Base pairing and structural insights into the 5-formylcytosine in RNA duplex

Rui Wang[1,2,†], Zhipu Luo[3,†], Kaizhang He[4], Michael O. Delaney[4], Doris Chen[1,2] and Jia Sheng[1,2,*]

[1]Department of Chemistry, University at Albany, State University of New York, Albany, NY 12222, USA, [2]The RNA Institute, University at Albany, State University of New York, Albany, NY 12222, USA, [3]Synchrotron Radiation Research Section, MCL National Cancer Institute, Argonne National Laboratory, Argonne, IL, 60439, USA and [4]Dharmacon, GE Healthcare, Lafayette, CO 80026, USA

ABSTRACT

5-Formylcytidine (f[5]C), a previously discovered natural nucleotide in the mitochondrial tRNA of many species including human, has been recently detected as the oxidative product of 5-methylcytidine (m[5]C) through 5-hydroxymethylcytidine (hm[5]C) in total RNA of mammalian cells. The discovery indicated that these cytosine derivatives in RNA might also play important epigenetic roles similar as in DNA, which has been intensively investigated in the past few years. In this paper, we studied the base pairing specificity of f[5]C in different RNA duplex contexts. We found that the 5-formyl group could increase duplex thermal stability and enhance base pairing specificity. We present three high-resolution crystal structures of an octamer RNA duplex [5′-GUA(f[5]C)GUAC-3′]$_2$ that have been solved under three crystallization conditions with different buffers and pH values. Our results showed that the 5-formyl group is located in the same plane as the cytosine base and forms an intra-residue hydrogen bond with the amino group in the N4 position. In addition, this modification increases the base stacking between the f[5]C and the neighboring bases while not causing significant global and local structure perturbations. This work provides insights into the effects of 5-formylcytosine on RNA duplex.

INTRODUCTION

RNA is involved in numerous biochemical and cellular processes as a carrier of genetic information, an adapter in protein synthesis, a structural scaffold in subcellular organelles and a functional catalyst and regulator of biochemical reactions (1–5). Over 140 naturally occurring chemical modifications have been identified in mRNA, rRNA, tRNA and other non-coding RNAs (6). Many of these modifications have been shown to play critical roles in maintaining and diversifying RNA structures and functions. Although it had been assumed that these modifications were stable and static, there is evidence that some of these posttranscriptional RNA modifications, similar to those DNA and protein epigenetic markers, can be very dynamic with unexpected regulatory and/or signaling functions. For example, N[6]-methyladenosine (m[6]A), the most common internal modification in eukaryotic mRNAs and some long non-coding RNAs, appears to be involved in epigenetic inheritance. m[6]A can undergo oxidative demethylation catalyzed by dioxygenases including FTO, AlkBH1, AlkBH3 and AlkBH5 (7–11). This methylation-demethylation process is dynamic and may be important in RNA structural switches, protein recognition, mRNA stability, miRNA processing, protein translation and stem cell differentiation (12,13).

Another potential 'RNA epigenetic' marker is the 5-methylcytosine (m[5]C), which was first discovered over a half century ago in cellular RNAs across all three domains of life (14–21). The structures and epigenetic functions of m[5]C and oxidative analogues 5-hydroxymethyl-, 5-formyl- and 5-carboxyl-cytosines (hm[5]C, f[5]C and ca[5]C, respectively) in DNA have been extensively studied in the past few years (22,23). In comparison, their corresponding investigations in RNA contexts are left far behind. It is known that the 5-methylation of cytosine in long non-coding RNAs affects the binding of chromatin-modifying complexes and is thought to regulate cellular epigenetic status (20). It was also demonstrated that the m[5]C is subject to *in vivo* oxidation to generate hm[5]C and f[5]C in total RNA from all domains of life and in polyA-enriched RNA fractions from mammalian cells (24). In addition, the *in vitro* formation of hm[5]C from m[5]C can be mediated by TET proteins (25), the 10-11 translocation family of enzymes that are

*To whom correspondence should be addressed. Tel: +1 518 437 4419; Fax: +1 518 442 3462; Email: jsheng@albany.edu
† These authors contributed equally to the paper as first authors.

the major players in the m^5C oxidative metabolism in DNA. Very recently, the transcriptome-wide distribution of hm^5C has been investigated in *Drosophila* and the RNA hydroxymethylation was found to favor mRNA translation (26). All these discoveries support the hypothesis that the transformative m^5C could serve as an RNA epigenetic marker. Therefore, studying the structures of its oxidative intermediates (Scheme 1) in different RNA contexts will help elucidate the biological roles of these naturally occurring modifications in RNA functions and regulation.

In the tRNAs of many species, including humans, f^5C is observed at the wobble position of the anticodon loop of mitochondrial methionine tRNA (27–31). This modification increases the dynamics of loop residues and affords the single $tRNA^{Met}$ the ability to decode both AUG and AUA in translational initiation and elongation sites of mRNA (32–34). Although the crystal structure of f^5C within the codon-anticodon complex in ribosomal A-site has been studied (35), a clear view of its local geometry remains elusive. In addition, the base recognition abilities and structural features of f^5C in RNA duplexes are also unknown. In this paper, we studied the base paring specificity of f^5C in different RNA duplex contexts and found that the 5-formyl modification could increase duplex stability relative to unmodified C and enhanced the discrimination for G over the other bases. We also present three high-resolution crystal structures of an octamer RNA duplex [5′-GUA(f^5C)GUAC-3′]$_2$ containing two consecutive f^5C:G pairs that have been solved in three different conditions. We found that the 5-formyl group does not cause significant local or global structural perturbations. The formyl group is located in the same plane as the cytosine base and forms an intra-residue hydrogen bond with the amino group at N4 position, increasing the base stacking between the f^5C and the neighboring bases. In addition, the f^5C interacts with neighboring residues through bridging water molecules.

MATERIALS AND METHODS

Synthesis and purification of f^5C-containing RNA oligonucleotides

f^5C-modified RNA sequences were synthesized on a 1-μmol scale using a MerMade 12 synthesizer through ACE chemistry procedures (36). The f^5C 2′-ACE-phosphoramidite (0.067 M in anhydrous acetonitrile), synthesized with the previously published method (32), was coupled to the growing polyribonucleotide chain with a coupling time of 3.5 min using 5-ethylthio-1H-tetrazole (0.5 M in anhydrous acetonitrile) as the activator. Once the synthesis of the polyribonucleotide chain was completed, the phosphate protecting groups were removed from the immobilized polyribonucleotide by treatment with disodium 2-carbamoyl-2-cyanoethylene-1,1-dithiolate-trihydrate in DMF for 15 min. The support was washed with water (5 ml), and the column was flushed with argon gas for 5 min to dry the support. The support was then transferred to a 2-ml Eppendorf tube, and the polyribonucleotide was cleaved from the support and the exocyclic amine protecting groups were removed by treatment with 1:3 (v/v) tert-butylamine:water for 6 h at 55°C. The sample

was cooled to room temperature, filtered and lyophilized to obtain the crude polyribonucleotide. The 2′-ACE group was removed with acetate/TEMED using the protocol recommended by Dharmacon. The oligonucleotides were desalted by ethanol precipitation and then purified by ion-exchange HPLC over a PA-100 column (Dionex). The oligonucleotides were eluted with a linear gradient of 0–35% buffer B (buffer A was pure water, and buffer B was 2 M ammonium acetate, pH 7.1) over 20 min at a flow rate of 1 ml/min. Collected fractions were lyophilized, desalted with Waters Sep-Pac C18 columns and re-concentrated. All the samples were verified by ESI-MS (mass spectra and sequences are shown in Supplementary Table S1 and Figures S1–3).

