PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2023, **25**, 241

5-Formylcytosine weakens the G–C pair and imparts local conformational fluctuations to DNA duplexes†

Manjula Jaisal,‡ Rajesh Kumar Reddy Sannapureddi,‡ Arjun Rana and Bharathwaj Sathyamoorthy *

DNA epigenetic modifications such as 5-methyl (^{5m}C), 5-hydroxymethyl (^{5hm}C), 5-formyl (^{5f}C) and 5-carboxyl (^{5ca}C) cytosine have unique and specific biological roles. Crystallographic studies of ^{5m}C containing duplexes were conducted in the A-, B- or the intermediate E-DNA polymorphic forms. ^{5f}C-modified duplexes initially observed in the disputed F-DNA architecture were subsequently crystallized in the A-form, suggesting that epigenetic modifications enable DNA sequences to adopt diverse conformational states that plausibly contribute to their function. Solution-state studies of these modifications were found in the B-DNA form, with marked differences in the conformational flexibility of ^{5f}C containing duplexes in comparison to C/^{5m}C containing duplexes, compromising the DNA duplex's stability. Herein, we systematically evaluate sensitive and commonly inaccessible NMR parameters to map the subtle differences between C, ^{5m}C, and their oxidized (^{5hm}C/^{5f}C) counterparts. We observe that ^{15}N/^{1}H chemical shifts effectively report on the weakening of ^{5f}C–G Watson–Crick base-pair H-bonding, extending the instability beyond any achievable within the sequence-specific changes in DNA. Triple ^{5f}C containing sequences propagate the destabilization farther from the site of modifications, explaining reduced duplex stability upon multiple modifications. Additionally, scalar and residual dipolar coupling measurements unravel local sugar pucker fluctuations. One-bond ^{13}C–^{1}H scalar coupling measurements point towards a significant deviation away from the anticipated C2′-*endo* pucker for the ^{5f}C modified nucleotide. Structural models obtained employing ^{13}C–^{1}H residual dipolar couplings and inter-proton distances corroborate the sugar pucker's deviation for ^{5f}C modified DNA duplexes. The changes in the sugar pucker equilibria remain local to the ^{5f}C modified nucleotide sans additive/long-range effects arising from multiple contiguous modifications. These observations highlight the impact of a major groove modification that alters the physical properties of DNA duplex without disturbing the Watson–Crick face. The changes observed in our studies for the ^{5f}C containing DNA contrast with the perturbations induced by damage/lesion highlight the varied conformational preferences that modified nucleobases impart to the DNA duplex. As sequence-specific DNA transactions are rooted in the base-pair stability and pucker deviations, the observed structural perturbations for ^{5f}C-modified DNA potentially play critical functional roles, such as protein-DNA recognition and interactions.

Received 16th October 2022,
Accepted 4th December 2022

DOI: 10.1039/d2cp04837j

rsc.li/pccp

Introduction

DNA methyltransferases robustly incorporate and maintain the epigenetic cytosine modifications in CpG dinucleotide steps.[1–3] Methylation at the 5th position of cytosine (5-methylcytosine, ^{5m}C, Fig. 1A) is the most common epigenetic marker in DNA, with ^{5m}C being regarded as the 5th abundant base in the genome.[4–8] ^{5m}C modified sites are recognized to play myriad roles in cells.[9–16] Ten-eleven translocation enzymes sequentially oxidize ^{5m}C to 5-hydroxymethylcytosine[17] (^{5hm}C), 5-formylcytosine[17] (^{5f}C), and 5-carboxylcytosine[18] (^{5ca}C), with thymidine DNA glycosylase and base excision repair enzymes providing a pathway towards demethylation of ^{5m}C[19,20] (Fig. 1B). Furthermore, each of these oxidized counterparts is increasingly identified to be semi-permanent, not just intermediates, and perform a wide range of unique, tissue-specific, and functional roles[21–23] in, including but not limited to, genome packaging,[24–26] gene expression,[27,28] replication modulation,[29]

Department of Chemistry, Indian Institute of Science Education and Research, Bhopal 462066, India. E-mail: bharathwaj@iiserb.ac.in

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2cp04837j

‡ Contributed equally to this work.

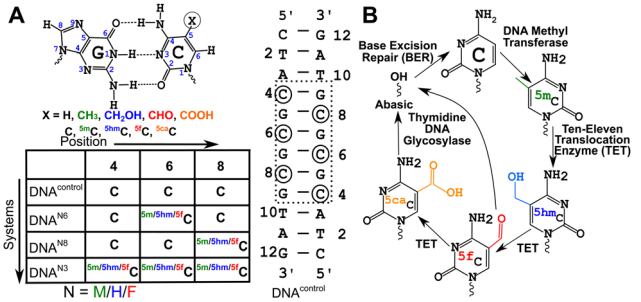

Fig. 1 (A) Chemical structure of the cytosine-guanosine Watson–Crick pair with characteristic hydrogen (H-)bonds. Epigenetic modifications in cytosine are induced by changing the functional group in the 5th position. The palindromic dodecamer duplex DNA sequence ($DNA^{control}$, (5′-CTACGCGCGTAG-3′)$_2$) studied in this work along with the suitable modifications ($DNA^{N\#}$, with N = M/H/F for $^{5m}C/^{5hm}C/^{5f}C$ and # = 8/6/3 depending upon the type of modification, see Experimental methods) is introduced. Changes are introduced in the CpG repeat "core" of the duplex to avoid end-fraying conformational dynamics. (B) Methylation and subsequent demethylation are carried out by enzymes that convert C to ^{5m}C, then to ^{5hm}C, ^{5f}C, and ^{5ca}C completing the cycle for the cytosine epigenetic modification.

mutability of neighboring nucleotides,[30] embryo development[31] and prognosis of cancer.[32] The structure–function paradigm of molecular biology thus motivates detailed biophysical characterization of these modifications.

Cytosine modifications in the major groove retain the conventional Watson–Crick hydrogen (H-)bonding pattern (Fig. 1A). X-ray crystallographic studies of singly hemi-modified $^{5m}C/^{5hm}C/^{5f}C$ in the CpG step of palindromic Drew-Dickerson dodecamer duplex DNA (5′-CGCGAATTNGCG-3′, referred to as DDD^N, N = $^{5m}C/^{5hm}C/^{5f}C$ modification, G indicates the N-G pair) showed minimal perturbation from the B-DNA architecture.[33–35] ^{5m}C incorporated in a G-C base-pair rich palindromic hexamer d(5′-GG^{5m}CGCC-3′)$_2$ was crystallized in an intermediate E-DNA form with bases being perpendicular to the helical axis (B-form like) while the sugars sample an A-form like the C3′-*endo* pucker.[36] The metastable E-DNA eventually equilibrates under crystallographic conditions to the A-DNA form.[36] On the other hand, the triply ^{5f}C modified palindromic dodecamer sequence (5′-CTA^{5f}CG^{5f}CG^{5f}CGTAG-3′, referred henceforth as DNA^{F3}, Fig. 1A) was crystallized in a form that alters the hydration pattern stabilizing propeller twist and base-pair opening parameters, that appeared to differ significantly from A- and B-DNA forms, and hence led to a newly proposed class of architecture called the F-DNA.[37] Such an observation correlated with differences in the circular dichroism (CD) signatures of DNA^{F3} compared to the unmodified DNA ($DNA^{control}$, Fig. 1A), in line with *in silico* modeling that predicts that the helical under-winding traps water molecules stabilizing the proposed F-DNA form.[38] However, a subsequent study showed that structures of both DNA^{F3} and $DNA^{control}$ sample the A-DNA form with no significant differences in the spatial arrangement of heavy atoms.[39] Previously reported differences in CD signatures between DNA^{F3} and $DNA^{control}$ were attributed to potential changes in the local electronic transition dipole moment rather than due to global structural perturbations of the DNA duplex.[39] Hence, the next question follows whether the structure observed in the crystal form would be retained or be any different in the solution-state conditions.

Solution-state ^{1}H-based nuclear magnetic resonance (NMR) studies of DNA^{F3} substantiated that the ^{5f}C modification maintained the B-DNA form, as adjudged from the inter-proton distance and ^{1}H–^{1}H scalar coupling measurements.[39] Interestingly, this study hinted at a deviation from the C2′-*endo* pucker only for the ^{5f}C-modified nucleotides. Imino ^{1}H-exchange NMR experiments performed on hemi-modified DDD^N (N = $^{5m}C/^{5hm}C/^{5f}C$) samples showed increased base-pair opening rates for ^{5f}C compared to the unmodified duplex suggesting subtle differences in their conformational landscape.[35] Single-molecule fluorescence-based DNA cyclization assays revealed that ^{5f}C modification imparts enhanced flexibility compared to unmodified cytosine-containing duplexes, while ^{5m}C rigidifies the duplex.[40] Steady-state and time-resolved infrared spectroscopy showed that ^{5f}C in DNA^{F3} increases base-pair fluctuations reducing the cooperativity of duplex formation and thereby increasing the double-strand dissociation rate constant.[41] The weakening of the duplex was attributed to the reduced pK_a of the N3 nitrogen atom in 5-formyl modified cytosine that accepts the proton from the pairing guanosine nucleobase (Fig. 1A).[42,43] Recently, solution-state ^{1}H-based relaxation dispersion measurements have demonstrated an increase in the population of the single-stranded form for the ^{5f}C containing DNA duplex[44] (5′-GCGAT^{5f}CGATCGC-3′). Additionally, it was reported that the destabilization propagates across the DNA duplex beyond the single ^{5f}C–G fully modified base-pair. These observations suggest that ^{5f}C modification might not alter the structure as much in comparison to cytosine or ^{5m}C, but may interfere with the conformational fluctuations due to its unique chemical properties.

While the effect of a single site modification has been characterized, the influence of multiple contiguous modifications on DNA duplex structure is yet to be explored. Additionally, cytosine nucleotides are known to exhibit enhanced sugar puckering dynamics in comparison to other canonical nucleotides catering towards sequence-specific recognition.[45,46] Therefore, a question arises whether these modifications alter such specific conformational dynamics of DNA duplexes, and whether can there be more NMR probes for measuring the same. Also, we sought to compare the destabilization/fluctuations achieved by the ^{5f}C–G pair to what is achievable within the canonical C–G framework without modifications by only altering the primary sequence. In this study, we present NMR probes to understand the effect of single and multiple cytosine modifications (^{5m}C, ^{5hm}C, and ^{5f}C) on the global structure and dynamics of DNA duplexes using solution-state NMR spectroscopy. Additionally, using these parameters we probe the presence/absence of differential sugar puckering of ^{5f}C-containing duplexes.

