Nucleic Acids Research, 2025, **53**, gkaf217
https://doi.org/10.1093/nar/gkaf217
Nucleic Acid Enzymes

A sequence-specific RNA acetylation catalyst

Supuni Thalalla Gamage[1], Shereen Howpay Manage[1], Aldema Sas-Chen[2], Ronit Nir[3], Brett W. Burkhart[4], Isita Jhulki[1], Courtney N. Link[1], Manini S. Penikalapati[1], Jane E. Jones[5], Lakshminarayan M. Iyer[6], L. Aravind[6], Thomas J. Santangelo[4], Schraga Schwartz[3], Jordan L. Meier[1,*]

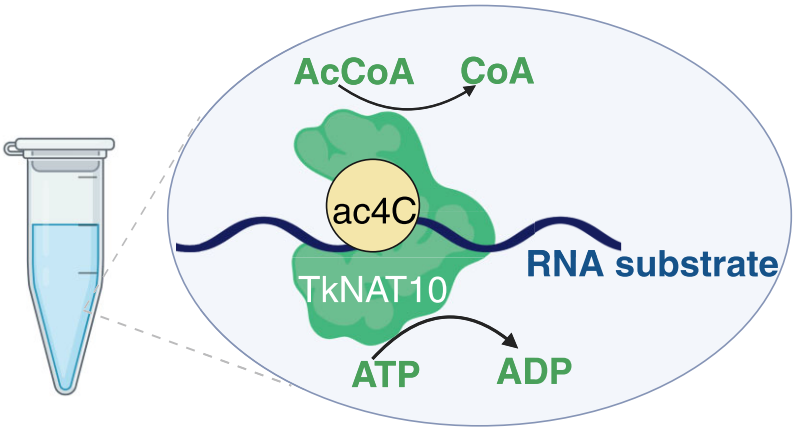
[1]Chemical Biology Laboratory, National Cancer Institute, Frederick, MD 21702, United States
[2]The Shmunis School of Biomedicine and Cancer Research, The George S. Wise Faculty of Life Sciences, Tel Aviv University, 6195001 Tel Aviv, Israel
[3]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7630031, Israel
[4]Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, CO 80523, United States
[5]Protein Expression Laboratory, Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, MD 21702, United States
[6]Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, United States

[*]To whom correspondence should be addressed. Email: jordan.meier@nih.gov

Abstract

N4-acetylcytidine (ac4C) is a ubiquitous RNA modification incorporated by cytidine acetyltransferase enzymes. Here, we report the biochemical characterization of *Thermococcus kodakarensis* Nat10 (TkNat10), an RNA acetyltransferase involved in archaeal thermotolerance. We demonstrate that TkNat10's catalytic activity is critical for *T. kodakarensis* fitness at elevated temperatures. Unlike eukaryotic homologs, TkNat10 exhibits robust stand-alone activity, modifying diverse RNA substrates in a temperature, ATP, and acetyl-CoA-dependent manner. Transcriptome-wide analysis reveals TkNat10 preferentially modifies unstructured RNAs containing a 5′-CCG-3′ consensus sequence. Using a high-throughput mutagenesis approach, we define sequence and structural determinants of TkNat10 substrate recognition. We find TkNat10 can be engineered to facilitate use of propionyl-CoA, providing insight into its cofactor specificity. Finally, we demonstrate TkNat10's utility for site-specific acetylation of RNA oligonucleotides, enabling analysis of ac4C-dependent RNA–protein interactions. Our findings establish a framework for understanding archaeal RNA acetylation and a new tool for studying the functional consequences of ac4C in diverse RNA contexts.

Graphical abstract



Introduction

Post-transcriptional modifications provide an evolutionary strategy to expand RNA function. One representative example is cytidine acetylation [1]. In bacteria, archaea, and eukaryotes, the formation of N4-acetylcytidine (ac4C) within RNA is catalyzed by the TmcA, Kre33, and Nat10 family of cytidine acetyltransferase enzymes (Fig. 1). The architecture of these proteins is highly homologous across all three kingdoms

Received: October 10, 2024. **Revised:** February 26, 2025. **Editorial Decision:** February 28, 2025. **Accepted:** March 7, 2025

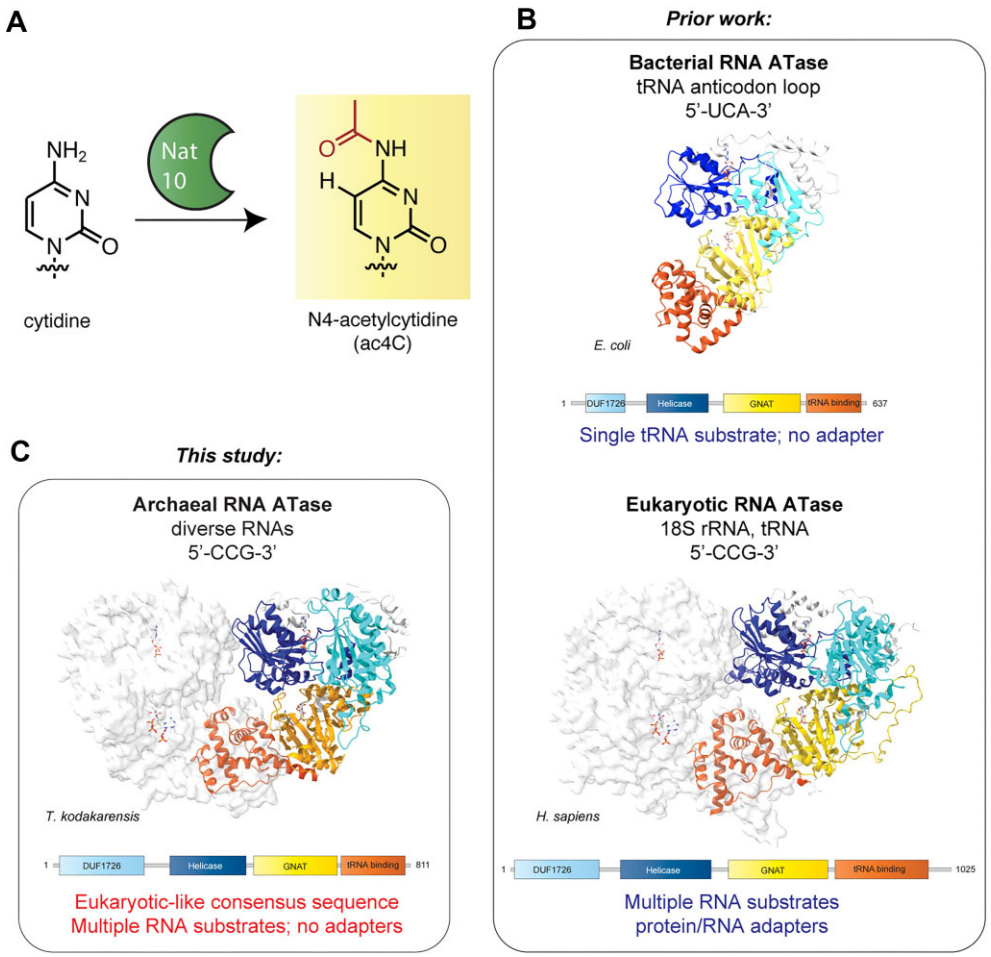


Figure 1. (**A**) RNA cytidine acetylation is catalyzed by the Nat10 family of enzymes. (**B**) Prior work has focused on characterization of monomeric bacterial RNA acetyltransferases (PDB 2ZPA) [10] that are highly specific or dimeric eukaryotic RNA acetyltransferases which use adapters to address 5′-CCG-3′ sequences in specific substrates. *Homo sapiens* NAT10 is an AlphaFold model superimposed on the cryogenic electron microscopy (cryo-EM) structure of Kre33 from *C. thermophilum* (PDB 6RXX) [52]. (**C**) This study characterizes an adapter-independent archaeal RNA acetyltransferase that modifies 5′-CCG-3′ consensus sequence in diverse substrates. *Thermococcus kodakarensis* Nat10 is depicted as a dimer based on AlphaFold Multimer predictions (Supplementary Fig. S1).

of life and comprises at a minimum an N-terminal helicase domain, a Gcn5-related N-acetyltransferase (GNAT), and an RNA-binding domain all joined within a single polypeptide scaffold ranging from ∼600 to 1000 amino acids (Fig. 1).

In bacteria, ac4C helps ensure accurate protein translation through the modification of a single substrate, the wobble position of initiator $tRNA^{Met}$ [2, 3]. Cytidine acetylation appears to also be highly controlled in eukaryotic organisms. Penetrant acetylation sites have thus far been validated within type II transfer RNA (tRNA) and 18S ribosomal RNA (rRNA), where they have been implicated in tRNA stability and ribosome biogenesis, respectively [4–6]. In contrast, archaea have demonstrated the potential for much more pervasive patterns of nucleic acid acetylation to exist. Applying nucleotide resolution ac4C sequencing as well as liquid chromatography–mass spectrometry (LC-MS) to the archaeal hyperthermophiles *T. kodakarensis* and *Pyrococcus furiosis*, our group and others have detected hundreds of ac4C sites distributed across diverse coding and noncoding RNA substrates [7–9]. In many substrates RNA acetylation was found to be directly proportional to growth temperature. This has led to the hypothesis that ac4C may aid fitness by increasing the robustness of RNA secondary structure at elevated temperature. Interestingly, in nearly all instances these sites occur at the central nucleobase of a 5′-CCG-3′ consensus sequence, the same motif that the human tRNA and rRNA sites are found within. How archaeal enzymes modify the same consensus sequence as eukaryotes—as well as the parameters of their vastly different substrate scope—remains to be established.

Despite much effort, biochemical studies of RNA cytidine acetyltransferase enzymes remain limited. *Escherichia coli* TmcA has been the most extensively characterized, revealing key determinants of $tRNA^{Met}$ recognition, stand-alone ATPase activity, and a potential mechanism for communication between the helicase and GNAT active sites [3, 10]. Eukaryotic Nat10-type acetyltransferases require protein and short nucleolar RNA (snoRNA) adapters to efficiently modify their cellular tRNA and rRNA substates, and thus present a more challenging target for reconstitution [11–13]. Our group and others have found sensitive radioactivity-based assays are necessary to detect the activity of reconstituted eukaryotic (fungal) Nat10 [5, 14, 15]. Although these studies have provided valuable biochemical and structural insights, their technical nature makes it difficult to analyze a large number of substrates,

while the limited turnover of eukaryotic enzymes prevents an obstacle to preparative applications [5]. Archaeal enzymes often provide tractable model systems useful for understanding more challenging eukaryotic counterparts as well as powerful tools for biotechnology [16, 17]. However, the reconstitution and functional properties of RNA acetyltransferases involved in archaeal thermotolerance are currently unknown.

Here we report the biochemical characterization of the sequence-specific RNA acetylation catalyst *T. kodakarensis* Nat10 (TkNat10). This enzyme exhibits stand-alone RNA cytidine acetyltransferase activity and is able to modify substrates in a temperature, ATP, and acetyl-CoA-dependent fashion. Profiling the transcriptome-wide modification of *T. kodakarensis* substrates reveals that the recombinant enzyme reaction directs acetylation to different substrates than are modified *in vivo*. A next-generation sequencing assay demonstrates a requirement for a 5′-CCG-3′ consensus sequence and a preference for unstructured substrates that can be influenced by secondary mutations. Finally, we probe the cofactor tolerance of this enzyme and demonstrate its promiscuous activity can be harnessed for site-specific acetylation of non-native RNA substrates, enabling us to study how ac4C perturbs RNA–protein interactions in its biologically relevant 5′-CCG-3′ context. Our studies define the properties of a novel component of the archaeal thermotolerance program and provide a tool for studying the functional consequences of cytidine acetylation in diverse RNAs.

