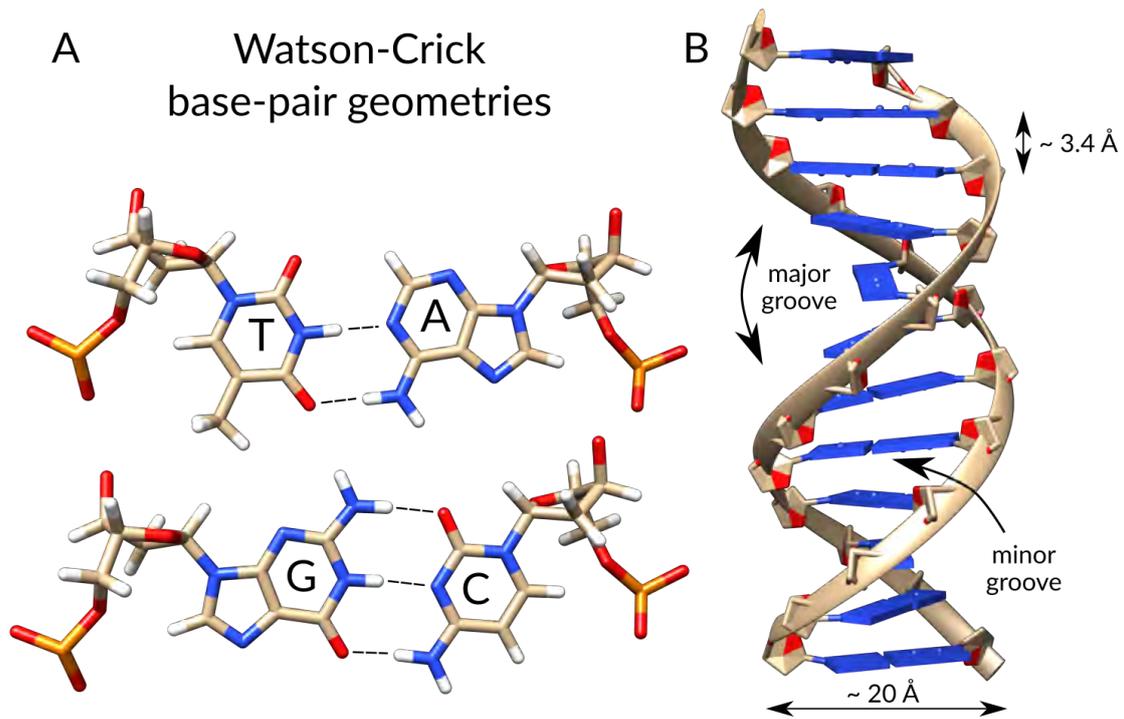
Figure 1.1: (A) Watson-Crick base pair (WC) canonical base pairing geometries for A:T (upper) and G:C (lower). (B) General structural representation of B-DNA. Key features as major and minor grooves, together with interbase distance and the total diameter, are highlighted in the figure.

quent double-strand configuration, nowadays commonly designated as B-DNA. Compositionally, the main features of B-DNA can be summarized using Chargaff's first and second parity rules, stating that a purine is ordinarily paired with a pyrimidine, *i.e.* the number of Thymine (Cytosine) nucleosides is equal to that of Adenosine (Guanine), both within double- and single-stranded DNA, respectively. [5] Such empirical observations found statistical confirmations four decades later by studying all sequences longer than 50000 nucleotides in the genetic sequence database. [6] From a structural perspective, B-DNA strands are arranged coiling around the same helical axis by interweaving with one another, a pattern normally related to as "plectonemic coil" (Figure 1.1B). The geometry of the intertwined strands generates two asymmetric apertures, characterized by different width and depth. Such grooves, a major and a minor groove, spiral around an imaginary axis and play key roles in the structure and conformational stability of B-DNA, with many reader proteins or enzymes selectively recognizing either of the two through a system of specific sets of H-bond acceptors and donors. One helical turn of B-DNA contains about 10.5 base pairs, concealed within the helix and arranged perpendicularly to the reference axis propagating

along the double-helical structure. In B-DNA, the distance between one base and the following (or preceding) one across the sequence is about 3.4 Å. Apart from such general and averaged-out features, B-DNA is in reality not entirely uniform and its exact structure and stability are tightly linked to the exact sequence of interest. For instance, G:C-rich B-DNA strands are characterized by higher melting points, as a result of the base-pair geometry involving three stable hydrogen bonds. [7]

Albeit B-DNA is by far the most common configuration, especially at neutral pH and physiological salt concentrations, less ordinary environments can energetically favour the formation of other right-handed helix conformations (A-DNA) or even left-handed ones (Z-DNA). [8, 9] Despite the proven biological relevance of these structural elements, such arrangements fall outside the scope of the present work and will not be further discussed.

1.2.2 The dynamical landscape of DNA

The discovery of the iconic B-DNA structure reported by Watson and Crick in 1953 went down in history as one of the most disruptive scientific accomplishments of the 20th century. While the finding was of monumental importance on its own, it also represented a starting point for several decades of intense research on nucleic acids involving further structural elucidation and, perhaps most importantly, the unraveling of its dynamical features. Indeed, DNA roles and functions cannot be enacted altogether by a unique, single static structure. Conversely, a high degree of conformational freedom is not only useful, but rather a strict requirement in order to allow for productive interactions to occur. As a result, considerable research efforts went into the elucidation of DNA annealing and melting, and more broadly into its dynamics and kinetics. This specific research field has enjoyed significant advances in the past few decades, as increasingly precise investigations made extensive use of both established and novel biophysical methodologies and techniques, such as Ultraviolet absorbance spectroscopy (UV) [10, 11], single-molecule fluorescence resonance energy transfer (smFRET) [12], imino proton exchange [13, 14] and chemical exchange relaxation NMR spectroscopy. [15, 16]

DNA double helices, similarly to most proteins and RNA molecules, are intrinsically dynamic entities and display a vast array of conformational freedom on both fast and slow timescales. In fact, the feature of deformability is a key charac-

teristic for the execution of DNA biological functions. [17] Macroscopically, torsional stress and supercoiling are crucial factors in transcription, chromosome architecture, DNA replication and its regulation. [18, 19] Similarly, motions on the microscopic scale such as DNA breathing or single base extrusions are essential in several recognition processes. [20]

Below, a synopsis of the annealing and melting process, and base-pair dynamics is presented. It is to be noted that other typical DNA conformational fluctuations, such as breathing, although relevant for gene expression by affecting binding and assembly of molecular machinery, is beyond the scope of this particular body of work, and will not be considered further. [21]

1.2.2.1 *Annealing and melting*

The intermediate and slow-intermediate (ranging from 10^{-3} to 10^0 s) time frame is a particularly important one for macroscopic processes such as DNA hybridization and melting. The former is the process of establishing non-covalent and base-specific bonds between two nucleic acids' strands, resulting into a single complex, a double strand (generally referred to as a "duplex"). The latter (also referred to as DNA denaturation) entails the unwinding of a duplex into its elementary constituents, *i.e.* into single-stranded polymeric entities, through the breaking of base-pairing patterns and hydrophobic stacking attractions between the bases. [22]

A number of core molecular biology concepts, including but not limited to DNA replication, transcription, Polymerase chain reaction (PCR) and DNA sequencing, are fundamentally reliant on the concepts of DNA hybridization and melting. For this reason, it is perhaps unsurprising that pioneering kinetics and dynamics studies focused on defining the features of these processes. [23–26] Below, the essential stages in the nucleation process are discussed.

BROWNIAN ENCOUNTER AND NUCLEATION   Assume two single DNA strands are diffusing in solution via temperature-induced, stochastic Brownian motions. In order for the annealing event to start, a diffusive encounter leading to nucleation is needed, first and foremost. Such an event is defined by a minimal set of hydrogen bond connectivity between the two random-coil strands (Fig. 1.2A). If this first, embryonic encounter event is sufficiently long-lived, more hydrogen-bonding events may subsequently occur (Fig. 1.2B).

ZIPPERING AND DUPLEX FORMATION An increasing amount of metastable hydrogen bonds leads to a higher likelihood of progressing towards zippering (Fig. 1.2C-D). Zippering happens as bases from random-coil, single-stranded DNA molecules successively pair onto the growing double helix. Once an end is reached, zippering towards the other end begins, ultimately forming the duplex (Fig. 1.2E). The thermodynamics and kinetics of DNA duplex melting and annealing are discussed later on, in Chap. 2.
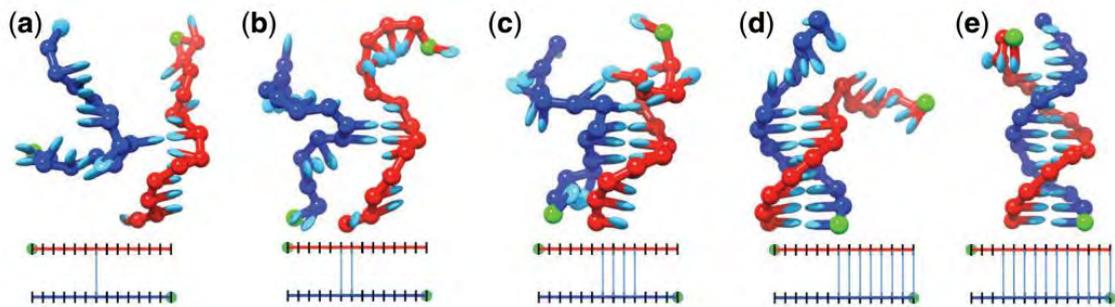
Figure 1.2: Adapted and reproduced with permission from Ouldridge *et al.* (Ref. 27): Exemplary stages of DNA hybridization. (a) Brownian motion-induced diffusive encounter, the incipient intermolecular hydrogen bond is established. (b-e) Successive H-bonds formation and zipping mechanism. Green dots indicate the 5'-end of each strand. Schematic diagrams underneath indicate the current state of the base pairs present in the system.

Conversely, the process of DNA melting essentially consists of progressive unzipping and eventual separation of the two strands. Base pairs positioned at the extremities of DNA molecules in solution are known to undergo frequent yet transient opening events (a phenomenon known as "fraying"), which can extend further across the sequence if sufficient energy is provided. [28] In vitro, dissociation can be facilitated by mainly two factors, namely either by pH-induced base ionization (disrupting the hydrogen-bond network), or heating the DNA-containing solution. Earlier investigations helped establishing the notion that the association process is dependent on concentration, highlighting a second order process which follows an anti-Arrhenius behaviour (*i.e.* inversely temperature-dependent). Conversely, the dissociation process fits to an Arrhenius-type behaviour (directly temperature-dependent). [29]

1.2.2.2 *Base-pair dynamics*

The double-helical design of B-DNA addresses a specific need for a structure whose main purpose is ensuring conformational and chemical stability. It may

not be surprising then to ascertain that evolution selected an architecture where the carrier of the genetic information, stored in the form of long sequence of alternating nucleobases, is placed at the very center of the helix, where it is protected from the solvent and other potentially interacting molecules. [30] Despite this strategy of concealing, section 1.2.2.1 detailed how the sturdy and robust B-DNA arrangement can indeed be subject to fraying and eventually melting, opening up the possibility for duplication events or interaction with other molecules. [31] Apart from melting and annealing dynamics, mobility and conformational pliancy at the base-pair level are essential aspects or nucleic acids chemistry, physics, biology and biotechnology. Indeed, indispensable DNA functions such as transcription, recombination and repair, together with several protein recognition and enzymatic events, depend entirely on such motions. In the past decade, a number of seminal contributions have shed light on the mechanism of base-pair opening and its broader biological implications. Colizzi *et al.* focused on the role of base extrusion in the context of double-helix unwinding initiation, identifying asymmetric base-pair dynamics as highly sequence and composition dependent (specifically with respect to its interactions with structure-remodelling enzymes) as opening of a base-pair follows a stepwise mechanism whose directionality is modulated by the specific nature of the sequence. [32] Indeed, they found that helicases can more readily unwind DNA strands with pyrimidine-rich sequences, highlighting how both composition and orientation of a target DNA sequence may influence gene regulation. Other studies focused on base-pair geometries alternatives to the classical, WC configuration. Such an arrangement is the Hoogsteen base pair (HG) (Fig. 1.3), which has been increasingly attracting the attention of the scientific community following its discovery in 1963. [33]

Hoogsteen base pairs are transiently and sparsely populated conformations that exist in thermal equilibrium with standard Watson–Crick base pairs and are believed to be crucially important for DNA-protein interactions and larger scale remodelling such as a B-DNA to Z-DNA transition. [34, 35] The past decade has witnessed several breakthrough contributions towards the characterization of such rare and ephemeral conformers. Using state-of-the-art bioNMR techniques, the Al-Hashimi group has achieved unprecedented structural, kinetic and thermodynamical precision describing the WC-HG equilibrium and transitions, information that revealed itself to be crucial in detailing HG base-pairs biological relevance. In a breakthrough paper, Nikolova *et al.* elaborate on how DNA is naturally prone to preexisting WC-to-HG equilibrium, with the HG conformation representing an energetically competent alternative to canonical base-
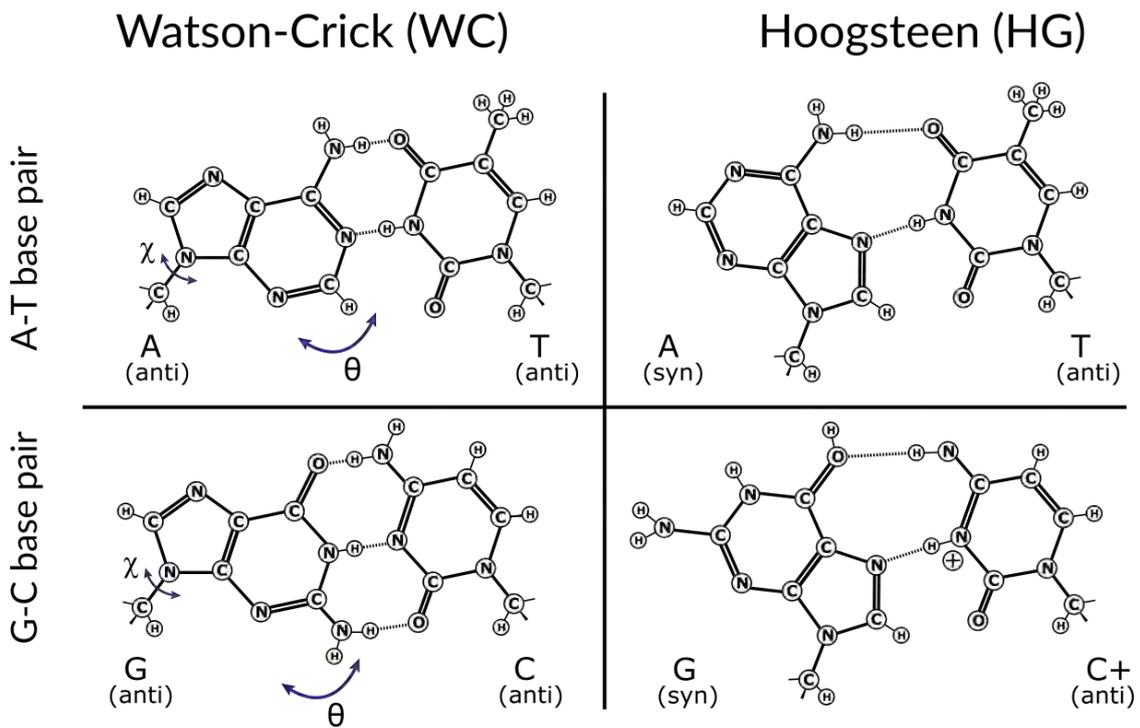
Figure 1.3: Chemical structures for WC (left) and HG (right) A:T (upper) and G:C$^+$ (lower) base pairs. Conformational switching is enabled by the simultaneous rotation of the base around the glycosidic bond (χ) and a base-flipping event (θ) (blue arrows).

pairing. [15] Such equilibrium expands the functional diversity of WC base pairs and is selectively recognized by a number of transcription factors and repair enzymes. [36, 37]

The existence of non-canonical DNA bases, *i.e.* chemically distinct moieties of the four canonical nucleobases, adds a further layer of complexity to the branch of DNA dynamics. The origin and role of these modifications are the focus of the following section, where I will be dwelling specifically on naturally occurring and biologically relevant cytosine modifications.

1.3 EPIGENETIC REGULATION OF THE GENOME

Epigenetics is a branch of biology that studies the modifications influencing gene expression and any heritable phenotypic change as a result of environmental factors and/or ordinary cellular development. [38] More specifically, the field of epigenetics is concerned with the mechanisms that act "on top" of the standard genetic bio-machinery. [39] In the following sections, I will discuss the foundational elements of this rapidly developing branch of biology in order to gain a better appreciation of the context which inspired the present work. Section 1.3.1 focuses on the description of DNA methylation and its role in silencing gene expression. Section 1.3.2 aims at introducing passive and active demethylation pathways. Section 1.3.3 discusses 5fC and 5caC potential involvement as semi-permanent standalone epigenetic markers, while section 1.3.4 focuses on in their impact on dsDNA structure and kinetics.

1.3.1 DNA methylation

The genome is composed of only four building blocks, and all cells in multicellular organisms share an identical set of genetic material. Despite this apparent yet deceptive redundancy, specific cells may perform entirely different tasks from one another. For instance, a keratinocyte (found in the outmost layer of the epidermis) is a type of cell which produces molecules to contain potentially dangerous radiation, heat and physical damage. DNA repair enzymes are also in place to rectify eventual genomic deterioration. [40]

On the other hand, a neuron is capable of receiving and emitting electric pulses in order to achieve synaptic communication. How do these cells perform such different tasks while still sharing an identical set of genetic instructions? They do so by *manifesting* a specific subset of active genes. [41] During cellular development, cells are directed to a specialization pathway by silencing unnecessary portions of the genome, and achieve this by cytosine methylation through the action of DNMT enzymes (Fig. 1.4). [42]

DNMT1 and DNMT3 (and their isoforms) are the most abundant DNA methyltransferase enzymes in humans and mammals in general. [43] Their purpose is to methylate canonical cytosine nucleotides within CpG islands (*i.e.*, portions of the genome exhibiting a high abundance of 5'-CG-3' repeats) at the C5 position. The mechanism for molecular recognition and enzymatic action of the
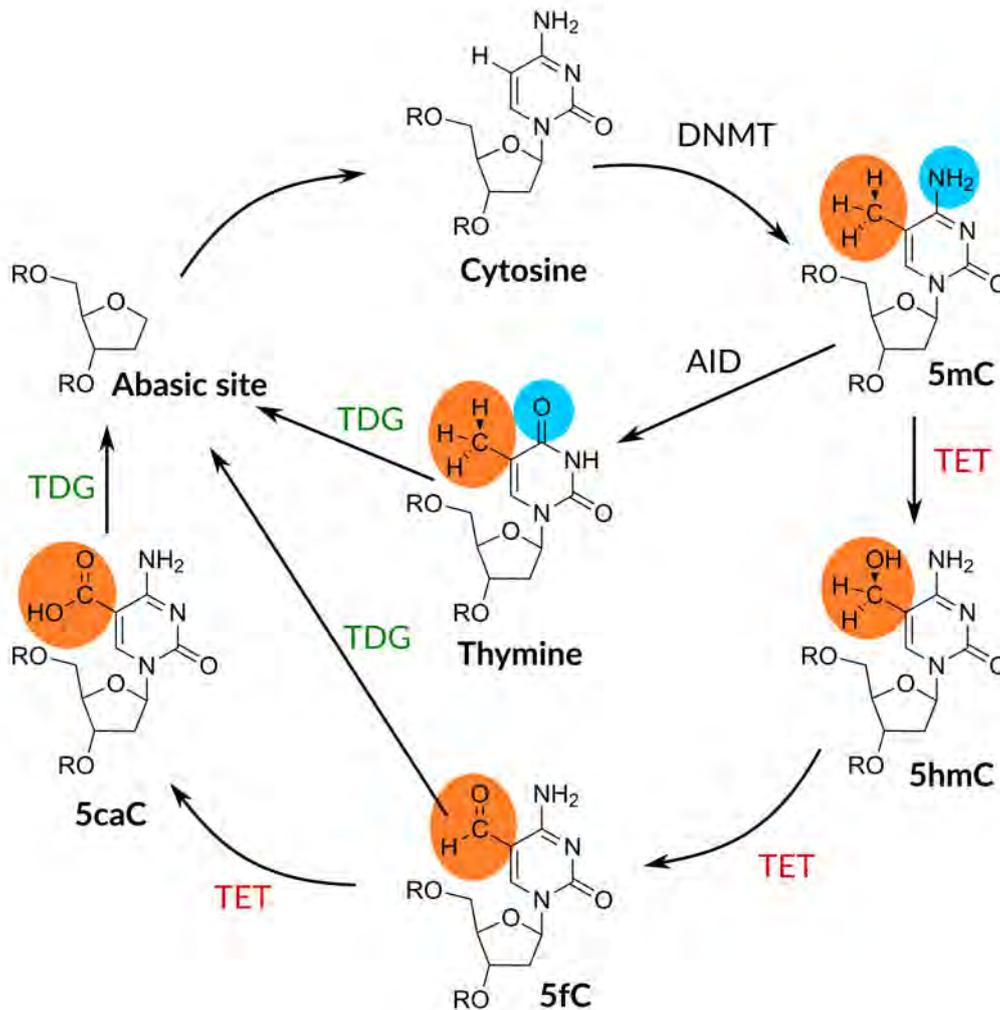
Figure 1.4: The cycle of active DNA demethylation. DNA methyltransferase (DNMT) initially methylates canonical C at the carbon C5. 5mC is then prone to either a deamination process via Activation-induced cytidine deaminase (AID) enzymes, or subsequent oxydation steps via Ten-eleven translocation enzyme (TET). Canonical T, 5fC and 5caC are substrates for Thymine DNA glycosylase (TDG) enzymes, producing an abasic site which can then be repaired though Base-excision repair (BER) cycle.

human variant of DNMT1, shown in Fig. 1.5), is based on the presence of a C on one strand and CpG on the counter-strand in duplex-DNA. [44] Cytosine's C5 and C6 positions are located inside the major groove of B-DNA, where reduced steric hinderance for a nucleophilic attack favours the reaction. After recognition of a flipped-out cytosine residue, the consensus mechanism proceeds by oxidation of the C6 position via Cys1226, yielding the first reaction intermediate. Subsequently, methyl transfer from S-adenosyl-L-methionine (SAM) to cytosine's C5 and proton abstraction yields 5mC.
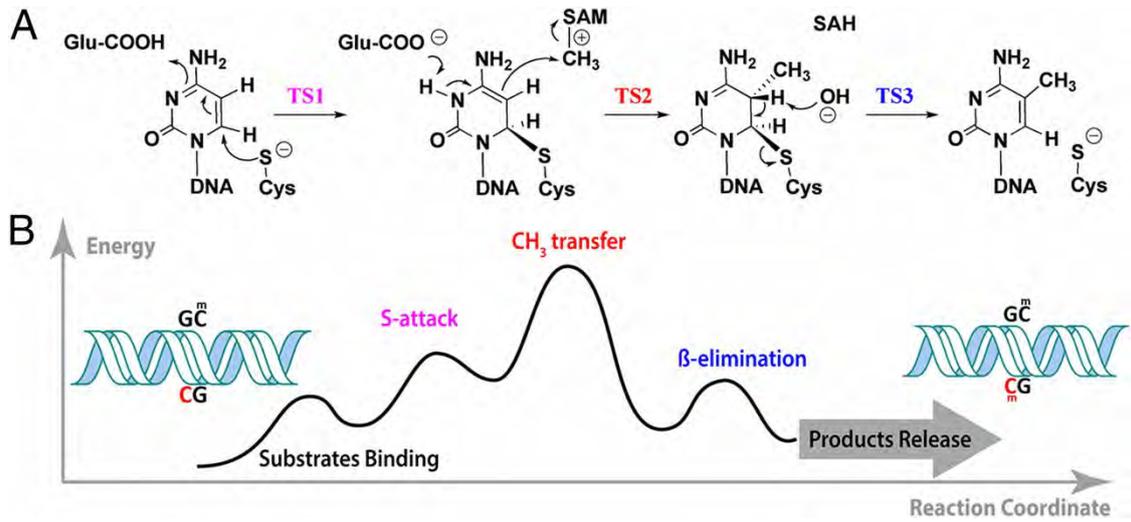
Figure 1.5: Adapted and reproduced with permission from Du *et al.* (Ref. 44). DNA methylation catalyzed by DNMT1. (A) Consensus catalytic mechanism for DNMT1. (B) Kinetic isotopic effect (KIE) analysis for the determination of the chemically rate-limiting step for DNMT1. The methyl transfer has a higher energy barrier than the thiol-attack and β-elimination steps.

The resulting methylation patterns are essential for determining cell proliferation, differentiation, individual development and gene expression. In other terms, 5mC dynamics involving the genome are responsible for the eventual fate of a differentiating cell and its further development and life cycle.

1.3.2 DNA demethylation pathways

While cytosine methylation results in a covalent, chemically stable modification, the reaction is reversible, and methylation levels are highly dynamic. [45] The reverse process, DNA demethylation, removes the methyl group from 5mC in DNA. DNA demethylation can undergo two distinct and independent mechanisms: a replication-based "passive" pathway, and enzyme-dependent "active" one. The former occurs by cell replication and subsequent dilution of methylation levels, resulting in a lack of inhibitory pressure on gene expression. On the other hand, the active mechanism can occur in both dividing and non-dividing cells, and requires a cascade of enzymatic-driven reactions involving a number of 5mC derivatives. An overview is offered in Fig. 1.4. [46, 47]

Passive DNA demethylation can be assimilated to a replication-driven dilution process in the absence of functional DNA methylation maintenance machinery. While its dynamics are well established and a consensus has been reached within

the scientific community regarding the passive mechanism, the evidence for how the active process occurs is far from unison and uncertainty remains surrounding the structural details of the methyl moiety removal. [48, 49] The present work dwells on naturally-occurring oxidized derivatives of 5-methylcytosine originating from the active demethylation pathway.

1.3.2.1 *Active DNA demethylation pathway*

The discovery of 5-hydroxymethylcytosine (5hmC) and its interplay with TET enzymes has been one of the key breakthrough discoveries in the investigation of the active demethylation pathway. In 2009, Tahiliani *et al.* found that over-expression of TET1 was linked to a depletion of methylated DNA. [50] Shortly after, purified TET enzymes were shown to modify 5mC-edited oligonucleotide substrates yielding 5hmC. [51] These seminal reports represented formidable starting points for a number of studies concerning the quantification of 5-hydroxymethylcytosine, and established this 5mC-oxidized derivative as the *de facto* 6th base of the genome. [52, 53] Further research highlighted that 5hmC is not the only possible product of TET-induced oxidation of methylated cytosine. In fact, novel chromatographic approaches in tandem with highly sensitive mass spectrometry methodologies allowed for the discovery of 5fC and 5caC in the context of TET overexpression and most importantly Embryonic stem cells (ESC). [54] Subsequent to the discovery of 5fC and 5caC in their capacity as TET-derived products, biologists started pondering about their potential role within DNA demethylation and beyond. The discovery that TDG is required for embryonic development, coupled with the recognition that some artificially modified cytosine bases could also be substrates for TDG, motivated a stint of intense biochemical research linking TDG with 5mC oxidized derivatives. [55, 56] While the interplay between 5mC and 5hmC, the role of TET, the *in vivo* equilibrium reached by active and passive demethylation mechanisms and the role of enzymatic activity in stem cells development and disease are all fundamentally important research items, for the purposes of introducing the research object of this thesis I will now focus on the two subsequent products of 5hmC oxidation.

1.3.3 5-formyl- and 5-carboxylcytosine's roles

The connection between the two oxidized 5hmC derivatives and TDG enzymes came soon afterwards their discovery. [57, 58] As a result, their role as TDG sub-

strates in completing the removal of genomic 5mC yielding the abasic site was soon recognized. [59] Further research on the genomic locations that present the highest concentration of 5fC and 5caC showed that dynamic TET/TDG-mediated 5hmC oxidation and demethylation could be used to maintain those key gene elements in an inactive or active state. [60] While 5fC and 5caC as transient intermediates within the active demethylation pathway is a widely accepted concept, demonstrated by their low levels even after TDG depletion, more recent research has entertained the possibility that they could serve as semi-permanent modification in their own right. [61] For instance, two independent studies led by Iurlaro and Spruijt showed how 5fC is recognized by more protein readers compared to its methylated and hydroxymethylated precursors. [62, 63] Further, Kellinger *et al.* concluded that both 5fC and 5caC reduce the rate and substrate specificity of RNA polymerase II transcription. [64] The identification and description of these yet undefined roles has been a sought-after research avenue and inspired a number of contributions from the structural biology and physical chemistry areas, which will be introduced and reviewed next.

1.3.4   Structural biology and biophysics of cytosine's epigenetic modifications

The characterization of the fundamental biophysical aspects of proteins and nucleic acids has gained considerable momentum since the advent of high-resolution methodologies and continues to provide indispensable insights for our ever-improving understanding of life and related processes. DNA oligomers carrying naturally occurring, epigenetically-relevant cytosine derivatives represent no exception, as they attracted considerable research interest since their discovery. In this paragraph I will be reviewing some of the developments that proceeded and inspired the work presented hereby.

1.3.4.1   *Impact on structure*

The impact of 5fC and 5caC on DNA structure and propensity to melt and anneal has been a debated topic in the scientific community for the better half of the past decade. In 2015, Raiber *et al.* reported compelling insights into how 5fC can affect the structure of DNA at the molecular level. [65] Their claims, which are substantiated by UV/Vis and Circular dichroism spectroscopy (CD) data combined with X-ray crystallography (Fig. 1.6), are that 5fC does not affect the thermodynamic stability of unmodified CpG repeat–containing oligomers. Instead, 5fC

drives DNA structures to a new distinct conformation, dubbed F-DNA, characterized by helical underwinding. The proposed biological implications would be that F-DNA may directly impact the interaction between 5fC and reader protein at formylated sites of the genome by recognition of the localized, distorted DNA geometry rather than the modified base itself.
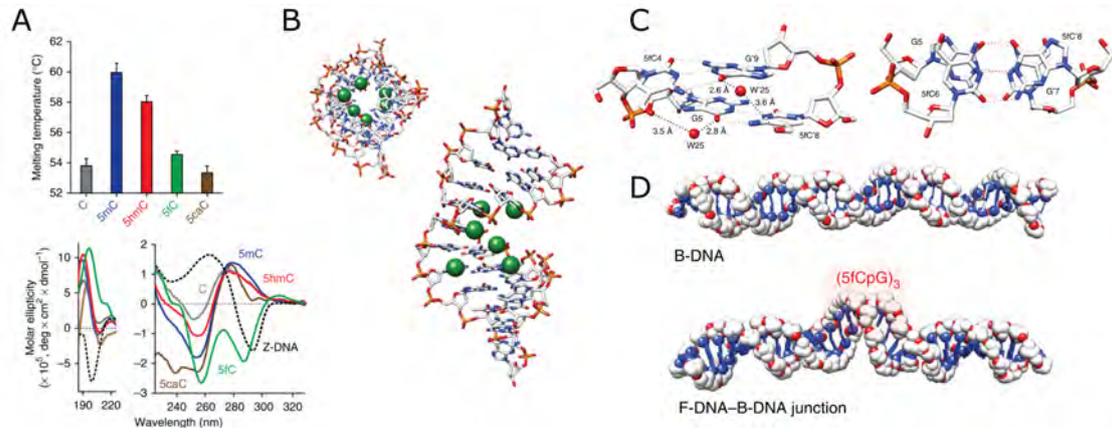
Figure 1.6: Adapted and reproduced with permission from Raiber *et al.* (Ref. 65). (A) Spectroscopic analysis of epigenetically modified oligomers strands. UV/Vis melting profiles show that the differences in melting temperature of 5fC, 5caC and C-containing dsDNA are not statistically significant. CD spectroscopy data shows a distinct spectroscopic signature associated with 5fC-containing oligonucleotides. (B) Crystal structure overview of a 5fC-containing dodecamer: formyl groups (green spheres) are solvent exposed in the major groove. (C) Zoom-in of the crystal structure showing base pair stacking between 5fC4-G'9 and G5-5fC'8, showing considerable overlap between the π-systems of 5fC and guanines resulting in the kinked geometry of the proposed F-DNA arrangement. (D) Model of a longer DNA oligomer highlighting the twisted helical trajectory by local variation of the grooves in F-DNA.

In 2017, Hardwick and coworkers reported a comparative study deploying NMR and X-ray crystallography methodologies. [66] The investigation, in sharp contrast, acknowledged that formylated cytosine has no significant influence on conformation in the crystal state, while attributing minor geometric alterations to the different oligomeric sequences under consideration.

A similar literature trend can be ascertained for 5caC, whose influence on DNA structure is reported to be either insignificant or extensive according to different studies. [67, 68] Given the impasse reached by consideration of structural elements alone, the identification of stabilization and destabilization effects, including that of potential asymmetric base-pair openings, gained the interest of structural biologists and biophysicists.

1.3.4.2 *Impact on stability and kinetics*

Considerable progress has been achieved over the past few years in the characterization of the physico-chemical influence of these modifications on dsDNA kinetics and dynamics. For instance, it has been shown that the incorporation of 5fC or 5hmC makes DNA more flexible, while 5mC stiffens it and 5caC was deemed almost indistinguishable from canonical C. [12] 5fC and 5caC were identified as responsible for the alteration of the stability of the X:G base-pair due to their Electro-withdrawing group (EWG) effect on the pKa of the N1 nitrogen. [69] Subsequently, by leveraging 2D transient temperature jump Fourier-transform infrared spectroscopy (FT-IR), the Tokmakoff laboratory determined and quantified the impact of oxidized 5mC derivatives by considering the dehybridization process. [10, 11] They found that incorporation of 5fC at any pH and 5caC in acidic conditions are linked to a decrease in melting cooperativity and a lowering of the free energy barrier to dissociation. In disagreement with previous reports, further analysis by NMR spectroscopy highlighted that 5mC may actually favour dsDNA melting by increasing the 5mC:G base-pair opening kinetic rate. [70] The same report argues that CpG repeats substituted asymmetrically may have a different, and occasionally opposite impact on DNA melting and base-pair stability when compared to their symmetric homologues.

1.4 NUCLEIC ACIDS NMR SPECTROSCOPY

Since its infancy, the field of NMR spectroscopy has enjoyed a steady and continuous development from both a methodological and engineering standpoint. [71, 72] In the late 90s and early 2000s, a confluence of advancements in NMR hardware and software technology, molecular biology tools and pulse sequence development made the investigations of structural features and dynamical phenomena in biomolecules suitable in solution-state and, later on, in solid-state applications. [73, 74] While proteins have been (and remain) the main focus in the last three decades, investigations revolving around nucleic acids have more recently been assuming an increasingly prominent role in the field. [16, 75]

Below, I will introduce a portion of the fundamental aspects of NMR spectroscopy, with a focus on phenomenological and experimental features of the methodologies employed in the present work, together with a number of notable applications. Section 1.4.1 was designed to provide a solid theoretical ground to understand the content of the following chapters. The spotlight is on the introduction of concepts such as the Larmor frequency, the vector model, NMR pulses, chemical shifts, spin relaxation and chemical exchange phenomena, which are central to appreciating the studies presented in Chapters 2 and 3. Sections 1.4.2 and 1.4.3 provide insight into the resonance assignment methods and common subsequent procedures, which are the first and foremost steps of a typical nucleic acids NMR study. Last, Section 1.4.4 builds upon the previous segments to present the necessary concepts and experiments required to detect and characterize the presence of sparsely and dynamically populated conformational ensembles of DNA, with a special focus on the microseconds-to-seconds regime.

1.4.1 Theoretical elements

NMR spectroscopy is a far-reaching analytical technique originating from the domains of electromagnetism and quantum physics. [76] Amongst the plethora of individual modern applications, it is notably suitable for the determination of molecular conformation, as well as studying physical properties at the molecular level and with atomic precision. This feature motivated its gradual expansion to synthetic chemistry first and then biology, designating NMR as a standard methodology in the characterization of simple and complex molecules. [77] In this section, I will be introducing the necessary concepts for the appreciation

of subsequent experimental work employing the traditional semi-classical approach. [78] Unless explicitly referenced otherwise, the following sections are based on Refs. 76, 77 and 78.

1.4.1.1   *The Larmor frequency*

All forms of spectroscopy require two or more distinct energy states between which transitions can be accommodated by the means of absorption or emission of a photon. NMR is no exception, as the phenomenon itself arises from the interaction of nuclear magnetic dipole moments with a strong external magnetic field (from now on referred to as B_0) in the order of several Tesla, coupled with bursts of radio frequency radiations. In turn, the dipole moment is a byproduct of a quantum-mechanical property termed spin angular momentum. The spin angular momentum S depends on the spin quantum number I according to the following relationship:

$$\vec{S} = \hbar\sqrt{I(I+1)} \tag{1.1}$$

where $\hbar$ is the reduced Planck constant. I is property of central importance in NMR spectroscopy and depends on the exact nuclear composition in terms of protons and neutrons. Isotopes with an odd mass number (such as ^{1}H, ^{13}C and ^{15}N) have $I = \frac{1}{2}$. On the other hand, isotopes with even mass number, but odd charge have $I = n$ where n is an integer. Isotopes with spin quantum number $\neq$ ½ and their features fall outside the scope of this work and will not be discussed further. The roots of eq. 1.1 for spin one-half nuclei identifying the value of the z-component of the angular moment S_z are $+\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$. The relative energy of these two states depends on the interaction between the dipoles and an external magnetic field, and can be quantified through the Hamiltonian (a quantum mechanical operator used to determine the energy of a given system). When an ensemble of nuclear spins (a molecule, for instance) is observed outside of a magnetic field, any possible orientation of a magnetic dipole is equal in energy, *i.e.* their energetic levels are degenerate. However, the situation differs when said molecule is exposed to a magnetic field, such as those generated by NMR spectrometers. In the latter conditions, the Hamiltonian dictates that:

$$H = -\gamma\hbar B_0 m_z \tag{1.2}$$

where $m_z \propto S_z$ and γ is an important isotope-specific constant (the gyromagnetic ratio). Application of eq. 1.2 identifies two energy levels for a given spin-½ nucleus as a function of the external magnetic field:

$$E = \pm \frac{\gamma \hbar B_0}{2} \tag{1.3}$$

where the positive value is generally referred to as α state and the negative one as β state. An important consequence is that the energy difference between these two states is:

$$\Delta E = E_\beta - E_\alpha = \gamma \hbar B_0 \tag{1.4}$$

and, by application of Planck's equation

$$E = \hbar \omega \tag{1.5}$$

one obtains the most fundamental relationship of NMR spectroscopy, the Larmor frequency equation:

$$\omega = \gamma B_0 \tag{1.6}$$

The relationship above identifies ω as the observed resonance frequency of the transition between two energy levels of a given nucleus when exposed to a particular magnetic field. In other words, the absorption frequency of an $\alpha \longrightarrow \beta$ transition is directly and linearly proportional to the magnetic field strength that the nucleus perceives. The difference in energy ΔE has the important consequence of determining a slight (order to 10^{-6}) inequality in population of the states α (lower energy) and β (higher energy). This minute difference, quantitatively definable through the Boltzmann relationship, constitutes the so-called "bulk magnetization" (M), and represents the imperative starting condition to any conceivable NMR experiment.

1.4.1.2 *The vector model*

The frequency of transitions between lower and higher energy spin states in nuclei is approximately nine orders of magnitude smaller compared to electronic ones, which constitute the theoretical underpinnings of techniques such as UV/Vis and FT-IR spectroscopy. By application of Einstein's law for spontaneous emission, the interesting consequence is that, in the nuclei, the lifetime of a higher energy state is sufficiently long to yield the characteristic sharp peaks one can normally observe in an NMR spectrum. Another feature is that this longer coherence lifetime can be artificially manipulated by the means of a secondary (but weaker) external magnetic field, from here on out referred to as B_1. A radio frequency wave (or, more properly, its oscillating magnetic field component) that is opportunely tuned to the appropriate Larmor frequency, depending on the target nucleus (affecting γ) and the intensity of B_0, serves as B_1 in any modern NMR experiment, effectively manipulating M by exciting transitions between the α and β states. It is precisely by this mechanism that one can experimentally control M and ultimately extract information about a given sample.

Before I present the key equations that govern the interaction between B_1 and M, it is key to visualize these phenomena in the context of a three-dimensional Cartesian coordinate system. In this space, M is established as parallel to the z-axis, perfectly aligned with the external, static magnetic field B_0. In order to account for the Larmor frequency, it is common to introduce the concept of a rotating frame of reference, meaning that the whole coordinate system rotates around the z-axis at the exact Larmor frequency. The rotating frame of reference is useful since it allows to more readily account for the effect of B_1 on the equilibrium state of the bulk magnetization. In the presence of this secondary magnetic field, and under the condition that said radio frequency is applied along either x- or y-axis, M will tilt away from the z-axis. The extent to which M is diverted away from it depends on the difference between the frequency of the pulse and the Larmor frequency. If this difference (the offset, Ω) is equal to (or close to) 0, then M will effectively be aligned to B_1. That is a 90° pulse. The applied radio frequency I just described will temporarily re-align M along either the x- or the y-axis, depending on the experimental specifications. This is called a *hard* pulse, and its function is to induce identical z-to-xy-axis transitions in every nucleus within a given (usually large) spectral window. Conversely, a *soft* pulse may not have the same effect on every single spin contributing to the bulk magnetization. In other terms, a soft pulse will not induce a complete 90° rotations in spins with

large Ω, and if Ω is large enough, then these spins could remain completely unaffected and continue precessing, unperturbed, around the z-axis. Soft (or weak) pulses are useful tools in NMR spectroscopy since they allow the NMR spectroscopist to selectively target only a particular region of the spectrum, and have found extensive use in recent applications.

1.4.1.3 *Chemical shift*

Another fundamental concept which needs to be introduced is that of "chemical shift", which is of central importance to this work and, more broadly, to the realm of NMR spectroscopy. Despite being exposed to the very same B_0, each nucleus within a molecule will perceive a slightly different magnetic field due to a phenomenon called "shielding". Shielding is mainly dependent on the electronic density. A higher density will "protect" the nucleus from B_0, which will precess at a lower Larmor frequency, and vice-versa. Because the impact of shielding is proportional to B_0, and NMR spectrometers routinely used for bioNMR applications span between 10 and 20 Tesla, it is common to express the frequency of any nucleus as a fraction relative to a reference:

$$\delta = 10^6 \left(\frac{\omega - \omega_{ref}}{\omega_{ref}} \right) \tag{1.7}$$

where δ indicates a given chemical shift value and ω the Larmor frequency of a nucleus. ω_{ref} is usually set equal to a characteristic resonance of a suitable internal or external standard. In practice, Sodium trimethylsilylpropanesulfonate (DSS) is widely used nowadays, mainly due to its ubiquitously high solubility in commonly employed solvents, including water-based solutions (Tetramethylsilane (TMS) is often reported in legacy literature for referencing purposes, instead). Chemical shifts are the premier observable in NMR spectroscopy and are informative of both molecular composition and conformational features.

1.4.1.4 *Spin relaxation*

Section 1.4.1.2 describes how the bulk magnetization (or a fraction of its individual components) can be tilted away from its equilibrium state along the z-axis, yielding a non-equilibrium state. Provided that the stimulus responsible for the generation of this higher-energy state is switched off, the magnetic moment of

the target nuclei will not keep precessing indefinitely around B_1. In fact, an NMR signal will exponentially decay over time according to the following phenomeno-logical relationship:

$$I(t) = I_0 e^{-t/T} \quad \text{or, equivalently} \quad I_0 e^{-Rt} \tag{1.8}$$

where I_0 is the signal immediately after B_1 excitation, t represents time and T is a characteristic time constant. In the NMR field, it is more common to refer to relaxation *rates*, so R is usually the preferred notation, which is simply the reciprocal of T.

Excited nuclear dipoles are subjected to two macroscopic relaxation processes. The first one arises from the re-alignment of M to the z-axis, regenerating the original state dictated by the Boltzmann population law. In legacy literature, this phenomenon is called "spin-lattice relaxation" and is commonly referred to as R_1 relaxation. The second relaxation process is responsible for the decay of the detectable signal in any NMR experiment, reducing the net magnetization present in the x-y plane. This "loss of coherence" between the spins participating to the signals is due to three fundamental mechanisms:

$$R_2 = R_2^{\Delta B} + R_2^{DD} + R_2^{CSA} \tag{1.9}$$

where:

- $R_2^{\Delta B}$ is due to microscopic inhomogeneities in the magnetic field B_0. Apart from being an unavoidable experimental limitation, this contribution is usually the smallest and will not be discussed further.

- R_2^{DD} arises from dipolar coupling between magnetic dipoles, and is responsible for the fundamentally important Nuclear Overhauser effect (NOE).

- R_2^{CSA}, where CSA stands for "chemical shift anisotropy", which arises due to the uneven distribution of the electron cloud around the nucleus.

This latter relaxation mechanism is fundamental in the measurement and description of chemical exchange in NMR spectroscopy.

1.4.1.5 *NMR chemical exchange*

The Larmor frequency of a given nucleus is primarily dependent on the nature of the isotopic species. The value of NMR spectroscopy stems from the ability to differentiate between different instances of the same isotope within a molecule, or multiple ones. This is possible due to the fact that δ is also extremely sensitive to the surrounding chemical environment. Take for instance the following a generic reaction:

$$\text{GS} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{ES} \tag{1.10}$$

Where Ground state (GS) identifies the more thermodynamically stable, lower energy species and Excited state (ES) the less thermodynamically stable, higher energy species (note that the reaction described in Eq. 1.10 can virtually describe any chemical or conformational transition). Because of this transition, the nuclei are exchanging between different chemical environments. If we assume that these two states are characterized by distinct chemical shift δ values due to their peculiar environments, then we obtain δ_{GS} and δ_{ES}. As a consequence, one defines the lifetime of GS and ES states as $\tau_{GS} = 1/k_1$ and $\tau_{ES} = 1/k_{-1}$. Pioneering studies leveraging this phenomenon were limited to the study of *e.g.* the isomerization of small organic molecules. [79] Since then, methodological and engineering advances have allowed the same theoretical fundamentals to be applied to the study of complex dynamics in proteins and nucleic acids. [80, 81]

In section 1.4.1.2, we noted that the extent to which M is diverted from the z-axis following the application of a radio frequency pulse depends on the difference between the frequency of the pulse and the Larmor frequency, *i.e.* the offset Ω. Since distinct δ equals to distinct Larmor frequencies it becomes implied, without loss of generality, the existence of Ω_{GS} and Ω_{ES}. Since chemical shift is the primary observable in NMR spectroscopy, chemical exchange affects the spectrum of the molecule.

Back to Eq. 1.10, the exchange rates of the reaction k_1 and k_{-1} can be combined to yield k_{ex}, to be understood as the characteristic rate of the reaction's dynamic equilibrium. Traditionally, NMR chemical exchange processes are catalogued into three groups depending on the relationship between the parameters $\Delta\Omega = |\delta_{GS} - \delta_{ES}|$ and $k_{ex} = k_1 + k_{-1}$:

- $k_{ex} \ll \Delta\Omega$, implying that τ_{ES} is in the order of seconds, is referred to as the "slow regime",

- $k_{ex} \approx \Delta\Omega$, implying that τ_{ES} is in the order of μs–ms, is referred to as the "intermediate regime", and finally

- $k_{ex} \gg \Delta\Omega$, implying that τ_{ES} is in the order of ps–ns, is referred to as the "fast regime"

The intermediate regime is of specific interest in this particular work.

1.4.1.6 *The intermediate regime*

Exchange processes occurring on the μs–ms time scale result in the broadening of spectral resonances in an NMR spectrum. For this reason, it has been traditionally challenging to characterize dynamics in this regime, as GS resonances may display an abnormally large line width, while ES resonances are broadened beyond detection. Due to these complications, studies concerned with the characterization of the μs–ms time scale are usually devoted to describing "invisible states". For the purposes of the present discussion it is convenient to further distinguish between a *slower*-intermediate exchange regime, and a *faster*-intermediate exchange regime.

SLOWER–INTERMEDIATE EXCHANGE   Defined by $k_{ex} < \Delta\Omega/2$, this condition identifies the millisecond regime, where transitions between GS and EG are comparatively infrequent as opposed to the faster-intermediate regime. "Leaps" between different precession frequencies lead to enhanced dephasing of transverse magnetization, causing motional-induced broadening in the NMR spectrum. The experimental frameworks devised for the analysis of this particular exchange regime are CEST and Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion. The former technique is relevant to this work and further considered in Section 1.4.4.1.

FASTER–INTERMEDIATE EXCHANGE   Defined by $k_{ex} > \Delta\Omega/2$, this condition identifies the microsecond regime. Since $GS \rightleftharpoons ES$ transitions happen more frequently, spins do not accumulate significant phase differences (motional narrowing). This narrowing ultimately leads to average precession frequencies of chemical environments, resulting in the fast regime (rotating frame relaxation disper-

sion ($R_{1\rho}$) is the premier methodology for investigation of faster-intermediate exchange regimes and is evaluated in Section 1.4.4.2).

1.4.2 Assignment of chemical shifts

Chemical shift assignment is an iterative process, requiring information from multiple nuclei (usually ^{1}H, ^{13}C and ^{15}N) in order to yield unambiguous and reliable information. Resonance assignment in NMR spectroscopy can be even more cumbersome for nucleic acids than for proteins, due to the poor distribution (*i.e.*, frequent overlaps) of signals in the spectra. Such a poor chemical shift dispersion is mainly due to two factors. Firstly, DNA only features four building blocks, as opposed to proteins, which are constituted by 20 chemically distinct amino acids. Secondly, DNA is characterized by a relatively scarce variety of conformations (which can essentially be catalogued as B-DNA, A-DNA and Z-DNA), in stark opposition to proteins, which can access a myriad of folded, semi-folded and unfolded architectures.

Below, I briefly describe the routine procedure for chemical shift assignments of natural abundance DNA samples of small to moderate molecular mass.

1.4.2.1 1D ^{1}H experiment

A 1D ^{1}H NMR spectrum is an important staring point and yields key insights about the status of the sample:

1. Considerations on sample concentration. Apart from an immediate qualitative assessment based on the signal-to-noise ratio, internal standards such as DSS can be used to precisely quantify the molarity of biomolecule in solution;

2. Purity of the sample. The presence of unusually sharp peaks characterized by intensities that differ substantially from the sample's peaks could be due to undesired organic impurities. Diffusion Ordered SpectroscopY (DOSY) is a useful technique to ascertain whether those peak actually pertain to the analyte, or not;

3. Water suppression efficacy. Suppressing the overwhelmingly strong H_2O signal in proton-based spectra is essential. While the identification of satisfactory conditions is now trivial thanks to efficient automated procedures,
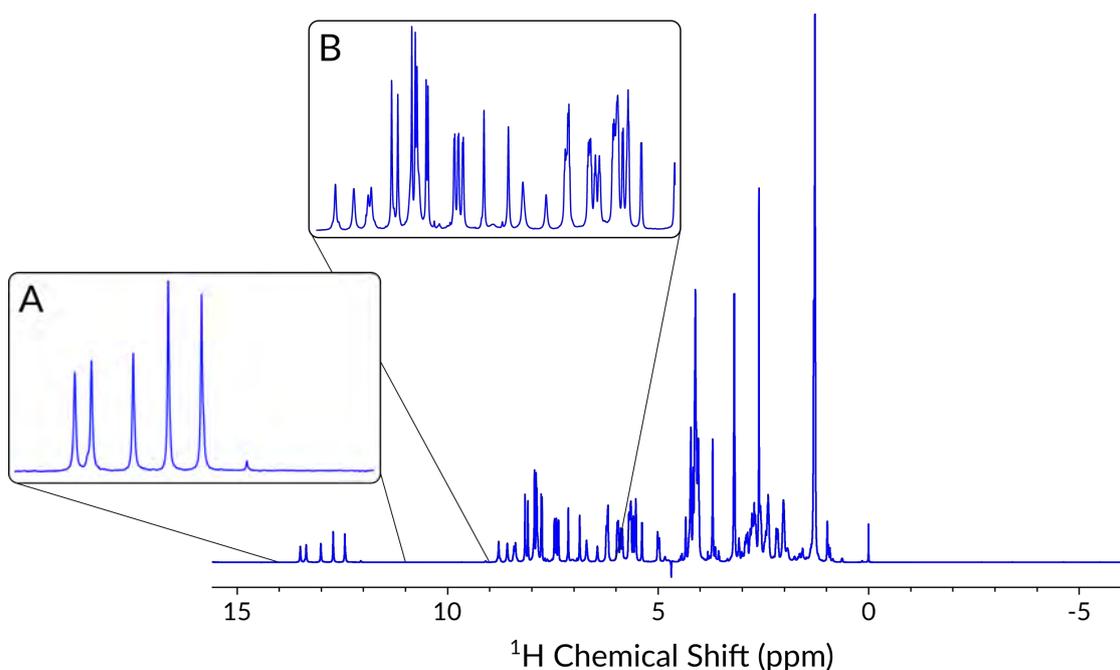
Figure 1.7: 1D ^{1}H spectrum of a 5caC-modified DNA sample measured at 37 °C in a 95% H_2O/5% D_2O phosphate buffer (pH 7.0) at ^{1}H Larmor frequency of 800 MHz. Insets (A) and (B) show enlargements of the imino and aromatic/amino proton regions, respectively. Peak at 0.00 ppm is due to the methyl protons of DSS which is used as an internal standard for spectral referencing purposes.

water suppression performance should be checked frequently as it is heavily dependent on pH and temperature conditions;

4. Imino protons chemical shift. Hydrogen-bonded imino protons in DNA are less electronically shielded compared to non-hydrogen-bonded ones. Annealed DNA can be recognized by the presence of peaks in the 12–15 ppm region. The absence (or weak and broad appearance) of these signals can indicate that DNA is not properly annealed.

1.4.2.2 *2D experiments*

Multi-dimensional NMR experiments are techniques that allow for the recording of spectra displaying correlations between two (or more) frequencies. The main building blocks common to any 2D NMR experiment are the following:

1. Preparation period, allowing the spins to return to thermodynamic equilibrium along the Z-axis. This period is commonly referred to as d_1 or "recovery time".

2. Evolution period (t_1), which serves to encode the spins' chemical shift evolution under the Hamiltonian. This period is responsible for the generation of the second dimension.

3. Mixing period, which is responsible for the magnetization transfer across spins, whether via J_{HH}/J_{XH} (through bond) coupling or dipolar (through space) coupling.

4. Detection period, when direct detection of the magnetization precessing in the x-y plane occurs. This period is usually referred to as t_2 or "acquisition time".

For the purpose of assigning resonances to natural abundance (*i.e.* isotopically unlabelled) DNA oligomers, three kinds of experiments were used:

1. ^1H–^1H J_{HH} (scalar) coupling-based methods (COSY, TOCSY);

2. ^1H–^1H dipolar coupling-based methods (NOESY);

3. ^1H–^{13}C, ^1H–^{15}N $^1J_{XH}$ scalar coupling-based methods (HSQC, HMQC).

In sections 1.4.2.3 and 1.4.2.4, I will briefly introduce and discuss the utilization of ^1H–^1H spectra in the context of the assignment of canonical and epigenetically modified DNA, which is the fundamental preliminary step preceding any further study.

1.4.2.3 *2D Homonuclear NMR experiments*
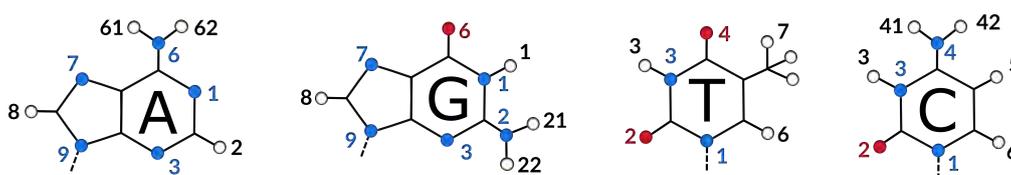
Figure 1.8: Standard IUPAC numbering, used throughout this chapter and following ones, for the canonical DNA building blocks Adenosine, Guanosine, Thymine and Cytosine. White, blue and red dots represent hydrogen, nitrogen and oxygen atoms, respectively. Carbon atoms are implied as junctions connecting black lines.

HOMONUCLEAR J_{HH} COUPLING METHODS ^1H–^1H J_{HH} coupling-based methods rely on the fact that sugar protons, together with cytosine's H5 and H6, belong to the same spin system. Two pulse sequences are mainly used for this purpose:

1. COSY, the simplest multidimensional NMR experiment, yields a spectrum correlating resonances of vicinal ($^3J_{HH}$) or geminal ($^2J_{HH}$) protons. It consists of a first excitation pulse to generate transverse (x-y plane) magnetization, a single frequency labelling period (t_1), a second excitation pulse that acts as the mixing period, and finally the detection period;

2. TOCSY, which generates a spectrum correlating all protons within a spin system, regardless of their respective individual relationship. It only differs from a COSY experiment because of the mixing period, where the 90° pulse is substituted by a series of 180° pulses (*i.e.* spin echo blocks), preventing chemical shifts evolution while retaining J_{HH} coupling evolution.

HOMONUCLEAR DIPOLAR COUPLING METHODS   While scalar coupling relies on the molecular orbitals' electrons mediation, dipolar coupling is due to the through-space interaction of a network of nuclear dipoles and it is regarded as the major source of information in NMR applications involving biologically relevant molecules. Consequently, ^{1}H–^{1}H dipolar coupling-based methods are the premier techniques for the assignment of small, unlabelled nucleic acids sample. They allow for the detection of reliable through-space connections which aid in the verification of the correctness of the DNA sequence, apart from providing an unambiguous starting point for the non-trivial process of correct aliphatic proton assignment. Based on the dipole-dipole cross-relaxation phenomenon, 2D NOESY is the ^{1}H–^{1}H correlation experiment of choice when it comes to moderately small systems.