Heteronuclear $^{13}C/^{15}N$ chemical shifts,[47–50] scalar couplings,[51,52] and partial anisotropic parameters, such as residual dipolar couplings[53–57] (RDCs), are sensitive in characterizing conformational

properties of DNA duplexes.[58] RDCs provide a relative orientation of bonds across the molecule and thus improve the global structure of DNA duplexes, that otherwise evade conventional characterization that employs inter-proton distances and ^{1}H–^{1}H scalar coupling measurements. The structural perturbations employing RDCs for duplexes have been well characterized for DNA comprising of A-tracts,[55] nucleotides with a locked sugar pucker,[56] and *N*1-methyladenine[57] (^{m1}A) modification. In particular, the damage modification ^{m1}A present in duplexes results in bending of the helical axis and contributes to local base-pair melting suggesting a pre-primed bent DNA for effective protein recognition toward damage repair.[57]

In this work, we employ an optimized sparse sampling methodology that reduces overall measurement times of two-dimensional NMR data by 75%, thus making it possible to measure heteronuclear (^{13}C/^{15}N) shifts and RDCs robustly at low concentration (~100 μM) in natural isotopic abundance samples (ESI†). Application of the optimized methods reveals that ^{15}N imino chemical shifts of the paired guanosine are sensitive to the weakening of the H-bond for ^{5f}C modified duplexes in comparison to DNA^{control}. The triply ^{5f}C modified sample (DNA^{F3}, Fig. 1A) shows a weakening of H-bonds farther than the singly modified samples (DNA$^{F6/F8}$, Fig. 1A) indicating propagation of base-pair destabilization. At the same time, no discernable effect is observed for the ^{5m}C/^{5hm}C analog. One-bond ^{13}C–^{1}H scalar coupling ($^{1}J_{CH}$) measurements for sugar C1′–H1′ bonds point towards deviation from the C2′-*endo* pucker confined to the ^{5f}C modified nucleotides. Structural models, obtained by employing inter-proton distances and one-bond ^{13}C–^{1}H residual dipolar couplings (RDCs, $^{1}D_{CH}$), indicate that ^{5f}C modified nucleotides' sugar moiety samples conformations away from the C2′-*endo* pucker, while C, ^{5m}C, and ^{5hm}C containing DNA duplexes do not display any appreciable excursions. Such sugar pucker perturbations are localized to ^{5f}C modified sites, with no additive effect arising from multiple modifications next to each other. The results highlight the impact that conformational changes due to ^{5f}C incorporation may potentially have on protein–DNA recognition.

Results

^{15}N/^{1}H chemical shifts indicate a weakened ^{5f}C–G H-bond beyond all possible sequence-specific contexts

^{13}C/^{15}N chemical shifts are NMR parameters that provide the necessary resolution to alleviate any chemical shift degeneracy in the ^{1}H dimension and contain critical structural information, such as the presence and strength of H-bonds, and changes to the sugar pucker and glycosyl dihedral angle.[47–50] Chemical shifts are perturbed by subtle changes in atomistic/molecular interactions, such as changes in H-bonding and/or π–π stacking.[49,50] To delineate chemical shift perturbations (CSP) that arise in the modified duplexes due to changes in H-bonding and ring current effects, single "fully" modified (DNA^{N6}, Fig. 1A) and single "hemi" modified (DNA^{N8}, Fig. 1A) samples were studied and compared with the control (DNA^{control}, 5′-CTACGCGCGTAG-3′, Fig. 1A). CSPs observed in the paired G5 for the hemi-modified DNA^{N8} samples (with C8 being modified with ^{5m}C/^{5hm}C/^{5f}C, Fig. 1A) would provide the change solely due to H-bonding, while the CSPs of G7 (5′-neighbor of C8, Fig. 1A) and G9 (3′-neighbor of C8, Fig. 1A) indicate the changes due to stacking/ring current effects for ^{5m}C/^{5hm}C/^{5f}C in comparison to unmodified cytosine. On the other hand, the G7 CSP from a single fully modified DNA^{N6} (Fig. 1A) would reflect the effect due to both H-bonding and ring current effects. Any differences in CSP observed in DNA^{N6} *versus* DNA^{N8} (Fig. 1A) would thus aid in pointing at the effect of hemi- *vs.* fully modified systems. Importantly, differences in CSPs measured from DNA^{N3} *versus* DNA^{N6} (Fig. 1A) would provide insights into potential long-range perturbations due to multiple contiguous modifications.

Firstly, the G5 imino chemical shift (associated with C8–G5 pairing) in DNA$^{M8/H8/F8}$ was examined to probe the influence of modifications solely on the base pairing. G5–N1/H1 resonances shift upfield by ~0.8/0.4 ppm and ~0.3/0.1 ppm for ^{5f}C and ^{5hm}C, respectively, in comparison to unmodified C, while ^{5m}C shows marginal downfield shifts of 0.05/0.05 ppm (Fig. 2A, B and Table S1, Fig. S3, ESI†). It is evident that amongst the C–G pairs, modification with ^{5f}C tends to shift both G–N1/H1 resonances significantly in contrast to the control and the other epigenetic modifications. The electron donating/withdrawing characteristics of the CH_3, CH_2OH, and CHO functional groups present in modified cytosine are correlated to the direction of the imino ^{1}H CSP. A chemical modification on the C alters the C[N3]–G[N1] H-bond distance which in turn causes deshielding/shielding of the G–N1/H1 spins affecting CSP relative to the unmodified cytosine.[48,59] The longer (shorter) the hydrogen bond, the higher (lower) the (de)shielding of the imino group.

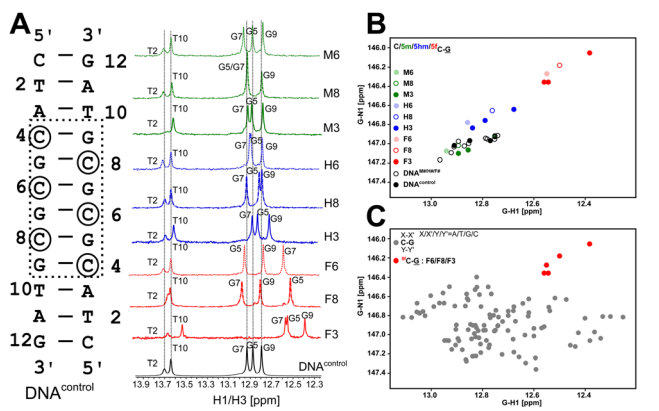

Fig. 2 (A) ^{1}H 1D NMR spectra acquired for DNA^{control} (bottom trace, black) and modified DNA$^{N\#}$ (N = M/H/F for ^{5m}C/^{5hm}C/^{5f}C, respectively, and # = 8/6/3 for hemi-/fully/triply modified samples) indicate stable duplex formation across samples. (B) Scatter plot of ^{15}N–^{1}H chemical shift correlation obtained for G–N1/H1 paired with C/^{5m}C/^{5hm}C/^{5f}C chemical shifts obtained for DNA^{control} (black filled circles) and DNA$^{N\#}$ (circles colored based on modification N = M/H/F). C–G pairs that are unmodified within modified sequences are also shown (open black squares) to indicate that only the modified cytosine experiences CSP. (C) Comparison of ^{5f}C modified G–N1/H1 shifts (red circles) to unmodified C–G pairs across all possible trinucleotide sequence contexts (gray circle).

Consequently, imino CSP is being upfield shifted for ^{5f}C/^{5hm}C and downfield shifted for ^{5m}C paired G–N1/H1 in comparison to unmodified C. Prior computational studies predict a correlated change in G–N1 and G–H1 chemical shifts due to the weakening of the C–G base pair upon chemical modification of the cytosine base.[48]

Having assessed the effect of cytosine modification on base pairing, next is to quantify the changes that may arise due to the stacking of a chemically altered base on the 5′- and 3′-neighbors. The G7–N1/H1 resonances in DNA^{N8} (5′-end neighbor of C8, Fig. 2 and Fig. S3, ESI†) are downfield shifted to 0.15/0.06 ppm for ^{5f}C, while a negligible change is observed for $^{5m/5hm}$C (Table S1 and Fig. S3, ESI†), suggesting either ring current or stacking change (or both) only for the ^{5f}C modification. These measurements would come in handy to interpret the chemical shift perturbation for DNA^{N6} modifications, wherein a mere arithmetic sum of H-bonding and ring current effects would then indicate no appreciable difference between single hemi-modified (*i.e.*, DNA^{N8}) and single fully modified (*i.e.*, DNA^{N6}) cases. The magnitude and directionality of G–N1/H1 chemical shift perturbation for the C6–G7 pair in DNA^{N6} are in line with the observation for C8–G5 in DNA^{N8} sequences across all modifications (N = M/H/F). Such an observation suggests that base pairing affects the chemical shifts more significantly than the effect of modified ring current effects. Importantly, the G7–N1/H1 shifts in DNA^{N6} (for all modifications) show a simple arithmetic sum of chemical shift perturbation due to H-bonding and 3′ neighbor effect, indicating no significant structural changes from single hemi-modified to single fully modified systems (Table S2, ESI†).

Next, the question arises whether single *versus* multiple modifications cause any differential effects on the DNA duplex. Like the observation in DNA^{N6} systems, G5–N1/H1 and G7–N1/H1 chemical shift changes in DNA^{N3} (for all modifications) are simple arithmetic sums of a single fully (6th position) and hemi-modified (8th position) chemical shift. The only exception is observed with the magnitude of the G9–N1/H1 chemical shift change that arises due to inherent differences in the dinucleotide step (A<u>C</u> *vs.* G<u>C</u>). Noticeably, in DNA^{F3}, the T10–N3/H3 and T2–N3/H3 nuclei experience a significant upfield shift to 0.25/0.13 ppm and 0.11/0.03 ppm suggesting a weakening of pairing that is two base pairs away from the sight of ^{5f}C modification (Fig. 2A and Table S1, Fig. S3, ESI†) for the triply ^{5f}C modified system. This observation is in agreement with complementary infra-red[41] and NMR[44] experiments, where the rate of duplex association is markedly reduced while that of dissociation is increased upon ^{5f}C incorporation.

It is intriguing to comprehend the implications of the upfield shift of imino resonances of ^{5f}C–G pairs in the context of the DNA duplex structure. Comparison of the measured shifts for the imino resonances of C–G pairs across primary sequence contexts would yield insights into how the ^{5f}C–G pair differs from the canonical unmodified C–G pair. This was carried out by generating DNA samples consisting of trinucleotide steps in the non-terminal regions of dodecamer duplexes with C–G being the middle base pair (*i.e.*, 5′-X<u>C</u>Y-3′ paired "•" with 5′-Y′<u>G</u>X′-3′) flanked by canonical Watson–Crick pairs (X–X′ and Y–Y′). The first nearest neighbors to the C–G pair on both 5′- and 3′-ends were sampled across all possible trinucleotides (X/X′/Y/Y′ = A/T/G/C) resulting in 16 combinations, with a minimum of four replicates for each combination (unpublished data). The average G–N1/H1 chemical shift for all C–G pairs is observed to be 146.9/12.75 ppm (110 data points, Fig. 2C), agreeing well with the data obtained for DNA^{control}. The ^{5m}C and ^{5hm}C modified G–N1/H1 resonate at 147.0/12.87 ppm and 146.7/12.79 ppm, respectively, with 5 data points each across DNA$^{M\#/H\#}$ (Fig. 2B). Interestingly, for the ^{5f}C modified base-pair, G–N1/H1 are well resolved from the entire cluster of C–G canonical pairs and resonate at 146.2/12.51 ppm (5 data points across DNA$^{F\#}$) – upfield shifted in both ^{15}N and ^{1}H dimensions (Fig. 2C). The significant average upfield shift for G–N1/H1 paired to ^{5f}C in comparison to ^{5m}C/^{5hm}C and the entire C–G cluster indicates that the destabilization achieved for C–G upon formylation is beyond the scope that is achievable for any given trinucleotide primary sequence of DNA. This is an important observation given the fact that C–G pairs tend to impart stability to the DNA duplex in comparison to A–T pairs. The ^{5f}C modification, in contrast, relaxes this property and contributes to the necessary level of destabilization beyond the scope achievable from the primary sequence, yet suitably retaining the Watson–Crick pairing that is essential for biomolecular processes.