Materials and methods

Expression and purification of recombinant TkNat10

The coding sequence (CDS) of *T. kodakarensis* Nat10 (TK0754) was codon-optimized for *E. coli* expression, custom-synthesized (ATUM), and inserted into a bacterial expression vector with an N-terminal His and maltose binding protein (MBP) tag (His6-MBP-tev-TkNat10 Wild type (WT); #24231-X02-566). The two TkNat10 mutants, His6-MBP-tev-TkNat10 S623G (#30594-X10-566) and His6-MBP-tev-TkNat10 S623G/W635A (#30594-X11-566), were similarly prepared. To prepare protein, each plasmid was transformed into BL21*pRARE competent *E. coli* cells (Protein Expression Laboratory, CCR/NCI). The proteins were expressed in cells grown at 37°C until they were at an OD600nm of 0.6 and then induced with 500 μM isopropyl β-D-1-thiogalactopyranoside (IPTG), and grown overnight at 15°C. After harvesting the cells by centrifugation for 30 min at 4°C at 6800 × *g*, they were lysed by sonication in a lysis buffer comprised of 50 mM Tris–HCl (pH 7.4), 150 mM sodium chloride, 10 mM imidazole, 0.1 mg/ml lysozyme (Thermo Scientific), 125 U/ml benzonase (Millipore), complete ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor cocktail tablets (Roche Diagnostics), and 2 mM tris(2-carboxyethyl)phosphine (TCEP). The lysate was pelleted in a pre-chilled centrifuge at 4°C at 12 000 rpm in a 14.5 rotor for 1 h. The resulting supernatant was filtered through a Millex HP 0.45 um, 33 mm filter unit and, was loaded onto a HisTrap HP Ni column (Cytiva) pre-equilibrated with lysis buffer using a peristaltic pump. After washing with His-binding buffer containing 50 mM Tris–HCl (pH 7.4), 150 mM sodium chloride and 25 mm imidazole, recombinant protein was eluted with a 10–250 mM gradient of imidazole in 50 mM Tris–HCl (pH 7.4), 150 mM sodium chloride. The collected fractions were further purified on a HiLoad 26/60 Superdex 200 column (GE Healthcare) pre-equilibrated with size exclusion chromatography buffer consisting of 50 mM Tris–HCl (pH 7.4), 300 mM NaCl, 10% glycerol, and 1 mm dithiothreitol. The purified protein was stored at −80°C until use.

Total RNA isolation from cells

Total RNA from human cells or *T. kodakarensis* cells [T559 strain, DTK0754:TkNat10 knockout (KO)] was extracted using TRIzol (Invitrogen, catalog no. 15–596-018) according to the manufacturer's protocol. Briefly, 1 ml TRIzol was used per 5–10 × 10^6 mammalian or 1 × 10^7 archaeal cells. After the cell lysis, samples were incubated at room temperature for 5 min followed by chloroform extraction and isopropanol precipitation. The RNA pellet was resuspended in water and quantified by ultraviolet (UV) absorbance and stored at −80°C.

Biochemical assessment of TkNat10 activity

In vitro TkNat10 activity assays were done by mixing 50 mM Tris–HCl (pH 6), 400 mM NaCl, 5 mM $MgCl_2$, 1 mM ATP, 0.2 mM acetyl CoA, RNA substrate, ∼1 μM recombinant TkNat10 enzyme (∼10 μg), and water up to 100 μl. The reaction mixture was incubated at 65°C for 3 h and the RNA was then isolated by using the Zymo RNA clean and concentrator-5 kit (Zymo Research, catalog no. R1013) according to the manufacturer's protocol. For acetylation control, reactions were set up without the enzyme or acetyl CoA. Enzyme activity was also assessed as a function of pH and temperature using a wide variety of substrates, as described in the manuscript.

Dot blot analysis

Dot blot assays were carried out to assess ac4C in RNA as described previously [18]. RNA was purified after the TkNat10 reaction and spotted onto Amersham Hybond-N + membranes (Cytiva, catalog no. RPN119B). Samples were crosslinked twice with 150 mJ/cm^2 in the UV_{254nm} Stratalinker 2400 (Stratagene, San Diego, CA). Following crosslinking, samples were blocked with 5% nonfat milk in TBST at room temperature for 30 min, and then probed with anti-ac4C (1:2000, Abcam) antibody in 5% nonfat milk in 1 × Tris-Buffered saline with Tween 20 (TBST) at 4°C overnight. Membranes were washed three times with TBST and incubated with HRP-conjugated secondary antirabbit IgG (1:10 000 dilution, Cell Signaling) for 1 h. Membranes were washed four times with 1 × TBST and developed with the SuperSignal ELISA Femto Maximum Sensitivity Substrate (Thermo Scientific, catalog no. 37 074) before detecting via chemiluminescent imaging.

In vitro transcription of TkNat10 substrates

In vitro transcription (IVT) was performed with the HiScribe T7 Kit (New England Biolabs, catalog no. E2040S), according to the manufacturer's instructions using DNA templates containing a T7 promoter upstream of the template sequence. Briefly, a 20 μl transcription reaction was set up with 100–200 ng DNA template, 10 mM of each NTP, 2 μl T7 RNA polymerase mix, and 1 × reaction buffer. The reaction was incubated for 3 h at 37°C, DNA was digested by adding 2 U of Turbo DNase (Invitrogen, catalog no. AM2238) directly to

the IVT reaction and incubating for 15 min at 37°C. Then the RNA was purified by using the Monarch RNA cleanup kit (NEB, catalog no. T2050L) according to the manufacturer's protocol.

IVT templates and products:

RNA name	DNA template sequence	RNA sequence
iorA	5'GCTAATACGACTCACT ATAGGGACAGCTTCCT CAGCTCGGTCGGCATG GGCGTTGAGGGAGGCT TCGTCATAATGGTTGC AGACGAC**CCG** AGCATGTGGTCTTCAC AAAACGAGCAGGACAC GCGCGTTTACGCAA3'	5'GGGACAGCUUCCUC AGCUCGGUCGGCAUGG GCGUUGAGGGAGGCUU CGUCAUAAUGGUUGCA GACGAC**CCG** AGCAUGUGGUCUUCAC AAAACGAGCAGGACAC GCGCGUUUACGCAA3'

Sanger sequencing-based analysis of TkNat10-catalyzed acetylation

Sanger-sequencing-based ac4C quantification was carried out as described previously [7]. *In vitro* transcribed RNA was incubated with Turbo DNase (Invitrogen, catalog no. AM2238) for 30 min at 37°C to remove any DNA contamination prior to the ac4C sequencing assay. Then, the iorA RNA (1 μg) was incubated with either sodium cyanoborohydride (100 mM in H_2O) or vehicle (H_2O) in a final reaction volume of 100 μl. Samples were incubated for 20 min at room temperature and quenched with 1 M Tris, pH 8 buffer (30 μl). Reactions were adjusted to 200 μl with H_2O, ethanol precipitated, briefly dried on Speedvac, and resuspended in H_2O. RNA from individual reactions (200 pg) was incubated with the iorA forward primer (ACAGCTTCCTCAGCTCGGTC, 0.2 uM) at 65°C for 5 min and transferred to ice for 3 min to facilitate annealing in 1 × Superscript III reaction buffer. After annealing, reverse transcription (RT) reactions were set up by adding 5 mM DTT, 200 units of Superscript III, optimized dNTP mix (250 μM GTP and 500 μM of ATP, CTP, TTP), 10 U of RNasin plus inhibitor (Promega), H_2O up to 20 μL to the tubes and incubating for 60 min at 55°C. Polymerase chain reaction (PCR) reactions using 2 μl of RT reactions were set up to amplify the complementary DNA (cDNA) product. To a 50 μl total reaction, 1 × HF buffer, 0.5 μM of each forward and reverse primer (iorA fwd primer: ACAGCTTCCTCAGCTCGGTC, iorA rev primer: TTTGCGTAAACGCGCGTG), 200 μM of each dNTP, 2 U of Phusion hot start enzyme, and H_2O were added. Thermocycling conditions: 30 s at 98°C, 33 cycles (10 s at 98°C, 30 s at 65°C, 20 s at 72°C) and 2 min extension at 72°C. PCR products were run on a 2% agarose gel, stained with SYBR safe and visualized on UV transilluminator at 302 nm. Bands of the desired size were excised from the gel and DNA extracted using QIA-quick gel extraction kit from Qiagen and submitted for Sanger sequencing (Azenta) using the forward PCR primer. Processed sequencing traces were viewed using 4Peaks software. Peak height for each base was measured at the CCG and the percent misincorporation was determined using the equation: "Percent misincorporation= (T base peak intensity)/(sum of C and T base peaks)*100%" for sodium cyanoborohydride treated and vehicle reactions. The % misincorporation for each sample was calculated by subtracting the vehicle from sodium cyanoborohydride treated and then normalizing to empty vector and WT values.

NGS-based analysis of TkNat10-catalyzed acetylation

To assess recombinant TkNat10's ability to modify total RNA isolated from a *T. kodakarensis* TK0754 KO strain, samples were treated with the enzyme (10 μg) followed by treatment with either $NaCNBH_3$ or a mock-treatment (water/HCl) as described in the ac4C-seq protocol [19]. RT was performed using TGIRT at 42°C for 16 h [20]. RNA was hydrolyzed and a second ligation was performed by adding a 3′ DNA oligo (5′ to the RNA strand) containing the Truseq R1 Illumina adapter and a 6 bp unique molecular identifier (UMI) for BY4741/HEK293T/3T3 or without a UMI for mouse liver samples. Barcoded primers were then used to amplify the sequencing library via PCR. Libraries were subsequently sequenced on Illumina NovaSeq 6000 platform with an SP100 kit with read-lengths split evenly between R1 and R2. Read processing, alignment, and analysis were performed as previously described [7].

High-throughput mutational profiling of TkNat10 model substrates

DNA templates for the preparation of a degenerate library were purchased from Twist Bioscience (Supplementary Table S4). Each template contained a T7 promoter at the 5′ end for IVT, along with either a human or *T. kodakarensis* h45 rRNA sequence, which was flanked by two stretches of DNA that did not contain any CCG motifs. The flanking DNA regions were added to increase RNA length for synthesis and sequencing purposes. The library contained h45 sequences with a single CCG motif (positions 4–6), which remained constant, while mutations were introduced into the surrounding h45 regions to generate single or double mutant substrates. Additionally, all single mutants that incorporated XCG or CCX sequences at the CCG motif were included in the library. Mutations that would result in the creation of additional CCG sites were excluded.

The DNA library was PCR-amplified using Phusion High-Fidelity (HF) DNA polymerase (annealing temperature = 56°C, 33 cycles) with a fixed forward primer (5′-TAATACGACTCACTATAGGGTCGG-3′) and a reverse primer (5′-GACCTTTCCAAGGGCATAGATC-3′), which were common to all templates. Briefly, a 50 μl reaction was set up containing 1× HF buffer, 0.5 μM each of the forward and reverse primers, 200 μM of each dNTP, 2 U of Phusion Hot Start DNA polymerase, ∼17 ng of the template DNA pool, and nuclease-free water. The thermocycling conditions were as follows: initial denaturation at 98°C for 30 s, followed by 33 cycles of 10 s at 98°C, 30 s at 56°C, and 20 s at 72°C, with a final extension at 72°C for 2 min. The amplified DNA library was run on a 2% agarose gel (expected product size: 170 bp), and the desired band was excised and purified using the NucleoSpin Gel and PCR Cleanup Kit (Macherey–Nagel, catalog no. 740609) following the manufacturer's protocol.

IVT was performed using the Superscript T7 Kit (New England Biolabs, catalog no. E2040S) according to the manufacturer's instructions. In brief, a 100 μl transcription reaction was assembled using 1 μg of the DNA template pool, 10 mM of each NTP, 10 μl of T7 RNA polymerase mix, and 1 × T7 reaction buffer. The reaction was incubated at 37°C for 3 h, after which 20 U of Turbo DNase (Invitrogen, catalog no. AM2238) were added to digest the DNA for 30 min at 37°C. The IVT product was purified by denaturing it at 95°C for

3 min in 1 × RNA loading dye, followed by running the sample on a 6% polyacrylamide gel electrophoresis (PAGE)–urea gel (14 W, 1 × Tris-borate-EDTA (TBE)). The RNA band (150 nucleotides) was visualized under UV light, excised, and subjected to a crush and soak method using 800 µl of buffer (500 mM ammonium acetate, 2 mM EDTA) at 4°C overnight. The RNA-containing supernatant was recovered and desalted using a 1 kDa molecular weight cutoff centrifugal device (Pall Corporation, catalog no. MCP001C41).

The RNA library was then subjected to *in vitro* acetylation using TkNat10, as described previously. Each reaction used 3 µg of RNA and was conducted at 37, 55, 65, 75, and 85°C for 3 h. For the time-course experiments at 65°C, reactions were incubated for 1, 3, and 5 h. Temperature-and time-dependent reactions were done in duplicate. Approximately 1 µg of RNA was used for $CNBH_3$ treatment and mock (H_2O treated) reactions (32 samples), followed by RT, as described earlier. Superscript III-based RT reactions were performed with 8 ng of RNA (both $CNBH_3$-treated and mock-treated) using the RT primer (5′-GACCTTTCCAAGGGCATAGATC-3′). The RT reactions were adjusted to a final volume of 25 µl with 5 µl of nuclease-free water, and the resulting cDNA was purified using MicroSpin G-25 columns (Cytiva, catalog no. 27532501) according to the manufacturer's instructions.