1.4.2.4 *Sequential walk assignment*

Before NMR data can be used to inform on the structure or dynamics of a given system, it is necessary to assign resonances in the above mentioned spectra to their respective ^{1}H, ^{13}C and ^{15}N nuclei. The standard procedure for signal assignments in relatively small systems (below $\sim$ 5 kDa) is the so-called "sequential walk assignment". [82] First, the obtained TOCSY and NOESY spectra are overlaid. If the DNA sample is properly annealed and the double-stranded conformation is the prevalent one, the two spectra will appear as partially superimposable. The shared peaks will be those that satisfy two conditions:

1. They are part of the same spin system, yielding a cross-peak on a TOCSY spectrum, and
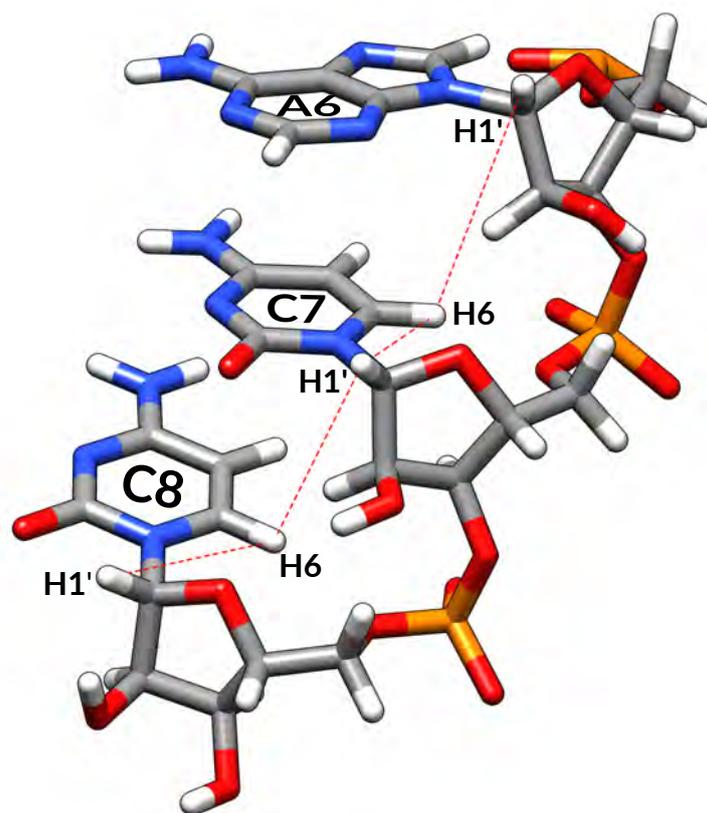
Figure 1.9: Generic dsDNA model visualized via UCSF Chimera software. [83] Red dashed lines highlight key C H6-X H1′ connectivities.

2. They are spatially contiguous, meaning they also develop a cross-peak in the NOESY spectrum, which should also vary in intensity depending on the mixing time deployed.

One particular set of protons that is especially prone to be unambiguously identified via this procedure is Cytosine's H6 and H5. Once H6-H5 cross-peaks have been successfully identified for each Cytosine in the sample, then the next step entails linking H6 to H1′ in proximity. This procedure, which can be further simplified with the aid of interactive visualisation software such as UCSF Chimera (ref. 83), is relatively straightforward due to the characteristic chemical shift values of H1′ protons (the only sugar protons signal above 5 ppm in DNA). Each but terminal Cytosine residues are expected to yield two H6-H1′ connectivities, as per Fig. 1.9.

Whenever an H1′ resonance only has a single cross-peak with H6 (or H8, in case of Adenosine and Thymine), that is a strong indication that such H1′ is part of a terminal nucleoside. With this information alone, following the sequential aromatic-H1′ proton path should suffice for a tentative identification (or verifica-

tion) of the 5'-3' dsDNA sequence, provided that the sample is sufficiently small, the sequence does not display highly repetitive segments and the cross-peaks are sufficiently sharp. Once these resonances are available, NOESY cross-peaks help identify other key protons. For instance, Cytosine's H5 tends to display strong NOESY contacts with H42 and H41 protons (ref. Fig 1.8), which are in turn very useful for explicit determination of resonances pertaining to H1 and H3 nuclei. On the other hand, TOCSY is necessary to identify spin systems starting from the H1' nucleus, while NOESY is often use to infer the stereochemistry (H2' vs H2", H5' vs H5"). Due to the high chemical shift degeneracy of sugar protons, it is often the case that not every signal is unambiguously identifiable, especially when the sample exceeds a certain size. Once the assignment of protons has reached a satisfying level, it is common procedure to record natural abundance ^{1}H–^{13}C and ^{1}H–^{15}N 2D correlation spectra.

1.4.2.5 *2D Heteronuclear NMR experiments*

Assignment of H signals allow for trivial identification of a subset of C and N nuclei by leveraging $^1J_{XH}$ scalar coupling by the means of HSQC and/or HMQC experiments. A particularly convenient set of experiments has been devised by Schanda *et al.*, allowing for the rapid acquisition of ^{1}H-detected heteronuclear experiments. [84] As mentioned above, for relatively small biomolecular systems a thorough resonance assignment can be obtained on the basis of 2D homonuclear experiments alone in favourable cases. Nonetheless, measuring ^{1}H–^{13}C and ^{1}H–^{15}N 2D correlation spectra, even if relatively time-consuming without isotope labelling, can be useful for *e.g.* quantitative comparison of chemical shifts.

1.4.3 Chemical shift metadata

The recording of ^{1}H–^{1}H, ^{1}H–^{13}C and ^{1}H–^{15}N correlation spectra is not only useful for assignment purposes. In fact, the chemical shift values of the assigned resonances can be further employed to inform on conformational preferences. [85] There are mainly two convenient methods to visualize this kind of information that pertain to the present work: chemical shift displacement diagrams and Chemical shift perturbation (CSP) studies. Both methodologies are briefly introduced below.

1.4.3.1 *Chemical shift displacement diagrams*

Chemical shift values are especially informative when matched against those obtained for identical or similar samples in comparable conditions. An example
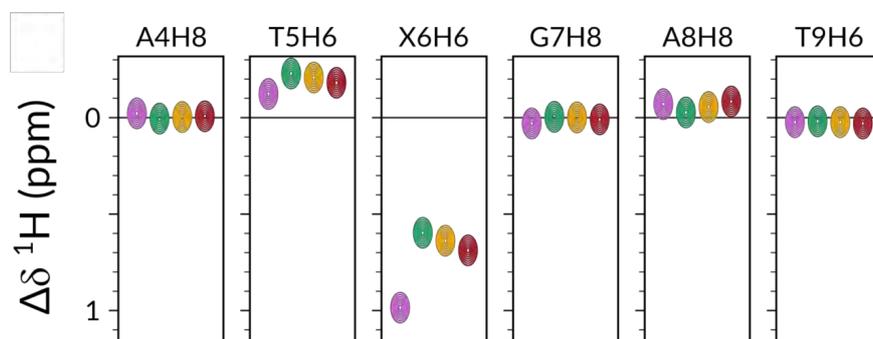
Figure 1.10: Adapted and reproduced with permission from Dubini *et al.* (Ref. 86). Chemical shift displacement studies carried on the same dsDNA sequence carrying epigenetically modified nucleobases and considered at distinct pH conditions. Purple, green, yellow and red dots indicate 5fC and 5caC at pH 7.0, 5.8 and 4.7, respectively. The black solid line is the reference chemical shift of an identical canonical cytosine-containing sample.

of such analysis is shown in Fig. 1.10. There, chemical shifts extracted from canonical cytosine-containing dsDNA are used as a reference value to inform on the long and short range implications of the inclusion of epigenetically modified nucleosides at different pH conditions. Similar diagrams have recently enjoyed widespread adoption in NMR-based studies of conformational and structural features in non-canonical nucleic acids. [87]

1.4.3.2 *Chemical shift perturbation*

If chemical shift data from multiple nuclei is available, then CSP becomes a viable strategy to expand the previous approach. Since its inception, CSP has quickly found widespread use within the bioNMR community due to its usefulness and cost-effectiveness. [88] Resonances of any nucleus in a molecule are influenced by a vast array of chemical bonds, whether of covalent or dipolar character, making them highly sensitive to the environment they perceive. For this reason, chemical shift values have been extensively used to characterize conformational changes in biomolecules due to variations in pH, solvent, temperature or even

the presence of ligands and mutations. [89] In practice, after assignment of the signals, the following equation is used to calculate the degree of perturbation:

$$\Delta\delta_{total} = \sqrt{\delta_H^2 + (\alpha \cdot \delta_X)^2} \tag{1.11}$$

where δ_H and δ_X identify the chemical shift values for ^{1}H and heteronuclei, respectively. The parameter α is usually determined by comparing the chemical shift range of hydrogen with the one of nitrogen or carbon, and literature values of 0.14 for nitrogen and 0.3 for carbon have been suggested, as per ref. 89.

1.4.4 Chemical exchange experiments

The detection of sparsely populated, transient alternative states is a key aspect in the characterization of biomolecules. As it has been long recognized that the biological functions of proteins and nucleic acids cannot be performed by static structures, the investigation of sparsely and temporarily populated conformations has captured the attention of the scientific community. A degree of conformational freedom is indeed a necessary condition for allowing molecular machines like enzymes, DNA and RNA to operate their respective roles. [90, 91] In other words, biomolecules that perform functions are generally best described by ensembles of energetically accessible structural arrangements, rather than by the most populated conformer exclusively.

Building on top of the considerations made in Section 1.4.1.5, below I provide an introduction to the NMR methodologies that have played key roles in the execution of the studies presented in the following chapters. The focus is hereby placed on the underlying concepts and phenomena useful to grasp the goals and the outcomes of these experiments, while a more technical description can be found in the references below, Methods and Materials sections and the relevant Supplementary Information files of the enclosed publications.

1.4.4.1 *Chemical exchange saturation transfer*

The CEST experiment is the most powerful tool for observing chemical exchange processes occurring on the slower-intermediate exchange regime. [92] First described in 1963 in the context of Magnetic resonance imaging (MRI), CEST has more recently found extensive adoption in high-resolution bioNMR. [93, 94] This specific technique is especially useful if one of the two (or more) states' pop-

ulations is low to the point that it becomes undetectable on an NMR spectrum. Interconverting conformers often differ in structural features, thus the magnetic environments of the directly involved nuclei, or adjacent ones, is distinct. This exchange process gives rise to a stochastic modulation of Larmor frequencies.

Since epigenetically modified DNA strands are of central interest to the present work, let us consider the process of DNA melting:

$$\text{dsDNA} \underset{k_a}{\overset{k_d}{\rightleftharpoons}} 2\,\text{ssDNA} \tag{1.12}$$

The key idea is to perturb the magnetization associated with the more populated conformer (*i.e.*, dsDNA, which is GS in eq. 1.10) by applying a weak radio frequency field. Because the melting and annealing processes can be observed in dynamic equilibrium conditions, the weak pulse will affect the magnetization of a second site (ssDNA, which is ES in eq. 1.10), under opportune conditions. While chemical exchange phenomena have been known since the early '60s, only exchanging conformations with similar populations have been reported until recently. If left unaddressed, this limitation would have hindered the characterization of virtually any biologically relevant system, where there often is a primary and one or several secondary ES. In 2000, Ward *et al.* realized that CEST had the potential for exposing exceedingly sparse chemical exchange phenomena resulting in ES populations so low that they would otherwise be well below the NMR detection limit. [95] Deploying the aqueous solvent signal as a reference, they could describe a proton exchange process of the imino proton of hydroxytryptophan in a 35 mM solution, accounting for a solute concentration of only 0.03%. Correspondingly, "invisible" resonances of biomolecules can be characterized through the relevant resonances of highly populated GS owing to a similar amplification effect, as long as the minor states have a distinct chemical shift, fractional populations greater than $\approx 0.5\%$ and lifetimes in the range of ≈ 3 to 50 ms.

From a practical standpoint, the CEST experiments employs a low power, long duration pulse (a "spin lock", typically hundreds of milliseconds long) to saturate different regions of a spectrum by varying its carrier frequency. In Fig. 1.11, a schematic representation of the CEST experiment is offered.

If a spectrum is measured without a spin-lock pulse (or if the spin-lock pulse is far enough off-resonance relative to our signal of interest), the GS signal is not impacted. Whenever the carrier frequency is instead moved closer towards
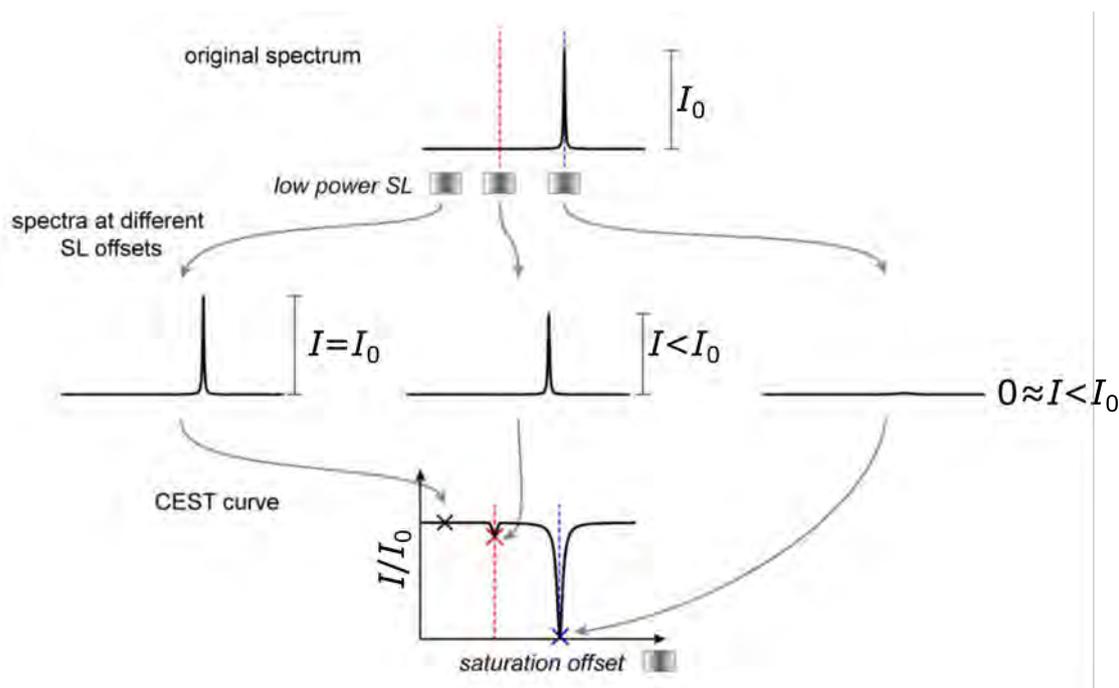
Figure 1.11: Adapted and reproduced with permission from Marusic *et al.* (Ref. 90. 1D spectrum representing the signal of a GS conformation engaging in a slower-intermediate exchange process with ES (undetectable, red line). Middle: a series of spectra are recorded featuring a weak and long pulse (spin-lock). Spin-locks affect the appearance of the spectra depending on its carrier frequency. Bottom: the ratio of the measured signal intensity I and the initial signal intensity I_0 (spectrum measured without a spin-lock) are plotted yielding a typical CEST profile.

either the ES or GS chemical shifts, GS signal intensity will decrease, provided that the aforementioned exchange conditions are met. The reduction of intensity of the major peak is much easier to detect than the minor exchanging peak in a conventional spectrum, therefore small populations can be detected.

From a CEST profile (Fig. 1.11, bottom), the exchange process (including the population and the chemical shift of the minor state together with the kinetic rate) can be extracted finding numerical solutions to the Bloch–McConnell equations. [96] In addition, classic theoretical frameworks in the context of thermodynamics can be implemented to further inform on standard Gibbs free energies of GS and ES, together with the activation parameters of the process and its transition state. The reader is directed to Chapter 2, where such aspects have been extensively discussed and applied to the study of epigenetically modified dsDNA strands.

Lastly, a recent research theme in bioNMR has been the adaptation of techniques initially devised to probe heteronuclei such as ^{13}C and ^{15}N to ^1H. The

main indisputable advantage of such an approach is the dramatic reduction in the cost of production sample, while a drawback may be represented by the reduced chemical shift dispersion of protons compared to carbon-13 and nitrogen-15. Below, the ^{1}H CEST pulse sequence used in Chapter 2 and 3 is presented (Fig. 1.12).
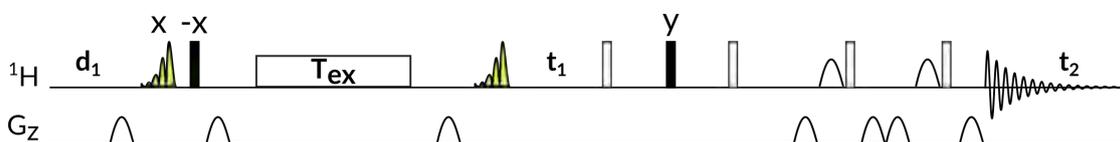
Figure 1.12: Pulse sequence for the SELectively Optimized Proton Experiment (SELOPE) CEST ^{1}H experiment. Thin and thick bars represent 90° and 180° hard pulses, respectively. Selective ^{1}H irradiation of the spectral region of interest was achieved via pulse shapes eBurp2 (selective excitation), labelled in yellow. Suppression of the water signal was achieved using the excitation sculpting scheme.

1.4.4.2 *Rotating-frame relaxation dispersion*

The $R_{1\rho}$ experiment is the premier experimental technique for probing chemical exchange processes on the faster-intermediate time scale. $R_{1\rho}$ relaxation dispersion has been extensively used in the description of conformational exchange occurring in proteins, RNA and DNA. [97–99] Similar to CEST, $R_{1\rho}$ is most useful whenever an "invisible" conformer is suspected to populate a higher-energy state.

From a practical perspective, decay rates are measured as exponential decays by using spin-locks of different intensities and durations. The resulting profiles can then be fitted using the Bloch–McConnell equations or analytical expressions (such as the Laguerre approximation) for simplified scenarios and specific exchange regimes. [100] Fig. 1.13 offers a schematic description of an on-resonance $R_{1\rho}$ experiment, while the two paragraphs below exemplify the two limiting cases usually employed: the no-exchange limit and the chemical exchange limit:

NO–EXCHANGE CASE   In a simple yet realistic scenario with two non-exchanging spins with different chemical shifts, a sufficiently strong spin-lock pulse will keep the nuclei bulk magnetizations in the transverse plane. Signal intensities will decay according to the respective R_2 rates. A weaker spin-lock will cause the bulk magnetization responsible for the red signal in Fig. 1.13 to gain a non-zero z-component, introducing an R_1 driven correction to the observed
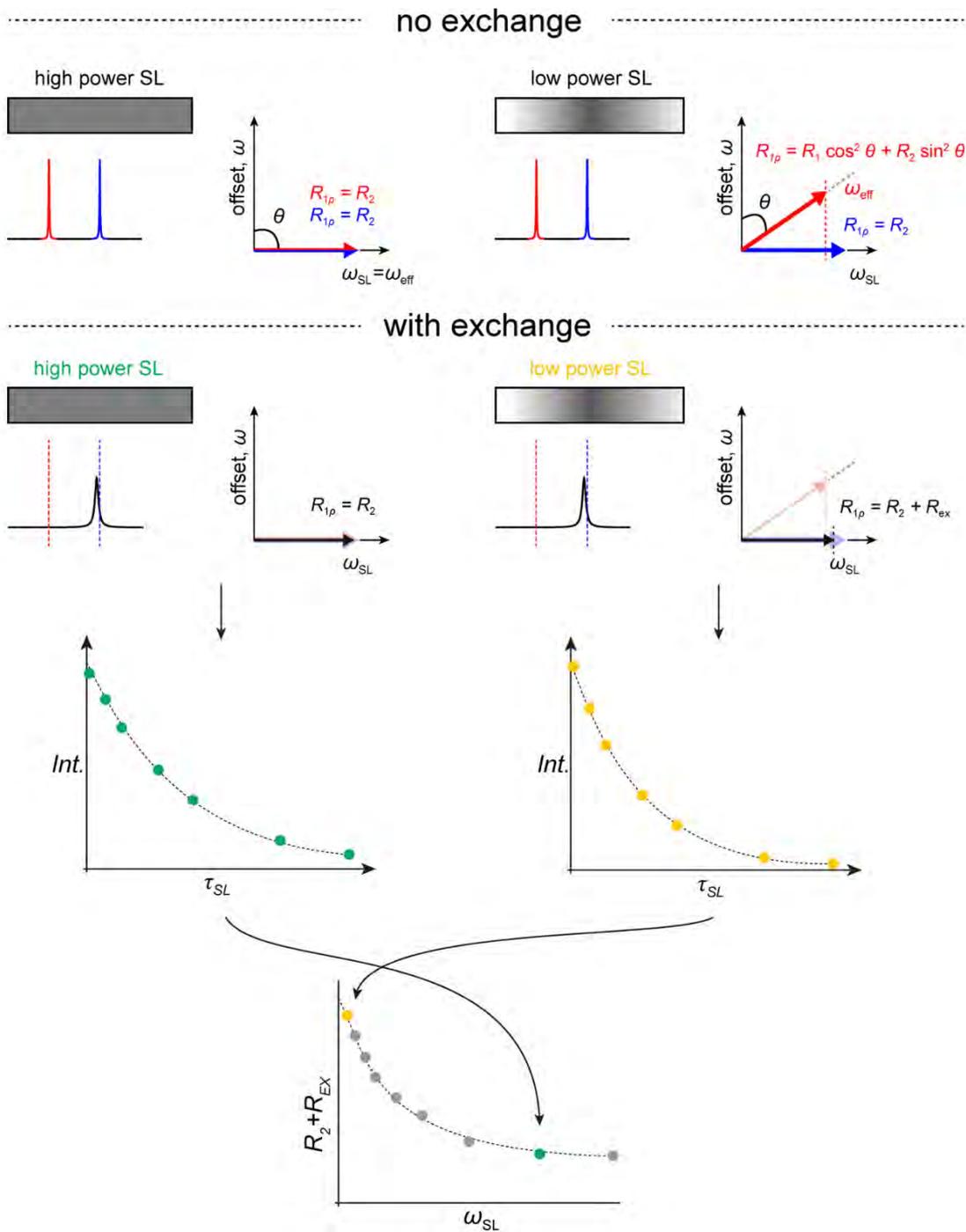
Figure 1.13: Adapted and reproduced with permission from Marusic *et al.* (Ref. 90). Effects of high- and low-power spin-locks on non-exchanging (upper row) and exchanging (lower rows) systems. Red and blue labelling indicates either distinct non-exchanging conformations (upper row) or a GS and ES (respectively, lower rows). Green and yellow data points indicate to $R_{1\rho}$ profiles recorded with either high- or low-power spin-lock pulses, respectively.

relaxation rate (now $R_{1\rho}$). This instance leads R_2 and R_1 to a squared trigonometric modulation.

CHEMICAL EXCHANGE CASE   If a chemical exchange phenomenon is present on the μs regime, the peak corresponding to GS is usually broadened, while the ES peak is undetectable. In case a high-power spin-lock is deployed, no information reporting on the chemical exchange can be extracted, as both GS and ES resonances will decay according to R_2 (the assumption is generally made that R_2 is identical for both GS and ES). Whenever progressively weaker spin-locks are applied, the off-resonance effect causing the tilt of ES magnetization away from the transverse plane and establishing an increasingly deviated effective field ω_{eff} will become more apparent. The difference in the perceived strength of the spin lock pulse (and hence, the alignment with the respective effective field) leads to a difference in relaxation due to the chemical exchange event. In other words, an R_{ex} contributes to the overall signal decay rate, in addition to R_2. R_{ex} is progressively larger in magnitude the weaker the spin-lock applied.

In practice, the recording of an $R_{1\rho}$ relaxation dispersion profile consists of measuring several decay rates at several spin-lock strengths. Similarly to CEST, also $R_{1\rho}$ has been adapted to ^{1}H-only measurements. Production of natural abundance RNA and DNA samples is usually one to two orders of magnitude cheaper compared to the same amount of isotopically labelled material, making ^{1}H measurements attractive. In addition, proton density in nucleic acids is substantially reduced compared to protein meaning that deleterious ^{1}H-^{1}H network effects, which are potentially harmful to CEST and $R_{1\rho}$ experiments, can more easily be avoided compared to protein-based applications. In contrast, the limited ^{1}H chemical shift dispersion is usually an insurmountable obstacle whenever measuring biologically relevant systems, making SELOPE-derived applications accessory techniques at best in the characterization of larger samples. Below, the ^{1}H $R_{1\rho}$ pulse sequence used in Chapter 2 and 3 is presented (Fig. 1.14).
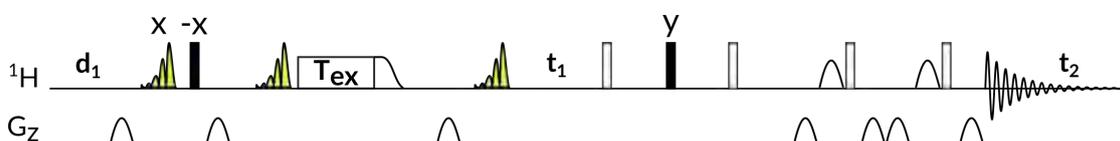
Figure 1.14: Pulse sequence for the SELOPE $R_{1\rho}$ ^{1}H experiment. Thin and thick bars represent 90° and 180° hard pulses, respectively. Selective ^{1}H irradiation of the spectral region of interest was achieved via pulse shapes eBurp2 (selective excitation), labelled in yellow. Suppression of the water signal was achieved using the excitation sculpting scheme. A heat compensation block (not shown) was included before d_1 such that the sample was exposed to the same total radio frequency power, independent of the specific spin-lock duration employed.

1.5 THESIS SYNOPSIS

The overarching goal of my doctoral work was the structural and dynamic characterization of cytosine-derived epigenetically modified nucleosides within ds-DNA. Projects inherent to this thesis are presented below. Chapter 2 introduces the 5fC modification and dwells on its impact on dsDNA structure and dynamics in comparison with canonical cytosine. Chapter 3 deals with the characterization of 5caC, with a special focus on its pH dependent behaviour. Each of chapters 2 and 3 are preceded by a "Foreword" section, whose function is to put each specific project in the broader context of this work. Finally, chapter 4 takes stock of the projects included in the PhD thesis and outlines futures research avenues.

2 | 5FC–INDUCED MELTING KINETICS BY CHEMICAL EXCHANGE

2.1 FOREWORD AND MAIN PUBLICATION

Epigenetically modified nucleobases reshape the molecular properties of nucleic acids and provide an additional regulatory information layer beyond the canonical base sequence. In particular, their distributional patters in distinct tissues, interactions with enzymes and impact on the B-DNA structure have been the most investigated aspects. The most abundant set of DNA modifications within the genome of any domain of life occur via functionalization of the C(5) position of canonical cytosine (see Fig. 1.8). Epigenetic modifications of cytosine have attracted considerable research interest in the last decade, but their subtle influence on genomic DNA has proven to be recalcitrant of thorough, comprehensive characterization.

This work has been devised to specifically address the thermodynamic and kinetic features of a DNA double strand including the naturally occurring 5fC base. Since its discovery, 5fC has been mostly considered and studied through the lenses of biology (tissue distribution, impact on enzymatic activity) and synthetic organic chemistry (synthesis of phosphoramidite precursors). Controversial results obtained from structural biology or physical chemistry studies, together with its prevalent presence within cells and tissues where gene expression is sharply enhanced, have evidenced the need for further investigations in consideration to the impact of 5fC on the double-helix dynamics of DNA. In particular slow and intermediate time scales, which are of utmost interest in relation to DNA replication and maintenance, have long eluded experimental characterization.

Recent developments in bioNMR methodology allowed for experimentally more accessible investigations accessing dynamics and kinetics information via natural abundance samples, avoiding costly isotopic enrichment procedures. Coupling state-of-the-art synthetic methodologies with the non-invasivity and site-specificity typical of NMR spectroscopy, we unravelled the conformational

transitions of 5fC-modified and canonical DNA strands on the millisecond timescale. Numerical comparison of the obtained kinetic and thermodynamic parameters yielded site specific quantification of the influence of 5fC in a biologically relevant setting, advancing our understanding of its role and its interactions with reader proteins.

In summary, the value of the present effort is two-fold. On one hand, we present an original theoretical framework capable of extracting thermodynamic and kinetic parameters using exclusively CEST-derived data. On the other hand, we apply this new methodology to unravel the dynamics of the epigenetically relevant 5fC base.

This work is integrative in nature, coupling synthetic organic chemistry and physical chemistry concepts. It was designed and carried out in collaboration with the Carell group within the SFB1309 framework. A copy of the open-access publication (available through the Creative Commons CC BY-NC license) is reprinted below and accessible following this link.

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

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 noninvasively 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

2 Nucleic Acids Research, 2020

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 under-

goes 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

4 Nucleic Acids Research, 2020
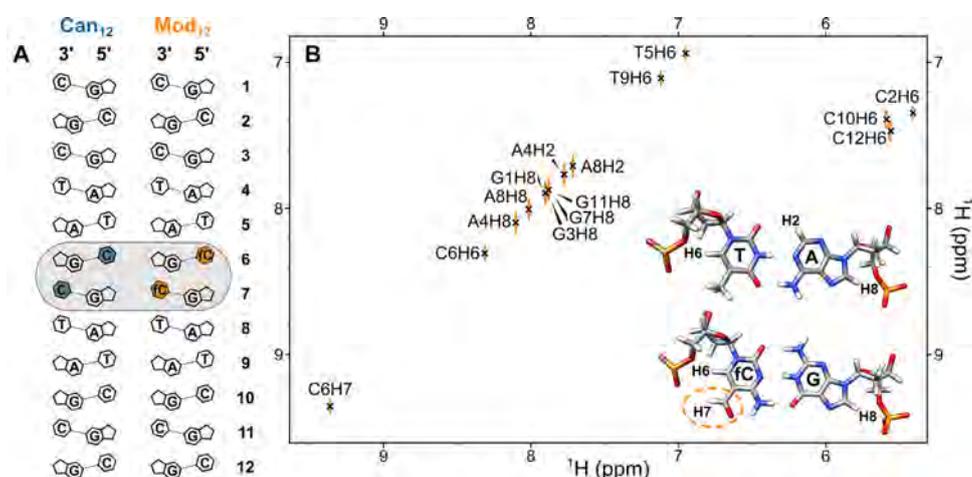
Figure 1. (**A**) DNA constructs of Can_{12} and Mod_{12}, gray box highlights the site of the 5fC incorporation. (**B**) SELOPE 2D ^1H–^1H correlation spectrum of the Mod_{12} aromatic region. H6 protons of canonical cytosine bases appear as crosspeaks due to their ^3J coupling to H5, while the H6 proton 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 ^1H–^{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 ^3J ^1H–^1H couplings of the sugar moieties confirmed the predominant presence of B-DNA conformation in both samples (58).

The chemical shift values for all ^1H 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 ^1H 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 ^1H 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 ($\Delta G°$, $\Delta H°$, $\Delta S°$) 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 $\Delta H°$ 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-
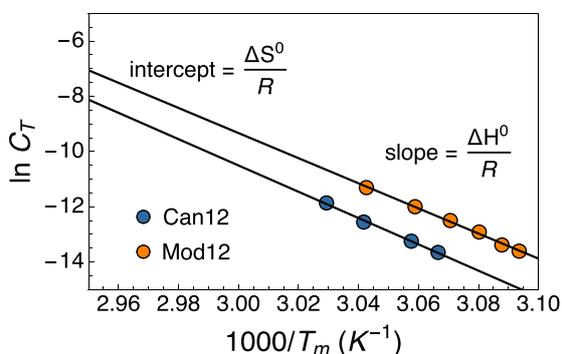
Figure 2. Concentration-dependent UV/Vis melting temperatures analysis of Can$_{12}$ (blue) and Mod$_{12}$ (orange). The reciprocal of the observed T_m is plotted as a function of the logarithm of the total DNA concentration (C_t). The slope and the intercept of the fitted linear corresponds to $\Delta H^{\circ}/R$, 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 entropy 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

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

6 Nucleic Acids Research, 2020

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
Mod₁₂	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						
Can₁₂	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:

$$k_a = \frac{k_{ex} p_D}{2 C_t (1 − p_D)} \qquad (2)$$

$$k_d = k_{ex}(1 − p_D) \qquad (3)$$

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

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

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 × 10⁵ 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 Mod₁₂ than for Can₁₂.

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 Mod₁₂ and Can₁₂, 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

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

where i stands for association (a) or dissociation (d), k_B is the Boltzmann constant, R is the universal gas constant,
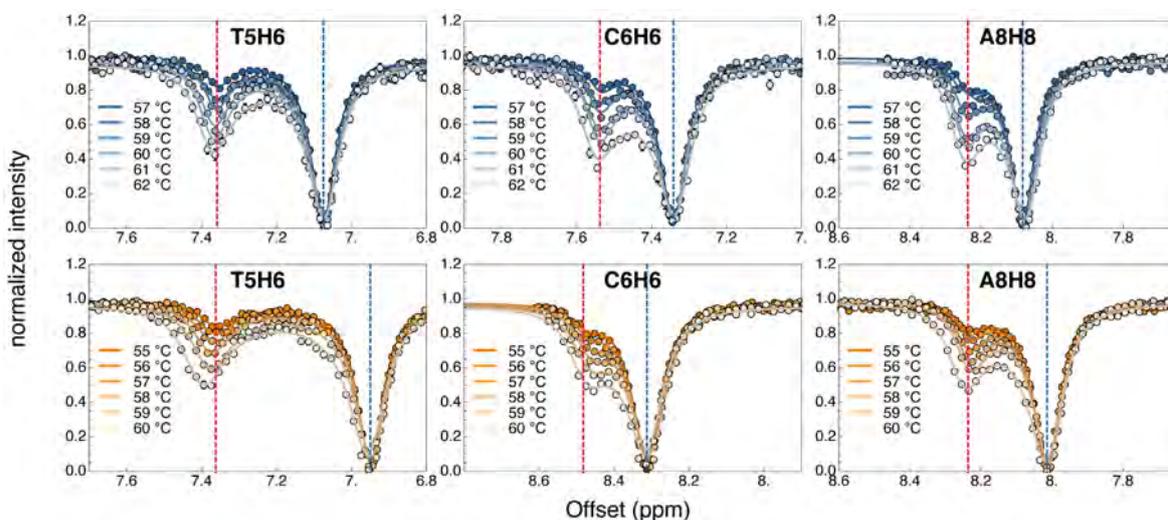
Figure 3. Temperature-dependent ^{1}H CEST profiles and fits of T5H6, C6H6 and A8H8 protons of Can$_{12}$ (blue, upper row) and Mod$_{12}$ (orange, bottom row). Dashed blue and red vertical lines indicate the chemical shifts of the double-stranded and single-stranded conformations, respectively. As expected, the population of ssDNA increases gradually with increasing temperature.
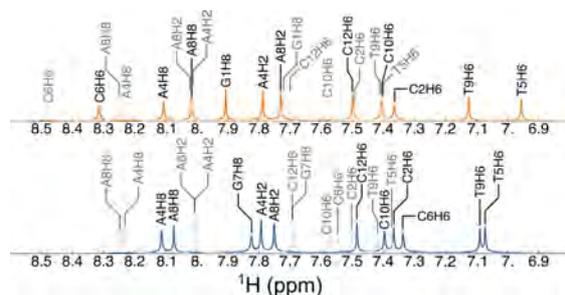
Figure 4. Simulated ^{1}H 1D spectra. Darker and lighter shades represent the ground-state (dsDNA) and excited-state (ssDNA) conformations. Only those sites are included that showed a second dip in the CEST profiles.
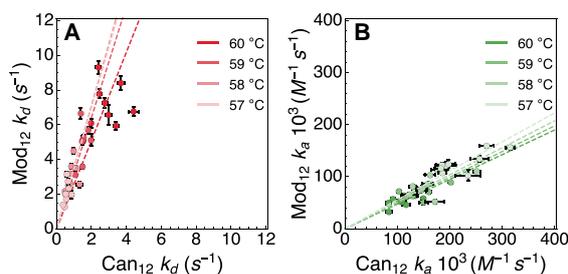
Figure 5. Correlation between the site-specific k_d (**A**) and k_a (**B**) kinetic rates between the observable sites of Mod$_{12}$ and Can$_{12}$ at 57, 58, 59 and 60°C. Dashed lines represents the linear fits at each temperature with a slope of 2.5–3.5 for k_d and 0.47–0.55 for k_a.

and κ is the transmission coefficient, assumed to be 1. The slope and intercept of the linear fit to $\ln k_i/T$ versus $1000(1/T - 1/T_{\mathrm{hm}})$ gives the activation enthalpy change

$(\Delta H_i^\ddagger)$ and the Gibbs free energy change at T_{hm} ($\Delta G_i^\ddagger(T_{\mathrm{hm}})$) respectively, from which the activation entropy ($\Delta S_i^\ddagger$) can be derived as

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

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 $\Delta H°$ and $\Delta S°$ 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-
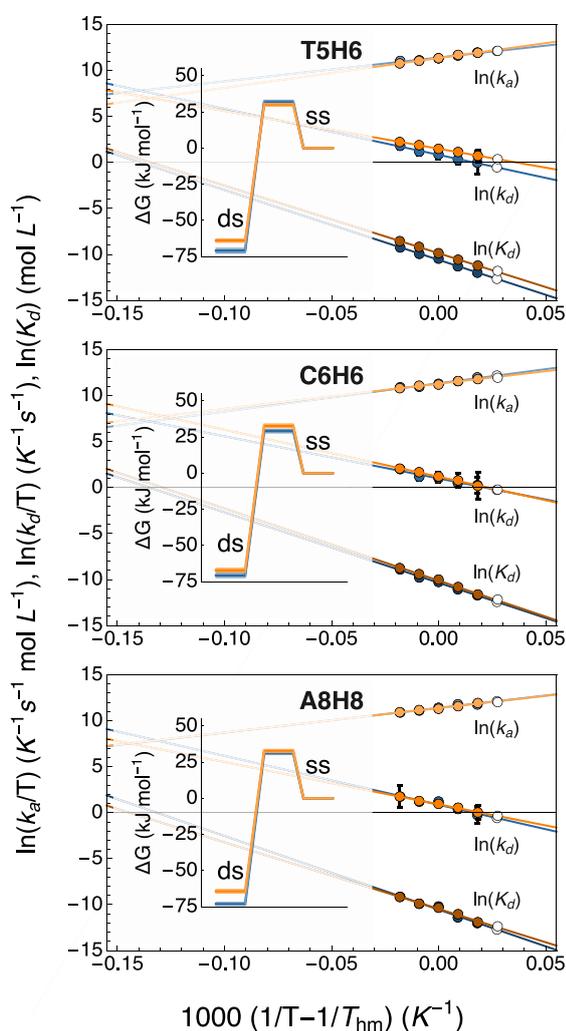
Figure 6. van't Hoff plots are displayed for three selected sites. Shades of orange and blue indicate data entries and linear fits for Mod_{12} and Can_{12}, respectively. White data points represent back-calculated values. Fitted values for the association and dissociation rate constants (k_a and k_d) 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
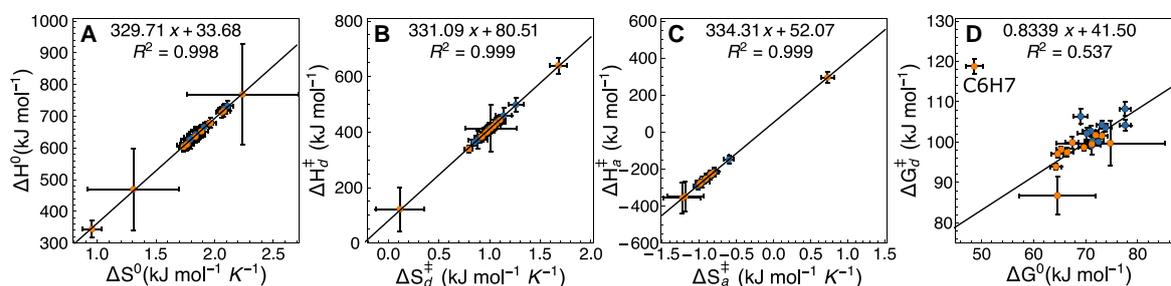
Figure 7. (A–C) Correlation between the enthalpy and entropy changes of the thermodynamic and kinetic processes for Can_{12} (blue) and Mod_{12} (orange). The solid lines represent a linear fit in the form of $\Delta H = T_c \Delta S + \Delta G$. (D) Correlation between the dissociation and equilibrium 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

($\Delta G_d^\ddagger$) 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 ($\Delta G°$, $\Delta H°$, $\Delta S°$, 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 time-scale 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

FUNDING

REFERENCES

1. Poater,J., Swart,M. and Bickelhaupt,F.M. (2014) B-DNA structure and stability: the role of hydrogen bonding, $\pi-\pi$ 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.

12 Nucleic Acids Research, 2020

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) m1A and m1G 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 N6-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.

2.2   SUPPLEMENTARY MATERIAL

The Supplementary Material file for the above publication contains:

- The derivation of mathematical relationships between UV/Vis, NMR observables and thermodynamic/kinetic data,

- Additional UV/Vis and NMR melting profile plots,

- Figures and schemes supporting the Discussion and Conclusion sections, and

- Comprehensive numerical results of the CEST fitting procedures.

Supporting Information for

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[1,3], Thomas Carell[1,3], Petra Rovó[1,2]*

*E-mail: petra.rovo@lmu.de

[1]*Faculty of Chemistry and Pharmacy, Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandstraße 5-13, 81377 Munich, Germany*
[2]*Center for Nanoscience (CeNS), Faculty of Physics, Ludwig-Maximilians-Universität München, Schellingstraße 4, 5th floor, 80799 Munich, Germany*
[3]*Center for Integrated Protein Science Munich (CIPSM), Department of Chemistry and Biochemistry, Ludwig-Maximilians-Universität München, Butenandtstraße 5, 81377 Munich, Germany*

Table of Contents

List of Figures

List of Tables

SI Text

DNA thermodynamics

The duplex (D) to single-strand (M) transition of the melting of short double-stranded DNA can be accurately modelled as an all-or-none two-state transition [1, 2, 3]

$$D \underset{k_a}{\overset{k_d}{\rightleftharpoons}} 2\,M \tag{1}$$

where the dissociation equilibrium constant (K_d) can be expressed with the concentration of the duplex and singe-stranded molecules, or equivalently, with the ratio of the dissociation (k_d) and association (k_a) rates

$$K_d = \frac{[M]^2}{[D]} = \frac{k_d}{k_a} \tag{2}$$

it is also known that

$$K_d = \exp(-\Delta G^0 / RT) \tag{3}$$

where the standard Gibbs free energy change (ΔG^0) is given as the difference of the standard enthalpy change (ΔH^0) and the temperature times the standard entropy change (ΔS^0).

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 \tag{4}$$

The total DNA concentration (C_t) can be given as

$$C_t = 2[D] + [M] \tag{5}$$

and hence, the fraction of molecules that are in the double (p_D) or single-stranded (p_M) states equals to

$$p_D = \frac{2[D]}{C_t} = 1 - p_M \tag{6}$$

$$p_M = \frac{[M]}{C_t} = 1 - p_D \tag{7}$$

from which the ssDNA and dsDNA concentration can be expressed as

$$[M] = C_t(1 - p_D) \tag{8}$$

$$[D] = \frac{p_D C_t}{2} \tag{9}$$

By replacing eq. 24 and 23 to eq. 2 we get

$$K_d = \frac{(C_t(1 - p_D))^2}{\frac{p_D C_t}{2}} = \frac{2C_t^2(1 - p_D)^2}{p_D C_t} = \frac{2C_t(1 - p_D)^2}{p_D} \tag{10}$$

This expression can be rearranged to p_D as

$$p_D = \frac{4C_t + K_d - \sqrt{K_d^2 + 8K_d C_t}}{4C_t} \tag{11}$$

At the melting temperature (T_m) $p_D = 0.5$, and from eq. 10 it follows that

$$K_d(T_m) = C_t \tag{12}$$

The fact that the equilibrium constant at the melting temperature equals to the total DNA concentration can be exploited in a DNA-concentration-dependent melting study to determine the standard enthalpy and entropy changes of the dissociation process.[1]

$$\Delta G^0(T_{\mathrm{m}}) = -RT \ln C_t \tag{13}$$

and hence

$$\ln C_t = \frac{-\Delta H^0}{RT_m} + \frac{\Delta S}{R} \tag{14}$$

Analysis of UV melting data

Temperature-dependent UV melting analysis was used to derive a first estimate about the DNA melting thermodynamics at low DNA concentration, and to extrapolate to the concentration regime of the NMR measurement. Explicit thermodynamic equation for the temperature-dependence of the fraction of double-stranded and single-stranded DNA is known from the literature. [2] Here we only outline the most important equations that are needed to arrive at the published relationships.

The van't Hoff equation gives the temperature dependence of the equilibrium constant in the following form

$$\frac{d \ln K_d}{d(1/T)} = \frac{\Delta H^0}{R} \tag{15}$$

The integral of this equation can be given as

$$\ln K_d = \int \frac{\Delta H^0}{R} d(1/T) = \frac{\Delta H^0}{RT} + C, \tag{16}$$

where C is an integration constant. At the melting temperature $K_d(T_m) = C_t$ thus

$$\ln K_d - \ln C_t = \frac{\Delta H^0}{R} \left(\frac{1}{T} - \frac{1}{T_m} \right) =: x \tag{17}$$

The equilibrium constant for a bimolecular reaction has been given in eq. 10, and hence

$$\ln \left(\frac{K_d}{C_t} \right) = \ln \frac{2(1 - p_{\mathrm{D}})^2}{p_{\mathrm{D}}} \tag{18}$$

Since we labeled the expression in eq. 17 as x, we can take its exponential and arrive at

$$\exp(x) = \frac{2(1 - p_{\mathrm{D}})^2}{p_{\mathrm{D}}} \tag{19}$$

Solving the above equation for p_{D} we obtain

$$p_{\mathrm{D}} = 1 - \frac{\exp(x)}{4} \left(\sqrt{1 + 8 \exp(-x)} - 1 \right) \tag{20}$$

the non-zero heat capacity change can be explicitly taken into account by defining x as

$$x = \frac{\Delta H^0(T_m)}{R} \left(\frac{1}{T} - \frac{1}{T_m} \right) - \frac{\Delta C_p}{R} \ln \frac{T}{T_m} + \frac{\Delta C_p}{R} \left(1 - \frac{T_m}{T} \right) \tag{21}$$

In UV melting studies the temperature dependence of the pure double- and single-stranded DNA should be taken into account. This is usually done assuming a linear relationship in the form of

$$A_{260nm} = \epsilon_{\mathrm{D}} p_{\mathrm{D}} + \epsilon_{\mathrm{M}} (1 - p_{\mathrm{D}}) \tag{22}$$

where ϵ_{D} and ϵ_{M} are the molar extinction coefficient of dsDNA and ssDNA, respectively.

Parameter estimation of UV melting experiment

We measured the DNA denaturation profiles at multiple total DNA concentration between the range of 1 and 12 μM, between 20 and 85 °C. To obtain the melting temperature at each concentration we fitted eq. 20 and eq. 22 simultaneously with x as it is defined in eq. 21, and fixed the heat-capacity change at 1320 J mol^{-1} K^{-1} which is a good estimate for GC-rich 12mer DNA.[4] Setting this parameter to zero did not affect the fitted value for the melting temperatures. Even though the fitting procedure includes the fit of the standard enthalphy change at each concentration, we found that this value is highly dependent on the quality of the baseline fit, and therefore we did not include them in any further analysis. Exemplary UV/Vis melting profiles are shown in Figure S2 and S3.

The error of the fitted T_m was determined from quadruplicate experiments and further propagated to the parameters obtained during the concentration-dependent thermodynamic analysis. The uncertainty of the total DNA concentration was assumed to be 10%. With Monte Carlo error estimation method we simulated 1000 synthetic data pairs (T_m and C_t) for each measurement condition by sampling the T_m and C_t around their mean value assuming normal distribution and fitted them according to eq. 14. The mean and standard deviation of the obtain slope and intercept of the fitted linears correspond to the mean and standard deviation of $\Delta H^0/R$ and $\Delta S^0/R$, from which ΔG^0 and its error was obtained according to eq. 4.

DNA melting kinetics

In the following we consider the reversible, steady-state dissociation and association kinetics of a DNA melting process (eq. 1) and describe it with perturbation method. [5] In equilibrium, the macroscopic changes of the dimer and monomer concentrations are zero:

$$[D]_{\mathrm{eq}}t = -k_d[D]_{\mathrm{eq}} + k_a[M]_{\mathrm{eq}}^2 = 0 \tag{23}$$

$$[M]_{\mathrm{eq}}t = k_d[D]_{\mathrm{eq}} - k_a[M]_{\mathrm{eq}}^2 = 0 \tag{24}$$

Upon perturbation, the system reaches a new equilibrium:

$$\begin{aligned}[D] &= [D]_{\mathrm{eq}} + [d] \\ [M] &= [M]_{\mathrm{eq}} + [m]\end{aligned} \tag{25}$$

where $[d]$ and $[m]$ are the small perturbations of the equilibrium values of $[D]_{\mathrm{eq}}$ and $[M]_{\mathrm{eq}}$. From the conservation law it follows that

$$[d] + 2[m] = 0 \tag{26}$$

The kinetic equation of $[D]$ and $[M]$

$$([D]_{\mathrm{eq}} + [d])t = -k_d\left([D]_{\mathrm{eq}} + [d]\right) + k_a\left([M]_{\mathrm{eq}} + [m]\right)^2 \tag{27}$$

$$([M]_{\mathrm{eq}} + [m])t = k_d\left([D]_{\mathrm{eq}} + [d]\right) - k_a\left([M]_{\mathrm{eq}} + [m]\right)^2 \tag{28}$$

By expanding these equations we get

$$[D]_{\mathrm{eq}}t + [d]t = -k_d[D]_{\mathrm{eq}} - k_d[d] + k_a[M]_{\mathrm{eq}}^2 + 2k_a[M]_{\mathrm{eq}}[m] + k_a[m]^2 \tag{29}$$

$$[M]_{\mathrm{eq}}t + [m]t = k_d[D]_{\mathrm{eq}} + k_d[d] - k_a[M]_{\mathrm{eq}}^2 - 2k_a[M]_{\mathrm{eq}}[m] - k_a[m]^2 \tag{30}$$

Subtracting eq. 23 and 24 from eq. 29 and 30 respectively we get

$$[d]t = -k_d[d] + 2k_a[M]_{\text{eq}}[m] + k_a[m]^2 \tag{31}$$

$$[m]t = k_d[d] - 2k_a[M]_{\text{eq}}[m] + k_a[m]^2 \tag{32}$$

and neglecting the second-order terms we arrive at the kinetic equation in the form of

$$[d]t = -k_d[d] + 2k_a[M]_{\text{eq}}[m] \tag{33}$$

$$[m]t = k_d[d] - 2k_a[M]_{\text{eq}}[m] \tag{34}$$

These above two equation could be written in a matrix form as

$$\frac{d}{dt}\begin{bmatrix}[d]\\{[m]}\end{bmatrix} = \begin{bmatrix}-k_d & +2[M]_{\text{eq}}k_a\\ k_d & -2[M]_{\text{eq}}k_a\end{bmatrix}\begin{bmatrix}[d]\\{[m]}\end{bmatrix} \tag{35}$$

and using the condition stated in eq. 23, 24, and 25 we arrive at

$$\frac{d}{dt}\begin{bmatrix}[D]\\{[M]}\end{bmatrix} = \begin{bmatrix}-k_d & +2[M]_{\text{eq}}k_a\\ k_d & -2[M]_{\text{eq}}k_a\end{bmatrix}\begin{bmatrix}[D]\\{[M]}\end{bmatrix} \tag{36}$$

In a chemical exchange experiment, the general differential equation system that accounts for the interconversion between two states is given as

$$\frac{d}{dt}\begin{bmatrix}[I]\\{[S]}\end{bmatrix} = \begin{bmatrix}-k_1 & +k_{-1}\\ k_1 & -k_{-1}\end{bmatrix}\begin{bmatrix}[I]\\{[S]}\end{bmatrix} \tag{37}$$

from which it follows that the forward and backward reaction will equal to k_d and $2[M]_{\text{eq}}k_a$, respectively

$$k_1 = k_d \qquad k_{-1} = 2[M]_{\text{eq}}k_a \tag{38}$$

Hence, the exchange rate is

$$k_{\text{ex}} = k_d + 2[M]_{\text{eq}}k_a \tag{39}$$

By rearranging eq. 39 and using eq. 2, 7, the dissociation and association rates can be expressed as

$$k_d = k_{\text{ex}}(1 - p_{\text{D}}) \tag{40}$$

$$k_a = \frac{k_d}{K_d} = \frac{k_{\text{ex}}(1 - p_{\text{D}})}{\frac{2C_t(1 - p_{\text{D}})^2}{p_{\text{D}}}} = \frac{k_{\text{ex}}p_{\text{D}}}{2C_t(1 - p_{\text{D}})} \tag{41}$$

Parameter estimation of CEST experiment

At each temperature, the evolution of the signal intensities in the ^{1}H on-resonance $R_{1\rho}$, in the ^{1}H inversion-recovery (R_1), and in the CEST experiments were analyzed and fitted concurrently to yield baseline $R_{1\rho}^0$, R_1 rates, as well as exchange rates (k_{ex}), populations (p_{D}, $1 - p_{\text{D}}$) and the chemical shifts (Ω_{D}, Ω_{M}) of the exchanging states. We assumed, that the exchanging states have the same transverse and longitudinal magnetization rates. Besides, fitted parameters included the intensity scales for R_1, $R_{1\rho}$ and CEST experiments, yielding a total number of nine fitted parameters at each temperature. The CEST profiles implied a slow exchange process, therefore the R_ρ rate constants which were measured with a 4 kHz on-resonance spin-lock field, were assumed to be equal to the R_2 rates as they appear in the Bloch-

McConnell equation (eq. 42). During the CEST data analysis, normalized peak intensities were extracted from the spectra recorded as a function of the irradiation offset. Signal intensity errors were estimated from the signal-to-noise ratio of each peak, which were then used to produce 100 synthetic data set for Monte Carlo error estimation. The simulated data were generated by randomly sampling 100 points of a normal distribution whose mean value equals to the measured normalized intensity and the standard deviation equals to the measured error of the normalized intensity.

To fit the CEST intensity profiles we used the homogeneous form of the Bloch-McConnell differential equation system which describes the time evolution of the exchanging spin-states (two-state model) as they experience a continuous wave irradiation (ω_x).

$$\frac{d}{dt}\begin{bmatrix} E/2 \\ D_x \\ D_y \\ D_z \\ M_x \\ M_y \\ M_z \end{bmatrix} = - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & R_2+k_1 & \Omega_{\mathrm{D}} & -\omega_y & -k_{-1} & 0 & 0 \\ 0 & -\Omega_{\mathrm{D}} & R_2+k_1 & \omega_x & 0 & -k_{-1} & 0 \\ -2\Theta_D & \omega_y & -\omega_x & R_1+k_1 & 0 & 0 & -k_{-1} \\ 0 & -k_1 & 0 & 0 & R_2+k_{-1} & \Omega_{\mathrm{M}} & -\omega_y \\ 0 & 0 & -k_1 & 0 & -\Omega_{\mathrm{M}} & R_2+k_{-1} & \omega_x \\ -2\Theta_M & 0 & 0 & -k_1 & \omega_y & -\omega_x & R_1+k_{-1} \end{bmatrix} \begin{bmatrix} E/2 \\ D_x \\ D_y \\ D_z \\ M_x \\ M_y \\ M_z \end{bmatrix} \quad (42)$$

Here, $\Theta_D = 2p_{\mathrm{D}}R_1$ and $\Theta_M = 2p_{\mathrm{M}}R_1$ are the thermal equilibrium magnetizations in the dimer D and monomer M states which are proportional to the populations $p_{\mathrm{D}} = k_{-1}/k_{\mathrm{ex}}$ and $p_{\mathrm{M}} = k_1/k_{\mathrm{ex}}$.

The CEST profiles were fitted with eq. 42 using in-house written Mathematica scripts (Wolfram Research) assuming an initial magnetization state in the form of $D_x = 0$, $D_y = 0$, $D_z = p_{\mathrm{D}}$, $M_x = 0$, $M_y = 0$, $M_z = p_{\mathrm{M}}$ using the *NonlinearModelFit* function of Mathematica. Fitted parameters and their errors were obtained as the the mean and standard deviation of the result of the 100 individual CEST fits. CEST profiles measured at 58, 59, 60, 61, and 62 °C for Can$_{12}$ and at 56, 57, 58, 59, and 60 °C for Mod$_{12}$ were used to obtain the populations and exchange rates of the DNA melting process. The population of the monomer state at 57 and 55 °C for Can$_{12}$ and Mod$_{12}$ was too low to produce a reliable parameter fit for p_{D} and k_{ex}, therefore we excluded them from the analysis, and used those values only to validate and back-predict the exchange rates and populations based on the obtained thermodynamics and kinetics parameters.

From the obtained k_{ex}, p_{D}, and total DNA concentration (C_t) we derived k_d, k_a rates and K_d equilibrium constant at each temperature according to eq. 40, 41, and 2. These values were further analysed using the Eyring-Polanyi and van't Hoff formalism as described in the main text. The parameter errors of the CEST experiments were further propagated to the fits of the kinetic and thermodynamic equations by fitting the derived 100 synthetic data individually.

Supplementary Discussion: Considerations regarding Adenosine H2 and H8 protons

H2 and H8 protons of adenosine bases of both samples Can$_{12}$ and Mod$_{12}$ show moderate but systematic variations in the obtained thermodynamic parameters, reflecting the fact that the differences between the melting thermodynamics observed via adenosine H2 and H8 protons are related to the base type and not to the sequential position, and also implies that the two-site exchange model is illsuited to describe the base-pair opening mechanism even on the individual base pair level.

By inspecting the equilibrium constants of adenine H2 and H8 protons, we could establish that K_{eq} of H8 protons are always $1.5 - 2.5$ times higher than that of H2 protons, i.e. the H8 protons have a higher propensity to be found in the excited state conformation than the H2 protons have. This difference in

K_{eq} translates to an apparent 1-2 °C lower melting temperature for H8 than for H2 (c.f. Table 1 of the main text). Such scenario can only be explained by extending the two-state melting model and including a third species into the description:

$$\mathrm{dsDNA} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{Intermediate} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{ssDNA} \tag{43}$$

The higher equilibrium constant of H8 suggests that in the intermediate species the position and the chemical environment of H8 is similar to its environment in the ssDNA conformation, while the chemical environment of H2 has not changed and thus resembles the dsDNA state, presumably because the H-bonds and the $\pi - \pi$ stacking interactions are still intact. To gather further evidence for the presence of an intermediate species on the potential energy surface one could potentially measure the melting kinetics/thermodynamics of the sugar protons. If the above assumption is correct and the intermediate species has higher backbone flexibility than the dsDNA state has then the melting thermodynamics of the sugar protons will reflect similar behavior as H8 does.