Thermal denaturation and CD experiments

Solutions of the duplex RNAs (0.5 μM) were prepared by dissolving the purified oligonucleotides in 10 mM sodium phosphate (pH 6.5) with 100 mM NaCl. The solutions were heated to 85°C for 3 min, then cooled slowly to room temperature, and stored at 4°C for overnight. Thermal denaturation was performed in a Cary 300 UV-Visible Spectrophotometer equipped with a temperature controller. The temperature reported is the block temperature. Each curve was acquired at 260 nm by heating and cooling from 5 to 85°C four times at a rate of 0.5°C/min. Experiments were repeated at least four times. Circular Dichroism (CD) studies were carried out in the same buffer utilizing a Jasco-815 CD spectrometer at room temperature in a quartz cell with a 10-mm path length. CD spectra were collected from 380 to 200 nm and with a scanning speed of 100 nm/min. The bandwidth was 1.0 nm, and the digital integration time was 1.0 s. All CD spectra were baseline-corrected for signal contributions due to the buffer.

Crystallization and diffraction data collection

RNA samples (0.5 mM duplex) were heated to 80°C for 3 min, cooled slowly to room temperature and placed at 4°C overnight before crystallization. Nucleic Acid Mini Screen Kits (Hampton Research), Natrix (Hampton Research) and Nuc-Pro-HTS (Jena Bioscience) were used to screen crystallization conditions at both 4 and 20°C using the hanging drop method. Perfluoropolyether was used as cryoprotectant for crystal mounting. Data was collected under a liquid nitrogen stream at −174°C. All diffraction data were collected at beam lines SER-CAT 22-BM at Advance Photon Source, Argonne National Laboratory, USA. A number of crystals that grew under different conditions were scanned, and the data were collected at a wavelength of 1.0 Å. All data were processed using HKL2000 and DENZO/SCALEPACK (37).

Structure determination and refinement

All of the three RNA structures presented here were solved by molecular replacement with PHASER using PDB structure 197D (the deoxy version of the same sequence) as the search model, followed by refinement using *REFMAC*. The refinement protocol included simulated annealing, positional refinement, restrained B-factor refinement and bulk
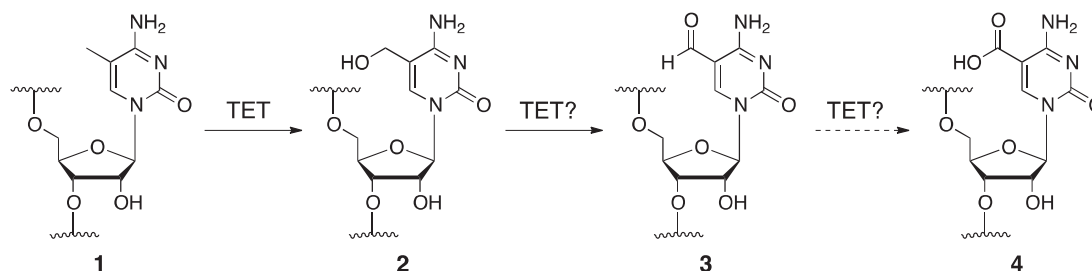
Scheme 1. Proposed oxidative metabolism of m^{5}C (**1**) through hm^{5}C (**2**), f^{5}C (**3**) and ca^{5}C (**4**). The hydroxylation of m^{5}C can be mediated by TET enzymes, which very likely could perform the following oxidation steps, similar to their functions in DNA.

solvent correction. The stereo-chemical topology and geometrical restraint parameters of DNA/RNA were applied (38). The topologies and parameters for 5-formylcytosine were constructed using Jligand (39). After several cycles of refinement, a number of highly ordered waters and metal ions were added. Anisotropic refinement was applied for high resolution data sets with space group $P2_1$ and $P3_2$. TLS refinement was carried out for dataset processed in $C2$ space group. Cross-validation (40) with a 5% test set was monitored during the refinement. The σA-weighted maps (41) of the (2m|Fo|-D|Fc|) and the difference (m|Fo|-D|Fc|) density maps were computed and used throughout the model building.

Quantum mechanical calculations

Quantum mechanical calculations of the 1-methyl-5-formyl cytosine were performed with the Gaussian09 package (42). Geometry optimization was carried out with 'tight' convergence criteria, which required that the maximum and average forces on an atom in the final iteration diminished to less than 1.5×10^{-5} and 1×10^{-5} Hartree/Bohr, respectively, and that the maximum and average displacements in the last two iterations were less than 6×10^{-5} and 4×10^{-5} Bohr, respectively. All density functional theory (DFT) calculations were performed using an 'ultra-fine' numerical integration grid, including 99 radial shells and 590 angular points per shell. DFT and second-order Møller-Plesset perturbation theory (MP2) were employed in conjunction with the 6–31G(d,p) basis set. For DFT calculations, the B3LYP (43–45) function was used. To account for solvation effects, all structures were fully optimized with the polarizable continuum model (PCM) (46,47). These calculations were designated as B3LYP/6-31G(d,p)/PCM and MP2/6-31G(d,p)/PCM, respectively. Harmonic vibrational frequency calculations at the same level verified that each structure was a minimum with zero imaginary frequency. Single point energy calculations were performed at the geometries optimized at the B3LYP/6-31G(d,p)/PCM level. The B3LYP was combined with the PCM in conjunction with the more complete 6–311++G(2d,2p) basis set. The zero point energy and thermal corrections obtained from the B3LYP/6-31G(d,p)/PCM level of theory were then added to the electronic energies to obtain free energies.

RESULTS AND DISCUSSION

Thermodynamic stabilities and spectroscopic properties of f^{5}C-containing RNA duplexes

We synthesized three sets of RNA oligonucleotides to investigate the thermodynamic stability and base pairing specificity of 5-formylcytosine-containing duplexes. As shown in Table 1 and Supplementary Figure S4A–E, duplexes containing the f^{5}C-G pair have significantly higher melting temperatures (T_m) than the native counterparts. In the 7-mer **duplex-1** system, the 5-formyl modification increased the T_m by 4.8°C relative to the duplex with only native Watson–Crick pairs (entry 2 versus 7). Consistent with this result, the T_m of the f^{5}C-modified 12-mer **duplex-2** was 4.1°C higher than the unmodified duplex (entry 12 versus 17). In the context of a self-complementary 8-mer (**duplex-3**), two consecutive f^{5}C-G pairs resulted in a 7.3°C increase in T_m relative to the unmodified duplex (entry 21 versus 22). Although a previous study found that the f^{5}C in the anticodon loop of tRNA^{Met} decreases the thermal stability of the codon-anticodon interaction due to the decreased base stacking relative to the unmodified tRNA (32), our results indicate that the f^{5}C modification might enhance the base stacking in RNA duplex context since the 5-formyl group is not directly involved in the hydrogen bonding interactions. This result is also consistent with a previous study of DNA duplexes modified with f^{5}C, in which a 5-formyl-2′-deoxycytosine-modified 25-mer duplex had a T_m 1°C higher than that of the unmodified duplex (48).

