מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

# Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping

**Document Version:**
Accepted author manuscript (peer-reviewed)

**Citation for published version:**
Sas-Chen, A, Thomas, JM, Matzov, D, Taoka, M, Nance, KD, Nir, R, Bryson, KM, Shachar, R, Liman, GLS, Burkhart, BW, Gamage, ST, Nobe, Y, Briney, CA, Levy, MJ, Fuchs, RT, Robb, GB, Hartmann, J, Sharma, S, Lin, Q, Florens, L, Washburn, MP, Isobe, T, Santangelo, TJ, Shalev-Benami, M, Meier, JL & Schwartz, S 2020, 'Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping', *Nature*, vol. 583, no. 7817, pp. 638-643. https://doi.org/10.1038/s41586-020-2418-2

*Total number of authors:*
26

**Digital Object Identifier (DOI):**
10.1038/s41586-020-2418-2

**Published In:**
Nature

**License:**
Other

**General rights**
@ 2020 This manuscript version is made available under the above license via The Weizmann Institute of Science Open Access Collection is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognize and abide by the legal requirements associated with these rights.

**How does open access to this work benefit you?**
Let us know @ library@weizmann.ac.il

**Take down policy**
The Weizmann Institute of Science has made every reasonable effort to ensure that Weizmann Institute of Science content complies with copyright restrictions. If you believe that the public display of this file breaches copyright please contact library@weizmann.ac.il providing details, and we will remove access to the work immediately and investigate your claim.

(article begins on next page)

# Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping

Aldema Sas-Chen[1,*], Justin M. Thomas[2,*], Donna Matzov[3,*], Masato Taoka[4], Kellie D. Nance[2], Ronit Nir[1], Keri M. Bryson[2], Ran Shachar[1], Geraldy L.S. Liman[5], Brett W. Burkhart[5], Supuni Thalalla Gamage[2], Yuko Nobe[4], Chloe A. Briney[2], Michaella J. Levy[6], Ryan T. Fuchs[7], G. Brett Robb[7], Jesse Hartmann[1], Sunny Sharma[8], Qishan Lin[9], Laurence Florens[6], Michael P. Washburn[6], Toshiaki Isobe[4], Thomas J. Santangelo[5], Moran Shalev-Benami[3,#], Jordan L. Meier[2,#], Schraga Schwartz[1,#]

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7610001, Israel

[2]National Cancer Institute, NIH, Frederick MD 21702-1201, USA

[3]Department of Structural Biology, Weizmann Institute of Science, Rehovot 7610001, Israel

[4]Department of Chemistry, Graduate School of Science and Engineering, Tokyo Metropolitan University, Minami-osawa 1-1, Hachioji-shi, Tokyo 192-0397, Japan

[5]Department of Biochemistry and Molecular Biology, Colorado State University, Colorado, USA

[6]Stowers Institute for Medical Research, Kansas City, USA

[7]RNA Research Division, New England Biolabs, Inc., 240 County Road, Ipswich, Massachusetts 01938 USA

[8]Department of Cell Biology and Neuroscience, Rutgers University, Piscataway NJ, 08854, USA.

[9]RNA Epitranscriptomics and Proteomics Resource, University at Albany, Albany NY, 12222, USA.

[*]These authors have contributed equally to this manuscript

[#]Corresponding authors

Moran Shalev-Benami - moransb@weizmann.ac.il

Jordan Meier - jordan.meier@nih.gov

Schraga Schwartz - schwartz@weizmann.ac.il

**N4-acetylcytidine ($ac^4C$) is an ancient and highly conserved RNA modification, present on tRNA, rRNA and recently investigated in eukaryotic mRNA[1–3]. We report $ac^4C$-seq, a chemical genomic method for single-nucleotide resolution, transcriptome-wide quantitative mapping of $ac^4C$. While we did not find detectable $ac^4C$ sites in human and yeast mRNAs, $ac^4C$ was induced via ectopic overexpression of eukaryotic acetyltransferase complexes, invariably at a conserved sequence motif. In contrast, cross-evolutionary profiling reveals unprecedented levels of $ac^4C$ across hundreds of residues in rRNA, tRNA, ncRNA and mRNA from hyperthermophilic archaea. $Ac^4C$ is dramatically induced in response to temperature, and acetyltransferase-deficient archaeal strains exhibit temperature-dependent growth defects. Cryo-EM visualization of WT and acetyltransferase-deficient archaeal ribosomes furnishes structural insights into the temperature-dependent distribution of $ac^4C$ and its potential thermoadaptive role. Our studies quantitatively define the $ac^4C$ landscape, providing a technical and conceptual foundation for unravelling this modification's role in biology and disease[4–6].**

Acetylation is an ancient mechanism for regulating biomolecular function. Perhaps the most well-conserved of these mechanisms is the enzymatic modification of RNA to form the acetylated nucleobase N4-acetylcytidine ($ac^4C$). $Ac^4C$ occurs in all domains of life and in humans is catalyzed by the acetyltransferase Nat10 (Kre33 in yeast)[1–3]. Nat10/Kre33 is essential in both human and yeast, and its four target sites in rRNA and tRNA are also conserved between these two distant eukaryotes[1–3]. Deposition of $ac^4C$ at its two tRNA targets (tRNA-Ser and tRNA-Leu) requires an additional adapter protein, Thumpd1/Tan1[1], and has been implicated in tRNA stability[7,8]. Conversely, Nat10 is guided towards its two target sites in rRNA by specialized snoRNAs[9]. Recently, antibody-based mapping suggested the existence of additional Nat10-regulated $ac^4C$ sites in human mRNAs;[10] however, the lack of base resolution quantification of any single $ac^4C$ site precluded orthogonal validation and functional prioritization based on modification stoichiometries. Thus, the quantitative distribution of $ac^4C$ amongst rRNA, tRNA, and mRNA remains to be comparatively defined in any organism.