Amino ^{1}H spins present in the cytosine nucleobase (C–H41/H42) also corroborate the above observations. ^{1}H chemical shifts of C–H41, which is also involved in the formation of Watson–Crick H-bonding, are relatively downfield shifted at the $^{5m/5hm/5f}$C nucleotide position. On the other hand, the chemical shift of C–H42 experiences an upfield shift for ^{5m}C (0.30–0.40 ppm) and ^{5hm}C (0.10–0.14 ppm), while ^{5f}C modification results in a significant downfield shift (~1.5 ppm). This observation supports the formation of a intranucleobase H-bond between the formyl group's carbonyl oxygen (C=O) and the amino proton (H42) of 5-formyl cytosine.[60] This intramolecular H-bonding of ^{5f}C restricts formyl substituent conformation and hence forces it to be in plane with the cytosine aromatic ring, consistent with the previous reports.[35] The small magnitude of chemical shift perturbation for $^{5m/5hm}$C indicates these bases do not make such type of H-bonding (CHO H-bond for instance), with prior crystallographic studies involving ^{5hm}C containing DNA providing evidence that the orientation of CH_2OH precludes such intramolecular H-bond formation with C(H42).[34] Such an intramolecular H-bond excludes the interaction of water molecules at this site, which is otherwise available with the CH_3 and CH_2OH modifications.[61]

Following the characterization of ^{15}N/^{1}H imino/amino shifts, changes in ^{13}C/^{1}H were pursued for the aromatic base [C–C6/H6 and G–C8/H8]. As anticipated, C–C6/H6 was highest for the modified base due to the change in the functional group present in the 5th position, with upfield shift (3.3/0.2 ppm) for ^{5m}C and downfield shifts for ^{5f}C (13.3/0.9 ppm) and ^{5hm}C (1.7/0.04 ppm) (Fig. S3, ESI†). Importantly, G5–C8/H8 nucleotide DNA^{F8} (^{5f}C pair) experiences a downfield shift of 0.3/0.04 ppm

(Fig. S3, ESI†), sensing the weakening of ^{5f}C–G H-bond strength propagated by the aromaticity of the nucleobase. Next, ^{13}C–C8 CSP of G7 in DNA^{N6} samples was analyzed to probe for any effects that may arise due to single contiguous modifications in the DNA duplex, *versus* a hemi-modified case (DNA^{N8}). We observe a simple arithmetic sum of the H-bonding and ring-current changes manifested by the 5′/3′-neighbor (as adjudged from DNA^{N8}) for all the cytosine modifications (Table S2, ESI†), without any exceptions. This suggests that the modifications do not confer any additive effect in terms of structural perturbations beyond the site of change. A similar observation is made when comparing ^{13}C-C8 CSP of G5, G7, and G9 for DNA^{N3} samples, potentially indicating minimal changes along the major groove of the DNA duplex due to multiple contiguous modifications present in the system. Like aromatic ^{13}C–H chemical shift perturbations, the furanose ring was most affected for the modified bases, with ^{5f}C–C1′/H1′ nuclei experiencing the highest magnitude of 0.7–0.9/∼0.02 ppm (Fig. S3, ESI†). Although C1′ shifts report on sugar pucker equilibria,[49,62] their interpretation, in this case, is affected due to the strong influence of ring current effects. Thus, furanose ^{13}C/^{1}H shifts are not further interpreted.

The magnitude of ^{13}C–^{1}H scalar coupling indicates a local deviation from the C2′-*endo* pucker at ^{5f}C modified sites

Prior NMR studies involving DNA^{F3} hinted at the deviation of the ^{5f}C sugars away from the C2′-*endo* pucker, adjudged from the cross-peak intensities observed in the NOESY data across furanose ring protons.[39] Scalar couplings between protons connected *via* three covalent bonds ($^{3}J_{HH}$) are immensely useful in characterizing ring puckers, especially for nucleic acids.[63,64] These are measured conventionally using the double-quantum filtered ^{1}H–^{1}H COSY experiment, where deoxyribose sugars populated heavily close to the C2′-*endo* pucker show substantial $\Sigma^{3}J_{HH}$ between H1′–H2′/H2″ ($\Sigma^{3}J_{H1'-H2'/2''}$, 10–15 Hz).[51,65] On the other hand, deoxyribose sugars averaging in their C3′-*endo* pucker are expected to display a reduction of such a measurement such that $\Sigma^{3}J_{H1'-H2'/2''} \sim 7-8$ Hz.[65] Previous report on ^{5f}C modified duplexes documented small reductions (0.5–1 Hz) in $\Sigma^{3}J_{H1'-H2'/2''}$, with the NOESY data indicating an excursion away from the C2′-*endo* pucker for the formyl-modified cytosines adjudged from the inter-proton distances obtained from the NOESY experiment.[39] However, $\Sigma^{3}J_{H1'-H2'/2''}$ (ΣH1′) measurements are relatively insensitive, requiring a significant population change (∼30%) away from the C2′-*endo* pucker to effect a substantial reduction of the coupling (∼1 Hz) given the precision of the measurements (∼0.5 Hz).[65] Thus, other probes would be convenient for mapping subtle pucker changes. One-bond heteronuclear scalar couplings (*e.g.*, $^{1}J_{C1'-H1'}$) are influenced by torsion angles (including pucker and glycosyl angle) and C–H bond lengths making them attractive probes to highlight sugar pucker changes.[52,66,67]

Beginning with the DNA^{control} system, we observe that the position of the cytosine in the sequence influences the magnitude of the $^{1}J_{C1'-H1'}$ coupling magnitude. For instance, $^{1}J_{C1'-H1'}$ for the cytosine nucleotide in the RCG (R = purine, A or G) trinucleotide step is found to be ∼166 Hz, while 5′-CT (cytosine positioned at the 5′-end of the DNA strand) averages ∼172 Hz. This is expected as conformational degrees of freedom allow 5′-terminal cytosine to sample a broader range of puckers and glycosyl torsion angles. No significant difference in $^{1}J_{C1'-H1'}$ ($\Delta^{1}J_{C1'-H1'}$, relative to DNA^{control}) is observed for all nucleotides present in DNA$^{M\#}$ and DNA$^{H\#}$ within the measurement uncertainty (±2 Hz) (Fig. 3A). On the other hand, $\Delta^{1}J_{C1'-H1'}$ for singly modified ^{5f}C6 (in DNA^{F6}) and ^{5f}C8 (DNA^{F8}) results in an increase of 5–6 Hz, while the unmodified cytosine nucleotides within these samples show no change (Fig. 3A). All ^{5f}C-modified nucleotides in DNA^{F3} also exhibit an increase of 3–6 Hz (Fig. 3A). No significant changes were observed for aromatic ^{13}C–^{1}H $^{1}J_{CH}$ (adenine C2–H2, pyrimidine C6–H6, purine C8–H8), indicating the reliability of the scalar coupling measurements (Fig. S4, ESI†). An increase in $^{1}J_{C1'-H1'}$ indicates a deviation from the C2′-*endo* sugar pucker as predicted from a computational study involving ribose sugars for a given *anti* glycosyl dihedral angle, with C3′-*endo* being predicted to have a coupling of 178 Hz, 10 Hz increase over the C2′-*endo* conditions.[52] NMR data analysis across 2D spectra (NOESY, HMQC, and HSQC) of ^{5f}C modified DNA (DNA$^{F\#}$) rules out any evidence of ^{5f}C/G *syn* orientation. Hence, the increased $^{1}J_{C1'-H1'}$ of ^{5f}C potentially arises due to the shift in sugar pucker

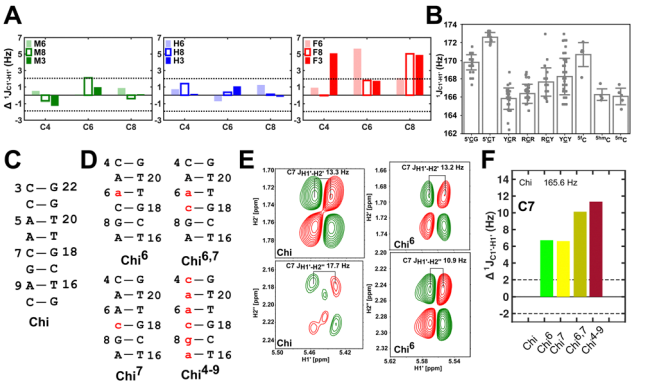

Fig. 3 (A) Changes to one-bond ^{13}C–^{1}H sugar C1′–H1′ heteronuclear scalar coupling magnitudes ($\Delta^{1}J_{C1'-H1'}$, in Hz) for nucleotide positions 4, 6, and 8 upon cytosine modification across DNA$^{N\#}$ samples. Measurement uncertainty (2 Hz) is marked with dotted lines, with ^{5f}C modification (in red) showing significant changes relative to unmodified cytosine. (B) $^{1}J_{C1'-H1'}$ scalar coupling magnitude for cytosine nucleotides juxtaposed between being purine (R)/pyrimidine (Y) neighbors within a trinucleotide step. 5′-Terminal cytosine (5′CG and 5′CT, ∼170–172 Hz) displays a higher magnitude relative to cytosine present in the core of the helix (166–168 Hz). ^{5m}C and ^{5hm}C show no significant difference (∼166 Hz), while ^{5f}C modification introduces a ∼6 Hz difference (RCR *versus* ^{5f}C). (C) Non-palindromic model system (*Chi*) was studied to chart the deviation of the sugar pucker from C2′-*endo* conformation by introducing ribose sugars. (D) Secondary structures of ribose containing the "*Chi*" system, with ribose sugars marked in red and with small alphabets. (E) Subset of DQF-COSY spectra highlighting the reduction in $\Sigma^{3}J_{H1'-H2''}$ for *Chi*^{6} (6th position adenine changed to ribose) in comparison to *Chi*. (F) Change in one-bond ^{13}C–^{1}H C1′–H1′ scalar coupling by 6–11 Hz upon single (*Chi*^{6}, *Chi*^{7}), double (*Chi*^{6,7}), and multiple (*Chi*$^{4-9}$) ribose incorporations (relative to *Chi*).

equilibrium from C2′-*endo* and plausibly subtle changes in the glycosidic dihedral angle.[52,66,67]

The $^{1}J_{C1'-H1'}$ magnitude is also influenced by the ^{13}C–^{1}H bond distance.[66] Formyl being an electron-withdrawing group might affect the bond lengths of base C6–H6 and furanose C1′–H1′ due to the resonance effect in aromatic rings. Although C6–H6 chemical shifts are most affected by C5 modifications of cytosine, $\Delta^{1}J_{C6-H6}$ for all nucleobases (including modified cytosine) remains within ±2 Hz across all systems ($DNA^{N\#}$, Fig. S4, ESI†). And, if it was the bond distance that caused a change in $\Delta^{1}J_{C1'-H1'}$, then irrespective of the position and across samples (*i.e.*, $DNA^{F6/F8}$ and DNA^{F3}) the magnitude of change would have remained constant. The mere fact that ^{5f}C modified in the sixth position in DNA^{F6} (∼6 Hz) and DNA^{F3} (∼3 Hz) are different suggests that the change in scalar coupling is not due to bond-distance changes. Additionally, a comparison of high-resolution (∼1 Å) crystal structures of the cytosine nucleotide (BOXGIE, CCDC 114593) and ^{5f}C (RAKLOG, CCDC 843055) showed no substantial increase in the C1′–H1′ bond length, supporting the fact that the $^{1}J_{C1'-H1'}$ change is not due to change in bond length but due to other structural factors (pucker and glycosyl dihedral angle).