We next evaluated the mismatched pairing with f^{5}C. The T_ms of f^{5}C-U and f^{5}C-C mismatch-containing **duplex-1** sequences were similar to those of the native mismatched duplexes (entry 3 versus 8 and 4 versus 9). The f^{5}C-A mismatched duplex was significantly more stable than the duplex containing the C-A mismatch (entry 5 versus 10). In the **duplex-2** system, f^{5}C stabilized all the three mismatched duplexes relative to duplexes with mismatches with C (entry 13 versus 18, 14 versus 19, and 15 versus 20). These data support our hypothesis that the f^{5}C has more favorable base stacking than cytosine, balancing the lack of optimal hydrogen bonding of these mispairs. Interestingly, the f^{5}C-G pairs have overall better discrimination against all the other mispairs when f^{5}C-G is compared to other mispairs. For example, the T_m of the f^{5}C-G-containing **duplex-2** is 23.3°C higher than that of the f^{5}C-A-containing duplex (entry 7 versus 10), whereas the difference between T_ms of the C-G- and C-A-containing duplexes is 21.9°C (entry 2 versus 5).

Table 1. Melting temperatures of native and f⁵C-modified RNA duplexes

	Entry	Sequences	Base pair	T_m (°C)[a]	ΔT_m (°C)[b]
Duplex-1	**1**	**I**: 5′-UAG**C**UCC-3′			
	2	**I** + 3′-AUC**G**AGG-5′	C:G	38.1	
	3	**I** + 3′-AUC**U**AGG-5′	C:U	23.9	−14.2
	4	**I** + 3′-AUC**C**AGG-5′	C:C	18.1	−20.0
	5	**I** + 3′-AUC**A**AGG-5′	C:A	16.2	−21.9
	6	**II**: 5′-UAGf**⁵C**UCC-3′			
	7	**II** + 3′-AUC**G**AGG-5′	f⁵C:G	42.9	**+4.8**
	8	**II** + 3′-AUC**U**AGG-5′	f⁵C:U	24.5	−13.6
	9	**II** + 3′-AUC**C**AGG-5′	f⁵C:C	18.6	−19.5
	10	**II** + 3′-AUC**A**AGG-5′	f⁵C:A	19.6	−18.5
Duplex-2	**11**	**III**: 5′-GGACU**C**CUGUAG -3′			
	12	**III** + 3′-CCUGA**G**GACAUC -3′	C:G	63.5	
	13	**III** + 3′-CCUGA**U**GACAUC -3′	C:U	46.9	−17.6
	14	**III** + 3′-CCUGA**C**GACAUC -3′	C:C	47.3	−17.2
	15	**III** + 3′-CCUGA**A**GACAUC -3′	C:A	48.7	−15.8
	16	**IV**: 5′-GGACUf**⁵C**CUGUAG -3′			
	17	**IV** + 3′-CCUGA**G**GACAUC -3′	f⁵C:G	67.6	**+4.1**
	18	**IV** + 3′-CCUGA**U**GACAUC -3′	f⁵C:U	49.8	−13.7
	19	**IV** + 3′-CCUGA**C**GACAUC -3′	f⁵C:C	49.1	−14.4
	20	**IV** + 3′-CCUGA**A**GACAUC -3′	f⁵C:A	52.0	−11.5
Duplex-3	21	**V**: (5′-GUA**C**GUAC-3′)₂	C:G	35.3	
	22	**VI**: (5′-GUAf**⁵C**GUAC-3′)₂	f⁵C:G	42.6	**+7.3**

[a] T_ms were measured in 10 mM sodium phosphate (pH 7.0) containing 100 mM NaCl. T_m values reported are the averages of four measurements.
[b] ΔT_m values are relative to the duplexes with only Watson–Crick pairs.

This demonstrates higher RNA base pairing specificity in the presence of 5-formyl group.

We also compared the circular dichroism spectra of RNA duplexes containing C-G and f⁵C-G pairs. As shown in Supplementary Figure S4F and G, unmodified **duplex-1** and **duplex-2** show characteristic peaks of the A-form helical conformation in solution: a strong positive peak around 270 nm and a weak negative peak around 240 nm, typical peaks of RNA duplexes with mostly GC pairs (49,50). The f⁵C-modified 12-mer duplex has a spectrum similarly to that of the native duplex (Supplementary Figure S4G), whereas the presence of f⁵C in the 7-mer **duplex-1** down-shifts the positive peak by about 10 nm and eliminated the negative peak at 240 nm (Supplementary Figure S4F). Both duplex and hairpin structures are present in the solution of self-complementary 8-mer **duplex-3** (51), and a down-field shift of the strong positive peak was observed in the spectrum of the modified oligonucleotide compared to the unmodified one (Supplementary Figure S4H). The CD spectra of nucleic acids are affected by many factors including sequence, base stacking and overall conformation; given this, our data indicate that the 5-formyl modification does not cause gross perturbations in folding but does have some impact on double helix parameters. This is in agreement with conclusions from a previous study of f⁵C (48) and the m⁵C, hm⁵C modified DNA duplexes (52).

Overall crystal structures of f⁵C RNA duplex

To investigate the structural features of f⁵C RNA duplexes and to explore the mechanisms of the enhancement of duplex stability due to the 5-formyl group, we obtained three crystal structures for the f⁵C-containing self-complementary **duplex-3**, a purine-pyrimidine alternating octamer [5′-GUA(f⁵C)GUAC-3′]₂, under three different

buffer and pH conditions. The detailed crystallization conditions and data collection and structure refinement statistics are summarized in Table 2. All the structures were solved by molecular replacement using a DNA duplex with the same sequence (PDB ID: 197D) as the search model (53). In HEPES buffer at pH 7, the RNA molecules crystallized in the space group of $P2_1$. In Tris buffer at pH 7.5, the space group was $P3_2$, and in sodium cacodylate at pH 6.0, the space group was $C2$. To determine whether the 5-formyl group is involved in molecular packing, we investigated the helix-helix interactions in each asymmetric unit. As shown in Figure 1A and C, the duplexes in the $C2$ and $P2_1$ space groups are stacked in head-to-tail fashion with the G1:C8 pair stacked on the C8:G1 pair (Figure 1B), forming essentially continuous helices that pack side-by-side (Figure 1A). In the $P3_2$ structure, there are three duplexes in each asymmetric unit that pack along three helical axes at an angle of ∼100° (Figure 1D), forming a 'bowl' shaped structure. Further analysis indicates that although the 5-formyl group is not directly involved in the duplex–duplex interactions, the 2′-hydroxyl of the f⁵C residue hydrogen bonds with a phosphate oxygen in the A3 residue of one adjacent duplex (Figure 1E).