## Nucleotide Resolution $Ac^4C$ Sequencing

To quantitatively study cytidine acetylation in the transcriptome we developed a chemical method to allow sensitive nucleotide resolution detection of $ac^4C$. Building on previous work[11], we found reaction of $ac^4C$ with sodium cyanoborohydride ($NaCNBH_3$) under acidic conditions forms a reduced nucleobase, N4-acetyl-tetrahydrocytidine, whose altered structure causes incorporation of non-cognate deoxynucleotide triphosphates (dNTPs) upon reverse transcription[11], which can be detected via cDNA sequencing. Compared to previous chemistries, this reaction shows faster kinetics and causes increased misincorporation at known $ac^4C$ sites in rRNA (**Supplementary Note 1, Extended Data Fig. 1**). Critically, $ac^4C$-dependent mutations are not observed when the modification is hydrolyzed (chemically deacetylated) using mild alkali prior to analysis[12] (**Fig. 1a**). Integrating these chemistries with next-generation sequencing allowed development of $ac^4C$-seq, a method enabling the transcriptome-wide, quantitative

analysis of ac$^4$C at single nucleotide resolution (**Fig. 1b** and **Methods**). Inspection of sequencing data revealed that $NaCNBH_3$ treatment caused C->T misincorporation at acetylated sites, which were reduced upon alkali-induced deacetylation (**Fig. 1c**). This guided development of an analytical pipeline for ac$^4$C detection, based on (1) C->T misincorporations upon acid/$NaCNBH_3$ treatment, (2) reduced C->T misincorporation upon the alkali pre-treatment, and (3) absence of C->T misincorporation in mock-treated RNA. These three requirements were formalized as two statistical tests, comparing misincorporations in $NaCNBH_3$-treated samples to those in alkali- or mock-treated controls. In practice, excellent signal to noise ratios could be obtained on the basis of the latter comparison, allowing the former to be used as an optional filter to increase confidence in identified sites (**Fig. 1d**). To evaluate our ability to quantitatively measure acetylation levels, we applied ac$^4$C-seq to four synthetic RNAs each harboring a single ac$^4$C site. In these synthetic RNAs, ac$^4$C was embedded within several sequence contexts, and spiked into complex RNA samples at varying stoichiometries (**Supplementary Table 1a**). We observed excellent *absolute* agreement between the synthesized ac$^4$C stoichiometries and the experimentally measured misincorporation levels (R= 0.99) across the entire range of stoichiometries (**Fig. 1e**). Thus, given sufficient read-depth, ac$^4$C-seq is able to detect and quantify even low-stoichiometry (4%) modifications with excellent accuracy and precision.

## Ac$^4$C in Eukaryotic RNA

We next explored the properties of ac$^4$C in eukaryotic RNA. To empower this study, we employed a cross-evolutionary approach, analyzing two human cell lines and the budding yeast, *S. cerevisiae*. Applying ac$^4$C-seq to total RNA from these organisms recapitulated both known sites of ac$^4$C modification in 18S rRNA, as well as the two known sites of ac$^4$C on serine- and leucine-tRNA (**Fig. 1f Extended Data Fig. 2a-c** and **Supplementary Table 2**).[1–3] No additional rRNA or tRNA sites met detection thresholds. Acetylation of rRNA and tRNA sites were reduced following disruption of human Nat10, and eliminated following mutation of yeast Kre33 (**Fig. 1d**, red bars and **Extended Data Fig. 2d)**. These results suggest eukaryotic rRNA and tRNA ac$^4$C is well-annotated, and that in these abundant RNAs, ac$^4$C-seq demonstrates outstanding sensitivity and specificity.

Next we explored the properties of ac$^4$C in eukaryotic mRNA[10,13]. Applying ac$^4$C-seq to poly(A)-enriched mRNA from HEK-293T cells readily identified the known sites on rRNA (**Extended Data Fig. 2e**). However, only 4 additional C->T misincorporations passed detection thresholds (**Fig. 1g**, blue bars and **Supplementary Table 2**), a number consistent with the anticipated false discovery rate **(Fig. 1g)**. To address the possibility that absence of detectable ac$^4$C in mRNA is unique to HEK-293T cells, we applied ac$^4$C-seq to poly(A)-RNA isolated from HeLa cells and the budding yeast *S. cerevisiae*, where ac$^4$C was previously suggested to be present using other approaches[1,10,13]. In both models we detected the known rRNA ac$^4$C sites **(Extended Data Fig. 2a,c** and **Supplementary Table 2)**. However, no additional sites passed detection thresholds in HeLa cells, and in yeast 3 additional sites were identified in mRNA, but were not eliminated following mutation of yeast Kre33, suggesting they do not represent ac$^4$C sites (**Supplementary Table 2**), and no enrichment was observed for C->T misincorporations. Thus, while these observations do not rule out rare or low stoichiometry acetylation sites (**Extended**