To put things in perspective regarding $^{1}J_{C1'-H1'}$ scalar coupling measurements, similar data were measured for cytosine present across trinucleotide repeats and in the 5′/3′ termini of duplex DNA (unpublished data). The presence of cytosine in the 5′-terminus observed for 5′-C̲G and 5′-C̲T results in 169.8 ± 0.8 and 172.6 ± 0.5 Hz, respectively, while the 3′-terminal G C̲-3′ displays an average of 167.3 ± 1.3 Hz (Fig. 3B). Penultimate to 5′/3′-termini results in reduction (166.7 ± 1.1 for 5′-GC̲C and 166.0 ± 1.1 Hz for TC̲G-3′) in the magnitude with respect to the termini by 1–3 Hz. Similar measurements across the RC̲R, RC̲Y, YC̲R, and YC̲Y (where R = purine and Y = pyrimidine) trinucleotide steps within the ''core'' of the duplex resulted in 166.4 ± 1.0, 167.6 ± 1.5, 165.8 ± 1.2, and 168.2 ± 2.0 Hz, respectively, with the highest magnitude and spread of measured scalar couplings for the YC̲Y (Fig. 3B) step. The observations are thus consistent with the fact that the cytosine nucleotide tends to sample a larger conformational pool[68] depending on the available degrees of freedom, with $^{1}J_{C1'-H1'}$ measurements reflecting the same. The increase in $^{1}J_{C1'-H1'}$ by 3–6 Hz suggests that ^{5f}C modification to the RC̲G step makes it behave like the YC̲Y step, the most conformationally flexible trinucleotide present.

To further validate the results obtained from $^{1}J_{C1'-H1'}$, control experiments were performed with ribose sugars in a non-palindromic DNA duplex (Fig. 3C, reference ''*Chi*'' system) anticipated to force pucker equilibria away from C2′-*endo*.[69,70] In this sequence, ribose sugars were strategically positioned to increase the population of the C3′-*endo* pucker on the cytosine nucleotide (C7). Positioning the ribose sugar in A6 (Fig. 3D, Chi^{6}) results in an increase of $\Delta^{1}J_{C1'-H1'}$ of ∼7 Hz, accompanied by a decrease in ΣH1′ (H1′–H2″) of ∼7 Hz (Fig. 3E) indicating the pucker equilibria shifting towards C3′-*endo*. This is validated by ribose sugar modification for *Chi* at positions C7 (Chi^{7}), A6 and C7 ($Chi^{6,7}$), and C4–A9 (Chi^{4-9}), where C7 $\Delta^{1}J_{C1'-H1'}$ increased by 7–12 Hz (Fig. 3F), and by the disappearance of the H1′–H2′ cross peak in the DQF-COSY spectrum. Hence a change in $\Delta^{1}J_{C1'-H1'}$ for ^{5f}C modified nucleotides indicates puckering away from C2′-*endo* by a small yet significant degree.

Residual dipolar coupling measurements reiterate that ^{5f}C modified sites deviate in pucker/glycosyl angle

RDC measurements have the capability of mapping global structural changes, in addition to local perturbations.[45,53,54,56] Comparison of RDCs for an A-tract DNA duplex *versus* a randomized sequence clearly indicates the helical bending observed in the former.[55,57,71–73] RDCs would further complement $^{1}J_{C1'-H1'}$ measurements in probing sugar pucker changes for ^{5f}C modified DNA duplexes. In particular, C1′–H1′, C2′–H2′/2″ and C3′–H3′ RDCs are sensitive to the changes in the pseudorotation angle.[45] Since sugar moieties display fast exchange across the different puckers, RDC measurements have been interpreted as a population-weighted average across C2′-*endo* and C3′-*endo* puckers. Such studies on DDD have shown that cytosine sugar present in the core ends up sampling 20–30% C3′-*endo* pucker, followed by thymidine (2–20%) and purines[45] (0–4%). RDCs measured for $DNA^{control}$ also reiterate their ability to discriminate pucker differences as C4 present in an AC̲G shows a lowered $^{1}D_{C1'-H1'}$ (3.5 Hz) in comparison to C6/C8 (11–13 Hz) that is present in the GC̲G step ($^{1}D_{C6-H6}$ for C4/C6/C8 19–21 Hz). Structure refinement of $DNA^{control}$ with NOE-derived distances and RDCs indicates that the C4 sugar pucker averages around the O4′-*endo* while C6/C8 sample the C1′-*exo* to C2′-*endo* pucker (see the next section).

Measuring RDCs and correlating the measured values across $DNA^{control}$ and modified systems ($DNA^{N\#}$) would aid in characterizing any global bending that may be present upon cytosine modification. To start with, a good RDC agreement (Pearson's coefficient of R^2 ∼ 0.95 and RDC RMSD ∼ 1.2 Hz, Fig. 4A) was observed for concentrated (2.7 mM, uniform Nyquist NMR data sampling and conventional Fourier transform processing) and diluted (500 μM, 25% sparse sampling and compressed sensing processing) $DNA^{control}$ samples indicating that the sparse methodology for limited concentration samples works as efficiently (within the experimental uncertainty of ∼2 Hz) as the routinely employed conventional methods.

RDCs measured for ^{5m}C and ^{5hm}C modified samples ($DNA^{M\#}$ and $DNA^{H\#}$) correlate well with $DNA^{control}$ (R^2 in the range of 0.86–0.91 and RMSD < 2 Hz, Fig. S9, ESI†), indicating similarity in their overall structure. Strikingly, significant RDC differences are observed for DNA^{F6} and DNA^{F3} (R^2 0.75–0.80, RMSD 3.0–3.5 Hz, Fig. 4B and D) but within the experimental uncertainty for DNA^{F8} (R^2 0.88, RMSD 2.0 Hz, Fig. 4C) pointing at differences between single hemi-modified (DNA^{F8}) and single fully modified (DNA^{F6}) systems. Noticeably, ^{5f}C–C1′–H1′ RDC is the only data point (indicated in pink color in Fig. 4B and D) that deviates by 6–10 Hz reduction in the correlation plot. Removal of these ^{5f}C C1′–H1′ RDC outliers improves the

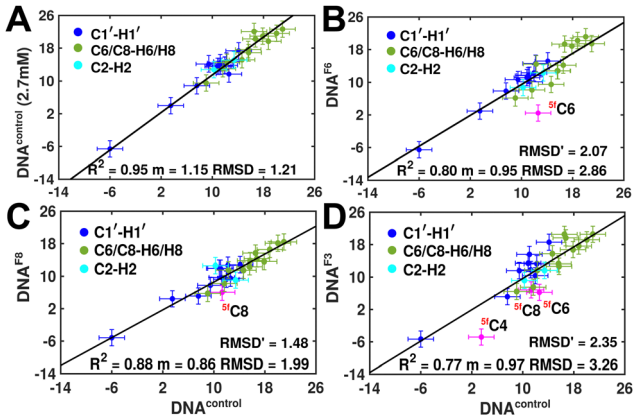

Fig. 4 Experimentally measured RDC correlation scatter plots to highlight the differences that arise between $DNA^{control}$ and $DNA^{F\#}$, with sugar (C1′–H1′, blue) and nucleobase (C6/C8–H6/H8 and C2–H2, in green and cyan, respectively) RDCs displayed. (A) Comparison of RDCs measured between $DNA^{control}$ (2.7 mM, *y*-axis) using conventional NMR data acquisition and $DNA^{control}$ (500 mM, *x*-axis) with 25% sparse sampling NMR methods. Data were best fit with a linear function (solid black line) without an intercept, with the slope varying depending upon subtle changes in Pf1 alignment media concentrations known to arise during sample preparation. RDC RMSD is calculated between the *x*- and *y*-axis to highlight that low-concentration sparse sampling methods work within experimental uncertainties (2 Hz, represented by error bars). Scatter of RDCs measured for DNA^{F6} (B), DNA^{F8} (C) and DNA^{F3} (D) plotted against $DNA^{control}$, with C1′–H1′ RDC of the ^{5f}C modified RDC marked in pink. RMSD′ reported in panels (B)–(D) indicates measurement difference with $DNA^{control}$ when ^{5f}C modified nucleotide measurement is removed.

correlation ($R^2 \sim 0.90$, RMSD′ < 2 Hz, Fig. S9, ESI†), implying only a change in the local structure for $DNA^{F6/F3}$ with no apparent helical bending that is any different from $DNA^{control}$.

The RDC measurement also helps rule out the possibility of C–H bond length changes for the C1′–H1′ bond vector. A back-of-the-envelope calculation suggests that a ~6 Hz decrease in RDC (given an alignment and B-DNA structure for DNA and $DNA^{N\#}$) requires an increase of ~0.25 Å in the C1′–H1′ bond length, which is rather unlikely. The ^{5f}C selective deviations corroborate with the ~6 Hz increase in $\Delta^1J_{C1'-H1'}$, suggesting a local structural perturbation induced by ^{5f}C plausibly due to changes in sugar pucker equilibria away from canonical C2′-*endo* conformation for B-DNA.

It is pertinent to note here that the magnitude of terminal 5′-CT C1′–H1′ RDCs is in the range of −5 to −8 Hz across $DNA^{control}$ and $DNA^{N\#}$ samples (Table S1, ESI†). This scenario yet again highlights that ^{5f}C alters the local structure in terms of pucker and glycosyl dihedral angle for the RCG step; however, it does not make it as flexible as the terminal cytosine nucleotides.

Structure refinement supports the change in the pucker at ^{5f}C modified sites

Following the detailed analysis of NMR parameters, the next step was to refine the structure using the NOESY and RDC data acquired for all the samples. Firstly, NOESY cross peak connectivity across the base (H6/H8) and sugar protons (H1′/H2′/H2″) qualitatively confirms that all DNA duplexes are in the right-handed helix in solution and close to B-form conformation.[58,74–76] The weak NOE cross-peak of inter and intranucleotide H6/H8–H1′ and intranucleotide H6/H8–H2″ and the strong intensity of intranucleotide H6/H8–H2′ qualitatively describe a high *anti* glycosyl torsional angle and a C2′-*endo* sugar conformation for $^{5m/5hm/5f}$C DNA.