The purified cDNA was then used in PCR to attach adapters for next-generation sequencing. PCR1 was performed using 5 µl of purified cDNA, 0.5 µM primers (forward primer: 5′-CCCTACACGACGCTCTTCCGAT CTTCGGCCCACGGCCCTGGC-3′, reverse primer: 5′-GACTGGAGTTCAGACGTGTGCTCTTCCGATCGACCT TTCCAAGGGCATAGATC-3′), and Q5 Hot Start High-Fidelity 2 × Master Mix (NEB, catalog no. M0494S) in a 50 µl reaction. The thermocycling conditions were as follows: initial denaturation at 98°C for 30 s, followed by 5 cycles of 10 s at 98°C, 30 s at 72°C, and 20 s at 72°C, with a final extension at 72°C for 2 min. The amplified DNA was purified using PureLink PCR microspin columns (Thermo Fisher, catalog no. K310050), eluted with 10 µl of nuclease-free water, and used as the template for PCR2.

In PCR2, i5 and i7 dual indexing primers were attached to the PCR1 amplicon using Q5 Hot Start High-Fidelity 2 × Master Mix. The PCR2 forward primer was 5′-AATGATACGGCGACCACCGAGATCTACAC[i5]AC ACTCTTTCCCTACACGACGCTCTTCCGATC*T-3′*, and the reverse primer was *5′-CAAGCAGAAGACGGCATACG AGAT[i7]GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCT*-3′. Thermocycling conditions were 30 s at 98°C, followed by 25 cycles of 10 s at 98°C, 30 s at 72°C, and 20 s at 72°C, with a final extension at 72°C for 2 min. The PCR products were purified using a 2% agarose gel (amplicon size: 283 bp), eluted with 15 µl of nuclease-free water, and diluted to 5 ng/µl (~50 nM). The quality of the libraries was assessed using a High Sensitivity DNA Assay (Agilent, catalog no. 5067-4626) on an Agilent Bioanalyzer. Library size and concentration were obtained from the assay, and an equimolar concentration of the 28 libraries was pooled for sequencing. The sequencing-ready pool was quantified by quantitative polymerase chain reaction (qPCR) before being subjected to a sequencing quality control (QC) run on an Illumina MiSeq sequencer. Data from the MiSeq run were analyzed, and adjustments to the libraries were made based on the read count for each sample. Following adjustment, the library pool was quantified again by qPCR, and a sequencing run was initiated on an Illumina NextSeq 2000 P2 paired-end 100-cycle sequencing platform. Basecalling was performed using Illumina RTA v3.10.30, and samples were demultiplexed using BCL2FASTQ v2.20. QC was performed using FastQC and FastQ_Screen, with all samples showing a Q30 base percentage >75%. The samples yielded between 12 and 34 million pass-filter reads.

Bioinformatic analysis of TkNat10 substrate library experiments

Sequences were uploaded and demultiplexed on the NCI Sequencing Facility's database system with read quality assessed using FastQC. Target sequences were extracted from paired end sequencing data using the custom scripts R1_extract.perl (my $left_flank= "GTCGTAACAAGG"; my $max_length = 24) and R2_extract.perl (my $left_flank= "CTCCCTTAATGATC"; my $right_flank = "CCTTGTTAC-GAC"). The R2 reads for each condition were transformed into their reverse complement (reverse_complement.perl), merged with the R1 reads for each condition (merge_files.perl) and the occurrence of each individual sequence in the merged reads for each condition counted (sequence_counts.perl; length = 24). The output of this analysis was then uploaded and further analyzed on the NIH HPC Biowulf Linux cluster. Counts for all 28 conditions were merged into a single parseable file (merge_counts.perl) and then filtered for matched occurrences of the 5′-CCG-3′ consensus sequence and 5′-CTG-3′ ac4C-seq sequence at positions 4–7. The percent CTG conversion was calculated by comparing the number of reads 5′-CTG-3′ at positions 4–7 to the total number of reads (5′-CCG-3′ + 5′-CTG-3′ for either (i) $NaCNBH_3$/HCl-treated versus vehicle-treated or (ii) TkNat10-treated versus untreated samples for each condition. In practice the use of either control yielded nearly identical results. To correct for the fact that three mutants (U17 → G17, U17 → C17, and G20 → C20) can create new TkNat10 consensus sequences, the percent CTG for these mutants was calculated combining the read counts for the converted and unconverted substrate sequences. Percent CCG remaining for each mutant was calculated analogously. Positional base frequencies were calculated from extracted 24-mers using Discriminative Regular Expression Motif Elicitation, part of the MEME Suite (https://meme-suite.org/meme/). Predicted free energies for library members were calculated using ViennaRNA version 2.3.1 (https://www.tbi.univie.ac.at/RNA/#) and the parameter 'RNAfold -p0 –noPS'. All other parameters were set to default settings. Scatterplots and heatmaps were plotted using Graphpad Prism. Custom perl scripts used in this study can be found at the GitHub repository, https://github.com/Meier-Lab-NCI/TkNat10.

Identifying modifications in RNA by MALDI mass spectrometry

To test the TkNat10 activity by Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS), a 10-mer RNA with the sequence CUUCCGUAGG (3136.9 Da) was purchased from Integrated DNA Technologies (IDT). *In vitro* TkNat10 reactions were carried out with either the WT TkNat10 enzyme, S623G or S623G/W635A variants in the presence of acetyl CoA or propionyl CoA cofactors using the activity assay described above. Control reactions were set up without CoA. After the reaction, the 10-mer RNA was puri-

fied using ziptip with C18 resin (Millipore, ZTC18S096) according to the manufacturer's protocol. Samples for MALDI-MS were prepared by mixing 1–2 μl of RNA (∼100 ng) with 1–2 μl of 3-HPA matrix (prepared by mixing nine parts 50 mg/ml 3-Hydroxypicolinic acid (3-HPA) in 50% acetonitrile (ACN)/ddH_2O with one part 50 mg/ml ammonium citrate in ddH_2O). Samples were left to air dry on the MALDI plate, and mass spectra were recorded on a Shimadzu Biotech Axima Performance mass spectrometer in negative reflectron mode.

LC-MS analysis of acetylated 10-mer RNA

Short 10-mer RNA containing a 5′-CCG-3′ sequence was incubated with TkNat10 and cofactors at 65°C for 3 h as described above. After the reaction, the RNA was purified using ziptip with C18 resin (Millipore, ZTC18S096) according to the manufacturer's protocol. Purified RNA was digested using the nucleotide digestion mix (Cat. no. M0649S, New England BioLabs) according to the manufacturer's protocol. Digested samples were directly used for LC-MS quantification of free nucleosides, as previously described [21]. The percent ac4C/(ac4C + C) was calculated.

RNA–protein pulldown

Both 3′-biotinylated and nonbiotinylated 10-mer RNA oligonucleotides were purchased from IDT. These RNAs were acetylated using the TkNat10 enzyme, following previously described protocols. Control reactions without acetylation were set up by omitting acetyl-CoA from the enzymatic reaction. RNA purification was performed using ZipTips (Millipore, catalog no. ZTC18S096), according to the manufacturer's instructions. Acetylation was confirmed via Matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectrometry.

For the pull-down experiments, whole-cell lysates were prepared from RAW 264.7 cells by resuspending cell pellets in 1× ice-cold phosphate-buffered saline containing a protease inhibitor cocktail (EDTA-free; Cell Signaling Technology, catalog no. 5871). Cells were lysed by sonication using a QSonica Q500 sonicator (3-s pulse, amplitude 20%, with 30-s rests on ice between pulses, for a total of five pulses). Lysates were clarified by centrifugation at 20 817 × *g* for 30 min at 4°C and quantified using a Qubit 4.0 Fluorometer with the Qubit Protein Assay Kit (Thermo Fisher Scientific, catalog no. Q33211). The lysates were diluted to a final concentration of 3 mg/ml in lysis buffer and stored at −80°C until further use. Prior to the pull-down assay, lysates were thawed on ice.

For each pull-down reaction, 500 μg of lysate was incubated with 10 μl of 1 × protein–RNA binding buffer, 30 μl of 50% glycerol, 100 pmol of yeast $tRNA^{Phe}$, 0.8 U/μl of RNasin, and nuclease-free water to a final volume of 100 μl. The RNA–protein pull-down assay was performed using the Pierce Magnetic RNA–Protein Pull-Down Kit, following the manufacturer's protocol. Magnetic streptavidin beads were first washed twice with 20 mM Tris–HCl (pH 7.5). Subsequently, 50 μl of 1 × RNA capture buffer (20 mM Tris–HCl, pH 7.5, 1 M NaCl, 1 mM EDTA) and 0.8 U/μl of RNasin were added to the beads, followed by incubation at room temperature for 30 min. After removing the supernatant, the beads were pre-blocked by incubating with 50 pmol of yeast $tRNA^{Phe}$ at 4°C for ∼16 h while rotating.

After pre-blocking, the supernatant was discarded, and 50 pmol of biotinylated RNA in 1 × RNA capture buffer was added to the beads, gently mixed, and incubated for 30 min at room temperature with agitation. Following RNA binding, the beads were washed twice with 20 mM Tris–HCl (pH 7.5) and once with 1 × protein–RNA binding buffer (0.2 M Tris–HCl, pH 7.5, 0.5 M NaCl, 20 mM $MgCl_2$, 1% Tween-20). The beads were then incubated with the prepared cell lysate solution for 1 h at 4°C. Control reactions were performed by incubating lysates with beads that had no biotinylated RNA bound. After protein capture, the beads were washed three times with wash buffer (20 mM Tris–HCl, pH 7.5, 10 mM NaCl, 1% Tween-20) and then exchanged into 50 mM HEPES–KOH (pH 7.5) for subsequent proteomic analysis.

LC/MS/MS and proteomic analysis of the RNA–protein pulldown samples

Digestion of biotinylated-10mer on-bead proteins and TMT-pro labeling

Each sample was incubated with 150 μl of digestion buffer (10 mM TCEP, 40 mM chloroacetamide, and 15 ng/μl trypsin/LysC in the lysis buffer) from the EasyPep kit (Thermo, catalog no. A40006) at 37°C, overnight in the dark shaking at 1000 rpm. After digestion 150 μl of each solution was transferred to a new tube and treated with 10 μl of 10 μg/μl TMTpro (Thermo, catalog no. A52045) reagent followed by a 1 h incubation at 25 °C while shaking. Excess TMTpro was quenched with 50 μl of 5% hydroxylamine and 20% formic acid (FA) for 10 min. Samples were combined and cleaned using EasyPep mini columns (Thermo, catalog no. A40006) following the manufacturer's protocol. Eluted peptides were dried in speed-vac.

LC/MS analysis of peptides

Dried down peptides were reconstituted in 50 μl of 0.1% FA and 5 μl of the peptide sample was analyzed with an EASY-Spray™ HPLC column (Thermo Scientific™ catalog no. ES900) in a Dionex UltiMate 3000 RSLC system combined with a Orbitrap Eclipse™ Tribrid™ Mass Spectrometer. The analysis was done in duplicates. Solvent A consisted of 0.1% FA in water and Solvent B consisted of 0.1% FA in 80% ACN. The loading pump with Solvent A was first operated at 7 μl/min for 6 min and subsequently decreased to 2 μl/min when the valve was switched to bring the trap column (Acclaim™ PepMap™ 100 C18 HPLC Column, 3 μm, 75 μm I.D., 2 cm, PN 164535) in-line with the analytical column (EasySpray C18 HPLC Column, 2 μm, 75 μm I.D., 25 cm, PN ES902). The flow rate of the gradient pump was 300 nl/min. Each run used a linear liquid chromatography gradient of 5%–7% B for 1 min, 7%–30% B for 83 min, 30%–50% B for 25 min, 50%–95% B for 4 min, holding at 95% B for 7 min, followed by re-equilibration of the analytical column at 5% B for 17 min. All mass spectrometry (MS) injections employed the TopSpeed method with three high-field asymmetric waveform ion mobility spectrometry compensation voltages (CVs) and a 1-s cycle time for each CV (3-s cycle time total) that consisted of the following: a spray voltage of 2200 V and an ion transfer temperature of 300 °C. MS1 scans were acquired in the Orbitrap with a resolution of 120 000 automatic gain control (AGC) of 4e5 ions, and a max injection time of 50 ms with a mass range of 350–1600 m/z. MS2 scans were acquired in the Orbitrap using the TurboTMT method with a resolution of 15 000 AGC of 1.25e5, a max injection time of 22 ms, an HCD energy of 38%, isolation width of

0.4 Da, an intensity threshold of 2.5e4 and charges 2–6 for MS2 selection. Advanced Peak Determination, Monoisotopic Precursor selection, and EASY-IC for internal calibration were enabled and dynamic exclusion was set to a count of 1 for 15 s. The only difference in the methods was the CVs used (the first method used CVs of −45, −60, −75, and the second used CVs of −50, −65, −80).