**Data Fig. 2f-h**), our studies provide no confirmatory evidence for the presence of $ac^4C$ in eukaryotic mRNA.

To understand the *potential* for eukaryotic cytidine acetyltransferases to modify mRNA, we co-overexpressed the Nat10/Thumpd1 complex in HEK-293T cells, and the orthologous Kre33/Tan1 complex in yeast (**Extended Data Fig. 3a-d**). Remarkably, overexpression of these complexes led to identification of 146 and 66 putative novel $ac^4C$ sites in human and yeast mRNA, respectively (**Fig. 1g**, **Extended Data Fig. 3e** and **Supplementary Table 2**). Misincorporation levels within mRNA remained modest (median: 7.7% and 4.9% in human and yeast, respectively) even when Nat10/Thumpd1 were co-overexpressed at very high levels (**Fig 1h, Extended Data Fig. 3a,d**). Targeted deep sequencing of 5 of these sites (median: 120,000 reads/site) recapitulated acetylation upon dual overexpression of Nat10/Thump1 (~3-4% misincorporation), while misincorporation rates in RNA from cells in which only one protein was overexpressed were on the order of 0.2%, identical to WT **(Fig. 1i** and **Extended Data Fig. 3e; Supplementary Note 2)**. To characterize Nat10/Thumpd1 substrates and explore elements directing their specificity, we performed additional analysis of induced eukaryotic $ac^4C$ sites. Remarkably, we found that 154 of 157 (98%) sites in human and 73 of 74 sites in yeast (98.6%) occurred at a 'CCG' motif, with the central 'C' being acetylated **(Fig. 1g,j,k and Extended Data Fig. 3e)**. Of note, all four $ac^4C$ sites previously identified in eukaryotic rRNA and tRNA occur within precisely this motif **(Extended Data Fig. 3f)**. Induced $ac^4C$ sites were randomly distributed across genes and displayed no preference for a particular position in a codon **(Extended Data Fig. 3g,h)**. The obligate nature of the CCG motif was validated by plasmid-based reconstitution of an inducible $ac^4C$ site, whose Nat10/Thumpd1-dependent acetylation was abolished by mutation of the 'G' immediately downstream of the acetylated site **(Fig. 1l)**. Systematic mutagenesis experiments further suggested that base-paired structural elements play a role in $ac^4C$ deposition, suggesting why 'CCG' is required, but not sufficient, for induced acetylation (**Extended Data Fig. 4**)**.** Overall, our studies define rRNA and tRNA as the predominant sites of $ac^4C$ in eukaryotes, suggest $ac^4C$ is absent or present at very low levels in endogenous eukaryotic mRNA, and demonstrate RNA acetylation can be induced at hundreds of sites via dual overexpression of Nat10/Thumpd1, invariably within a C**C**G motif.

## Unprecedented $Ac^4C$ Levels in Archaeal RNA

A cross-evolutionary analysis of total RNA by LC-MS (liquid chromatography–mass spectrometry) revealed high concentrations of $ac^4C$ in the archaeal hyperthermophile *Thermococcus kodakarensis* (**Fig. 2a**)[14]. Motivated by this, we applied $ac^4C$-seq to quantitatively map cytidine acetylation in *T. kodakarensis* cultured at its optimal growth temperature of 85˚C. We found an unprecedented number (404) of $ac^4C$ sites spread across rRNA, tRNA, non-coding RNAs, and mRNA (**Extended Data Fig. 5a**). 99% of sites occurred within C**C**G motifs and were highly enriched for C->T misincorporation signatures (**Fig. 2b**). To validate these identifications, we performed quantitative tandem LC-MS analysis of purified and partially digested *T. kodakarensis* rRNA[15]. This revealed 25 uniquely mapped $ac^4C$ sites, fully overlapping with positions identified via $ac^4C$-seq (**Fig. 2c, Supplementary Data 1 and Supplementary Table 3**). LC-MS-based estimates of modification stoichiometry agreed excellently (R= 0.97) with

ac$^4$C-seq (**Fig. 2c, Supplementary Table 4**). Deletion of the *Nat10* homologue, *TK0754* (hereafter: 'TkNat10'; recently reported to acetylate *T. kodakarensis* tRNA[16]), but not of *Thumpd1* homolog, *TK2097* ('TkThumpd1'), caused complete loss of ac$^4$C in all RNA substrates (**Fig. 2d**), a result confirmed by ac$^4$C-specific northern blotting and MS analysis (**Extended Data Fig. 5b-g**, **Supplementary Table 5a**). To understand whether pervasive RNA acetylation is a common feature of archaeal extremophiles, we used ac$^4$C-seq to profile *Pyrococcus furiosus* and *Thermoccus* sp. *AM4*, close euryarchaeal relatives of *T. kodakarensis* within the order Thermococcales, and the more phylogenetically distant species *Methanocaldococcus jannaschii* (Euryarchaeota/Methanococcales) and *Saccharolobus solfataricus* (Crenarchaeota/Sulfolobales), for evolutionary-breadth. This revealed ac$^4$C is widespread within each of the *Thermococcales* species, occurring at hundreds of sites across diverse RNA types (**Fig. 2e and Supplementary Table 2**), almost exclusively within C**C**G consensus motifs. In *T. AM4* and *P. furiosus*, ac$^4$C was not only widely present, but the precise sites and stoichiometry of ac$^4$C were also highly conserved (**Fig. 2f,g and Extended Data Fig. 5h**). In contrast, ac$^4$C detected in *S. solfataricus* was confined to 41 C**C**G sites mostly in tRNAs (**Fig. 2e** and **Supplementary Table 2**), while *M. jannaschii* lacked ac$^4$C entirely, consistent with the absence of an apparent *Nat10* homologue in this organism[17]. These studies establish the existence and regulation of prevalent RNA acetylation in the archaeal order Thermococcales.

## Dynamic Acetylation of Archaeal RNA

To investigate how ac$^4$C responds to environmental cues, we applied ac$^4$C-seq to RNA from *T. kodakarensis* cultures grown at 55-95˚C, which spans the range of temperatures at which this organism can be cultivated. This revealed ac$^4$C across all classes of RNA increases dramatically with temperature, (**Fig. 3a**), which was validated by northern blotting and LC-MS (**Fig. 3b and Extended Data Fig. 6a**). Proteomic analysis indicates TkNat10 expression is elevated at high temperatures (**Extended Data Fig. 6b,c and Supplementary Table 5b**), consistent with increased ac$^4$C. These temperature-dependent patterns of ac$^4$C in rRNA, tRNA, ncRNA and mRNA are described in further detail in **Figure 3c**, **Extended Data Fig. 6d-h**, and **Supplementary Note 3**. Strikingly, the ΔTkNat10 showed a temperature dependent growth lag in comparison to the WT strain, beginning at 75˚C and peaking at 95˚C (**Fig. 3d**). The reduced fitness of ΔTkNat10 strains at higher temperatures parallels the increased ac$^4$C of WT strains under these conditions, suggesting ac$^4$C is particularly required for high temperature growth. If cytidine acetylation is a response to thermal stress, we might expect closely-related organisms to also employ this mechanism. Indeed, induced acetylations at higher temperatures were also conserved in *P. furiosus* and *T. AM4*, two species closely related to *T. kodakarensis* (**Fig. 3e, Extended Data Fig. 6i**). Moreover, the precise sites and stoichiometries at which ac$^4$C was induced were also highly conserved in these organisms (**Extended Data Fig. 6j**). These studies suggest temperature-dependent cytidine acetylation as a unique adaptive survival strategy employed by the archaeal order Thermococcales.