Next, the characterization of the structures sampled by $DNA^{control}$ and $DNA^{N\#}$ was pursued using inter-proton distances and RDCs as constraints. As the number of measurements/constraints are significantly small given the total number of degrees of freedom available for nucleic acids,[45] the aim here was to avoid overfitting the NMR data yet obtain a (low-resolution) conformational model for $DNA^{control}$ and $DNA^{N\#}$ that may highlight any differences in the DNA duplex upon modification. Also, as the modifications are in the major groove with no effect on Watson–Crick pairing, the unmodified cytosine nucleobase was refined against the measured NMR parameters for each of the $DNA^{N\#}$ modified sequences. Thus, the measured data (inter-proton distances and RDCs, Table S3, ESI†) were supplied to refine initialized from "idealized" B-DNA geometry using the XPLOR-NIH structure refinement program[77] (see Experimental methods).

Upon refinement, DNA systems studied ($DNA^{control}$ and $DNA^{N\#}$) continue to sample an overall B-DNA as anticipated and predicted in previous studies (Fig. S5, ESI†).[39] Notably, RDCs refine the B-DNA structure where back-prediction of RDCs measured for $DNA^{N\#}$ with the $DNA^{control}$ structure (and *vice versa*) yields experimentally derived correlations (Table S5, ESI†). It indicates that refined structures mimic conformations sampled across these modifications. Structural analysis of refined conformers was performed to determine base pairs, base-pair step parameters, sugar pucker using 3DNA, and Curves+ to determine DNA helical curvature (methods, Table S4, ESI†). Parameters that are used to define intra-basepair[78] (shear, stretch, stagger, buckle, propeller, and opening) and inter-basepairs[78] (shift, slide, rise, roll, tilt, and twist) and dihedral angles (backbone: α, β, γ, δ, ε, ζ; glycosidic dihedral angle χ; and sugar: ν_0–ν_4) follow the anticipated distribution about the canonical B-DNA geometry without any exceptions. No differences between average helical bending (within the measurement uncertainty and structural noise) and major groove widths were observed between $DNA^{control}$ and $DNA^{N\#}$.

Sugar pucker analysis of the refined structures agrees with the inferences derived from one-bond scalar ($^1J_{C1'-H1'}$) and residual ($^1D_{C1'-H1'}$) dipolar coupling measurements. Sugar puckers in B-DNA are known to sample conformations about the C2′-*endo* puckers, with drifts commonly observed towards O4′-*endo*. This expectation is preserved for $DNA^{control}$ and $DNA^{M\#/H\#}$ systems (Fig. 5A). Mainly, the ACG ($^1D_{C1'-H1'} \sim 4$ Hz for C4) *versus* GCG ($^1D_{C1'-H1'}$ 11–14 Hz for C6 and C8) trinucleotide step indicates a discernable difference in the pucker equilibria corroborating the RDC measurements for these steps in $DNA^{control}$ (Fig. 5A).

In the single ^{5f}C-modified systems, it is observed that the C6 nucleotide in DNA^{F6} shows more extensive excursions towards

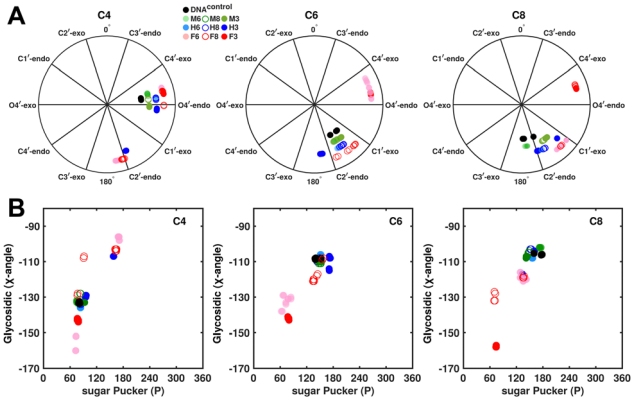

Fig. 5 (A) Pseudorotation phase angle plots at cytosine nucleotides C4, C6, and C8 of refined DNA structures to compare the sugar conformation of unmodified and modified DNA. (B) Variation in the glycosidic dihedral angle as a function of the sugar pucker for the refined DNA structures. Black, green, blue, and red colored data points correspond to $DNA^{control}$, $DNA^{M\#}$, $DNA^{H\#}$, and $DNA^{F\#}$, respectively.

O4′-*endo* compared to $DNA^{control}$. In contrast, DNA^{F8} shows to a lesser extent, in agreement with the coupling measurements and highlights the difference between single hemi-modified and single fully modified ^{5f}C systems. DNA^{F3} alters the pucker clearly for C6 and C8 away from C2′-*endo*, while C4, which is already at O4′-*endo*, is altered to a smaller extent. Additionally, pucker changes tend to affect the glycosidic torsional (χ-)angle, as observed for A- (C3′-*endo*, $\chi = -150^{\circ}$) and B-DNA (C2′-*endo*, $\chi = -110^{\circ}$). A correlation was plotted between sugar pucker and χ (Fig. 5B) for the refined DNA structures to see whether a similar effect persists upon ^{5f}C modification. Indeed, for nucleotides C6 (DNA^{F6}) and C8 (DNA^{F3}), C4 is affected in $DNA^{control}$ and $DNA^{N\#}$ due to its presence in the ACG step (Fig. 5B). In contrast, all complementary base-paired guanosine nucleotides (*i.e.*, G5, G7, and G9) exist in C2′-*endo* with χ near -100°, pointing to the relative orientation between base and sugar changing locally at the ^{5f}C site.

Further, to assess whether any correlated change occurs in the phosphate backbone due to alteration in the pucker, the phosphate backbone dihedral angles ε and ζ were measured from the refined structures to see whether B_I ($\varepsilon - \zeta < 0$) and B_{II} ($\varepsilon - \zeta > 0$) equilibria get affected. The correlation of the sugar pucker to $\varepsilon - \zeta$ indicates that all cytosine nucleotides in $DNA^{control}$ and $DNA^{N\#}$ are in B_I backbone conformation (Fig. S6, ESI†), without exceptions. Indeed, the results are analyzed conservatively, as without ^{31}P chemical shifts and scalar coupling ($^3J_{P-H3'}$ and $^4J_{P-H4'}$) measurements the observations cannot be further refined/validated. Thus, ^{5f}C modification in duplex DNA alters sugar pucker equilibria without significant changes to other conformational and structural properties.

Discussion

The effect of $^{5m/5hm/5f}C$ on the stability and structural properties of the DNA duplexes has been studied employing various spectroscopic techniques. Thermal melting studies show that ^{5m}C increases the duplex stability by $\sim 5\ ^{\circ}C$, and $^{5hm/5f}C$ tends to reverse the impact of stability afforded by ^{5m}C.[37,41,42,79] ^{5hm}C has a melting temperature similar to that of unmodified DNA, whereas ^{5f}C destabilizes the DNA duplex by $\sim 3\ ^{\circ}C$.[41,42,44,57] Contrastingly, the presence of ^{5f}C in duplex RNA results in increased stability with a $\sim 5\ ^{\circ}C$ increase in the melting temperature, due to increased stacking interactions with neighboring base pairs.[80] In addition to DNA and RNA duplexes, formation of i-motifs in cytosine-rich DNA sequences is also altered by the presence of these epigenetic modifications where $C–C^+$ pairs are formed.[81] The fact that additional protonation is required to stably form $C–C^+$ pairs, the addition of CH_2OH and CHO groups stabilizes i-motifs at a lower pH (~ 0.1 units relative to unmodified cytosine), while ^{5m}C increases the same by 0.1–0.2 units.[81]

Prior studies have pointed out that CHO (^{5f}C) and COOH (^{5ca}C) modifications in cytosine change the pK_a of the H-bond accepting N3 nitrogen atom that was predicted to cause a weakening of the H-bond for DNA duplexes.[42,43,82] Computational studies performed on such modified cytosine duplex systems report that the calculated isotropic chemical shift of both the imino proton (^1H) and nitrogen (^{15}N) shows a correlated change with the increasing or decreasing H-bond distance in the C–G base pair.[48] Geometry optimized and energy minimized structures of C–G pairs predict an increase in the G:N1–H1···N3:C distance upon varying C from ^{5m}C to ^{5hm}C, ^{5f}C, and ^{5ca}C, the longest being for the ^{5f}C–G base pair.[59] Such a weakening of the H-bond is attributed to enhanced base-pair opening rates[35] and increased population of single-stranded DNA.[41,44] However, direct measurement of structural changes in duplex DNA upon ^{5f}C modification would be convenient and aid in characterizing other pertinent modifications in nucleic acids.

Our results of $^{15}N/^1H$ chemical shifts of the guanosine base paired with the modified cytosine provide an unbiased way of assessing local structural changes. Notably, the measurements are made without the need for ^{15}N-isotopically enriched samples, demonstrating $^{13}C/^{15}N$ chemical shift measurements to be a viable approach to studying modified nucleotides – an unexplored treasure trove in terms of epigenetics, damage/lesion, and epitranscriptomics. $^{15}N/^1H$ chemical shifts measured from the complementary G paired to ^{5m}C, and ^{5f}C modified nucleotides show significant downfield and upfield shifts, respectively, indicating the strengthening and weakening of the H-bond. In addition, the weakening of the ^{5f}C–G base-pair propagates beyond the modification site, as reported for DNA^{F3}, substantiating the previous findings that ^{5f}C destabilizes the whole DNA duplex.[44,82] Thus, measurement of ^{15}N chemical shifts could proxy as an indicator of strengthening/weakening akin to the chemical exchange saturation transfer type experiments. This also explains that ^{5f}C containing DNA templates display reduced substrate specificity of dGTP incorporation as observed experimentally.[30] The insertion of dGMP opposite to ^{5f}C is less efficient in comparison with the insertion of dGMP opposite to unmodified C, with dAMP/dTMP being more frequently misincorporated.[83]

DNA duplexes are known to exhibit exchange across lowly populated conformational states (such as Hoogsteen and tautomeric forms) that have been implicated in various functional roles.[84–88] As G–C pair Hoogsteen pair formation requires C–N3 protonation, we speculate that lowered pK_a for cytosine (4.5 units) upon 5-formyl incorporation (2.1 units) would reduce the Hoogsteen population. Also, prior studies have indicated that 5-formyl substitution could potentially drive cytosine to a lesser-known imino tautomer rather than the conventional amino form.[89] To keep the three H-bonds between the G–C pair, then such a change would force the paired guanosine to sample the enol (G^{enol}) form away from the keto form. Interestingly, the formation of G^{enol} has been documented to shift the G–N1 chemical shift (in the context of the dG·dT wobble pair) downfield by 30–50 ppm.[90,91] However, we observe for the ^{5f}C–G pair a moderate 0.8 ppm upfield shift of the ^{15}N–N1 paired guanosine indicating that such a tautomeric base pair formation appears less likely.