In the whole unit cell level, the molecular packing in different space groups looks very similar (Supplementary Figure S5). Investigating the intermolecular interactions between duplexes shows uniform hydrogen-bonds networks. In $C2$ space group, there are very few intermolecular interactions observed at medium resolution (2.4 Å). For higher resolution (1.5 Å) data in $P3_2$ space group, very strong H-bond networks were formed by OP1 of f⁵C4 and O2′ and O3′ of A3 mediated by two water molecules which located around the 3-fold axis of the cell (Supplementary Figure S6A), which can be assigned into three parallel panels (Supplementary Figure S6B). Notably, it is possible that the top
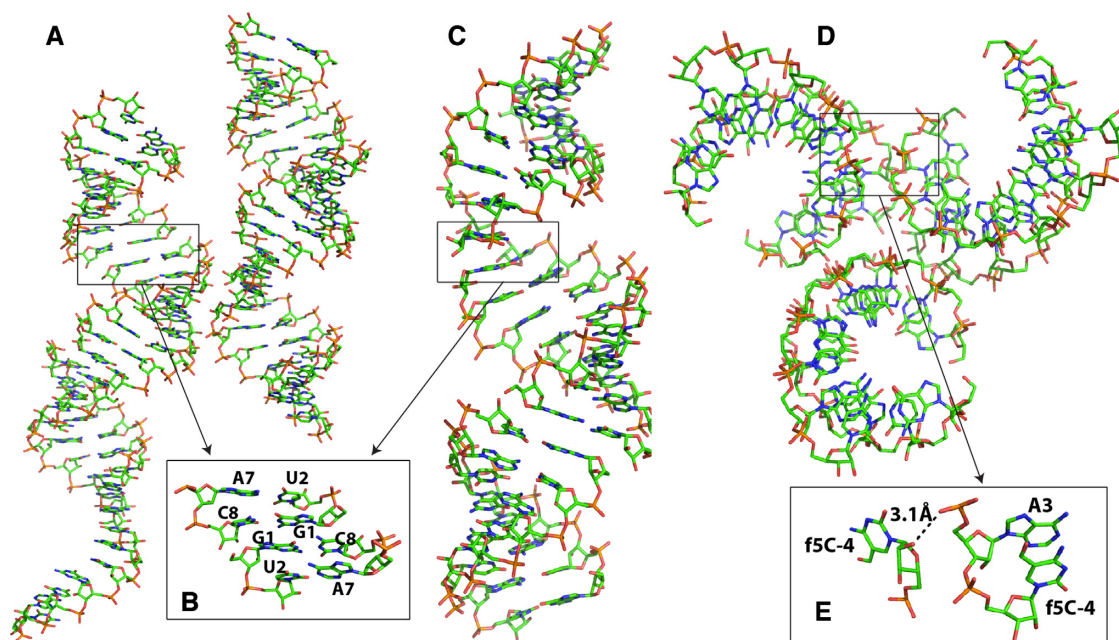
Figure 1. Molecular packing patterns of f^5C-RNA in asymmetric unit cells. (**A**) Side view of duplex stacking in one asymmetric unit with *C*2 space group. (**B**) Zoom in view of the duplex terminal junction showing head-to-tail stacking of two terminal C8:G1 pairs; this occurs in both *C*2 and *P*2$_1$ space group asymmetric units. (**C**) Side view of duplex stacking in one asymmetric unit with *P*2$_1$ space group. (**D**) Top view of duplex stacking in one asymmetric unit with *P*3$_2$ space group. (**E**) Zoom in view of interaction between by 2′ hydroxyl of f^5C and the phosphate oxygen of A3 in adjacent duplexes.

Table 2. X-ray data and structure refinement statistics of the f^5C-containing self-complementary **duplex-3** crystals grown under three different conditions

	Condition 1 (PDB ID: 5HNJ)	Condition 2 (PDB ID:5HN2)	Condition 3 (PDB ID:5HNQ)
Crystallization	50 mM Na•HEPES (pH 7.0), 50 mM MgSO$_4$, 1.6 M Li$_2$SO$_4$	10% MPD, 50 mM Tris•HCl (pH 7.5), 50 mM NH$_4$Ac, 10 mM MgCl$_2$	10% MPD, 40 mM sodium cacodylate(pH 6.0), 12 mM spermine tetrahydrochloride, 80 mM NaCl, 12 mM KCl, 20 mM MgCl$_2$
Scaling			
Space group	*P*2$_1$	*P*3$_2$	*C*2
Unit cell parameters, Å	30.7 × 45.9 × 45.4	42.3 × 42.3 × 58.6	138.9 × 44.3 × 50.6
Unit cell parameters, degrees	90, 100.9, 90	90, 90, 120	90, 102.8, 90
Resolution range, Å (last shell)	30–1.24 (1.28–1.24)	30–1.5 (1.55–1.5)	50–2.4 (2.49–2.4)
Unique reflections (last shell)	35 304 (3511)	18 577 (1829)	12 285 (1225)
Completeness, %	100 (100)	98 (96.7)	99.6 (99.2)
R$_{merge}$,%	4.6 (35.5)	5.8 (36.5)	4.9 (33.0)
<I/σ(I)>	35.4 (2.1)	35.9 (2.1)	22.3 (2.0)
Redundancy	7.2 (5.7)	5.3 (5.3)	3.7 (3.8)
Refinement			
Molecules per asymmetric unit	3 duplexes	3 duplexes	6 duplexes and 1 single strand
Resolution range, Å	25.2–1.24	22.9–1.5	28.7–2.4
Number of reflections	33 515	17 551	11 304
Completeness, %	99.9	98.1	99.7
R$_{work}$, %	13.4	12.3	22.1
R$_{free}$, %	15.5	14.0	26.1
Bond length rmsd, Å	0.020	0.022	0.016
Bond angle rmsd, degrees	2.2	2.5	2.4
Overall B-factor with water, Å^2	18.6	22.2	103.5

water molecule in Supplementary Figure S6B can be assigned to NH$_4^+$ ion (54) since we used ammonium acetate in the crystallization buffer and it is technically very difficult to differentiate the two residues based on the density map. The terminal O2′ and O3′ of chain A, C and E can also form H-bond networks by direct interaction or by water-mediated interactions (Supplementary Figure S6C and D). In addition, eight water molecules were also observed to form in line mediating H-bond networks between chain A, B, E and F (Supplementary Figure S7). At much higher resolution (1.24 Å) in *P*2$_1$ space group, the terminal O2′ and O3′ in chain A, C and E can form very strong H-bond networks without any water molecules (Supplementary Figure S6E). These different inter-strand H-bond networks might have resulted in different space groups and further confirm that the 5-formyl group is not directly interacting with other residues.

Conformation of f⁵C in the RNA duplex

The root mean square deviations (r.m.s.d) between the three structures are less than 0.4 Å (Figure 2A), and the conformations of the f⁵C:G pairs from the three structures are almost identically superimposable (Figure 2B). This implies that the f⁵C has a very similar conformation over a pH range from 6.0 to 7.5. The density map of the f⁵C clearly shows that the 5-formyl group is located in the same plane as the cytosine base, with the bond of the carbonyl group parallel to the C4-N4 bond (Figure 2C); this effectively expands the conjugation system of cytosine ring. This structural feature is different from the geometry of f⁵C observed in tRNA codon-anticodon interactions (35), in which the carbonyl group turns to its 5′-phosphate oxygen and forms a dihedral angle of ∼60° with the cytosine plane (Figure 2D), although the density map of 5-formyl group in this structure is not clear due to the low resolution of the overall complex structure. The different 5-formyl conformations might also be attributed to the presence of ribosomal binding, which results in different structure context than duplex. Notably, the short distance between the formyl oxygen and the N4 atom (2.7 Å) indicates a strong hydrogen bonding interaction that might stabilize the observed f⁵C conformation.