## Profiling Ac$^4$C in an Archaeal Ribosome

The dynamics of ac$^4$C on the *T. kodakarensis* ribosome are unprecedented, with both the number of sites and their stoichiometry of modification increasing dramatically with temperature (**Fig. 3a**)**.** In comparison, characterized eukaryotic ribosomes have at most two ac$^4$C sites[18] while their bacterial counterparts have none[18–20]. To visualize the distribution of ac$^4$C in *T. kodakarensis* rRNA we obtained cryo-EM structures of ribosomes derived from WT and ΔTkNat10 strains with nominal resolutions of 2.95 Å and 2.65 Å, respectively (**Extended Data Figs 7-8**, **Supplementary Table 6**). This resolution allowed full delineation of the *T. kodakarensis* 70S ribosome architecture, including assignment of the three RNA constituents, associated core proteins, and visualization of modified nucleotides (**Fig. 4a-b**, **Extended Data Fig. 8b** and **Supplementary Tables 7,8**). Comparing ribosome structures of WT and ΔTkNat10 strains we found that the density associated with ac$^4$C was exclusively observed in WT ribosomes (**Fig. 4b-c** and **Extended Data Fig. 8b**). Cryo-EM maps directly supported the presence of 69 ac$^4$C sites in the *T. kodakarensis* ribosome grown at 85˚C (**Fig. 4a** and **Supplementary Table 4**). The ability to visualize these residues via cryo-EM was consistent with the high stoichiometry estimated at these sites on the basis of the ac$^4$C-seq measurements (**Supplementary Table 4; Extended Data Fig. 9a**).The unbiased nature of this analysis augmented ac$^4$C-seq by identifying six locations of the doubly modified nucleoside ac$^4$Cm (**Extended Data Fig. 9b-e**), previously hypothesized to play a role in thermostability[21,22]. To explore ac$^4$C dynamics using cryo-EM, we also determined the structure of ribosomes derived from WT *T. kodakarensis* grown at 65 ˚C (2.55 Å, **Extended Data Figs 7-8** and **Supplementary Tables 6-7**). Consistent with ac$^4$C-seq, the 65 ˚C structure exhibited dramatically lower ac$^4$C levels, with only five cytidine residues showing a clear density for acetylation (**Extended Data Fig. 8b** and **Supplementary Table 4**).

A remarkable feature of ac$^4$C in the *T. kodakarensis* ribosome is that acetylation appears spread across core and surface residues in both subunits (**Fig. 4a**). This contrasts starkly with rRNA base modifications in eukaryotes and bacteria, which are enriched at functional regions near the ribosome core (**Extended Data Fig. 9f**). Nonetheless, inspection of modification level as a function of temperature revealed a striking pattern of ac$^4$C in archaeal rRNA (**Fig. 4d**). The seven ac$^4$C residues detected at low temperatures (herein termed ‘core’ sites) were found to concentrate at the interface between the two ribosomal subunits making direct interactions with the ribosomal substrates (**Fig. 4e**). Six of these sites envelop an inter-subunit bridge comprised of the large-subunit (LSU) ribosomal protein eL41, whereas an additional site localized at the ribosome exit tunnel (**Fig. 4e**). Of note, the eukaryotic homologue of eL41 (RPL41) also localizes in an environment enriched in modified nucleosides[18]. ‘Core’ sites were acetylated at remarkably high levels across all temperatures (median of 77% misincorporation at 85˚C, **Extended Data Fig. 9g,h**), and also modified at high levels in *T. AM4* and *P. furiosus* (**Fig. 4f**), emphasizing a potential role in ribosome function. In contrast, ac$^4$C sites detected only at higher temperatures were modified at lower levels (median 18% at 85 ˚C) and distributed widely across the ribosome, suggesting a non-catalytic ‘auxiliary’ role (**Extended Data Fig. 9g,h**). Considering physical mechanisms impacted by ac$^4$C, we noted that in the vast majority of sites visualized by cryo-EM (64/70, 91%), the N4-acetyl group present in WT ribosomes is replaced by an ordered solvent molecule in the deletion strain (**Fig. 4b,c, Extended Data Fig. 8b** and **Supplementary**

**Table 4**). Similar replacement was observed in unmodified positions from the strain grown at 65 ˚C (**Extended Data Fig. 8b**). Ordered solvent molecules are often visualized in near atomic resolution structures and can contribute to the structural integrity of protein and RNA architecture, and it is tempting to speculate ac$^4$C may have evolved as a covalent installation to replace tightly bound solvent molecules that might otherwise undergo displacement at high temperatures. Concomitantly, we identified a small subset of positions in which cytidine acetylation created the potential for unique RNA-protein interactions. Representative examples are the interaction of O(7) of ac$^4$C1459, a core site located in helix 45 of the *T. kodakarensis* small subunit, with Arg15 of eL41 (**Fig. 4g**) and ac$^4$C1434 of LSU with OP2 of A1786 (**Extended Data Fig. 9i**). In these examples, the ordered solvent molecule bridges the interactions that are otherwise mediated by the acetyl group (**Fig. 4g**). Examining the potential influence of ac$^4$C on RNA-RNA interactions, we found the vast majority (68/69, 99%) of modified residues lie in duplexed rRNA and engage in canonical C-G base pairing. Consistent with the potential for acetylation to strengthen these interactions, biophysical analyses of a synthetic ribosomal RNA hairpin found replacement of C with ac$^4$C enhances thermal stability (**Fig. 4h**, **Extended Data Fig. 9j**)[23,24]. Overall, our structural survey highlights multiple avenues by which dynamic cytidine acetylation at higher temperatures may alter the catalytic properties and physical robustness of the archaeal ribosome.