Supplementary Figures
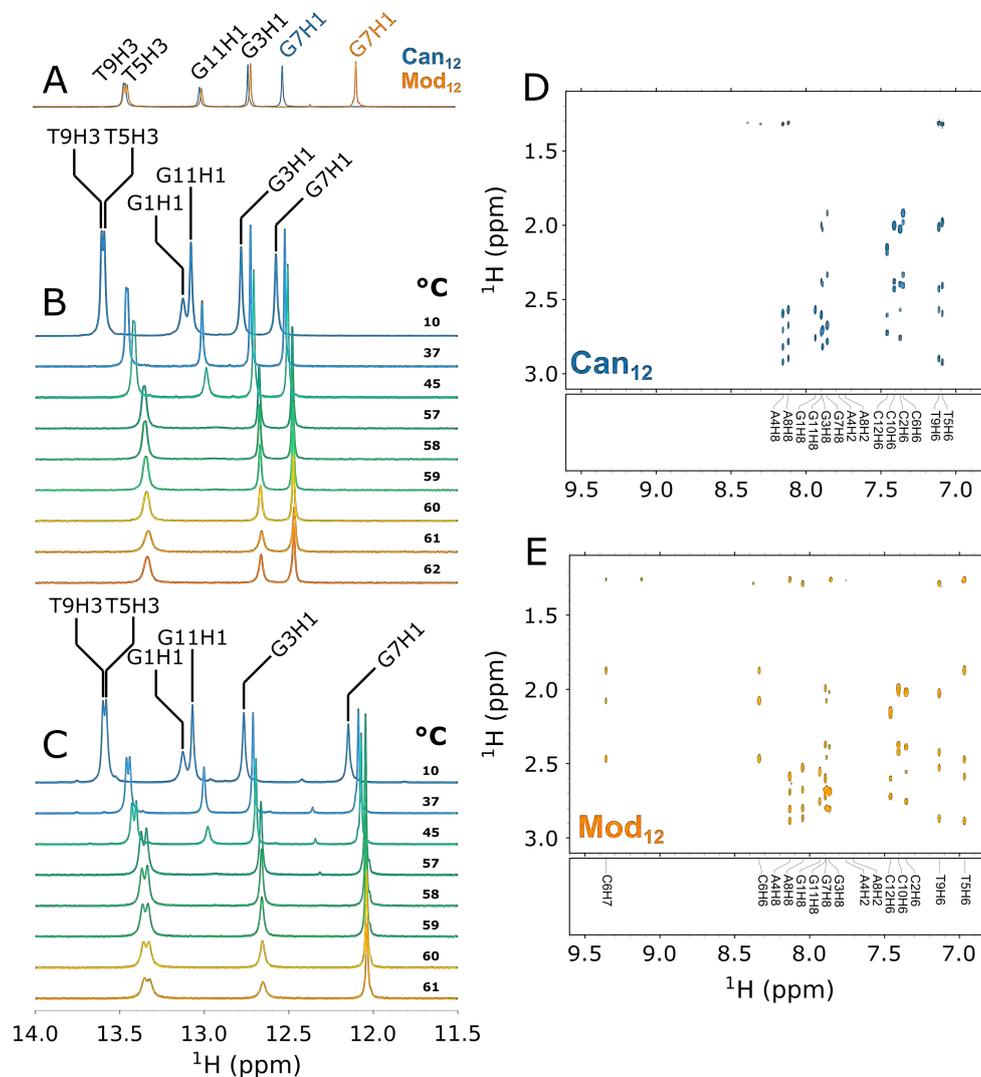
Figure S1: **A** 1D ^{1}H NMR spectrum of the ^{1}H imino region of Can$_{12}$ (blue) and Mod$_{12}$ (orange) measured at 37 °C at 800 MHz ^{1}H Larmor frequency. **B** and **C** NMR melting profiles of the imino proton region of Can$_{12}$ and Mod$_{12}$ between 10 and 62°C. **D** and **E** Exerts from the^{1}H-^{1}H NOESY spectra of Can$_{12}$ (blue) and Mod$_{12}$ (orange) showing the assignment of the H6/H2/H8 protons, measured at 37 °C.
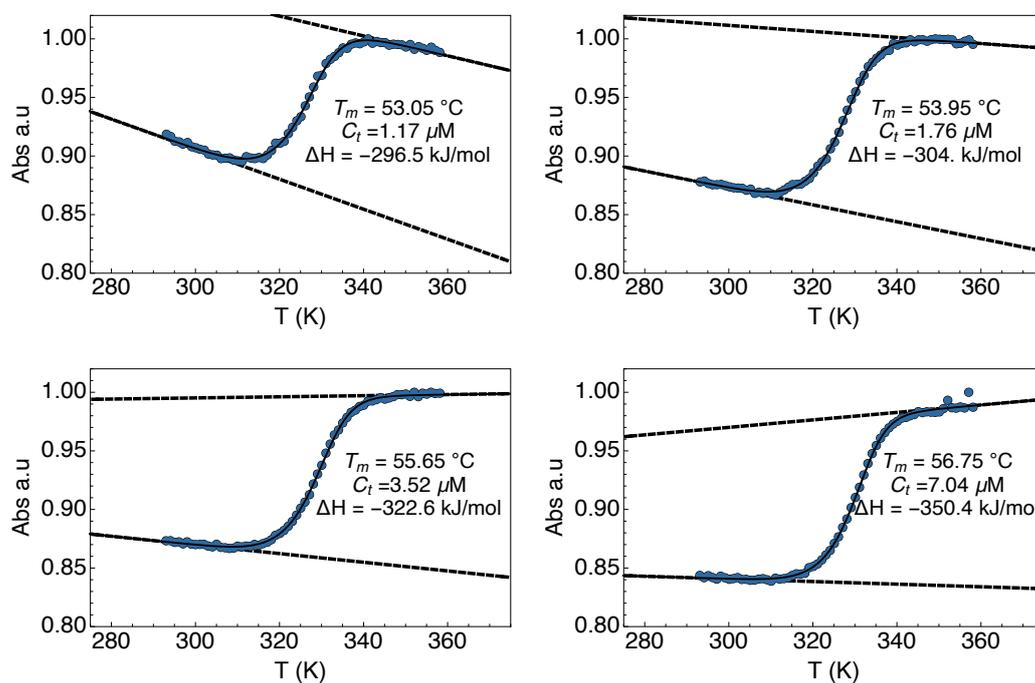
Figure S2: Representative UV melting profiles of Can_{12} measured at 1.17, 1.76, 3.52, and 7.04 μM total DNA concentration at 260 nm. Solid line represents the fitted melting function according to eq. 20, eq. 22, and eq. 21, while dashed lines serve as representation of the two baselines (upper and lower), needed in order to describe unfolded and folded forms, respectively, and identification of the T_m through the calculation of the median.
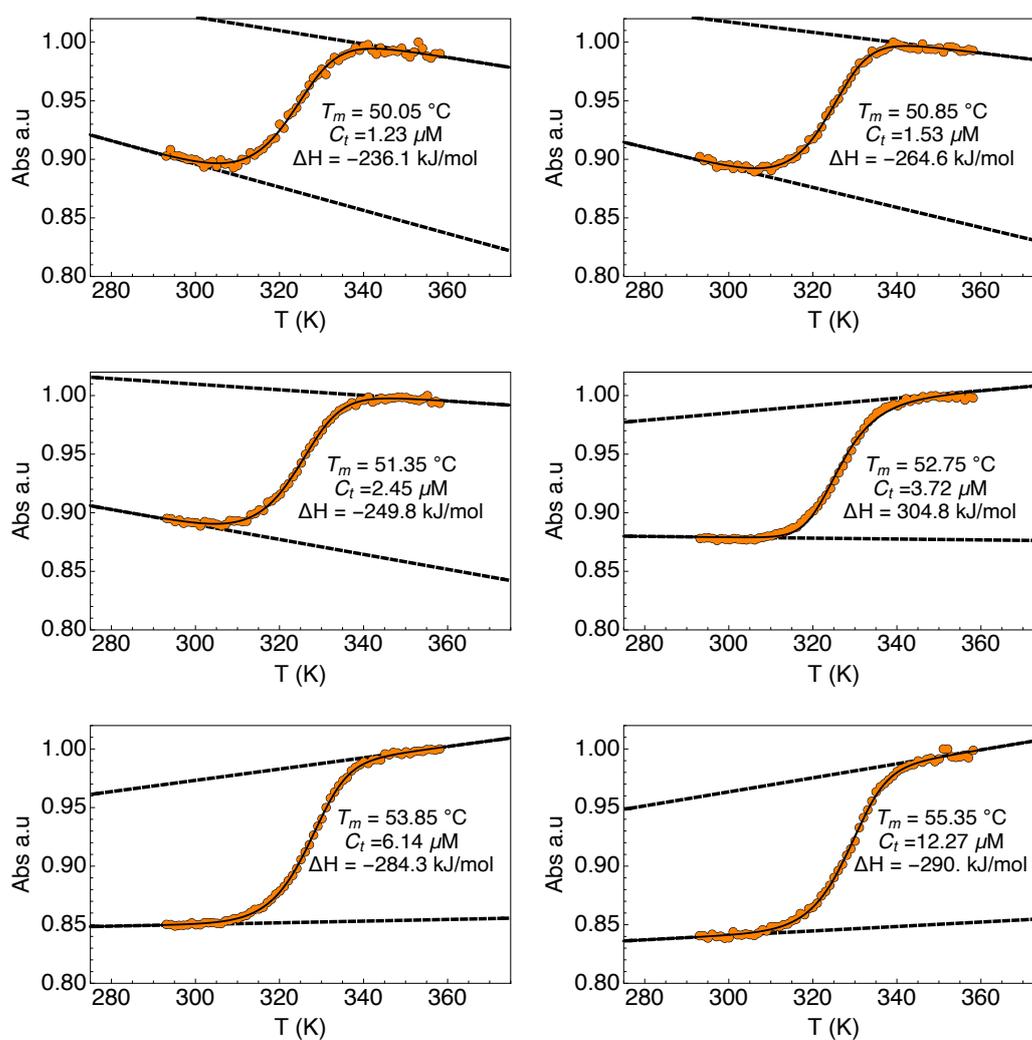
Figure S3: Representative UV melting profiles of Mod_{12} measured at 1.23, 1.53, 2.45, 3.72, 6.14, and 12.27 μM total DNA concentration at 260 nm. Solid line represents the fitted melting function according to eq. 20, eq. 22, and eq. 21, while dashed lines serve as representation of the two baselines (upper and lower), needed in order to describe unfolded and folded forms, respectively, and identification of the T_m through the calculation of the median.
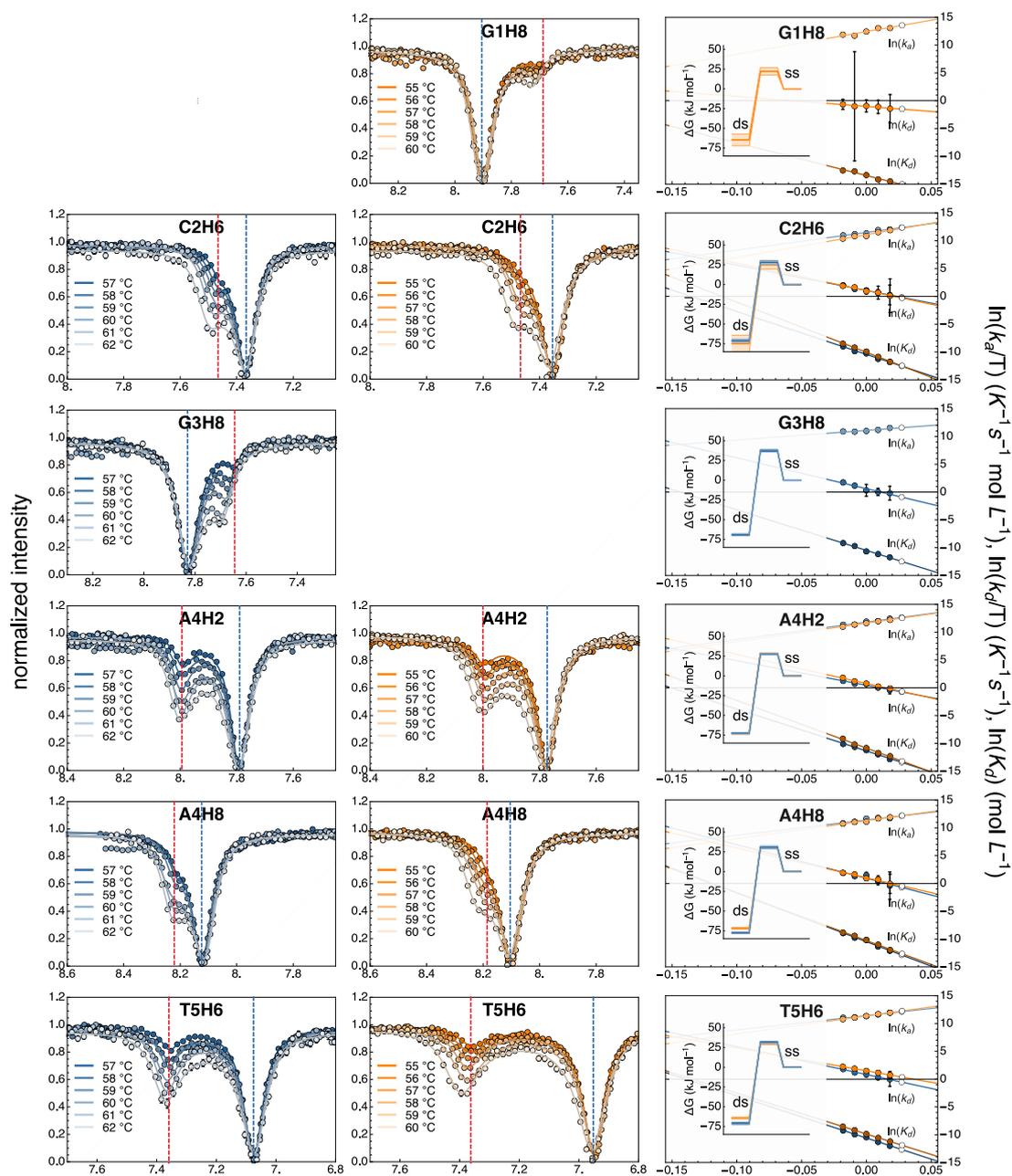
Figure S4: Continued on next page

S12
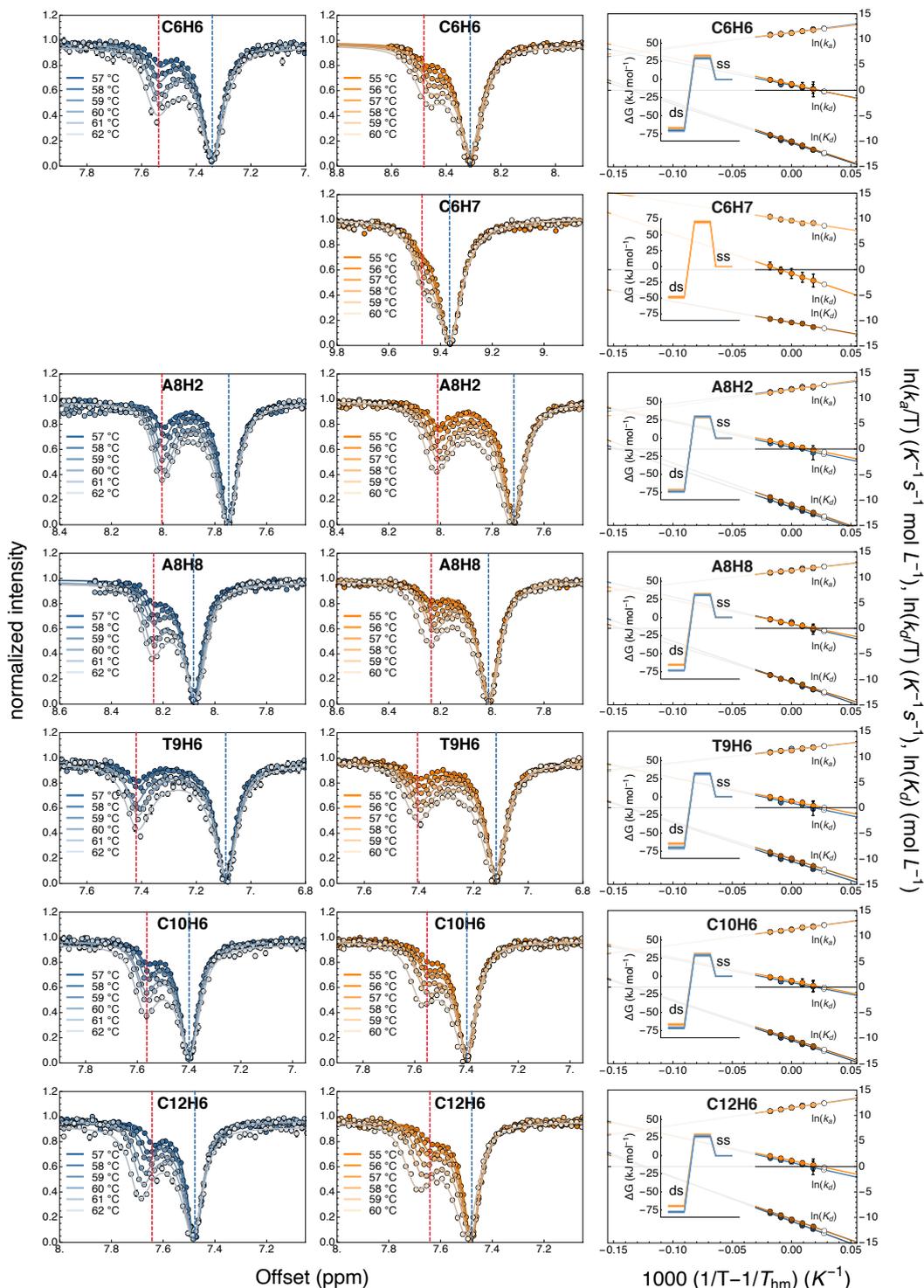
Figure S4: First and second column: Temperature-dependent ^{1}H CEST profiles of all uniquely accessible H2/H6/H8 sites of Can$_{12}$ (first column, blue) and Mod$_{12}$ (second column, orange). Third column: van't Hoff and Eyring plots of all sites displayed on the left. Shades of orange and blue indicate data entries and linear fits for Can$_{12}$ and Mod$_{12}$, respectively. White data points represent back-calculated values. Fitted values for the association and dissociation rate constants (k_a and k_d) and the dissociation equilibrium constant K_d are listed in Table S5 and S6. Insets present the relevant Gibbs free energy plots at each site at 37 °C.
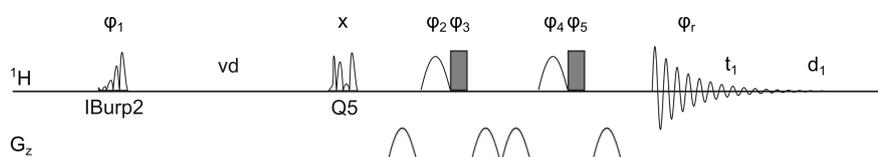
Figure S5: Pulse scheme for ^{1}H R_1 inversion recovery determination. Rates were measured in pseudo-2D interleaved fashion using fixed time increments labelled as "vd" in the pulse sequence. Filled grey symbols denote hard 180° pulses, while filled white items represent selective 180° pulses at water resonance. Selective ^{1}H irradiation of the spectral region of interest was achieved via pulse shapes IBurp2 (selective inversion) and Q5 (selective excitation), as labelled. Suppression of the water signal was achieved using the excitation sculpting scheme. The phase cycle was ϕ_1: x,-x; ϕ_2: x, y; ϕ_3: -x, -y; ϕ_4: x, x, y, y; ϕ_5: -x, -x, -y, -y; ϕ_r: x, -x, -x, x.
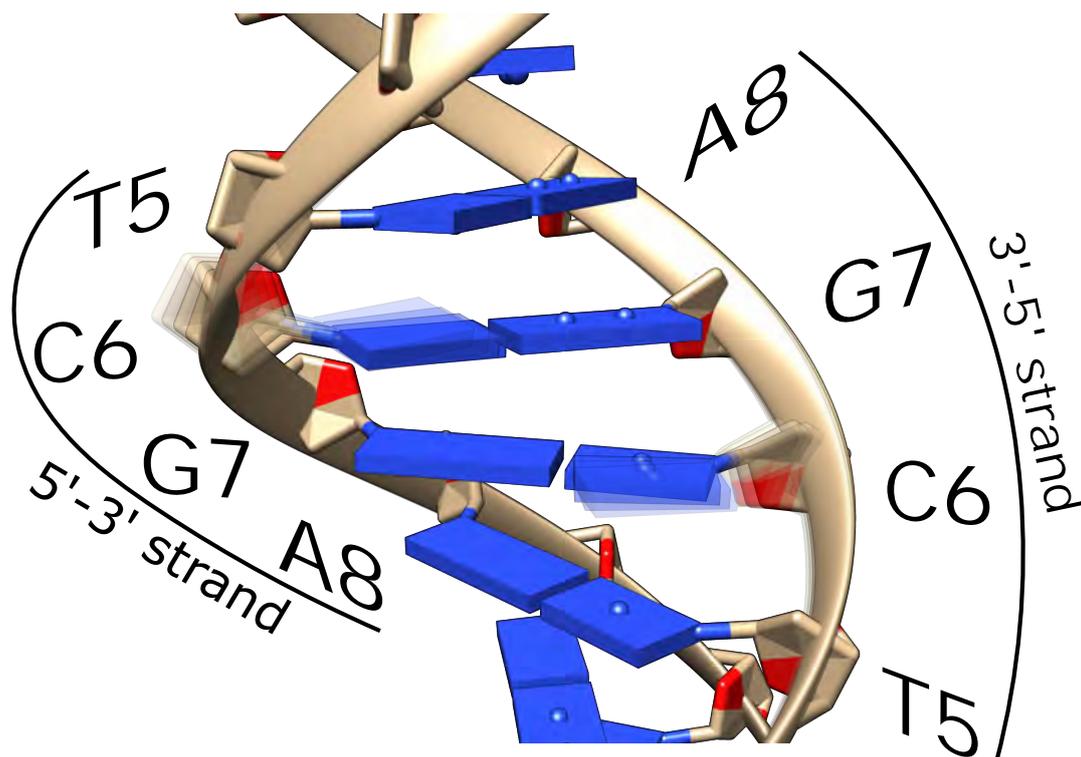
Figure S6: Molecular model for the visualization of 5fC-induced destabilization. The lower stability of the 5fC-G base pair implies a higher propensity for nucleotide vibrations around its canonical, Watson-Crick paired position.
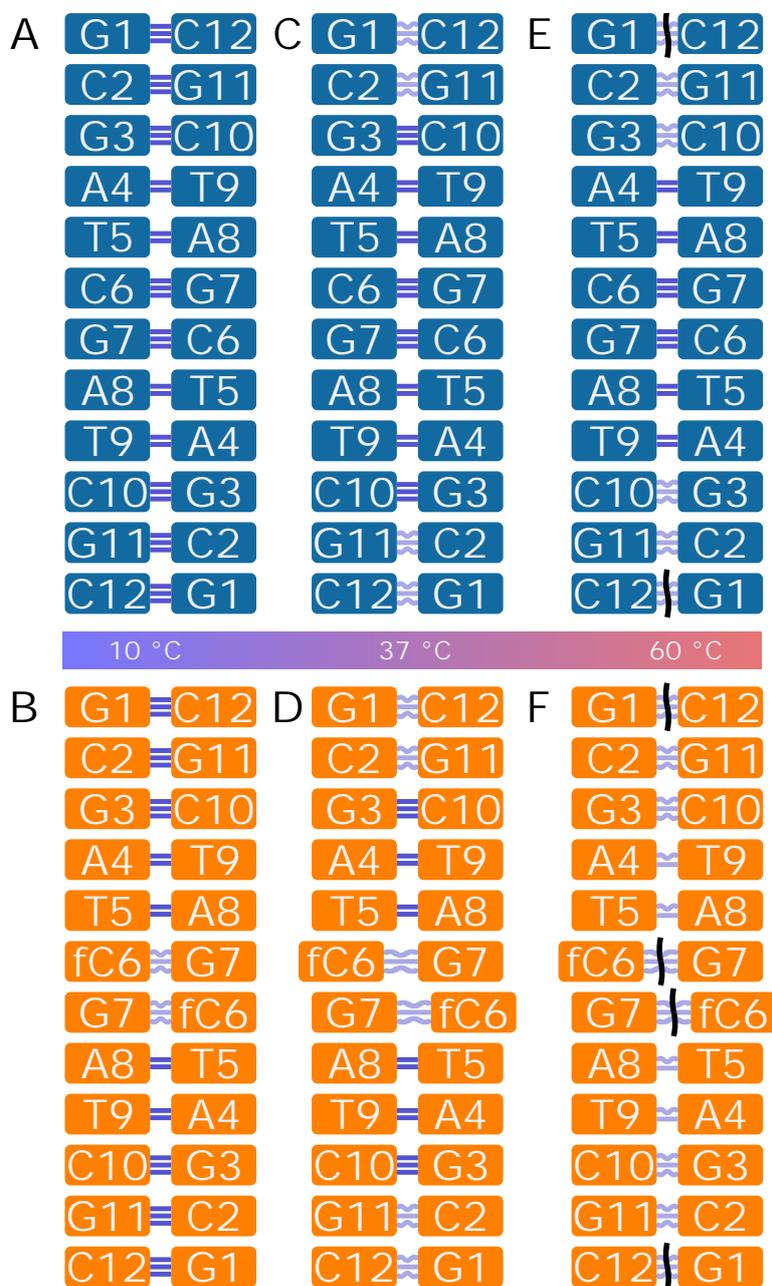
Figure S7: Schematic representation of the impact of 5fC on the melting dynamics. Blue bars symbolize Watson-Crick hydrogen bonding, gray wiggles identified weakened Watson-Crick hydrogen bonding and the black vertical wiggles indicate positions in the DNA that are prone to initiate dsDNA melting. (A, B) At 10 °C, all base pairs in both Can_{12} (top row) and Mod_{12} (bottom row) are formed and stable. (C, D) At intermediate temperatures, the 5'- and 3'-ends start fraying. The lower H-bonding strength associated with the 5fC-G base pair as compared with the canonical C-G pair generates instability in the middle of the sequence, in addition to the fraying at the termini. (E, F) Close to the melting temperatures, 5fC-G base pair is kinetically more prone to base-pair opening. This phenomenon accounts for the ~ 2 °C lower melting temperature of Mod_{12} with respect to Can_{12}.

Supplementary Tables

Table S1: Measurement conditions for the ^{1}H CEST experiments. The profiles obtained at 37, 45, and 53 °C were not used for parameters fitting or the followup thermodynamic analysis.

	Carrier frequency	Profile resolution	Condition
	Range, spacing (ppm)	Sampled points (−)	Temperature (°C)
	$6.7 - 8.2$ and $9.1 - 10.1$, 0.02	127	37
	$6.7 - 8.2$ and $9.1 - 10.1$, 0.02	127	45
	$6.75 - 8.47$ and $9.1 - 10.1$, 0.02	137	53
	$6.8 - 8.8$ and $9.1 - 10.1$, 0.02	152	55
	$6.8 - 8.8$ and $9.1 - 10.1$, 0.02	152	56
	$6.8 - 8.8$ and $9.1 - 10.1$, 0.02	152	57
Mod_{12}	$6.8 - 8.8$ and $9.1 - 10.1$, 0.02	152	58
	$6.8 - 8.8$ and $9.1 - 10.1$, 0.02	152	59
	$6.8 - 8.8$ and $9.1 - 10.1$, 0.02	152	60
	$6.9 - 8.4$, 0.02	76	37
	$7.1 - 8.4$, 0.02	66	45
	$6.8 - 8.4$, 0.02	81	53
	$7.0 - 8.6$, 0.02	81	57
	$7\text{-}0 - 8.4$, 0.02	81	58
Can_{12}	$7.0 - 8.6$, 0.02	81	59
	$7.0 - 8.6$, 0.02	81	60
	$7.0 - 8.6$, 0.02	81	61
	$7.0 - 8.6$, 0.02	81	62

Table S2: Comparison between random coil (calculated as per References 6, 7 using the program DSHIFT, available at http://www.chem.cuhk.edu.hk/DSHIFT/ran_input.asp) and experimentally determined ss-DNA for both Mod_{12} and Can_{12} at 60 °C. Errors are given as one standard deviation. Data labelled as "N/A" could not be determined either because of *(i)* lack of a suitable reference for the random coil predictor, or *(ii)* the site could not be uniquely assigned in our CEST experiment.

	Chemical shift		
	δ(Random coil prediction) (ppm)	$\delta(\text{Mod}_{12}$ ssDNA) (ppm)	$\delta(\text{Can}_{12}$ ssDNA) (ppm)
G1H8	N/A	7.722 ± 0.0004	N/A
C2H6	7.558	7.493 ± 0.0022	7.485 ± 0.0012
G3H8	7.773	N/A	
A4H2	8.051	8.012 ± 0.0017	8.004 ± 0.00073
A4H8	8.285	8.228 ± 0.0012	8.237 ± 0.0016
T5H6	7.406	7.399 ± 0.0036	7.368 ± 0.0016
C6H6	7.557	8.478 ± 0.0019	7.542 ± 0.0014
C6H7	N/A	9.466 ± 0.0016	N/A
G7H8	7.798	N/A	7.675 ± 0.048
A8H2	8.051	8.022 ± 0.0011	8.005 ± 0.00071
A8H8	8.285	8.25 ± 0.0012	8.246 ± 0.00091
T9H6	7.406	7.412 ± 0.0025	7.416 ± 0.0019
C10H6	7.590	7.581 ± 0.0019	7.571 ± 0.0012
G11H8	7.961	N/A	N/A
C12H6	N/A	7.699 ± 0.0016	7.679 ± 0.001

S17

Table S3: Results of the fit of ^{1}H CEST measured at 800 MHz ^{1}H Larmor frequency. The temperature of each measurement is indicated on the leftmost column of the table. Chemical shifts relevant to dsDNA and ssDNA are given together with the relaxation parameters R_1 and R_2 for each available reporter of Mod$_{12}$. Errors are given as one standard deviation.

		Chemical shift		Relaxation rates	
		δ(dsDNA) (ppm)	δ(ssDNA) (ppm)	R_1 (s^{-1})	R_2 (s^{-1})
	G1H8-H8	7.905 ± 0.34	7.688 ± 0.0015	0.882 ± 0.01	3.27 ± 0.04
	C2H6-H5	7.354 ± 5.9	7.468 ± 0.024	1.94 ± 0.04	8.92 ± 0.1
	A4H2-H2	7.773 ± 0.15	8.001 ± 0.00036	0.946 ± 0.02	2.8 ± 0.04
	A4H8-H8	8.1 ± 2.2	8.186 ± 0.021	1.69 ± 0.04	6.67 ± 0.04
Mod$_{12}$	T5H6-H6	6.952 ± 0.31	7.362 ± 0.0015	2.18 ± 0.05	6.53 ± 0.05
55 °C	C6H6-H6	8.313 ± 0.26	8.482 ± 0.0025	2.16 ± 0.05	9.18 ± 0.05
	A8H2-H2	7.716 ± 0.18	8.011 ± 0.00038	0.91 ± 0.02	2.95 ± 0.04
	A8H8-H8	8.016 ± 0.15	8.236 ± 0.00096	1.57 ± 0.03	6.01 ± 0.05
	T9H6-H6	7.12 ± 0.25	7.406 ± 0.0013	1.96 ± 0.04	6.09 ± 0.06
	C10H6-H5	7.399 ± 0.16	7.552 ± 0.0012	1.89 ± 0.04	5.87 ± 0.05
	C12H6-H5	7.48 ± 0.17	7.641 ± 0.0012	1.49 ± 0.04	4.03 ± 0.06
	G1H8	7.903 ± 0.27	7.685 ± 0.0009	0.857 ± 0.01	3.16 ± 0.04
	C2H6	7.354 ± 0.85	7.465 ± 0.021	1.86 ± 0.04	8.62 ± 0.2
	A4H2	7.774 ± 0.1	8.001 ± 0.00047	0.934 ± 0.02	2.85 ± 0.04
	A4H8	8.104 ± 0.067	8.206 ± 0.0011	1.66 ± 0.03	6.3 ± 0.04
	T5H6	6.952 ± 0.19	7.368 ± 0.0018	2.08 ± 0.04	6.31 ± 0.06
Mod$_{12}$	C6H6	8.313 ± 0.11	8.47 ± 0.0012	2.09 ± 0.04	9.69 ± 0.07
56 °C	C6H7	9.365 ± 0.047	9.472 ± 0.00072	1.09 ± 0.02	7.11 ± 0.08
	A8H2	7.717 ± 0.095	8.01 ± 0.0004	0.887 ± 0.01	2.66 ± 0.03
	A8H8	8.015 ± 0.1	8.23 ± 0.00077	1.52 ± 0.03	6.05 ± 0.06
	T9H6	7.12 ± 0.14	7.401 ± 0.0012	1.92 ± 0.05	5.87 ± 0.07
	C10H6	7.4 ± 0.096	7.557 ± 0.0011	1.83 ± 0.04	5.92 ± 0.06
	C12H6	7.481 ± 0.1	7.649 ± 0.0016	1.45 ± 0.03	3.93 ± 0.06
	G1H8	7.901 ± 0.21	7.701 ± 0.00092	0.859 ± 0.01	3.36 ± 0.04
	C2H6	7.355 ± 0.065	7.467 ± 0.0018	1.84 ± 0.04	8.68 ± 0.07
	A4H2	7.776 ± 0.084	$8. \pm 0.00039$	0.938 ± 0.02	3.03 ± 0.04
	A4H8	8.103 ± 0.055	8.21 ± 0.00076	1.63 ± 0.03	6.36 ± 0.04
	T5H6	6.954 ± 0.12	7.372 ± 0.0015	$2. \pm 0.04$	6.16 ± 0.05
Mod$_{12}$	C6H6	8.311 ± 0.061	8.471 ± 0.00096	2.06 ± 0.04	9.71 ± 0.05
57 °C	C6H7	9.367 ± 0.033	9.469 ± 0.00075	1.07 ± 0.02	7.63 ± 0.07
	A8H2	7.718 ± 0.063	8.013 ± 0.00038	0.886 ± 0.01	2.78 ± 0.03
	A8H8	8.014 ± 0.061	8.235 ± 0.00061	1.49 ± 0.03	6.07 ± 0.05
	T9H6	7.12 ± 0.081	7.405 ± 0.0012	1.87 ± 0.04	5.89 ± 0.05
	C10H6	7.398 ± 0.059	7.558 ± 0.0011	1.79 ± 0.04	5.82 ± 0.04
	C12H6	7.484 ± 0.076	7.659 ± 0.00089	1.43 ± 0.04	3.84 ± 0.05
	G1H8	7.9 ± 0.15	7.698 ± 0.00063	0.837 ± 0.01	3.08 ± 0.04
	C2H6	7.357 ± 0.029	7.475 ± 0.00098	1.83 ± 0.03	8.27 ± 0.06
	A4H2	7.778 ± 0.052	8.001 ± 0.00056	0.963 ± 0.02	3.09 ± 0.04
	A4H8	8.101 ± 0.024	8.206 ± 0.00065	1.62 ± 0.03	6.33 ± 0.04
	T5H6	6.96 ± 0.08	7.378 ± 0.0018	1.99 ± 0.04	5.91 ± 0.06
Mod$_{12}$	C6H6	8.312 ± 0.036	8.466 ± 0.00087	2.05 ± 0.05	9.26 ± 0.05
58 °C	C6H7	9.368 ± 0.033	9.47 ± 0.00088	1.04 ± 0.02	7.73 ± 0.07
	A8H2	7.721 ± 0.047	8.014 ± 0.00057	0.89 ± 0.01	2.75 ± 0.04
	A8H8	8.014 ± 0.044	8.234 ± 0.00076	1.48 ± 0.03	5.83 ± 0.05
	T9H6	7.123 ± 0.057	7.403 ± 0.0014	1.88 ± 0.04	5.77 ± 0.06
	C10H6	7.401 ± 0.039	7.564 ± 0.00084	1.81 ± 0.04	5.64 ± 0.05
	C12H6	7.487 ± 0.039	7.671 ± 0.00083	1.42 ± 0.03	3.72 ± 0.04
	G1H8	7.897 ± 3.4	7.708 ± 0.0061	0.823 ± 0.01	3.19 ± 0.2
Mod$_{12}$	C2H6	7.359 ± 0.022	7.478 ± 0.0014	1.78 ± 0.03	7.94 ± 0.09
59 °C	A4H2	7.781 ± 0.041	7.999 ± 0.00079	0.962 ± 0.02	3.06 ± 0.05

(*Continues on the next page...*)

Table S3: Results of the fit of ^{1}H CEST measured at 800 MHz ^{1}H Larmor frequency. The temperature of each measurement is indicated on the leftmost column of the table. Chemical shifts relevant to dsDNA and ssDNA are given together with the relaxation parameters R_1 and R_2 for each available reporter of Mod_{12}. Errors are given as one standard deviation.

		Chemical shift		Relaxation rates	
		δ(dsDNA) (ppm)	δ(ssDNA) (ppm)	R_1 (s^{-1})	R_2 (s^{-1})
	A4H8	8.102 ± 0.023	8.214 ± 0.0011	1.57 ± 0.03	6.05 ± 0.05
	T5H6	6.969 ± 0.061	7.381 ± 0.0029	1.85 ± 0.04	5.98 ± 0.1
	C6H6	8.311 ± 0.03	8.471 ± 0.0015	1.93 ± 0.04	9.05 ± 0.07
	C6H7	9.37 ± 0.026	9.466 ± 0.00089	1.02 ± 0.01	7.85 ± 0.06
	A8H2	7.724 ± 0.039	8.012 ± 0.00074	0.874 ± 0.01	2.96 ± 0.04
	A8H8	8.015 ± 0.034	8.235 ± 0.00092	1.43 ± 0.02	5.71 ± 0.05
	T9H6	7.126 ± 0.045	7.404 ± 0.0019	1.79 ± 0.03	5.49 ± 0.08
	C10H6	7.401 ± 0.024	7.561 ± 0.0011	1.72 ± 0.03	5.42 ± 0.05
	C12H6	7.492 ± 0.033	7.675 ± 0.0011	1.38 ± 0.03	3.83 ± 0.06
	G1H8	7.906 ± 0.093	7.722 ± 0.0004	0.829 ± 0.02	3.51 ± 0.06
	C2H6	7.376 ± 0.019	7.493 ± 0.0022	1.62 ± 0.03	7.85 ± 0.1
	A4H2	7.8 ± 0.041	8.012 ± 0.0017	1.03 ± 0.04	3.17 ± 0.06
	A4H8	8.115 ± 0.015	8.228 ± 0.0012	1.49 ± 0.03	5.79 ± 0.07
	T5H6	6.994 ± 0.042	7.399 ± 0.0036	1.7 ± 0.03	5.59 ± 0.1
Mod_{12}	C6H6	8.326 ± 0.021	8.478 ± 0.0019	1.84 ± 0.03	7.94 ± 0.1
60 °C	C6H7	9.369 ± 0.037	9.466 ± 0.0016	$1. \pm 0.01$	7.63 ± 0.1
	A8H2	7.742 ± 0.029	8.022 ± 0.0011	0.87 ± 0.01	3.13 ± 0.05
	A8H8	8.028 ± 0.027	8.25 ± 0.0012	1.3 ± 0.03	5.6 ± 0.07
	T9H6	7.147 ± 0.031	7.412 ± 0.0025	1.67 ± 0.03	5.06 ± 0.1
	C10H6	7.417 ± 0.022	7.581 ± 0.0019	1.57 ± 0.03	5.35 ± 0.1
	C12H6	7.512 ± 0.022	7.699 ± 0.0016	1.28 ± 0.03	3.57 ± 0.08

Table S4: Results of the fit of ^{1}H CEST measured at 800 MHz ^{1}H Larmor frequency. The temperature of each measurement is indicated on the leftmost column of the table. Chemical shifts relevant to dsDNA and ssDNA are given together with the relaxation parameters R_1 and R_2 for each available reporter of Can$_{12}$. Errors are given as one standard deviation.

		Chemical shift		Relaxation rates	
		δ(dsDNA) (ppm)	δ(ssDNA) (ppm)	$R_1\ (s^{-1})$	$R_2\ (s^{-1})$
Can$_{12}$ 57 °C	C2H6	7.367 ± 0.095	7.466 ± 0.0011	1.74 ± 0.03	5.21 ± 0.05
	A4H2	7.79 ± 0.12	7.994 ± 0.0004	0.787 ± 0.01	2.07 ± 0.03
	A4H8	8.124 ± 0.39	8.221 ± 0.0051	1.46 ± 0.03	5.35 ± 0.09
	T5H6	7.076 ± 0.2	7.359 ± 0.0012	1.77 ± 0.04	5.3 ± 0.09
	C6H6	7.344 ± 0.16	7.538 ± 0.0012	1.82 ± 0.04	6.15 ± 0.09
	G7H8	7.828 ± 0.17	7.645 ± 0.0011	1.58 ± 0.03	8.07 ± 0.08
	A8H2	7.747 ± 0.14	8.005 ± 0.00032	0.785 ± 0.01	2.09 ± 0.03
	A8H8	8.084 ± 0.097	8.237 ± 0.00065	1.5 ± 0.03	4.67 ± 0.06
	T9H6	7.094 ± 0.26	7.42 ± 0.0013	1.79 ± 0.03	5.59 ± 0.08
	C10H6	7.401 ± 0.11	7.563 ± 0.00083	1.75 ± 0.04	4.99 ± 0.07
	C12H6	7.477 ± 0.095	7.643 ± 0.00078	1.41 ± 0.02	3.71 ± 0.05
Can$_{12}$ 58 °C	C2H6	7.368 ± 0.06	7.475 ± 0.0014	1.73 ± 0.03	5.22 ± 0.06
	A4H2	7.792 ± 0.089	7.997 ± 0.00045	0.787 ± 0.01	2.17 ± 0.03
	A4H8	8.122 ± 0.052	8.218 ± 0.00084	1.47 ± 0.03	4.6 ± 0.05
	T5H6	7.076 ± 0.14	7.36 ± 0.001	1.76 ± 0.04	5.33 ± 0.09
	C6H6	7.343 ± 0.093	7.542 ± 0.0012	1.81 ± 0.04	6.25 ± 0.09
	G7H8	7.826 ± 0.066	7.658 ± 0.001	1.56 ± 0.03	8.45 ± 0.09
	A8H2	7.749 ± 0.17	7.998 ± 0.00063	0.783 ± 0.01	2.35 ± 0.04
	A8H8	8.083 ± 0.052	8.238 ± 0.00049	1.5 ± 0.03	4.6 ± 0.06
	T9H6	7.094 ± 0.17	7.419 ± 0.0013	1.79 ± 0.03	5.37 ± 0.08
	C10H6	7.401 ± 0.072	7.559 ± 0.00096	1.74 ± 0.04	5.05 ± 0.07
	C12H6	7.48 ± 0.083	7.659 ± 0.00095	1.4 ± 0.02	4.16 ± 0.05
Can$_{12}$ 59 °C	C2H6	7.368 ± 0.041	7.482 ± 0.00047	1.66 ± 0.04	4.87 ± 0.05
	A4H2	7.793 ± 0.049	8.001 ± 0.00028	0.774 ± 0.01	1.89 ± 0.03
	A4H8	8.12 ± 0.027	8.223 ± 0.00037	1.41 ± 0.03	4.45 ± 0.04
	T5H6	7.077 ± 0.068	7.363 ± 0.0007	1.68 ± 0.04	5.28 ± 0.08
	C6H6	7.344 ± 0.037	7.541 ± 0.0005	1.75 ± 0.03	5.98 ± 0.07
	G7H8	7.823 ± 0.03	7.664 ± 0.00052	1.5 ± 0.03	7.75 ± 0.09
	A8H2	7.75 ± 0.066	8.005 ± 0.00031	0.757 ± 0.01	1.98 ± 0.03
	A8H8	8.081 ± 0.037	8.244 ± 0.00029	1.45 ± 0.03	4.44 ± 0.04
	T9H6	7.097 ± 0.075	7.424 ± 0.00097	1.73 ± 0.04	4.94 ± 0.08
	C10H6	7.401 ± 0.037	7.566 ± 0.00037	1.7 ± 0.04	4.91 ± 0.06
	C12H6	7.481 ± 0.045	7.668 ± 0.00046	1.34 ± 0.02	3.34 ± 0.04
Can$_{12}$ 60 °C	C2H6	7.371 ± 0.027	7.485 ± 0.0012	1.68 ± 0.04	4.34 ± 0.08
	A4H2	7.794 ± 0.04	8.004 ± 0.00073	0.773 ± 0.01	1.29 ± 0.06
	A4H8	8.12 ± 0.024	8.237 ± 0.0016	1.41 ± 0.03	4.61 ± 0.06
	T5H6	7.08 ± 0.065	7.368 ± 0.0016	1.68 ± 0.04	5.1 ± 0.1
	C6H6	7.348 ± 0.034	7.542 ± 0.0014	1.72 ± 0.03	5.86 ± 0.1
	G7H8	7.822 ± 0.48	7.675 ± 0.048	1.48 ± 0.03	8.01 ± 0.3
	A8H2	7.751 ± 0.058	8.005 ± 0.00071	0.757 ± 0.01	1.8 ± 0.04
	A8H8	8.081 ± 0.028	8.246 ± 0.00091	1.44 ± 0.03	4.77 ± 0.06
	T9H6	7.098 ± 0.069	7.416 ± 0.0019	1.72 ± 0.04	5.04 ± 0.1
	C10H6	7.403 ± 0.038	7.571 ± 0.0012	1.69 ± 0.04	4.77 ± 0.09
	C12H6	7.486 ± 0.04	7.679 ± 0.001	1.34 ± 0.02	3.13 ± 0.08
Can$_{12}$ 61 °C	C2H6	7.372 ± 0.026	7.492 ± 0.0014	1.52 ± 0.03	4.34 ± 0.2
	A4H2	7.798 ± 0.044	7.998 ± 0.00094	0.765 ± 0.01	2.16 ± 0.06
	A4H8	8.122 ± 0.014	8.228 ± 0.00078	1.32 ± 0.02	3.99 ± 0.08
	T5H6	7.085 ± 0.053	7.37 ± 0.0022	1.5 ± 0.03	4.88 ± 0.2
	C6H6	7.35 ± 0.035	7.542 ± 0.0018	1.58 ± 0.03	5.52 ± 0.2
	G7H8	7.817 ± 0.016	7.689 ± 0.00098	1.39 ± 0.02	7.03 ± 0.2

(*Continues on the next page...*)

Table S4: Results of the fit of ^{1}H CEST measured at 800 MHz ^{1}H Larmor frequency. The temperature of each measurement is indicated on the leftmost column of the table. Chemical shifts relevant to dsDNA and ssDNA are given together with the relaxation parameters R_1 and R_2 for each available reporter of Can$_{12}$. Errors are given as one standard deviation.

		Chemical shift		Relaxation rates	
		δ(dsDNA) (ppm)	δ(ssDNA) (ppm)	R_1 (s^{-1})	R_2 (s^{-1})
	A8H2	7.754 ± 0.064	8.006 ± 0.0012	0.729 ± 0.02	1.88 ± 0.07
	A8H8	8.083 ± 0.024	8.246 ± 0.00087	1.35 ± 0.03	4.28 ± 0.1
	T9H6	7.102 ± 0.054	7.416 ± 0.002	1.54 ± 0.03	5.49 ± 0.2
	C10H6	7.404 ± 0.028	7.57 ± 0.0012	1.53 ± 0.03	5.22 ± 0.1
	C12H6	7.492 ± 0.028	7.683 ± 0.0013	1.25 ± 0.02	2.87 ± 0.1
	C2H6	7.379 ± 0.023	7.502 ± 0.0025	1.47 ± 0.03	4.55 ± 0.2
	A4H2	7.8 ± 0.036	8.004 ± 0.0014	0.762 ± 0.01	2.34 ± 0.1
	A4H8	8.124 ± 0.011	8.231 ± 0.0012	1.31 ± 0.02	3.85 ± 0.1
	T5H6	7.094 ± 0.059	7.374 ± 0.0042	1.51 ± 0.03	4.11 ± 0.2
	C6H6	7.356 ± 0.028	7.548 ± 0.0026	1.5 ± 0.04	5.46 ± 0.2
Can$_{12}$	G7H8	7.815 ± 0.022	7.691 ± 0.0021	1.36 ± 0.02	7.26 ± 0.2
62 °C	A8H2	7.76 ± 0.047	8.007 ± 0.0017	0.722 ± 0.02	2.24 ± 0.1
	A8H8	8.087 ± 0.018	8.248 ± 0.0015	1.3 ± 0.03	4.42 ± 0.1
	T9H6	7.115 ± 0.068	7.418 ± 0.0062	1.51 ± 0.05	5.11 ± 0.3
	C10H6	7.412 ± 0.025	7.575 ± 0.0025	1.51 ± 0.03	4.91 ± 0.2
	C12H6	7.5 ± 0.031	7.693 ± 0.0026	1.22 ± 0.02	3.25 ± 0.2

Table S5: Results of the thermodynamic and kinetic analysis based on of ^{1}H CEST fitted data measured at 800 MHz ^{1}H Larmor frequency. The relevant temperature for each of the parameters' sets is indicated on the leftmost column of the table. The population of dsDNA, the kinetic rates and the dissociation equilibrium constants are given for each available reporter of Mod$_{12}$. Errors are given as one standard deviation.

		dsDNA population	Kinetic rates				Eq. constants
		p_D $(-)$	k_{ex} (s^{-1})	k_d (s^{-1})	k_a $(10^3\ M^{-1}s^{-1})$	k_1 (s^{-1})	K_{eq} $(10^{-6}\ M^{-1})$
Mod$_{12}$ 56 °C	G1H8	0.9874 ± 0.00018	18.99 ± 0.61	0.2392 ± 0.0089	472.5 ± 15.	18.75 ± 0.6	0.5063 ± 0.015
	C2H6	0.9512 ± 0.016	23.9 ± 3.2	1.168 ± 0.41	182. ± 170.	22.73 ± 3.1	8.796 ± 5.4
	A4H2	0.9629 ± 0.00061	24.09 ± 0.32	0.8945 ± 0.019	198.3 ± 4.4	23.19 ± 0.31	4.513 ± 0.15
	A4H8	0.956 ± 0.0012	21.15 ± 1.6	0.9317 ± 0.081	145.8 ± 11.	20.22 ± 1.5	6.393 ± 0.37
	T5H6	0.9365 ± 0.0019	32.59 ± 1.3	2.071 ± 0.11	152.6 ± 7.5	30.52 ± 1.2	13.59 ± 0.83
	C6H6	0.9477 ± 0.0018	23.32 ± 1.7	1.219 ± 0.1	134.4 ± 10.	22.1 ± 1.6	9.088 ± 0.64
	C6H7	0.9382 ± 0.0017	1.915 ± 0.14	0.1184 ± 0.0093	9.233 ± 0.76	1.796 ± 0.14	12.85 ± 0.71
	A8H2	0.9654 ± 0.00039	24.44 ± 0.28	0.8458 ± 0.014	216.5 ± 3.5	23.6 ± 0.27	3.908 ± 0.09
	A8H8	0.9547 ± 0.001	22.76 ± 0.76	1.03 ± 0.043	152.5 ± 6.	21.73 ± 0.72	6.765 ± 0.3
	T9H6	0.9436 ± 0.0017	26.05 ± 1.3	1.469 ± 0.086	138.5 ± 8.	24.58 ± 1.2	10.62 ± 0.64
	C10H6	0.9465 ± 0.0016	24.48 ± 1.4	1.31 ± 0.083	137.7 ± 9.1	23.17 ± 1.3	9.532 ± 0.59
	C12H6	0.9518 ± 0.0013	27.83 ± 1.6	1.343 ± 0.085	174.4 ± 11.	26.48 ± 1.5	7.711 ± 0.42
Mod$_{12}$ 57 °C	G1H8	0.9851 ± 0.00021	22.26 ± 0.43	0.3308 ± 0.008	468.4 ± 11.	21.93 ± 0.43	0.7066 ± 0.02
	C2H6	0.9247 ± 0.0036	26.12 ± 2.5	1.968 ± 0.22	102. ± 11.	24.15 ± 2.3	19.36 ± 1.9
	A4H2	0.9506 ± 0.001	25.57 ± 0.38	1.263 ± 0.034	156.3 ± 3.8	24.3 ± 0.36	8.086 ± 0.34
	A4H8	0.9297 ± 0.003	28.69 ± 1.2	2.015 ± 0.11	120.8 ± 8.4	26.67 ± 1.2	16.77 ± 1.5
	T5H6	0.9136 ± 0.0025	36.38 ± 1.2	3.144 ± 0.15	122.2 ± 5.1	33.24 ± 1.1	25.78 ± 1.6
	C6H6	0.9232 ± 0.0024	28.44 ± 1.5	2.185 ± 0.13	108.6 ± 6.8	26.25 ± 1.4	20.17 ± 1.3
	C6H7	0.9146 ± 0.0023	2.583 ± 0.23	0.2206 ± 0.02	8.792 ± 0.84	2.363 ± 0.21	25.14 ± 1.4
	A8H2	0.9489 ± 0.00058	26.99 ± 0.33	1.378 ± 0.022	159.3 ± 2.9	25.61 ± 0.32	8.654 ± 0.2
	A8H8	0.9311 ± 0.0015	24.97 ± 0.72	1.721 ± 0.065	107.2 ± 3.8	23.25 ± 0.67	16.08 ± 0.74
	T9H6	0.9173 ± 0.0023	32.91 ± 1.3	2.723 ± 0.14	115.9 ± 5.3	30.19 ± 1.2	23.53 ± 1.4
	C10H6	0.9259 ± 0.0023	31.28 ± 1.7	2.32 ± 0.15	124.1 ± 7.5	28.96 ± 1.6	18.72 ± 1.2
	C12H6	0.9334 ± 0.0021	30.11 ± 0.8	2.006 ± 0.074	134.1 ± 6.3	28.11 ± 0.76	15. ± 0.97
Mod$_{12}$ 58 °C	G1H8	0.978 ± 0.00037	17.86 ± 0.39	0.3922 ± 0.01	252.6 ± 7.6	17.47 ± 0.39	1.554 ± 0.053
	C2H6	0.8822 ± 0.0039	21.79 ± 1.3	2.566 ± 0.17	51.89 ± 3.8	19.23 ± 1.2	49.59 ± 3.5
	A4H2	0.9242 ± 0.0016	28.52 ± 0.31	2.163 ± 0.052	110.4 ± 2.8	26.36 ± 0.29	19.61 ± 0.86
	A4H8	0.8956 ± 0.0027	24.04 ± 0.91	2.509 ± 0.12	65.55 ± 3.	21.53 ± 0.81	38.34 ± 2.1
	T5H6	0.8791 ± 0.0029	37.08 ± 1.2	4.485 ± 0.17	85.62 ± 3.7	32.6 ± 1.	52.46 ± 2.7
	C6H6	0.8868 ± 0.0033	29.79 ± 1.7	3.374 ± 0.21	74.14 ± 4.9	26.42 ± 1.5	45.6 ± 2.8
	C6H7	0.9044 ± 0.0032	4.944 ± 0.48	0.4728 ± 0.048	14.86 ± 1.5	4.471 ± 0.43	31.89 ± 2.2
	A8H2	0.9254 ± 0.001	29. ± 0.34	2.163 ± 0.038	114.2 ± 2.1	26.83 ± 0.31	18.94 ± 0.53
	A8H8	0.9041 ± 0.0018	27.16 ± 0.68	2.605 ± 0.084	81.31 ± 2.5	24.56 ± 0.61	32.06 ± 1.2
	T9H6	0.8832 ± 0.003	30.75 ± 1.1	3.592 ± 0.14	73.84 ± 3.6	27.15 ± 0.96	48.74 ± 2.7
	C10H6	0.8875 ± 0.0033	31.13 ± 1.1	3.503 ± 0.15	78.01 ± 3.8	27.63 ± 0.96	45. ± 2.8
	C12H6	0.9059 ± 0.002	33.73 ± 0.79	3.174 ± 0.1	103.1 ± 3.5	30.55 ± 0.72	30.82 ± 1.4
Mod$_{12}$ 59 °C	G1H8	0.9679 ± 0.019	11.14 ± 4.1	0.3622 ± 0.24	146.4 ± 250.	10.78 ± 4.	4.623 ± 4.3
	C2H6	0.836 ± 0.0059	31.1 ± 1.8	5.1 ± 0.34	50.42 ± 3.6	26. ± 1.5	101.5 ± 8.1
	A4H2	0.8972 ± 0.0019	31.17 ± 0.45	3.205 ± 0.073	86.4 ± 2.1	27.97 ± 0.41	37.12 ± 1.4
	A4H8	0.8518 ± 0.0046	34.31 ± 1.6	5.083 ± 0.26	62.71 ± 4.	29.23 ± 1.4	81.29 ± 5.5
	T5H6	0.8337 ± 0.0043	39.97 ± 1.4	6.646 ± 0.31	63.68 ± 2.9	33.32 ± 1.2	104.5 ± 6.
	C6H6	0.8506 ± 0.0047	35.7 ± 1.8	5.336 ± 0.34	64.56 ± 3.8	30.36 ± 1.6	82.82 ± 5.7
	C6H7	0.8869 ± 0.004	7.626 ± 0.61	0.8628 ± 0.075	19.01 ± 1.7	6.764 ± 0.54	45.53 ± 3.4
	A8H2	0.9002 ± 0.0015	31.16 ± 0.45	3.108 ± 0.063	89.29 ± 2.	28.05 ± 0.41	34.83 ± 1.1
	A8H8	0.881 ± 0.0021	30.08 ± 0.75	3.581 ± 0.12	70.7 ± 2.	26.5 ± 0.65	50.68 ± 1.9
	T9H6	0.8533 ± 0.0032	38.78 ± 1.2	5.689 ± 0.2	71.64 ± 3.	33.09 ± 1.	79.51 ± 3.8
	C10H6	0.8491 ± 0.0037	34.7 ± 1.1	5.237 ± 0.2	62.01 ± 2.7	29.46 ± 0.92	84.6 ± 4.5

(*Continues on the next page...*)

Table S5: Results of the thermodynamic and kinetic analysis based on of ^{1}H CEST fitted data measured at 800 MHz ^{1}H Larmor frequency. The relevant temperature for each of the parameters' sets is indicated on the leftmost column of the table. The population of dsDNA, the kinetic rates and the dissociation equilibrium constants are given for each available reporter of Mod_{12}. Errors are given as one standard deviation.