Next, we calculated energies of the 5-formylcytosine base using both DFT and MP2 quantum mechanics methods. Both bond lengths (C5-C7 and C7-O7) and angles (C5-C7-O7) from the minimized 1-methyl-5-formylcytosine structures are consistent with those in the crystal structures (Supplementary Figure S8). Our calculations also indicate that the intramolecular hydrogen bonding contributes to the conformational stability of the f⁵C residue. As shown in Supplementary Figure S9, although the 5-formyl group is still coplanar with the cytosine ring after rotating for 180°, which stands for another relatively stable f⁵C conformation without the internal hydrogen bonding, this rotated conformation results in the overall free energy increase of 4.7 kcal/mol compared to the present one in the crystal structures.

We then compared the f⁵C-containing duplex structures to an ideal RNA structure model with the same sequence that was generated using the software Coot (55). The duplexes align well with an r.m.s.d of about 1.2 Å; major differences are localized in the two terminal base pairs (Figure 2E). The comparison of C-G pairs in the f⁵C-containing duplex with the ideal duplex showed slightly different extents of base-pairing buckle with almost identical hydrogen bonding strengths (Figure 2F). In addition, we also compared our f⁵C-containing duplex structure to the crystal structure of a DNA with the same sequence (PDB ID: 197D) and another RNA structure with similar sequence (5′-(GUAUAUAC)₂-3′; PDB ID: 246D). All duplexes and local C-G pairs have very similar structures (Supplementary Figure S10). The further comparison between our ribo-f⁵C residue and the deoxyribo-f⁵C from previously published DNA structures (PDB ID: 1VE8, 4QC7 and 4QKK) (56,57) also showed very high structural isomorphism with similar internal hydrogen bonding strength (Supplementary Figure S11), although the sugar puckers are different in A-form RNA and B-form DNA.

We further compared the geometric parameters of all the base pairs and base-pair steps in the f⁵C-containing duplex with the ones in the ideal RNA duplex using 3DNA software tools (58). The average values of these parameters are generally quite similar in the modified and native counterparts (Supplementary Tables S2 and 3). Figure 3 depicts the most characteristic and significant structural changes induced by the 5-formyl group. There were changes in buckle, opening, shift and tip, although it is possible that these parameters may also be partially impacted by the crystallization conditions. The f⁵C-G pair shows approximately a 5° positive buckle and a 2° negative opening (Figure 3A and B), different as the other pairs in both duplexes. The f⁵C-G pair involved steps show the most negative shifts of 0.2–0.3 Å and the most positive tip values (Figure 3C and D). Adjacent pairs compensate for these local effects resulting in similar average parameters for the modified and native duplexes. As the hydrogen bonding strength of f⁵C-G pair is very similar to that of the native C-G pair, the integrated local structural changes caused by the 5-formyl group together with the expanded planar system of f⁵C most likely increase the base-pair stacking interactions, resulting in enhanced thermal stability of the f⁵C-containing duplex. Indeed the f⁵C-involved steps have larger overlap areas compared to those in the unmodified duplex, especially in the region where the 5-formyl group lies on the top of the 5′ neighboring A3 residue (Table 3 and Supplementary Figure S12). As a result, the total overlap area of all base steps in f⁵C-containing duplex is 5 Å² more than that in the unmodified duplex.