## Conclusion

Here we describe ac$^4$C-seq, a method for the quantitative nucleotide resolution profiling of RNA cytidine acetylation. This method leverages acid-catalyzed reactivity enhancement to achieve an efficient chemical reduction of ac$^4$C, which was integrated with next-generation sequencing to enable transcriptome-wide detection of ac$^4$C in diverse organisms and RNA species. Applied to eukaryotes, our studies define rRNA and tRNA as the major physiological repositories of ac$^4$C, and suggest cytidine acetylation is absent or present at very low levels in endogenous eukaryotic mRNA. This diverges substantially from previous reports using antibody-based enrichment[10]. It remains to be established whether this discrepancy originates from technical differences in the methods (**Supplementary Note 2a**) or artifacts caused by antibody promiscuity, the latter of which has substantial precedent in the field (**Supplementary Note 2b**)[25–28].

Application of ac$^4$C-seq in archaea revealed pervasive programs of RNA acetylation. In the context of rRNA base modifications, ac$^4$C in Thermococcales is unprecedented in its prevalence and responsiveness to environmental cues. The dynamic and widespread distribution of ac$^4$C in the *T. kodakarensis* ribosome challenges our orthodox view of rRNA modifications, in which target sites of rRNA modifying enzymes are classically conceptualized as being *deterministic*, i.e. each RNA modifying enzyme catalyzes the modification of one or more highly specific sites. The high number and partial modification of ‘auxiliary’ sites in the *T. kodakarensis* ribosome instead raises the possibility that ac$^4$C catalysis at these positions may be *statistical*, i.e. each site harbors a predefined *probability* of being targeted by the acetyltransferase, and contributes in an *additive* manner to overall rRNA function. It remains to be addressed whether such deposition is primarily required for the function of *mature* ribosomes or to facilitate rRNA *folding and processing* under elevated temperatures. Our results further imply that such ‘statistical’

deposition of ac$^4$C is not limited to rRNA, but also widespread in other highly structured RNAs. Collectively, our studies define the ac$^4$C landscape across archaeal and eukaryotic lineages, providing a technical and conceptual foundation for unravelling this modification's role in biology and disease[4–6].

## References


1. Sharma, S. *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.* **43**, 2242–2258 (2015).

2. Ito, S. *et al.* A single acetylation of 18 S rRNA is essential for biogenesis of the small ribosomal subunit in Saccharomyces cerevisiae. *J. Biol. Chem.* **289**, 26201–26212 (2014).

3. Ito, S. *et al.* Human NAT10 Is an ATP-dependent RNA Acetyltransferase Responsible for N4-Acetylcytidine Formation in 18 S Ribosomal RNA (rRNA). *J. Biol. Chem.* **289**, 35724–35730 (2014).

4. Larrieu, D., Britton, S., Demir, M., Rodriguez, R. & Jackson, S. P. Chemical inhibition of NAT10 corrects defects of laminopathic cells. *Science* **344**, 527–532 (2014).

5. Tschida, B. R. *et al.* Insertional Mutagenesis in Mice Identifies Drivers of Steatosis-Associated Hepatic Tumors. *Cancer Res.* **77**, 6576–6588 (2017).

6. Zhang, H. *et al.* GSK-3 -Regulated N-Acetyltransferase 10 Is Involved in Colorectal Cancer Invasion. *Clinical Cancer Research* vol. 20 4717–4729 (2014).

7. Kotelawala, L., Grayhack, E. J. & Phizicky, E. M. Identification of yeast tRNA Um(44) 2'-O-methyltransferase (Trm44) and demonstration of a Trm44 role in sustaining levels of specific tRNA(Ser) species. *RNA* **14**, 158–169 (2008).

8. Dewe, J. M., Whipple, J. M., Chernyakov, I., Jaramillo, L. N. & Phizicky, E. M. The yeast rapid tRNA decay pathway competes with elongation factor 1A for substrate tRNAs and acts on tRNAs lacking one or more of several modifications. *RNA* **18**, 1886–1896 (2012).

9. Sharma, S. *et al.* Specialized box C/D snoRNPs act as antisense guides to target RNA base acetylation. *PLoS Genet.* **13**, e1006804 (2017).

10. Arango, D. *et al.* Acetylation of Cytidine in mRNA Promotes Translation Efficiency. *Cell* **175**, 1872–1886.e24 (2018).

11. Thomas, J. M. *et al.* A Chemical Signature for Cytidine Acetylation in RNA. *J. Am. Chem. Soc.* **140**, 12667–12670 (2018).

12. Sinclair, W. R. *et al.* Profiling Cytidine Acetylation with Specific Affinity and Reactivity. *ACS Chem. Biol.* **12**, 2922–2926 (2017).

13. Tardu, M., Jones, J. D., Kennedy, R. T., Lin, Q. & Koutmou, K. S. Identification and Quantification of Modified Nucleosides in Saccharomyces cerevisiae mRNAs. *ACS Chemical Biology* (2019) doi:10.1021/acschembio.9b00369.

14. Kowalak, J. A., Dalluge, J. J., McCloskey, J. A. & Stetter, K. O. The role of posttranscriptional modification in stabilization of transfer RNA from hyperthermophiles. *Biochemistry* **33**, 7869–7876 (1994).

15. Taoka, M. *et al.* Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.* (2018) doi:10.1093/nar/gky811.

16. Orita, I. *et al.* Random mutagenesis of a hyperthermophilic archaeon identified tRNA modifications associated with cellular hyperthermotolerance. *Nucleic Acids Res.* **47**, 1964–1976 (2019).

17. Yu, N. *et al.* tRNA Modification Profiles and Codon-Decoding Strategies in Methanocaldococcus jannaschii. *J. Bacteriol.* **201**, (2019).

18. Sharma, S. & Lafontaine, D. L. J. 'View From A Bridge': A New Perspective on Eukaryotic rRNA Base Modification. *Trends Biochem. Sci.* **40**, 560–575 (2015).

19. Fischer, N. *et al.* Structure of the E. coli ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. *Nature* **520**, 567–570 (2015).

20. Polikanov, Y. S., Melnikov, S. V., Söll, D. & Steitz, T. A. Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nat. Struct. Mol. Biol.* **22**, 342–344 (2015).

21. Kawai, G. *et al.* Conformational Rigidity of N4-Acetyl-2′-O-methylcytidine Found in tRNA of Extremely Thermophilic Archaebacteria (Archaea). *Nucleosides and Nucleotides* **11**, 759–771 (1992).