		dsDNA population	Kinetic rates				Eq. constants
		p_D $(-)$	k_{ex} (s^{-1})	k_d (s^{-1})	k_a $(10^3\ M^{-1}s^{-1})$	k_1 (s^{-1})	K_{eq} $(10^{-6}\ M^{-1})$
	C12H6	0.8604 ± 0.0035	36.07 ± 0.98	5.034 ± 0.19	70.65 ± 2.8	31.04 ± 0.85	71.35 ± 3.9
Mod_{12} 60 °C	G1H8	0.9669 ± 0.00058	15.62 ± 0.49	0.5165 ± 0.018	145.1 ± 5.4	15.1 ± 0.48	3.563 ± 0.13
	C2H6	0.7762 ± 0.0095	29.34 ± 2.1	6.568 ± 0.56	32.37 ± 2.9	22.77 ± 1.7	$203.9 \pm 20.$
	A4H2	0.8335 ± 0.0054	36.55 ± 0.69	6.086 ± 0.25	58.15 ± 2.3	30.46 ± 0.57	104.9 ± 7.4
	A4H8	0.806 ± 0.0045	34.85 ± 1.1	6.76 ± 0.26	45.99 ± 1.9	28.09 ± 0.87	147.2 ± 7.6
	T5H6	0.7807 ± 0.0054	42.45 ± 1.4	9.308 ± 0.37	48.01 ± 2.2	33.14 ± 1.1	$194.3 \pm 11.$
	C6H6	0.7985 ± 0.0061	39.67 ± 2.2	7.996 ± 0.56	49.94 ± 3.1	31.67 ± 1.8	$160.4 \pm 11.$
	C6H7	0.8681 ± 0.0072	15.46 ± 1.7	2.039 ± 0.24	32.43 ± 4.2	13.42 ± 1.5	63.39 ± 7.4
	A8H2	0.8528 ± 0.0022	34.41 ± 0.61	5.066 ± 0.12	63.27 ± 1.5	29.34 ± 0.52	80.12 ± 2.6
	A8H8	0.8338 ± 0.004	35.7 ± 0.95	5.934 ± 0.21	56.87 ± 2.3	29.76 ± 0.81	104.5 ± 5.5
	T9H6	0.7961 ± 0.0056	41.2 ± 1.4	8.4 ± 0.39	51.12 ± 2.2	32.8 ± 1.1	$164.6 \pm 10.$
	C10H6	0.8065 ± 0.0056	37.52 ± 1.3	7.258 ± 0.3	49.72 ± 2.6	30.26 ± 1.1	146.4 ± 9.5
	C12H6	0.8065 ± 0.0049	$40.25 \pm 1.$	7.786 ± 0.26	53.31 ± 2.3	32.46 ± 0.88	146.3 ± 8.3

Table S6: Results of the thermodynamic and kinetic analysis based on of ^{1}H CEST fitted data measured at 800 MHz ^{1}H Larmor frequency. The relevant temperature for each of the parameters' sets is indicated on the leftmost column of the table. The population of dsDNA, the kinetic rates and the dissociation equilibrium constants are given for each available reporter of Can$_{12}$. Errors are given as one standard deviation.

		dsDNA population	Kinetic rates				Eq. constants
		p_D $(-)$	k_{ex} (s^{-1})	k_d (s^{-1})	k_a $(10^3\ M^{-1}s^{-1})$	k_1 (s^{-1})	K_{eq} $(10^{-6}\ M^{-1})$
Can$_{12}$ 58 °C	C2H6	0.94 ± 0.0019	21.9 ± 2.3	1.315 ± 0.15	171.7 ± 18.	20.59 ± 2.2	7.665 ± 0.49
	A4H2	0.9641 ± 0.00056	18.68 ± 0.42	0.67 ± 0.019	251.1 ± 6.9	18.01 ± 0.41	2.669 ± 0.085
	A4H8	0.9491 ± 0.0014	13.2 ± 1.5	0.6716 ± 0.076	123.2 ± 14.	12.53 ± 1.4	5.462 ± 0.3
	T5H6	0.9459 ± 0.0015	18.08 ± 1.3	0.9795 ± 0.079	158. ± 11.	17.1 ± 1.2	6.203 ± 0.36
	C6H6	0.9383 ± 0.0018	21.21 ± 1.1	1.31 ± 0.08	161.3 ± 9.5	19.9 ± 1.	8.13 ± 0.48
	G7H8	0.9387 ± 0.0014	13.47 ± 0.97	0.8254 ± 0.064	103.2 ± 7.6	12.64 ± 0.91	8.007 ± 0.38
	A8H2	0.9635 ± 0.00089	13.23 ± 0.68	0.4834 ± 0.028	174.6 ± 9.9	12.75 ± 0.65	2.771 ± 0.14
	A8H8	0.9465 ± 0.001	14.6 ± 0.66	0.7808 ± 0.039	129.2 ± 6.2	13.82 ± 0.62	6.045 ± 0.23
	T9H6	0.9475 ± 0.0015	16.11 ± 1.1	0.8462 ± 0.061	145.4 ± 10.	15.26 ± 1.	5.829 ± 0.34
	C10H6	0.9408 ± 0.0015	18.62 ± 1.1	1.102 ± 0.071	148. ± 9.8	17.51 ± 1.	7.462 ± 0.38
	C12H6	0.9527 ± 0.00095	16.12 ± 0.99	0.7618 ± 0.05	162.5 ± 10.	15.36 ± 0.94	4.69 ± 0.19
Can$_{12}$ 59 °C	C2H6	0.9221 ± 0.0022	25.33 ± 1.1	1.973 ± 0.1	150.1 ± 8.1	23.36 ± 1.1	13.17 ± 0.76
	A4H2	0.9454 ± 0.00077	18.77 ± 0.23	1.025 ± 0.02	162.4 ± 2.9	17.74 ± 0.21	6.311 ± 0.18
	A4H8	0.9268 ± 0.0016	20.41 ± 1.2	1.495 ± 0.1	129.2 ± 8.	18.92 ± 1.1	11.58 ± 0.52
	T5H6	0.9245 ± 0.0016	18.42 ± 0.59	1.391 ± 0.048	112.8 ± 4.9	17.03 ± 0.55	12.35 ± 0.55
	C6H6	0.9156 ± 0.0015	19.68 ± 0.85	1.661 ± 0.081	106.8 ± 4.8	18.02 ± 0.77	15.56 ± 0.58
	G7H8	0.9191 ± 0.0013	12.59 ± 0.62	1.018 ± 0.053	71.53 ± 3.8	11.57 ± 0.57	14.24 ± 0.49
	A8H2	0.9499 ± 0.00068	21.28 ± 0.21	1.065 ± 0.017	201.9 ± 3.7	20.21 ± 0.2	5.28 ± 0.15
	A8H8	0.9279 ± 0.0014	20.61 ± 0.55	1.486 ± 0.05	132.7 ± 4.3	19.12 ± 0.51	11.21 ± 0.45
	T9H6	0.9102 ± 0.0022	20.32 ± 0.73	1.825 ± 0.085	103.1 ± 4.3	18.5 ± 0.66	17.73 ± 0.92
	C10H6	0.9216 ± 0.0016	19.71 ± 0.64	1.546 ± 0.062	115.8 ± 4.3	18.17 ± 0.59	13.36 ± 0.57
	C12H6	0.9345 ± 0.0011	22.39 ± 0.52	1.465 ± 0.038	159.9 ± 5.2	20.93 ± 0.49	9.171 ± 0.33
Can$_{12}$ 60 °C	C2H6	0.8752 ± 0.004	23.84 ± 1.1	2.975 ± 0.15	83.71 ± 5.3	20.87 ± 0.98	35.65 ± 2.4
	A4H2	0.9219 ± 0.0013	25.34 ± 0.54	1.979 ± 0.052	149.6 ± 4.2	23.36 ± 0.5	13.24 ± 0.46
	A4H8	0.8716 ± 0.0039	34.35 ± 1.9	4.41 ± 0.28	116.7 ± 7.6	29.93 ± 1.7	37.88 ± 2.4
	T5H6	0.887 ± 0.0033	21.17 ± 1.1	2.393 ± 0.15	83.12 ± 5.	18.78 ± 0.99	28.84 ± 1.8
	C6H6	0.8765 ± 0.0028	22.9 ± 1.3	2.828 ± 0.17	81.29 ± 5.1	20.07 ± 1.1	34.83 ± 1.7
	G7H8	0.8986 ± 0.017	13.52 ± 1.6	1.369 ± 0.27	62.18 ± 16.	12.15 ± 1.4	23.67 ± 7.8
	A8H2	0.9274 ± 0.0014	21.06 ± 0.53	1.528 ± 0.047	134.6 ± 4.4	19.53 ± 0.49	11.36 ± 0.44
	A8H8	0.8727 ± 0.0026	26.73 ± 0.93	3.404 ± 0.14	91.67 ± 3.7	23.33 ± 0.81	37.16 ± 1.6
	T9H6	0.878 ± 0.0035	30.19 ± 1.6	3.683 ± 0.21	108.7 ± 7.2	26.51 ± 1.4	33.96 ± 2.1
	C10H6	0.8795 ± 0.0035	22.9 ± 1.	2.759 ± 0.14	83.68 ± 4.8	20.14 ± 0.9	33.04 ± 2.1
	C12H6	0.8973 ± 0.0023	23.88 ± 0.81	2.453 ± 0.099	104.4 ± 4.4	21.43 ± 0.73	23.53 ± 1.1
Can$_{12}$ 61 °C	C2H6	0.8508 ± 0.0059	25.18 ± 1.5	3.759 ± 0.27	71.9 ± 5.3	21.42 ± 1.3	52.47 ± 4.5
	A4H2	0.8962 ± 0.0028	19.84 ± 0.7	2.061 ± 0.091	85.69 ± 4.	17.78 ± 0.63	24.09 ± 1.4
	A4H8	0.8454 ± 0.003	23.74 ± 0.88	3.672 ± 0.15	64.92 ± 2.9	20.07 ± 0.75	56.61 ± 2.4
	T5H6	0.8566 ± 0.004	22.82 ± 1.2	3.272 ± 0.18	68.22 ± 4.4	19.55 ± 1.	48.08 ± 2.9
	C6H6	0.8455 ± 0.0045	21.09 ± 1.4	3.259 ± 0.25	57.75 ± 4.2	17.83 ± 1.2	56.54 ± 3.6
	G7H8	0.8404 ± 0.0036	19.74 ± 0.96	3.152 ± 0.18	51.99 ± 2.7	16.59 ± 0.8	60.68 ± 3.
	A8H2	0.8975 ± 0.003	21.87 ± 0.64	2.24 ± 0.09	95.87 ± 4.3	19.63 ± 0.58	23.41 ± 1.4
	A8H8	0.8543 ± 0.0031	22.75 ± 0.76	3.314 ± 0.15	66.73 ± 2.5	19.43 ± 0.64	49.71 ± 2.3
	T9H6	0.8632 ± 0.0029	21.78 ± 1.2	2.978 ± 0.17	68.79 ± 4.5	18.8 ± 1.1	43.36 ± 2.
	C10H6	0.8547 ± 0.0038	20.73 ± 1.2	3.01 ± 0.18	61.03 ± 4.	17.72 ± 1.	49.42 ± 2.8
	C12H6	0.858 ± 0.003	24.7 ± 0.88	3.507 ± 0.15	74.69 ± 3.2	21.2 ± 0.76	47.01 ± 2.1
Can$_{12}$ 62 °C	C2H6	0.78 ± 0.01	33.01 ± 3.	7.26 ± 0.72	58.71 ± 6.5	25.75 ± 2.4	124.5 ± 13.
	A4H2	0.8674 ± 0.0036	25.87 ± 0.99	3.429 ± 0.16	84.68 ± 4.2	22.44 ± 0.86	40.57 ± 2.3
	A4H8	0.7753 ± 0.0051	28.58 ± 1.1	6.424 ± 0.32	49.34 ± 2.1	22.16 ± 0.84	130.4 ± 6.8

(Continues on the next page...)

Table S6: Results of the thermodynamic and kinetic analysis based on of ^{1}H CEST fitted data measured at 800 MHz ^{1}H Larmor frequency. The relevant temperature for each of the parameters' sets is indicated on the leftmost column of the table. The population of dsDNA, the kinetic rates and the dissociation equilibrium constants are given for each available reporter of Can_{12}. Errors are given as one standard deviation.

	dsDNA population	Kinetic rates				Eq. constants
	$p_\mathbf{D}$ (−)	$k_\mathbf{ex}$ (s^{-1})	k_d (s^{-1})	k_a ($10^3\ M^{-1}s^{-1}$)	k_1 (s^{-1})	$K_\mathbf{eq}$ ($10^{-6}\ M^{-1}$)
T5H6	0.8027 ± 0.0079	31.09 ± 1.7	6.135 ± 0.41	63.36 ± 4.8	24.96 ± 1.4	97.27 ± 8.6
C6H6	0.7715 ± 0.0085	32.43 ± 1.8	7.411 ± 0.48	54.86 ± 4.1	25.02 ± 1.4	135.7 ± 12.
G7H8	0.7958 ± 0.0082	28.2 ± 2.5	5.763 ± 0.59	55.02 ± 5.1	22.44 ± 1.9	105. ± 9.6
A8H2	0.864 ± 0.0034	22.83 ± 0.93	3.106 ± 0.15	72.56 ± 3.6	19.73 ± 0.8	42.87 ± 2.3
A8H8	0.7951 ± 0.0056	27.84 ± 1.2	5.706 ± 0.28	54.06 ± 3.	22.14 ± 0.96	105.8 ± 6.5
T9H6	0.7909 ± 0.011	29.63 ± 1.6	6.2 ± 0.51	56.2 ± 4.5	23.43 ± 1.3	111.1 ± 14.
C10H6	0.7785 ± 0.0086	28.69 ± 1.4	6.354 ± 0.37	50.52 ± 3.8	22.34 ± 1.2	126.4 ± 11.
C12H6	0.7962 ± 0.0073	25.75 ± 1.5	5.247 ± 0.33	50.39 ± 3.9	20.5 ± 1.2	104.6 ± 8.4

References

[1] Jean-louis Mergny and Laurent Lacroix. Analysis of Thermal Melting Curves. *Oligonucleotides*, 13:515–537, 2003.

[2] Albrecht Böttcher, Danny Kowerko, and Roland K.O. Sigel. Explicit analytic equations for multi-molecular thermal melting curves. *Biophys. Chem.*, 202:32–39, 2015.

[3] Ryan J Menssen and Andrei Tokmakoff. Length-Dependent Melting Kinetics of Short DNA Oligonucleotides Using Temperature-Jump IR Spectroscopy. *J. Phys. Chem. B*, 123:756–767, 2019.

[4] Anatoliy Dragan, Peter Privalov, and Colyn Crane-Robinson. Thermodynamics of DNA: heat capacity changes on duplex unfolding. *Eur. Biophys. J.*, 48(8):773–779, 2019.

[5] B. Nöltig. *Protein Folding Kinetics, Biophysical Methods*, volume 53. Springer: Berlin, 2006.

[6] Sik Lok Lam, Lai Nang Ip, Xiaodai Cui, and Cheuk Nang Ho. Random coil proton chemical shifts of deoxyribonucleic acids. *J. Biomol. NMR*, 24:329–337, 2002.

[7] Sik Lok Lam. DSHIFT: A web server for predicting DNA chemical shifts. *Nucleic Acids Res.*, 35(SUPPL.2):713–717, 2007.

3 | MECHANISTIC BASIS OF 5CAC-INDUCED DE-/STABILIZATION

3.1 FOREWORD AND MAIN PUBLICATION

5caC and 5fC, the two most oxidized derivatives of 5mC, share a number of similarities. They are both genomically stable, semipermanent although exceedingly sparse epigenetically nucleobases. Within the active demethylation pathway, they both are substrates for TDG enzymes, acting as necessary intermediates in erasing methylated portions of DNA. In contrast, a key difference is represented by the intrinsic chemical nature of the formyl versus the carboxyl functionalities. While the former acts as an electron-withdrawing group under any physiological condition, the latter's impact ultimately depends on the acidity, basicity or neutrality of the pH of the solvent. The chameleonic behaviour of 5caC has been a recurring item of research interest in enzymology, structural biology and biophysical chemistry studies in the past decade.

The work presented below, which can be considered as an expansion of the investigation displayed in Chapter 2, aims at the characterization of 5caC incorporation into canonical DNA strands at multiple pH and temperature conditions. By adopting identical experimental conditions as previously performed for 5fC, structural, kinetic and thermodynamic data could be accurately compared, allowing for a comprehensive picture of the similarities and differences between the two. In addition to the slower millisecond time scale previously discussed, we hereby performed a complementary analysis of the microsecond domain. The insights yielded by said analysis have been instrumental in the understanding of the de-/stabilization mechanisms operated by 5fC and 5caC.

In summary, the significance of this study lies in the successful dissection of the impact of protonated carboxylcytosine against its deprotonated counterpart. The ability to numerically compare previously obtained results for 5fC and canonical cytosine dsDNA oligomers, together with an exploratory examination of the μs time scale, yielded valuable insights into how the two epigenetic marks operate and influence dsDNA.

This work is integrative in nature, coupling synthetic organic chemistry and physical chemistry concepts. It was designed and carried out in collaboration with the Carell group within the SFB1309 framework. A copy of the open-access publication (available through the Creative Commons CC BY license) is re-printed below and accessible following this link.

Article

¹H NMR Chemical Exchange Techniques Reveal Local and Global Effects of Oxidized Cytosine Derivatives

Romeo C. A. Dubini, Eva Korytiaková, Thea Schinkel,[§] Pia Heinrichs,[§] Thomas Carell, and Petra Rovó*

Cite This: https://doi.org/10.1021/acsphyschemau.1c00050

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: 5-Carboxycytosine (5caC) is a rare epigenetic modification found in nucleic acids of all domains of life. Despite its sparse genomic abundance, 5caC is presumed to play essential regulatory roles in transcription, maintenance and base-excision processes in DNA. In this work, we utilize nuclear magnetic resonance (NMR) spectroscopy to address the effects of 5caC incorporation into canonical DNA strands at multiple pH and temperature conditions. Our results demonstrate that 5caC has a pH-dependent global destabilizing and a base-pair mobility enhancing local impact on dsDNA, albeit without any detectable influence on the ground-state B-DNA structure. Measurement of hybridization thermodynamics and kinetics of 5caC-bearing DNA duplexes highlighted how acidic environment (pH 5.8 and 4.7) destabilizes the double-stranded structure by ~10−20 kJ mol⁻¹ at 37 °C when compared to the same sample at neutral pH. Protonation of 5caC results in a lower activation energy for the dissociation process and a higher barrier for annealing. Studies on conformational exchange on the microsecond time scale regime revealed a sharply localized base-pair motion involving exclusively the modified site and its immediate surroundings. By direct comparison with canonical and 5-formylcytosine (5fC)-edited strands, we were able to address the impact of the two most oxidized naturally occurring cytosine derivatives in the genome. These insights on 5caC's subtle sensitivity to acidic pH contribute to the long-standing questions of its capacity as a substrate in base excision repair processes and its purpose as an independent, stable epigenetic mark.
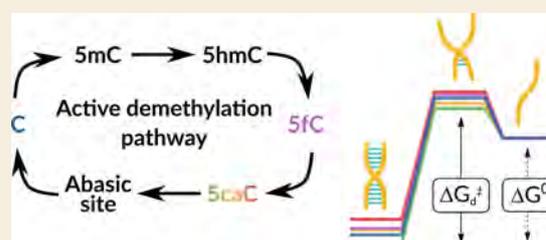
KEYWORDS: DNA, Epigenetic modifications, 5-formylcytosine, 5-carboxycytosine, NMR, Dynamics

INTRODUCTION

Apart from the four canonical bases, DNA may also contain modified versions of cytosine, thymine, or adenosine nucleosides.[1,2] Such naturally occurring DNA modifications are broadly named epigenetically modified bases, and they constitute an additional regulatory layer by extending genomic complexity and have been shown to play several crucial roles.[3] The discovery of ten-eleven translocation (TET)-induced oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) has been a key catalyst for the exploration and subsequent characterization of both 5-formylcytosine (5fC) and 5-carboxycytosine (5caC).[4]

5fC and 5caC retain a yet undefined number of functional roles, besides being intermediates within the active demethylation pathway,[5,6] as they represent genomically stable, semipermanent modifications with clearly defined tissue distributional patterns.[7−9] Both have been reported as abnormally abundant in prostate, breast and plasma cells cancer.[10−13] Their biological significance is not limited to the initial appearance and progression of diseases, since 5fC and 5caC are transiently accumulated during lineage specification of neural stem cells (NSCs) in culture and *in vivo*,[14] reduce the rate and substrate specificity of RNA polymerase II tran-

scription,[15] or can be selectively recognized by specialized proteins.[16]

Despite extensive research efforts in recent years, it is yet unclear how reader proteins recognize 5fC and 5caC with high specificity and selectivity regardless of their sparse genomic abundance and their chemical similarity to canonical, methylated or hydroxymethylated cytosines. 5fC has been more extensively studied compared to its carboxylated counterpart. Recent research endeavors clarified initial contradictory reports about its impact on structure and stability, leaning toward the notion that 5fC does not affect the B-DNA form.[17−19] As a result of the destabilizing effect that formylation imparts on the C−G base pair, 5fC was found to facilitate melting and hinders annealing, although without affecting the structure to any measurable extent, an aspect which has been investigated via FT-IR and NMR spectroscopy.[20,21]

pubs.acs.org/physchemau

Article

From a structural perspective, it has been recently suggested that 5caC is able to induce a kink in dsDNA, and such geometric alteration has been deemed essential for recognition and enzymatic action by Thymine DNA Glycosylase (TDG).[22] On the other hand, disagreement arises concerning its impact on dsDNA melting and annealing, with studies reporting inconsistent de/stabilization-related properties even when studied under nearly identical conditions.[23−25] While 5fC-induced weakening of the base pair was found to be independent of the mildly basic or acidic conditions naturally occurring within different tissues and/or organisms, with the introduction of the titrable carboxyl group, a pH-dependent behavior emerges for 5caC.[25,26] Protonation/deprotonation events of the carboxyl group induce 5caC to act as an electron-withdrawing group (EWG)/electron-donating group (EDG), respectively. These aspects have been reported to be biologically significant. Indeed, the carboxyl group protonation state has a subtle impact on hydrogen bonding, a behavior that was rationalized via the pK_a values of the two solvent-exposed sites: the nitrogen atom N3 and the carboxyl group itself.[21,26,27] Correspondingly, activity studies on DNA polymerases also concluded that 5caC acts as a base-pair mismatch during DNA replication, signaling that protonated 5caC is a highly destabilizing entity in the context of dsDNA.[28] In addition, structural biology investigations have suggested that the degree of protonation of 5caC's carboxyl group might be a key factor in the mechanism of excision operated by TDG.[26]

Even though other studies have considerably advanced our understanding of epigenetically modified DNA bases within the context of nucleic acids, a number of aspects remain unsettled. In this paper, we seek to shed light on the extent to which 5caC-edited dsDNA's structural features and kinetics-related phenomena deviate from its formylated and canonical equivalents. By employing state-of-the-art methodologies and analytical frameworks in solution-state NMR spectroscopy, we aimed at providing a noninvasive, label-free and site-specific description of the structure and dynamics of these samples using three distinct approaches. First, we compared the impact of the 5caC modification on dsDNA at pH 7.0, 5.8, and 4.7 with respect to 5fC and canonical C by measuring ^{1}H, ^{13}C, and ^{15}N chemical shift perturbations. Second, we applied temperature-dependent ^{1}H chemical exchange saturation transfer (CEST) experiments to assess the impact of 5caC on dsDNA melting and annealing processes providing site-specific parameters for dissociation ($\Delta G_{d,37°C}^{\ddagger}$, $\Delta H_d^{\ddagger}$, $\Delta S_d^{\ddagger}$) and association kinetics ($\Delta G_{a,37°C}^{\ddagger}$, $\Delta H_a^{\ddagger}$, $\Delta S_a^{\ddagger}$) and for melting thermodynamics ($\Delta G_{37°C}^{\circ}$, ΔH°, ΔS°).[20,29] Third, we explored the microsecond time scale exchange kinetics via ^{1}H on-resonance $R_{1\rho}$ relaxation dispersion (RD) targeting potential base-specific motions. Watson−Crick (WC) to Hoogsteen (HG) base-pair exchange processes have been characterized in the context of canonical DNA and protein−DNA complexes,[30−32] but no other exchange phenomena are known to occur in undamaged, non-mismatched DNA helices which are not interacting with a reader protein or enzyme.[33,34]

Collectively, our findings reveal that 5caC's pH-induced chameleonic behavior is not due to any permanent structural changes, opposite to previously reported results.[22] In fact, we observed that the repercussions of 5caC and 5fC incorporation into otherwise canonical dsDNA are uniquely perceptible in the context of conformational dynamics affecting at least two distinct time scales: a slower one (tens of milliseconds) and a faster one (hundreds of microseconds). In addition, we found that carboxycytosine does noticeably affect dsDNA annealing and melting phenomena when exposed to progressively lower pH conditions. Our unified analysis provides a highly comparable and comprehensive overview of the mechanistic details involving both global (such as energetics of DNA strands association and dissociation) and local (site specific motions involving a single base-pair) dynamics phenomena, and sets the stage for future studies of protein−DNA interactions.

RESULTS

We considered the pH-dependent structural features, thermodynamic stability, and site-specific dynamics of a self-complementary 12mer homocarboxylated DNA with the sequence of 5′-GCGATXGATCGC-3′ where X stands for 5caC (Figure 1). The corresponding samples were named
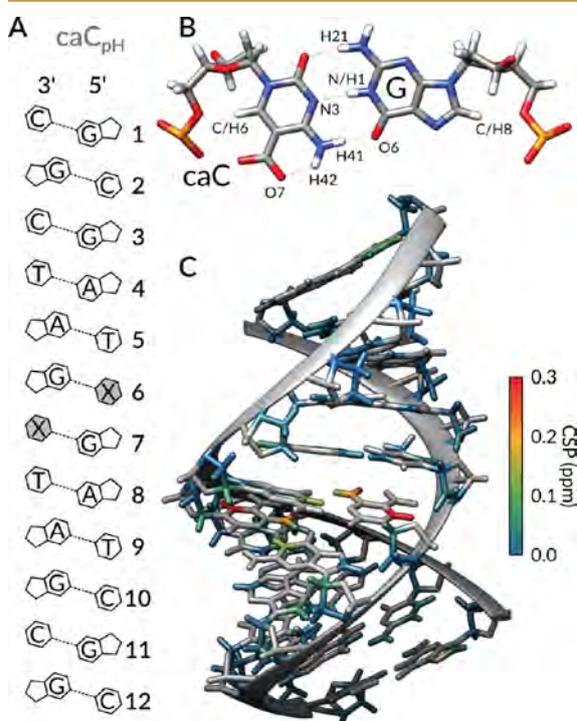
Figure 1. (A) DNA sequence of caC$_{pH}$'s model sequence. The modified cytosine nucleoside is highlighted in gray, marked as X. (B) Structural model showing the expected 5caC−G base pair conformation, consistent with ^{1}H, ^{13}C, and ^{15}N chemical shift values. (C) Absolute CSP values comparing samples at pH 7.0 and 4.7 are schematically displayed onto the B-DNA structural model of caC$_{pH}$.

caC$_{7.0}$, caC$_{5.8}$, and caC$_{4.7}$ indicating the pH at which they were studied. In order to evaluate the influence of cytosine carboxylation in a broader context, we compare the chemical shifts and the NMR-derived thermodynamics, kinetics, and dynamics parameters to analogous values obtained for canonical (C$_{7.0}$) and 5fC-modified (fC$_{7.0}$) samples featuring a sequence identical as what we considered here, which were previously reported in ref 20.

Structural Impact by Chemical Shift Analysis

To spot any unusual features induced by inclusion of carboxylated cytosine within otherwise canonical DNA, we recorded a comprehensive set of homo- (^{1}H−^{1}H NOESY) and heteronuclear (^{1}H−^{13}C HSQC, ^{1}H−^{15}N SOFAST HMQC) 2D spectra for resonance assignment and chemical shift analysis purposes.[35]

Previous crystallographic studies have reported contrasting results concerning a potential geometric alteration of the dsDNA helix induced by 5caC.[22,23] Our results, in support of ref 23, indicate that the 5caC nucleobase does not induce any detectable, permanent deviation from the canonical B-DNA structure. At pH 7.0, when compared to the 5fC−G base-pairing interactions, 5caC−G appears to be equally well-tolerated within a canonical double-helix architecture: all of ^{1}H−^{13}C and ^{1}H−^{15}N cross-peaks are superimposable among the canonical, and 5fC and 5caC-containing 12mer DNA constructs, with the sole exception represented by those nuclei in direct proximity to the epigenetic modification (Figures 2, S1, and S2).

An analogous comparison of caC$_{7.0}$ with caC$_{5.8}$ and caC$_{4.7}$ points to no detectable structural change upon acidification. ^{1}H, ^{13}C, and ^{15}N resonances, $^3J_{HH}$ couplings and ^{1}H−^{1}H NOESY cross-peak patterns suggest that carboxylated cytosine leaves the B-DNA structure entirely unperturbed at both neutral and acidic pH values, resulting in no substantial structural difference from the canonical or formylated cytosine-bearing constructs in any studied condition (Figures S3−S6).

Figure 2 compares the perturbations of representative ^{1}H, ^{13}C, and ^{15}N chemical shifts of fC$_{7.0}$, caC$_{7.0}$, caC$_{5.8}$, and caC$_{4.7}$ with respect to the shifts of the canonical 12mer sequence. Figure 2A shows imino protons, which are sensitive reporters of the strength of base-pairing: downfield (upfield) shifted signals, appearing at higher (lower) ppm values, revealing stronger (weaker) intermolecular hydrogen bonds. Interestingly, the only responsive site to pH changes is G7H1, which reports on the X:G base pair, where X is either C, 5fC, or 5caC at the three distinct pH conditions we selected. Substantial chemical shift perturbations are also visible for aromatic ^{1}H sites T5H6 and X6H6, ^{13}C X6C6, and ^{15}N G7N1 (Figures 2B−D and S7−S10). Among those nuclei, the trends suggest that caC$_{4.7}$'s resonances resemble fC$_{7.0}$ the most, while in neutral conditions caC$_{7.0}$ is comparable with canonical C$_{7.0}$.

These pH-induced shifts were most prominent for H-bond forming protons, namely, for caC6H41, H42, and G7H1. The nature of the effect is related to the reshuffling of electron densities around the base-pairing atoms due to altered Pauli repulsion between occupied atomic orbitals.[36] An increased ^{1}H chemical shift is associated with a decreased electron density (^{1}H shielding) around the proton and hence with a stronger H-bond, or shorter H−X distance. As the carboxyl group of 5caC becomes progressively protonated with decreasing pH, the intrabase H-bond between caC6H42 and the carboxyl oxygen O7 gets weaker (caC6H42 $\Delta\delta$ = −0.3 ppm), while the interbase H-bond formed by caC6H41 and G7O6 (Figure 1B) gets stronger (caC6H41 $\Delta\delta$ = +0.1 ppm). Meanwhile, hydrogen bonding between G7H1 and caC6N3 weakens (G7H1 $\Delta\delta$ = −0.15 ppm) due to the decreased electron density at N3 caused by the EWG properties of the protonated carboxyl group, compensating for the increased base-pairing stability gained by the H41−O6 interaction (Figures S7−S10). The fine balance between the strengths of the intra- and
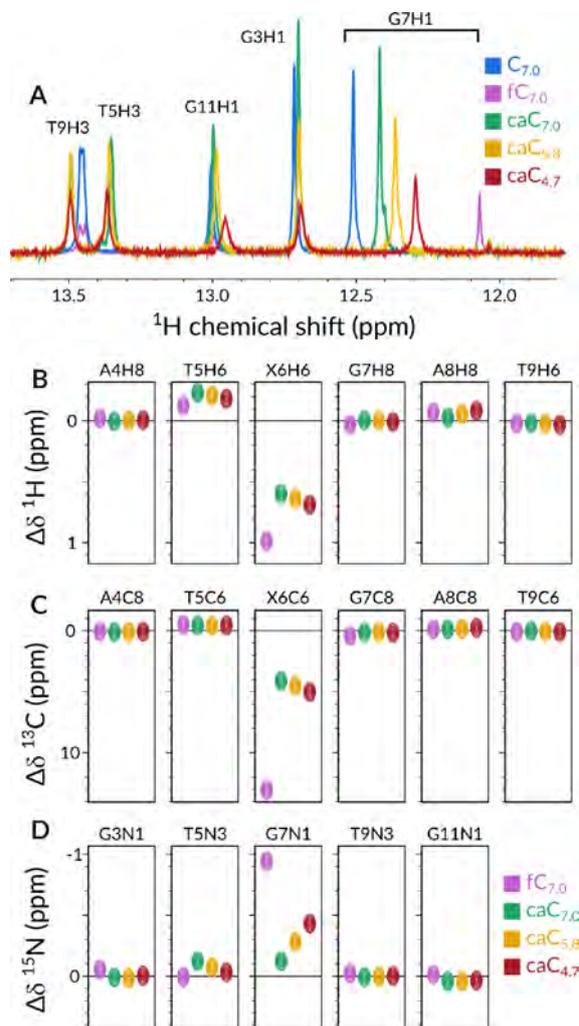
Figure 2. Comparison of chemical shift perturbations $\Delta\delta$ for representative imino ^{1}H (A), aromatic ^{1}H (B), aromatic ^{13}C (C), and imino ^{15}N (D) resonances with respect to the chemical shifts of the canonical 12mer sample (blue spectrum in A, black horizontal line in B−D). Magenta, green, yellow, and red symbols represent fC$_{7.0}$, caC$_{7.0}$, caC$_{5.8}$, and caC$_{4.7}$, respectively. A full comparison of all comparable chemical shifts between the five samples is displayed in Figures S7−S10.

interbases hydrogen bonds in the 5caC−G base pair leads to close to optimal base-pairing both at neutral and mildly acidic conditions, leaving the B-DNA structure unperturbed through the entire studied pH range.

5caC's Influence on DNA Melting and Annealing

A thorough characterization of nucleic acids' global structural rearrangements, such as folding, melting, annealing, and binding should entail a comprehensive analysis of kinetic events occurring on the millisecond to second time scale.[37] CEST experiments have found adoption in modern biomolecular NMR, allowing for quantitative and site-specific determination of population, chemical shift, and exchange kinetics of sparsely populated conformations.[29,38] When measured in a temperature-dependent fashion, the shift in
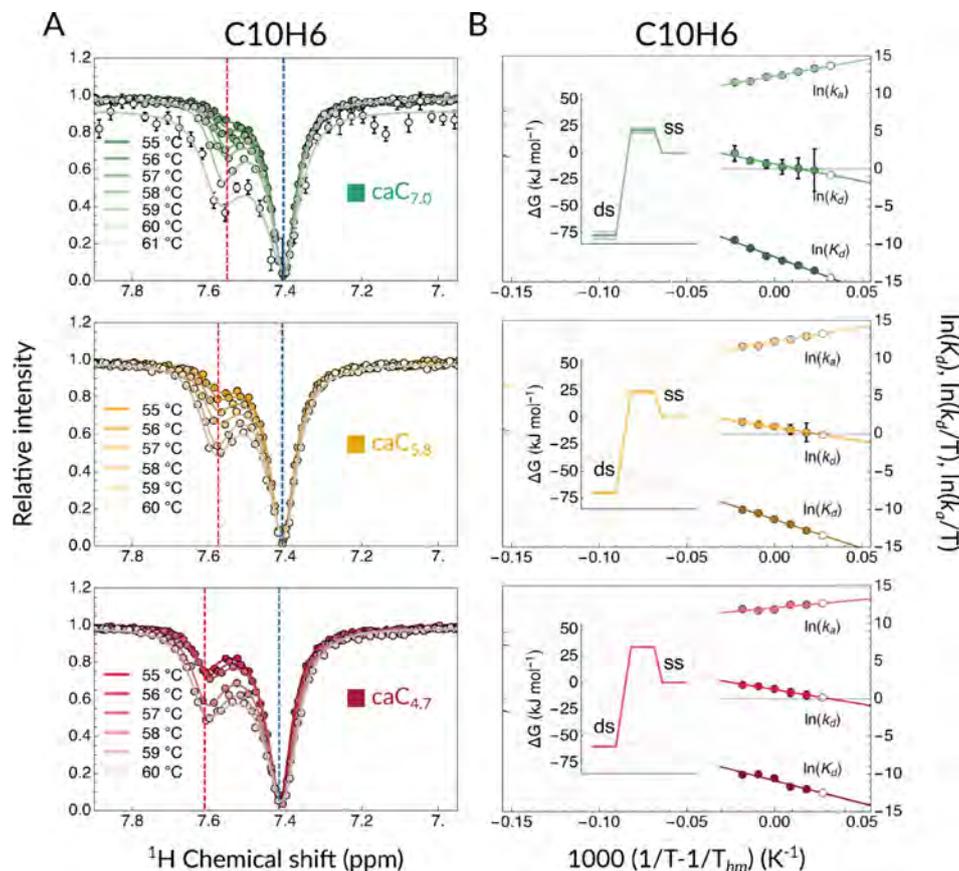
Figure 3. (A) Temperature-dependent CEST melting profiles together with the obtained two-site exchange fits for proton C10H6. $caC_{7.0}$, $caC_{5.8}$, and $caC_{4.7}$ are shown in shades of green, yellow, and red, respectively. Dashed blue and red lines indicate dsDNA and ssDNA chemical shift values. (B) van't Hoff plots relative to the CEST melting profiles. Shades of green, yellow, and red indicate data entries and linear fits of $\ln k_a$, $\ln k_d$, and $\ln K_d$ vs $1/T - 1/T_{hm}$, where T_{hm} is the harmonic mean of the measured temperatures, for $caC_{7.0}$, $caC_{5.8}$, and $caC_{4.7}$, respectively. White data points represent back-calculated values for 55 °C that were used to validate the CEST fits. Insets present the relevant Gibbs free energy plots at each pH condition at 37 °C.

exchange parameters can reveal atomistic details about the melting thermodynamics and kinetics of the studied system providing unprecedented insights into the molecular processes. In pursuance of the study of 5caC-induced DNA destabilization, we proceeded by recording CEST profiles for the aromatic protons in all caC_{pH} samples in the 55−61 °C range. As an example, Figure 3A displays the melting CEST profiles for C10H6 at three pH values (profiles for other comparable protons can be found in Figures S11−S18). The appearance of a distinctive secondary dip at increasingly higher temperatures indicates the presence of an alternative conformer, which we identify as the single-stranded conformation (ssDNA) as per comparative chemical shift analysis.

To obtain a more quantitative comparison, we fit the CEST profiles to a two-state exchange model (dsDNA ⇌ 2 ssDNA) that yielded a numerical estimation of 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 highest temperatures, while we used the lowest temperature of the ensemble for each sample (55 °C) to compare the predicted back-calculated value from the fits to the experimental data points. ^{1}H longitudinal (R_1) and transverse (R_2) relaxation

rates were measured separately at multiple temperatures and used as inputs for the CEST fits assuming that the rate constants of the dsDNA and ssDNA states are the same.

As an example, Figure 3B shows the logarithm of the obtained kinetic rates and equilibrium constants for C10H6. Consistently with results obtained for $fC_{7.0}$ and $C_{7.0}$, two observations can be made: (i) a linear fit could be identified for kinetic rates and equilibrium constants against T^{-1}, suggesting a single transition state is plausible for the melting and annealing phenomena, and (ii) all sites show thermally activated kinetics with a positive dissociation barrier (Arrhenius behavior) and a negative association barrier (anti-Arrhenius).

In addition, CEST measurements run at multiple distinct temperatures allow for the extraction of Gibbs free energy and related parameters by leveraging the temperature-dependent nature of kinetics and thermodynamic phenomena. In order to achieve an estimation of the degree of protonation-induced destabilization to dsDNA, we applied our recently described methodological framework for the decomposition of traditional CEST output into enthalpic and entropic stability and activation changes.[20] Succinctly, we assumed the observed

dynamic equilibrium is the reversible melting/annealing process of a dsDNA strand into two single-stranded DNA sequences. If the concentration of DNA is known, then kinetics of association (k_a) and dissociation (k_d) can be extracted. From the temperature dependence of k_a and k_d, activation barriers for both process ($\Delta G_{d,37°C}^{\ddagger}$ and $\Delta G_{a,37°C}^{\ddagger}$) can be derived. Analogously, the ratio between $k_a(T)$ and $k_d(T)$ rates allows for the determination of the equilibrium dissociation constant (K_d) and consequently the quantification of thermodynamic parameters such as $\Delta G_{37°C}^{\circ}$.

This analysis allows us to directly compare caC$_{7.0}$, caC$_{5.8}$, and caC$_{4.7}$ not only between them, but also with otherwise identical canonical and formylated samples (C$_{7.0}$ and fC$_{7.0}$, respectively), which we previously discussed in ref 20. In Table 1, we compare three proton reporters across all five samples

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[a]

sample		$\Delta G_{37°C}^{\circ}$ (kJ mol^{-1})	$\Delta G_{d,37°C}^{\ddagger}$ (kJ mol^{-1})	$\Delta G_{a,37°C}^{\ddagger}$ (kJ mol-1)
C$_{7.0}$	C2H6	71.1 ± 1.7	99.2 ± 2.2	28.1 ± 2.1
	T9H6	70.2 ± 1.8	102.2 ± 1.5	31.9 ± 1.5
	C10H6	71.3 ± 1.4	100.3 ± 1.3	29.0 ± 1.3
caC$_{7.0}$	C2H6	73.3 ± 7.4	93.0 ± 4.3	19.7 ± 4.1
	T9H6	75.7 ± 0.8	99.2 ± 0.9	23.4 ± 0.9
	C10H6	77.3 ± 3.6	98.6 ± 3.2	21.3 ± 2.9
caC$_{5.8}$	C2H6	71.1 ± 1.2	94.8 ± 1.4	23.7 ± 1.3
	T9H6	68.0 ± 0.9	96.6 ± 0.9	28.6 ± 1.0
	C10H6	67.4 ± 1.4	92.4 ± 1.2	24.9 ± 1.1
caC$_{4.7}$	C2H6	58.6 ± 0.7	93.1 ± 0.7	34.5 ± 0.9
	T9H6	62.1 ± 0.4	95.2 ± 0.4	33.1 ± 0.5
	C10H6	56.8 ± 0.6	92.9 ± 0.6	36.1 ± 0.6
fC$_{7.0}$	C2H6	74.7 ± 10.4	99.7 ± 5.6	25. ± 5.5
	T9H6	64.6 ± 1.2	97.1 ± 1.0	32.5 ± 1.0
	C10H6	66.4 ± 1.2	97.5 ± 1.0	31.1 ± 1.1

[a]Errors are given as one standard deviation. An extended version of Table 1 can be found in the Supporting Information file (Table S7).

(comprehensive tables including fitting results for all proton reporters can be found in Tables S1−S6, while an extended version of Table 1 is available in Table S7). From a thermodynamic perspective, among the C-modified samples, caC$_{7.0}$ scores as the most stable one, which is highly comparable to C$_{7.0}$, as observed by UV/vis spectroscopy (Figure S19) as well as in other studies.[23,27] Acidification of the buffer to pH 5.8 destabilizes the double-stranded conformer by ~2−10 kJ mol^{-1} and further by another ~6−13 kJ mol^{-1} when the pH is decreased from 5.8 to 4.7. Interestingly, $\Delta G_{37°C}^{\circ}$ for caC$_{5.8}$'s proton reporters are very similar to those we obtained for fC$_{7.0}$. Kinetics of dissociation data, as expected, support the notion that caC$_{7.0}$ and C$_{7.0}$ require the most energy for undergoing a dsDNA → 2 ssDNA conformational transition. According to this metric, fC$_{7.0}$ is slower in undergoing the melting process when compared to caC$_{5.8}$ and caC$_{4.7}$, as $\Delta G_{d,37°C}^{\ddagger}$ is higher by ~3−7 kJ mol^{-1}. Lastly, kinetics of association suggest that caC$_{4.7}$ is the slowest in performing an annealing process, followed by fC$_{7.0}$. Compellingly, $\Delta G_{a,37°C}^{\ddagger}$ data indicate that the rate of association of caC$_{7.0}$ is ~8 kJ mol^{-1} less energetically demanding when compared to C$_{7.0}$, a

behavior that can be rationalized considering the EDG nature of the carboxylate substituent (5caC at neutral pH) when compared to a proton (canonical C).

In Figure 4, we show the correlation between the dissociation and equilibrium free energy changes across all
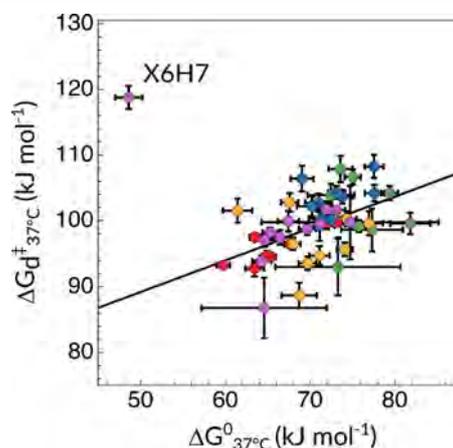
Figure 4. Correlation plot between the changes of the Gibbs free energy of activation for the dissociation process and the equilibrium free energy of previously reported values for fC$_{7.0}$ and C$_{7.0}$ (magenta and blue, respectively) together with 5caC-containing samples (caC$_{7.0}$ as green, caC$_{5.8}$ as yellow, and caC$_{4.7}$ as red) at 37 °C. fC$_{7.0}$ X6H7 (featuring much higher activation free energies and lower equilibrium energies than the rest of the molecule) is an outlier due to its stable intramolecular hydrogen-bond between the formyl O7 atom and the adjacent amino H42 proton.

samples and conditions, for every proton reporter. Here, $\Delta G_d^{\ddagger}$ is plotted as a function of ΔG° for all five samples across all eligible proton reporters. Data points accounting for caC$_{7.0}$ and C$_{7.0}$ tend to cluster at the upper right-hand corner of the plot. Conversely, $\Delta G_{37°C}^{\circ}$ and $\Delta G_{d,37°C}^{\ddagger}$ values are substantially decreased whenever fC$_{7.0}$, caC$_{5.8}$, or caC$_{4.7}$ is considered, as elaborated above.

Protonation-Induced Microsecond Dynamics

^{1}H and ^{15}N chemical shift values for G7H1/N1 nuclei, both reporters of the centrally positioned 5caC6-G7 base pair's stability, have evidenced that the protonation state of the exocyclic carboxylic group 5caC has a selective impact on this imino proton resonance (Figure 2A, D), which have long been regarded as key indicators of hydrogen bond strength.[23,39] Because of this observation, we aimed at investigating whether this protonation-induced, localized weakening is accompanied by increased probability of local fast time-scale motions.

Our CEST-based kinetic and thermodynamic analysis has established that caC$_{5.8}$, caC$_{4.7}$, and fC$_{7.0}$ destabilize the double-stranded DNA structure without any apparent static, persistent impact on its helical architecture. In order to investigate the presence of a potentially localized conformational exchange which might contribute to the destabilization, we interrogated the faster, microsecond time scale by applying ^{1}H $R_{1\rho}$ relaxation dispersion (RD) methods.[37,40]

In Figure 5, we show X6H6 (where X = C, 5fC or 5caC, depending on the sample under current consideration) ^{1}H on-resonance $R_{1\rho}$ RD profiles measured at 55 °C, a condition that ensures that the melting process is still rather sparse and
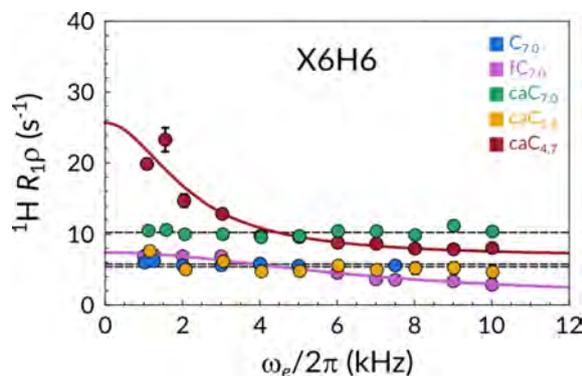
Figure 5. $R_{1\rho}$ relaxation dispersion profiles of $C_{7.0}$ (blue), $fC_{7.0}$ (magenta), $caC_{7.0}$ (green), $caC_{5.8}$ (yellow), and $caC_{4.7}$ (red) recorded at 55 °C. Lines represent best fits to models either accounting for (color-coded) or discounting chemical exchange (black).

infrequent. The data sets recorded for $caC_{7.0}$, $caC_{5.8}$, and $C_{7.0}$ are best fit to a no-exchange model, resulting in flat profiles (black lines). Conversely, profiles for $caC_{4.7}$ and $fC_{7.0}$ fit best to the exchange model, exposing a chemical exchange contribution to its R_2 relaxation rate. Such motions, consistent with a τ_{ex} in the order of hundreds of microseconds, fall within the intermediate exchange regime and are approximately 2 orders of magnitude faster compared to the overall melting process as characterized by our CEST measurements. This result is especially interesting when comparing the H6 proton of the modified base in samples $caC_{4.7}$ and $fC_{7.0}$ to any other available ^{1}H nucleus. No other profile is consistent with an exchange phenomenon on this interval (Figures S20−S24), suggesting that the detected phenomenon is unrelated to the previously reported Watson−Crick to Hoogsteen base-pair exchange process.[30−32] In other words, this motions appears to be sharply localized, leaving all other bases unaffected.

This result can be rationalized by considering our previous speculations on the mechanism by which 5fC (and by extension protonated 5caC) weakens the dsDNA conformer.[20] Deprotonated (or not protonated to a sufficient extent) 5caC proved itself to be either a neutral or even a stabilizing factor in terms of base-pair strength and agrees well both with NOESY-derived chemical shift values and the literature. Instead, whenever the pH of the buffer is sufficiently acidified, the weakening of the 5caC-G hydrogen bond induces a chemical exchange process in the intermediate microsecond time scale. The fact that this motion is sharply localized at the modified site seems to suggest that the protonation of the exocyclic carboxyl moiety and the consequent lowered basicity of the hydrogen-bonded N3 atom could be deemed responsible for generating a sparse and transitory fraying event in the middle of the DNA strand, presumably synergistic with the global destabilization effect on the whole structure.

■ **DISCUSSION**

In pursuance of understanding the role of 5caC, we addressed the structural and dynamic features of oligomeric DNA double strands carrying a single version of such modification on each strand. Incorporation of carboxycytosine into DNA is a naturally occurring phenomenon that has been mainly discussed within two biological frameworks: (i) its recently devised (and to date mostly obscure) semipermanent

epigenetic role and (ii) as a DNA lesion that undergoes the base excision repair (BER) process.[2−4,41]

Considerations on 5caC Protonation Sites

In their report on divergent mechanisms of enzymatic excision for 5caC and 5fC, Maiti et al. have elaborated how, in a TDG-DNA complex, nitrogen N3 of 5caC (Figure 1B) is likely more basic than the carboxyl group and thus undergoes protonation before the carboxylic exocyclic moiety does.[26] In contrast, despite assigning a pK_{a,COO^-} of 4.7 and $pK_{a,N3}$ of 2.1 for the isolated nucleoside, a distinct infrared spectroscopy and quantum mechanical analysis suggests that, when in the context of a dsDNA strand, the first protonation site is the carboxyl group.[25] Our results support the latter idea. In Figure 2A, we show that no other base pair is affected by the pH change over the entire 7.0−4.7 interval. Indeed, T9H3, T5H3, and G3H1 display negligible chemical shift differences, while G11H1 shifts upfield by ∼0.05 ppm likely due to its proximity to the fraying ends of the oligomeric model system. In sharp contrast, the signal reporting on G7H1 consistently shifts upfield with decreasing pH. While this preference for protonation of the weaker base (i.e., 5caC's carboxylate group over N3) is apparently counterintuitive, we reason that in a dsDNA setting the N3 site is well protected from the solvent environment, both for steric and electrostatic reasons.

5fC and 5caC as Semipermanent Modifications

For a long time, 5fC and 5caC have been mainly considered as transient intermediates within the active demethylation pathway.[42] However, research toward their capacity as standalone epigenetic marks has gained increasingly more traction. For instance, both 5fC and 5caC overlap with H3K4me1 marked regions, associated with active transcription.[43] Also, several developmental and metabolic related genes show 5fC enrichment on promoters before gene upregulation,[44] while 5caC has been reported to transiently accumulate at promoter regions preceding gene expression during lineage specification and differentiation.[45] Lastly, a number of cancerous diseases are correlated with a significant enrichment of such oxidized cytosine epigenetic modifications.[10−13]

From a structural and conformational perspective, both 5fC and 5caC have been reported to induce structural changes localized in the proximity of the modified nucleoside,[22,26] while the pH dependence of 5caC characteristics on dsDNA stability has been previously discussed in the context of short oligomers carrying several clustered modifications.[21,27] In the context of longer (90 bp) DNA strands, Ngo et al. reported a 3-fold enhancement of cyclization rate for 5fC containing strands, while the sample carrying 5caC at pH 8.0 showed no detectable difference when compared to canonical cytosine.[46]

In contrast to refs 22 and 26, our results indicate that, identically to 5fC, 5caC does not induce any permanent structural modification detectable by NMR spectroscopy at any pH condition under consideration (Figures 2 and S2−10). In fact, in analogy to 5fC, 5caC was found to affect dsDNA melting and annealing equilibrium and kinetics, rather than B-DNA average structure. On average, and across all sites, $caC_{5.8}$ and $caC_{4.7}$ resemble $fC_{7.0}$, both thermodynamically and kinetically, when it comes to annealing and melting. In contrast, and consistently with previous cyclization essays and FRET studies, the behavior of deprotonated 5caC is most similar to canonical cytosine.

5mC oxidation derivatives have been shown to accumulate and persist in a relatively stable state in certain biological contexts. For this reason, they have been suggested to carry out additional roles apart from being intermediates in biochemical pathways. In light of the foregoing, we contend that our results could correlate with the relative abundance of 5caC and 5fC in cancerous tissues. How protonation of the exocyclic carboxyl moiety affects the thermodynamics and kinetics of melting and annealing *in vitro* has been discussed in this and previous reports.[20,21,25,27] Our results corroborate the protonation-induced destabilization of the double strand, and we speculate that a similar effect might take place in cancerous tissues, triggered by the low pH environment. Cell proliferation is notoriously accelerated in cancer, and the acidic microenvironment where such diseases thrive is widely recognized as a phenotypic trait, making the presence of the two most oxidized cytosine epigenetic derivatives more than circumstantial.[47,48]

5fC- and 5caC-Driven Enzymatic Recognition

In mammals, active DNA demethylation takes place via an enzymatic tandem of TET and TDG, which together govern the initial stages of the BER pathway. In order to successfully complete the removal of 5mC, the presence of either 5fC or 5caC is ultimately necessary, as they are substrates for TDG, which generates the abasic site.[26] Apart from TET and TDG, several additional proteins are able to selectively recognize, bind, and exert their respective enzymatic activity upon 5fC and 5caC.[12,13,49] The mechanism by which different reader enzymes selectively recognize cytosine's epigenetic modifications has long been established as a crucial theme in chemical biology.[6] Across the proposed enzymatic mechanisms, many rely on a specific residue to initiate the base-extrusion process into the active site. For instance, the "pinch−push−pull" mechanism proposed for TDG leans on Arg275 to promote the breakage of an X:G base pair, where X = T, 5fC, or protonated 5caC.[50,51] Alternative studies suggest that partially extruded nucleotide conformations, which are sparse but naturally occurring events, might play a role in recognition and base excision.[52] This second mechanism of action seems to agree with DNA replication studies, which highlighted how 5caC:G base-pair can behave as a DNA lesion.[28]

Although the data hereby presented is not conclusive, we believe our on-resonance $R_{1\rho}$ RD data assist in providing one more piece of evidence in this complex puzzle. By observing site selective and spontaneous (i.e., not triggered by the enzyme) base-flipping of protonated 5caC:G, especially in the context of a weakened dsDNA strand as evidenced in our CEST analysis, we present further experimental evidence that 5fC and protonated 5caC could indeed act as DNA lesions.[33] We hypothesize that, albeit undetected in our study, such kinetic processes could be present at physiological temperatures and decisively impact enzyme recognition and mode of action.

CONCLUSIONS

In this work, we have considered the impact of 5caC incorporation into a model dsDNA oligomer (caC$_{pH}$) in three pH conditions, namely, at pH 7.0, 5.8, and 4.7. We obtained chemical shift, melting/annealing, and microsecond conformational exchange data which we could reliably compare with our recent study focusing on 5fC and canonical cytosine. Assignment and chemical shift studies on comparable ^{1}H, ^{13}C, and ^{15}N nuclei have shown that there is no evidence

of a permanent structural change: all caC$_{pH}$ samples, together with C$_{7.0}$ and fC$_{7.0}$, are compatible with a standard B-DNA helical arrangement. CEST-derived kinetic and thermodynamic data suggested that the reduced cohesion of the X6:G7 base pair, evidenced by chemical shift studies, affects the extent to which nearby bases are able to cooperatively stabilize one another. caC$_{4.7}$ and fC$_{7.0}$ emerged as the most destabilized samples of the cohort, while caC$_{7.0}$ and C$_{7.0}$ showed remarkably similar properties overall.

caC$_{4.7}$ and fC$_{7.0}$ also revealed a detectable chemical exchange process at or in the proximity of the modified nucleoside X6 on the microsecond time scale. The data hereby presented indicate that 5caC's impact on B-DNA is only evident through the lenses of conformational dynamics, as protonation of the exocyclic carboxyl moiety affects the melting-annealing equilibrium as well as induces sparse and localized micro-second time scale base-pair dynamics. We believe our findings are relevant in the context of several open questions concerning this sparse epigenetic mark. Future investigations may consider expanding our initial exploration of the microsecond time scale by recording off-resonance $R_{1\rho}$ RD experiments or studying protein−DNA interactions featuring isotopic labeled 5fC or 5caC nucleosides to unravel the exact mechanistic details of the interaction between TET, TDG (and other enzymes), and cytosine's oxidized derivatives.

◼ MATERIALS AND METHODS

Sample Preparation

The cadC-phosphoramidite (cadC-PA) and subsequently the modified dsDNA samples caC$_{pH}$ were prepared via phosphoramidite chemistry as previously reported.[53] Solid phase synthesis of oligonucleotides containing cadC 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 cadC-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.[53] 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. Samples caC$_{7.0}$ and caC$_{5.8}$ were dissolved in aqueous buffers consisting of 15 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$ (pH 7.0 and 5.8, respectively), 25 mM NaCl in H$_2$O. Sample caC$_{4.7}$ was prepared by titrating a 1 M HCl solution into the same buffer described above. The thermal stability of the buffer between room temperature and 60 °C was ascertained by pH-meter measurements. Annealing was performed by heating the dsDNA-containing buffer solution to 90 °C for 5 min and slowly cooling it to 5 °C in approximately 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$_2$O, resulting in final sample concentrations of ∼0.66 mM for all samples, as determined via UV spectrophotometric measure-ments 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. $caC_{7.0}$ was measured at four concentrations (1.25, 2.50, 5.00, and 10.00 μM), while $fC_{7.0}$ and $C_{7.0}$ were measured as described in ref 20. 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 TCI probe. Standard 2D NOESY (mixing time 250 ms) spectra were recorded at 37 °C for resonance assignment. Natural abundance [1]H−^{13}C and [1]H−^{15}N HSQC and HMQC spectra were recorded using standard fast-pulsing pulse sequences.[54] Site-selective spin relaxation measurements were performed following the SELOPE scheme; these included ^{1}H CEST, on-resonance [1]H $R_{1\rho}$, recorded either with a single spin-lock strength of 10 kHz or as an entire RD profile ranging from 1 to 10 kHz, and [1]H R_1 experiments at temperatures between 37 and 60 °C. The employed on-resonance $R_{1\rho}$ and ^{1}H CEST pulse sequences have been modified from Schlagnitweit et al.[40] CEST profiles and pseudo-2D [1]H R_1 and pseudo-3D on-resonance $R_{1\rho}$ experiments were performed, processed, and analyzed as previously described.[20]

■ ASSOCIATED CONTENT

ⓈⒾ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsphyschemau.1c00050.