Database search and postprocessing analysis

Both injections were coupled together as fractions and screened with Proteome Discoverer 2.4 using the Sequest node. Data were queried against the Uniprot Human database (February 2020) using a full tryptic digest, 2 max missed cleavages, a minimum peptide length of 6 amino acids and a maximum peptide length of 40 amino acids, an MS1 mass tolerance of 10 ppm, an MS2 mass tolerance of 0.02 Da, variable oxidation on methionine (+15.995 Da), fixed modifications of carbamidomethyl on cysteine (+57.021) and TMTpro (+304.207) on lysine. Peptide N-terminus Percolator was used for false discovery rate (FDR) analysis and TMTpro reporter ions were quantified using the Reporter Ion Quantifier node and normalized on the total peptide intensity of each channel. The TMTpro channel assignment for conditions were as follows: (i) Control-biot-10mer: 126, 127N, 128N; (ii) ac4c-biot-10mer: 129N, 130N, 131N; (iii) Lysate: 132N, 133N, 134N.

Biochemical reconstitution of an archaeal RNA acetyltransferase

As a preface to *T. kodakarensis* Nat10 reconstitution, we first sought to determine the necessity for its catalytic activity *in vivo*. Prior research has shown that genetic deletion of TkNat10 causes decreased fitness at elevated temperatures [7, 8]. One caveat to this finding is that TkNat10 contains multiple protein domains (Fig. 1), making it unclear whether this phenotype was specifically due to loss of acetyltransferase activity. To address this, we engineered a *T. kodakarensis* strain in which the highly conserved histidine (H506) of the 'HY' motif in the GNAT domain was mutated to alanine (Fig. 2A). This motif, present in all characterized Nat10-type GNAT domains, is essential for the activity of bacterial and eukaryotic enzymes [6, 10, 15]. Analysis of ac4C content in total RNA isolated from the H506A strain confirmed complete loss of acetylation (Supplementary Fig. S2A). Monitoring growth as a function of time revealed that strains lacking RNA acetylation—either due to catalytic mutation or genetic deletion—exhibit significantly decreased growth at 85°C and 95°C (Fig. 2A and B and Supplementary Fig. S2B and C). These findings validate the critical role of TkNat10's cytidine acetyltransferase activity in *T. kodakarensis* fitness under thermal stress.

Having established the *in vivo* importance of TkNat10, we proceeded to its *in vitro* reconstitution. An *E. coli* codon-optimized version of the TkNat10 gene (TK0754) was synthesized and inserted into a Gateway cloning vector and used to produce N-terminal His and His/MBP tagged versions of the enzyme via bacterial overexpression. Soluble protein was purified by immobilized metal chromatography and fast protein liquid chromatography (FPLC). The His and His-MBP versions displayed roughly equivalent activity (vide infra); for practicality we will describe the characterization of His-MBP-tagged TkNat10, which consistently gave higher yields of soluble protein. The purified 1207 amino acid enzyme showed a molecular weight of ∼137 kDa on sodium dodecyl sulfate–polyacrylamide gel electrophoresis, aligning with its predicted mass (Fig. 2C and Supplementary Fig. S3).

TkNat10 is known to modify hundreds of RNA targets *in vivo* [1], which complicates the selection of an optimal biochemical substrate. To circumvent this, we developed a nonradioactive assay for RNA acetyltransferase activity using total RNA isolated from *T. kodakarensis* strains deficient in TkNat10 (TkNat10 KO). Our rationale was that RNA from these strains, being entirely deficient in ac4C [7], should contain many substrates amenable to TkNat10 modification that could be readily detected by anti-ac4C dot blot (Fig. 2D) [18, 22]. Incubation of recombinant TkNat10, acetyl-CoA (AcCoA), and ATP with KO RNA at 65°C produced a robust ac4C signal (Fig. 2E). Omission of any single reaction component (enzyme, AcCoA, ATP) or heat abolished the signal. TkNat10-catalyzed acetylation was temperature-dependent, with RNA modification increasing over a temperature gradient of 55–95°C (Fig. 2F). RNA acetylation was further found to be installed in a time-dependent manner with an optimal pH of 6 -7 (Supplementary Fig. S4). These studies establish a strategy for heterologous expression and biochemical reconstitution of archaeal RNA acetyltransferase activity.

Transcriptome-wide analysis of recombinant RNA acetyltransferase activity

To gain a higher resolution picture of the activity of recombinant TkNat10 we next analyzed the biochemically modified pool using quantitative ac4C sequencing (ac4C-seq; Fig 3A) [19]. This approach exploits the unique electronic character of ac4C which renders it susceptible to acid-catalyzed sodium cyanoborohydride reduction. The product of this reaction—reduced ac4C—causes misincorporations during RT and produces a strong mutational signature at ac4C sites that can be detected upon cDNA sequencing [23]. Consistent with dot blot data, applying ac4C-seq to TkNat10 KO RNA treated with enzyme and cofactors at 65°C revealed the formation of ac4C misincorporation signals (>5%) across 405 RNA sites sequenced with a depth of at least 50 reads (Supplementary Table S1). In line with the known cellular specificity of this enzyme, these signals were observed exclusively within 5′-CCG-3′ sequences. Among RNAs containing a 5′-CCG-3′ and sequenced with this depth, 85% (749/878) showed evidence of modification, consistent with this motif's ability to serve as a strong determinant of cytidine acetyltransferase catalysis.

To systematically compare the modification profile of biochemical and cellular TkNat10, we took advantage of the quantitative nature of ac4C-seq which allows misincorporation percentage to be used as a proxy for modification level. We analyzed 239 ac4C sites within the *T. kodakarensis* transcriptome that were quantified (depth ≥ 50, misincorporation > 0) both in the *in vivo* experiment [7] as well as upon enzymatic TkNat10 treatment. *In vivo* many of the most penetrant ac4C sites lie within rRNA and tRNA (Fig. 3B, y-axis). In contrast, the highest levels of misincorporation driven by recombinant TkNat10 treatment were found in messenger RNA (mRNA; Fig. 3B, x-axis). It is important to specify that the graph in Fig. 3B does not depict the 726 sites exclusively modified in the enzymatic experiment or the 44 sites observed exclusively modified *in vivo* (Supplementary Table S1), the

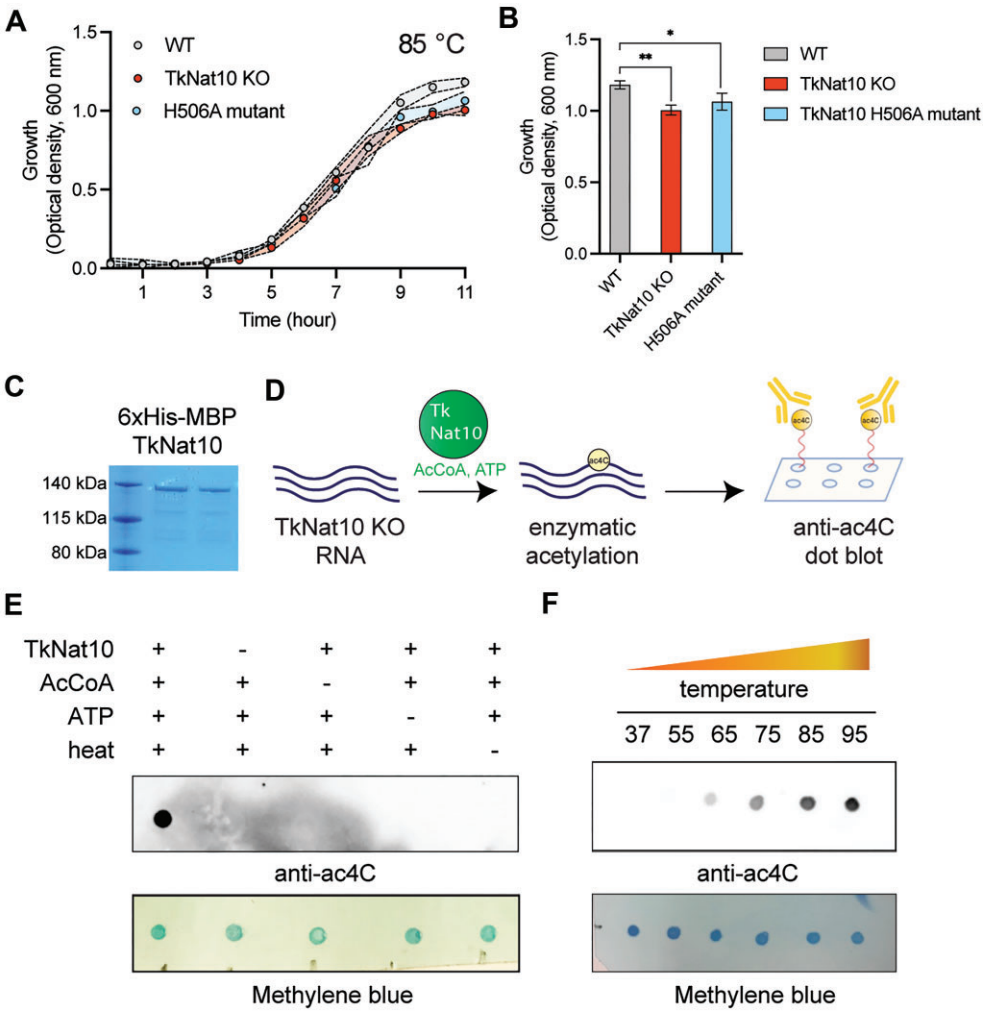


Figure 2. (**A**) Growth rate of WT *T. kodakarensis*, TkNat10 deletion strain, and strain containing a TkNat10 H506A catalytic mutant at 85°C. (**B**) Differential growth of *T. kodakarensis* strains at 85°C for 11 h. Values represent ≥3 replicates, analyzed by two-tailed Student's *t*-test (**P* < .05, ***P* < .01). (**C**) Purification of 6xHis-MBP-tagged TkNat10 enzyme. (**D**) Schematic for qualitative analysis of TkNat10 activity by anti-ac4C dot blot. (**E**) Acetylation catalyzed by TkNat10 is enzyme, acetyl-CoA, ATP, and heat-dependent. (**F**) TkNat10 catalysis as a function of temperature.

imbalance of which emphasizes the promiscuity of the free enzyme.

The mRNA modification sites identified were localized to the CDS, although this primarily reflects the lack of untranslated region (UTR) annotation in archaeal genome browsers. Detailed annotation of archaeal UTRs does not fall within the focus of this manuscript. To validate the ability of TkNat10 to modify one of these newly identified CDS sites, we *in vitro* transcribed a highly modified mRNA encoding indolepyruvate ferredoxin oxidoreductase, alpha subunit (iorA; Fig. 3C). Consistent with our transcriptome-wide analysis, we found that treatment with recombinant TkNat10 catalyzed significant modification of this IVT-produced substrate (Fig. 3C). Interestingly, secondary structure prediction indicates the modification site in iorA occurs in an unpaired 5′-CCG-3′ motif adjacent to two stem loops.

As coding transcripts are often less structured than noncoding RNAs, we next asked: what is the relationship between folding energy and TkNat10 modification? To assess this, we divided 805 RNAs modified by TkNat10 into five bins based on their degree of misincorporation, ranging from high to low. The Vienna RNA package was then used to calculate the MFE of a 40-nucleotide fragment of each RNA centered on the modified 5′-CCG-3′ site [24]. RNAs exhibiting the highest modification by TkNat10 (e.g. bin 1; Fig. 3D) on average had the least negative MFE, suggestive of less structure. This trend held when equivalent calculations were made using a 100-bp folding context or a folding temperature of 75°C (Supplementary Fig. S5). This finding evokes a previous study of yeast PUS7, which was found to preferentially incorporate pseudouridine into UGUAG motifs within RNAs lacking secondary structure [25]. These data indicate the wide substrate scope of recombinant TkNat10 and suggest its stand-alone activity may be anticorrelated with RNA secondary structure.