Crystal structures of the DNA duplex containing ^{5m}C[36] and ^{5f}C[37] have reported significant deviations from B-DNA. However, prior solution NMR studies refuted such claims based on NOE-based distances, indicating only subtle differences in the ^{5f}C-modified nucleotides.[39] In our studies, complementing NOEs, heteronuclear ^{13}C/^{15}N chemical shifts, and coupling-based measurements aid in confirming that the overall structure of $^{5m/5hm/5f}$C DNA does not deviate from that of canonical B-DNA. RDCs are effective probes for global structural perturbations and our results provide no evidence favoring the presence of E- or F-DNA forms under solution conditions. Heteronuclear scalar and residual dipolar couplings aid in capturing subtle variations in the local structure upon ^{5f}C incorporation. Combined analysis across various NMR parameters shows that ^{5f}C influences the local nucleotide structure in the sugar pucker and the glycosyl dihedral angle.

Contrary to common misconception, the DNA duplex embeds subtle differences on top of the uniform double-helix structure based on the primary sequence. For instance, sequence-specific variation in structure is essential for indirect DNA readout carried out by regulatory proteins.[92] Conformational flexibility of DNA allows for the torsion angles to sample sparsely populated states and is often functionally relevant. Hoogsteen base pair formation for A–T and C–G pairs is a good example and is known to induce helical bending and increase the propensity of DNA damage in the Watson–Crick phase.[57,88,93] Similarly, in B-DNA, 2′-deoxyriboses sugar moieties primarily pucker proximal to the C2′-*endo* region, transgressing to the C3′-*endo* conformation at 5–20% population based on the nucleobase type.[94] This is not surprising given that the C2′-*endo* form in B-form DNA is only marginally more stable than the C3′-*endo* form by $\sim$1 kcal mol^{-1}, with transitions occurring in the pico-nanoseconds timescale (energy barrier 2–5 kcal mol^{-1}).[68,95–97] Molecular dynamics simulation shows that C2′-*endo* to C3′-*endo* transitions occur stochastically and are uncooperative.[94] Hence, individual sugar puckering is rapid and such effects cannot be directly studied by spectroscopy as they do not dramatically impact the average duplex structure. Importantly, C3′-*endo* conformations are more commonly observed in pyrimidine (especially for C) nucleotides than in purine.[45,46] The lifetime and population of C3′-*endo* conformation increase to 20% for C located in the CG, CA, and TG steps compared to other dinucleotide steps, with CA, TG, TA, and CG being the most flexible steps in the DNA duplex.[46,98] ^{5f}C exploits this unique property of C, enhances the flexibility of DNA and establishes itself as a distinct cytosine modification over the other ^{5m}C and ^{5hm}C. Such a facet of ^{5f}C, in addition to weakened H-bonds, enables duplex DNA containing the modification to transiently sample locally melted and flexible states that results in faster duplex cyclization rates for ^{5f}C in comparison to C/^{5m}C/^{5hm}C. The rate increases with multiple ^{5f}C modifications in the sequence.[40]

It is well documented now that the chemical structure of the modifications in the 5th position of the cytosine base serves as a mode of recognition and binding of proteins.[25,99–102] For instance, ^{5f}C modification strongly interacts with transcriptional regulators, DNA repair factors and chromatin regulators.[25] The CHO group present in ^{5f}C is known to form covalent interactions with the amine groups present in proteins such as methyltransferases[103] and histones.[104] The motivation in our study was to interrogate the plausible effects that transcend the chemical structure and potentially drive conformational changes that modulate the properties of the double helical DNA structure. Our results unequivocally indicate that ^{5f}C introduction into the DNA duplex results in the sampling of C–G conformations that are not accessible within any sequence context. Hence, the weakening of H-bond strength achieved due to the formyl modification in the ^{5f}C–G pair enhances the base opening rate,[35] local fluctuations,[41] and double-strand DNA dissociation constant resulting in reduced DNA duplex stability[44] in comparison to any possible canonical primary sequence containing Watson–Crick base pairs. This is important as transcription factors are known to exploit the weakened base pair towards recognition.[105] Hence, because of base-pair wobbling around the ^{5f}C–G base-pair, the duplex achieves an enhanced degree of flexibility. Weakening of the ^{5f}C–G H-bond increases the probability of ^{5f}C base flipping and un-base stacking over the other ^{5m}C and ^{5hm}C, which may assist TDG in recognizing. Therefore, the base flipping into the catalytic pocket of the thymidine DNA glycosylase/base-excision repair[106] enzymes is plausibly facilitated.

Another factor to highlight here is the difference between epigenetic and damage modifications in duplex DNA. For instance, 1-methyadenine (^{m1}A) is a known form of DNA damage with a methyl group inhibiting Watson–Crick pairing and facilitating Hoogsteen pairing.[57] Such a modification is found to enhance local fluctuations in the millisecond time scale. In contrast, ^{5f}C epigenetic modification enhances conformational flexibility in the faster pico-nanosecond time scale motion (as no appreciable resonance broadening is observed in the NMR spectra of DNA$^{F\#}$) contrasting the effect of epigenetic *versus* a damage (^{m1}A) modification in the conformational landscape of DNA duplexes. This potentially underlines the fact that damage modifications that severely affect the function

of DNA duplexes cause more alarming conformational changes in comparison to epigenetic modifications that play more than one given role in the biological context. A thorough structural mapping of damage and natural modifications would aid in testing/refining this hypothesis.

Conclusions

Cytosine epigenetic modifications are reported to sample a wide range of polymorphic structures. Our study shows that all the cytosine modifications do not deviate from the B-DNA duplex structure, although prior crystallographic reports have suggested the same. We present heteronuclear chemical shifts and scalar couplings as effective probes to map subtle variations arising from chemical modifications in DNA. These NMR probes reveal the weakening of the G–C H-bonding upon formyl modification. The subtle differences between single and multiple ^{5f}C modifications are evidently observed with these measurements. Notably, the change in the pucker/glycosyl angle for the ^{5f}C modified duplexes highlights the fact that cytosine uniquely manages to change the local flexibility of the duplex thereby enhancing its functionality within the context of duplex DNA. Such a feature is brought about with no change in the canonical base pair, hence not affecting the integral function of DNA. Also, the fundamental paradigm of structure–function within molecular biology is expanded to include conformational flexibility that provides distinctive avenues for encoding information within the limited chemical space of nucleotides. Their alterations to the physical properties of duplex DNA upon ^{5f}C modification throw light on the role of epigenetic modifications in their biological function.

Experimental methods

Choice of the primary sequence

DNA oligonucleotides were prepared with the palindromic sequence 5′-CTAC̲GC̲GC̲GTAG-3′; 4th/6th/8th positions were modified with ^{5m}C/^{5hm}C/^{5f}C in various samples (Fig. 1A). The choice of the sequence was motivated by the $(CpG)_n$ repeat sequence that also has abundant data available across crystallography,[37,39] solution-state NMR,[39] infrared spectroscopy,[41] and computational studies.[38] Additionally, the system enables careful dissection of chemical shift perturbations that arise solely due to base-pairing (8th position, single hemi-modified) and a combination of base-pairing and stacking (6th position, single fully modified). The sequence also sports a CpG repeat sequence that allows one to understand the effect of single *versus* multiple contiguous modifications. ^{5m}C/^{5hm}C/^{5f}C modified duplexes are labeled as DNA$^{M\#}$/DNA$^{H\#}$/DNA$^{F\#}$ (# = 6, 8, or 3 for single fully, single hemi-modified, or triple modification, respectively). The sample without any modifications serving as the control is denoted as DNA^{control}.

Sample preparation

DNA^{control} was purchased from Integrated DNA Technologies (IDT USA) and modified DNA$^{N\#}$ (N = M/H/F) from Keck Oligonucleotide Synthesis Resource (W. M. Keck Foundation) synthesized using phosphoramidite chemistry[107] and purified with RP-HPLC (purity >99% from mass spectrometry). DNA oligonucleotides were used as is, without any further purification. Duplexes were annealed by heating single-strands (~200 μM concentration) in pure water to 95 °C for 10 min and cooling the sample at room temperature. The duplexes were then subjected to centrifugal concentration using 3 kDa cut-off filters (EMD Millipore) with the NMR buffer (15 mM sodium phosphate pH 7.4, 25 mM sodium chloride, 0.1 mM ethylene diamine tetraacetate (EDTA), 10% D_2O for field-frequency locking, 50 μM trimethylsilyl propanoic acid (TSP) as an internal standard for chemical shift referencing). The final duplex DNA concentrations for DNA^{control} and modified DNA$^{N\#}$ were between 90 and 250 μM. Partial anisotropic alignment was achieved by adding 20–25 mg mL^{-1} filamentous Pf1 phage[108] to the sample, keeping the DNA duplex concentration as similar as possible to the isotropic condition.

NMR spectroscopy

NMR experiments were performed employing a 700 MHz ^{1}H Larmor precession frequency Bruker Avance-III spectrometer equipped with a cryogenically cooled triple {^{1}H, ^{13}C}, ^{15}N channel resonance probe at 298 K. Chemical shifts were referenced using TSP to 0 ppm on the indirect ^{13}C dimension (following appropriate spectral aliasing) and direct ^{1}H dimension. The ^{1}H imino 1D NMR spectra of the DNA^{control} and DNA$^{N\#}$ samples show characteristic resonances between 12 and 14 ppm (Fig. 2A), indicating stable duplex formation facilitated by Watson–Crick pairing. The ^{1}H chemical shifts of DNA^{control} and DNA^{F3} were observed to be in excellent agreement (±0.02 ppm) with previously published values.[39] ^{13}C–C7 shifts of modified C̲H$_3$, C̲H$_2$OH, and C̲HO groups fall in the expected ~15, ~60, and ~191 ppm indicating their proper incorporation in the sites of interest in all the DNA systems (DNA$^{N\#}$) studied. ^{1}H shifts of the aldehyde CH̲O proton resonating at 9.2 ppm in DNA$^{F\#}$ samples against 9.5 ppm for the free base indicate stacking accompanied by duplex formation. In addition, observation of the CH̲O resonance at 9.2 ppm indicates the insignificant population of the geminal diol C(OH̲)$_2$ form, which resonates at ~5 ppm.[109]

Data were acquired using TopSpin 3.6pl5, with sparse Poisson-Gap[110] sampling scheduling done using the macro 'nusPGSv3' (PGS_TS3.2 distribution) obtained from the Wagner's lab (gwagner.med.harvard.edu). Two-dimensional (2D) heteronuclear correlations ^{13}C–^{1}H and ^{15}N-^{1}H were obtained using the sensitivity-enhanced adiabatic heteronuclear single quantum coherence (HSQC with ^{13}C adiabatic pulses with water flip-back)[111] and band-Selective Optimized Flip Angle Short Transient (SOFAST-) heteronuclear multiple quantum coherence (HMQC)[112,113] spectroscopy, respectively, from the Bruker pulse program library. The ^{13}C and ^{15}N spectral widths (with carrier position) were optimized to obtain maximal resolution (64 ms $t_{1,max}$) to 8 (83) and 16 (153) ppm, respectively, by spectral aliasing with minimal signal overlap/loss. The scheduling lists were generated with 5–30% (5% increments), 50%, 75%, and

95% sampling to obtain the optimum level of sampling, providing a robust measurement of chemical shifts and scalar couplings. Data were then processed using multi-dimensional decomposition[114] (qMDD 2.5 v3b) followed by NMRPipe[115] and analyzed using NMRFAM-SPARKY.[116] The details of the performance of sparse sampling methodology to measure chemical shifts and couplings robustly and reliably are provided in the ESI.†

The 2D nuclear Overhauser effect (NOESY, 100, 150, and 200 ms mixing time) and double-quantum filtered correlation (DQF-COSY) spectra were acquired with the 3-9-19 WATERGATE water suppression scheme and uniform sampling with an inter-scan delay of 2.5 and 1.5 s, respectively.[111] ^{1}H–^{1}H correlation 2D data were acquired using conventional Nyquist sampling. $^{1}J_{CH}$ and $^{1}D_{CH}$ couplings were measured for samples under isotropic and anisotropic conditions, respectively, from the frequency difference between the doublets obtained from ^{13}C–^{1}H 2D HSQC without decoupling in the direct detect ^{1}H dimension.