> [1]H−[1]H, [1]H−[13]C, and [1]H−[15]N 2D correlation spectra, schematic comparison of ^{1}H, ^{13}C, and ^{15}N chemical shift perturbations, concentration-dependent UV/vis melting analysis, temperature-dependent CEST profiles, and van't Hoff plots for all available sites, on-resonance $R_{1\rho}$ relaxation dispersion profiles for all samples, CEST-derived dsDNA and ssDNA chemical shifts and relaxation parameters (R_1 and R_2), and CEST-derived thermodynamic and kinetic parameters for all samples (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Petra Rovó − *Faculty of Chemistry and Pharmacy, Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany; Center for Nanoscience (CeNS), Faculty of Physics, Ludwig-Maximilians-Universität München, 80799 Munich, Germany; Institute of Science and Technology Austria (ISTA), 3400 Klosterneuburg, Austria;* ⓞ orcid.org/0000-0001-8729-7326; Email: petra.rovo@ist.ac.at

Authors

Romeo C. A. Dubini − *Faculty of Chemistry and Pharmacy, Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany; Center for Nanoscience (CeNS), Faculty of Physics, Ludwig-Maximilians-Universität München, 80799 Munich, Germany;* ⓞ orcid.org/0000-0001-6045-271X

Eva Korytiaková − *Faculty of Chemistry and Pharmacy, Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany*

Thea Schinkel − *Faculty of Chemistry and Pharmacy, Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany;* ⓞ orcid.org/0000-0002-1891-2660

Pia Heinrichs − *Faculty of Chemistry and Pharmacy, Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany;* ⓞ orcid.org/0000-0002-3128-6676

Thomas Carell − *Faculty of Chemistry and Pharmacy, Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsphyschemau.1c00050

Author Contributions

§T.S. and P.H. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Kumar, S.; Chinnusamy, V.; Mohapatra, T. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front. Genet.* **2018**, *9*, 1−14.

(2) Bilyard, M. K.; Becker, S.; Balasubramanian, S. Natural, modified DNA bases. *Curr. Opin. Chem. Biol.* **2020**, *57*, 1−7.

(3) Chen, Y.; Hong, T.; Wang, S.; Mo, J.; Tian, T.; Zhou, X. Epigenetic modification of nucleic acids: from basic studies to medical applications. *Chem. Soc. Rev.* **2017**, *46*, 2844−2872.

(4) Carell, T.; Kurz, M. Q.; Müller, M.; Rossa, M.; Spada, F. Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA Angewandte. *Angew. Chem., Int. Ed.* **2018**, *57*, 4296−4312.

(5) 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*, 1300−1304.

(6) Wu, X.; Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **2017**, *18*, 517−534.

(7) Bachman, M.; Uribe-Lewis, S.; Yang, X.; Burgess, H. E.; Iurlaro, M.; Reik, W.; Murrell, A.; Balasubramanian, S. 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **2015**, *11*, 555−557.

(8) Su, M.; Kirchner, A.; Stazzoni, S.; Müller, M.; Wagner, M.; Schröder, A.; Carell, T. 5-Formylcytosine Could Be a Semipermanent Base in Specific Genome Sites. *Angew. Chem., Int. Ed.* **2016**, *55*, 11797−11800.

(9) Koivunen, P.; Laukka, T. The TET enzymes. *Cell. Mol. Life Sci.* **2018**, *75*, 1339−1348.

(10) Guo, M.; Li, X.; Zhang, L.; Liu, D.; Du, W.; Yin, D.; Lyu, N.; Zhao, G.; Guo, C.; Tang, D. Accurate quantification of 5-

Methylcytosine, 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine in genomic DNA from breast cancer by chemical derivatization coupled with ultra performance liquid chromatography-electrospray quadrupole time. *Oncotarget* **2017**, *8*, 91248−91257.

(11) Eleftheriou, M.; Pascual, A. J.; Wheldon, L. M.; Perry, C.; Abakir, A.; Arora, A.; Johnson, A. D.; Auer, D. T.; Ellis, I. O.; Madhusudan, S.; et al. 5-Carboxylcytosine levels are elevated in human breast cancers and gliomas. *Clin. Epigenet.* **2015**, *7*, 88.

(12) Hashimoto, H.; Olanrewaju, Y. O.; Zheng, Y.; Wilson, G. G.; Zhang, X.; Cheng, X. Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes Dev.* **2014**, *28*, 2304−2313.

(13) Wang, D.; Hashimoto, H.; Zhang, X.; Barwick, B. G.; Lonial, S.; Boise, L. H.; Vertino, P. M.; Cheng, X. MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Res.* **2017**, *45*, 2396−2407.

(14) Wheldon, L. M.; Abakir, A.; Ferjentsik, Z.; Dudnakova, T.; Strohbuecker, S.; Christie, D.; Dai, N.; Guan, S.; Foster, J. M.; Corrêa, I. R.; et al. Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. *Cell Rep* **2014**, *7*, 1353−1361.

(15) Kellinger, M. W.; Song, C. X.; Chong, J.; Lu, X. Y.; He, C.; Wang, D. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **2012**, *19*, 831−833.

(16) Song, J.; Pfeifer, G. P. Are there specific readers of oxidized 5-methylcytosine bases? *BioEssays* **2016**, *38*, 1038−1047.

(17) Raiber, E.-A.; Murat, P.; Chirgadze, D. Y.; Beraldi, D.; Luisi, B. F.; Balasubramanian, S. 5-formylcytosine alters the structure of the DNA double helix. *Nat. Struct Mol. Biol.* **2015**, *22*, 44−49.

(18) Wang, S.; Long, Y.; Wang, J.; Ge, Y.; Guo, P.; Liu, Y.; Tian, T.; Zhou, X. Systematic Investigations of Different Cytosine Modifications on CpG Dinucleotide Sequences: The Effects on the B-Z. Transition. *J. Am. Chem. Soc.* **2014**, *136*, 56.

(19) Hardwick, J. S.; Ptchelkine, D.; El-Sagheer, A. H.; Tear, I.; Singleton, D.; Phillips, S. E.; Lane, A. N.; Brown, T. 5-Formylcytosine does not change the global structure of DNA. *Nat. Struct. Mol. Biol.* **2017**, *24*, 544−552.

(20) Dubini, R. C.; Schön, A.; Müller, M.; Carell, T.; Rovó, P. Impact of 5-formylcytosine on the melting kinetics of DNA by 1H NMR chemical exchange. *Nucleic Acids Res.* **2020**, *48*, 8796−8807.

(21) 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. *J. Phys. Chem. B* **2020**, *124*, 1160−1174.

(22) Fu, T.; Liu, L.; Yang, Q.-l.; Wang, Y.; Xu, P.; Zhang, L.; Liu, S.; Dai, Q.; Ji, Q.; Xu, G.-l.; et al. Thymine DNA glycosylase recognizes the geometry alteration of minor grooves induced by 5-formylcytosine and 5-carboxylcytosine. *Chem. Sci.* **2019**, *10*, 7407−7417.

(23) 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.

(24) 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.

(25) 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.* **2016**, *11*, 470−477.

(26) 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.

(27) Ashwood, B.; Sanstead, P. J.; Dai, Q.; He, C.; Tokmakoff, A. 5-Carboxylcytosine and Cytosine Protonation Distinctly Alter the Stability and Dehybridization Dynamics of the DNA Duplex. *J. Phys. Chem. B* **2020**, *124*, 627−640.

(28) Shibutani, T.; Ito, S.; Toda, M.; Kanao, R.; Collins, L. B.; Shibata, M.; Urabe, M.; Koseki, H.; Masuda, Y.; Swenberg, J. A.; et al. Guanine-5-carboxylcytosine base pairs mimic mismatches during DNA replication. *Sci. Rep.* **2015**, *4*, 5220.

(29) Vallurupalli, P.; Bouvignies, G.; Kay, L. E. Studying "Invisible" Excited Protein States in Slow Exchange with a Major State Conformation. *J. Am. Chem. Soc.* **2012**, *134*, 8148−8161.

(30) Liu, B.; Rangadurai, A.; Shi, H.; Al-Hashimi, H. M. Rapid assessment of Watson−Crick to Hoogsteen exchange in unlabeled DNA duplexes using high-power SELOPE imino ^{1}H CEST. *Magn. Reson.* **2021**, *2*, 715−731.

(31) Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M. Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **2011**, *470*, 498−502.

(32) Zhou, H.; Sathyamoorthy, B.; Stelling, A.; Xu, Y.; Xue, Y.; Pigli, Y. Z.; Case, D. A.; Rice, P. A.; Al-Hashimi, H. M. Characterizing Watson-Crick versus Hoogsteen Base Pairing in a DNA-Protein Complex Using Nuclear Magnetic Resonance and Site-Specifically 13C- and 15N-Labeled DNA. *Biochemistry* **2019**, *58*, 1963−1974.

(33) Galindo-Murillo, R.; Roe, D. R.; Cheatham, T. E., III On the absence of intrahelical DNA dynamics on the μs to ms timescale. *Nat. Commun.* **2014**, *5*, 5152.

(34) Choi, S.-R.; Kim, N.-H.; Jin, H.-S.; Seo, Y.-J.; Lee, J.; Lee, J.-H. Base-pair Opening Dynamics of Nucleic Acids in Relation to Their Biological Function. *Comput. Struct. Biotechnol. J.* **2019**, *17*, 797−804.

(35) Schanda, P.; Brutscher, B. Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds. *J. Am. Chem. Soc.* **2005**, *127*, 8014−8015.

(36) Zarycz, M. N. C.; Fonseca Guerra, C. NMR ^{1}H-Shielding Constants of Hydrogen-Bond Donor Reflect Manifestation of the Pauli Principle. *J. Phys. Chem. Lett.* **2018**, *9*, 3720−3724.

(37) Marušič, M.; Schlagnitweit, J.; Petzold, K. RNA Dynamics by NMR Spectroscopy. *ChemBioChem.* **2019**, *20*, 2685−2710.

(38) Shi, H.; Liu, B.; Nussbaumer, F.; Rangadurai, A.; Kreutz, C.; Al-Hashimi, H. M. NMR Chemical Exchange Measurements Reveal That N^{6}-Methyladenosine Slows RNA Annealing. *J. Am. Chem. Soc.* **2019**, *141*, 19988−19993.

(39) Wang, Y.; Han, G.; Jiang, X.; Yuwen, T.; Xue, Y. Chemical shift prediction of RNA imino groups: application toward characterizing RNA excited states. *Nat. Commun.* **2021**, *12*, 1595.

(40) Schlagnitweit, J.; Steiner, E.; Karlsson, H.; Petzold, K. Efficient Detection of Structure and Dynamics in Unlabeled RNAs: The SELOPE Approach. *Chem.—Eur. J.* **2018**, *24*, 6067−6070.

(41) Hardwick, J. S.; Lane, A. N.; Brown, T. Epigenetic Modifications of Cytosine: Biophysical Properties, Regulation, and Function in Mammalian DNA. *BioEssays* **2018**, *40*, 1700199.

(42) Zhu, Q.; Stöger, R.; Alberio, R. A Lexicon of DNA Modifications: Their Roles in Embryo Development and the Germline. *Front. Cell Dev. Biol.* **2018**, *6*, 24.

(43) Wu, H.; Wu, X.; Shen, L.; Zhang, Y. Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat. Biotechnol.* **2014**, *32*, 1231−1240.

(44) Zhu, C.; Gao, Y.; Guo, H.; Xia, B.; Song, J.; Wu, X.; Zeng, H.; Kee, K.; Tang, F.; Yi, C. Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution. *Cell Stem Cell* **2017**, *20*, 720−731.

(45) Lewis, L. C.; Lo, P. C. K.; Foster, J. M.; Dai, N.; Corrêa, I. R.; Durczak, P. M.; Duncan, G.; Ramsawhook, A.; Aithal, G. P.; Denning, C.; et al. Dynamics of 5-carboxylcytosine during hepatic differentiation: Potential general role for active demethylation by DNA repair in lineage specification. *Epigenetics* **2017**, *12*, 277−286.

(46) 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.

(47) Lee, S. H.; Griffiths, J. R. How and Why Are Cancers Acidic? Carbonic Anhydrase IX and the Homeostatic Control of Tumour Extracellular pH. *Cancers (Basel)* **2020**, *12*, 1616.

(48) Swietach, P.; Vaughan-Jones, R. D.; Harris, A. L.; Hulikova, A. The chemistry, physiology and pathology of pH in cancer. *Philos. Trans. R. Soc. B Biol. Sci.* **2014**, *369*, 20130099.

(49) Spruijt, C. G.; Gnerlich, F.; Smits, A. H.; Pfaffeneder, T.; Jansen, P. W.; Bauer, C.; Münzel, M.; Wagner, M.; Müller, M.; Khan, F.; et al. Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell* **2013**, *152*, 1146−1159.

(50) Dodd, T.; Yan, C.; Kossmann, B. R.; Martin, K.; Ivanov, I. Uncovering universal rules governing the selectivity of the archetypal DNA glycosylase TDG. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 5974−5979.

(51) Coey, C. T.; Malik, S. S.; Pidugu, L. S.; Varney, K. M.; Pozharski, E.; Drohat, A. C. Structural basis of damage recognition by thymine DNA glycosylase: Key roles for N-terminal residues. *Nucleic Acids Res.* **2016**, *44*, 10248−10258.

(52) Kanaan, N.; Imhof, P. Interactions of the DNA Repair Enzyme Human Thymine DNA Glycosylase with Cognate and Noncognate DNA. *Biochemistry* **2018**, *57*, 5654−5665.

(53) Schröder, A. S.; Steinbacher, J.; Steigenberger, B.; Gnerlich, F. A.; Schiesser, S.; Pfaffeneder, T.; Carell, T. Synthesis of a DNA Promoter Segment Containing All Four Epigenetic Nucleosides: 5-Methyl-, 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxy- 2'-Deoxy-cytidine. *Angew. Chem., Int. Ed.* **2014**, *53*, 315−318.

(54) Favier, A.; Brutscher, B. NMRlib: user-friendly pulse sequence tools for Bruker NMR spectrometers. *J. Biomol. NMR* **2019**, *73*, 199−211.

3.2 SUPPLEMENTARY MATERIAL

The Supplementary Material file for the above publication contains:

- Additional ^{1}H-^{1}H NOESY, ^{1}H-^{13}C HSQC and ^{1}H-^{15}N HMBC spectra for chemical shift comparisons,

- $\Delta\delta$ plots for all available nuclei and samples,

- CEST and $R_{1\rho}$ profiles for all available nuclei and samples, and

- Comprehensive numerical results of the CEST fitting procedures.

Supporting Information:

^{1}H NMR chemical exchange techniques reveal local and global effects of oxidized cytosine derivatives

Romeo C. A. Dubini,[†,‡] Eva Korytiaková,[†] Thea Schinkel,[†,§] Pia Heinrichs,[†,§]

Thomas Carell,[†] and Petra Rovó[*,†,‡,¶]

†*Faculty of Chemistry and Pharmacy, Department of Chemistry,*

Ludwig-Maximilians-Universität München, Butenandstraße 5-13, 81377 Munich, Germany

‡*Center for Nanoscience (CeNS), Faculty of Physics, Ludwig-Maximilians-Universität*

München, Schellingstraße 4, 5th floor, 80799 Munich, Germany

¶*Institute of Science and Technology (IST), Am Campus 1, 3400 Klosterneuburg, Austria*

§*Contributed equally to this work*

E-mail: petra.rovo@ist.ac.at

List of Figures

List of Tables

Figure S1: ^{1}H-^{13}C HSQC aromatic region spectrum of fC$_{7.0}$, caC$_{7.0}$ and C$_{7.0}$ recorded at 800 MHz at 37 ℃.
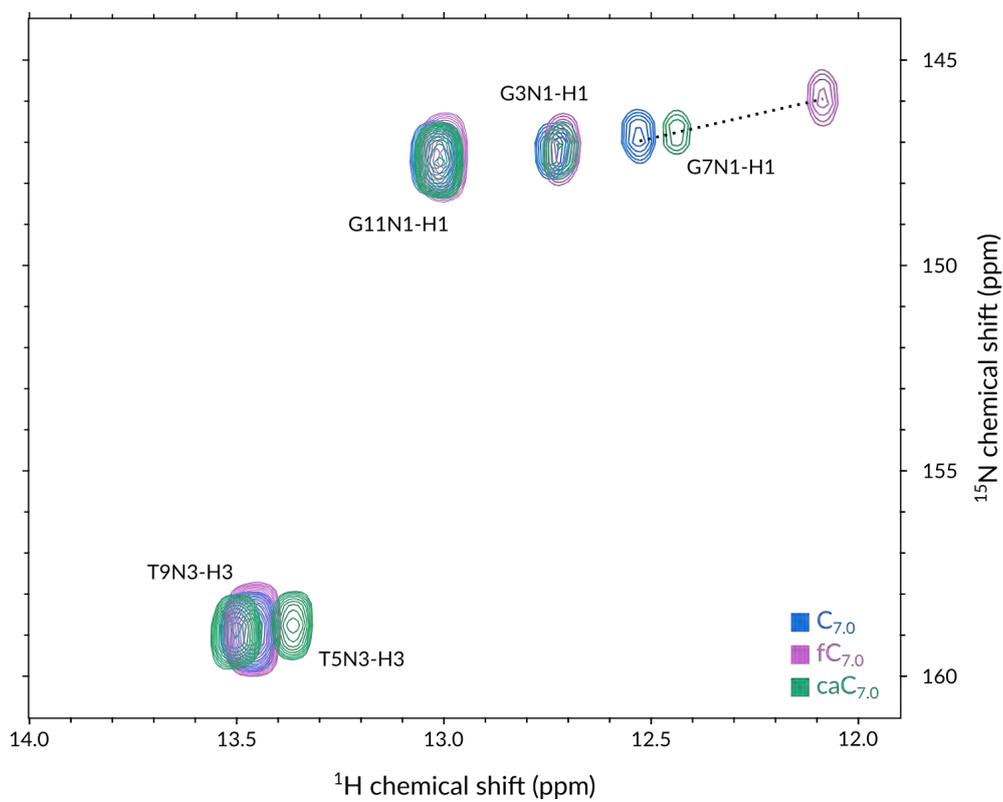
Figure S2: ^{1}H-^{15}N HMQC imino region spectrum of fC$_{7.0}$, caC$_{7.0}$ and C$_{7.0}$ recorded at 800 MHz at 37 ℃.
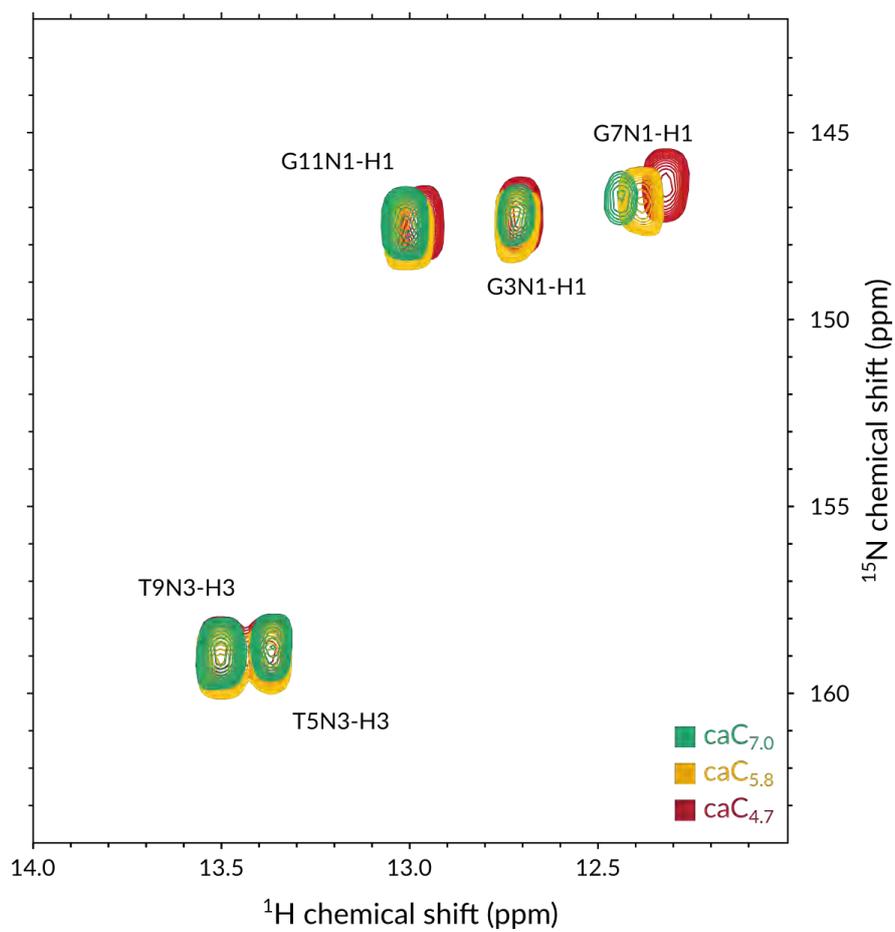
Figure S3: ^{1}H-^{15}N HMQC imino region spectrum of caC$_{7.0}$, caC$_{5.8}$ and caC$_{4.7}$ recorded at 800 MHz at 37 ℃.
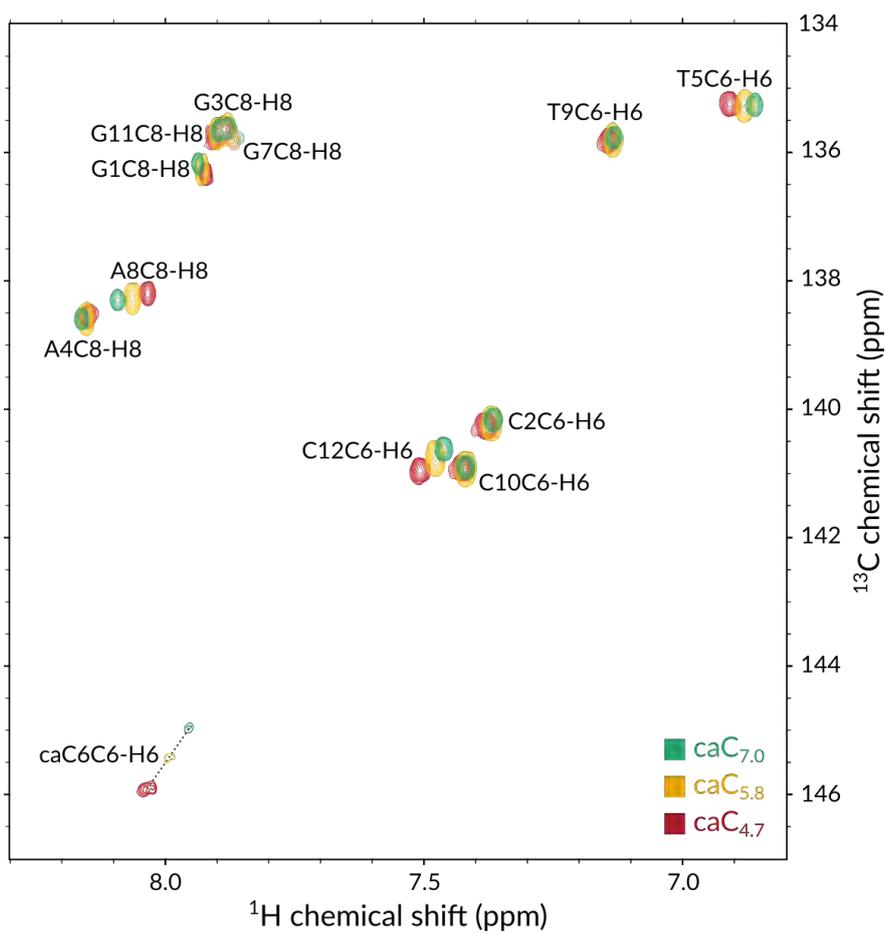
Figure S4: ^{1}H-^{13}C HSQC aromatic region spectrum of caC$_{7.0}$, caC$_{5.8}$ and caC$_{4.7}$ recorded at 800 MHz at 37 ℃.
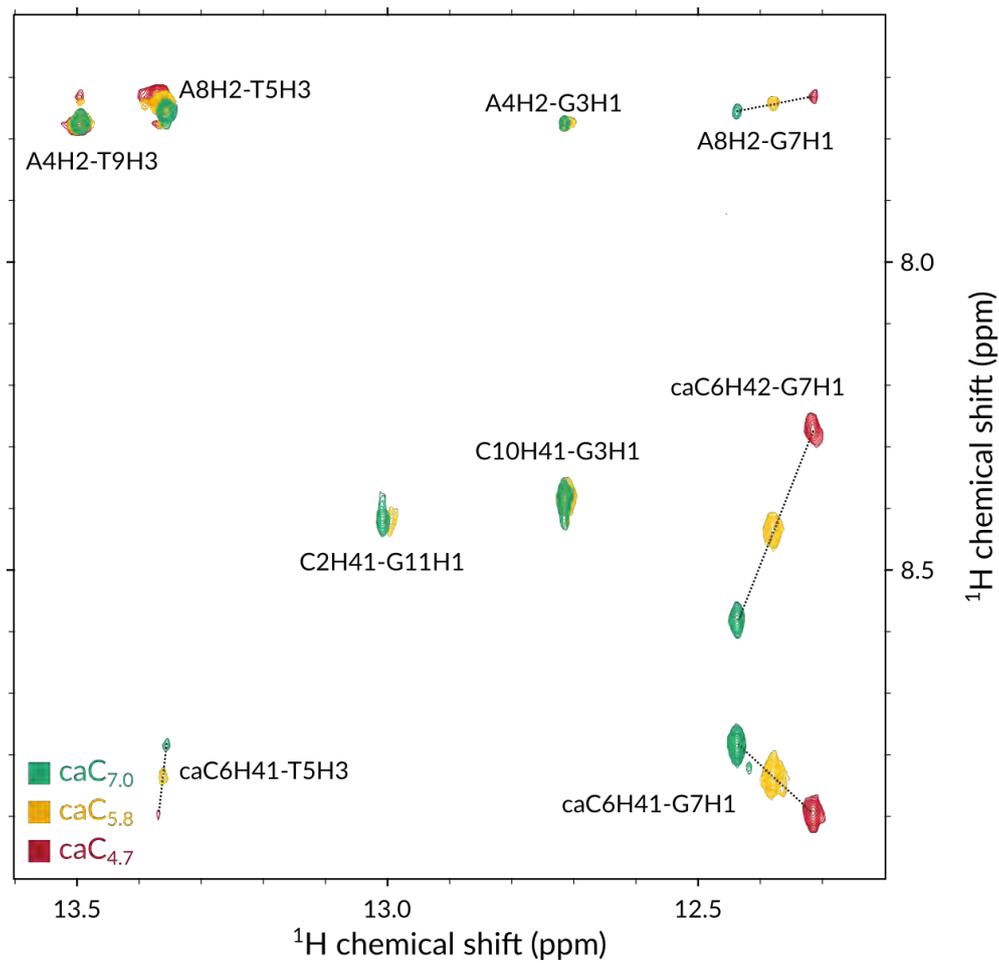
Figure S5: ^{1}H-^{1}H NOESY imino region spectrum of caC$_{7.0}$, caC$_{5.8}$ and caC$_{4.7}$ recorded at 800 MHz at 37 ℃.
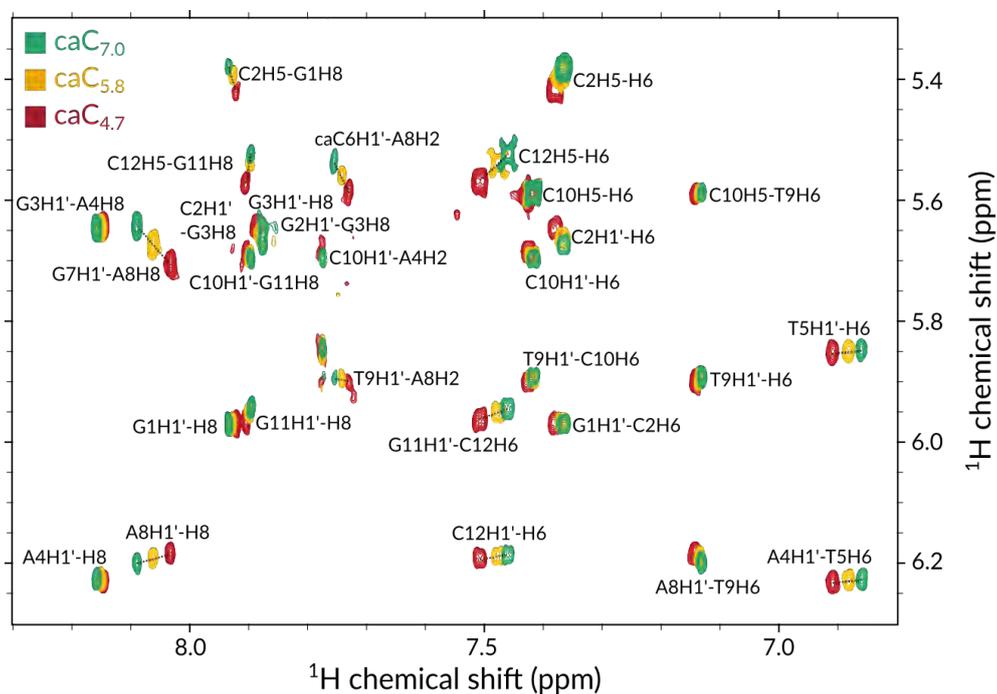
Figure S6: ^{1}H-^{1}H NOESY aromatic region spectrum of caC$_{7.0}$, caC$_{5.8}$ and caC$_{4.7}$ recorded at 800 MHz at 37 ℃.
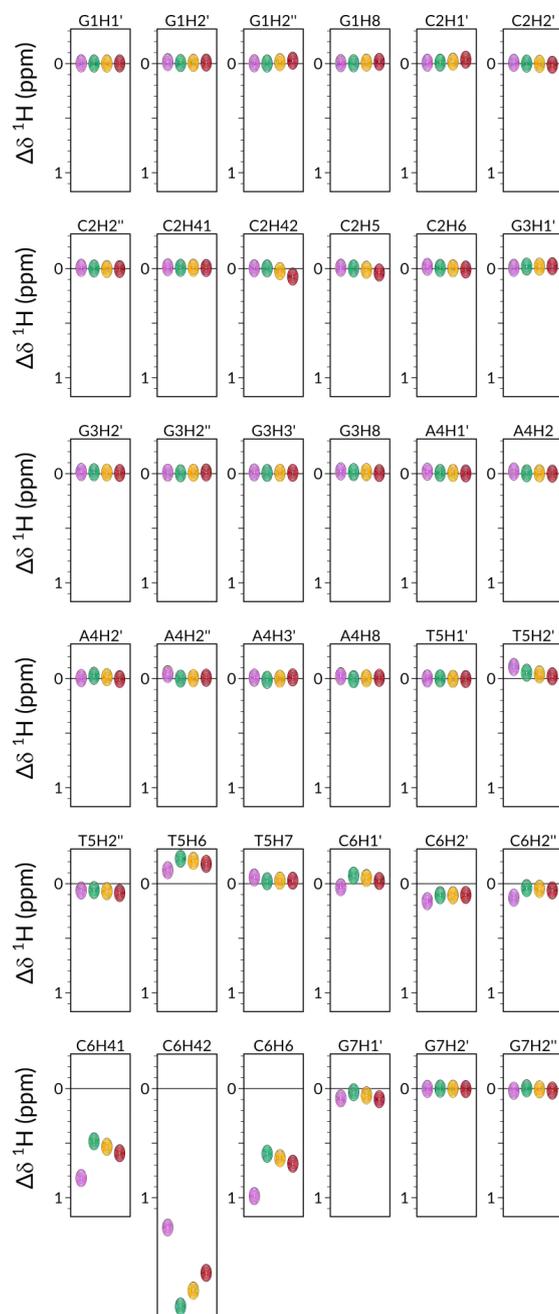
Figure S7: Comparison of chemical shift perturbations $\Delta\delta$ for all comparable aliphatic and aromatic ^{1}H nuclei across all samples in identical conditions. Resonances are reported with respect to the chemical shifts of $C_{7.0}$. Magenta, green, yellow and red symbols represent $fC_{7.0}$, $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$, respectively. Remainder of the figure follows in Fig. S3.

S9
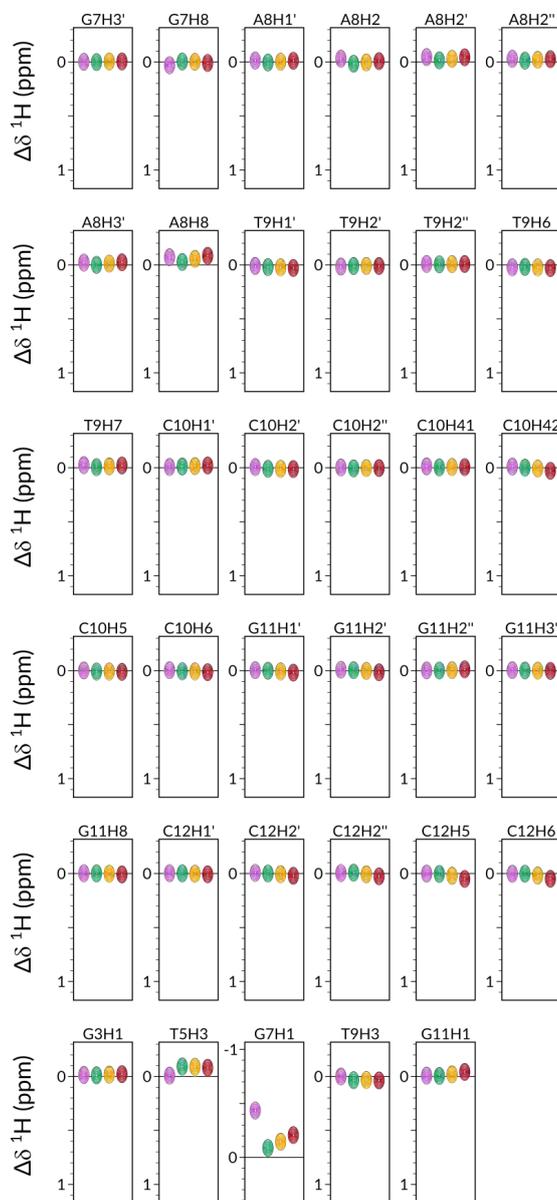
Figure S8: Comparison of chemical shift perturbations $\Delta\delta$ for all comparable aliphatic and aromatic ^{1}H nuclei across all samples in identical conditions. Resonances are reported with respect to the chemical shifts of $C_{7.0}$. Magenta, green, yellow and red symbols represent $fC_{7.0}$, $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$, respectively.
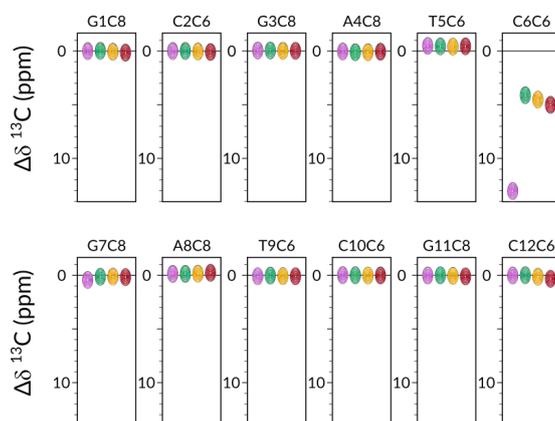
Figure S9: Comparison of chemical shift perturbations $\Delta\delta$ for all comparable aromatic ^{13}C nuclei across all samples in identical conditions. Resonances are reported with respect to the chemical shifts of $C_{7.0}$. Magenta, green, yellow and red symbols represent $fC_{7.0}$, $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$, respectively.
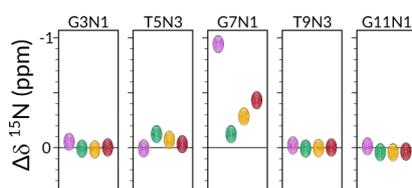
Figure S10: Comparison of chemical shift perturbations $\Delta\delta$ for all comparable aromatic ^{15}N nuclei across all samples in identical conditions. Resonances are reported with respect to the chemical shifts of $C_{7.0}$. Magenta, green, yellow and red symbols represent $fC_{7.0}$, $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$, respectively.
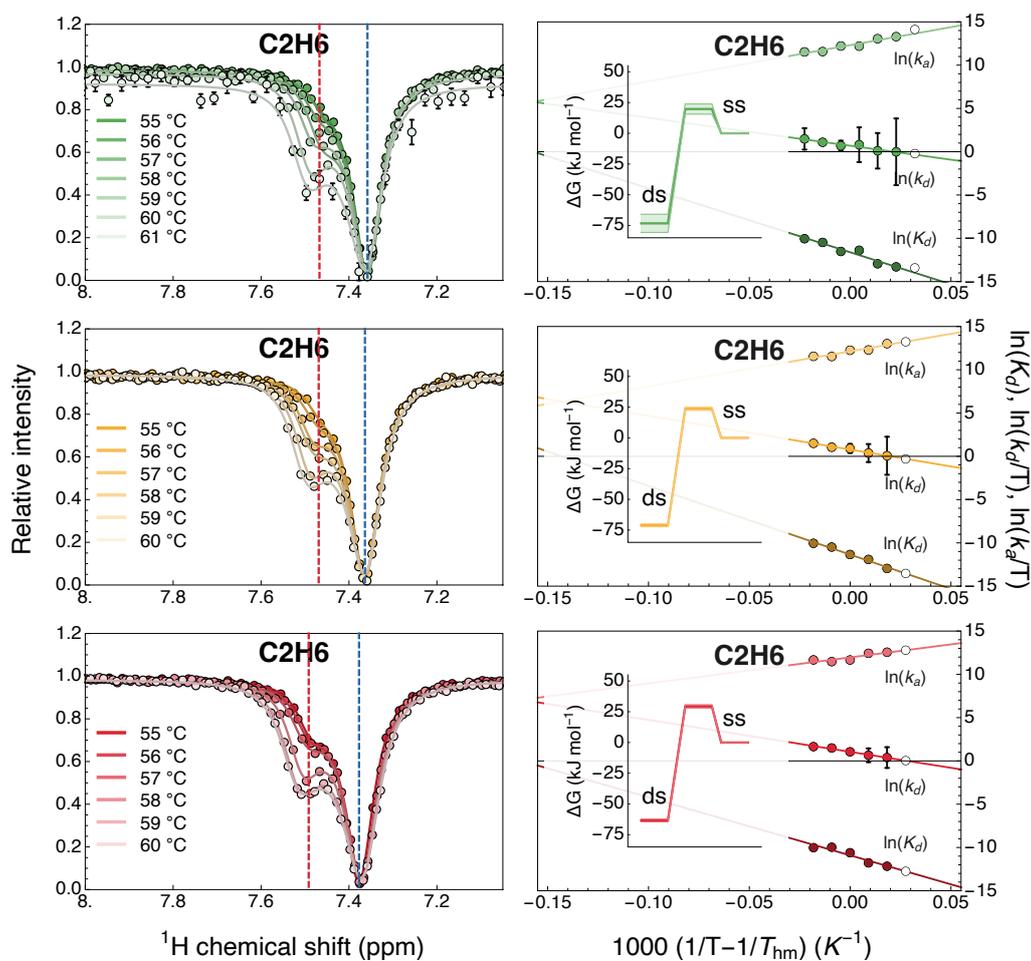
Figure S11: Temperature-dependent CEST melting profiles and van't Hoff plots for proton C2H6. $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$ are shown in shades of green, yellow and red, respectively. Dashed blue and red lines indicate GS and ES chemical shift values. In the van't Hoff plots, shades of green, yellow and red indicate data entries and linear fits for $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$, respectively. White data points represent back-calculated values. Insets present the relevant Gibbs free energy plots at each pH condition at 37 °C.
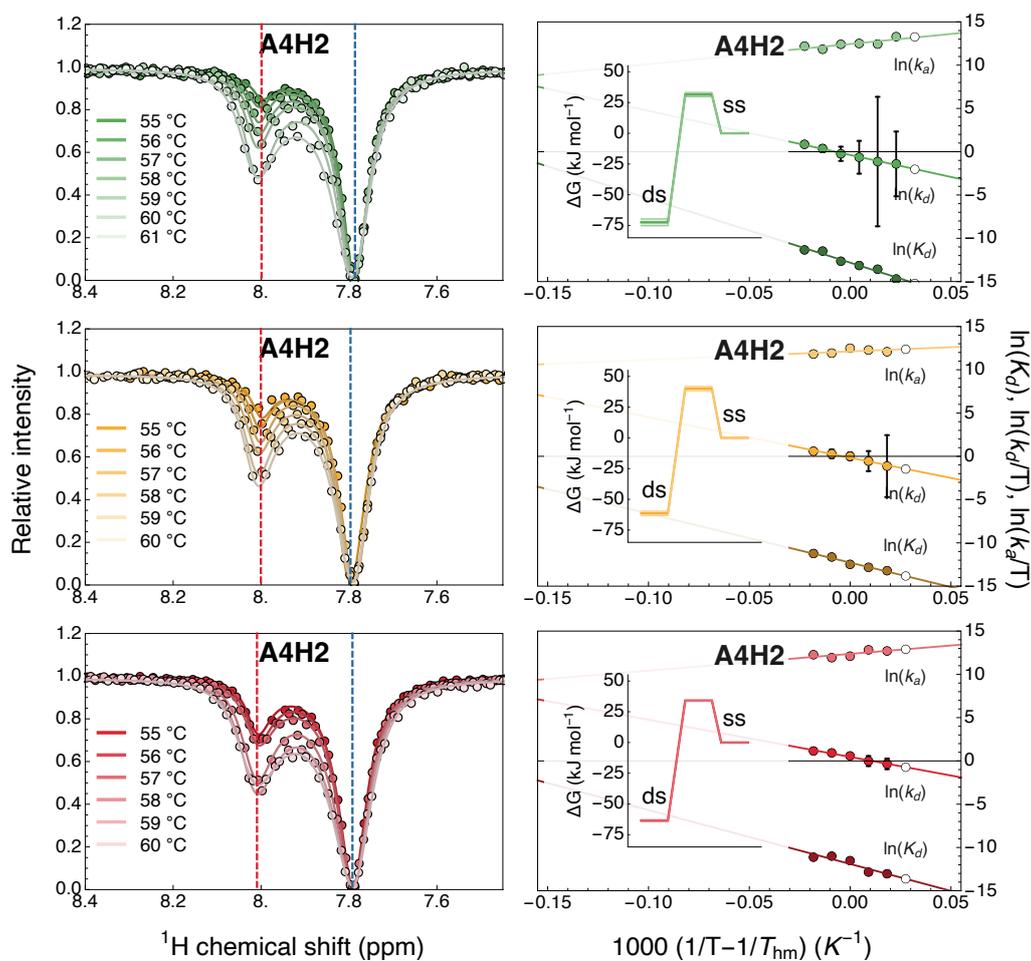
Figure S12: Temperature-dependent CEST melting profiles and van't Hoff plots for proton A4H2. $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$ are shown in shades of green, yellow and red, respectively. Dashed blue and red lines indicate GS and ES chemical shift values. In the van't Hoff plots, shades of green, yellow and red indicate data entries and linear fits for $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$, respectively. White data points represent back-calculated values. Insets present the relevant Gibbs free energy plots at each pH condition at 37 °C.
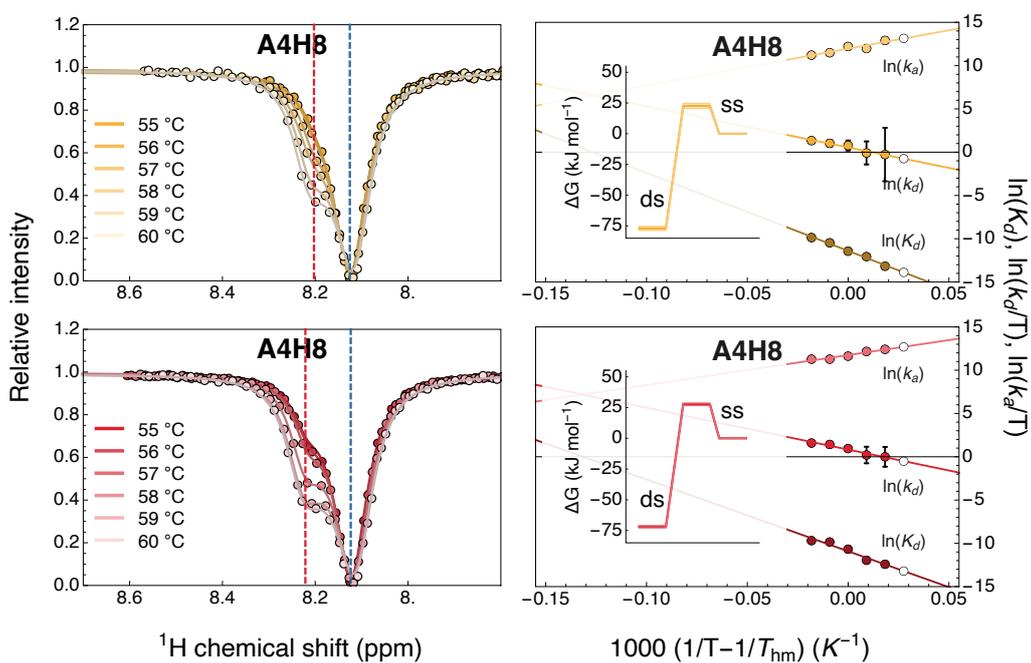
Figure S13: Temperature-dependent CEST melting profiles and van't Hoff plots for proton A4H8. $caC_{5.8}$ and $caC_{4.7}$ are shown in shades of yellow and red, respectively. Dashed blue and red lines indicate GS and ES chemical shift values. In the van't Hoff plots, shades of yellow and red indicate data entries and linear fits for $caC_{5.8}$ and $caC_{4.7}$, respectively. White data points represent back-calculated values. Insets present the relevant Gibbs free energy plots at each pH condition at 37 °C.
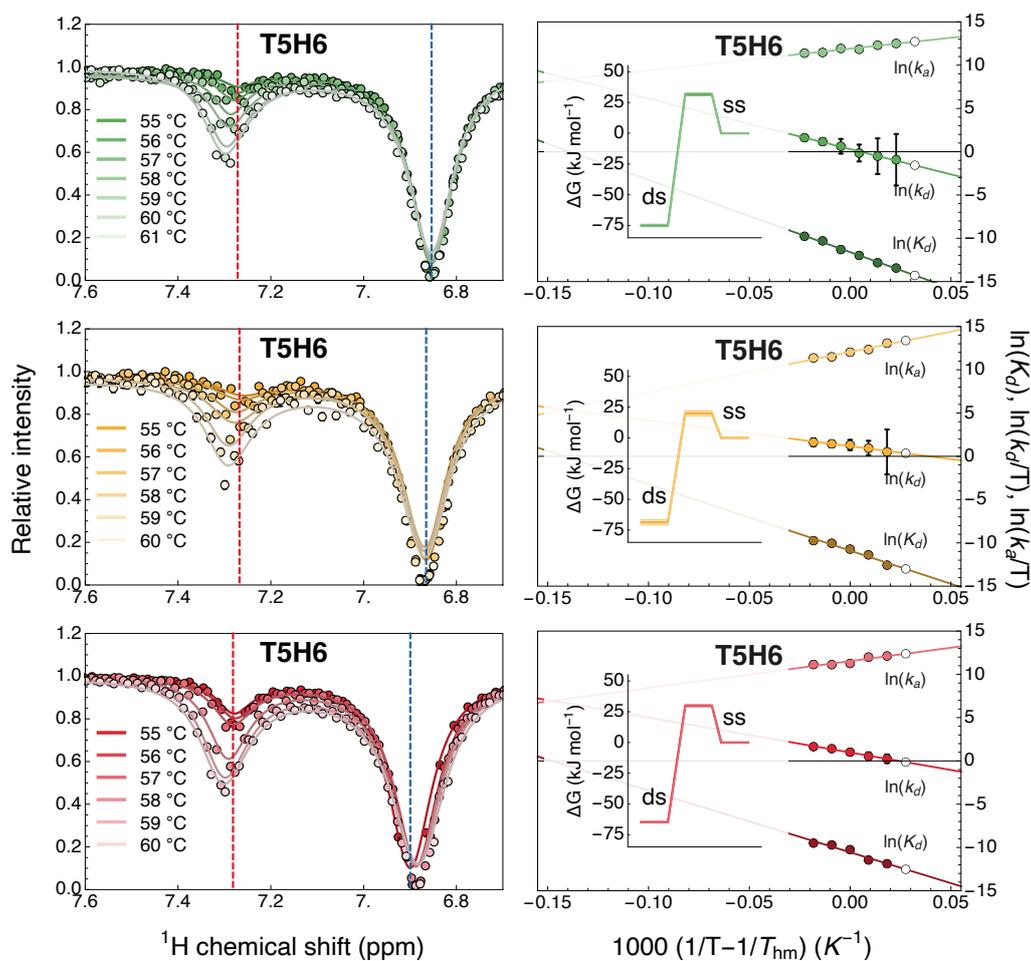
Figure S14: Temperature-dependent CEST melting profiles and van't Hoff plots for proton T5H6. $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$ are shown in shades of green, yellow and red, respectively. Dashed blue and red lines indicate GS and ES chemical shift values. In the van't Hoff plots, shades of green, yellow and red indicate data entries and linear fits for $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$, respectively. White data points represent back-calculated values. Insets present the relevant Gibbs free energy plots at each pH condition at 37 °C.
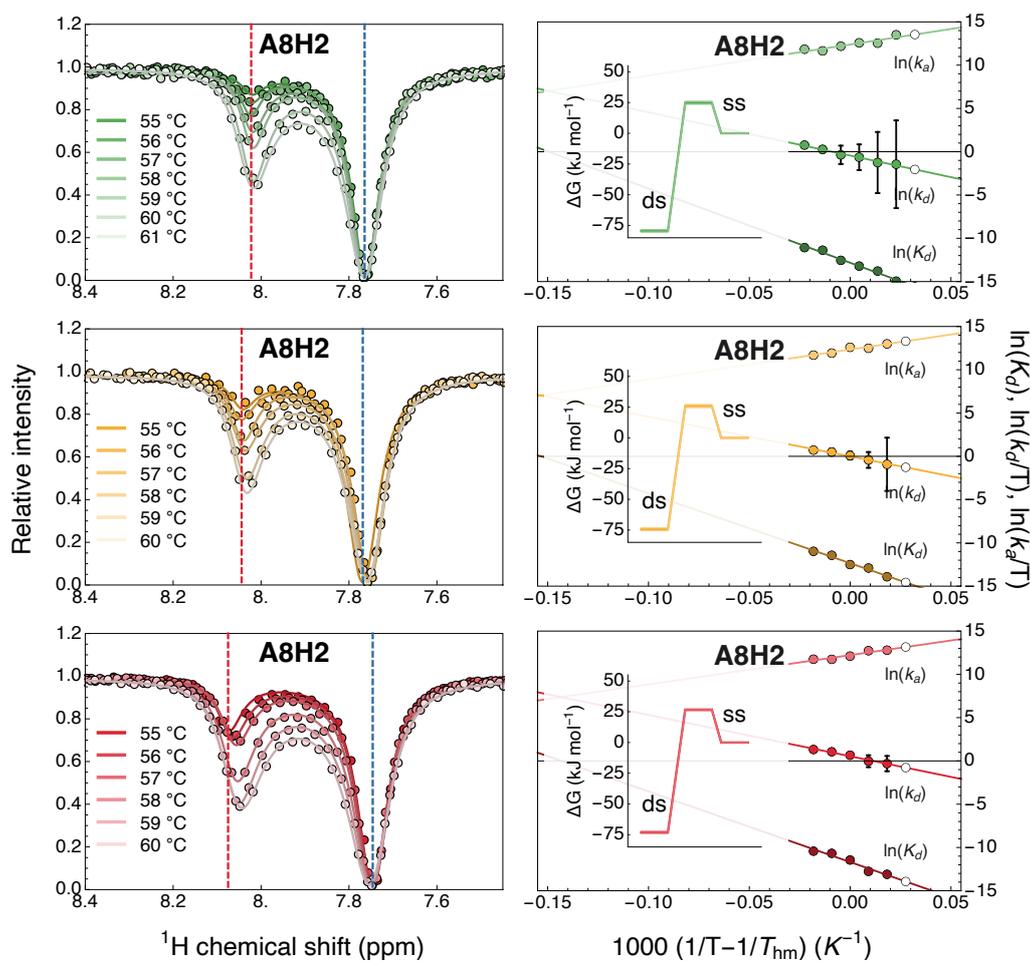
Figure S15: Temperature-dependent CEST melting profiles and van't Hoff plots for proton A8H2. $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$ are shown in shades of green, yellow and red, respectively. Dashed blue and red lines indicate GS and ES chemical shift values. In the van't Hoff plots, shades of green, yellow and red indicate data entries and linear fits for $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$, respectively. White data points represent back-calculated values. Insets present the relevant Gibbs free energy plots at each pH condition at 37 ℃.
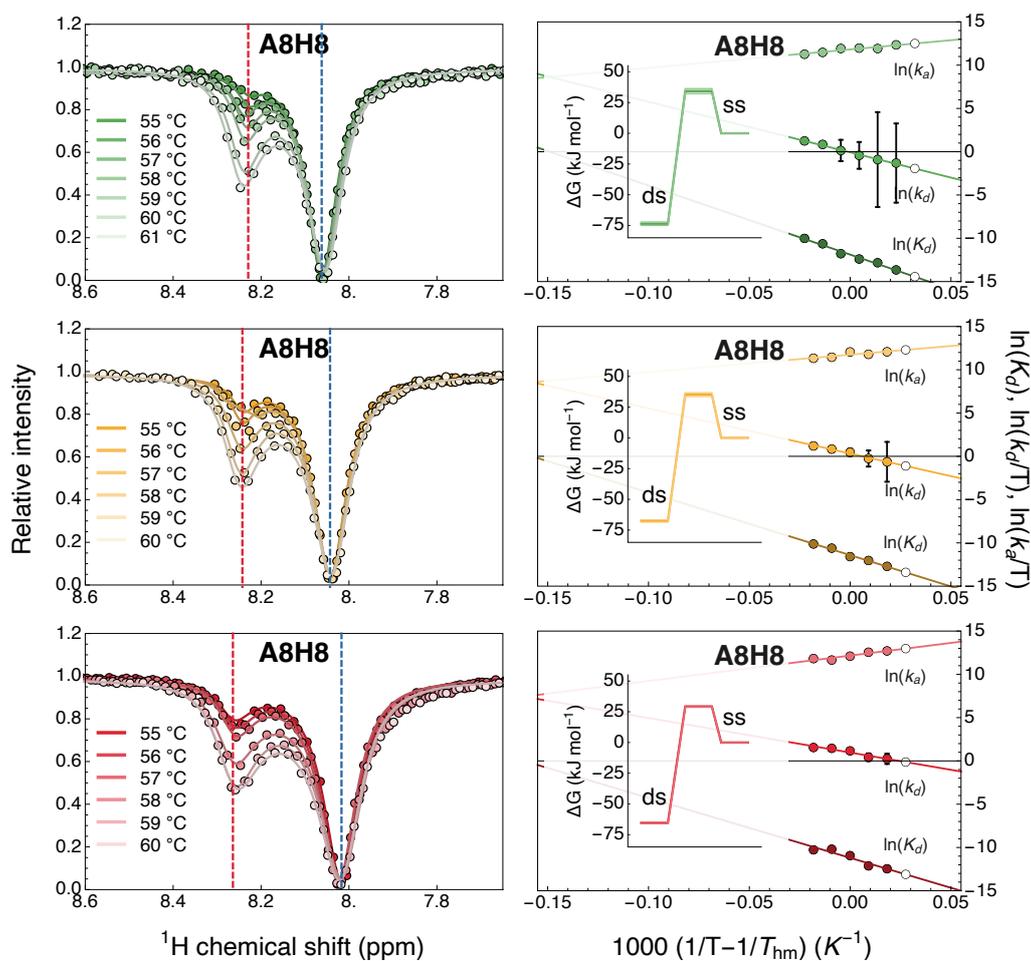
Figure S16: Temperature-dependent CEST melting profiles and van't Hoff plots for proton A8H8. $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$ are shown in shades of green, yellow and red, respectively. Dashed blue and red lines indicate GS and ES chemical shift values. In the van't Hoff plots, shades of green, yellow and red indicate data entries and linear fits for $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$, respectively. White data points represent back-calculated values. Insets present the relevant Gibbs free energy plots at each pH condition at 37 °C.
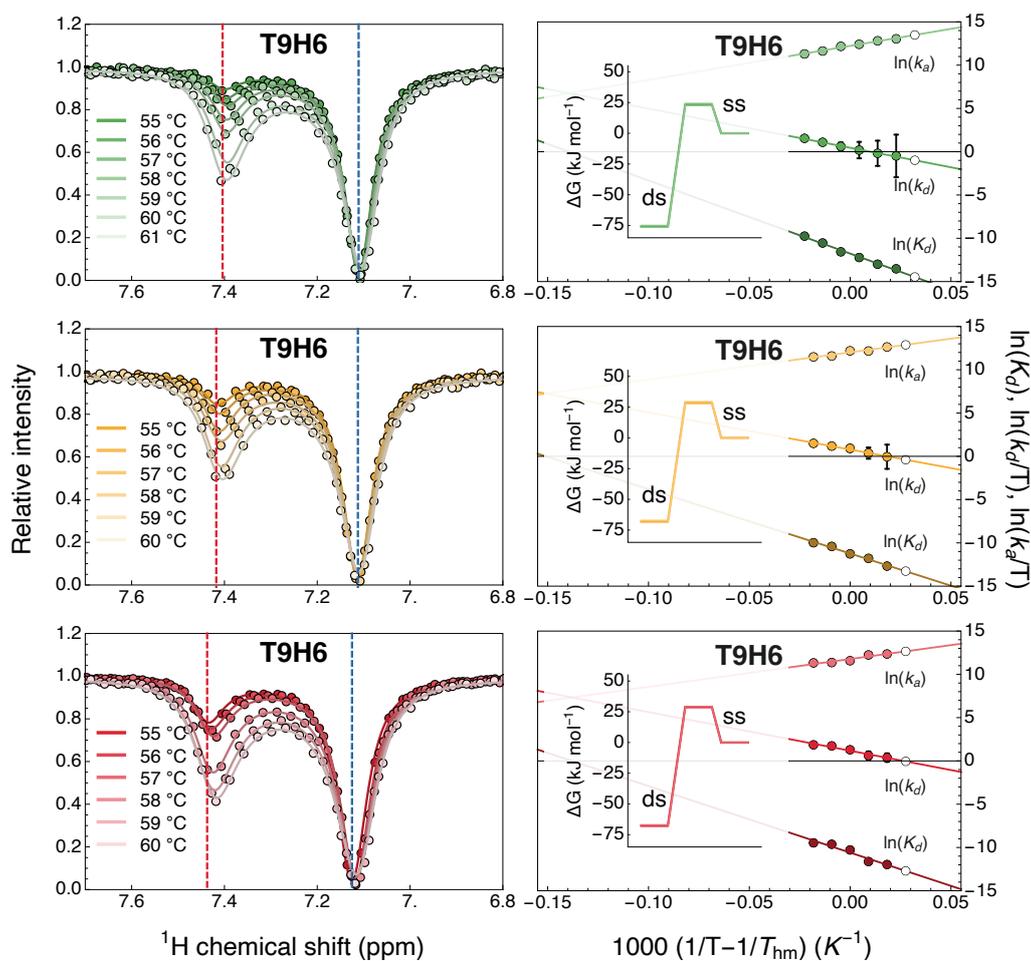
Figure S17: Temperature-dependent CEST melting profiles and van't Hoff plots for proton T9H6. $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$ are shown in shades of green, yellow and red, respectively. Dashed blue and red lines indicate GS and ES chemical shift values. In the van't Hoff plots, shades of green, yellow and red indicate data entries and linear fits for $caC_{7.0}$, $caC_{5.8}$ and $caC_{4.7}$, respectively. White data points represent back-calculated values. Insets present the relevant Gibbs free energy plots at each pH condition at 37 ℃.
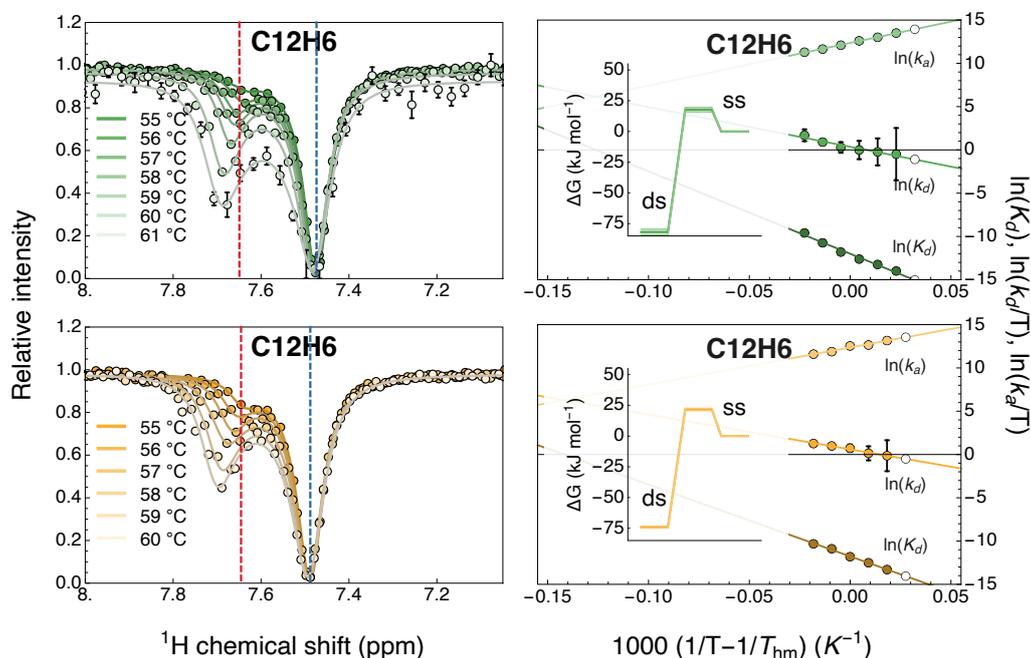
Figure S18: Temperature-dependent CEST melting profiles and van't Hoff plots for proton C12H6. $caC_{7.0}$, and $caC_{5.8}$ are shown in shades of green and yellow, respectively. Dashed blue and red lines indicate GS and ES chemical shift values. In the van't Hoff plots, shades of green and yellow indicate data entries and linear fits for $caC_{7.0}$ and $caC_{5.8}$, respectively. White data points represent back-calculated values. Insets present the relevant Gibbs free energy plots at each pH condition at 37 °C.
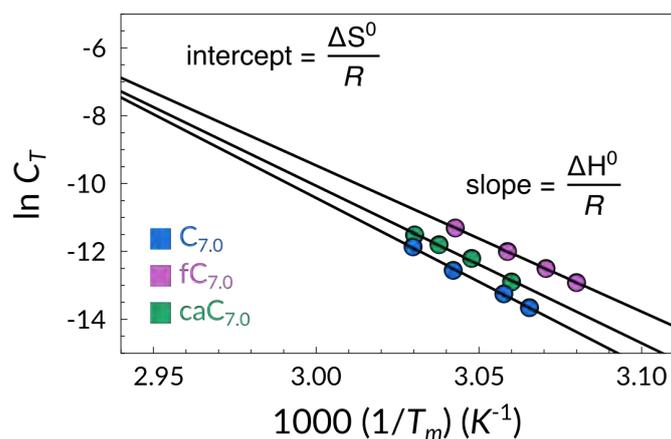
Figure S19: Concentration-dependent UV/Vis melting temperatures analysis of $C_{7.0}$ (blue), $fC_{7.0}$ (magenta) and $caC_{7.0}$ (green). The reciprocal of the observed T_m is plotted as a function of the logarithm of the total DNA concentration (C_t). The slope and the intercept of the fitted linear corresponds to $\Delta H^0/R$, and $\Delta S^0/R$, respectively.
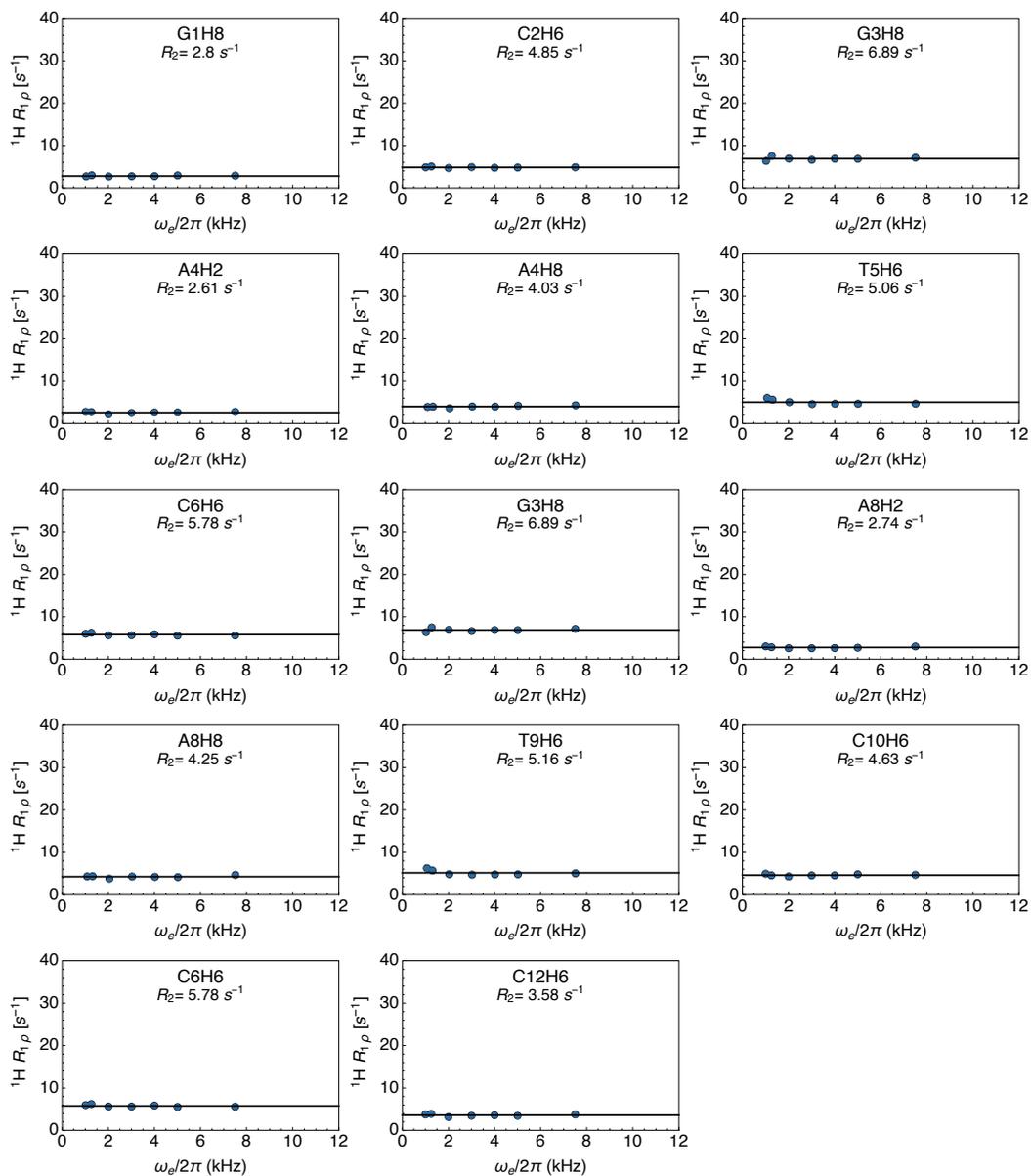
Figure S20: On-resonance $R_{1\rho}$ relaxation dispersion profiles for each available proton reported of $C_{7.0}$ recorded at 55 °C. Lines represent best fits either to models accounting for (color-coded) or discounting chemical exchange (black). Fitted R_2 relaxation rates are displayed for each site.
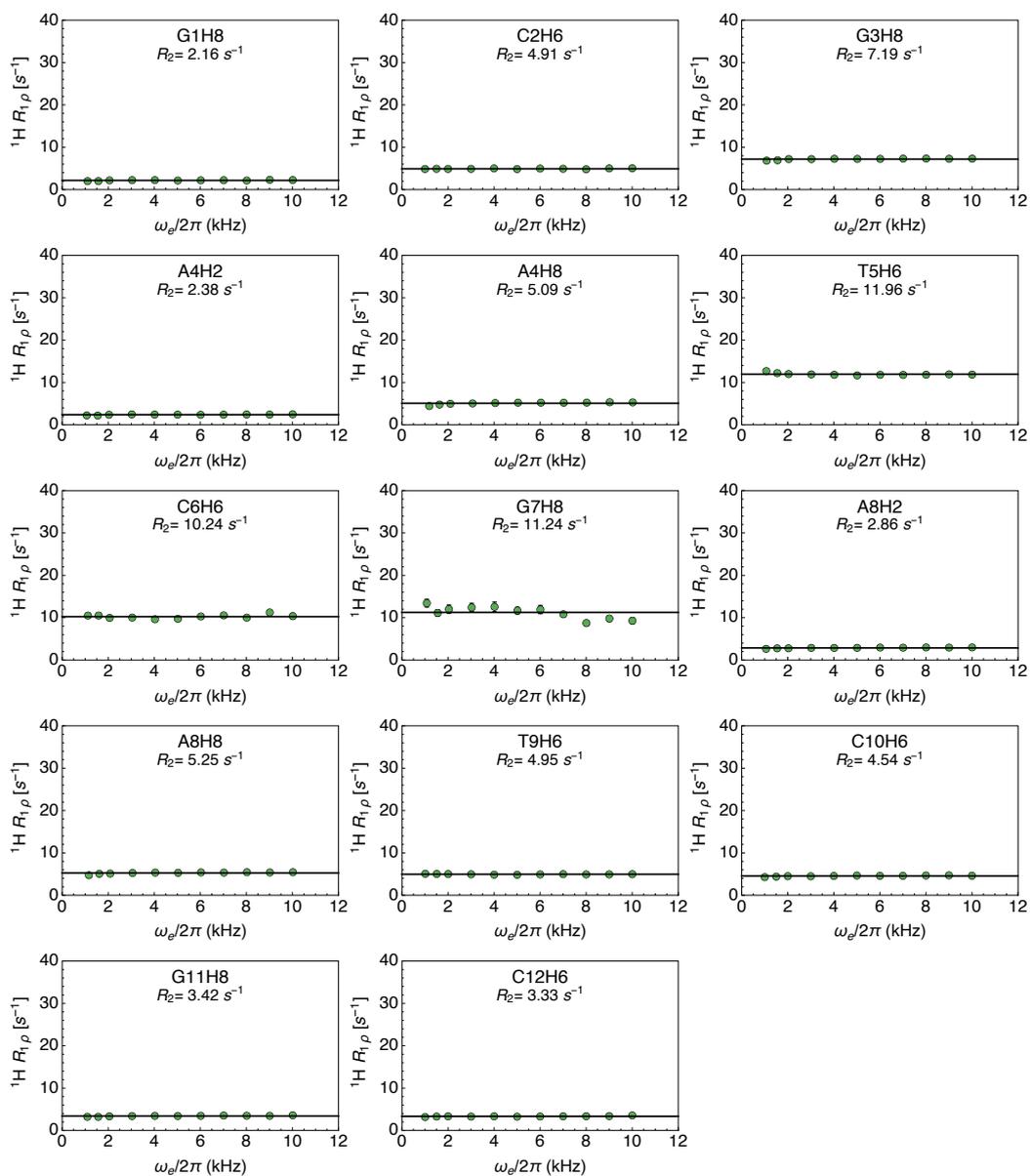
Figure S21: On-resonance $R_{1\rho}$ relaxation dispersion profiles for each available proton reported of caC$_{7.0}$ recorded at 55 °C. Lines represent best fits either to models accounting for (color-coded) or discounting chemical exchange (black). Fitted R_2 relaxation rates are displayed for each site.
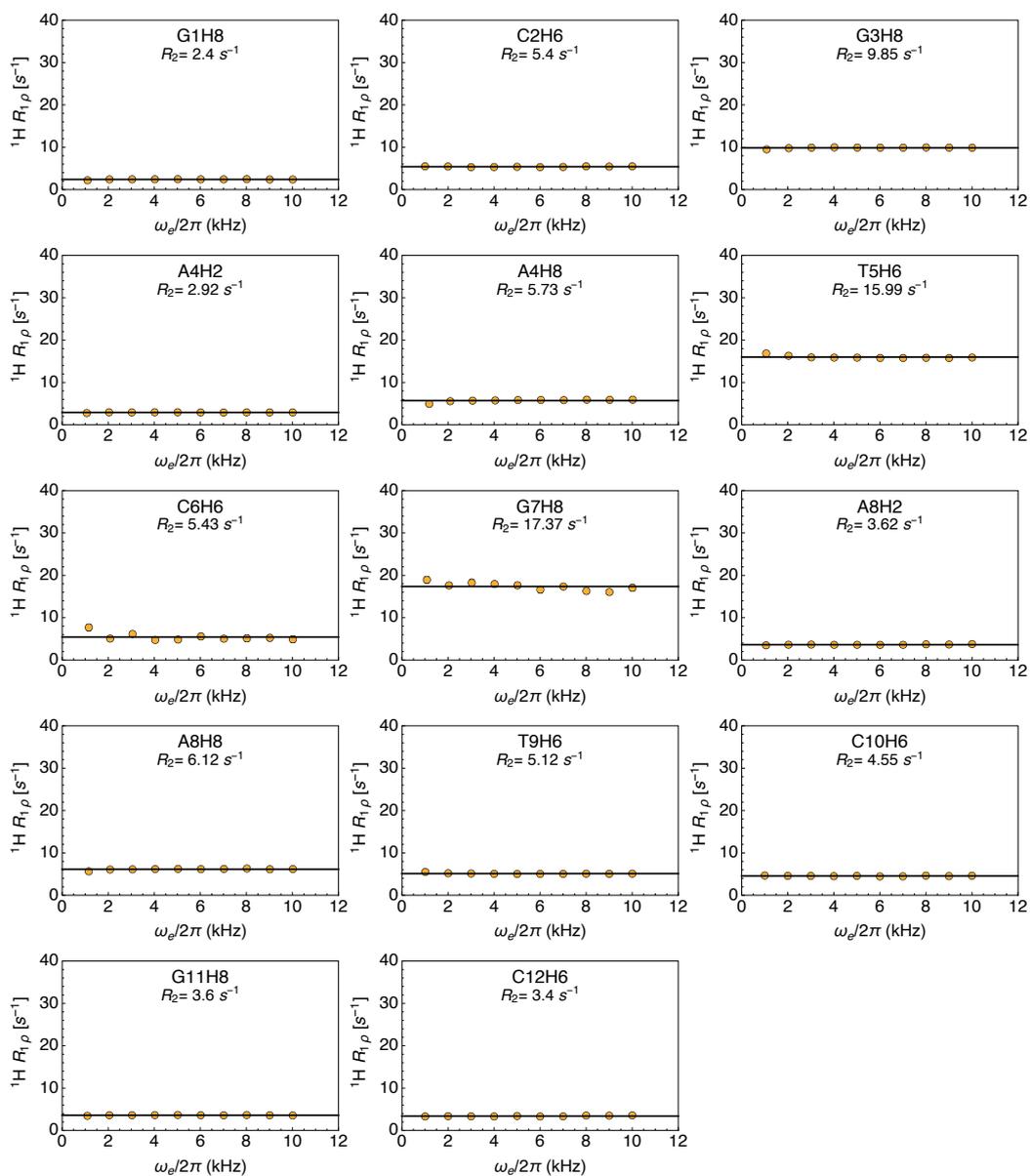
Figure S22: On-resonance $R_{1\rho}$ relaxation dispersion profiles for each available proton reported of $caC_{5.8}$ recorded at 55 °C. Lines represent best fits either to models accounting for (color-coded) or discounting chemical exchange (black). Fitted R_2 relaxation rates are displayed for each site.
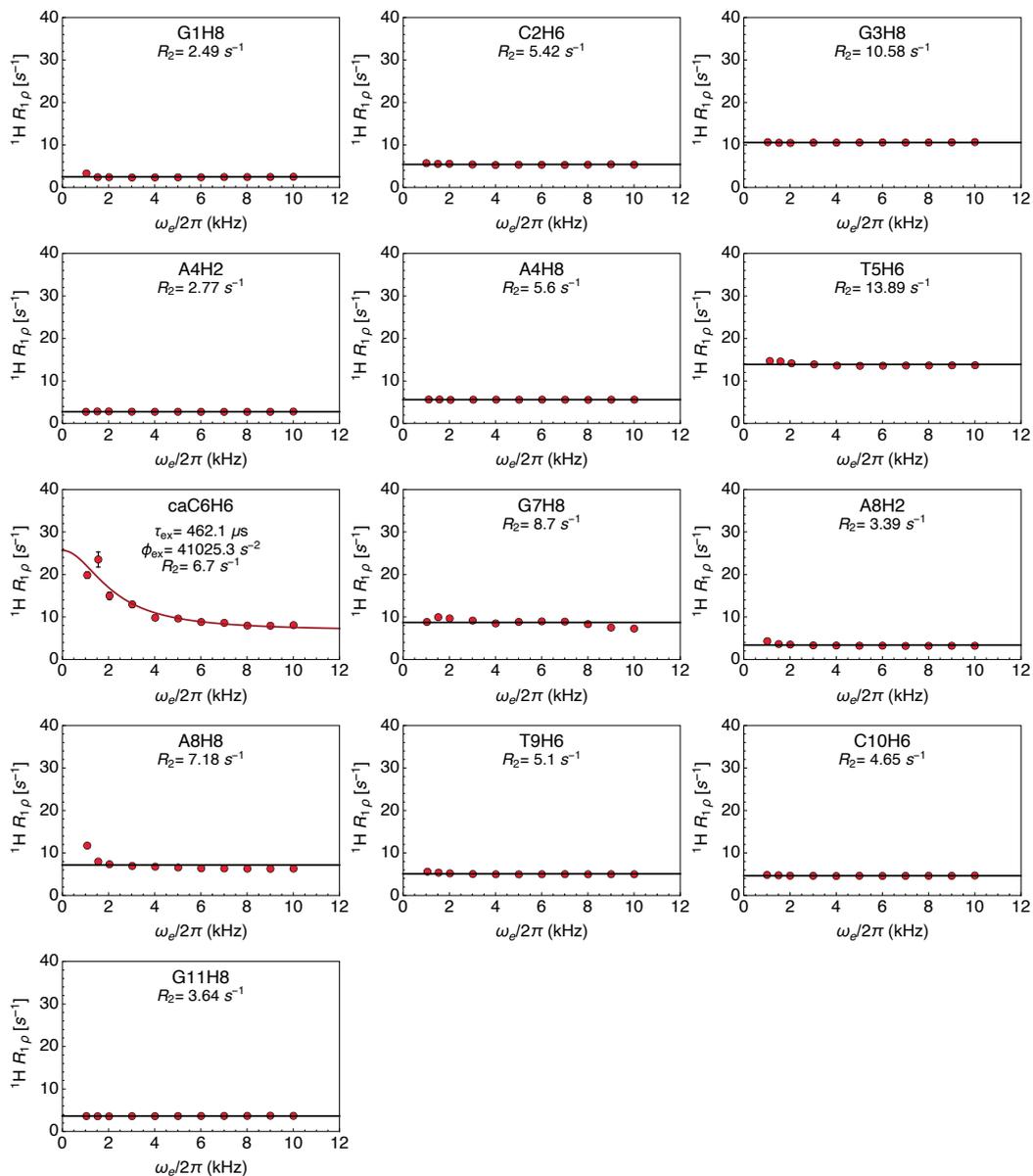
Figure S23: On-resonance $R_{1\rho}$ relaxation dispersion profiles for each available proton reported of caC$_{4.7}$ recorded at 55 °C. Lines represent best fits either to models accounting for (color-coded) or discounting chemical exchange (black). Fitted R_2 relaxation rates are displayed for each site, while exchange parameters such as τ_{ex} and ϕ_{ex} are only available for profiles best fit to the exchange model.
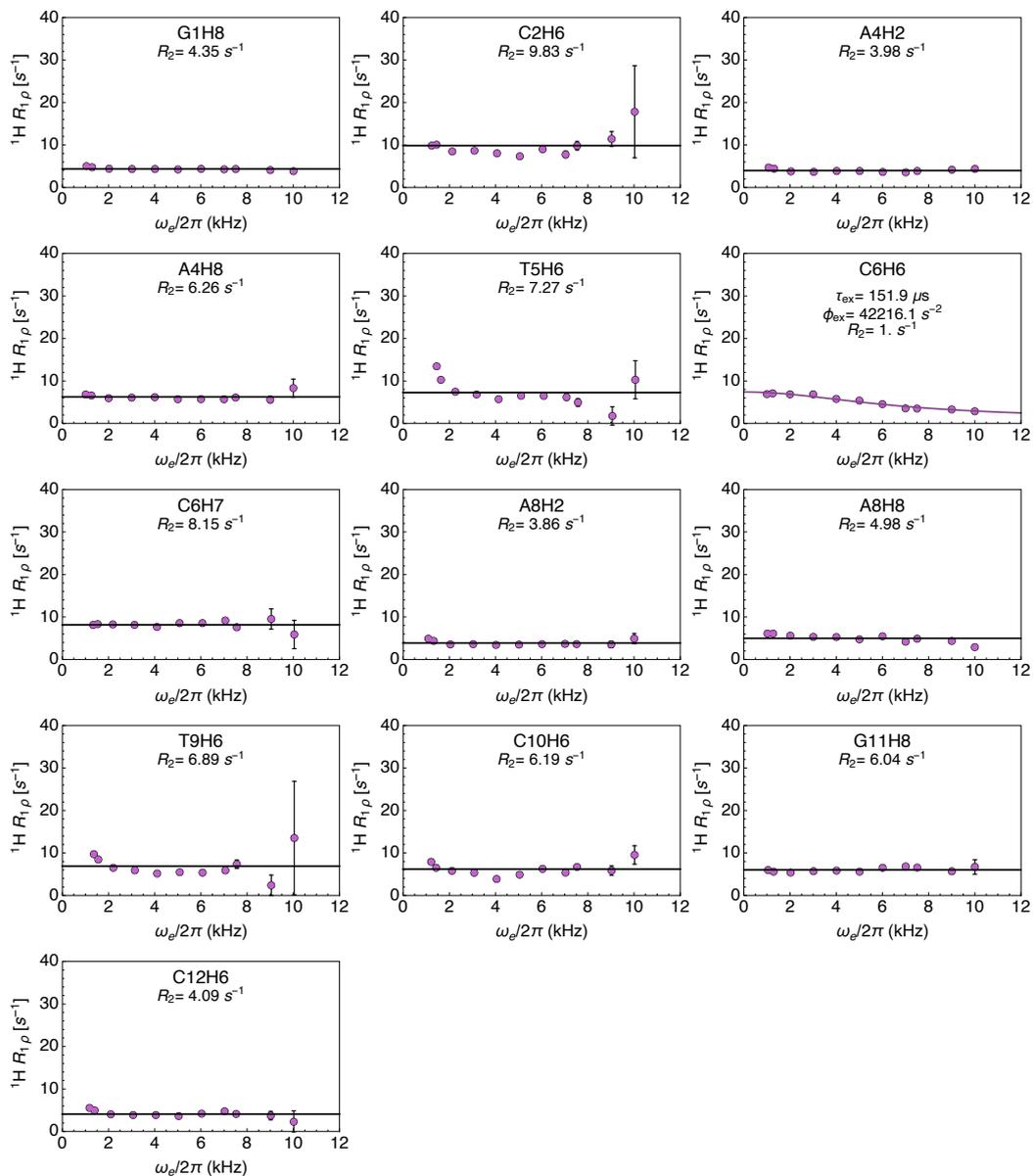
Figure S24: On-resonance $R_{1\rho}$ relaxation dispersion profiles for each available proton reported of $fC_{7.0}$ recorded at 55 °C. Lines represent best fits either to models accounting for (color-coded) or discounting chemical exchange (black). Fitted R_2 relaxation rates are displayed for each site, while exchange parameters such as τ_{ex} and ϕ_{ex} are only available for profiles best fit to the exchange model.