Hydration of f⁵C in major groove

The formyl group is located in the major groove where it may affect hydration pattern and duplex hydrophobicity of RNAs, which influence their biochemical properties and enzyme recognition processes. Therefore, we investigated the hydration pattern along A3-f⁵C4-G5 and their interactions with waters in the major groove within the typical hydrogen bonding range of 2.8–3.4 Å. The locations of these waters in the structure of $P3_2$ space group are shown in Figure 4. A water molecule (W1) bridges from the 5-formyl oxygen to the phosphate backbone of the A3 residue. The additional water molecules, W2 and W3, that links the N7 of A3 further expand the interactions of f⁵C with the neighboring residues. There are two more water molecules (W4 and W5) that connect the N4 of f⁵C and the O6 of G5′, its pairing partner on the opposite strand. In addition, f⁵C also has interactions with its 3′ neighboring G5 residue through the networking of W4-W8-W9-W10, which are further bound to the two oxygen atoms in the phosphate backbone. Very similar hydration pattern in the $P2_1$ structure (Supplementary Figure S13) under different crystallization conditions is also observed, indicating this water-network is highly conserved and not pH-dependent. In addition, the f⁵C in DNA duplex (PDB ID: 4QKK) also shows a similar water-bridging network as showed in Supplementary Figure S14. These hydrogen bonds might pre-organize the conformation of the 5-formyl group and RNA backbone, rigidifying the single stranded RNA and reducing the en-
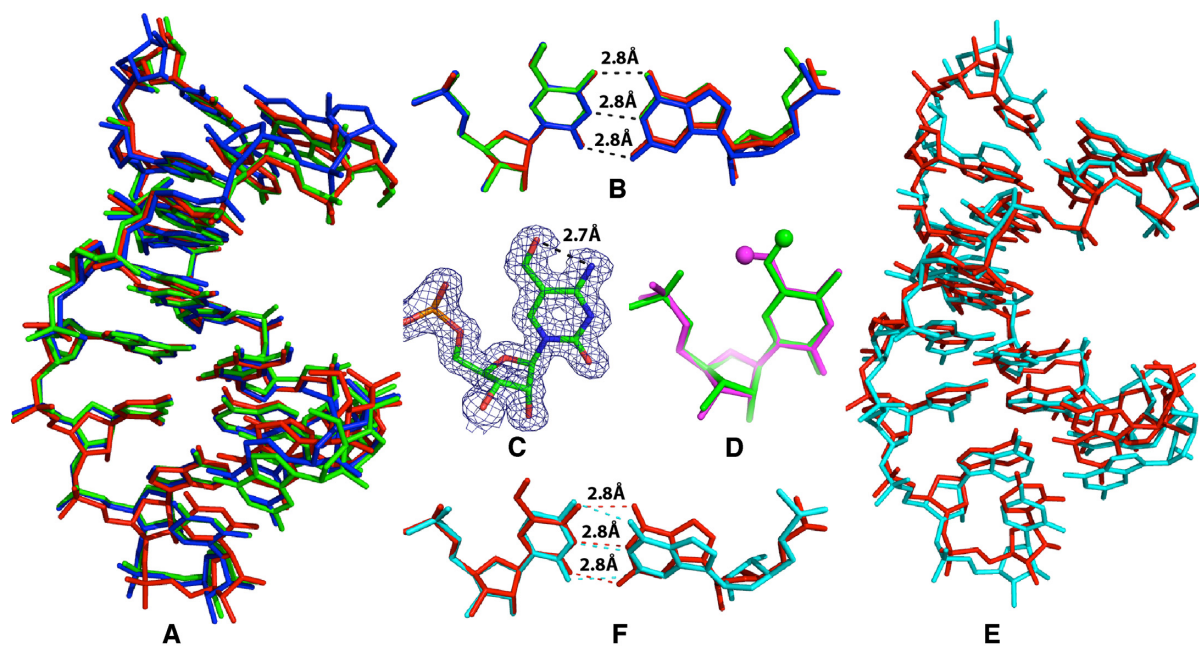
Figure 2. Overview and comparison of the structures obtained under different crystallization conditions. (**A** and **B**) Superimposed (A) duplexes and (B) f5C-G pairs from $P3_2$ (red), $P2_1$ (blue) and $C2$ (green) space groups. (**C**) Electron density map of f5C residue with the oxygen atom pointing up in the same plane as the cytosine (2Fo-Fc map with σ 1.0). (**D**) Comparison of f5C residue in $P3_2$ (green) and the one in the tRNA codon–anticodon interaction (magenta; PDB ID: 4GKK). The two spheres represent the 5-formyl oxygen atoms. (**E** and **F**) Superimposed (E) duplexes and (F) f5C-G pairs from $P3_2$ space group structure (red) and unmodified duplex structure.
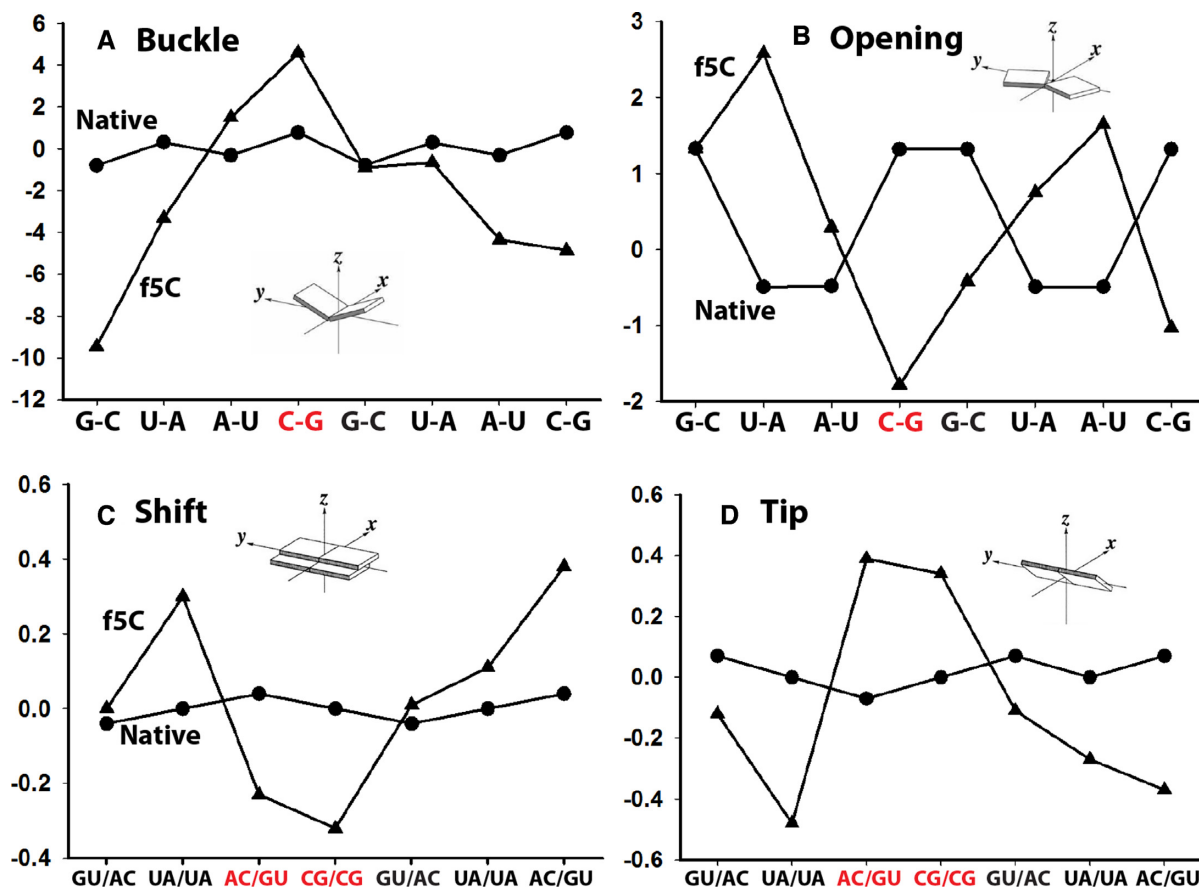

Figure 3. Local base pair and base-pair step parameters (**A**) buckle (°), (**B**) opening (°), (**C**) shift (Å) and (**D**) tip (Å) for f5C-containing (triangle) and native (circle) duplexes. The f5C involved base pairs are labeled in red.

Table 3. Overlap of base pair steps in native and f5C-containing duplexes

Entry	Steps	Total overlap area (Å^2)[a]	
		Native	f5C duplex
1	GU/AC	9.67	10.03
2	UA/UA	2.08	2.04
3	AC/GU	9.68	12.17
4	CG/CG	4.58	4.11
5	GU/AC	9.68	12.30
6	UA/UA	2.09	1.79
7	AC/GU	9.68	10.26
Overall		47.46	52.7

[a]The total overlap area includes both intra-strand and inter-strand overlap within the four bases of each base pair step.
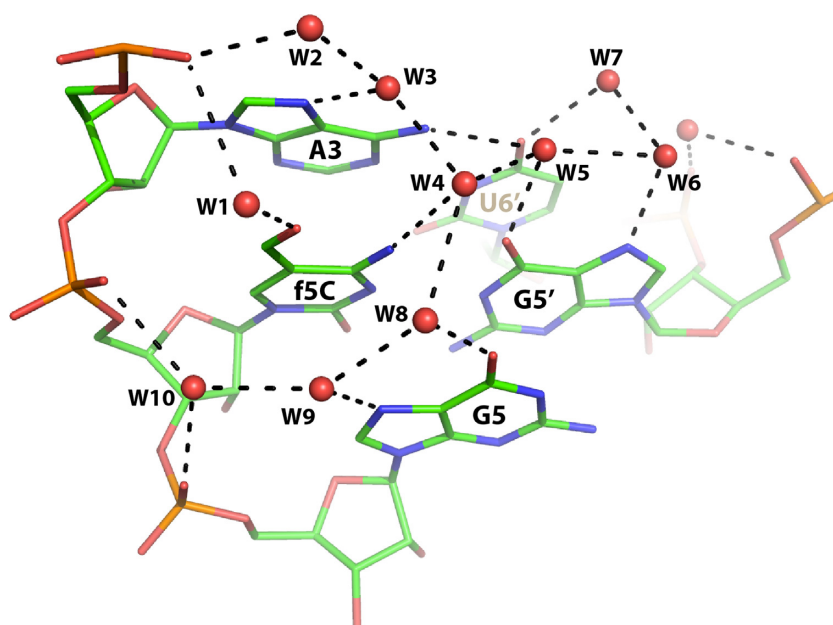
Figure 4. Local hydration pattern of A3-f5C4-G5 in the major groove of the duplex structure with the space group $P3_2$. The hydrogen bonds with water molecules within the range of 2.8–3.4 Å are represented by the dashed lines.

tropy cost during the duplex formation, therefore resulting in the higher stability of the f5C-modified duplex.