High-throughput mutagenesis of a TkNat10 substrate

While our transcriptome-wide experiment describes the ability of recombinant TkNat10 to modify a wide variety of *T. kodakarensis* RNAs, it does not provide a detailed view into how the sequence of any individual substrate impacts acetylation. To evaluate this in a systematic fashion, we developed a strategy to apply ac4C-seq for high-throughput mutational profiling of model substrates. The approach harnesses advances in parallel oligonucleotide synthesis to construct large

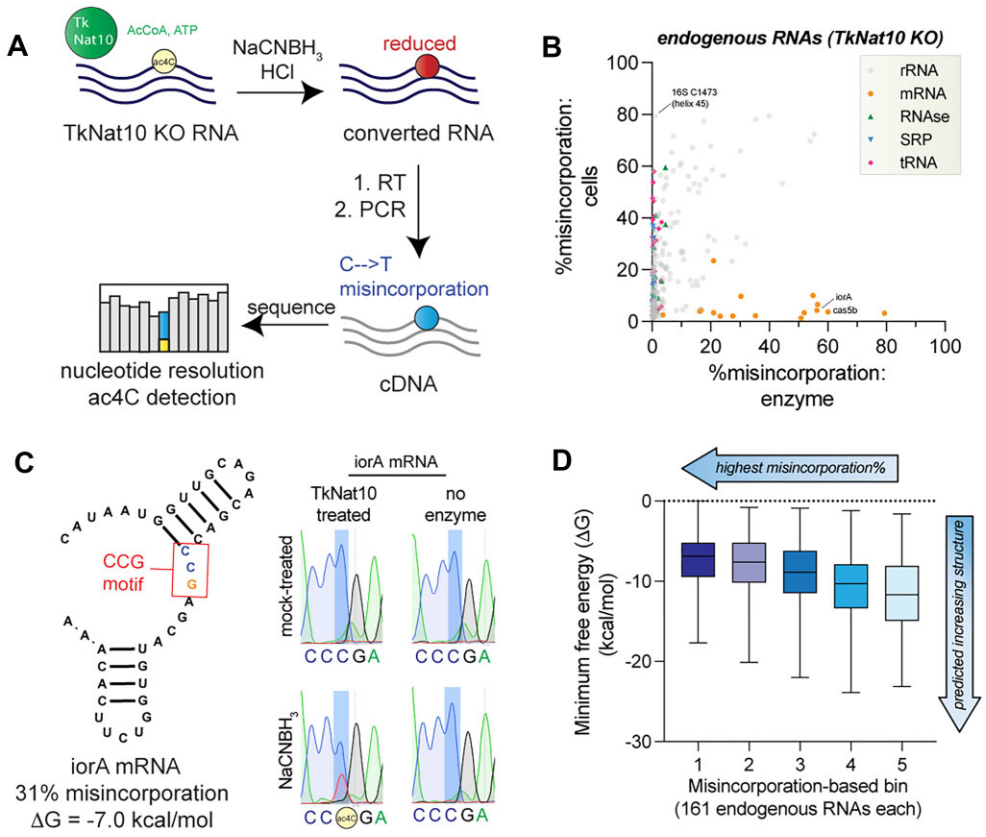


Figure 3. (**A**) Schematic for transcriptome-wide identification of recombinant TkNat10 substrates in total RNA isolated from a TK0754 KO strain. (**B**) Correlation of misincorporation percentages observed upon treatment of KO RNA with recombinant TkNat10 enzyme (x-axis) or in RNA isolated from wild-type *T. kodakarensis* (y-axis). (**C**) Example of an enzyme-catalyzed misincorporation site in iorA messenger RNA (mRNA) which was validated by amplicon-based Sanger sequencing. (**D**) Correlation of ac4C-seq misincorporation percentage with minimum free energy (MFE) of the 40 base region surrounding the ac4C site calculated by ViennaRNA.

libraries of substrate sequences which can be *in vitro* transcribed, treated with TkNat10 and cofactors, and subjected to ac4C-seq. As initial targets for analysis, we focused on the highly conserved ac4C sites found in helix 45 of *T. kodakarensis* and *H. sapiens* small subunit rRNA (Fig. 4A). Our hypothesis was that these two models would enable analysis of evolutionary conservation and divergence of substrate recognition in a well-defined structural context. Furthermore, the presence of two noncanonical G • U base pairs in the stem of the human substrate would allow us to test the effect of strengthening base pairs on TkNat10 activity.

Accordingly, we first designed 24-mer sequences which kept the 5′-CCG-3′ found at positions 4–6 constant (numbering from the base of the helix) while sampling all single mutations for both models. Mutations that would create potential 5′-CCG-3′ substrate sequences were omitted to simplify analysis. To enhance our analysis of human helix 45, we additionally surveyed all possible pair-wise mutants as well and positions 4/6, with the aim of identifying epistatic mutations and probing TkNat10's necessity for the 5′-CCG-3′ consensus, respectively. Sequences with mutations in positions 4 and 6—which flank the modified cytidine at 5′-CCG-3′—displayed >98% less modification than the parent, confirming the enzyme's preference for this motif (Supplementary Table S2 and Supplementary Fig. S6A). T7 templates encoding the designed rRNA hairpin as well as PCR adapters were purchased, PCR-amplified, and subjected to IVT. Purified RNA was incubated with recombinant TkNat10 and cofactors and subjected to ac4C-seq (Fig. 4B). In addition to a 'no reduction' control, we examined the effect of time (1–5 h) and temperature (37–85°C) on conversion. Analyzing a control library where no conversion was expected (no reduction, 37°C) revealed adequate coverage (>200 reads) of all single mutants of the parent sequence, with a median depth of 2264 reads per library member for this single condition (Supplementary Table S2).

To extract TkNat10 activity from our sequencing data reads containing 5′-CCG-3′ and 5′-CTG-3′ at bases 4–6 were paired and used to calculate '% CCG conversion' (Supplementary Table S2). This ratiometric measure normalizes for variable read depth caused by the differential representation of each sequence in the initial library. Our expectation was that the most efficient TkNat10 substrates would display a high '% CCG conversion'. To correct our data for heat-catalyzed deamination, another potential contributor to conversion, we subtracted the conversion rate of samples treated with TkNat10/cofactors/heat but not reduced. STREME [26] analysis was used to identify 3-mers overrepresented in nonreduced versus sodium cyanoborohydride-reduced samples and readily identified the 5′-CCG-3′ acetylation consensus sequence, consistent with the expected activity of the enzyme (Fig. 4C) [27]. Sequences with mutations in positions 4 and 6—which flank the modified cytidine at 5′-CCG-3′—displayed >98% less modification than

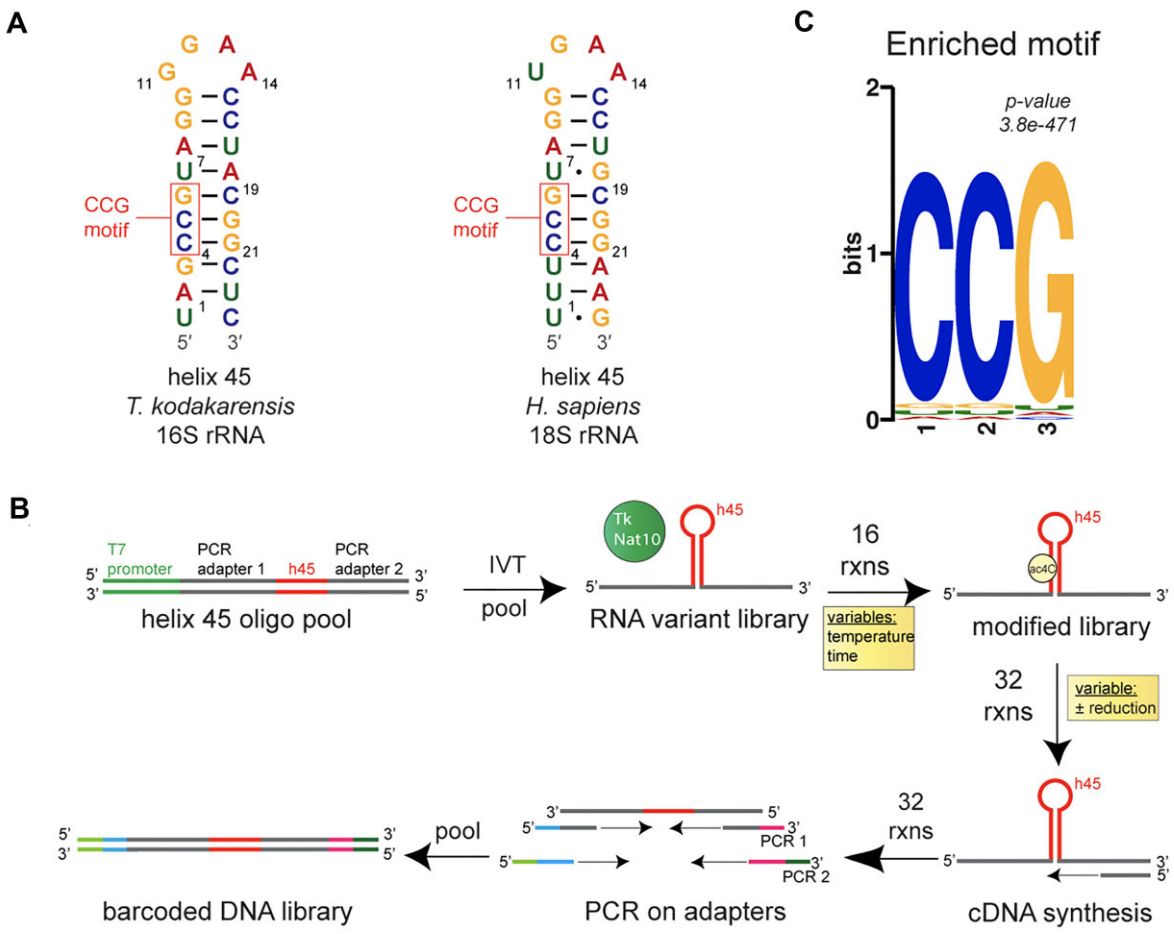


Figure 4. (**A**) Structure of 16S and 18S SSU helix 45 rRNA model substrates. (**B**) Schematic for high-throughput analysis of TkNat10 substrate-specificity using a hairpin substrate. (**C**) Simple, Thorough, Rapid, Enriched Motif Elicitation (STREME) analysis of 3-bp motifs differentially enriched in reaction ($NaCNBH_3$-treated) and control (mock-treated) libraries. Both libraries were subjected to equivalent ac4C sequencing reactions.

the parent, confirming the enzyme's preference for this motif (Supplementary Table S2 and Supplementary Fig. S6A).

Comparing activity of TkNat10 towards the parent *T. kodakarensis* and *H. sapiens* SSU rRNA sequence (Fig. 5A and B, '0′ left-most column) some immediate trends emerged. Both substrates show temperature-dependent modification that is driven to near completion after 3 h of incubation at 85°C. The human helix 45 surrogate displays time-dependent modification at 65°C, while the more GC-rich *T. kodakarensis* hairpin is significantly modified at 75°C and not lower temperatures. None of the mutations enabled modification at 37°C. Scanning across the 5′ hairpin stem, we observe two mutations, U1 → G1 in *T. kodakarensis* and U1 → C1 in *H. sapiens* that clearly decrease ac4C formation (Fig. 5C). U7 → C7 in *H. sapiens* shows an even more dominant effect, limiting ac4C deposition even at 75°C. Each of these mutations install a more stable G • C pair within the helix. This indicates stabilized duplex hairpin structures are less susceptible to TkNat10-catalyzed modification. A similar, albeit less consistent, effect can be seen upon mutation of *H. sapiens* G18 → A and G24 → A, which replace the G • U with a canonical A • U pair. All other mutations in our library targeted to the stem disrupt canonical base-pairing and largely had neutral or positive effects on acetylation. Common stem mutations which improved Circular arc plot depicting interaction of secondary mutations with C1, C7, and U20 mutations. Interaction scores were calculated from difference in summed % CCG conversion for parent and modification of both the *T. kodakarensis* and *H. sapiens* substrates were observed at positions 16, 17, 19, 20, and 21 in the 3′ stem (Fig. 5A and B). The latter three (19–20) are positioned antiparallel to the 5′-CCG-3′ consensus sequence ('anti-CCG') and appeared among the most conducive to modification (Fig. 5C, right). TkNat10's ability to modify an unpaired substrate cytidine stands in marked contrast to the recently characterized Kre33/Tan1 complex, which shows an absolute requirement for base-pairing at the modified base [14]. Tetraloop mutations influenced acetylation to a relatively minor extent, with the highest ranking sequences containing purine pairs at the internal position (Supplementary Fig. S6B).