Table S1: Results of the fit of ^{1}H CEST measured at 800 MHz ^{1}H Larmor frequency. The temperature of each measurement is indicated on the leftmost column of the table. Chemical shifts relevant to dsDNA and ssDNA are given together with the relaxation parameters R_1 and R_2 for each available reporter of sample caC$_{7.0}$. Errors are given as one standard deviation.

		Chemical shift		Relaxation rates	
		δ(dsDNA) (ppm)	δ(ssDNA) (ppm)	R_1 (s^{-1})	R_2 (s^{-1})
	C2H6	7.36 ± 2.17	7.462 ± 0.012	1.74 ± 0.02	4.82 ± 0.01
	A4H2	7.79 ± 0.44	8.008 ± 0.001	0.73 ± 0.02	1.94 ± 0.03
	T5H6	6.85 ± 0.37	7.276 ± 0.003	2.1 ± 0.02	11.51 ± 0.03
56 ℃	A8H2	7.76 ± 0.62	8.025 ± 0.001	0.74 ± 0.02	2.2 ± 0.04
	A8H8	8.06 ± 0.27	8.229 ± 0.001	1.47 ± 0.02	5.08 ± 0.01
	T9H6	7.11 ± 0.29	7.402 ± 0.001	1.84 ± 0.03	4.86 ± 0.01
	C10H6	7.41 ± 0.18	7.557 ± 0.001	1.79 ± 0.03	4.47 ± 0.01
	C12H6	7.47 ± 0.27	7.649 ± 0.002	1.36 ± 0.02	3.17 ± 0.01
	C2H6	7.36 ± 0.12	7.473 ± 0.002	1.67 ± 0.03	4.78 ± 0.04
	A4H2	7.79 ± 2.44	8.008 ± 0.004	0.73 ± 0.03	2.1 ± 0.07
	T5H6	6.86 ± 0.5	7.281 ± 0.002	2.08 ± 0.02	11.36 ± 0.03
57 ℃	A8H2	7.76 ± 0.54	8.026 ± 0.001	0.76 ± 0.02	2.41 ± 0.05
	A8H8	8.06 ± 0.4	8.24 ± 0.002	1.44 ± 0.02	5.04 ± 0.01
	T9H6	7.11 ± 0.23	7.398 ± 0.001	1.77 ± 0.02	4.73 ± 0.01
	C10H6	7.4 ± 0.11	7.559 ± 0.001	1.73 ± 0.02	4.39 ± 0.01
	C12H6	7.48 ± 0.13	7.66 ± 0.002	1.31 ± 0.02	3.14 ± 0.01
	C2H6	7.36 ± 2.32	7.467 ± 0.047	1.62 ± 0.04	4.57 ± 0.04
	A4H2	7.79 ± 0.14	8.008 ± 0.001	0.74 ± 0.02	2.24 ± 0.01
	T5H6	6.86 ± 0.24	7.288 ± 0.002	2.09 ± 0.03	10.95 ± 0.03
58 ℃	A8H2	7.76 ± 0.21	8.022 ± 0.001	0.73 ± 0.03	2.22 ± 0.06
	A8H8	8.06 ± 0.1	8.235 ± 0.001	1.41 ± 0.02	4.96 ± 0.01
	T9H6	7.11 ± 0.12	7.4 ± 0.001	1.73 ± 0.02	4.63 ± 0.01
	C10H6	7.4 ± 0.08	7.561 ± 0.001	1.69 ± 0.02	4.29 ± 0.01
	C12H6	7.48 ± 0.1	7.665 ± 0.001	1.28 ± 0.02	3.1 ± 0.01

(*Continues on the next page...*)

59 ℃	C2H6	7.36 ± 0.03	7.485 ± 0.001	1.6 ± 0.02	4.4 ± 0.05
	A4H2	7.79 ± 0.13	8.011 ± 0.001	0.72 ± 0.03	2. ± 0.05
	T5H6	6.86 ± 0.18	7.293 ± 0.003	2.04 ± 0.04	10.69 ± 0.11
	A8H2	7.76 ± 0.17	8.02 ± 0.001	0.71 ± 0.02	2.21 ± 0.06
	A8H8	8.06 ± 0.21	8.238 ± 0.002	1.39 ± 0.02	4.71 ± 0.1
	T9H6	7.11 ± 0.06	7.399 ± 0.001	1.7 ± 0.02	4.75 ± 0.05
	C10H6	7.4 ± 0.05	7.566 ± 0.001	1.66 ± 0.03	4.29 ± 0.05
	C12H6	7.48 ± 0.05	7.673 ± 0.001	1.26 ± 0.02	3.1 ± 0.04
60 ℃	C2H6	7.36 ± 0.02	7.491 ± 0.001	1.52 ± 0.02	4.18 ± 0.02
	A4H2	7.79 ± 0.06	8.01 ± 0.001	0.74 ± 0.03	2.08 ± 0.02
	T5H6	6.87 ± 0.07	7.298 ± 0.003	2.19 ± 0.04	10.31 ± 0.06
	A8H2	7.77 ± 0.06	8.02 ± 0.001	0.72 ± 0.02	2.43 ± 0.02
	A8H8	8.06 ± 0.03	8.241 ± 0.001	1.33 ± 0.02	4.53 ± 0.02
	T9H6	7.12 ± 0.04	7.398 ± 0.001	1.58 ± 0.02	4.23 ± 0.02
	C10H6	7.41 ± 0.02	7.568 ± 0.001	1.58 ± 0.03	3.9 ± 0.02
	C12H6	7.49 ± 0.04	7.687 ± 0.001	1.21 ± 0.02	2.73 ± 0.02
61 ℃	C2H6	7.37 ± 0.07	7.498 ± 0.005	1.48 ± 0.03	4.59 ± 0.24
	A4H2	7.8 ± 0.06	8.014 ± 0.001	0.73 ± 0.02	2.52 ± 0.04
	T5H6	6.88 ± 0.03	7.301 ± 0.001	2.14 ± 0.04	10.77 ± 0.13
	A8H2	7.77 ± 0.06	8.021 ± 0.001	0.72 ± 0.02	2.62 ± 0.05
	A8H8	8.06 ± 0.02	8.244 ± 0.001	1.34 ± 0.02	4.86 ± 0.04
	T9H6	7.12 ± 0.02	7.4 ± 0.001	1.6 ± 0.02	4.59 ± 0.04
	C10H6	7.42 ± 0.12	7.578 ± 0.01	1.53 ± 0.12	4.3 ± 0.24
	C12H6	7.5 ± 0.1	7.693 ± 0.008	1.17 ± 0.03	2.92 ± 0.25

Table S2: Results of the fit of ^{1}H CEST measured at 800 MHz ^{1}H Larmor frequency. The temperature of each measurement is indicated on the leftmost column of the table. Chemical shifts relevant to dsDNA and ssDNA are given together with the relaxation parameters R_1 and R_2 for each available reporter of sample caC$_{5.8}$. Errors are given as one standard deviation.

		Chemical shift		Relaxation rates	
		δ(dsDNA) (ppm)	δ(ssDNA) (ppm)	R_1 (s^{-1})	R_2 (s^{-1})
	C2H6	7.36 ± 0.11	7.48 ± 0.002	1.74 ± 0.025	5.23 ± 0.01
	A4H2	7.79 ± 0.38	8.005 ± 0.001	0.83 ± 0.015	2.64 ± 0.
	A4H8	8.12 ± 0.12	8.205 ± 0.002	1.53 ± 0.02	5.5 ± 0.01
	T5H6	6.87 ± 0.51	7.272 ± 0.004	2.67 ± 0.045	14.4 ± 0.05
56 ℃	A8H2	7.75 ± 0.39	8.047 ± 0.001	0.82 ± 0.017	3.11 ± 0.02
	A8H8	8.04 ± 0.14	8.244 ± 0.001	1.48 ± 0.02	5.88 ± 0.01
	T9H6	7.11 ± 0.14	7.415 ± 0.001	1.81 ± 0.028	4.89 ± 0.01
	C10H6	7.41 ± 0.1	7.579 ± 0.001	1.76 ± 0.025	4.54 ± 0.01
	C12H6	7.49 ± 0.12	7.661 ± 0.002	1.35 ± 0.024	3.29 ± 0.01
	C2H6	7.36 ± 0.05	7.486 ± 0.001	1.67 ± 0.03	5.08 ± 0.05
	A4H2	7.79 ± 0.11	8.005 ± 0.001	0.8 ± 0.02	2.22 ± 0.03
	A4H8	8.12 ± 0.03	8.207 ± 0.001	1.49 ± 0.02	5.51 ± 0.03
	T5H6	6.87 ± 0.17	7.281 ± 0.003	2.96 ± 0.05	15.21 ± 0.17
57 ℃	A8H2	7.76 ± 0.14	8.042 ± 0.001	0.8 ± 0.01	2.52 ± 0.04
	A8H8	8.04 ± 0.07	8.244 ± 0.001	1.44 ± 0.02	5.66 ± 0.04
	T9H6	7.12 ± 0.1	7.413 ± 0.001	1.75 ± 0.02	4.86 ± 0.05
	C10H6	7.41 ± 0.06	7.582 ± 0.001	1.74 ± 0.03	4.48 ± 0.06
	C12H6	7.49 ± 0.07	7.67 ± 0.001	1.33 ± 0.02	3.18 ± 0.05
	C2H6	7.36 ± 0.03	7.488 ± 0.001	1.59 ± 0.03	4.64 ± 0.04
	A4H2	7.79 ± 0.07	8.009 ± 0.001	0.76 ± 0.01	2.09 ± 0.04
	A4H8	8.12 ± 0.03	8.212 ± 0.001	1.45 ± 0.02	4.93 ± 0.04
58 ℃	T5H6	6.88 ± 0.1	7.284 ± 0.003	2.91 ± 0.05	16.28 ± 0.28
	A8H2	7.76 ± 0.09	8.04 ± 0.001	0.76 ± 0.01	2.53 ± 0.05
	A8H8	8.04 ± 0.04	8.247 ± 0.001	1.38 ± 0.02	5.29 ± 0.05

(*Continues on the next page...*)

	T9H6	7.12 ± 0.06	7.412 ± 0.001	1.71 ± 0.02	4.47 ± 0.05
58 ℃	C10H6	7.41 ± 0.04	7.584 ± 0.001	1.65 ± 0.03	4.1 ± 0.05
	C12H6	7.49 ± 0.04	7.683 ± 0.001	1.27 ± 0.02	2.81 ± 0.04
	C2H6	7.37 ± 0.02	7.492 ± 0.001	1.56 ± 0.03	5.01 ± 0.06
	A4H2	7.79 ± 0.06	8.009 ± 0.001	0.8 ± 0.01	2.49 ± 0.04
	A4H8	8.12 ± 0.02	8.213 ± 0.001	1.43 ± 0.02	5.44 ± 0.06
	T5H6	6.89 ± 0.08	7.295 ± 0.004	2.4 ± 0.06	14.47 ± 0.24
59 ℃	A8H2	7.76 ± 0.07	8.036 ± 0.001	0.79 ± 0.02	2.93 ± 0.06
	A8H8	8.05 ± 0.02	8.248 ± 0.001	1.36 ± 0.02	5.84 ± 0.06
	T9H6	7.12 ± 0.04	7.41 ± 0.001	1.62 ± 0.02	4.81 ± 0.07
	C10H6	7.41 ± 0.03	7.585 ± 0.001	1.58 ± 0.03	4.39 ± 0.06
	C12H6	7.5 ± 0.04	7.693 ± 0.001	1.24 ± 0.02	2.96 ± 0.06
	C2H6	7.37 ± 0.02	7.499 ± 0.001	1.5 ± 0.02	4.57 ± 0.06
	A4H2	7.8 ± 0.04	8.01 ± 0.001	0.79 ± 0.02	2.52 ± 0.05
	A4H8	8.12 ± 0.02	8.215 ± 0.001	1.41 ± 0.02	5.33 ± 0.07
	T5H6	6.9 ± 0.06	7.296 ± 0.004	1.98 ± 0.06	13.11 ± 0.21
60 ℃	A8H2	7.77 ± 0.05	8.035 ± 0.001	0.77 ± 0.01	2.99 ± 0.06
	A8H8	8.05 ± 0.02	8.249 ± 0.001	1.32 ± 0.02	5.71 ± 0.06
	T9H6	7.13 ± 0.03	7.409 ± 0.002	1.57 ± 0.02	4.67 ± 0.06
	C10H6	7.41 ± 0.02	7.588 ± 0.001	1.53 ± 0.03	4.43 ± 0.06
	C12H6	7.5 ± 0.02	7.702 ± 0.001	1.19 ± 0.02	2.96 ± 0.06

S28

Table S3: Results of the fit of ^{1}H CEST measured at 800 MHz ^{1}H Larmor frequency. The temperature of each measurement is indicated on the leftmost column of the table. Chemical shifts relevant to dsDNA and ssDNA are given together with the relaxation parameters R_1 and R_2 for each available reporter of sample caC$_{4.7}$. Errors are given as one standard deviation.

		Chemical shift		Relaxation rates	
		δ(dsDNA) (ppm)	δ(ssDNA) (ppm)	R_1 (s^{-1})	R_2 (s^{-1})
	C2H6	7.37 ± 0.06	7.496 ± 0.002	1.67 ± 0.03	5.21 ± 0.03
	A4H2	7.79 ± 0.08	8.008 ± 0.	0.82 ± 0.01	2.34 ± 0.03
	A4H8	8.12 ± 0.04	8.219 ± 0.001	1.5 ± 0.02	5.33 ± 0.03
56 ℃	T5H6	6.89 ± 0.07	7.28 ± 0.001	2.22 ± 0.02	13.5 ± 0.06
	A8H2	7.75 ± 0.15	8.066 ± 0.001	0.81 ± 0.01	2.65 ± 0.04
	A8H8	8.02 ± 0.07	8.262 ± 0.001	1.49 ± 0.02	6.4 ± 0.04
	T9H6	7.12 ± 0.07	7.431 ± 0.001	1.81 ± 0.02	5.01 ± 0.04
	C10H6	7.41 ± 0.05	7.606 ± 0.001	1.76 ± 0.03	4.63 ± 0.04
	C2H6	7.37 ± 0.06	7.502 ± 0.002	1.64 ± 0.03	4.97 ± 0.04
	A4H2	7.79 ± 0.09	8.01 ± 0.001	0.81 ± 0.01	2.29 ± 0.04
	A4H8	8.12 ± 0.03	8.22 ± 0.001	1.48 ± 0.02	5.31 ± 0.03
57 ℃	T5H6	6.89 ± 0.07	7.286 ± 0.001	2.22 ± 0.03	13.62 ± 0.06
	A8H2	7.75 ± 0.13	8.06 ± 0.001	0.8 ± 0.02	2.64 ± 0.05
	A8H8	8.03 ± 0.05	8.26 ± 0.001	1.42 ± 0.02	6. ± 0.04
	T9H6	7.12 ± 0.07	7.428 ± 0.001	1.77 ± 0.03	4.87 ± 0.04
	C10H6	7.41 ± 0.05	7.605 ± 0.001	1.76 ± 0.03	4.52 ± 0.04
	C2H6	7.38 ± 0.02	7.507 ± 0.001	1.5 ± 0.02	4.92 ± 0.05
	A4H2	7.79 ± 0.04	8.012 ± 0.001	0.79 ± 0.01	2.49 ± 0.03
	A4H8	8.12 ± 0.01	8.224 ± 0.	1.4 ± 0.02	5.11 ± 0.03
58 ℃	T5H6	6.9 ± 0.03	7.294 ± 0.001	1.83 ± 0.02	13.18 ± 0.08
	A8H2	7.75 ± 0.05	8.057 ± 0.001	0.78 ± 0.01	2.73 ± 0.04
	A8H8	8.02 ± 0.04	8.264 ± 0.001	1.34 ± 0.02	6.24 ± 0.04
	T9H6	7.13 ± 0.03	7.428 ± 0.001	1.65 ± 0.02	4.77 ± 0.04
	C10H6	7.42 ± 0.02	7.609 ± 0.001	1.58 ± 0.02	4.4 ± 0.04

(*Continues on the next page...*)

	C2H6	7.38 ± 0.02	7.514 ± 0.001	1.43 ± 0.02	4.75 ± 0.05
	A4H2	7.8 ± 0.03	8.013 ± 0.001	0.79 ± 0.01	2.74 ± 0.04
	A4H8	8.12 ± 0.01	8.225 ± 0.001	1.36 ± 0.02	4.97 ± 0.03
59 ℃	T5H6	6.91 ± 0.02	7.301 ± 0.001	1.67 ± 0.02	12.85 ± 0.07
	A8H2	7.76 ± 0.04	8.054 ± 0.001	0.77 ± 0.02	2.88 ± 0.04
	A8H8	8.03 ± 0.02	8.262 ± 0.001	1.27 ± 0.02	6.4 ± 0.04
	T9H6	7.14 ± 0.02	7.426 ± 0.001	1.59 ± 0.02	4.44 ± 0.04
	C10H6	7.42 ± 0.02	7.61 ± 0.001	1.49 ± 0.02	4.19 ± 0.04
	C2H6	7.38 ± 0.02	7.518 ± 0.001	1.37 ± 0.03	4.59 ± 0.02
	A4H2	7.8 ± 0.04	8.016 ± 0.001	0.8 ± 0.01	2.97 ± 0.01
	A4H8	8.12 ± 0.01	8.226 ± 0.001	1.32 ± 0.02	4.99 ± 0.01
60 ℃	T5H6	6.92 ± 0.02	7.304 ± 0.002	1.61 ± 0.02	13.16 ± 0.04
	A8H2	7.76 ± 0.04	8.048 ± 0.001	0.76 ± 0.01	3.05 ± 0.01
	A8H8	8.03 ± 0.02	8.26 ± 0.001	1.14 ± 0.02	5.2 ± 0.02
	T9H6	7.14 ± 0.02	7.422 ± 0.002	1.48 ± 0.02	4.27 ± 0.01
	C10H6	7.42 ± 0.02	7.609 ± 0.001	1.37 ± 0.02	3.79 ± 0.02

Table S4: Results of the thermodynamic and kinetic analysis based on of ^{1}H CEST fitted data measured at 800 MHz ^{1}H Larmor frequency. The relevant temperature for each of the parameters' sets is indicated on the leftmost column of the table. The population of dsDNA, the kinetic rates and the dissociation equilibrium constants are given for each available reporter of sample caC$_{7.0}$. Errors are given as one standard deviation.

		dsDNA population	Kinetic rates				Eq. constants
		p_D $(-)$	k_{ex} (s^{-1})	k_d (s^{-1})	k_a $(10^3\ M^{-1}s^{-1})$	k_1 (s^{-1})	K_{eq} $(10^{-6}\ M^{-1})$
56 ℃	C2H6	0.9652 ± 0.0249	28.36 ± 3.58	0.977 ± 0.694	635.849 ± 2014.48	27.387 ± 3.641	2.554 ± 2.693
	A4H2	0.9825 ± 0.0007	13.99 ± 0.88	0.245 ± 0.018	596.901 ± 45.487	13.75 ± 0.868	0.412 ± 0.032
	T5H6	0.9673 ± 0.0009	12.14 ± 1.1	0.397 ± 0.038	272.921 ± 25.826	11.747 ± 1.064	1.457 ± 0.085
	A8H2	0.9845 ± 0.0005	15.27 ± 1.26	0.237 ± 0.021	736.862 ± 67.056	15.03 ± 1.241	0.322 ± 0.023
	A8H8	0.9705 ± 0.0014	9.18 ± 1.26	0.271 ± 0.04	229.485 ± 33.069	8.912 ± 1.226	1.187 ± 0.117
	T9H6	0.9688 ± 0.0011	19.76 ± 1.54	0.617 ± 0.054	465.649 ± 38.815	19.142 ± 1.489	1.328 ± 0.095
	C10H6	0.9687 ± 0.0009	27.21 ± 2.38	0.852 ± 0.08	639.56 ± 58.795	26.356 ± 2.302	1.334 ± 0.082
	C12H6	0.9749 ± 0.0009	24.67 ± 1.67	0.619 ± 0.047	727.494 ± 55.507	24.046 ± 1.626	0.854 ± 0.062
57 ℃	C2H6	0.958 ± 0.0019	26.28 ± 2.61	1.103 ± 0.118	456.048 ± 51.193	25.18 ± 2.503	2.431 ± 0.219
	A4H2	0.9695 ± 0.0114	10.61 ± 2.14	0.325 ± 0.146	333.599 ± 285.054	10.285 ± 2.064	1.446 ± 0.948
	T5H6	0.9559 ± 0.0027	13.45 ± 1.19	0.594 ± 0.068	221.954 ± 22.426	12.86 ± 1.137	2.692 ± 0.334
	A8H2	0.9724 ± 0.0015	10.33 ± 0.9	0.285 ± 0.029	277.255 ± 29.969	10.049 ± 0.879	1.035 ± 0.114
	A8H8	0.9549 ± 0.0045	9.15 ± 2.4	0.412 ± 0.115	148.669 ± 42.321	8.74 ± 2.288	2.831 ± 0.58
	T9H6	0.9595 ± 0.0012	20.5 ± 1.3	0.83 ± 0.061	368.712 ± 24.499	19.669 ± 1.239	2.255 ± 0.136
	C10H6	0.9562 ± 0.0014	24.68 ± 1.48	1.083 ± 0.079	408.743 ± 26.322	23.601 ± 1.415	2.653 ± 0.175
	C12H6	0.9638 ± 0.0009	21.87 ± 1.2	0.792 ± 0.049	442.122 ± 26.453	21.082 ± 1.16	1.793 ± 0.093
58 ℃	C2H6	0.8951 ± 0.1962	25.58 ± 4.8	2.585 ± 5.082	-65.025 ± 2421.53	22.997 ± 7.102	93.137 ± 142.913
	A4H2	0.9622 ± 0.0009	13.91 ± 1.25	0.526 ± 0.048	269.025 ± 25.801	13.387 ± 1.207	1.957 ± 0.1
	T5H6	0.9333 ± 0.0027	12.97 ± 0.82	0.865 ± 0.069	137.934 ± 9.757	12.104 ± 0.76	6.288 ± 0.526
	A8H2	0.9635 ± 0.0011	14.67 ± 0.76	0.536 ± 0.033	293.994 ± 17.539	14.137 ± 0.728	1.826 ± 0.112
	A8H8	0.9457 ± 0.0018	11.92 ± 1.09	0.647 ± 0.062	157.797 ± 15.945	11.276 ± 1.037	4.113 ± 0.284
	T9H6	0.9406 ± 0.0014	20.32 ± 1.26	1.206 ± 0.081	244.39 ± 16.079	19.112 ± 1.187	4.941 ± 0.241
	C10H6	0.9399 ± 0.0016	21.3 ± 1.17	1.281 ± 0.079	252.837 ± 15.681	20.023 ± 1.104	5.075 ± 0.279
	C12H6	0.9513 ± 0.0013	20.25 ± 1.08	0.986 ± 0.06	300.413 ± 17.509	19.265 ± 1.027	3.286 ± 0.175
59 ℃	C2H6	0.917 ± 0.0018	24.11 ± 1.11	2.001 ± 0.102	202.267 ± 10.274	22.112 ± 1.015	9.902 ± 0.441
	A4H2	0.952 ± 0.0015	16.41 ± 0.55	0.787 ± 0.034	247.313 ± 12.106	15.623 ± 0.524	3.191 ± 0.205
	T5H6	0.908 ± 0.0036	20.34 ± 1.32	1.872 ± 0.134	152.582 ± 12.567	18.472 ± 1.21	12.318 ± 1.007
	A8H2	0.9493 ± 0.0017	13.77 ± 0.69	0.698 ± 0.043	195.652 ± 11.632	13.068 ± 0.656	3.577 ± 0.238
	A8H8	0.9267 ± 0.0072	15.92 ± 1.39	1.168 ± 0.158	154.099 ± 20.951	14.749 ± 1.289	7.736 ± 1.554
	T9H6	0.9188 ± 0.0015	22.54 ± 0.78	1.831 ± 0.075	193.472 ± 7.373	20.706 ± 0.718	9.468 ± 0.357

(Continues on the next page...)

59 ℃	C10H6	0.9195 ± 0.0022	24.05 ± 0.98	1.935 ± 0.096	208.583 ± 10.249	22.11 ± 0.898	9.294 ± 0.535
	C12H6	0.9277 ± 0.0013	18.38 ± 0.8	1.329 ± 0.061	178.896 ± 8.83	17.049 ± 0.744	7.438 ± 0.286
60 ℃	C2H6	0.8642 ± 0.0031	21.9 ± 1.04	2.975 ± 0.159	105.764 ± 5.641	18.924 ± 0.902	28.158 ± 1.368
	A4H2	0.9156 ± 0.0026	16.88 ± 0.63	1.425 ± 0.07	138.985 ± 6.919	15.454 ± 0.578	10.277 ± 0.654
	T5H6	0.8522 ± 0.0034	$21.62 \pm 1.$	3.195 ± 0.164	94.702 ± 5.234	18.429 ± 0.859	33.786 ± 1.705
	A8H2	0.9112 ± 0.0021	14.77 ± 0.54	1.311 ± 0.06	114.997 ± 4.952	13.454 ± 0.493	11.416 ± 0.571
	A8H8	0.8741 ± 0.0032	18.34 ± 0.81	2.308 ± 0.109	96.711 ± 5.381	16.031 ± 0.716	23.91 ± 1.292
	T9H6	0.8686 ± 0.0027	22.5 ± 0.89	2.956 ± 0.139	112.928 ± 4.929	19.545 ± 0.77	26.199 ± 1.16
	C10H6	0.868 ± 0.0029	22.96 ± 0.91	3.029 ± 0.134	114.639 ± 5.506	19.927 ± 0.798	26.458 ± 1.254
	C12H6	0.8795 ± 0.0028	20.41 ± 0.65	2.46 ± 0.097	113.092 ± 4.655	17.951 ± 0.57	21.777 ± 1.095
61 ℃	C2H6	0.8345 ± 0.0169	27.62 ± 5.47	4.591 ± 1.13	106.44 ± 22.406	23.024 ± 4.464	43.923 ± 9.729
	A4H2	0.9097 ± 0.0022	25.89 ± 0.56	2.338 ± 0.078	197.974 ± 6.55	23.551 ± 0.513	11.826 ± 0.598
	T5H6	0.8125 ± 0.0023	26.93 ± 0.75	5.05 ± 0.156	88.52 ± 2.756	21.876 ± 0.61	57.066 ± 1.56
	A8H2	0.8974 ± 0.0026	20.9 ± 0.43	2.144 ± 0.065	138.812 ± 5.201	18.754 ± 0.396	15.469 ± 0.819
	A8H8	0.8325 ± 0.0025	20.74 ± 0.5	3.473 ± 0.094	78.231 ± 2.417	17.264 ± 0.42	44.421 ± 1.44
	T9H6	0.8106 ± 0.0029	24.7 ± 0.48	4.678 ± 0.123	80.247 ± 2.036	20.024 ± 0.389	58.325 ± 1.987
	C10H6	0.7795 ± 0.0382	35.77 ± 7.03	7.918 ± 2.21	100.316 ± 35.853	27.853 ± 5.511	86.126 ± 30.785
	C12H6	0.7996 ± 0.0241	26.83 ± 3.43	5.374 ± 0.932	82.729 ± 16.851	21.453 ± 2.85	67.735 ± 18.616

Table S5: Results of the thermodynamic and kinetic analysis based on of ^{1}H CEST fitted data measured at 800 MHz ^{1}H Larmor frequency. The relevant temperature for each of the parameters' sets is indicated on the leftmost column of the table. The population of dsDNA, the kinetic rates and the dissociation equilibrium constants are given for each available reporter of sample caC$_{5.8}$. Errors are given as one standard deviation.

		dsDNA population	Kinetic rates				Eq. constants
		p_D $(-)$	k_{ex} (s^{-1})	k_d (s^{-1})	k_a $(10^3\ M^{-1}s^{-1})$	k_1 (s^{-1})	K_{eq} $(10^{-6}\ M^{-1})$
	C2H6	0.9586 ± 0.0015	25.94 ± 2.35	1.075 ± 0.105	455.759 ± 44.699	24.861 ± 2.25	2.364 ± 0.169
	A4H2	0.9636 ± 0.0026	8.89 ± 1.27	0.322 ± 0.047	180.032 ± 31.536	8.564 ± 1.233	1.818 ± 0.265
	A4H8	0.963 ± 0.0018	$20.71 \pm 2.$	0.767 ± 0.08	409.896 ± 46.152	19.946 ± 1.926	1.882 ± 0.186
	T5H6	0.9498 ± 0.0028	33.6 ± 4.24	1.687 ± 0.244	483.734 ± 63.779	31.914 ± 4.018	3.506 ± 0.401
56 ℃	A8H2	0.974 ± 0.0009	15.23 ± 1.09	0.396 ± 0.031	433.742 ± 35.156	14.832 ± 1.06	0.914 ± 0.062
	A8H8	0.9535 ± 0.0016	11.34 ± 1.25	0.527 ± 0.062	176.776 ± 20.179	10.816 ± 1.187	2.988 ± 0.209
	T9H6	0.9528 ± 0.001	20.02 ± 1.45	0.946 ± 0.072	306.444 ± 23.03	19.076 ± 1.386	3.09 ± 0.127
	C10H6	0.9556 ± 0.001	25.59 ± 1.56	1.137 ± 0.075	417.699 ± 27.049	24.452 ± 1.494	2.725 ± 0.126
	C12H6	0.965 ± 0.0009	25.11 ± 1.74	0.879 ± 0.067	525.351 ± 38.026	24.227 ± 1.68	1.675 ± 0.086
	C2H6	0.9308 ± 0.0019	21.11 ± 1.66	1.46 ± 0.12	215.67 ± 18.415	19.646 ± 1.55	6.782 ± 0.378
	A4H2	0.9559 ± 0.0011	13.19 ± 0.67	0.582 ± 0.035	217.205 ± 11.695	12.613 ± 0.637	2.681 ± 0.14
	A4H8	0.9359 ± 0.0014	14.24 ± 1.3	0.913 ± 0.084	157.915 ± 15.15	13.331 ± 1.216	5.789 ± 0.261
	T5H6	0.9115 ± 0.0033	29.17 ± 1.81	2.58 ± 0.164	228.542 ± 18.855	26.594 ± 1.674	11.338 ± 0.895
57 ℃	A8H2	0.9575 ± 0.0009	15.58 ± 0.58	0.662 ± 0.029	266.515 ± 11.086	14.914 ± 0.552	2.485 ± 0.109
	A8H8	0.9352 ± 0.0017	$12. \pm 0.79$	0.778 ± 0.057	131.466 ± 8.881	11.226 ± 0.733	5.926 ± 0.317
	T9H6	0.9269 ± 0.0016	19.95 ± 0.95	1.459 ± 0.079	192.037 ± 9.941	18.493 ± 0.882	7.602 ± 0.356
	C10H6	0.9335 ± 0.0016	25.57 ± 1.07	1.701 ± 0.079	272.579 ± 13.87	23.874 ± 1.003	6.248 ± 0.309
	C12H6	0.9483 ± 0.0009	22.09 ± 0.81	1.143 ± 0.049	307.314 ± 12.101	20.947 ± 0.769	3.721 ± 0.135
	C2H6	0.9095 ± 0.0024	27.5 ± 1.21	2.489 ± 0.126	209.887 ± 11.376	25.014 ± 1.108	11.877 ± 0.649
	A4H2	0.9478 ± 0.001	19.29 ± 0.46	1.007 ± 0.032	265.775 ± 8.164	18.283 ± 0.439	3.793 ± 0.149
	A4H8	0.9138 ± 0.0023	24.93 ± 1.2	2.149 ± 0.116	200.719 ± 11.692	22.782 ± 1.104	10.724 ± 0.611
	T5H6	0.8789 ± 0.0043	30.35 ± 2.1	3.678 ± 0.304	167.271 ± 12.64	26.674 ± 1.836	22.038 ± 1.689
58 ℃	A8H2	0.9484 ± 0.0009	21.12 ± 0.47	1.089 ± 0.034	294.777 ± 7.476	20.029 ± 0.44	3.696 ± 0.13
	A8H8	0.9198 ± 0.0014	19.81 ± 0.71	1.589 ± 0.061	172.334 ± 7.293	18.219 ± 0.658	9.231 ± 0.337
	T9H6	0.9065 ± 0.0021	26.73 ± 1.01	2.498 ± 0.111	196.801 ± 8.739	24.228 ± 0.915	12.705 ± 0.593
	C10H6	0.9085 ± 0.0019	28.22 ± 0.94	2.582 ± 0.099	212.718 ± 8.872	25.641 ± 0.858	12.153 ± 0.523
	C12H6	0.9276 ± 0.0013	28.43 ± 0.88	2.06 ± 0.071	276.303 ± 10.472	26.372 ± 0.817	7.46 ± 0.284
	C2H6	0.8646 ± 0.0031	20.81 ± 1.18	2.817 ± 0.179	100.874 ± 6.092	17.99 ± 1.017	27.958 ± 1.392
59 ℃	A4H2	0.9212 ± 0.0018	16.96 ± 0.65	1.337 ± 0.059	150.568 ± 6.987	15.627 ± 0.601	8.888 ± 0.425
	A4H8	0.864 ± 0.0042	20.34 ± 1.03	2.766 ± 0.158	98.167 ± 6.298	17.575 ± 0.9	28.252 ± 1.904

(Continues on the next page...)