The solvation energies of the f5C-modified and native RNA duplexes were calculated using the 'solvate' module of the SEQMOL program. The f5C-containing RNA duplex had lower solvation energy and ~200 Å^2 less accessible surface area than the native duplex (Supplementary Table S4). This indicates that the 5-formyl group actually reduced hydration of the duplex probably by disrupting the water molecules in the narrow and deep major groove of A-form RNA duplex (59), despite the water molecules observed bound to the f5C residue. The decreased solvation might also contribute the duplex stability, as it has been shown to do for RNA duplexes modified with 2-thiouridine and with 2-thiocytosine (60,61).

CONCLUSION

In this work, we have studied the thermodynamic stability and base pairing specificity of f5C in different RNA duplexes. We found that relative to the natural cytosine, the 5-formyl modification increased duplex stability and en-

hanced the base pairing discrimination for C:G pair relative to other mispairs. Analysis of three high-resolution crystal structures of an f5C-modified octamer RNA duplex [5′-GUA(f5C)GUAC-3′]$_2$ obtained under different conditions indicated that the 5-formyl group does not cause significant overall or local structural perturbations. The formyl group is located in the same plane as cytosine and forms a pH-independent hydrogen bond with the N4 position, expanding the conjugation system of the ring and increasing the base stacking between the f5C and the neighboring bases. Although the 5-formyl group reduced the overall duplex solvation energy, the f5C interacts with its neighboring residues through bridging water molecules, which might further contribute to the duplex thermal stability and the biochemical roles of f5C in RNA. We speculate that the duplex-stabilizing water-mediated interactions between the 5-formyl group and neighboring residues together with the conformation of 5-formyl group influences its potential epigenetic roles, which are based on the recognition of f5C residue by enzymes that oxidize this modified nucleotide to ca5C. The identification of the working enzymes, the crystal structure of a ca5C-containing RNA duplex, and energy

required to catalyze the reaction from f^{5}C to ca^{5}C in the duplex context are currently under investigation.

ACCESSION NUMBERS

The three RNA structures have been deposited in Protein Data Bank (www.rcsb.org) with the PDB IDs: 5HNJ (*P2$_1$* duplex structure), 5HN2 (*P3$_3$* duplex structure) and 5HNQ (*C*2 duplex structure).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Fedor,M.J. (2009) Comparative enzymology and structural biology of RNA self-cleavage. *Annu. Rev. Biophys.*, **38**, 271–299.
2. Doudna,J.A. and Lorsch,J.R. (2005) Ribozyme catalysis: not different, just worse. *Nat. Struct. Mol. Biol.*, **12**, 395–402.
3. Henkin,T.M. (2008) Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev.*, **22**, 3383–3390.
4. Roth,A. and Breaker,R.R. (2009) The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.*, **78**, 305–334.
5. Montange,R.K., Mondragon,E., van Tyne,D., Garst,A.D., Ceres,P. and Batey,R.T. (2010) Discrimination between closely related cellular metabolites by the SAM-I riboswitch. *J. Mol. Biol.*, **396**, 761–772.
6. Machnicka,M.A., Milanowska,K., Osman Oglou,O., Purta,E., Kurkowska,M., Olchowik,A., Januszewski,W., Kalinowski,S., Dunin-Horkawicz,S., Rother,K.M. *et al.* (2013) MODOMICS: a database of RNA modification pathways–2013 update. *Nucleic Acids Res.*, **41**, D262–D267.
7. Jia,G., Fu,Y., Zhao,X., Dai,Q., Zheng,G., Yang,Y., Yi,C., Lindahl,T., Pan,T., Yang,Y.G. *et al.* (2011) N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.*, **7**, 885–887.
8. Zheng,G., Dahl,J.A., Niu,Y., Fedorcsak,P., Huang,C.M., Li,C.J., Vagbo,C.B., Shi,Y., Wang,W.L., Song,S.H. *et al.* (2013) ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell*, **49**, 18–29.
9. Yi,C., Yang,C.G. and He,C. (2009) A non-heme iron-mediated chemical demethylation in DNA and RNA. *Acc. Chem. Res.*, **42**, 519–529.
10. Trewick,S.C., Henshaw,T.F., Hausinger,R.P., Lindahl,T. and Sedgwick,B. (2002) Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature*, **419**, 174–178.
11. Falnes,P.O., Johansen,R.F. and Seeberg,E. (2002) AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature*, **419**, 178–182.
12. Fu,Y., Dominissini,D., Rechavi,G. and He,C. (2014) Gene expression regulation mediated through reversible m(6)A RNA methylation. *Nat. Rev. Genet.*, **15**, 293–306.
13. Wang,X., Lu,Z., Gomez,A., Hon,G.C., Yue,Y., Han,D., Fu,Y., Parisien,M., Dai,Q., Jia,G. *et al.* (2014) N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, **505**, 117–120.
14. Hotchkiss,R.D. (1948) The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J. Biol. Chem.*, **175**, 315–332.
15. Wyatt,G.R. (1950) Occurrence of 5-methylcytosine in nucleic acids. *Nature*, **166**, 237–238.
16. Fox,J.J., Yung,N., Davoll,J. and Brown,G.B. (1956) Pyrimidine nucleosides. I. A new route for the synthesis of thymine nucleosides1. *J. Am. Chem. Soc.*, **78**, 2117–2122.
17. Motorin,Y., Lyko,F. and Helm,M. (2010) 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. *Nucleic Acids Res.*, **38**, 1415–1430.
18. Squires,J.E., Patel,H.R., Nousch,M., Sibbritt,T., Humphreys,D.T., Parker,B.J., Suter,C.M. and Preiss,T. (2012) Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.*, **40**, 5023–5033.
19. Hussain,S., Sajini,A.A., Blanco,S., Dietmann,S., Lombard,P., Sugimoto,Y., Paramor,M., Gleeson,J.G., Odom,D.T., Ule,J. *et al.* (2013) NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep.*, **4**, 255–261.
20. Amort,T., Soulière,M.F., Wille,A., Jia,X.-Y., Fiegl,H., Wörle,H., Micura,R. and Lusser,A. (2013) Long non-coding RNAs as targets for cytosine methylation. *RNA Biol.*, **10**, 1002–1008.
21. Militello,K.T., Chen,L.M., Ackerman,S.E., Mandarano,A.H. and Valentine,E.L. (2014) A map of 5-methylcytosine residues in Trypanosoma brucei tRNA revealed by sodium bisulfite sequencing. *Mol. Biochem. Parasitol.*, **193**, 122–126.
22. Tahiliani,M., Koh,K.P., Shen,Y., Pastor,W.A., Bandukwala,H., Brudno,Y., Agarwal,S., Iyer,L.M., Liu,D.R., Aravind,L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
23. Kriaucionis,S. and Heintz,N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
24. Huber,S.M., van Delft,P., Mendil,L., Bachman,M., Smollett,K., Werner,F., Miska,E.A. and Balasubramanian,S. (2015) Formation and abundance of 5-hydroxymethylcytosine in RNA. *Chembiochem*, **16**, 752–755.
25. Fu,L., Guerrero,C.R., Zhong,N., Amato,N.J., Liu,Y., Liu,S., Cai,Q., Ji,D., Jin,S.G., Niedernhofer,L.J. *et al.* (2014) Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.*, **136**, 11582–11585.
26. Delatte,B., Wang,F., Ngoc,L.V., Collignon,E., Bonvin,E., Deplus,R., Calonne,E., Hassabi,B., Putmans,P., Awe,S. *et al.* (2016) Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science*, **351**, 282–285.
27. Moriya,J., Yokogawa,T., Wakita,K., Ueda,T., Nishikawa,K., Crain,P.F., Hashizume,T., Pomerantz,S.C., McCloskey,J.A., Kawai,G. *et al.* (1994) A novel modified nucleoside found at the first position of the anticodon of methionine tRNA from bovine liver mitochondria. *Biochemistry*, **33**, 2234–2239.
28. Watanabe,Y., Tsurui,H., Ueda,T., Furushima,R., Takamiya,S., Kita,K., Nishikawa,K. and Watanabe,K. (1994) Primary and higher order structures of nematode (Ascaris suum) mitochondrial tRNAs lacking either the T or D stem. *J. Biol. Chem.*, **269**, 22902–22906.
29. Tomita,K., Ueda,T. and Watanabe,K. (1997) 5-formylcytidine (f5C) found at the wobble position of the anticodon of squid mitochondrial tRNA(Met)CAU. *Nucleic Acids Symp. Ser.*, **37**, 197–198.
30. Takemoto,C., Ueda,T., Miura,K. and Watanabe,K. (1999) Nucleotide sequences of animal mitochondrial tRNAs(Met) possibly recognizing both AUG and AUA codons. *Nucleic Acids Symp. Ser.*, **42**, 77–78.
31. Tomita,K., Ueda,T., Ishiwa,S., Crain,P.F., McCloskey,J.A. and Watanabe,K. (1999) Codon reading patterns in Drosophila melanogaster mitochondria based on their tRNA sequences: a unique wobble rule in animal mitochondria. *Nucleic Acids Res.*, **27**, 4291–4297.