22. Bruenger, E. *et al.* 5S rRNA modification in the hyperthermophilic archaea Sulfolobus solfataricus and Pyrodictium occultum. *FASEB J.* **7**, 196–200 (1993).

23. Kumbhar, B. V., Kamble, A. D. & Sonawane, K. D. Conformational preferences of modified nucleoside N(4)-acetylcytidine, ac4C occur at 'wobble' 34th position in the anticodon loop of tRNA. *Cell Biochem. Biophys.* **66**, 797–816 (2013).

24. Parthasarathy, R., Ginell, S. L., De, N. C. & Chheda, G. B. Conformation of N4-acetylcytidine, a modified nucleoside of tRNA, and stereochemistry of codon-anticodon interaction. *Biochem. Biophys. Res. Commun.* **83**, 657–663 (1978).

25. Safra, M. *et al.* The m(1)A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* (2017) doi:10.1038/nature24456.

26. Li, X. *et al.* Base-Resolution Mapping Reveals Distinct m1A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. *Mol. Cell* **68**, 993–1005.e9 (2017).

27. Grozhik, A. V. *et al.* Antibody cross-reactivity accounts for widespread appearance of m1A in 5' UTRs. doi:10.1101/648345.

28. Helm, M., Lyko, F. & Motorin, Y. Limited antibody specificity compromises epitranscriptomic analyses. *Nat. Commun.* **10**, 5669 (2019).

29. Deutsch, E. W. *et al.* The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res.* **45**, D1100–D1106 (2017).

30. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

## Acknowledgments

The authors thank Dr. S. Fox (Laboratory of Proteomics and Analytical Technology) for LC-MS analyses, Dr. C. Grose (Protein Expression Laboratory) for assisting with cloning and preparation of plasmid DNA, Profs. P. Blum (University of Nebraska) and Prof. B. Mukhopadhyay (Virginia Tech) for archaeal RNA samples, and Dr. N. Elad (EM unit, Weizmann Institute of Science) for his assistance in setting up the cryo-EM data collection. We thank the

Biophysics Resource in the Structural Biophysics Laboratory, Center for Cancer Research, NCI for assistance with DSC and CD spectroscopy studies. The authors further thank Prof. E. Westhof for his insights into RNA structure.

## Author Contributions

**Conceptualization**: A.S-C., J.M.T., D.M., M.S-B., J.L.M. and S.S.; **Methodology development**: A.S-C. and J.M.T.; **Cryo-EM and ribosome purification**: D.M and M.S-B.; **Archaeal biology and genetics**: G.L.S.L, B.W.B. and T.J.S.; **Proteomics**: M.J.L., L.F., and M.P.W.; **LC-MS of partially digested ribosomes**: M.T., Y.N., and T.I.; **Eukaryotic overexpression analyses**: A.S-C. and K.D.N; **Systematic mutagenesis screen**: A.S-C, R.N, and S.S; **Computational analysis**: A.S-C, and S.S; **Biophysical studies**: J.M.T., K.D.N., and and S.T.G; **Validation experiments and follow-ups**: K.M.B., R.S., C.A.B., S.T.G., Q.L., R.T.F., G.B.R., J.H., S.Sh., and Q.L; **Writing**: A.S-C., M.S-B., J.L.M, and SS with input from J.M.T. D.M, and T.J.S. **Supervision and funding acquisition**: M.S-B., J.L.M and S.S.

## Data availability

ac$^4$C-seq datasets generated in this manuscript have been deposited in the GEO under accession number GSE135826. The cryo-EM density maps have been deposited in the Electron Microscopy Data Bank (EMDB) under accession codes EMD-10223 and EMD-10503 for the WT strain grown at 85 ˚C and 65 ˚C, respectively and EMD-10224 for the TkNat10 deletion strain. Model coordinates have been deposited in the Protein Data Bank (PDB) under accession numbers 6skf, 6th6 and 6skg. Mass spectrometry proteomics data has been deposited to the ProteomeXchange Consortium via Pride[29,30] partner repository with the dataset identifier PXD014814 and 10.6019/PXD014814.

**Competing interests** The authors declare no financial or non-financial competing interests.

## Figure legends

**Figure 1. Development and application of ac$^4$C-seq in human and yeast.** (**a**) Depiction of chemistries reducing or deacetylating ac$^4$C. (**b**) Ac$^4$C-seq: RNA is deacetylated (or mock) pre-treated followed by $NaCNBH_3$ (or mock) treatment. Following library-prep as illustrated, ac$^4$C is detected via C->T misincorporation. (**c**) Misincorporation rates in HeLa cell total RNA are shown for known sites in 18S (blue: cytidine, red: thymidine). (**d**) Misincorporation rates in 18S sites in WT and Nat10 depleted cells (Bars: mean of 3 biological samples, dots: individual measurements). (**e**) Ac$^4$C-seq based misincorporation levels of 4 synthetic spikes (Y axis)

plotted against $ac^4C$ levels as measured by MS (X axis), R-Pearson, n=1 experiment. (**f**) Statistical significance plotted against the difference in misincorporation rates between $NaCNBH_3$ and mock treated total RNA from HeLa cells. Dashed lines; vertical: 5%, horizontal: *p*=0.05 (Chi-squared test). n=3 biological samples. (**g**) Frequency of the 12 possible misincorporation patterns (Y axis) for sites found in Poly(A)-enriched RNA from WT and Nat10/Thumpd1 overexpressing (OE) HEK-293T cells. Pie chart: the proportion of sites harboring C->T misincorporations within a C**C**G motif. (**h**) Misincorporation rate at $ac^4C$ sites within CCG motifs identified in (g) in presence versus absence of Nat10/Thumpd1 overexpression, shown for RNA treated with $NaCNBH_3$ and indicated controls (n=2 biological samples for overexpression $NaCNBH_3$- or mock-treated and 1 samples for the rest). boxplot's center and boundaries: median, 25th and 75th percentiles. Whiskers: ±1.5xIQR. Outliers: individual dots. (**i**) $Ac^4C$-seq based misincorporation level at amplicons spanning $ac^4C$ sites in HEK-293T cells, depicted as in (d). n=2 biological samples. (**j-k**) Sequence motif surrounding the sites identified in indicated organisms. (**l**) Misincorporation rate at two WT and mutated $ac^4C$ sites in Nat10/Thumpd1 overexpressing HEK-293T cells, quantified via targeted $ac^4C$-sequencing, depicted as in (d). n=2 biological samples.