	T5H6	0.834 ± 0.0057	25.69 ± 1.78	4.267 ± 0.335	98.031 ± 7.735	21.427 ± 1.483	43.647 ± 3.256
	A8H2	0.9132 ± 0.0022	18.9 ± 0.69	1.641 ± 0.076	150.991 ± 6.546	17.261 ± 0.625	10.879 ± 0.572
59 ℃	A8H8	0.8727 ± 0.0015	18.12 ± 0.67	2.307 ± 0.089	94.281 ± 3.729	15.815 ± 0.586	24.476 ± 0.621
	T9H6	0.8601 ± 0.0027	23.13 ± 1.02	3.235 ± 0.162	107.96 ± 5.162	19.893 ± 0.873	29.987 ± 1.267
	C10H6	0.8654 ± 0.0032	23.69 ± 0.96	3.189 ± 0.132	115.648 ± 6.297	20.503 ± 0.855	27.626 ± 1.415
	C12H6	0.8886 ± 0.0025	23.93 ± 0.69	2.665 ± 0.097	144.971 ± 5.468	21.265 ± 0.616	18.401 ± 0.86
	C2H6	0.8346 ± 0.0036	$27.2 \pm 1.$	4.5 ± 0.193	104.195 ± 4.7	22.704 ± 0.843	43.24 ± 2.073
	A4H2	0.9048 ± 0.0019	18.86 ± 0.58	1.795 ± 0.063	136.131 ± 5.405	17.067 ± 0.527	13.2 ± 0.568
	A4H8	0.8202 ± 0.0061	21.31 ± 1.17	3.831 ± 0.257	73.834 ± 4.823	17.474 ± 0.954	52.028 ± 3.937
	T5H6	0.8098 ± 0.0061	27.3 ± 2.47	5.195 ± 0.519	88.278 ± 8.334	22.109 ± 1.99	58.968 ± 4.268
60 ℃	A8H2	0.8922 ± 0.0025	18.88 ± 0.63	2.036 ± 0.083	118.625 ± 4.984	16.847 ± 0.56	17.185 ± 0.861
	A8H8	0.8404 ± 0.0033	20.8 ± 0.93	3.318 ± 0.159	83.174 ± 4.375	17.48 ± 0.787	39.946 ± 1.823
	T9H6	0.8293 ± 0.0035	26.22 ± 0.82	4.476 ± 0.167	96.738 ± 3.863	21.749 ± 0.684	46.319 ± 2.13
	C10H6	0.8351 ± 0.0042	27.91 ± 0.94	4.603 ± 0.191	107.333 ± 4.898	23.31 ± 0.796	42.956 ± 2.396
	C12H6	0.8545 ± 0.0026	24.19 ± 0.87	3.52 ± 0.15	107.825 ± 4.132	20.672 ± 0.736	32.664 ± 1.244

S34

Table S6: Results of the thermodynamic and kinetic analysis based on of ^{1}H CEST fitted data measured at 800 MHz ^{1}H Larmor frequency. The relevant temperature for each of the parameters' sets is indicated on the leftmost column of the table. The population of dsDNA, the kinetic rates and the dissociation equilibrium constants are given for each available reporter of sample caC$_{4.7}$. Errors are given as one standard deviation.

		dsDNA population	Kinetic rates				Eq. constants
		p_D (−)	k_{ex} (s^{-1})	k_d (s^{-1})	k_a ($10^3 \ M^{-1}s^{-1}$)	k_1 (s^{-1})	K_{eq} ($10^{-6} \ M^{-1}$)
56 ℃	C2H6	0.9392 ± 0.0016	24.26 ± 1.78	1.475 ± 0.115	284.68 ± 22.629	22.789 ± 1.677	5.189 ± 0.287
	A4H2	0.9602 ± 0.0006	17.92 ± 0.44	0.713 ± 0.022	328.342 ± 8.772	17.205 ± 0.419	2.171 ± 0.069
	A4H8	0.9462 ± 0.0011	18.42 ± 1.25	0.992 ± 0.068	245.787 ± 17.88	17.431 ± 1.183	4.039 ± 0.167
	T5H6	0.93 ± 0.0008	18.28 ± 0.54	1.28 ± 0.041	184.298 ± 5.944	17. ± 0.506	6.945 ± 0.171
	A8H2	0.9613 ± 0.0007	19.05 ± 0.64	0.736 ± 0.027	359.525 ± 14.06	18.31 ± 0.612	2.049 ± 0.072
	A8H8	0.9474 ± 0.0009	24.13 ± 0.68	1.271 ± 0.044	329.571 ± 10.293	22.862 ± 0.641	3.857 ± 0.135
	T9H6	0.9332 ± 0.001	21.86 ± 0.74	1.46 ± 0.051	231.693 ± 8.922	20.398 ± 0.69	6.307 ± 0.188
	C10H6	0.9348 ± 0.001	24.84 ± 0.96	1.62 ± 0.068	270.121 ± 11.293	23.219 ± 0.9	6.002 ± 0.195
57 ℃	C2H6	0.9261 ± 0.0026	25.92 ± 1.4	1.914 ± 0.12	246.919 ± 16.586	24.003 ± 1.303	7.774 ± 0.554
	A4H2	0.9563 ± 0.0008	22.53 ± 0.63	0.985 ± 0.034	374.083 ± 12.66	21.547 ± 0.603	2.635 ± 0.104
	A4H8	0.9325 ± 0.0013	18.41 ± 1.17	1.243 ± 0.083	193.036 ± 12.685	17.167 ± 1.089	6.441 ± 0.252
	T5H6	0.914 ± 0.0014	20.02 ± 0.82	1.722 ± 0.077	161.46 ± 7.203	18.298 ± 0.754	10.671 ± 0.364
	A8H2	0.9545 ± 0.0009	21.29 ± 0.55	0.968 ± 0.03	339.092 ± 11.859	20.317 ± 0.529	2.858 ± 0.12
	A8H8	0.9379 ± 0.0009	24.8 ± 0.66	1.539 ± 0.047	284.431 ± 8.727	23.258 ± 0.615	5.413 ± 0.164
	T9H6	0.9208 ± 0.0014	23.46 ± 0.97	1.857 ± 0.081	207.039 ± 9.648	21.599 ± 0.895	8.977 ± 0.326
	C10H6	0.9234 ± 0.0015	29.96 ± 1.15	2.294 ± 0.102	274.297 ± 11.575	27.667 ± 1.062	8.368 ± 0.343
58 ℃	C2H6	0.872 ± 0.0022	22.3 ± 0.81	2.856 ± 0.111	115.282 ± 4.869	19.448 ± 0.709	24.792 ± 0.915
	A4H2	0.917 ± 0.0014	21.5 ± 0.32	1.785 ± 0.039	180.22 ± 4.521	19.716 ± 0.301	9.913 ± 0.355
	A4H8	0.8764 ± 0.002	20.67 ± 0.7	2.556 ± 0.102	111.197 ± 4.036	18.116 ± 0.611	22.997 ± 0.814
	T5H6	0.8492 ± 0.002	18.15 ± 0.51	2.737 ± 0.09	77.547 ± 2.308	15.412 ± 0.428	35.309 ± 1.014
	A8H2	0.914 ± 0.0013	22.38 ± 0.45	1.925 ± 0.049	180.554 ± 4.705	20.459 ± 0.41	10.667 ± 0.345
	A8H8	0.8907 ± 0.002	29.11 ± 0.77	3.18 ± 0.107	180.118 ± 5.762	25.927 ± 0.68	17.669 ± 0.699
	T9H6	0.8518 ± 0.0024	24.37 ± 0.57	3.612 ± 0.096	106.331 ± 3.394	20.763 ± 0.497	33.997 ± 1.184
	C10H6	0.8685 ± 0.0021	28.54 ± 0.81	3.754 ± 0.119	143.003 ± 4.943	24.786 ± 0.706	26.271 ± 0.893
59 ℃	C2H6	0.8287 ± 0.0041	26.31 ± 1.01	4.505 ± 0.204	96.638 ± 4.609	21.8 ± 0.841	46.691 ± 2.459
	A4H2	0.8935 ± 0.0017	23.69 ± 0.51	2.523 ± 0.071	150.858 ± 4.054	21.168 ± 0.457	16.734 ± 0.575
	A4H8	0.8191 ± 0.0041	22.98 ± 0.73	4.158 ± 0.147	79.013 ± 3.655	18.826 ± 0.623	52.709 ± 2.665
	T5H6	0.807 ± 0.0024	21.37 ± 0.63	4.125 ± 0.133	67.812 ± 2.227	17.248 ± 0.507	60.86 ± 1.689
	A8H2	0.8761 ± 0.0024	23.26 ± 0.51	2.882 ± 0.085	124.813 ± 3.949	20.374 ± 0.452	23.109 ± 0.977
	A8H8	0.8457 ± 0.0021	27.76 ± 0.76	4.282 ± 0.134	115.458 ± 3.565	23.473 ± 0.644	37.105 ± 1.096

59 ℃	T9H6	0.8005 ± 0.0029	28.46 ± 0.57	5.678 ± 0.144	86.667 ± 2.27	22.782 ± 0.459	65.546 ± 2.127
	C10H6	0.834 ± 0.0031	31.81 ± 1.03	5.28 ± 0.177	121.359 ± 5.283	26.535 ± 0.89	43.561 ± 1.797
60 ℃	C2H6	0.8323 ± 0.0039	31.1 ± 0.86	5.216 ± 0.191	117.175 ± 4.576	25.885 ± 0.724	44.581 ± 2.305
	A4H2	0.8998 ± 0.0018	31.34 ± 0.52	3.14 ± 0.079	213.587 ± 5.433	28.198 ± 0.465	14.713 ± 0.564
	A4H8	0.8052 ± 0.0049	25.08 ± 0.71	4.886 ± 0.17	78.74 ± 3.559	20.198 ± 0.603	62.173 ± 3.523
	T5H6	0.7877 ± 0.0027	24.53 ± 0.59	5.208 ± 0.143	69.072 ± 1.993	19.324 ± 0.467	75.442 ± 2.221
	A8H2	0.8607 ± 0.0029	27.09 ± 0.69	3.774 ± 0.125	127.096 ± 4.477	23.32 ± 0.597	29.724 ± 1.363
	A8H8	0.8508 ± 0.0023	31.44 ± 0.58	4.691 ± 0.111	136.03 ± 3.623	26.746 ± 0.502	34.507 ± 1.172
	T9H6	0.7852 ± 0.0038	29.81 ± 0.51	6.405 ± 0.172	82.693 ± 2.145	23.408 ± 0.396	77.523 ± 3.164
	C10H6	0.8392 ± 0.0031	38.61 ± 0.69	6.21 ± 0.163	152.914 ± 4.437	32.401 ± 0.589	40.652 ± 1.734

S36

Table S7: Thermodynamic and kinetic parameters of the dsDNA melting process obtained from the van't Hoff and Eyring analysis of the CEST-derived exchange parameters. Errors are given as one standard deviation.

		Thermodynamics				Dissociation kinetics			Association kinetics		
		ΔG^0_{37C}	ΔH^0	ΔS^0	T_m (1 mM)	$\Delta G^{\ddagger}_{d,37C}$	$\Delta H^{\ddagger}_d$	$\Delta S^{\ddagger}_d$	$\Delta G^{\ddagger}_{a,37C}$	$\Delta H^{\ddagger}_a$	$\Delta S^{\ddagger}_a$
		kJ mol^{-1}	kJ mol^{-1}	J K^{-1} mol^{-1}	°C	kJ mol^{-1}	kJ mol^{-1}	J K^{-1} mol^{-1}	kJ mol^{-1}	kJ mol^{-1}	J K^{-1} mol^{-1}
C$_{7.0}$	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	68.8 ± 1.5						
caC$_{7.0}$	C2H6	73.3 ± 7.4	669 ± 106	1920 ± 318	65.2 ± 1.	93. ± 4.3	287 ± 64	625 ± 192	19.7 ± 4.1	-382 ± 61	-1295 ± 183
	A4H2	72.4 ± 2.8	607 ± 38	1725 ± 113	67.7 ± 0.4	103.9 ± 1.7	412 ± 24	993 ± 70	31.5 ± 1.7	-195 ± 24	-731 ± 71
	T5H6	75. ± 0.9	696 ± 14	2002 ± 42	64.8 ± 0.1	106.6 ± 1.1	481 ± 16	1208 ± 49	31.6 ± 1.1	-215 ± 16	-794 ± 48
	A8H2	79.4 ± 1.0	715 ± 14	2048 ± 44	66.2 ± 0.2	104.2 ± 1.1	414 ± 16	998 ± 47	24.8 ± 1.1	-301 ± 17	-1050 ± 50
	A8H8	73.6 ± 1.4	663 ± 21	1899 ± 63	65.5 ± 0.2	107.9 ± 2.	485 ± 29	1217 ± 88	34.3 ± 2.1	-177 ± 30	-682 ± 91
	T9H6	75.7 ± 0.8	700 ± 13	2014 ± 38	65. ± 0.1	99.2 ± 0.9	372 ± 13	879 ± 40	23.4 ± 0.9	-328 ± 13	-1134 ± 40
	C10H6	77.3 ± 3.6	727 ± 59	2096 ± 180	64.7 ± 0.6	98.6 ± 3.2	371 ± 51	879 ± 156	21.3 ± 2.9	-356 ± 46	-1217 ± 140
	C12H6	81.8 ± 2.4	785 ± 38	2267 ± 114	64.6 ± 0.3	99.6 ± 1.7	374 ± 26	884 ± 80	17.7 ± 1.9	-411 ± 30	-1382 ± 91
	UV/Vis	52.9 ± 2.9	388 ± 52	1079.8 ± 158	67.7 ± 1.7						
caC$_{5.8}$	C2H6	71.1 ± 1.2	660 ± 19	1899 ± 56	64.2 ± 0.2	94.8 ± 1.4	324 ± 22	739 ± 67	23.7 ± 1.3	-336 ± 19	-1160 ± 58
	A4H2	61.4 ± 1.7	470 ± 26	1316 ± 80	68.8 ± 0.6	101.5 ± 1.8	392 ± 28	937 ± 84	40.1 ± 1.7	-78 ± 26	-379 ± 79
	A4H8	77. ± 1.7	750 ± 26	2170 ± 77	63.6 ± 0.2	99.6 ± 1.8	391 ± 28	939 ± 84	22.6 ± 2.1	-359 ± 32	-1231 ± 97
	T5H6	68.7 ± 2.1	641 ± 32	1844 ± 97	63.8 ± 0.3	88.7 ± 1.9	244 ± 30	501 ± 89	20. ± 2.	-397 ± 32	-1343 ± 95
	A8H2	74.4 ± 1.1	671 ± 17	1923 ± 51	65.6 ± 0.2	100.3 ± 1.2	379 ± 18	897 ± 54	25.9 ± 1.2	-292 ± 18	-1026 ± 55
	A8H8	67.6 ± 1.	602 ± 15	1722 ± 47	65. ± 0.2	102.8 ± 1.4	432 ± 21	1062 ± 64	35.2 ± 1.5	-170 ± 23	-660 ± 69
	T9H6	68. ± 0.9	615 ± 14	1765 ± 43	64.5 ± 0.2	96.6 ± 0.9	352 ± 13	825 ± 40	28.6 ± 1.	-263 ± 15	-940 ± 45
	C10H6	69.7 ± 1.	639 ± 16	1834 ± 50	64.4 ± 0.2	93.6 ± 1.	309 ± 16	694 ± 47	23.9 ± 1.1	-330 ± 17	-1140 ± 51
	C12H6	74.1 ± 0.9	690 ± 14	1985 ± 42	64.6 ± 0.1	95.7 ± 1.	331 ± 16	760 ± 48	21.6 ± 1.	-358 ± 16	-1225 ± 47
caC$_{4.7}$	C2H6	58.6 ± 0.7	483 ± 11	1369 ± 32	65.6 ± 0.2	93.1 ± 0.7	309 ± 11	695 ± 34	34.5 ± 0.9	-175 ± 14	-674 ± 42
	A4H2	58.5 ± 0.5	441 ± 8	1232 ± 24	68.5 ± 0.2	96.5 ± 0.5	339 ± 7	782 ± 22	38. ± 0.4	-102 ± 7	-451 ± 21
	A4H8	64.9 ± 0.7	583 ± 11	1669 ± 34	64.3 ± 0.1	96.6 ± 0.8	357 ± 12	840 ± 37	31.8 ± 0.8	-226 ± 12	-829 ± 35
	T5H6	60. ± 0.4	519 ± 6	1479 ± 19	64.4 ± 0.1	94.2 ± 0.5	324 ± 8	740 ± 24	34.3 ± 0.6	-195 ± 9	-739 ± 26
	A8H2	64.5 ± 0.6	546 ± 9	1552 ± 27	66. ± 0.1	98.1 ± 0.6	368 ± 9	870 ± 29	33.6 ± 0.6	-178 ± 9	-682 ± 27
	A8H8	60.8 ± 0.5	505 ± 8	1433 ± 25	65.8 ± 0.1	92.3 ± 0.5	294 ± 8	650 ± 24	31.5 ± 0.5	-211 ± 8	-783 ± 23
	T9H6	62.1 ± 0.4	550 ± 7	1575 ± 21	64.1 ± 0.1	95.2 ± 0.4	347 ± 7	813 ± 20	33.1 ± 0.5	-203 ± 7	-761 ± 21
	C10H6	56.8 ± 0.6	454 ± 10	1280 ± 29	66.1 ± 0.2	92.9 ± 0.6	312 ± 9	708 ± 26	36.1 ± 0.6	-142 ± 9	-573 ± 28
fC$_{7.0}$	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						

4 | CLOSING REMARKS

4.1 SUMMARY

This work represents a pursuit of the characterization of sparse epigenetically modified DNA cytosine modifications. Despite their eerily similar chemical structure with respect to canonical cytosine, 5mC and its oxidized derivatives have a noticeable impact on DNA. Assignment and chemical shift analysis of low-energy conformers revealed that no permanent detectable change in B-DNA structure is induced by the presence of 5fC or 5caC. On the other hand, an in-depth investigation of the slower- and faster-intermediate regimes has evidenced that these modifications impact the extent to which dsDNA is able to undergo molecular motions, modulating the population size of higher-energy conformers and the kinetic rates characterizing the dynamic equilibrium process.

The work presented in chapter 2 focuses on the effect of the inclusion of 5fC within its naturally occurring settings and conditions in the context of a double-stranded DNA oligomer. We obtained a quantitative evaluation of 5fC-induced destabilization by the means of ^{1}H CEST techniques, ultimately resulting in the identification of ES chemical shifts, and the overall exchange kinetic rates. Further, k_{ex} could be decomposed into k_a and k_d, allowing for the discrimination of the annealing and melting mechanisms, respectively. Eyring-Polanyi and van't Hoff treatment showed that 5fC generates a tenuous degree of destabilization in the duplex state without affecting the transition state architecture.

An extension of the study above is presented in chapter 3, where the effects of 5caC incorporation into canonical DNA strands at multiple pH and temperature conditions is addressed. The investigation was designed to characterize how protonation of the exocyclic carboxyl moiety affects the melting-annealing equilibrium as well as potentially inducing sparse and localized microsecond time scale base-pair dynamics. Applying a similar methodological approach as per our previous investigation, 5caC was considered at three acidic conditions: pH 7.0, 5.8, and 4.7. CEST-derived kinetic and thermodynamic data suggested that

the reduced degree of of cohesion of the 5caC6:G7 base pair, also evidenced by chemical shift studies, affects the extent to which nearby bases are able to cooperatively stabilize one another. An exploration of conformational exchange on the microsecond time scale regime revealed a protonation-triggered, sharply localized base-pair motion involving exclusively the modified site and its immediate surroundings, which could be subsequently identified in the 5fC-containing sample as well.

The overall significance is two-fold. On one hand, our site-specific biophysical characterization of cytosine derivatives by the means of commonly utilized thermodynamic and kinetic frameworks allowed for the extraction of reliable and comparable parameters, which utility might apply, for instance, to the interpretation of enzymatic mechanisms and activity. On the other hand, our results provide additional pieces to the puzzle concerning the specific roles of epigenetically modified DNA bases. It is my conviction that the most relevant dilemmas in biology necessarily require an holistic and interdisciplinary approach, and the work hereby presented was performed with the intention and hope that one day, the larger questions regarding the nature, the function and the finality of epigenetics might be clarified and successfully engineered to the betterment of humankind.

4.2 FUTURE DIRECTIONS

This work specifically focuses on the investigation of natural abundance epigenetically modifed dsDNA oligomers. While our results were also interpreted in the light of current structural biology literature, they are ultimately ill-suited for commenting on the larger scale processes occurring within cells. Below, I list a number of potential perspective studies that might take advantage of the work hereby presented and further advance our understanding of oxidized methylcytosine derivatives.

A central conundrum remains the recognition and activity mechanisms of reader enzymes acting upon dsDNA. While the pinch-push-pull mechanism for base-extrusion and base-cleaving was proposed by Maiti *et al.* in 2013, there is currently little structural evidence supporting the hypothesis. In this respect, one natural extension of this work might be represented by (i) a detailed NMR characterization of TDG enzymes and (ii) the investigation of the interaction between a modified dsDNA strand and and equimolar amount of TDG in solution.

On a different note, a relatively new field within bioNMR spectroscopy is represented by in-cell NMR. This young and fast-developing methodology has gradually emerged as the link between structural and cellular biology, filling the critical gap between in vitro-oriented structural techniques, such as NMR spectroscopy, ultra-high resolution cellular imaging techniques, such as super resolution microscopy and cryo-electron tomography. [101] Nucleosomes, the basic unit of chromatin, help the packaging of genetic material while controlling access to the genetic information. 5fC has been shown to affect nucleosome organization by increasing occupancy *in vitro* and *in vivo*. [102] Incorporation of isotopically labelled 5fC into DNA to be wrapped around a histone might provide clues into their roles and functions.

5 | APPENDIX

5.1 DSDNA CHEMICAL SHIFT ASSIGNMENT

Below are reported the BMRB-compliant chemical shift values for the dsDNA samples used for studies presented in Chapters 2 and 3. The published files, which also include information about experimental details (buffer conditions, software used to perform the assignment and the spectrometers employed to record the spectra) can be reached online at the following links:

1. Publication in Chapter 2, including C- and 5fC-containing dsDNA assignments,

2. Publication in Chapter 3, including 5caC-containing dsDNA assignments.

5.1.1 NMR–STAR assignment of 5caC pH 4.7

```
 save_assigned_chemical_shifts_3
_Assigned_chem_shift_list.Sf_category                 assigned_chemical_shifts
_Assigned_chem_shift_list.Sf_framecode                assigned_chemical_shifts_3
_Assigned_chem_shift_list.Entry_ID                    51337
_Assigned_chem_shift_list.ID                          3
_Assigned_chem_shift_list.Name                        '5caC-containing 12mer dsDNA pH 4.7'
_Assigned_chem_shift_list.Sample_condition_list_ID    1
_Assigned_chem_shift_list.Sample_condition_list_label $sample_conditions_1
_Assigned_chem_shift_list.Chem_shift_reference_ID     1
_Assigned_chem_shift_list.Chem_shift_reference_label  $chem_shift_reference_1
_Assigned_chem_shift_list.Chem_shift_1H_err           .
_Assigned_chem_shift_list.Chem_shift_13C_err          .
_Assigned_chem_shift_list.Chem_shift_15N_err          .
_Assigned_chem_shift_list.Chem_shift_31P_err          .
_Assigned_chem_shift_list.Chem_shift_2H_err           .
_Assigned_chem_shift_list.Chem_shift_19F_err          .
_Assigned_chem_shift_list.Error_derivation_method     .
_Assigned_chem_shift_list.Details                     .
_Assigned_chem_shift_list.Text_data_format            .
```

```
_Assigned_chem_shift_list.Text_data                    .


loop_
   _Chem_shift_experiment.Experiment_ID
   _Chem_shift_experiment.Experiment_name
   _Chem_shift_experiment.Sample_ID
   _Chem_shift_experiment.Sample_label
   _Chem_shift_experiment.Sample_state
   _Chem_shift_experiment.Entry_ID
   _Chem_shift_experiment.Assigned_chem_shift_list_ID


   7   '2D 1H-13C HSQC'   .   .   .   51337   3
   8   '2D 1H-15N HSQC'   .   .   .   51337   3
   9   '2D 1H-1H NOESY'   .   .   .   51337   3
stop_


loop_
   _Chem_shift_software.Software_ID
   _Chem_shift_software.Software_label
   _Chem_shift_software.Method_ID
   _Chem_shift_software.Method_label
   _Chem_shift_software.Entry_ID
   _Chem_shift_software.Assigned_chem_shift_list_ID


   1   $software_1   .   .   51337   3
stop_


loop_
   _Atom_chem_shift.ID
   _Atom_chem_shift.Assembly_atom_ID
   _Atom_chem_shift.Entity_assembly_ID
   _Atom_chem_shift.Entity_assembly_asym_ID
   _Atom_chem_shift.Entity_ID
   _Atom_chem_shift.Comp_index_ID
   _Atom_chem_shift.Seq_ID
   _Atom_chem_shift.Comp_ID
   _Atom_chem_shift.Atom_ID
   _Atom_chem_shift.Atom_type
   _Atom_chem_shift.Atom_isotope_number
   _Atom_chem_shift.Val
   _Atom_chem_shift.Val_err
   _Atom_chem_shift.Assign_fig_of_merit
   _Atom_chem_shift.Ambiguity_code
   _Atom_chem_shift.Ambiguity_set_ID
```

```
_Atom_chem_shift.Occupancy
_Atom_chem_shift.Resonance_ID
_Atom_chem_shift.Auth_entity_assembly_ID
_Atom_chem_shift.Auth_asym_ID
_Atom_chem_shift.Auth_seq_ID
_Atom_chem_shift.Auth_comp_ID
_Atom_chem_shift.Auth_atom_ID
_Atom_chem_shift.Details
_Atom_chem_shift.Entry_ID
_Atom_chem_shift.Assigned_chem_shift_list_ID


1  . 1 . 1 1  1  DG  H1'   H 1   5.968   0.000 . 1 . . . . . 1  G  H1'   . 51337 3
2  . 1 . 1 1  1  DG  H2'   H 1   2.560   0.001 . 1 . . . . . 1  G  H2'   . 51337 3
3  . 1 . 1 1  1  DG  H2''  H 1   2.731   0.002 . 1 . . . . . 1  G  H2''  . 51337 3
4  . 1 . 1 1  1  DG  H3'   H 1   4.824   0.001 . 1 . . . . . 1  G  H3'   . 51337 3
5  . 1 . 1 1  1  DG  H8    H 1   7.921   0.000 . 1 . . . . . 1  G  H8    . 51337 3
6  . 1 . 1 1  1  DG  C8    C 13  136.341 0.000 . 1 . . . . . 1  G  C8    . 51337 3
7  . 1 . 1 2  2  DC  H1'   H 1   5.646   0.000 . 1 . . . . . 2  C  H1'   . 51337 3
8  . 1 . 1 2  2  DC  H2'   H 1   2.037   0.000 . 1 . . . . . 2  C  H2'   . 51337 3
9  . 1 . 1 2  2  DC  H2''  H 1   2.390   0.001 . 1 . . . . . 2  C  H2''  . 51337 3
10 . 1 . 1 2  2  DC  H3'   H 1   4.842   0.001 . 1 . . . . . 2  C  H3'   . 51337 3
11 . 1 . 1 2  2  DC  H5    H 1   5.418   0.001 . 1 . . . . . 2  C  H5    . 51337 3
12 . 1 . 1 2  2  DC  H6    H 1   7.379   0.000 . 1 . . . . . 2  C  H6    . 51337 3
13 . 1 . 1 2  2  DC  H41   H 1   8.422   0.000 . 1 . . . . . 2  C  H41   . 51337 3
14 . 1 . 1 2  2  DC  H42   H 1   6.515   0.000 . 1 . . . . . 2  C  H42   . 51337 3
15 . 1 . 1 2  2  DC  C6    C 13  140.263 0.000 . 1 . . . . . 2  C  C6    . 51337 3
16 . 1 . 1 3  3  DG  H1    H 1   12.706  0.001 . 1 . . . . . 3  G  H1    . 51337 3
17 . 1 . 1 3  3  DG  H1'   H 1   5.642   0.001 . 1 . . . . . 3  G  H1'   . 51337 3
18 . 1 . 1 3  3  DG  H2'   H 1   2.700   0.001 . 1 . . . . . 3  G  H2'   . 51337 3
19 . 1 . 1 3  3  DG  H2''  H 1   2.805   0.001 . 1 . . . . . 3  G  H2''  . 51337 3
20 . 1 . 1 3  3  DG  H3'   H 1   5.010   0.001 . 1 . . . . . 3  G  H3'   . 51337 3
21 . 1 . 1 3  3  DG  H8    H 1   7.885   0.001 . 1 . . . . . 3  G  H8    . 51337 3
22 . 1 . 1 3  3  DG  C8    C 13  135.640 0.000 . 1 . . . . . 3  G  C8    . 51337 3
23 . 1 . 1 3  3  DG  N1    N 15  147.209 0.000 . 1 . . . . . 3  G  N1    . 51337 3
24 . 1 . 1 4  4  DA  H1'   H 1   6.232   0.002 . 1 . . . . . 4  A  H1'   . 51337 3
25 . 1 . 1 4  4  DA  H2    H 1   7.777   0.000 . 1 . . . . . 4  A  H2    . 51337 3
26 . 1 . 1 4  4  DA  H2'   H 1   2.597   0.001 . 1 . . . . . 4  A  H2'   . 51337 3
27 . 1 . 1 4  4  DA  H2''  H 1   2.918   0.001 . 1 . . . . . 4  A  H2''  . 51337 3
28 . 1 . 1 4  4  DA  H3'   H 1   4.984   0.000 . 1 . . . . . 4  A  H3'   . 51337 3
29 . 1 . 1 4  4  DA  H8    H 1   8.147   0.000 . 1 . . . . . 4  A  H8    . 51337 3
30 . 1 . 1 4  4  DA  C8    C 13  138.549 0.000 . 1 . . . . . 4  A  C8    . 51337 3
31 . 1 . 1 5  5  DT  H1'   H 1   5.852   0.002 . 1 . . . . . 5  T  H1'   . 51337 3
32 . 1 . 1 5  5  DT  H2'   H 1   1.960   0.000 . 1 . . . . . 5  T  H2'   . 51337 3
33 . 1 . 1 5  5  DT  H2''  H 1   2.493   0.000 . 1 . . . . . 5  T  H2''  . 51337 3
```

```
34 . 1 . 1 5  5  DT  H3   H 1  13.372  0.000 . 1 . . . . . 5  T   H3   . 51337 3
35 . 1 . 1 5  5  DT  H3'  H 1  4.828   0.001 . 1 . . . . . 5  T   H3'  . 51337 3
36 . 1 . 1 5  5  DT  H6   H 1  6.909   0.000 . 1 . . . . . 5  T   H6   . 51337 3
37 . 1 . 1 5  5  DT  H71  H 1  1.291   0.000 . 1 . . . . . 5  T   H71  . 51337 3
38 . 1 . 1 5  5  DT  H72  H 1  1.291   0.000 . 1 . . . . . 5  T   H72  . 51337 3
39 . 1 . 1 5  5  DT  H73  H 1  1.291   0.000 . 1 . . . . . 5  T   H73  . 51337 3
40 . 1 . 1 5  5  DT  C6   C 13 135.251 0.000 . 1 . . . . . 5  T   C6   . 51337 3
41 . 1 . 1 5  5  DT  N3   N 15 158.863 0.000 . 1 . . . . . 5  T   N3   . 51337 3
42 . 1 . 1 6  6  1CC H1'  H 1  5.586   0.003 . 1 . . . . . 6  caC H1'  . 51337 3
43 . 1 . 1 6  6  1CC H2'  H 1  2.022   0.002 . 1 . . . . . 6  caC H2'  . 51337 3
44 . 1 . 1 6  6  1CC H2'' H 1  2.394   0.001 . 1 . . . . . 6  caC H2'' . 51337 3
45 . 1 . 1 6  6  1CC H3'  H 1  4.803   0.000 . 1 . . . . . 6  caC H3'  . 51337 3
46 . 1 . 1 6  6  1CC H6   H 1  8.040   0.003 . 1 . . . . . 6  caC H6   . 51337 3
47 . 1 . 1 6  6  1CC H41  H 1  8.894   0.000 . 1 . . . . . 6  caC H41  . 51337 3
48 . 1 . 1 6  6  1CC H42  H 1  8.272   0.001 . 1 . . . . . 6  caC H42  . 51337 3
49 . 1 . 1 6  6  1CC C6   C 13 145.920 0.000 . 1 . . . . . 6  caC C6   . 51337 3
50 . 1 . 1 7  7  DG  H1   H 1  12.317  0.003 . 1 . . . . . 7  G   H1   . 51337 3
51 . 1 . 1 7  7  DG  H1'  H 1  5.706   0.001 . 1 . . . . . 7  G   H1'  . 51337 3
52 . 1 . 1 7  7  DG  H2'  H 1  2.679   0.002 . 1 . . . . . 7  G   H2'  . 51337 3
53 . 1 . 1 7  7  DG  H2'' H 1  2.798   0.000 . 1 . . . . . 7  G   H2'' . 51337 3
54 . 1 . 1 7  7  DG  H3'  H 1  4.981   0.001 . 1 . . . . . 7  G   H3'  . 51337 3
55 . 1 . 1 7  7  DG  H8   H 1  7.867   0.001 . 1 . . . . . 7  G   H8   . 51337 3
56 . 1 . 1 7  7  DG  C8   C 13 135.814 0.000 . 1 . . . . . 7  G   C8   . 51337 3
57 . 1 . 1 7  7  DG  N1   N 15 146.438 0.000 . 1 . . . . . 7  G   N1   . 51337 3
58 . 1 . 1 8  8  DA  H1'  H 1  6.183   0.001 . 1 . . . . . 8  A   H1'  . 51337 3
59 . 1 . 1 8  8  DA  H2   H 1  7.730   0.000 . 1 . . . . . 8  A   H2   . 51337 3
60 . 1 . 1 8  8  DA  H2'  H 1  2.530   0.001 . 1 . . . . . 8  A   H2'  . 51337 3
61 . 1 . 1 8  8  DA  H2'' H 1  2.872   0.000 . 1 . . . . . 8  A   H2'' . 51337 3
62 . 1 . 1 8  8  DA  H3'  H 1  4.944   0.001 . 1 . . . . . 8  A   H3'  . 51337 3
63 . 1 . 1 8  8  DA  H8   H 1  8.032   0.001 . 1 . . . . . 8  A   H8   . 51337 3
64 . 1 . 1 8  8  DA  C8   C 13 138.199 0.000 . 1 . . . . . 8  A   C8   . 51337 3
65 . 1 . 1 9  9  DT  H1'  H 1  5.900   0.001 . 1 . . . . . 9  T   H1'  . 51337 3
66 . 1 . 1 9  9  DT  H2'  H 1  2.022   0.000 . 1 . . . . . 9  T   H2'  . 51337 3
67 . 1 . 1 9  9  DT  H2'' H 1  2.423   0.001 . 1 . . . . . 9  T   H2'' . 51337 3
68 . 1 . 1 9  9  DT  H3   H 1  13.500  0.003 . 1 . . . . . 9  T   H3   . 51337 3
69 . 1 . 1 9  9  DT  H3'  H 1  4.830   0.000 . 1 . . . . . 9  T   H3'  . 51337 3
70 . 1 . 1 9  9  DT  H6   H 1  7.142   0.000 . 1 . . . . . 9  T   H6   . 51337 3
71 . 1 . 1 9  9  DT  H71  H 1  1.290   0.000 . 1 . . . . . 9  T   H71  . 51337 3
72 . 1 . 1 9  9  DT  H72  H 1  1.290   0.000 . 1 . . . . . 9  T   H72  . 51337 3
73 . 1 . 1 9  9  DT  H73  H 1  1.290   0.000 . 1 . . . . . 9  T   H73  . 51337 3
74 . 1 . 1 9  9  DT  C6   C 13 135.819 0.000 . 1 . . . . . 9  T   C6   . 51337 3
75 . 1 . 1 9  9  DT  N3   N 15 158.910 0.000 . 1 . . . . . 9  T   N3   . 51337 3
76 . 1 . 1 10 10 DC  H1'  H 1  5.686   0.001 . 1 . . . . . 10 C   H1'  . 51337 3
77 . 1 . 1 10 10 DC  H2'  H 1  2.013   0.000 . 1 . . . . . 10 C   H2'  . 51337 3
```

```
78 . 1 . 1 10 10 DC  H2''  H 1   2.378    0.001 . 1 . . . . . 10 C   H2''  . 51337 3
79 . 1 . 1 10 10 DC  H3'   H 1   4.827    0.002 . 1 . . . . . 10 C   H3'   . 51337 3
80 . 1 . 1 10 10 DC  H5    H 1   5.591    0.001 . 1 . . . . . 10 C   H5    . 51337 3
81 . 1 . 1 10 10 DC  H6    H 1   7.425    0.000 . 1 . . . . . 10 C   H6    . 51337 3
82 . 1 . 1 10 10 DC  H41   H 1   8.383    0.001 . 1 . . . . . 10 C   H41   . 51337 3
83 . 1 . 1 10 10 DC  H42   H 1   6.728    0.002 . 1 . . . . . 10 C   H42   . 51337 3
84 . 1 . 1 10 10 DC  C6    C 13 140.920 0.000 . 1 . . . . . 10 C   C6    . 51337 3
85 . 1 . 1 11 11 DG  H1    H 1  12.971    0.000 . 1 . . . . . 11 G   H1    . 51337 3
86 . 1 . 1 11 11 DG  H1'   H 1   5.962    0.004 . 1 . . . . . 11 G   H1'   . 51337 3
87 . 1 . 1 11 11 DG  H2'   H 1   2.623    0.000 . 1 . . . . . 11 G   H2'   . 51337 3
88 . 1 . 1 11 11 DG  H2''  H 1   2.710    0.002 . 1 . . . . . 11 G   H2''  . 51337 3
89 . 1 . 1 11 11 DG  H3'   H 1   4.970    0.000 . 1 . . . . . 11 G   H3'   . 51337 3
90 . 1 . 1 11 11 DG  H8    H 1   7.905    0.000 . 1 . . . . . 11 G   H8    . 51337 3
91 . 1 . 1 11 11 DG  C8    C 13 135.751 0.000 . 1 . . . . . 11 G   C8    . 51337 3
92 . 1 . 1 11 11 DG  N1    N 15 147.437 0.000 . 1 . . . . . 11 G   N1    . 51337 3
93 . 1 . 1 12 12 DC  H1'   H 1   6.193    0.000 . 1 . . . . . 12 C   H1'   . 51337 3
94 . 1 . 1 12 12 DC  H2'   H 1   2.173    0.002 . 1 . . . . . 12 C   H2'   . 51337 3
95 . 1 . 1 12 12 DC  H2''  H 1   2.213    0.001 . 1 . . . . . 12 C   H2''  . 51337 3
96 . 1 . 1 12 12 DC  H3'   H 1   4.492    0.000 . 1 . . . . . 12 C   H3'   . 51337 3
97 . 1 . 1 12 12 DC  H5    H 1   5.569    0.001 . 1 . . . . . 12 C   H5    . 51337 3
98 . 1 . 1 12 12 DC  H6    H 1   7.507    0.000 . 1 . . . . . 12 C   H6    . 51337 3
99 . 1 . 1 12 12 DC  C6    C 13 140.953 0.000 . 1 . . . . . 12 C   C6    . 51337 3
stop_
save_
```

5.1.2 NMR-STAR assignment of 5caC pH 5.8

```
 save_assigned_chemical_shifts_2
_Assigned_chem_shift_list.Sf_category                    assigned_chemical_shifts
_Assigned_chem_shift_list.Sf_framecode                   assigned_chemical_shifts_2
_Assigned_chem_shift_list.Entry_ID                       51337
_Assigned_chem_shift_list.ID                             2
_Assigned_chem_shift_list.Name                           '5caC-containing 12mer dsDNA pH 5.8'
_Assigned_chem_shift_list.Sample_condition_list_ID       1
_Assigned_chem_shift_list.Sample_condition_list_label    $sample_conditions_1
_Assigned_chem_shift_list.Chem_shift_reference_ID        1
_Assigned_chem_shift_list.Chem_shift_reference_label     $chem_shift_reference_1
_Assigned_chem_shift_list.Chem_shift_1H_err              .
_Assigned_chem_shift_list.Chem_shift_13C_err             .
_Assigned_chem_shift_list.Chem_shift_15N_err             .
_Assigned_chem_shift_list.Chem_shift_31P_err             .
_Assigned_chem_shift_list.Chem_shift_2H_err              .
_Assigned_chem_shift_list.Chem_shift_19F_err             .
_Assigned_chem_shift_list.Error_derivation_method        .
_Assigned_chem_shift_list.Details                        .
_Assigned_chem_shift_list.Text_data_format               .
_Assigned_chem_shift_list.Text_data                      .

loop_
   _Chem_shift_experiment.Experiment_ID
   _Chem_shift_experiment.Experiment_name
   _Chem_shift_experiment.Sample_ID
   _Chem_shift_experiment.Sample_label
   _Chem_shift_experiment.Sample_state
   _Chem_shift_experiment.Entry_ID
   _Chem_shift_experiment.Assigned_chem_shift_list_ID

   4   '2D 1H-13C HSQC'   .   .   .   51337   2
   5   '2D 1H-15N HSQC'   .   .   .   51337   2
   6   '2D 1H-1H NOESY'   .   .   .   51337   2
stop_

loop_
   _Chem_shift_software.Software_ID
   _Chem_shift_software.Software_label
   _Chem_shift_software.Method_ID
   _Chem_shift_software.Method_label
   _Chem_shift_software.Entry_ID
   _Chem_shift_software.Assigned_chem_shift_list_ID
```

```
    1   $software_1   .   .   51337   2
stop_


loop_
    _Atom_chem_shift.ID
    _Atom_chem_shift.Assembly_atom_ID
    _Atom_chem_shift.Entity_assembly_ID
    _Atom_chem_shift.Entity_assembly_asym_ID
    _Atom_chem_shift.Entity_ID
    _Atom_chem_shift.Comp_index_ID
    _Atom_chem_shift.Seq_ID
    _Atom_chem_shift.Comp_ID
    _Atom_chem_shift.Atom_ID
    _Atom_chem_shift.Atom_type
    _Atom_chem_shift.Atom_isotope_number
    _Atom_chem_shift.Val
    _Atom_chem_shift.Val_err
    _Atom_chem_shift.Assign_fig_of_merit
    _Atom_chem_shift.Ambiguity_code
    _Atom_chem_shift.Ambiguity_set_ID
    _Atom_chem_shift.Occupancy
    _Atom_chem_shift.Resonance_ID
    _Atom_chem_shift.Auth_entity_assembly_ID
    _Atom_chem_shift.Auth_asym_ID
    _Atom_chem_shift.Auth_seq_ID
    _Atom_chem_shift.Auth_comp_ID
    _Atom_chem_shift.Auth_atom_ID
    _Atom_chem_shift.Details
    _Atom_chem_shift.Entry_ID
    _Atom_chem_shift.Assigned_chem_shift_list_ID


    1  . 1 . 1 1  1  DG  H1'  H 1  5.969    0.001 . 1 . . . . . 1  G   H1'  . 51337 2
    2  . 1 . 1 1  1  DG  H2'  H 1  2.563    0.001 . 1 . . . . . 1  G   H2'  . 51337 2
    3  . 1 . 1 1  1  DG  H2'' H 1  2.745    0.002 . 1 . . . . . 1  G   H2'' . 51337 2
    4  . 1 . 1 1  1  DG  H3'  H 1  4.832    0.003 . 1 . . . . . 1  G   H3'  . 51337 2
    5  . 1 . 1 1  1  DG  H5'  H 1  3.700    0.000 . 1 . . . . . 1  G   H5'  . 51337 2
    6  . 1 . 1 1  1  DG  H8   H 1  7.928    0.000 . 1 . . . . . 1  G   H8   . 51337 2
    7  . 1 . 1 1  1  DG  C8   C 13 136.266  0.000 . 1 . . . . . 1  G   C8   . 51337 2
    8  . 1 . 1 2  2  DC  H1'  H 1  5.664    0.003 . 1 . . . . . 2  C   H1'  . 51337 2
    9  . 1 . 1 2  2  DC  H2'  H 1  2.029    0.001 . 1 . . . . . 2  C   H2'  . 51337 2
    10 . 1 . 1 2  2  DC  H2'' H 1  2.390    0.001 . 1 . . . . . 2  C   H2'' . 51337 2
    11 . 1 . 1 2  2  DC  H3'  H 1  4.845    0.000 . 1 . . . . . 2  C   H3'  . 51337 2
    12 . 1 . 1 2  2  DC  H5   H 1  5.393    0.000 . 1 . . . . . 2  C   H5   . 51337 2
```

```
13 . 1 . 1 2   2 DC  H6   H 1  7.368   0.001 . 1 . . . . . 2 C   H6   . 51337 2
14 . 1 . 1 2   2 DC  H41  H 1  8.420   0.001 . 1 . . . . . 2 C   H41  . 51337 2
15 . 1 . 1 2   2 DC  H42  H 1  6.464   0.001 . 1 . . . . . 2 C   H42  . 51337 2
16 . 1 . 1 2   2 DC  C6   C 13 140.221 0.000 . 1 . . . . . 2 C   C6   . 51337 2
17 . 1 . 1 3   3 DG  H1   H 1  12.712  0.003 . 1 . . . . 3 G   H1   . 51337 2
18 . 1 . 1 3   3 DG  H1'  H 1  5.646   0.002 . 1 . . . . 3 G   H1'  . 51337 2
19 . 1 . 1 3   3 DG  H2'  H 1  2.696   0.001 . 1 . . . . 3 G   H2'  . 51337 2
20 . 1 . 1 3   3 DG  H2'' H 1  2.808   0.001 . 1 . . . . 3 G   H2'' . 51337 2
21 . 1 . 1 3   3 DG  H3'  H 1  5.010   0.000 . 1 . . . . 3 G   H3'  . 51337 2
22 . 1 . 1 3   3 DG  H8   H 1  7.876   0.006 . 1 . . . . 3 G   H8   . 51337 2
23 . 1 . 1 3   3 DG  C8   C 13 135.640 0.000 . 1 . . . . 3 G   C8   . 51337 2
24 . 1 . 1 3   3 DG  N1   N 15 147.226 0.000 . 1 . . . . 3 G   N1   . 51337 2
25 . 1 . 1 4   4 DA  H1'  H 1  6.228   0.000 . 1 . . . . 4 A   H1'  . 51337 2
26 . 1 . 1 4   4 DA  H2   H 1  7.774   0.000 . 1 . . . . 4 A   H2   . 51337 2
27 . 1 . 1 4   4 DA  H2'  H 1  2.580   0.000 . 1 . . . . 4 A   H2'  . 51337 2
28 . 1 . 1 4   4 DA  H2'' H 1  2.922   0.001 . 1 . . . . 4 A   H2'' . 51337 2
29 . 1 . 1 4   4 DA  H3'  H 1  4.996   0.001 . 1 . . . . 4 A   H3'  . 51337 2
30 . 1 . 1 4   4 DA  H8   H 1  8.152   0.001 . 1 . . . . 4 A   H8   . 51337 2
31 . 1 . 1 4   4 DA  C8   C 13 138.595 0.000 . 1 . . . . 4 A   C8   . 51337 2
32 . 1 . 1 5   5 DT  H1'  H 1  5.849   0.001 . 1 . . . . 5 T   H1'  . 51337 2
33 . 1 . 1 5   5 DT  H2'  H 1  1.941   0.001 . 1 . . . . 5 T   H2'  . 51337 2
34 . 1 . 1 5   5 DT  H2'' H 1  2.476   0.001 . 1 . . . . 5 T   H2'' . 51337 2
35 . 1 . 1 5   5 DT  H3   H 1  13.364  0.002 . 1 . . . . 5 T   H3   . 51337 2
36 . 1 . 1 5   5 DT  H3'  H 1  4.831   0.001 . 1 . . . . 5 T   H3'  . 51337 2
37 . 1 . 1 5   5 DT  H6   H 1  6.881   0.001 . 1 . . . . 5 T   H6   . 51337 2
38 . 1 . 1 5   5 DT  H71  H 1  1.291   0.000 . 1 . . . . 5 T   H71  . 51337 2
39 . 1 . 1 5   5 DT  H72  H 1  1.291   0.000 . 1 . . . . 5 T   H72  . 51337 2
40 . 1 . 1 5   5 DT  H73  H 1  1.291   0.000 . 1 . . . . 5 T   H73  . 51337 2
41 . 1 . 1 5   5 DT  C6   C 13 135.282 0.000 . 1 . . . . 5 T   C6   . 51337 2
42 . 1 . 1 5   5 DT  N3   N 15 158.821 0.000 . 1 . . . . 5 T   N3   . 51337 2
43 . 1 . 1 6   6 1CC H1'  H 1  5.559   0.001 . 1 . . . . 6 caC H1'  . 51337 2
44 . 1 . 1 6   6 1CC H2'  H 1  2.025   0.000 . 1 . . . . 6 caC H2'  . 51337 2
45 . 1 . 1 6   6 1CC H2'' H 1  2.380   0.001 . 1 . . . . 6 caC H2'' . 51337 2
46 . 1 . 1 6   6 1CC H3'  H 1  4.808   0.001 . 1 . . . . 6 caC H3'  . 51337 2
47 . 1 . 1 6   6 1CC H6   H 1  7.990   0.002 . 1 . . . . 6 caC H6   . 51337 2
48 . 1 . 1 6   6 1CC H41  H 1  8.835   0.000 . 1 . . . . 6 caC H41  . 51337 2
49 . 1 . 1 6   6 1CC H42  H 1  8.436   0.001 . 1 . . . . 6 caC H42  . 51337 2
50 . 1 . 1 7   7 DG  H1   H 1  12.381  0.003 . 1 . . . . 7 G   H1   . 51337 2
51 . 1 . 1 7   7 DG  H1'  H 1  5.672   0.001 . 1 . . . . 7 G   H1'  . 51337 2
52 . 1 . 1 7   7 DG  H2'  H 1  2.676   0.002 . 1 . . . . 7 G   H2'  . 51337 2
53 . 1 . 1 7   7 DG  H2'' H 1  2.787   0.000 . 1 . . . . 7 G   H2'' . 51337 2
54 . 1 . 1 7   7 DG  H3'  H 1  4.982   0.000 . 1 . . . . 7 G   H3'  . 51337 2
55 . 1 . 1 7   7 DG  H8   H 1  7.859   0.005 . 1 . . . . 7 G   H8   . 51337 2
56 . 1 . 1 7   7 DG  C8   C 13 135.768 0.000 . 1 . . . . 7 G   C8   . 51337 2
```

```
57 . 1 . 1 7  7  DG N1   N 15 146.584 0.000 . 1 . . . . . 7  G N1   . 51337 2
58 . 1 . 1 8  8  DA H1'  H 1  6.192   0.000 . 1 . . . . . 8  A H1'  . 51337 2
59 . 1 . 1 8  8  DA H2   H 1  7.742   0.000 . 1 . . . . . 8  A H2   . 51337 2
60 . 1 . 1 8  8  DA H2'  H 1  2.544   0.000 . 1 . . . . . 8  A H2'  . 51337 2
61 . 1 . 1 8  8  DA H2'' H 1  2.878   0.001 . 1 . . . . . 8  A H2'' . 51337 2
62 . 1 . 1 8  8  DA H3'  H 1  4.956   0.000 . 1 . . . . . 8  A H3'  . 51337 2
63 . 1 . 1 8  8  DA H8   H 1  8.062   0.001 . 1 . . . . . 8  A H8   . 51337 2
64 . 1 . 1 8  8  DA C8   C 13 138.277 0.000 . 1 . . . . . 8  A C8   . 51337 2
65 . 1 . 1 9  9  DT H1'  H 1  5.894   0.000 . 1 . . . . . 9  T H1'  . 51337 2
66 . 1 . 1 9  9  DT H2'  H 1  2.021   0.000 . 1 . . . . . 9  T H2'  . 51337 2
67 . 1 . 1 9  9  DT H2'' H 1  2.424   0.001 . 1 . . . . . 9  T H2'' . 51337 2
68 . 1 . 1 9  9  DT H3   H 1  13.499  0.003 . 1 . . . . . 9  T H3   . 51337 2
69 . 1 . 1 9  9  DT H3'  H 1  4.829   0.000 . 1 . . . . . 9  T H3'  . 51337 2
70 . 1 . 1 9  9  DT H6   H 1  7.135   0.000 . 1 . . . . . 9  T H6   . 51337 2
71 . 1 . 1 9  9  DT H71  H 1  1.298   0.000 . 1 . . . . . 9  T H71  . 51337 2
72 . 1 . 1 9  9  DT H72  H 1  1.298   0.000 . 1 . . . . . 9  T H72  . 51337 2
73 . 1 . 1 9  9  DT H73  H 1  1.298   0.000 . 1 . . . . . 9  T H73  . 51337 2
74 . 1 . 1 9  9  DT C6   C 13 135.807 0.000 . 1 . . . . . 9  T C6   . 51337 2
75 . 1 . 1 9  9  DT N3   N 15 158.913 0.000 . 1 . . . . . 9  T N3   . 51337 2
76 . 1 . 1 10 10 DC H1'  H 1  5.692   0.001 . 1 . . . . . 10 C H1'  . 51337 2
77 . 1 . 1 10 10 DC H2'  H 1  2.011   0.001 . 1 . . . . . 10 C H2'  . 51337 2
78 . 1 . 1 10 10 DC H2'' H 1  2.378   0.001 . 1 . . . . . 10 C H2'' . 51337 2
79 . 1 . 1 10 10 DC H3'  H 1  4.827   0.000 . 1 . . . . . 10 C H3'  . 51337 2
80 . 1 . 1 10 10 DC H5   H 1  5.587   0.001 . 1 . . . . . 10 C H5   . 51337 2
81 . 1 . 1 10 10 DC H6   H 1  7.419   0.001 . 1 . . . . . 10 C H6   . 51337 2
82 . 1 . 1 10 10 DC H41  H 1  8.382   0.000 . 1 . . . . . 10 C H41  . 51337 2
83 . 1 . 1 10 10 DC H42  H 1  6.708   0.001 . 1 . . . . . 10 C H42  . 51337 2
84 . 1 . 1 10 10 DC C6   C 13 140.934 0.000 . 1 . . . . . 10 C C6   . 51337 2
85 . 1 . 1 11 11 DG H1   H 1  12.995  0.002 . 1 . . . . . 11 G H1   . 51337 2
86 . 1 . 1 11 11 DG H1'  H 1  5.951   0.000 . 1 . . . . . 11 G H1'  . 51337 2
87 . 1 . 1 11 11 DG H2'  H 1  2.611   0.001 . 1 . . . . . 11 G H2'  . 51337 2
88 . 1 . 1 11 11 DG H2'' H 1  2.718   0.001 . 1 . . . . . 11 G H2'' . 51337 2
89 . 1 . 1 11 11 DG H3'  H 1  4.968   0.000 . 1 . . . . . 11 G H3'  . 51337 2
90 . 1 . 1 11 11 DG H8   H 1  7.899   0.000 . 1 . . . . . 11 G H8   . 51337 2
91 . 1 . 1 11 11 DG C8   C 13 135.705 0.000 . 1 . . . . . 11 G C8   . 51337 2
92 . 1 . 1 11 11 DG N1   N 15 147.442 0.000 . 1 . . . . . 11 G N1   . 51337 2
93 . 1 . 1 12 12 DC H1'  H 1  6.188   0.000 . 1 . . . . . 12 C H1'  . 51337 2
94 . 1 . 1 12 12 DC H2'  H 1  2.158   0.000 . 1 . . . . . 12 C H2'  . 51337 2
95 . 1 . 1 12 12 DC H2'' H 1  2.194   0.000 . 1 . . . . . 12 C H2'' . 51337 2
96 . 1 . 1 12 12 DC H3'  H 1  4.492   0.000 . 1 . . . . . 12 C H3'  . 51337 2
97 . 1 . 1 12 12 DC H5   H 1  5.540   0.000 . 1 . . . . . 12 C H5   . 51337 2
98 . 1 . 1 12 12 DC H6   H 1  7.477   0.001 . 1 . . . . . 12 C H6   . 51337 2
99 . 1 . 1 12 12 DC C6   C 13 140.764 0.000 . 1 . . . . . 12 C C6   . 51337 2
stop_
```