Analysis of the NOESY spectra, structure refinement and analysis

2D NOESY data were analyzed for all samples to obtain inter-proton distances required for structure refinement protocols.[57,58] Briefly, the H5–H6 distance in cytosine was referenced to 2.45 Å, the methyl cross-peaks were calibrated with the H6–H7# distance in thymine to 3.00 Å, and the H2′–H2″ distances to 1.76 Å.[117] The distances obtained were then relaxed by 50% to obtain the lower and upper limit constraints for the structure refinement, as described earlier.[57]

XPLOR-NIH[77] version 2.41 was used for structure refinement following a simulated annealing protocol. As DNA^{control} and DNA$^{N\#}$ are palindromic in nature, the C_2-axis of symmetry was input as a constraint. While data for the modified systems were used, the unmodified cytosine base was employed for the structure refinement protocols as a proxy for ^{5m}C, ^{5hm}C, and ^{5f}C modifications, as only the trends of structural perturbations were sought from such refinements. Alignment tensor parameters (D_a and D_r – the axial and rhombic components of the tensor) were optimized for the DNA duplexes based on the measured RDC datasets.[54] As imino ^{1}H shifts were observed in the characteristic 12–14 ppm region indicative of Watson–Crick base pairs, H-bond constraints were incorporated in the structure refinement protocol. Dihedral angles (except for ε and ζ angles) were constrained as described earlier. Phosphate backbone dihedral angles were not constrained to assess changes in the B_I/B_{II} populations upon modified cytosine incorporation. Fifty structures were annealed starting from the idealized B-DNA geometry, and the five structures having no restraint violations were used for further structural analysis. The number of restraints and the summary of structure refinement for each system are listed in Table S3 (ESI†).

Structural analysis of the refined conformers was performed to determine inter- and intra-base pair parameters using 3DNA,[89] while helical bending was assessed using CURVES+.[19] RDC comparisons (Table S5, ESI†) were generated by fitting experimental RDCs to refined DNA structures with the module calcTensor (single value decomposition for best-fitting experimental measurements to back-predicted values) present in XPLOR-NIH.[77]

Author contributions

B. S. conceptualized, acquired funding, supervised the investigation, methodology and formal analysis, and wrote the manuscript. M. J. carried out the methods, data curation and analysis, with R. K. R. S. sharing the load and validating the datasets across the entire project. M. J. and R. K. R. S worked in editing the manuscript. A. R. performed analysis of a subsection of the dataset in this project, with supervision from M. J. and R. K. R. S.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank the Indian Institute of Science Education and Research (IISER) Bhopal for providing the necessary research infrastructure. We would like to thank IISER Bhopal for allowing access to the 700 MHz NMR facility at IISER Bhopal and Mr Rajbeer Singh for timely support in the maintenance of the spectrometer. This work was supported by the Science and Engineering Research Board *via* the Early Career Research grant (ECR/2016/001196) and the start-up research grant (INST/CHM/2016047) from IISER Bhopal to B. S. M. J. thanks CSIR for the fellowship and research support. R. K. R. S. thanks IISER Bhopal for the research fellowship.

Notes and references

1 T. H. Bester, *Gene*, 1988, **74**, 9–12.
2 M. Okano, S. Xie and E. Li, *Nat. Genet.*, 1998, **19**, 219–220.
3 S. Xie, Z. Wang, M. Okano, M. Nogami, Y. Li, W.-W. He, K. Okumura and E. Li, *Gene*, 1999, **236**, 87–95.
4 G. R. Wyatt, *Nature*, 1950, **166**, 237–238.
5 M. Ehrlich and R. Y. Wang, *Science*, 1981, **212**, 1350–1357.
6 M. Ehrlich, M. A. Gama-Sosa, L. H. Huang, R. M. Midgett, K. C. Kuo, R. A. McCune and C. Gehrke, *Nucleic Acids Res.*, 1982, **10**, 2709–2721.
7 A. P. Bird, *Nature*, 1986, **321**, 209–213.
8 R. Lister and J. R. Ecker, *Genome Res.*, 2009, **19**, 959–966.
9 T. Mohandas, R. S. Sparkes and L. J. Shapiro, *Science*, 1981, **211**, 393–396.
10 J. L. Swain, T. A. Stewart and P. Leder, *Cell*, 1987, **50**, 719–727.
11 W. Reik, A. Collick, M. L. Norris, S. C. Barton and M. A. Surani, *Nature*, 1987, **328**, 248–251.
12 E. Li, C. Beard and R. Jaenisch, *Nature*, 1993, **366**, 362–365.
13 A. P. Wolffe and M. A. Matzke, *Science*, 1999, **286**, 481–486.

14 P. A. Jones and D. Takai, *Science*, 2001, **293**, 1068–1070.
15 P. A. Jones, *Nat. Rev. Genet.*, 2012, **13**, 484–492.
16 D. P. Barlow and M. S. Bartolomei, *Cold Spring Harb Perspect Biol*, 2014, **6**, a018382.
17 M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind and A. Rao, *Science*, 2009, **324**, 930–935.
18 S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He and Y. Zhang, *Science*, 2011, **333**, 1300–1303.
19 C. Blanchet, M. Pasi, K. Zakrzewska and R. Lavery, *Nucleic Acids Res.*, 2011, **39**, W68–73.
20 R. M. Kohli and Y. Zhang, *Nature*, 2013, **502**, 472–479.
21 M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell and S. Balasubramanian, *Nat. Chem.*, 2014, **6**, 1049–1055.
22 M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell and S. Balasubramanian, *Nat. Chem. Biol.*, 2015, **11**, 555–557.
23 T. Carell, M. Q. Kurz, M. Muller, M. Rossa and F. Spada, *Angew. Chem., Int. Ed.*, 2018, **57**, 4296–4312.
24 J. S. Choy, S. Wei, J. Y. Lee, S. Tan, S. Chu and T. H. Lee, *J. Am. Chem. Soc.*, 2010, **132**, 1782–1783.
25 M. Iurlaro, G. Ficz, D. Oxley, E. A. Raiber, M. Bachman, M. J. Booth, S. Andrews, S. Balasubramanian and W. Reik, *Genome Biol.*, 2013, **14**, R119.
26 C. X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S. Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G. L. Xu, P. Jin and C. He, *Cell*, 2013, **153**, 678–691.
27 E. A. Raiber, D. Beraldi, G. Ficz, H. E. Burgess, M. R. Branco, P. Murat, D. Oxley, M. J. Booth, W. Reik and S. Balasubramanian, *Genome Biol.*, 2012, **13**, R69.
28 F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. Dal Bello and S. Oliviero, *Cell Rep.*, 2015, **10**, 674–683.
29 D. Ji, C. You, P. Wang and Y. Wang, *Chem. Res. Toxicol.*, 2014, **27**, 1304–1309.
30 M. W. Kellinger, C. X. Song, J. Chong, X. Y. Lu, C. He and D. Wang, *Nat. Struct. Mol. Biol.*, 2012, **19**, 831–833.
31 C. O'Neill, *Animal Front.*, 2015, **5**, 42–49.
32 T. M. Storebjerg, S. H. Strand, S. Hoyer, A. S. Lynnerup, M. Borre, T. F. Orntoft and K. D. Sorensen, *Clin Epigenetics*, 2018, **10**, 105.
33 D. Renciuk, O. Blacque, M. Vorlickova and B. Spingler, *Nucleic Acids Res.*, 2013, **41**, 9891–9900.
34 L. Lercher, M. A. McDonough, A. H. El-Sagheer, A. Thalhammer, S. Kriaucionis, T. Brown and C. J. Schofield, *Chem. Commun.*, 2014, **50**, 1794–1796.
35 M. W. Szulik, P. S. Pallan, B. Nocek, M. Voehler, S. Banerjee, S. Brooks, A. Joachimiak, M. Egli, B. F. Eichman and M. P. Stone, *Biochemistry*, 2015, **54**, 1294–1305.
36 J. M. Vargason, B. F. Eichman and P. S. Ho, *Nat. Struct. Biol.*, 2000, **7**, 758–761.
37 E. A. Raiber, P. Murat, D. Y. Chirgadze, D. Beraldi, B. F. Luisi and S. Balasubramanian, *Nat. Struct. Mol. Biol.*, 2015, **22**, 44–49.
38 K. Krawczyk, S. Demharter, B. Knapp, C. M. Deane and P. Minary, *Bioinformatics*, 2018, **34**, 41–48.
39 J. S. Hardwick, D. Ptchelkine, A. H. El-Sagheer, I. Tear, D. Singleton, S. E. V. Phillips, A. N. Lane and T. Brown, *Nat. Struct. Mol. Biol.*, 2017, **24**, 544–552.
40 T. T. Ngo, J. Yoo, Q. Dai, Q. Zhang, C. He, A. Aksimentiev and T. Ha, *Nat. Commun.*, 2016, **7**, 10813.
41 P. J. Sanstead, B. Ashwood, Q. Dai, C. He and A. Tokmakoff, *J. Phys. Chem. B*, 2020, **124**, 1160–1174.
42 Q. Dai, P. J. Sanstead, C. S. Peng, D. Han, C. He and A. Tokmakoff, *ACS Chem. Biol.*, 2016, **11**, 470–477.
43 D. Herschlag and M. M. Pinney, *Biochemistry*, 2018, **57**, 3338–3352.
44 R. C. A. Dubini, A. Schon, M. Muller, T. Carell and P. Rovo, *Nucleic Acids Res.*, 2020, **48**, 8796–8807.
45 Z. Wu, F. Delaglio, N. Tjandra, V. B. Zhurkin and A. Bax, *J. Biomol. NMR*, 2003, **26**, 297–315.
46 E. N. Nikolova, G. D. Bascom, I. Andricioaei and H. M. Al-Hashimi, *Biochemistry*, 2012, **51**, 8654–8664.
47 Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He and G. L. Xu, *Science*, 2011, **333**, 1303–1307.
48 J. Czernek, R. Fiala and V. Sklenar, *J. Magn. Reson.*, 2000, **145**, 142–146.
49 S. L. Lam and L. M. Chi, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2010, **56**, 289–310.
50 J. M. Fonville, M. Swart, Z. Vokacova, V. Sychrovsky, J. E. Sponer, J. Sponer, C. W. Hilbers, F. M. Bickelhaupt and S. S. Wijmenga, *Chemistry*, 2012, **18**, 12372–12387.
51 S. S. Wijmenga and B. N. M. van Buuren, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1998, **32**, 287–387.
52 S. Nozinovic, P. Gupta, B. Furtig, C. Richter, S. Tullmann, E. Duchardt-Ferner, M. C. Holthausen and H. Schwalbe, *Angew. Chem., Int. Ed.*, 2011, **50**, 5397–5400.
53 M. R. Hansen, L. Mueller and A. Pardi, *Nat. Struct. Biol.*, 1998, **5**, 1065–1074.
54 A. Vermeulen, H. Zhou and A. Pardi, *J. Am. Chem. Soc.*, 2000, **122**, 9638–9647.
55 D. MacDonald, K. Herbert, X. Zhang, T. Pologruto and P. Lu, *J. Mol. Biol.*, 2001, **306**, 1081–1098.
56 Z. Wu, M. Maderia, J. J. Barchi, Jr., V. E. Marquez and A. Bax, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 24–28.
57 B. Sathyamoorthy, H. Shi, H. Zhou, Y. Xue, A. Rangadurai, D. K. Merriman and H. M. Al-Hashimi, *Nucleic Acids Res.*, 2017, **45**, 5586–5601.
58 B. Sathyamoorthy, R. K. R. Sannapureddi, D. Negi and P. Singh, *J Magn Reson Open*, 2022, **10–11**, 100035.
59 J. Jerbi and M. Springborg, *J. Comput. Chem.*, 2017, **38**, 1049–1056.
60 M. Munzel, U. Lischke, D. Stathis, T. Pfaffeneder, F. A. Gnerlich, C. A. Deiml, S. C. Koch, K. Karaghiosoff and T. Carell, *Chemistry*, 2011, **17**, 13782–13788.
61 H. Hashimoto, Y. O. Olanrewaju, Y. Zheng, G. G. Wilson, X. Zhang and X. Cheng, *Genes Dev.*, 2014, **28**, 2304–2313.
62 K. L. Greene, Y. Wang and D. Live, *J. Biomol. NMR*, 1995, **5**, 333–338.