**Figure 2. $Ac^4C$ is present at unprecedented levels across diverse RNA species in archaea.** (**a**) Relative quantification of $ac^4C$ in total RNA isolated from *H. sapiens*, *S. cerevisiae*, *S. solfataricus* and *T. kodakarensis*. Mean of n = 3 technical replicates. *H. sapiens* total RNA was isolated from HeLa cells. (**b**) Misincorporation distribution (as in Fig. 1g) across all identified sites in *T. kodakarensis*. 99% of C->T misincorporation sites are embedded within a C**C**G motif. (**c**) Pearon's correlation between $ac^4C$ levels as measured by $ac^4C$-seq and LC-MS, shown for 25 sites that were quantified by both methodologies. n = 2 and 1 independent samples for LC-MS and ac4C-seq experiments, respectively. (**d**) $Ac^4C$-seq quantification of sites identified in WT and ΔTkNat10 strains. Boxplot visualization parameters are as in Fig. 1h. n = 4 and 2 independent biological samples for WT and ΔTkNat10, respectively. (**e**) The number of identified $ac^4C$ sites in the different RNA types as found in total RNA of different archaeal species. Note that for *T. kodakarensis* - but not for others - $ac^4C$-seq was applied also to rRNA-depleted RNA. ncRNAs reflect sites in RNaseP RNA, SRP RNA and snoRNAs, the last being present only in *P. furiosus*. The phylogenetic tree represents evolutionary distance between the species. (**f-g**) correlation between misincorporation level in ncRNA of *T. kodakarensis* compared to *P. furiosus* (f) and *T. AM4* (g) identified by $ac^4C$-seq. Pearson's correlation coefficient is shown. n = 4 and 1 independent biological samples for *T. kodakarensis* and other archaea, respectively. Shading indicates 95% confidence level interval for predictions from a linear model. WT: wild-type; KO: knock-out; LC-MS: Liquid chromatography–mass spectrometry; cryo-EM: Cryogenic electron microscopy.

**Figure 3. $Ac^4C$ accumulates in a temperature dependent manner across all RNA species in archaea and is required for growth at higher temperatures.** (**a**) Distributions of misincorporation level at $ac^4C$ sites across temperatures ranging from 55 to 95 ˚C. Boxplot visualization parameters are as in Fig. 1h. n = 4 biologically independent samples for 85 ˚C, n=2 for 65 and 75 ˚C and n=1 for 55 and 95 ˚C. (**b**) Anti-$ac^4C$ immuno-northern blot in *T. kodakarensis* total RNA as a function of temperature. Ethidium bromide staining is used to

visualize total RNA. Results are representative of two biological replicates. For gel source data, see Supplementary Data 3. (**c**) Schematic representation of a tRNA molecule. A total of 77 $ac^4C$ sites found within 19 tRNAs species (indicated by name of amino acids) were distributed across 13 distinct positions within the tRNA molecule. Each modified position is indicated by an orange circle. Numbers indicate position within the tRNA. Note that positions in the variable region are not numbered. (**d**) *T. kodakarensis* WT and ΔTkNat10 cells were grown across diverse temperatures (65-95 ˚C). OD600 was measured hourly. The average curves of replicates is shown (thick line, n=11 for 95 ˚C and n=12 for each of 65-85 ˚C, respectively) along with individual replicates (thin lines). (**e**) $Ac^4C$-seq quantification of total RNA collected from cells grown in a range of temperatures. Shown are misincorporation levels for $ac^4C$ sites identified in *P. furiosus* and *T. AM4*, respectively. Boxplot visualization parameters are as in Fig. 1h. n=1 biological sample per condition.

**Figure 4. Cryo-EM structure of WT and $ac^4C$-deficient *T. kodakarensis* ribosomes.** (**a**) $Ac^4C$ distribution as observed by cryo-EM of WT *T. kodakarensis* grown at 85˚C. Modified residues are highlighted orange, rRNA in grey and r-proteins are contoured in black. (**b**) $Ac^4Cs$ participate in Watson-Crick pairs with guanine residues. Example of $ac^4C$ in density is presented in (**b**). Residues correspond to $ac^4C2159$ and G2725 of LSU. Acetate is highlighted yellow and pointed by an arrow. The absence of acetate at the same position of the ΔTkNat10 strain is demostrated in (**c**) indicating that in the mutant, the acetyl moiety is replaced by structured solvent molecule. (**d**) $Ac^4C$ in *T. kodakarensis* ribosomes derived from archaea grown at different temperatures (identified by $ac^4C$-seq and LC-MS). (**e**) 'Core' $ac^4Cs$ (red) present in high stoichiometries across temperatures are enriched in the intersubunit interface and are in proximity to eL41 and to the ribosomal substrates. Functional ribosome regions indicated: decoding center (DC), the peptidyl-transferase center (PTC) and the protein exit tunnel (ET). tRNA and mRNA are highlighted yellow, eL41 - purple. The tRNA and mRNA coordinates are from PDB 4v5d. (**f**) Misincorporation level at core and auxiliary sites from *T. kodakarensis* and their conserved counterparts, *P. furiosus* and *T. AM4,* grown at optimal growth temperatures (85 ˚C for *T.kod* and *T.AM4* and 95 ˚C for *P.fur*) are shown (n = 4 and 1 independent biological samples for *T. kodakarensis* and other archaea, respectively). Boxplot visualization parameters are as in Fig. 1h. (**g**) Representative example of electrostatic interaction between $ac^4C$ and ribosomal proteins is shown between O(7) of $ac^4C1459$ at h45 of small-subunit (SSU) and R15 of eL41. The same position at the ΔTkNat10 strain (bottom) implicates a solvent molecule that serves to mediate the same interaction network in the absence of an acetyl group. (**h**) RNA thermal melting curves of synthetic RNA hairpin containing C (black) or $ac^4C$ (red) recorded with differential scanning calorimetry (DSC). Values represent the mean and standard deviation of n = 3 independent experiments.