save_

5.1.3 NMR-STAR assignment of 5caC pH 7.0

```
 save_assigned_chemical_shifts_1
_Assigned_chem_shift_list.Sf_category                    assigned_chemical_shifts
_Assigned_chem_shift_list.Sf_framecode                   assigned_chemical_shifts_1
_Assigned_chem_shift_list.Entry_ID                       51337
_Assigned_chem_shift_list.ID                             1
_Assigned_chem_shift_list.Name                           '5caC-containing 12mer dsDNA pH 7.0'
_Assigned_chem_shift_list.Sample_condition_list_ID       1
_Assigned_chem_shift_list.Sample_condition_list_label    $sample_conditions_1
_Assigned_chem_shift_list.Chem_shift_reference_ID        1
_Assigned_chem_shift_list.Chem_shift_reference_label     $chem_shift_reference_1
_Assigned_chem_shift_list.Chem_shift_1H_err              .
_Assigned_chem_shift_list.Chem_shift_13C_err             .
_Assigned_chem_shift_list.Chem_shift_15N_err             .
_Assigned_chem_shift_list.Chem_shift_31P_err             .
_Assigned_chem_shift_list.Chem_shift_2H_err              .
_Assigned_chem_shift_list.Chem_shift_19F_err             .
_Assigned_chem_shift_list.Error_derivation_method        .
_Assigned_chem_shift_list.Details                        .
_Assigned_chem_shift_list.Text_data_format               .
_Assigned_chem_shift_list.Text_data                      .

loop_
   _Chem_shift_experiment.Experiment_ID
   _Chem_shift_experiment.Experiment_name
   _Chem_shift_experiment.Sample_ID
   _Chem_shift_experiment.Sample_label
   _Chem_shift_experiment.Sample_state
   _Chem_shift_experiment.Entry_ID
   _Chem_shift_experiment.Assigned_chem_shift_list_ID

   1   '2D 1H-13C HSQC'   .   .   .   51337   1
   2   '2D 1H-15N HSQC'   .   .   .   51337   1
   3   '2D 1H-1H NOESY'   .   .   .   51337   1
stop_

loop_
   _Chem_shift_software.Software_ID
   _Chem_shift_software.Software_label
   _Chem_shift_software.Method_ID
   _Chem_shift_software.Method_label
   _Chem_shift_software.Entry_ID
   _Chem_shift_software.Assigned_chem_shift_list_ID
```

```
   1   $software_1   .   .   51337   1
stop_


loop_
   _Atom_chem_shift.ID
   _Atom_chem_shift.Assembly_atom_ID
   _Atom_chem_shift.Entity_assembly_ID
   _Atom_chem_shift.Entity_assembly_asym_ID
   _Atom_chem_shift.Entity_ID
   _Atom_chem_shift.Comp_index_ID
   _Atom_chem_shift.Seq_ID
   _Atom_chem_shift.Comp_ID
   _Atom_chem_shift.Atom_ID
   _Atom_chem_shift.Atom_type
   _Atom_chem_shift.Atom_isotope_number
   _Atom_chem_shift.Val
   _Atom_chem_shift.Val_err
   _Atom_chem_shift.Assign_fig_of_merit
   _Atom_chem_shift.Ambiguity_code
   _Atom_chem_shift.Ambiguity_set_ID
   _Atom_chem_shift.Occupancy
   _Atom_chem_shift.Resonance_ID
   _Atom_chem_shift.Auth_entity_assembly_ID
   _Atom_chem_shift.Auth_asym_ID
   _Atom_chem_shift.Auth_seq_ID
   _Atom_chem_shift.Auth_comp_ID
   _Atom_chem_shift.Auth_atom_ID
   _Atom_chem_shift.Details
   _Atom_chem_shift.Entry_ID
   _Atom_chem_shift.Assigned_chem_shift_list_ID


   1  . 1 . 1 1  1  DG  H1'  H 1  5.969    0.001 . 1 . . . . . 1  G   H1'  . 51337 1
   2  . 1 . 1 1  1  DG  H2'  H 1  2.565    0.001 . 1 . . . . . 1  G   H2'  . 51337 1
   3  . 1 . 1 1  1  DG  H2'' H 1  2.757    0.001 . 1 . . . . . 1  G   H2'' . 51337 1
   4  . 1 . 1 1  1  DG  H3'  H 1  4.836    0.002 . 1 . . . . . 1  G   H3'  . 51337 1
   5  . 1 . 1 1  1  DG  H5'  H 1  3.704    0.000 . 1 . . . . . 1  G   H5'  . 51337 1
   6  . 1 . 1 1  1  DG  H8   H 1  7.934    0.001 . 1 . . . . . 1  G   H8   . 51337 1
   7  . 1 . 1 1  1  DG  C8   C 13 136.175  0.000 . 1 . . . . . 1  G   C8   . 51337 1
   8  . 1 . 1 2  2  DC  H1'  H 1  5.673    0.001 . 1 . . . . . 2  C   H1'  . 51337 1
   9  . 1 . 1 2  2  DC  H2'  H 1  2.027    0.001 . 1 . . . . . 2  C   H2'  . 51337 1
   10 . 1 . 1 2  2  DC  H2'' H 1  2.393    0.003 . 1 . . . . . 2  C   H2'' . 51337 1
   11 . 1 . 1 2  2  DC  H3'  H 1  4.848    0.001 . 1 . . . . . 2  C   H3'  . 51337 1
   12 . 1 . 1 2  2  DC  H5   H 1  5.381    0.001 . 1 . . . . . 2  C   H5   . 51337 1
```

```
13 . 1 . 1 2  2  DC  H6    H  1  7.365   0.000 . 1 . . . . . 2  C   H6   . 51337 1
14 . 1 . 1 2  2  DC  H41   H  1  8.421   0.000 . 1 . . . . . 2  C   H41  . 51337 1
15 . 1 . 1 2  2  DC  H42   H  1  6.439   0.000 . 1 . . . . . 2  C   H42  . 51337 1
16 . 1 . 1 2  2  DC  C6    C 13 140.163 0.000 . 1 . . . . . 2  C   C6   . 51337 1
17 . 1 . 1 3  3  DG  H1    H  1 12.717   0.002 . 1 . . . . . 3  G   H1   . 51337 1
18 . 1 . 1 3  3  DG  H1'   H  1  5.646   0.003 . 1 . . . . . 3  G   H1'  . 51337 1
19 . 1 . 1 3  3  DG  H2'   H  1  2.693   0.000 . 1 . . . . . 3  G   H2'  . 51337 1
20 . 1 . 1 3  3  DG  H2''  H  1  2.813   0.001 . 1 . . . . . 3  G   H2'' . 51337 1
21 . 1 . 1 3  3  DG  H3'   H  1  5.011   0.000 . 1 . . . . . 3  G   H3'  . 51337 1
22 . 1 . 1 3  3  DG  H8    H  1  7.876   0.001 . 1 . . . . . 3  G   H8   . 51337 1
23 . 1 . 1 3  3  DG  C8    C 13 135.606 0.000 . 1 . . . . . 3  G   C8   . 51337 1
24 . 1 . 1 3  3  DG  N1    N 15 147.219 0.000 . 1 . . . . . 3  G   N1   . 51337 1
25 . 1 . 1 4  4  DA  H1'   H  1  6.228   0.002 . 1 . . . . . 4  A   H1'  . 51337 1
26 . 1 . 1 4  4  DA  H2    H  1  7.774   0.001 . 1 . . . . . 4  A   H2   . 51337 1
27 . 1 . 1 4  4  DA  H2'   H  1  2.567   0.000 . 1 . . . . . 4  A   H2'  . 51337 1
28 . 1 . 1 4  4  DA  H2''  H  1  2.926   0.001 . 1 . . . . . 4  A   H2'' . 51337 1
29 . 1 . 1 4  4  DA  H3'   H  1  5.007   0.001 . 1 . . . . . 4  A   H3'  . 51337 1
30 . 1 . 1 4  4  DA  H8    H  1  8.159   0.001 . 1 . . . . . 4  A   H8   . 51337 1
31 . 1 . 1 4  4  DA  C8    C 13 138.596 0.000 . 1 . . . . . 4  A   C8   . 51337 1
32 . 1 . 1 5  5  DT  H1'   H  1  5.845   0.005 . 1 . . . . . 5  T   H1'  . 51337 1
33 . 1 . 1 5  5  DT  H2'   H  1  1.926   0.001 . 1 . . . . . 5  T   H2'  . 51337 1
34 . 1 . 1 5  5  DT  H2''  H  1  2.466   0.004 . 1 . . . . . 5  T   H2'' . 51337 1
35 . 1 . 1 5  5  DT  H3    H  1 13.360   0.003 . 1 . . . . . 5  T   H3   . 51337 1
36 . 1 . 1 5  5  DT  H3'   H  1  4.837   0.000 . 1 . . . . . 5  T   H3'  . 51337 1
37 . 1 . 1 5  5  DT  H6    H  1  6.860   0.000 . 1 . . . . . 5  T   H6   . 51337 1
38 . 1 . 1 5  5  DT  H71   H  1  1.296   0.000 . 1 . . . . . 5  T   H71  . 51337 1
39 . 1 . 1 5  5  DT  H72   H  1  1.296   0.000 . 1 . . . . . 5  T   H72  . 51337 1
40 . 1 . 1 5  5  DT  H73   H  1  1.296   0.000 . 1 . . . . . 5  T   H73  . 51337 1
41 . 1 . 1 5  5  DT  C6    C 13 135.260 0.000 . 1 . . . . . 5  T   C6   . 51337 1
42 . 1 . 1 5  5  DT  N3    N 15 158.771 0.000 . 1 . . . . . 5  T   N3   . 51337 1
43 . 1 . 1 6  6  1CC H1'   H  1  5.537   0.001 . 1 . . . . . 6  caC H1'  . 51337 1
44 . 1 . 1 6  6  1CC H2'   H  1  2.024   0.002 . 1 . . . . . 6  caC H2'  . 51337 1
45 . 1 . 1 6  6  1CC H2''  H  1  2.373   0.001 . 1 . . . . . 6  caC H2'' . 51337 1
46 . 1 . 1 6  6  1CC H3'   H  1  4.811   0.001 . 1 . . . . . 6  caC H3'  . 51337 1
47 . 1 . 1 6  6  1CC H6    H  1  7.949   0.003 . 1 . . . . . 6  caC H6   . 51337 1
48 . 1 . 1 6  6  1CC H41   H  1  8.785   0.000 . 1 . . . . . 6  caC H41  . 51337 1
49 . 1 . 1 6  6  1CC H42   H  1  8.581   0.001 . 1 . . . . . 6  caC H42  . 51337 1
50 . 1 . 1 7  7  DG  H1    H  1 12.437   0.000 . 1 . . . . . 7  G   H1   . 51337 1
51 . 1 . 1 7  7  DG  H1'   H  1  5.644   0.001 . 1 . . . . . 7  G   H1'  . 51337 1
52 . 1 . 1 7  7  DG  H2'   H  1  2.675   0.002 . 1 . . . . . 7  G   H2'  . 51337 1
53 . 1 . 1 7  7  DG  H2''  H  1  2.777   0.001 . 1 . . . . . 7  G   H2'' . 51337 1
54 . 1 . 1 7  7  DG  H3'   H  1  4.987   0.000 . 1 . . . . . 7  G   H3'  . 51337 1
55 . 1 . 1 7  7  DG  H8    H  1  7.854   0.003 . 1 . . . . . 7  G   H8   . 51337 1
56 . 1 . 1 7  7  DG  C8    C 13 135.782 0.000 . 1 . . . . . 7  G   C8   . 51337 1
```

```
57 . 1 . 1 7  7  DG  N1   N 15 146.746 0.000 . 1 . . . . . 7  G  N1   . 51337 1
58 . 1 . 1 8  8  DA  H1'  H 1  6.199   0.001 . 1 . . . . . 8  A  H1'  . 51337 1
59 . 1 . 1 8  8  DA  H2   H 1  7.754   0.000 . 1 . . . . . 8  A  H2   . 51337 1
60 . 1 . 1 8  8  DA  H2'  H 1  2.558   0.001 . 1 . . . . . 8  A  H2'  . 51337 1
61 . 1 . 1 8  8  DA  H2'' H 1  2.886   0.000 . 1 . . . . . 8  A  H2'' . 51337 1
62 . 1 . 1 8  8  DA  H3'  H 1  4.968   0.001 . 1 . . . . . 8  A  H3'  . 51337 1
63 . 1 . 1 8  8  DA  H8   H 1  8.090   0.000 . 1 . . . . . 8  A  H8   . 51337 1
64 . 1 . 1 8  8  DA  C8   C 13 138.302 0.000 . 1 . . . . . 8  A  C8   . 51337 1
65 . 1 . 1 9  9  DT  H1'  H 1  5.892   0.001 . 1 . . . . . 9  T  H1'  . 51337 1
66 . 1 . 1 9  9  DT  H2'  H 1  2.022   0.001 . 1 . . . . . 9  T  H2'  . 51337 1
67 . 1 . 1 9  9  DT  H2'' H 1  2.426   0.000 . 1 . . . . . 9  T  H2'' . 51337 1
68 . 1 . 1 9  9  DT  H3   H 1  13.498  0.003 . 1 . . . . . 9  T  H3   . 51337 1
69 . 1 . 1 9  9  DT  H3'  H 1  4.829   0.001 . 1 . . . . . 9  T  H3'  . 51337 1
70 . 1 . 1 9  9  DT  H6   H 1  7.131   0.000 . 1 . . . . . 9  T  H6   . 51337 1
71 . 1 . 1 9  9  DT  H71  H 1  1.306   0.000 . 1 . . . . . 9  T  H71  . 51337 1
72 . 1 . 1 9  9  DT  H72  H 1  1.306   0.000 . 1 . . . . . 9  T  H72  . 51337 1
73 . 1 . 1 9  9  DT  H73  H 1  1.306   0.000 . 1 . . . . . 9  T  H73  . 51337 1
74 . 1 . 1 9  9  DT  C6   C 13 135.760 0.000 . 1 . . . . . 9  T  C6   . 51337 1
75 . 1 . 1 9  9  DT  N3   N 15 158.918 0.000 . 1 . . . . . 9  T  N3   . 51337 1
76 . 1 . 1 10 10 DC  H1'  H 1  5.695   0.001 . 1 . . . . . 10 C  H1'  . 51337 1
77 . 1 . 1 10 10 DC  H2'  H 1  2.011   0.002 . 1 . . . . . 10 C  H2'  . 51337 1
78 . 1 . 1 10 10 DC  H2'' H 1  2.381   0.001 . 1 . . . . . 10 C  H2'' . 51337 1
79 . 1 . 1 10 10 DC  H3'  H 1  4.826   0.002 . 1 . . . . . 10 C  H3'  . 51337 1
80 . 1 . 1 10 10 DC  H5   H 1  5.587   0.000 . 1 . . . . . 10 C  H5   . 51337 1
81 . 1 . 1 10 10 DC  H6   H 1  7.417   0.001 . 1 . . . . . 10 C  H6   . 51337 1
82 . 1 . 1 10 10 DC  H41  H 1  8.386   0.000 . 1 . . . . . 10 C  H41  . 51337 1
83 . 1 . 1 10 10 DC  H42  H 1  6.699   0.000 . 1 . . . . . 10 C  H42  . 51337 1
84 . 1 . 1 10 10 DC  C6   C 13 140.901 0.000 . 1 . . . . . 10 C  C6   . 51337 1
85 . 1 . 1 11 11 DG  H1   H 1  13.009  0.001 . 1 . . . . . 11 G  H1   . 51337 1
86 . 1 . 1 11 11 DG  H1'  H 1  5.945   0.000 . 1 . . . . . 11 G  H1'  . 51337 1
87 . 1 . 1 11 11 DG  H2'  H 1  2.604   0.000 . 1 . . . . . 11 G  H2'  . 51337 1
88 . 1 . 1 11 11 DG  H2'' H 1  2.721   0.002 . 1 . . . . . 11 G  H2'' . 51337 1
89 . 1 . 1 11 11 DG  H3'  H 1  4.967   0.000 . 1 . . . . . 11 G  H3'  . 51337 1
90 . 1 . 1 11 11 DG  H8   H 1  7.896   0.000 . 1 . . . . . 11 G  H8   . 51337 1
91 . 1 . 1 11 11 DG  C8   C 13 135.652 0.000 . 1 . . . . . 11 G  C8   . 51337 1
92 . 1 . 1 11 11 DG  N1   N 15 147.442 0.000 . 1 . . . . . 11 G  N1   . 51337 1
93 . 1 . 1 12 12 DC  H1'  H 1  6.185   0.000 . 1 . . . . . 12 C  H1'  . 51337 1
94 . 1 . 1 12 12 DC  H2'  H 1  2.152   0.000 . 1 . . . . . 12 C  H2'  . 51337 1
95 . 1 . 1 12 12 DC  H2'' H 1  2.179   0.000 . 1 . . . . . 12 C  H2'' . 51337 1
96 . 1 . 1 12 12 DC  H3'  H 1  4.492   0.001 . 1 . . . . . 12 C  H3'  . 51337 1
97 . 1 . 1 12 12 DC  H5   H 1  5.523   0.000 . 1 . . . . . 12 C  H5   . 51337 1
98 . 1 . 1 12 12 DC  H6   H 1  7.460   0.001 . 1 . . . . . 12 C  H6   . 51337 1
99 . 1 . 1 12 12 DC  C6   C 13 140.623 0.000 . 1 . . . . . 12 C  C6   . 51337 1
stop_
```

save_

5.1.4 NMR-STAR assignment of 5fC pH 7.0

```
 save_assigned_chemical_shifts_1
_Assigned_chem_shift_list.Sf_category                    assigned_chemical_shifts
_Assigned_chem_shift_list.Sf_framecode                   assigned_chemical_shifts_1
_Assigned_chem_shift_list.Entry_ID                       50278
_Assigned_chem_shift_list.ID                             1
_Assigned_chem_shift_list.Name                           Mod12_assignment_37C_NMR-STAR
_Assigned_chem_shift_list.Sample_condition_list_ID       1
_Assigned_chem_shift_list.Sample_condition_list_label    $sample_conditions_1
_Assigned_chem_shift_list.Chem_shift_reference_ID        1
_Assigned_chem_shift_list.Chem_shift_reference_label     $chem_shift_reference_1
_Assigned_chem_shift_list.Chem_shift_1H_err              .
_Assigned_chem_shift_list.Chem_shift_13C_err             .
_Assigned_chem_shift_list.Chem_shift_15N_err             .
_Assigned_chem_shift_list.Chem_shift_31P_err             .
_Assigned_chem_shift_list.Chem_shift_2H_err              .
_Assigned_chem_shift_list.Chem_shift_19F_err             .
_Assigned_chem_shift_list.Error_derivation_method        .
_Assigned_chem_shift_list.Details                        .
_Assigned_chem_shift_list.Text_data_format               .
_Assigned_chem_shift_list.Text_data                      .


loop_
   _Chem_shift_experiment.Experiment_ID
   _Chem_shift_experiment.Experiment_name
   _Chem_shift_experiment.Sample_ID
   _Chem_shift_experiment.Sample_label
   _Chem_shift_experiment.Sample_state
   _Chem_shift_experiment.Entry_ID
   _Chem_shift_experiment.Assigned_chem_shift_list_ID


   3   '2D 1H-1H NOESY'   .   .   .   50278   1
stop_


loop_
   _Chem_shift_software.Software_ID
   _Chem_shift_software.Software_label
   _Chem_shift_software.Method_ID
   _Chem_shift_software.Method_label
   _Chem_shift_software.Entry_ID
   _Chem_shift_software.Assigned_chem_shift_list_ID


   1   $software_1   .   .   50278   1
```

```
stop_


loop_
    _Atom_chem_shift.ID
    _Atom_chem_shift.Assembly_atom_ID
    _Atom_chem_shift.Entity_assembly_ID
    _Atom_chem_shift.Entity_assembly_asym_ID
    _Atom_chem_shift.Entity_ID
    _Atom_chem_shift.Comp_index_ID
    _Atom_chem_shift.Seq_ID
    _Atom_chem_shift.Comp_ID
    _Atom_chem_shift.Atom_ID
    _Atom_chem_shift.Atom_type
    _Atom_chem_shift.Atom_isotope_number
    _Atom_chem_shift.Val
    _Atom_chem_shift.Val_err
    _Atom_chem_shift.Assign_fig_of_merit
    _Atom_chem_shift.Ambiguity_code
    _Atom_chem_shift.Ambiguity_set_ID
    _Atom_chem_shift.Occupancy
    _Atom_chem_shift.Resonance_ID
    _Atom_chem_shift.Auth_entity_assembly_ID
    _Atom_chem_shift.Auth_asym_ID
    _Atom_chem_shift.Auth_seq_ID
    _Atom_chem_shift.Auth_comp_ID
    _Atom_chem_shift.Auth_atom_ID
    _Atom_chem_shift.Details
    _Atom_chem_shift.Entry_ID
    _Atom_chem_shift.Assigned_chem_shift_list_ID


    1  . 1 . 1 1  1  DG  H1'  H 1 5.968 0.001 . 1 . . . . . 1 G H1'  . 50278 1
    2  . 1 . 1 1  1  DG  H2'  H 1 2.556 0.001 . 1 . . . . . 1 G H2'  . 50278 1
    3  . 1 . 1 1  1  DG  H2'' H 1 2.755 0.001 . 1 . . . . . 1 G H2'' . 50278 1
    4  . 1 . 1 1  1  DG  H4'  H 1 3.697 0.000 . 1 . . . . . 1 G H4'  . 50278 1
    5  . 1 . 1 1  1  DG  H8   H 1 7.933 0.000 . 1 . . . . . 1 G H8   . 50278 1
    6  . 1 . 1 2  2  DC  H1'  H 1 5.674 0.000 . 1 . . . . . 2 C H1'  . 50278 1
    7  . 1 . 1 2  2  DC  H2'  H 1 2.020 0.001 . 1 . . . . . 2 C H2'  . 50278 1
    8  . 1 . 1 2  2  DC  H2'' H 1 2.387 0.002 . 1 . . . . . 2 C H2'' . 50278 1
    9  . 1 . 1 2  2  DC  H4'  H 1 4.133 0.029 . 1 . . . . . 2 C H4'  . 50278 1
    10 . 1 . 1 2  2  DC  H5   H 1 5.374 0.001 . 1 . . . . . 2 C H5   . 50278 1
    11 . 1 . 1 2  2  DC  H6   H 1 7.355 0.000 . 1 . . . . . 2 C H6   . 50278 1
    12 . 1 . 1 2  2  DC  H41  H 1 6.438 0.000 . 1 . . . . . 2 C H41  . 50278 1
    13 . 1 . 1 2  2  DC  H42  H 1 8.416 0.000 . 1 . . . . . 2 C H42  . 50278 1
    14 . 1 . 1 3  3  DG  H1'  H 1 5.657 0.004 . 1 . . . . . 3 G H1'  . 50278 1
```

```
15 . 1 . 1 3  3  DG  H2'   H 1 2.691 0.001 . 1 . . . . . 3 G H2'   . 50278 1
16 . 1 . 1 3  3  DG  H2''  H 1 2.807 0.001 . 1 . . . . . 3 G H2''  . 50278 1
17 . 1 . 1 3  3  DG  H3'   H 1 5.006 0.000 . 1 . . . . . 3 G H3'   . 50278 1
18 . 1 . 1 3  3  DG  H4'   H 1 4.343 0.001 . 1 . . . . . 3 G H4'   . 50278 1
19 . 1 . 1 3  3  DG  H8    H 1 7.869 0.000 . 1 . . . . . 3 G H8    . 50278 1
20 . 1 . 1 4  4  DA  H1'   H 1 6.216 0.001 . 1 . . . . . 4 A H1'   . 50278 1
21 . 1 . 1 4  4  DA  H2    H 1 7.758 0.000 . 1 . . . . . 4 A H2    . 50278 1
22 . 1 . 1 4  4  DA  H2'   H 1 2.588 0.001 . 1 . . . . . 4 A H2'   . 50278 1
23 . 1 . 1 4  4  DA  H2''  H 1 2.884 0.003 . 1 . . . . . 4 A H2''  . 50278 1
24 . 1 . 1 4  4  DA  H3'   H 1 4.987 0.000 . 1 . . . . . 4 A H3'   . 50278 1
25 . 1 . 1 4  4  DA  H4'   H 1 4.232 0.000 . 1 . . . . . 4 A H4'   . 50278 1
26 . 1 . 1 4  4  DA  H8    H 1 8.132 0.000 . 1 . . . . . 4 A H8    . 50278 1
27 . 1 . 1 5  5  DT  H1'   H 1 5.845 0.001 . 1 . . . . . 5 T H1'   . 50278 1
28 . 1 . 1 5  5  DT  H2'   H 1 1.873 0.001 . 1 . . . . . 5 T H2'   . 50278 1
29 . 1 . 1 5  5  DT  H2''  H 1 2.470 0.002 . 1 . . . . . 5 T H2''  . 50278 1
30 . 1 . 1 5  5  DT  H6    H 1 6.966 0.001 . 1 . . . . . 5 T H6    . 50278 1
31 . 1 . 1 5  5  DT  H71   H 1 1.262 0.002 . 1 . . . . . 5 T H7    . 50278 1
32 . 1 . 1 5  5  DT  H72   H 1 1.262 0.002 . 1 . . . . . 5 T H7    . 50278 1
33 . 1 . 1 5  5  DT  H73   H 1 1.262 0.002 . 1 . . . . . 5 T H7    . 50278 1
34 . 1 . 1 6  6  5FC H1'   H 1 5.643 0.001 . 1 . . . . . 6 C H1'   . 50278 1
35 . 1 . 1 6  6  5FC H2'   H 1 2.078 0.000 . 1 . . . . . 6 C H2'   . 50278 1
36 . 1 . 1 6  6  5FC H2''  H 1 2.461 0.002 . 1 . . . . . 6 C H2''  . 50278 1
37 . 1 . 1 6  6  5FC H4'   H 1 4.120 0.000 . 1 . . . . . 6 C H4'   . 50278 1
38 . 1 . 1 6  6  5FC H6    H 1 8.337 0.000 . 1 . . . . . 6 C H6    . 50278 1
39 . 1 . 1 6  6  5FC H7    H 1 9.360 0.000 . 1 . . . . . 6 C H7    . 50278 1
40 . 1 . 1 6  6  5FC H41   H 1 7.857 0.000 . 1 . . . . . 6 C H41   . 50278 1
41 . 1 . 1 6  6  5FC H42   H 1 9.122 0.000 . 1 . . . . . 6 C H42   . 50278 1
42 . 1 . 1 7  7  DG  H1'   H 1 5.699 0.000 . 1 . . . . . 7 G H1'   . 50278 1
43 . 1 . 1 7  7  DG  H2'   H 1 2.677 0.000 . 1 . . . . . 7 G H2'   . 50278 1
44 . 1 . 1 7  7  DG  H2''  H 1 2.800 0.000 . 1 . . . . . 7 G H2''  . 50278 1
45 . 1 . 1 7  7  DG  H3'   H 1 4.985 0.001 . 1 . . . . . 7 G H3'   . 50278 1
46 . 1 . 1 7  7  DG  H4'   H 1 4.451 0.000 . 1 . . . . . 7 G H4'   . 50278 1
47 . 1 . 1 7  7  DG  H8    H 1 7.888 0.000 . 1 . . . . . 7 G H8    . 50278 1
48 . 1 . 1 8  8  DA  H1'   H 1 6.181 0.001 . 1 . . . . . 8 A H1'   . 50278 1
49 . 1 . 1 8  8  DA  H2    H 1 7.708 0.000 . 1 . . . . . 8 A H2    . 50278 1
50 . 1 . 1 8  8  DA  H2'   H 1 2.528 0.000 . 1 . . . . . 8 A H2'   . 50278 1
51 . 1 . 1 8  8  DA  H2''  H 1 2.869 0.000 . 1 . . . . . 8 A H2''  . 50278 1
52 . 1 . 1 8  8  DA  H3'   H 1 4.947 0.001 . 1 . . . . . 8 A H3'   . 50278 1
53 . 1 . 1 8  8  DA  H4'   H 1 4.406 0.000 . 1 . . . . . 8 A H4'   . 50278 1
54 . 1 . 1 8  8  DA  H8    H 1 8.047 0.000 . 1 . . . . . 8 A H8    . 50278 1
55 . 1 . 1 9  9  DT  H1'   H 1 5.881 0.001 . 1 . . . . . 9 T H1'   . 50278 1
56 . 1 . 1 9  9  DT  H2'   H 1 2.027 0.003 . 1 . . . . . 9 T H2'   . 50278 1
57 . 1 . 1 9  9  DT  H2''  H 1 2.423 0.000 . 1 . . . . . 9 T H2''  . 50278 1
58 . 1 . 1 9  9  DT  H4'   H 1 4.153 0.000 . 1 . . . . . 9 T H4'   . 50278 1
```

```
59 . 1 . 1 9  9  DT  H6   H 1 7.135 0.001 . 1 . . . . . 9 T H6   . 50278 1
60 . 1 . 1 9  9  DT  H71  H 1 1.287 0.000 . 1 . . . . . 9 T H7   . 50278 1
61 . 1 . 1 9  9  DT  H72  H 1 1.287 0.000 . 1 . . . . . 9 T H7   . 50278 1
62 . 1 . 1 9  9  DT  H73  H 1 1.287 0.000 . 1 . . . . . 9 T H7   . 50278 1
63 . 1 . 1 10 10 DC  H1'  H 1 5.700 0.000 . 1 . . . . . 10C H1'  . 50278 1
64 . 1 . 1 10 10 DC  H2'  H 1 1.995 0.005 . 1 . . . . . 10C H2'  . 50278 1
65 . 1 . 1 10 10 DC  H2'' H 1 2.373 0.001 . 1 . . . . . 10C H2'' . 50278 1
66 . 1 . 1 10 10 DC  H3'  H 1 4.984 0.000 . 1 . . . . . 10C H3'  . 50278 1
67 . 1 . 1 10 10 DC  H4'  H 1 4.098 0.002 . 1 . . . . . 10C H4'  . 50278 1
68 . 1 . 1 10 10 DC  H5   H 1 5.577 0.001 . 1 . . . . . 10C H5   . 50278 1
69 . 1 . 1 10 10 DC  H6   H 1 7.407 0.001 . 1 . . . . . 10C H6   . 50278 1
70 . 1 . 1 10 10 DC  H41  H 1 6.689 0.001 . 1 . . . . . 10C H41  . 50278 1
71 . 1 . 1 10 10 DC  H42  H 1 8.376 0.001 . 1 . . . . . 10C H42  . 50278 1
72 . 1 . 1 11 11 DG  H1'  H 1 5.940 0.001 . 1 . . . . . 11G H1'  . 50278 1
73 . 1 . 1 11 11 DG  H2'  H 1 2.600 0.000 . 1 . . . . . 11G H2'  . 50278 1
74 . 1 . 1 11 11 DG  H2'' H 1 2.720 0.001 . 1 . . . . . 11G H2'' . 50278 1
75 . 1 . 1 11 11 DG  H3'  H 1 4.964 0.000 . 1 . . . . . 11G H3'  . 50278 1
76 . 1 . 1 11 11 DG  H8   H 1 7.895 0.000 . 1 . . . . . 11G H8   . 50278 1
77 . 1 . 1 12 12 DC  H1'  H 1 6.183 0.001 . 1 . . . . . 12C H1'  . 50278 1
78 . 1 . 1 12 12 DC  H2'  H 1 2.146 0.001 . 1 . . . . . 12C H2'  . 50278 1
79 . 1 . 1 12 12 DC  H2'' H 1 2.183 0.001 . 1 . . . . . 12C H2'' . 50278 1
80 . 1 . 1 12 12 DC  H3'  H 1 4.962 0.000 . 1 . . . . . 12C H3'  . 50278 1
81 . 1 . 1 12 12 DC  H4'  H 1 4.040 0.001 . 1 . . . . . 12C H4'  . 50278 1
82 . 1 . 1 12 12 DC  H5   H 1 5.520 0.000 . 1 . . . . . 12C H5   . 50278 1
83 . 1 . 1 12 12 DC  H6   H 1 7.460 0.001 . 1 . . . . . 12C H6   . 50278 1
   stop_
save_
```

5.1.5 NMR-STAR assignment of C pH 7.0

```
 save_assigned_chemical_shifts_2
_Assigned_chem_shift_list.Sf_category                    assigned_chemical_shifts
_Assigned_chem_shift_list.Sf_framecode                   assigned_chemical_shifts_2
_Assigned_chem_shift_list.Entry_ID                       50278
_Assigned_chem_shift_list.ID                             2
_Assigned_chem_shift_list.Name                           Can12_assignment_37C_NMR-STAR
_Assigned_chem_shift_list.Sample_condition_list_ID       2
_Assigned_chem_shift_list.Sample_condition_list_label    $sample_conditions_2
_Assigned_chem_shift_list.Chem_shift_reference_ID        1
_Assigned_chem_shift_list.Chem_shift_reference_label     $chem_shift_reference_1
_Assigned_chem_shift_list.Chem_shift_1H_err              .
_Assigned_chem_shift_list.Chem_shift_13C_err             .
_Assigned_chem_shift_list.Chem_shift_15N_err             .
_Assigned_chem_shift_list.Chem_shift_31P_err             .
_Assigned_chem_shift_list.Chem_shift_2H_err              .
_Assigned_chem_shift_list.Chem_shift_19F_err             .
_Assigned_chem_shift_list.Error_derivation_method        .
_Assigned_chem_shift_list.Details                        .
_Assigned_chem_shift_list.Text_data_format               .
_Assigned_chem_shift_list.Text_data                      .

loop_
   _Chem_shift_experiment.Experiment_ID
   _Chem_shift_experiment.Experiment_name
   _Chem_shift_experiment.Sample_ID
   _Chem_shift_experiment.Sample_label
   _Chem_shift_experiment.Sample_state
   _Chem_shift_experiment.Entry_ID
   _Chem_shift_experiment.Assigned_chem_shift_list_ID

   4   '2D 1H-1H NOESY'   .   .   .   50278   2
stop_

loop_
   _Chem_shift_software.Software_ID
   _Chem_shift_software.Software_label
   _Chem_shift_software.Method_ID
   _Chem_shift_software.Method_label
   _Chem_shift_software.Entry_ID
   _Chem_shift_software.Assigned_chem_shift_list_ID

   1   $software_1   .   .   50278   2
```

```
stop_


loop_
    _Atom_chem_shift.ID
    _Atom_chem_shift.Assembly_atom_ID
    _Atom_chem_shift.Entity_assembly_ID
    _Atom_chem_shift.Entity_assembly_asym_ID
    _Atom_chem_shift.Entity_ID
    _Atom_chem_shift.Comp_index_ID
    _Atom_chem_shift.Seq_ID
    _Atom_chem_shift.Comp_ID
    _Atom_chem_shift.Atom_ID
    _Atom_chem_shift.Atom_type
    _Atom_chem_shift.Atom_isotope_number
    _Atom_chem_shift.Val
    _Atom_chem_shift.Val_err
    _Atom_chem_shift.Assign_fig_of_merit
    _Atom_chem_shift.Ambiguity_code
    _Atom_chem_shift.Ambiguity_set_ID
    _Atom_chem_shift.Occupancy
    _Atom_chem_shift.Resonance_ID
    _Atom_chem_shift.Auth_entity_assembly_ID
    _Atom_chem_shift.Auth_asym_ID
    _Atom_chem_shift.Auth_seq_ID
    _Atom_chem_shift.Auth_comp_ID
    _Atom_chem_shift.Auth_atom_ID
    _Atom_chem_shift.Details
    _Atom_chem_shift.Entry_ID
    _Atom_chem_shift.Assigned_chem_shift_list_ID


    1  . 2 . 2 1  1  DG H1'  H 1 5.970  0.001 . 1 . . . . . 1  G H1'  . 50278 2
    2  . 2 . 2 1  1  DG H2'  H 1 2.570  0.001 . 1 . . . . . 1  G H2'  . 50278 2
    3  . 2 . 2 1  1  DG H2'' H 1 2.758  0.006 . 1 . . . . . 1  G H2'' . 50278 2
    4  . 2 . 2 1  1  DG H8   H 1 7.937  0.000 . 1 . . . . . 1  G H8   . 50278 2
    5  . 2 . 2 2  2  DC H1'  H 1 5.682  0.001 . 1 . . . . . 2  C H1'  . 50278 2
    6  . 2 . 2 2  2  DC H2'  H 1 2.027  0.001 . 1 . . . . . 2  C H2'  . 50278 2
    7  . 2 . 2 2  2  DC H2'' H 1 2.394  0.001 . 1 . . . . . 2  C H2'' . 50278 2
    8  . 2 . 2 2  2  DC H5   H 1 5.382  0.001 . 1 . . . . . 2  C H5   . 50278 2
    9  . 2 . 2 2  2  DC H6   H 1 7.371  0.000 . 1 . . . . . 2  C H6   . 50278 2
    10 . 2 . 2 2  2  DC H41  H 1 6.441  0.001 . 1 . . . . . 2  C H41  . 50278 2
    11 . 2 . 2 2  2  DC H42  H 1 8.428  0.000 . 1 . . . . . 2  C H42  . 50278 2
    12 . 2 . 2 3  3  DG H1   H 1 12.727 0.000 . 1 . . . . . 3  G H1   . 50278 2
    13 . 2 . 2 3  3  DG H1'  H 1 5.664  0.000 . 1 . . . . . 3  G H1'  . 50278 2
    14 . 2 . 2 3  3  DG H2'  H 1 2.706  0.000 . 1 . . . . . 3  G H2'  . 50278 2
```

```
15 . 2 . 2 3  3  DG H2''  H 1 2.816  0.002 . 1 . . . . . 3  G H2''  . 50278 2
16 . 2 . 2 3  3  DG H3'   H 1 5.017  0.000 . 1 . . . . . 3  G H3'   . 50278 2
17 . 2 . 2 3  3  DG H8    H 1 7.890  0.001 . 1 . . . . . 3  G H8    . 50278 2
18 . 2 . 2 4  4  DA H1'   H 1 6.234  0.001 . 1 . . . . . 4  A H1'   . 50278 2
19 . 2 . 2 4  4  DA H2    H 1 7.777  0.000 . 1 . . . . . 4  A H2    . 50278 2
20 . 2 . 2 4  4  DA H2'   H 1 2.593  0.002 . 1 . . . . . 4  A H2'   . 50278 2
21 . 2 . 2 4  4  DA H2''  H 1 2.926  0.005 . 1 . . . . . 4  A H2''  . 50278 2
22 . 2 . 2 4  4  DA H3'   H 1 4.995  0.000 . 1 . . . . . 4  A H3'   . 50278 2
23 . 2 . 2 4  4  DA H8    H 1 8.154  0.000 . 1 . . . . . 4  A H8    . 50278 2
24 . 2 . 2 5  5  DT H1'   H 1 5.847  0.001 . 1 . . . . . 5  T H1'   . 50278 2
25 . 2 . 2 5  5  DT H2'   H 1 1.980  0.001 . 1 . . . . . 5  T H2'   . 50278 2
26 . 2 . 2 5  5  DT H2''  H 1 2.407  0.004 . 1 . . . . . 5  T H2''  . 50278 2
27 . 2 . 2 5  5  DT H3    H 1 13.452 0.000 . 1 . . . . . 5  T H3    . 50278 2
28 . 2 . 2 5  5  DT H6    H 1 7.089  0.000 . 1 . . . . . 5  T H6    . 50278 2
29 . 2 . 2 5  5  DT H71   H 1 1.318  0.000 . 1 . . . . . 5  T H7    . 50278 2
30 . 2 . 2 5  5  DT H72   H 1 1.318  0.000 . 1 . . . . . 5  T H7    . 50278 2
31 . 2 . 2 5  5  DT H73   H 1 1.318  0.000 . 1 . . . . . 5  T H7    . 50278 2
32 . 2 . 2 6  6  DC H1'   H 1 5.610  0.001 . 1 . . . . . 6  C H1'   . 50278 2
33 . 2 . 2 6  6  DC H2'   H 1 1.918  0.001 . 1 . . . . . 6  C H2'   . 50278 2
34 . 2 . 2 6  6  DC H2''  H 1 2.332  0.001 . 1 . . . . . 6  C H2''  . 50278 2
35 . 2 . 2 6  6  DC H5    H 1 5.514  0.001 . 1 . . . . . 6  C H5    . 50278 2
36 . 2 . 2 6  6  DC H6    H 1 7.352  0.000 . 1 . . . . . 6  C H6    . 50278 2
37 . 2 . 2 6  6  DC H41   H 1 6.584  0.001 . 1 . . . . . 6  C H41   . 50278 2
38 . 2 . 2 6  6  DC H42   H 1 8.302  0.000 . 1 . . . . . 6  C H42   . 50278 2
39 . 2 . 2 7  7  DG H1    H 1 12.524 0.000 . 1 . . . . . 7  G H1    . 50278 2
40 . 2 . 2 7  7  DG H1'   H 1 5.610  0.001 . 1 . . . . . 7  G H1'   . 50278 2
41 . 2 . 2 7  7  DG H2'   H 1 2.675  0.001 . 1 . . . . . 7  G H2'   . 50278 2
42 . 2 . 2 7  7  DG H2''  H 1 2.780  0.002 . 1 . . . . . 7  G H2''  . 50278 2
43 . 2 . 2 7  7  DG H3'   H 1 4.986  0.000 . 1 . . . . . 7  G H3'   . 50278 2
44 . 2 . 2 7  7  DG H8    H 1 7.858  0.001 . 1 . . . . . 7  G H8    . 50278 2
45 . 2 . 2 8  8  DA H1'   H 1 6.194  0.001 . 1 . . . . . 8  A H1'   . 50278 2
46 . 2 . 2 8  8  DA H2    H 1 7.738  0.001 . 1 . . . . . 8  A H2    . 50278 2
47 . 2 . 2 8  8  DA H2'   H 1 2.573  0.007 . 1 . . . . . 8  A H2'   . 50278 2
48 . 2 . 2 8  8  DA H2''  H 1 2.898  0.002 . 1 . . . . . 8  A H2''  . 50278 2
49 . 2 . 2 8  8  DA H3'   H 1 4.968  0.000 . 1 . . . . . 8  A H3'   . 50278 2
50 . 2 . 2 8  8  DA H8    H 1 8.117  0.000 . 1 . . . . . 8  A H8    . 50278 2
51 . 2 . 2 9  9  DT H1'   H 1 5.871  0.001 . 1 . . . . . 9  T H1'   . 50278 2
52 . 2 . 2 9  9  DT H2'   H 1 2.011  0.002 . 1 . . . . . 9  T H2'   . 50278 2
53 . 2 . 2 9  9  DT H2''  H 1 2.431  0.005 . 1 . . . . . 9  T H2''  . 50278 2
54 . 2 . 2 9  9  DT H3    H 1 13.463 0.000 . 1 . . . . . 9  T H3    . 50278 2
55 . 2 . 2 9  9  DT H6    H 1 7.112  0.001 . 1 . . . . . 9  T H6    . 50278 2
56 . 2 . 2 9  9  DT H71   H 1 1.311  0.000 . 1 . . . . . 9  T H7    . 50278 2
57 . 2 . 2 9  9  DT H72   H 1 1.311  0.000 . 1 . . . . . 9  T H7    . 50278 2
58 . 2 . 2 9  9  DT H73   H 1 1.311  0.000 . 1 . . . . . 9  T H7    . 50278 2
```

```
59 . 2 . 2 10 10 DC H1'   H 1 5.707  0.001 . 1 . . . . . 10 C H1'   . 50278 2
60 . 2 . 2 10 10 DC H2'   H 1 2.001  0.001 . 1 . . . . . 10 C H2'   . 50278 2
61 . 2 . 2 10 10 DC H2''  H 1 2.375  0.003 . 1 . . . . . 10 C H2''  . 50278 2
62 . 2 . 2 10 10 DC H5    H 1 5.579  0.000 . 1 . . . . . 10 C H5    . 50278 2
63 . 2 . 2 10 10 DC H6    H 1 7.411  0.000 . 1 . . . . . 10 C H6    . 50278 2
64 . 2 . 2 10 10 DC H41   H 1 6.700  0.001 . 1 . . . . . 10 C H41   . 50278 2
65 . 2 . 2 10 10 DC H42   H 1 8.389  0.001 . 1 . . . . . 10 C H42   . 50278 2
66 . 2 . 2 11 11 DG H1    H 1 13.013 0.000 . 1 . . . . . 11 G H1    . 50278 2
67 . 2 . 2 11 11 DG H1'   H 1 5.944  0.001 . 1 . . . . . 11 G H1'   . 50278 2
68 . 2 . 2 11 11 DG H2'   H 1 2.608  0.002 . 1 . . . . . 11 G H2'   . 50278 2
69 . 2 . 2 11 11 DG H2''  H 1 2.722  0.003 . 1 . . . . . 11 G H2''  . 50278 2
70 . 2 . 2 11 11 DG H3'   H 1 4.968  0.000 . 1 . . . . . 11 G H3'   . 50278 2
71 . 2 . 2 11 11 DG H8    H 1 7.897  0.000 . 1 . . . . . 11 G H8    . 50278 2
72 . 2 . 2 12 12 DC H1'   H 1 6.186  0.001 . 1 . . . . . 12 C H1'   . 50278 2
73 . 2 . 2 12 12 DC H2'   H 1 2.152  0.001 . 1 . . . . . 12 C H2'   . 50278 2
74 . 2 . 2 12 12 DC H2''  H 1 2.187  0.000 . 1 . . . . . 12 C H2''  . 50278 2
75 . 2 . 2 12 12 DC H5    H 1 5.519  0.001 . 1 . . . . . 12 C H5    . 50278 2
76 . 2 . 2 12 12 DC H6    H 1 7.458  0.001 . 1 . . . . . 12 C H6    . 50278 2
   stop_

save_
```

BIBLIOGRAPHY

1. Dahm, R. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum. Genet.* **122,** 565–581 (2008).

2. Lorenz, M. G. & Wackernagel, W. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev.* **58,** 563–602 (1994).

3. Astbury, W. T. & Bell, F. O. Some Recent Developments in the X-Ray Study of Proteins and Related Structures. *Cold Spring Harb. Symp. Quant. Biol.* **6,** 109–121 (1938).

4. Crick, F. & Watson, J. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171,** 737–738 (1953).

5. Forsdyke, D. R. & Mortimer, J. R. Chargaff's legacy. *Gene* **261,** 127–137 (2000).

6. Prabhu, V. V. Symmetry observations in long nucleotide sequences. *Nucleic Acids Res.* **21,** 2797–2800 (1993).

7. Neidle, S. *Principles of Nucleic Acid Structure* (Academic press, New York, USA, 2008).

8. Lee, J., Kim, Y. G., Kim, K. K. & Seok, C. Transition between B-DNA and Z-DNA: Free energy landscape for the B-Z junction propagation. *J. Phys. Chem. B* **114,** 9872–9881 (2010).

9. Sung, C. H., Lowenhaupt, K., Rich, A., Kim, Y. G. & Kyeong, K. K. Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature* **437,** 1183–1186 (2005).

10. 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. *J. Phys. Chem. B* **124,** 1160–1174 (2020).

11. Ashwood, B., Sanstead, P. J., Dai, Q., He, C. & Tokmakoff, A. 5-Carboxylcytosine and Cytosine Protonation Distinctly Alter the Stability and Dehybridization Dynamics of the DNA Duplex. *J. Phys. Chem. B* **124,** 627–640 (2020).

12. Ngo, T. T. *et al.* Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* **7,** 1–9 (2016).

13. Szulik, M. W., Voehler, M. & Stone, M. P. NMR analysis of base-pair opening kinetics in DNA. *Curr. Protoc. Nucleic Acid Chem.* **59,** 1–18 (2014).

14. Coman, D. & Russu, I. M. A Nuclear Magnetic Resonance Investigation of the Energetics of Basepair Opening Pathways in DNA. *Biophys. J.* **89,** 3285–3292 (2005).

15. Nikolova, E. N. *et al.* Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470,** 498–502 (2011).

16. Liu, B., Rangadurai, A., Shi, H. & Al-Hashimi, H. M. Rapid assessment of Watson–Crick to Hoogsteen exchange in unlabeled DNA duplexes using high-power SELOPE imino 1H CEST. *Magn. Reson.* **2,** 715–731 (2021).

17. Lankas, F. DNA sequence-dependent deformability-insights from computer simulations. *Biopolymers* **73,** 327–339 (2003).

18. Lavelle, C. Pack, unpack, bend, twist, pull, push: The physical side of gene expression. *Curr. Opin. Genet. Dev.* **25,** 74–84 (2014).

19. Baranello, L., Levens, D., Gupta, A. & Kouzine, F. The importance of being supercoiled: How DNA mechanics regulate dynamic processes. *Biochim. Biophys. Acta* **1819,** 632–638 (2012).

20. Von Hippel, P. H., Johnson, N. P. & Marcus, A. H. Fifty years of DNA "breathing": Reflections on old and new approaches. *Biopolymers* **99,** 923–954 (2013).

21. Fei, J. & Ha, T. Watching DNA breath one molecule at a time. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 17173–17174 (2013).

22. Felsenfeld, G. & Miles, H. T. The physical and chemical properties of nucleic acids. *Annu. Rev. Biochem.* **36,** 407–448 (1967).

23. Pörschke, D., Uhlenbeck, O. C. & Martin, F. H. Thermodynamics and kinetics of the helix-coil transition of oligomers containing GC base pairs. *Biopolymers* **12,** 1313–1335 (1973).

24. Wetmur, J. G. & Davidson, N. Kinetics of renaturation of DNA. *J. Mol. Biol.* **31,** 349–370 (1968).

25. Crothers, D. M. The kinetics of DNA denaturation. *J. Mol. Biol.* **9,** 712–733 (1964).

26. Vologodskii, A. & Frank-Kamenetskii, M. D. DNA melting and energetics of the double helix. *Phys. Life Rev.* **25,** 1–21 (2018).

27. Ouldridge, T. E., Šulc, P., Romano, F., Doye, J. P. K. & Louis, A. A. DNA hybridization kinetics : zippering , internal displacement and sequence dependence. *Nucleic Acids Res.* **41,** 8886–8895 (2013).

28. Zgarbová, M., Otyepka, M., Šponer, J., Lankaš, F. & Jurečka, P. Base pair fraying in molecular dynamics simulations of DNA and RNA. *J. Chem. Theory Comput.* **10,** 3177–3189 (2014).

29. Mergny, J. L. & Lacroix, L. Analysis of Thermal Melting Curves. *Oligonucleotides* **13,** 515–537 (2003).

30. Choi, S.-R. *et al.* Base-pair Opening Dynamics of Nucleic Acids in Relation to Their Biological Function. *Comput. Struct. Biotechnol. J.* **17,** 797–804 (2019).

31. Roy, N. S. *et al.* Enhanced basepair dynamics pre-disposes protein-assisted flips of key bases in DNA strand separation during transcription initiation. *Phys. Chem. Chem. Phys.* **20,** 9449–9459 (2018).

32. Colizzi, F. *et al.* Asymmetric base-pair opening drives helicase unwinding dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **116,** 22471–22477 (2019).

33. Hoogsteen, K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **16,** 907–916 (1963).

34. Aishima, J. *et al.* A Hoogsteen base pair embedded in undistorted B-DNA. *Nucleic Acids Res.* **30,** 5244–5252 (2002).

35. Segers-Nolten, G. M., Sijtsema, N. M. & Otto, C. Evidence for Hoogsteen GC base pairs in the proton-induced transition from right-handed to left-handed poly(dG-dC)·poly(dG-dC). *Biochemistry* **36,** 13241–13247 (1997).

36. Patikoglou, G. A. *et al.* TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* **13,** 3217–3230 (1999).

37. Kitayner, M. *et al.* Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nat. Struct. Mol. Biol.* **17,** 423–429 (2010).

38. Peixoto, P., Cartron, P. F., Serandour, A. A. & Hervouet, E. From 1957 to nowadays: A brief history of epigenetics. *Int. J. Mol. Sci.* **21,** 1–18 (2020).

39. Zhao, L. Y., Song, J., Liu, Y., Song, C. X. & Yi, C. Mapping the epigenetic modifications of DNA and RNA. *Protein Cell* **11,** 792–808 (2020).

40. Yang, W. & Gao, Y. Translesion and Repair DNA Polymerases: Diverse Structure and Mechanism. *Annu. Rev. Biochem.* **87,** 239–261 (2018).

41. Bonasio, R., Tu, S. & Reinberg, D. Molecular signals of epigenetic states. *Science* **330,** 612–616 (2010).

42. Lyko, F. The DNA methyltransferase family: A versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.* **19,** 81–92 (2018).

43. Chen, Z. *et al.* Role of Mammalian DNA Methyltransferases in Development. *Annu. Rev. Biochem.* **89,** 135–158 (2020).

44. Du, Q., Wang, Z. & Schramm, V. L. Human DNMT1 transition state structure. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 2916–2921 (2016).

45. Liang Wang Xingang Li, D. T. & Wang, W. *Genome Plasticity in Health and Disease* (Elsevier, Amsterdam, NE, 2020).

46. Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **18,** 517–534 (2017).

47. Bochtler, M., Kolano, A. & Xu, G. L. DNA demethylation pathways: Additional players and regulators. *BioEssays* **39,** 1–13 (2017).

48. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502,** 472–479 (2013).

49. Ooi, S. K. & Bestor, T. H. The Colorful History of Active DNA Demethylation. *Cell* **133,** 1145–1148 (2008).

50. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935 (2009).

51. Ito, S. *et al.* Role of tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466,** 1129–1133 (2010).

52. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* **5,** 1–9 (2010).

53. Song, C. X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* **29,** 68–75 (2011).

54. Ito, S. *et al.* Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* **333,** 1300–1304 (2011).

55. Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146,** 67–79 (2011).

56. Bennett, M. T. *et al.* Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *J. Am. Chem. Soc.* **128,** 12510–12519 (2006).

57. Zhang, L. *et al.* Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.* **8,** 328–330 (2012).

58. Song, C. X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153,** 678–691 (2013).

59. Song, C. X., Yi, C. & He, C. Mapping recently identified nucleotide variants in the genome and transcriptome. *Nat. Biotechnol.* **30,** 1107–1116 (2012).

60. Shen, L. *et al.* Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153,** 692–706 (2013).

61. Su, M. *et al.* 5-Formylcytosine Could Be a Semipermanent Base in Specific Genome Sites. *Angew. Chem. Int. Ed.* **55,** 11797–11800 (2016).

62. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14,** 1–11 (2013).

63. Spruijt, C. G. *et al.* Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell* **152,** 1146–1159 (2013).

64. Kellinger, M. W. *et al.* 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **19,** 831–833 (2012).

65. Raiber, E. A. *et al.* 5-formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **22,** 44–49 (2015).

66. Hardwick, J. S. *et al.* 5-Formylcytosine does not change the global structure of DNA. *Nat. Struct. Mol. Biol.* **24,** 544–552 (2017).

67. Szulik, M. W. *et al.* 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 (2015).

68. Fu, T. *et al.* Thymine DNA glycosylase recognizes the geometry alteration of minor grooves induced by 5-formylcytosine and 5-carboxylcytosine. *Chem. Sci.* **10,** 7407–7417 (2019).

69. Dai, Q. *et al.* Weakened N3 Hydrogen Bonding by 5-Formylcytosine and 5-Carboxylcytosine Reduces Their Base-Pairing Stability. *ACS Chem. Biol.* **11,** 470–477 (2016).

70. Furukawa, A., Walinda, E., Arita, K. & Sugase, K. Structural dynamics of double-stranded DNA with epigenome modification. *Nucleic Acids Res.* **49,** 1152–1162 (2021).

71. Gardner, K. H. & Kay, L. E. The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **27,** 357–406 (1998).

72. Moser, E., Laistler, E., Schmitt, F. & Kontaxis, G. Ultra-high field NMR and MRI-the role of magnet technology to increase sensitivity and specificity. *Front. Phys.* **5,** 1–15 (2017).

73. Kay, L. E. NMR studies of protein structure and dynamics. *J. Magn. Reson.* **173,** 193–207 (2005).

74. Vasa, S. K., Rovó, P. & Linser, R. Protons as Versatile Reporters in Solid-State NMR Spectroscopy. *Acc. Chem. Res.* **51,** 1386–1395 (2018).

75. Zidek, L., Stefl, R. & Sklenar, V. NMR methodology for the study of nucleic acids. *Curr. Opin. Struct. Biol.* **11,** 275–281 (2001).

76. Levitt, M. H. *Spin dynamics: basics of nuclear magnetic resonance* (Academic Press, New York, USA, 2001).

77. Keeler, J. *Understanding NMR Spectroscopy* (University of Cambridge, Cambridge, UK, 2002).

78. Rule, G. S. & Hitchens, T. K. *Fundamentals of protein NMR spectroscopy* (Springer, Dordrecht, NE, 2006).

79. Hasha, D. L., Eguchi, T. & Jonas, J. Dynamical effects on conformational isomerization of cyclohexane. *J. Chem. Phys.* **75,** 1571–1573 (1981).

80. Dayie, T. K., Olenginski, L. T. & Taiwo, K. M. Isotope Labels Combined with Solution NMR Spectroscopy Make Visible the Invisible Conformations of Small-to-Large RNAs. *Chem. Rev.* (2022).

81. Bouvignies, G., Vallurupalli, P. & Kay, L. E. Visualizing side chains of invisible protein conformers by solution NMR. *J. Mol. Biol.* **426,** 763–774 (2014).

82. Scheek, R. M., Russo, N., Boelens, R., Kaptein, R. & Boom, V. Sequential resonance assignments in DNA proton NMR spectra by two-dimensional NOE spectroscopy. *J. Am. Chem. Soc.* **9,** 2914–2916 (1983).

83. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

84. Schanda, P. & Brutscher, B. Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds. *J. Am. Chem. Soc.* **127,** 8014–8015 (2005).

85. Nerli, S., McShan, A. C. & Sgourakis, N. G. Chemical shift-based methods in NMR structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **106-107,** 1–25 (2018).

86. Dubini, R. C. A. *et al.* 1 H NMR Chemical Exchange Techniques Reveal Local and Global Effects of Oxidized Cytosine Derivatives. *ACS Phys. Chem. Au* (2022).

87. Rangadurai, A., Szymanski, E. S., Kimsey, I., Shi, H. & Al-Hashimi, H. M. Probing conformational transitions towards mutagenic Watson–Crick-like G·T mismatches using off-resonance sugar carbon $R_{1\rho}$ relaxation dispersion. *J. Biomol. NMR* **74,** 457–471 (2020).

88. Mureddu, L. & Vuister, G. W. Simple high-resolution NMR spectroscopy as a tool in molecular biology. *FEBS J.* **286,** 2035–2042 (2019).

89. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73,** 1–16 (2013).

90. Marušič, M., Schlagnitweit, J. & Petzold, K. RNA Dynamics by NMR Spectroscopy. *ChemBioChem* **20,** 2685–2710 (2019).

91. Jarymowycz, V. A. & Stone, M. J. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem. Rev.* **106,** 1624–1671 (2006).

92. Vallurupalli, P., Sekhar, A., Yuwen, T. & Kay, L. E. Probing conformational dynamics in biomolecules via chemical exchange saturation transfer: a primer. *J. Biomol. NMR* **67,** 243–271 (2017).

93. Forsén, S. & Huffman, R. A. Study of moderately rapid chemical exchange reactions by means of nuclear magnetic double resonance. *J. Chem. Phys.* **39,** 2892–2901 (1963).

94. Vallurupalli, P., Bouvignies, G. & Kay, L. E. Studying "Invisible" Excited Protein States in Slow Exchange with a Major State Conformation. *J. Am. Chem. Soc.* **134,** 8148–8161 (2012).

95. Ward, K. M., Aletras, A. H. & Balaban, R. S. A New Class of Contrast Agents for MRI Based on Proton Chemical Exchange Dependent Saturation Transfer (CEST). *J. Magn. Reson.* **143,** 79–87 (2000).

96. McConnell, H. M. Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.* **28,** 430–431 (1958).

97. Mulder, F. A. A., Mittermaier, A., Hon, B., Dahlquist, F. W. & Kay, L. E. Studying excited states of proteins by NMR. *Nat. Struct. Mol. Biol.* **8,** 932–935 (2001).

98. Lee, J., Dethoff, E. A. & Al-Hashimi, H. M. Invisible RNA state dynamically couples distant motifs. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 9485–9490 (2014).

99. Hoogstraten, C. G., Wank, J. R. & Pardi, A. Active Site Dynamics in the Lead-Dependent Ribozyme. *Biochemistry* **39,** 9951–9958 (2000).

100. Miloushev, V. Z. & Palmer, A. G. R1ρ relaxation for two-site chemical exchange: General approximations and some exact solutions. *J. Magn. Reson.* **177,** 221–227 (2005).

101. Luchinat, E. & Banci, L. In-cell NMR: A topical review. *IUCrJ* **4,** 108–118 (2017).

102. Raiber, E. A. *et al.* 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse. *Nat. Chem* **10,** 1258–1266 (2018).