Our inclusion of all double mutants in our *H. sapiens* helix 45 library provides the opportunity to understand the interplay of individual variants. We began by examining the two variants that form new C • G base pairs, U1 → C1 and U7 → C7. To evaluate the effect of secondary mutations on modification rate, we calculated the difference in summed % CCG conversion for parent and double mutant across three conditions (3 h 55°C, 3 h 65°C, and 5 h 65°C), using this as a measure of relative enzyme modification rate. These values were graphed as arc plots, with double mutants showing increased modification rate relative to parent in orange, those similar in effect in gray, and those showing decreased modification rate

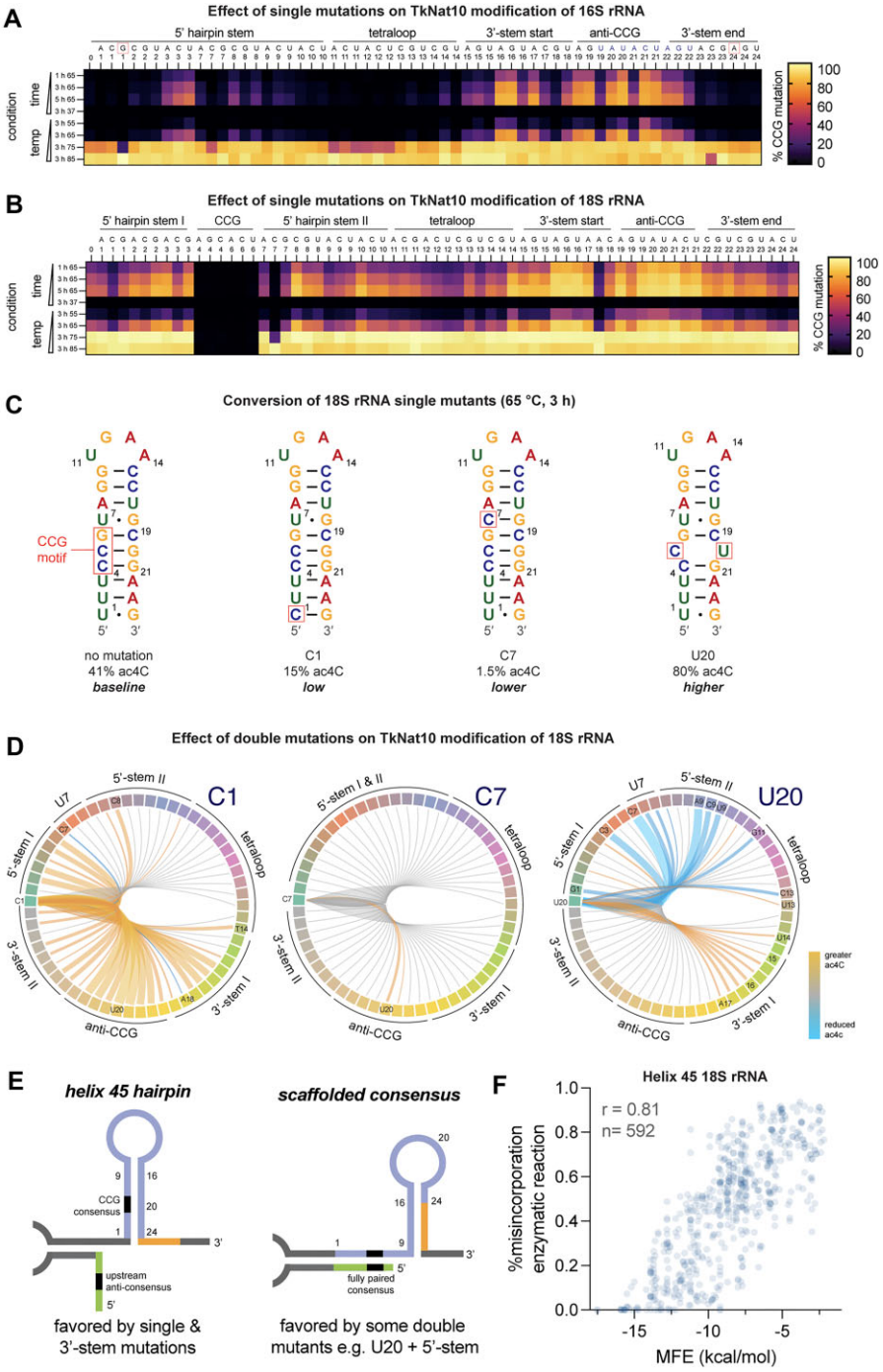


Figure 5. (**A**) Heatmap depicting % misincorporation ('% CCG mutation') as a function of time and temperature for 16S rRNA substrate. Each column corresponds to a mutant sequence with the mutation indicated in the top row of the header and the position in the bottom row of the header. Data represent the average of two replicate experiments. Raw data are given in Supplementary Table S2. (**B**) Heatmap depicting % misincorporation ('% CCG mutation') as a function of time and temperature for 18S rRNA substrate. Each column corresponds to a mutant sequence with the mutation indicated in the top row of the header and the position in the bottom row of the header. Data represent the average of two replicate experiments. Raw data are given in Supplementary Table S2. (**C**) Predicted secondary structure of 18S SSU helix 45 rRNA substate as well as C1, C7, and U20 mutants. (**D**) double mutant across three conditions (3 h 55°C, 3 h 65°C, and 5 h 65°C). The width and intensity of each arc is scaled to the strength of the interaction, with mutations that stimulate ac4C modification in orange and those that reduce it in blue. Secondary mutations with a small effect (averaging 10% or less) were depicted in gray. (**E**) Simplified schematic indicating the potential for 18S helix 45 mutation to form alternative RNA structures by interacting with scaffolding RNA. Left: The expected helix 45 surrogate forms and is predictably modulated by single mutations and double mutants in the 3′ stem. Right: When weakened helix 45 variants (e.g. U20) acquire a second mutation in the 5′ stem—particularly at position 9—they favor an alternative structure in which the consensus sequence becomes fully paired with an upstream sequence, reducing susceptibility to TkNat10 modification. Detailed exemplary structures provided in Supplementary Fig. S8. (**F**) Correlation of ac4C-seq misincorporation percentage with MFE for 18S rRNA variants. To focus exclusively on the helix 45 hairpin this analysis was tailored to mutants not predicted to form the alternative structure and included (i) all single mutants, (ii) mutants that strengthen the helix by forming new base pairs (e.g. C1, A18), and (iii) position 15–16 or 19–21 mutants with a secondary mutation in stem II. The misincorporation is derived from incubation of TkNat10 and cofactors with RNA at 3 h 65°C. C7 mutants are uniformly not modified under these conditions and thus were omitted from the graph for clarity.

relative to parent in blue (Fig. 5D). The decreased modification caused by C1 mutation is rescued by adjacent mutations at positions 2 and 3 as well as a wide variety of 3′-stem variants, most notably at positions 16, 20, and 21 (Fig. 5D, left, and Supplementary Fig. S7A). In contrast, the C7 mutation was more dominant, with ac4C deposition only being significantly restored by secondary A20 and U20 mutations (Fig. 5D, center, and Supplementary Fig. S7B).

Given the apparent anticorrelation between modification and base-pairing, we expected disruption of additional base pairs within U20 variants would enable even more efficient enzyme catalysis. Instead, we observed several mutants in stem I whose effects dominated that of U20 (Fig. 5D, right, and Supplementary Fig. S7C). One striking example came from analysis of position 9, whose mutation was uniformly associated with decreased modification of U20 helices (Fig. 5D, blue lines). In contrast, mutation of its pairing partner 16 increases acetylation (Fig. 5D, orange lines). Since our design embeds *H. sapiens* helix 45 in a larger RNA to facilitate RT and PCR, we hypothesized co-mutation of U20 and position 9 could be driving formation of a new secondary structure in which helix 45 was disrupted, but the 5′-CCG-3′ substrate was fully base-paired with an upstream 5′-CGG-3′ in the scaffold sequence. In line with this view, *in silico* structural prediction indicates the A9/U20 mutant is capable of forming an alternative structure in which the 5′-CCG-3′ consensus is housed in a fully paired duplex ('scaffolded consensus'; Fig. 5E and Supplementary Fig. S8). The alternative structure is predicted to be destabilized by mutations in positions 15–17, 23, and 24 in the 3′-stem, potentially explaining the differential effects of mutating the two stems. Further supporting this view, when the modification rate of U20 double mutants is corelated with MFE we find that 5′-stem mutations follow a parabolic distribution, with acetylation of those that highly enforce (C7) or abrogate structure (A9, C9, and U9) disfavored (Supplementary Fig. S9, blue). This suggests that 5′-stem mutations that are overly destabilizing favor the alternative structure, a relatively poor substrate. In contrast, for 3′-stem mutations MFE and modification are linearly correlated (Supplementary Fig. S9, green), as would be expected if TkNat10 displayed a straightforward preference for unstructured substrates. Analysis of modification and MFE among all single stem mutants and 3′-stem double mutants (592 data points) further confirms this anticorrelation of modification and negative free energy (Fig. 5F). This preference has some interesting parallels with SHAPE reagents [28, 29], reactive chemicals which similarly report on RNA accessibility. Overall, these studies demonstrate the utility of ac4C-seq for high-throughput profiling of RNA acetyltransferase–substrate interactions and specify the 5′-CCG-3′ consensus and RNA secondary structure as important determinants and antideterminants of *in vitro* modification by TkNat10.

Evaluating archaeal enzymes as a tool for sequence-specific RNA acetylation

A major challenge in the study of RNA cytidine acetylation has been a lack of methods to install ac4C into physiological consensus sequences. Recently, we reported a synthetic approach to address this obstacle which enabled the production of sequence-defined RNA oligomers harboring acetylation at specific positions [30]. One caveat to this approach is that it requires access to custom synthesized phosphoramidite building blocks and significant expertise with automated oligonucleotide synthesis. Considering this, we wondered whether the promiscuous activity of TkNat10 may be amenable to the acetylation of non-native substrates, providing a preparative route to RNAs containing the acetylated 5′-CCG-3′ consensus motif for functional studies. To test this, we incubated a short 10-mer RNA oligonucleotide containing a 5′-CCG-3′ sequence with TkNat10 and cofactors at 65°C for 3 h (Fig. 6A). MALDI-TOF confirmed a +42 Da mass shift consistent with the expected mass of acetylation, which was absent in the enzyme-free control reaction (Fig. 6B). LC-MS quantification of the nuclease-digested oligomer revealed 26% of ac4C modification.

Protein lysine acetyltransferases are able to use alternative acyl-CoA cofactors to catalyze formation of a wide range of acylations including propionylation and crotonylation [31–33]. The ability of RNA acetyltransferases to catalyze similarly diverse acylations has not been explored. To evaluate TkNat10's cofactor promiscuity we incubated enzyme with propionyl-CoA and ATP and analyzed modification of our short RNA oligonucleotide substrate by MALDI. Propionyl-CoA was chosen due to its relatively small footprint and known ability by many GNAT-containing protein acetyltransferases [34]. In contrast to acetyl-CoA, no modification was observed in the propionyl-CoA reaction, suggesting a strict requirement for the endogenous cofactor. To evaluate the molecular basis for this specificity we attempted to relax the specificity by enlarging the active site, expressing and purifying mutants in which Ser623 was mutated to a glycine either alone (S623G) or in combination with Trp635 (S623G/W635A; Fig. 6B, bottom, and Supplementary Fig. S10A). Analogous mutations have been shown to enable the GNAT histone acetyltransferase KAT2A to use biorthogonal azide and alkyne-containing acyl-CoAs [35]. Analyzing reaction products by MALDI, we found the TkNat10 S623G was able to use propionyl-CoA as a substrate, with the propionylation reaction appearing to proceed to a lesser extent than acetylation (Supplementary Fig. S10B). Neither enzyme was able to use larger biorthogonal acyl-CoAs as cofactors, and the dual S623G/W635A mutant was completely inactive (Supplementary Fig. S10C). These results indicate the strict specificity of cytidine acetyltransferases for acetyl-CoA and provide a mutational starting point for using these enzymes to study non-native N4-cytidine acylations.