63 F. J. Van de Ven and C. W. Hilbers, *Eur. J. Biochem.*, 1988, **178**, 1–38.
64 R. V. Hosur, G. Govil and H. T. Miles, *Magn. Reson. Chem.*, 1988, **26**, 927–944.
65 L. J. Rinkel, M. R. Sanderson, G. A. van der Marel, J. H. van Boom and C. Altona, *Eur. J. Biochem.*, 1986, **159**, 85–93.
66 A. S. Serianni, J. Wu and I. Carmichael, *J. Am. Chem. Soc.*, 2002, **117**, 8645–8650.
67 J. T. Fischer and U. M. Reinscheid, *Eur. J. Org. Chem.*, 2006, 2074–2080.
68 N. Foloppe and A. D. MacKerell, *Biophys. J.*, 1999, **76**, 3206–3218.
69 B. Schneider, Z. Moravek and H. M. Berman, *Nucleic Acids Res.*, 2004, **32**, 1666–1677.
70 J. S. Richardson, B. Schneider, L. W. Murray, G. J. Kapral, R. M. Immormino, J. J. Headd, D. C. Richardson, D. Ham, E. Hershkovits, L. D. Williams, K. S. Keating, A. M. Pyle, D. Micallef, J. Westbrook, H. M. Berman and R. N. A. O. Consortium, *RNA*, 2008, **14**, 465–481.
71 A. Barbic, D. P. Zimmer and D. M. Crothers, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 2369–2373.
72 K. McAteer, A. Aceves-Gaona, R. Michalczyk, G. W. Buchko, N. G. Isern, L. A. Silks, J. H. Miller and M. A. Kennedy, *Biopolymers*, 2004, **75**, 497–511.
73 R. Stefl, H. Wu, S. Ravindranathan, V. Sklenar and J. Feigon, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 1177–1182.
74 J. Feigon, J. M. Wright, W. Leupin, W. A. Denny and D. R. Kearns, *J. Am. Chem. Soc.*, 1982, **104**, 5540–5541.
75 D. R. Hare, D. E. Wemmer, S.-H. Chou, G. Drobny and B. R. Reid, *J. Mol. Biol.*, 1983, **171**, 319–336.
76 M. A. Weiss, D. J. Patel, R. T. Sauer and M. Karplus, *Proc. Natl. Acad. Sci. U. S. A.*, 1984, **81**, 130–134.
77 C. Schwieters, J. Kuszewski and G. Mariusclore, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2006, **48**, 47–62.
78 X. J. Lu and W. K. Olson, *Nat. Protoc.*, 2008, **3**, 1213–1227.
79 A. Thalhammer, A. S. Hansen, A. H. El-Sagheer, T. Brown and C. J. Schofield, *Chem. Commun.*, 2011, **47**, 5325–5327.
80 R. Wang, Z. Luo, K. He, M. O. Delaney, D. Chen and J. Sheng, *Nucleic Acids Res.*, 2016, **44**, 4968–4977.
81 E. P. Wright, M. A. S. Abdelhamid, M. O. Ehiabor, M. C. Grigg, K. Irving, N. M. Smith and Z. A. E. Waller, *Nucleic Acids Res.*, 2020, **48**, 55–62.
82 R. C. A. Dubini, E. Korytiakova, T. Schinkel, P. Heinrichs, T. Carell and P. Rovo, *ACS Phys Chem Au*, 2022, **2**, 237–246.
83 N. Karino, Y. Ueno and A. Matsuda, *Nucleic Acids Res.*, 2001, **29**, 2456–2463.
84 E. N. Nikolova, E. Kim, A. A. Wise, P. J. O'Brien, I. Andricioaei and H. M. Al-Hashimi, *Nature*, 2011, **470**, 498–502.
85 H. S. Alvey, F. L. Gottardo, E. N. Nikolova and H. M. Al-Hashimi, *Nat. Commun.*, 2014, **5**, 4786.
86 A. L. Stelling, A. Y. Liu, W. Zeng, R. Salinas, M. A. Schumacher and H. M. Al-Hashimi, *Angew. Chem., Int. Ed.*, 2019, **58**, 12010–12013.
87 H. Zhou, B. Sathyamoorthy, A. Stelling, Y. Xu, Y. Xue, Y. Z. Pigli, D. A. Case, P. A. Rice and H. M. Al-Hashimi, *Biochemistry*, 2019, **58**, 1963–1974.
88 Y. Xu, A. Manghrani, B. Liu, H. Shi, U. Pham, A. Liu and H. M. Al-Hashimi, *J. Biol. Chem.*, 2020, **295**, 15933–15947.
89 M. Banyay, M. Sarkar and A. Gräslund, *Biophys. Chem.*, 2003, **104**, 477–488.
90 I. J. Kimsey, K. Petzold, B. Sathyamoorthy, Z. W. Stein and H. M. Al-Hashimi, *Nature*, 2015, **519**, 315–320.
91 E. S. Szymanski, I. J. Kimsey and H. M. Al-Hashimi, *J. Am. Chem. Soc.*, 2017, **139**, 4326–4329.
92 R. Rohs, S. M. West, A. Sosinsky, P. Liu, R. S. Mann and B. Honig, *Nature*, 2009, **461**, 1248–1253.
93 E. N. Nikolova, H. Zhou, F. L. Gottardo, H. S. Alvey, I. J. Kimsey and H. M. Al-Hashimi, *Biopolymers*, 2013, **99**, 955–968.
94 A. Perez, F. J. Luque and M. Orozco, *J. Am. Chem. Soc.*, 2007, **129**, 14739–14745.
95 A. Saran, D. Perahia and B. Pullman, *Theor. Chim. Acta*, 1973, **30**, 31–44.
96 N. Foloppe and A. D. MacKerell, *J. Phys. Chem. B*, 1998, **102**, 6669–6678.
97 W. K. Olson, *J. Am. Chem. Soc.*, 2002, **104**, 278–286.
98 M. A. el Hassan and C. R. Calladine, *J. Mol. Biol.*, 1996, **259**, 95–103.
99 A. M. Deaton and A. Bird, *Genes Dev.*, 2011, **25**, 1010–1022.
100 O. Yildirim, R. Li, J. H. Hung, P. B. Chen, X. Dong, L. S. Ee, Z. Weng, O. J. Rando and T. G. Fazzio, *Cell*, 2011, **147**, 1498–1510.
101 M. Mellen, P. Ayata, S. Dewell, S. Kriaucionis and N. Heintz, *Cell*, 2012, **151**, 1417–1430.
102 C. Rausch, F. D. Hastert and M. C. Cardoso, *J. Mol. Biol.*, 2019, **432**(6), 1731–1746.
103 K. Sato, K. Kawamoto, S. Shimamura, S. Ichikawa and A. Matsuda, *Bioorg. Med. Chem. Lett.*, 2016, **26**, 5395–5398.
104 F. Li, Y. Zhang, J. Bai, M. M. Greenberg, Z. Xi and C. Zhou, *J. Am. Chem. Soc.*, 2017, **139**, 10617–10620.
105 A. Afek, H. Shi, A. Rangadurai, H. Sahay, A. Senitzki, S. Xhani, M. Fang, R. Salinas, Z. Mielko, M. A. Pufall, G. M. K. Poon, T. E. Haran, M. A. Schumacher, H. M. Al-Hashimi and R. Gordan, *Nature*, 2020, **587**, 291–296.
106 W. Yang, *Cell Res.*, 2008, **18**, 184–197.
107 A. A. Tanpure and S. Balasubramanian, *ChemBioChem*, 2017, **18**, 2236–2241.
108 G. M. Clore, M. R. Starich and A. M. Gronenborn, *J. Am. Chem. Soc.*, 1998, **120**, 10571–10572.
109 F. L. Zott, V. Korotenko and H. Zipse, *ChemBioChem*, 2022, **23**, e202100651.
110 S. G. Hyberts, K. Takeuchi and G. Wagner, *J. Am. Chem. Soc.*, 2010, **132**, 2145–2147.
111 J. Cavanagh, N. Skelton, W. Fairbrother, M. Rance and I. Palmer, Arthur, *Protein NMR Spectroscopy*, Academic Press, 2006.
112 J. Farjon, J. Boisbouvier, P. Schanda, A. Pardi, J. P. Simorre and B. Brutscher, *J. Am. Chem. Soc.*, 2009, **131**, 8571–8577.
113 B. Sathyamoorthy, J. Lee, I. Kimsey, L. R. Ganser and H. Al-Hashimi, *J. Biomol. NMR*, 2014, **60**, 77–83.

114 K. Kazimierczuk and V. Y. Orekhov, *Angew. Chem., Int. Ed.*, 2011, **50**, 5556–5559.
115 F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax, *J. Biomol. NMR*, 1995, **6**, 277–293.
116 W. Lee, M. Tonelli and J. L. Markley, *Bioinformatics*, 2015, **31**, 1325–1327.
117 J. D. Baleja, M. W. Germann, J. H. van de Sande and B. D. Sykes, *J. Mol. Biol.*, 1990, **215**, 411–428.