32. Lusic,H., Gustilo,E.M., Vendeix,F.A., Kaiser,R., Delaney,M.O., Graham,W.D., Moye,V.A., Cantara,W.A., Agris,P.F. and Deiters,A. (2008) Synthesis and investigation of the 5-formylcytidine modified, anticodon stem and loop of the human mitochondrial tRNAMet. *Nucleic Acids Res.*, **36**, 6548–6557.

33. Bilbille,Y., Gustilo,E.M., Harris,K.A., Jones,C.N., Lusic,H., Kaiser,R.J., Delaney,M.O., Spremulli,L.L., Deiters,A. and Agris,P.F. (2011) The human mitochondrial tRNAMet: structure/function relationship of a unique modification in the decoding of unconventional codons. *J. Mol. Biol.*, **406**, 257–274.

34. Takemoto,C., Spremulli,L.L., Benkowski,L.A., Ueda,T., Yokogawa,T. and Watanabe,K. (2009) Unconventional decoding of the AUA codon as methionine by mitochondrial tRNAMet with the anticodon f5CAU as revealed with a mitochondrial in vitro translation system. *Nucleic Acids Res.*, **37**, 1616–1627.

35. Cantara,W.A., Murphy,F.V.t., Demirci,H. and Agris,P.F. (2013) Expanded use of sense codons is regulated by modified cytidines in tRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 10964–10969.

36. Hartsel,S.A., Kitchen,D.E., Scaringe,S.A. and Marshall,W.S. (2005) RNA oligonucleotide synthesis via 5′-silyl-2′-orthoester chemistry. *Methods Mol. Biol.*, **288**, 33–50.

37. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.

38. Parkinson,G., Vojtechovsky,J., Clowney,L., Brunger,A.T. and Berman,H.M. (1996) New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallogr. D Biol. Crystallogr.*, **52**, 57–64.

39. Lebedev,A.A., Young,P., Isupov,M.N., Moroz,O.V., Vagin,A.A. and Murshudov,G.N. (2012) JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 431–440.

40. Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.

41. Read,R.J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst. A*, **42**, 140–149.

42. Frisch,M.J., Trucks,G.W., Schlegel,H.B., Scuseria,G.E., Robb,M.A., Cheeseman,J.R., Scalmani,G., Barone,V., Mennucci,B., Petersson,G.A. *et al.* (2009) *Gaussian 09, Revision E.01*. Gaussian, Inc., Wallingford.

43. Becke,A.D. (1993) Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, **98**, 5648–5652.

44. Lee,C., Yang,W. and Parr,R.G. (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens. Matter.*, **37**, 785–789.

45. Stephens,P.J., Devlin,F.J., Chabalowski,C.F. and Frisch,M.J. (1994) Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.*, **98**, 11623–11627.

46. Tomasi,J., Mennucci,B. and Cammi,R. (2005) Quantum mechanical continuum solvation models. *Chem. Rev.*, **105**, 2999–3093.

47. Scalmani,G. and Frisch,M.J. (2010) Continuous surface charge polarizable continuum models of solvation. I. General formalism. *J. Chem. Phys.*, **132**, 114110.

48. Karino,N., Ueno,Y. and Matsuda,A. (2001) Synthesis and properties of oligonucleotides containing 5-formyl-2′-deoxycytidine: in vitro DNA polymerase reactions on DNA templates containing 5-formyl-2′-deoxycytidine. *Nucleic Acids Res.*, **29**, 2456–2463.

49. Hall,K.B. and McLaughlin,L.W. (1991) Thermodynamic and structural properties of pentamer DNA.DNA, RNA.RNA, and DNA.RNA duplexes of identical sequence. *Biochemistry*, **30**, 10606–10613.

50. Ratmeyer,L., Vinayak,R., Zhong,Y.Y., Zon,G. and Wilson,W.D. (1994) Sequence specific thermodynamic and structural properties for DNA.RNA duplexes. *Biochemistry*, **33**, 5298–5304.

51. Kypr,J., Kejnovska,I., Renciuk,D. and Vorlickova,M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, **37**, 1713–1725.

52. Renciuk,D., Blacque,O., Vorlickova,M. and Spingler,B. (2013) Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine. *Nucleic Acids Res.*, **41**, 9891–9900.

53. Langlois d'Estaintot,B., Dautant,A., Courseille,C. and Precigoux,G. (1993) Orthorhombic crystal structure of the A-DNA octamer d(GTACGTAC). Comparison with the tetragonal structure. *Eur. J. Biochem.*, **213**, 673–682.

54. Safaee,N., Noronha,A.M., Rodionov,D., Kozlov,G., Wilds,C.J., Sheldrick,G.M. and Gehring,K. (2013) Structure of the parallel duplex of poly(A) RNA: evaluation of a 50 year-old prediction. *Angew. Chem. Int. Ed. Engl.*, **52**, 10370–10373.

55. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.

56. 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.

57. 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*, **54**, 1294–1305.

58. Lu,X.J. and Olson,W.K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.*, **3**, 1213–1227.

59. Auffinger,P. and Westhof,E. (1998) Hydration of RNA base pairs. *J. Biomol. Struct. Dyn.*, **16**, 693–707.

60. Testa,S.M., Disney,M.D., Turner,D.H. and Kierzek,R. (1999) Thermodynamics of RNA-RNA duplexes with 2- or 4-thiouridines: implications for antisense design and targeting a group I intron. *Biochemistry*, **38**, 16655–16662.

61. Siegfried,N.A., Kierzek,R. and Bevilacqua,P.C. (2010) Role of unsatisfied hydrogen bond acceptors in RNA energetics and specificity. *J. Am. Chem. Soc.*, **132**, 5342–5344.