Enzymatic modification enables analysis of ac4C-dependent RNA–protein interactions

Nucleobase modifications can act as potent regulators of RNA–protein interactions [36–38]. Previous studies have demonstrated that using IVT to homogenously replace cytidine with N4-acetylcytidine in synthetic mRNAs can alter the RNA–protein interactomes of these molecules [39]. However, this approach presents several caveats. It homogenously installs ac4C in a variety of non-native contexts, cannot access a 5′-CCG-3′ sequence in which ac4C is flanked by an unmodified cytidine nucleotide, and produces a variety of byproducts such as double-stranded antisense transcripts that may confound results [40]. To overcome these limitations, studies of RNA modifications such as N1-methylpseudouridine (m1Y) have benefitted from the use of synthetic m1Y-containing oligonucleotides to determine, for example, how this modification effects the immune response [41]. Since short syn-

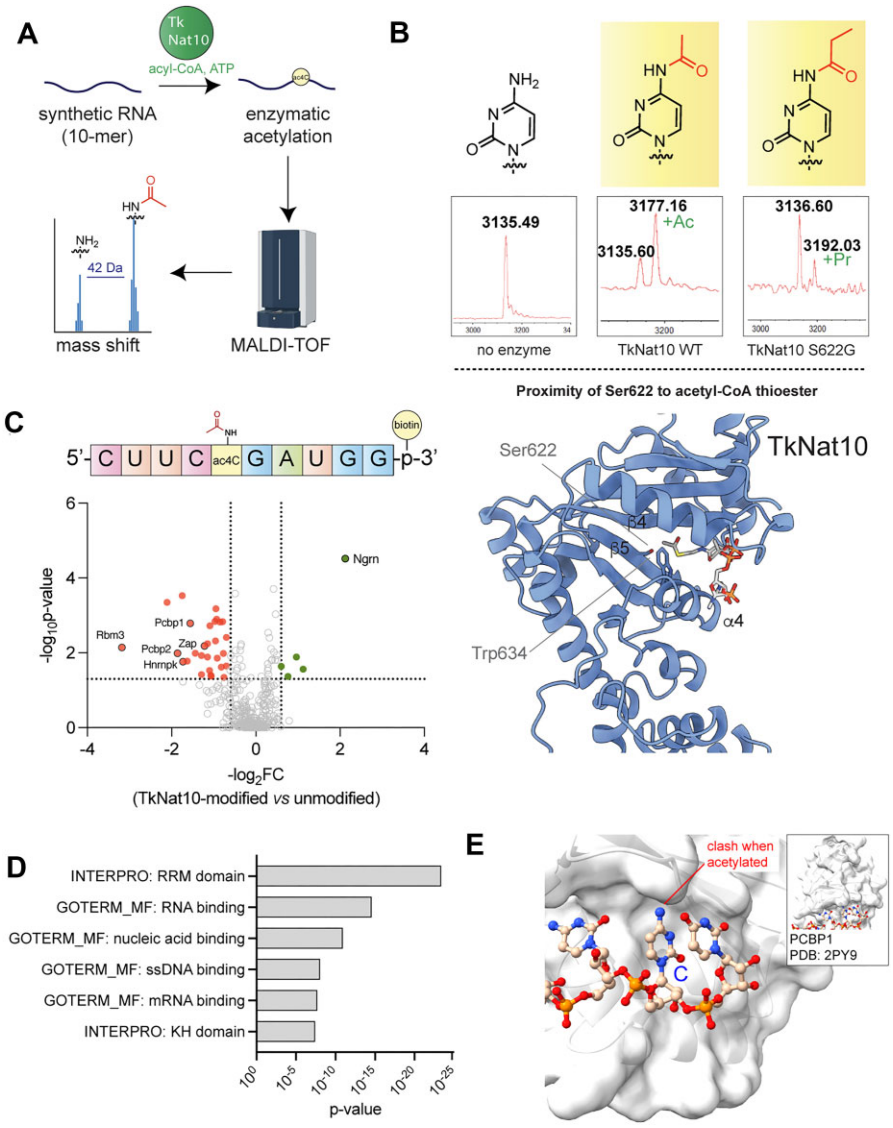


Figure 6. (**A**) Schematic for MALDI-TOF assay assessing TkNat10 RNA oligonucleotide modification. (**B**) Top: MALDI-TOF analysis of a single-stranded RNA oligonucleotide treated with no enzyme (left), TkNat10 WT + acetyl-CoA (middle), or TkNat10 S623G + propionyl-CoA (right). Full MALDI traces are provided in the Supplementary Information. Bottom: Predicted structure of TkNat10, indicating bulky residues in active site. Mutation of homologous residues in the human GNAT histone acetyltransferase enzyme KAT2A expands cofactor utilization. (**C**) LC-MS analysis comparing proteins enriched by an unmodified biotinylated RNA oligonucleotide versus a TkNat10 acetylated RNA oligonucleotide. (**D**) Gene ontology analysis of proteins preferentially enriched by an unmodified biotinylated RNA oligonucleotide versus a TkNat10-modified acetylated RNA oligonucleotide. (**E**) Crystal structure of PCBP1 bound to cytidine-containing RNA oligonucleotide. Modeling of N4-acetylation indicates a potential steric clash caused by cytidine modification.

thetic ac4C-containing oligonucleotides are not available to most researchers, we asked: could we use postsynthetic enzymatic modification to study how ac4C affects protein–RNA interactions?

To address this, a biotinylated 10-mer RNA containing a 5′-CCG-3′ sequence was obtained via commercial phosphoramidite synthesis. This substrate was incubated with TkNat10 and cofactors at 65°C for 3 h and acetylation was verified by MALDI (Supplementary Fig. S11). To test ac4C's ability to affect recognition by RNA-binding proteins (RBPs), we individually incubated two biotinylated RNAs—one with TkNat10-catalyzed ac4C and one without—with RAW 264.7 cell lysates, enriched over streptavidin, and analyzed the proteins captured by liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Fig 6C and Supplementary Table S3). Our rationale for using a short (10 bp) capture probe was to focus on differential capture driven by ac4C, as opposed to ac4C-dependent changes in RNA structure. Focusing on proteins identified by >1 peptide, we identified 18 proteins whose capture was significantly impeded by ac4C (P-value ≤ .05, $\log_2$ fold change ≤ −1) and four proteins whose capture was facilitated by ac4C. Interestingly, the majority of proteins whose capture was impeded by ac4C were RBPs (Fig 6D), while the majority of proteins whose capture was facilitated by ac4C have been identified as common contaminants in proteomics experiments [42]. One potential explanation for the latter observation is that RNA acetylation impedes RBP/resin interactions, which are then saturated by nonspecific protein binding. Among the RBPs whose interaction is disrupted by ac4C are known PCBP1, PCBP2, and ZC3HAV1, all of which have demonstrated evidence of cytidine-dependent RNA–protein interactions [43, 44]. The

potential for ac4C to interfere with cytidine-dependent protein binding is illustrated by modeling the modification into a crystal structure of RNA bound to PCBP1 (Fig 6E). Our findings provide further evidence for the ability of cytidine acetylation to influence RNA interactomes and highlight archaeal cytidine acetyltransferase enzymes as a powerful tool for functional analysis of 5′-CCG-3′ acetylation.

Discussion

Here we report the reconstitution, biochemical profiling, and functional application of a *T. kodakarensis* RNA acetyltransferase. Our studies shed light on unanswered questions around these catalysts, including the specific relevance of their acetyltransferase activity to archaeal fitness, their substrate and cofactor selectivity, and potential applications in preparative production of ac4C oligonucleotides. We demonstrate the utility of quantitative ac4C sequencing for high-throughput profiling of RNA acetyltransferase substrates. This study confirmed TkNat10's preference for single-stranded RNA substrates, suggesting a potential application of this enzyme as a structural probe. Finally, we show this promiscuous catalyst can be applied to N4-acetylate the 5′-CCG-3′ consensus motif short RNA oligonucleotide contexts and use this to examine the impact of site-specific ac4C on RNA–protein interactions.

Our results also highlight features warranting further investigation. One surprising result was the apparent incongruity between the preferred cellular versus biochemical targets of TkNat10. Applying an identical ac4C sequencing method, we found that in cells rRNA and tRNAs are the most highly modified targets of the enzyme [7], while in biochemical experiments unstructured RNA substrates are preferred. Several factors may contribute to this observation. First, our assay conditions undoubtedly present different concentrations of cofactors and ions than the endogenous intracellular environment of *T. kodakarensis*. The ability to detect modifications on coding versus noncoding RNAs can be influenced by differential sequencing depth [45]. Along similar lines, our recombinant enzyme preparations were applied to a deproteinized *T. kodakarensis* transcriptome, which is expected to have a large impact on both the presentation of rRNA [46] as well as mRNA–protein interactions that could influence mRNA substrate accessibility [47]. Perhaps the most intriguing hypothesis for this dissimilarity may come from consideration of eukaryotic RNA acetylation, where snoRNA and protein adapters are used to guide ac4C to specific sites [11, 12]. Archaea have equivalent sno-like RNA (sRNA) mechanisms which help carry out modifications such as 2′-O-methylation [48]. Future studies aiming to define the input of sRNAs as well as other proteins into archaeal ac4C deposition should help to better understand this phenomenon. Our observation that TkNat10 requires engineering to accept propionyl-CoA implies stringent cofactor specificity, suggesting catalysis of these alternative acylations in RNA may be evolutionarily constrained. TkNat10's preference for unfolded substrates is interesting in light of recent studies indicating ac4C can serve to enforce RNA secondary structure [7, 30]. This could conceivably allow the stand-alone enzyme to perform a sentinel function by sensing unfolded RNAs and installing a modification to promote folding and—potentially—fitness. Testing this hypothesis will benefit from additional research exploring ac4C's capacity to function as an RNA structure switch, which should in turn be facilitated by the enzymatic modification strategies described here.

It is also important to specify some limitations of our study and future work that may help address them. Our high-throughput studies explored TkNat10's activity within the context of only two hairpin substrates and used MFE as a simple proxy for RNA structure. Integrating analysis of more diverse substrates structure probing data could provide a more comprehensive understanding of archaeal RNA acetyltransferase activity and guide future applications. A barrier faced in our biochemical characterization was the requirement for heating, which hindered application of conventional coupled enzyme techniques to determine kinetic constants for TkNat10. We anticipate this may be remedied in the future by using stop-point assays in combination with highly specific and sensitive tools for acetyl-CoA or CoA detection [49, 50]. Finally, while we were able to demonstrate near stoichiometric enzymatic modification of a short RNA oligonucleotide, TkNat10's utility as a tool for site-specific RNA acetylation is currently limited to contexts where a consensus motif exists. Although the degree of modification can vary quite widely on a quantitative level, our finding that 85% of deeply sequenced 5′-CCG-3′ sites in the *T. kodakarensis* transcriptome can be modified by TkNat10 (75°C) suggests most RNAs containing this sequence should be at least partially modifiable. Efforts to broaden the substrate selectivity of this enzyme may benefit from directed evolution strategies capable of converting modifications into functional changes in RNA or protein activity, examples of which have been reported [51]. Overall, our studies provide new insights into the biochemistry of a highly conserved enzyme family and a foundation for studying the role of site-specific RNA acetylation in biology, therapeutics, and disease.

Acknowledgements

We thank Ronald Holewinski, Xia Xu, and Thorkell Andresson (Protein Characterization Laboratory, Frederick National Laboratory for Cancer Research) for LC/MS/MS and proteomic analysis. The graphical abstract was created using BioRender and can be accessed at https://BioRender.com/y54y323. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

Author contributions: Supuni Thalalla Gamage (Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Validation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]), Shereen Howpay Manage (Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Validation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]), Aldema Sas-Chen (Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Validation [equal], Visualization [equal]), Ronit Nir (Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Validation [equal], Visualization [equal]), Brett W. Burkhart (Data curation [equal], Formal analysis [equal], Methodology [equal], Validation [equal], Visualization [equal]), Isita Jhulki (Formal analysis [equal], Methodology [equal], Validation [equal], Visualiza-

tion [equal]), Courtney N. Link (Data curation [equal], Formal analysis [equal], Methodology [equal], Validation [equal], Visualization [equal]), Manini S. Penikalapati (Data curation [equal], Methodology [equal]), Jane E. Jones (Methodology [equal], Resources [equal]), Lakshminarayan M. Iyer (Formal analysis [equal], Investigation [equal], Software [equal]), L. Aravind (Formal analysis [equal], Investigation [equal], Resources [equal], Supervision [equal]), Thomas J. Santangelo (Conceptualization [equal], Funding acquisition [equal], Investigation [equal], Methodology [equal], Project administration [equal], Resources [equal], Supervision [equal], Validation [equal], Visualization [equal], Writing—original draft [equal]), Schraga Schwartz (Conceptualization [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Project administration [equal], Resources [equal], Software [equal], Supervision [equal], Writing—original draft [equal], Writing—review & editing [equal]), and Jordan L. Meier (Conceptualization [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Project administration [equal], Resources [equal], Software [equal], Supervision [equal], Validation [equal], Writing—original draft [equal], Writing—review & editing [equal])

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work is supported by the Intramural Research Program of the National Institutes of Health (NIH), the National Cancer Institute, and The Center for Cancer Research (ZIA BC011488) (J.L.M.). S.S. is funded by the Israel Science Foundation (grant no. 913/21), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant no. 101000970). A.S.C. is supported by the Israel Science Foundation (grant no. 1260/23) and by the Azrieli Foundation. T.J.S is supported by the USA National Science Foundation, award #2022065. L.A. was supported by the Intramural Research Program of the National Library of Medicine at the National Institutes of Health. In addition, this research was carried out whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract 75N91019D00024 (J.E.J.).

Data availability

All the raw data of high throughput sequencing has been deposited at the NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA1172712. Proteomics data are available at MassIVE under accession number MSV000096084. Custom perl scripts used in this study can be found at the GitHub repository, https://github.com/Meier-Lab-NCI/TkNat10 and in Zenodo, https://doi.org/10.5281/zenodo.14961019.

References

1. Thalalla Gamage S, Howpay Manage SA, Chu TT *et al.* Cytidine acetylation across the tree of life. *Acc Chem Res* 2024;**57**:338–48. https://doi.org/10.1021/acs.accounts.3c00673
2. Stern L, Schulman LH. The role of the minor base N4-acetylcytidine in the function of the *Escherichia coli* noninitiator methionine transfer RNA. *J Biol Chem* 1978;**253**:6132–9. https://doi.org/10.1016/S0021-9258(17)34590-8
3. Ikeuchi Y, Kitahara K, Suzuki T. The RNA acetyltransferase driven by ATP hydrolysis synthesizes N4-acetylcytidine of tRNA anticodon. *EMBO J* 2008;**27**:2194–203. https://doi.org/10.1038/emboj.2008.154
4. Johansson MJ, Bystrom AS. The Saccharomyces cerevisiae TAN1 gene is required for N4-acetylcytidine formation in tRNA. *RNA* 2004;**10**:712–9. https://doi.org/10.1261/rna.5198204
5. Ito S, Akamatsu Y, Noma A *et al.* A single acetylation of 18 S rRNA is essential for biogenesis of the small ribosomal subunit in Saccharomyces cerevisiae. *J Biol Chem* 2014;**289**:26201–12. https://doi.org/10.1074/jbc.M114.593996
6. Sharma S, Langhendries JL, Watzinger P *et al.* Yeast Kre33 and human NAT10 are conserved 18S rRNA cytosine acetyltransferases that modify tRNAs assisted by the adaptor Tan1/THUMPD1. *Nucleic Acids Res* 2015;**43**:2242–58. https://doi.org/10.1093/nar/gkv075
7. Sas-Chen A, Thomas JM, Matzov D *et al.* Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. *Nature* 2020;**583**:638–43. https://doi.org/10.1038/s41586-020-2418-2
8. Orita I, Futatsuishi R, Adachi K *et al.* Random mutagenesis of a hyperthermophilic archaeon identified tRNA modifications associated with cellular hyperthermotolerance. *Nucleic Acids Res* 2019;**47**:1964–76. https://doi.org/10.1093/nar/gky1313
9. Coureux PD, Lazennec-Schurdevin C, Bourcier S *et al.* Cryo-EM study of an archaeal 30S initiation complex gives insights into evolution of translation initiation. *Commun Biol* 2020;**3**:58. https://doi.org/10.1038/s42003-020-0780-0
10. Chimnaronk S, Suzuki T, Manita T *et al.* RNA helicase module in an acetyltransferase that modifies a specific tRNA anticodon. *EMBO J* 2009;**28**:1362–73. https://doi.org/10.1038/emboj.2009.69
11. Sharma S, Yang J, van Nues R *et al.* Specialized box C/D snoRNPs act as antisense guides to target RNA base acetylation. *PLoS Genet* 2017;**13**:e1006804. https://doi.org/10.1371/journal.pgen.1006804
12. Bortolin-Cavaille ML, Quillien A, Thalalla Gamage S *et al.* Probing small ribosomal subunit RNA helix 45 acetylation across eukaryotic evolution. *Nucleic Acids Res* 2022;**50**:6284–99. https://doi.org/10.1093/nar/gkac404
13. Thalalla Gamage S, Bortolin-Cavaille ML, Link C *et al.* Antisense pairing and SNORD13 structure guide RNA cytidine acetylation. *RNA* 2022;**28**:1582–96.
14. Ma CR, Liu N, Li H *et al.* Activity reconstitution of Kre33 and Tan1 reveals a molecular ruler mechanism in eukaryotic tRNA acetylation. *Nucleic Acids Res* 2024;**52**:5226–40. https://doi.org/10.1093/nar/gkae262
15. Zhou M, Gamage ST, Tran KA *et al.* Molecular basis for RNA cytidine acetylation by NAT10. bioRxiv, https://doi.org/10.1101/2024.03.27.587050, 27 March 2024, preprint: not peer reviewed.
16. van den Burg B. Extremophiles as a source for novel enzymes. *Curr Opin Microbiol* 2003;**6**:213–8. https://doi.org/10.1016/S1369-5274(03)00060-2
17. Ye K, Jia R, Lin J *et al.* Structural organization of box C/D RNA-guided RNA methyltransferase. *Proc Natl Acad Sci USA* 2009;**106**:13808–13. https://doi.org/10.1073/pnas.0905128106
18. Bryson KM, Thalalla-Gamage S, Meier JL. Visualizing RNA cytidine acetyltransferase activity by northern blotting. *Curr Protoc Chem Biol* 2020;**12**:e89. https://doi.org/10.1002/cpch.89

19. Thalalla Gamage S, Sas-Chen A, Schwartz S *et al.* Quantitative nucleotide resolution profiling of RNA cytidine acetylation by ac4C-seq. *Nat Protoc* 2021;**16**:2286–307. https://doi.org/10.1038/s41596-021-00501-9
20. Behrens A, Nedialkova DD. Experimental and computational workflow for the analysis of tRNA pools from eukaryotic cells by mim-tRNAseq. *STAR Protoc* 2022;**3**:101579. https://doi.org/10.1016/j.xpro.2022.101579
21. Cavalli G, Heard E. Advances in epigenetics link genetics to the environment and disease. *Nature* 2019;**571**:489–99. https://doi.org/10.1038/s41586-019-1411-0
22. Sinclair WR, Arango D, Shrimp JH *et al.* Profiling cytidine acetylation with specific affinity and reactivity. *ACS Chem Biol* 2017;**12**:2922–6. https://doi.org/10.1021/acschembio.7b00734
23. Thomas JM, Briney CA, Nance KD *et al.* A chemical signature for cytidine acetylation in RNA. *J Am Chem Soc* 2018;**140**:12667–70. https://doi.org/10.1021/jacs.8b06636
24. Lorenz R, Bernhart SH, Siederdissen CHZ *et al.* ViennaRNA package 2.0. *Algorithms Mol Biol* 2011;**6**:26. https://doi.org/10.1186/1748-7188-6-26
25. Purchal MK, Eyler DE, Tardu M *et al.* Pseudouridine synthase 7 is an opportunistic enzyme that binds and modifies substrates with diverse sequences and structures. *Proc Natl Acad Sci USA* 2022;**119**:e2109708119. https://doi.org/10.1073/pnas.2109708119
26. Bailey TL. STREME: accurate and versatile sequence motif discovery. *Bioinformatics* 2021;**37**:2834–40. https://doi.org/10.1093/bioinformatics/btab203
27. Bailey TL. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* 2011;**27**:1653–9. https://doi.org/10.1093/bioinformatics/btr261
28. Weeks KM, Mauger DM. Exploring RNA structural codes with SHAPE chemistry. *Acc Chem Res* 2011;**44**:1280–91. https://doi.org/10.1021/ar200051h
29. Spitale RC, Crisalli P, Flynn RA *et al.* RNA SHAPE analysis in living cells. *Nat Chem Biol* 2013;**9**:18–20. https://doi.org/10.1038/nchembio.1131
30. Bartee D, Nance KD, Meier JL. Site-specific synthesis of N(4)-acetylcytidine in RNA reveals physiological duplex stabilization. *J Am Chem Soc* 2022;**144**:3487–96. https://doi.org/10.1021/jacs.1c11985
31. Montgomery DC, Sorum AW, Guasch L *et al.* Metabolic regulation of histone acetyltransferases by endogenous acyl-CoA cofactors. *Chem Biol* 2015;**22**:1030–9. https://doi.org/10.1016/j.chembiol.2015.06.015
32. Ringel AE, Wolberger C. Structural basis for acyl-group discrimination by human Gcn5L2. *Acta Crystallogr D Struct Biol* 2016;**72**:841–8. https://doi.org/10.1107/S2059798316007907
33. Sabari BR, Tang Z, Huang H *et al.* Intracellular crotonyl-CoA stimulates transcription through p300-catalyzed histone crotonylation. *Mol Cell* 2015;**58**:203–15. https://doi.org/10.1016/j.molcel.2015.02.029
34. Simithy J, Sidoli S, Yuan ZF *et al.* Characterization of histone acylations links chromatin modifications with metabolism. *Nat Commun* 2017;**8**:1141. https://doi.org/10.1038/s41467-017-01384-9
35. Yang C, Mi J, Feng Y *et al.* Labeling lysine acetyltransferase substrates with engineered enzymes and functionalized cofactor surrogates. *J Am Chem Soc* 2013;**135**:7791–4. https://doi.org/10.1021/ja311636b
36. Edupuganti RR, Geiger S, Lindeboom RGH *et al.* N(6)-methyladenosine (m^6A) recruits and repels proteins to regulate mRNA homeostasis. *Nat Struct Mol Biol* 2017;**24**:870–8. https://doi.org/10.1038/nsmb.3462
37. Yang X, Yang Y, Sun BF *et al.* 5-methylcytosine promotes mRNA export – NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. *Cell Res* 2017;**27**:606–25. https://doi.org/10.1038/cr.2017.55
38. Nance KD, Meier JL. Modifications in an emergency: the role of N1-methylpseudouridine in COVID-19 vaccines. *ACS Cent Sci* 2021;**7**:748–56. https://doi.org/10.1021/acscentsci.1c00197
39. Nance KD, Gamage ST, Alam MM *et al.* Cytidine acetylation yields a hypoinflammatory synthetic messenger RNA. *Cell Chem Biol* 2022;**29**:312–20. https://doi.org/10.1016/j.chembiol.2021.07.003
40. Mu X, Greenwald E, Ahmad S *et al.* An origin of the immunogenicity of *in vitro* transcribed RNA. *Nucleic Acids Res* 2018;**46**:5239–49. https://doi.org/10.1093/nar/gky177
41. Nelson J, Sorensen EW, Mintri S *et al.* Impact of mRNA chemistry and manufacturing process on innate immune activation. *Sci Adv* 2020;**6**:eaaz6893. https://doi.org/10.1126/sciadv.aaz6893
42. Mellacheruvu D, Wright Z, Couzens AL *et al.* The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods* 2013;**10**:730–6. https://doi.org/10.1038/nmeth.2557
43. Du Z, Fenn S, Tjhen R *et al.* Structure of a construct of a human poly(C)-binding protein containing the first and second KH domains reveals insights into its regulatory mechanisms. *J Biol Chem* 2008;**283**:28757–66. https://doi.org/10.1074/jbc.M803046200
44. Meagher JL, Takata M, Goncalves-Carneiro D *et al.* Structure of the zinc-finger antiviral protein in complex with RNA reveals a mechanism for selective targeting of CG-rich viral sequences. *Proc Natl Acad Sci USA* 2019;**116**:24303–9. https://doi.org/10.1073/pnas.1913232116
45. Kong Y, Mead EA, Fang G. Navigating the pitfalls of mapping DNA and RNA modifications. *Nat Rev Genet* 2023;**24**:363–81. https://doi.org/10.1038/s41576-022-00559-5
46. Rouskin S, Zubradt M, Washietl S *et al.* Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. *Nature* 2014;**505**:701–5. https://doi.org/10.1038/nature12894
47. Corley M, Burns MC, Yeo GW. How RNA-binding proteins interact with RNA: molecules and mechanisms. *Mol Cell* 2020;**78**:9–29. https://doi.org/10.1016/j.molcel.2020.03.011
48. Dennis PP, Omer A, Lowe T. A guided tour: small RNA function in Archaea. *Mol Microbiol* 2001;**40**:509–19. https://doi.org/10.1046/j.1365-2958.2001.02381.x
49. Xue L, Schnacke P, Frei MS *et al.* Probing coenzyme A homeostasis with semisynthetic biosensors. *Nat Chem Biol* 2023;**19**:346–55. https://doi.org/10.1038/s41589-022-01172-7
50. Lieberman WK, Brown ZA, Kantner DS *et al.* Chemoproteomics yields a selective molecular host for acetyl-CoA. *J Am Chem Soc* 2023;**145**:16899–905. https://doi.org/10.1021/jacs.3c05489
51. Zhou H, Rauch S, Dai Q *et al.* Evolution of a reverse transcriptase to map N^1-methyladenosine in human messenger RNA. *Nat Methods* 2019;**16**:1281–8. https://doi.org/10.1038/s41592-019-0550-4
52. Cheng J, Bassler J, Fischer P *et al.* Thermophile 90S pre-ribosome structures reveal the reverse order of Co-transcriptional 18S rRNA subdomain integration. *Mol Cell* 2019;**75**:1256–69. https://doi.org/10.1016/j.molcel.2019.06.032

Received: October 10, 2024. **Revised:** February 26, 2025. **Editorial Decision:** February 28, 2025. **Accepted:** March 7, 2025