## Extended Data Figures Legends

**Extended Data Figure 1. An Optimized Reaction for Sequencing of N4-Acetylcytidine in RNA.** (**a**) Protonation under acidic conditions hyperactivates $ac^4C$, increasing its reactivity with $NaCNBH_3$. Efficient reduction manifests as quantitative misincorporation of deoxynucleotide triphosphates at $ac^4C$ upon reverse transcription. (**b**) $NaCNBH_3$-dependent misincorporation at the known $ac^4C$ site in human helix 45 is increased at more acidic pH. Percent misincorporation at $ac^4C$ sites following chemical reduction, reverse transcription, and PCR was quantified from Sanger sequencing data. Representative of 1 independent experiment. (**c**) Kinetic analysis of $ac^4C$ reduction. Reaction progress was assessed by monitoring the disappearance of $ac^4C$ absorbance at 300 nm in the presence of first and second-generation hydride donors. Reaction conditions: $ac^4C$ (0.1 mM, free nucleoside), reductant (20 mM), $H_2O$. $NaBH_4$ reactions were carried out at pH 10, while $NaCNBH_3$ reactions were adjusted to pH 1 using HCl prior to initiation. Representative of 3 independent experiments. (**d**) Kinetic analysis of $ac^4C$ hydrolysis at pH values used in $NaBH_4$ (pH 10) and $NaCNBH_3$ (pH 1) reduction reactions. Reaction progress was assessed by monitoring the disappearance of $ac^4C$ absorbance at 300 nm. Acid- and base-catalyzed hydrolysis occurred at similar rates, and were slow compared to $ac^4C$ reduction by $NaBH_4$ and $NaCNBH_3$. Representative of 3 independent experiments. (**e**) LC-MS/MS analysis confirms reduction of $ac^4C$ to reduced $ac^4C$ in the presence of $NaCNBH_3$. Reaction conditions: $ac^4C$ (0.1 mM, free nucleoside), $NaCNBH_3$ (20 mM), HCl pH 1. Representative of 2 independent experiments. (**f**) Exact mass of reduced $ac^4C$ and deamination product observed in LC-MS/MS experiments. (**g**) Primer extension analysis of $ac^4C$-containing RNAs following $NaCNBH_3$ treatment (100 mM, pH 1, 37 °C, 1 h). (**h**) Sanger sequence traces of a known $ac^4C$ sites in helix 45 of human HAP1 cells. C->T misincorporation is exclusively observed at the $ac^4C$ site in reduced ($NaBH_4$ and $NaCNBH_3$) but not in mock treated samples. $ac^4C$ sites is highlighted in yellow.

**Extended Data Figure 2. $Ac^4C$ in eukaryotic cells with WT Nat10 expression.** (**a-c**) $ac^4C$-seq was conducted on RNA from *S. cerevisiae* (a,c) and HeLa cells (b). Statistical significance of Chi-squared test is plotted against the difference in misincorporation rates (corresponding to $ac^4C$ levels) between $NaCNBH_3$ and mock treatment RNA from *S. cerevisiae* (a), RNA from WT and Nat10 depleted cells (b) or from WT *S. cerevisiae* cells and a strain expressing a catalytic mutant of Kre33, treated with $NaCNBH_3$. Sites with a differential misincorporation level > 5% and a *p* value < 0.05 are labeled and marked in red. For HeLa cells (b) an additional comparison between $NaCNBH_3$ and deacetylation pre-treatment was conducted. Sites that do not pass significance under these conditions are marked with a plus sign (shown only for sites found significant between $NaCNBH_3$ and mock treatment). Significant sites are labelled with the identity of the molecule and the relative position (or helix) of $ac^4C$. (n=3 biologically independent samples for all but Nat10 depleted HeLa cells for which n= 2 biologically independent samples) (**d-e**) Misincorporation level in the two known sites in 18S (helix 34 and helix45), compared to controls in poly(A)-enriched RNA from WT and a catalytic mutant of Kre33 *S. cerevisiae* cells (d) and from WT and Nat10/Thumpd1-overexpressing HEK-293T (e). (**f-g**) ac4C-seq data from poly(A)-enriched RNA from HEK-293T cells overexpressing Nat10/Thumpd1 on 'ac4C peaks' identified by Arango et al[10] as harboring ac4C. (f) Distribution of % misincorporation across each of 57 'ac4C peaks' that had a coverage of >400 reads in >80% of the cytosines in the peak. For each peak the cytosine harboring the highest misincorporation rate is indicated in color,

presented in blue if it harbors a CCG motif and red otherwise. (g) IGV browser traces of three such genes, with highest coverage in the ac4C-seq data. For each gene the 15 bases motif identified by Arango et al is presented. Numbers above each cytidine indicate the number of bases (A, C, G and T) observed in our data at that position. (**h**) Power analysis for $ac^4C$ detection, as a function of sequencing depths and stoichiometries. Each datapoint in each curve is based on 1000 simulations. For each sampled depth, numbers in legend indicate the sequencing depth (which was kept identical for treatment and control samples). In addition, the legend indicates the number of CCG sites found in HEK-293T WT samples that have such a minimal depth and the percentage of these detectable CCG sites from all CCG sites in the transcriptome.

**Extended Data Figure 3. $Ac^4C$ in eukaryotic cells with manipulated Nat10 expression.** (**a**) RNA expression levels of Nat10 and Thumpd1 in HEK293T cells overexpressing both genes compared to WT cells. Shown are TMM normalized read counts. Numbers above bars indicate fold increase from WT. (**b**) Immunoblotting analysis of Nat10 and Thumpd1 overexpression in HEK-293T cells. Representative of 3 independent experiments with similar results. For gel source data, see Supplementary Data 3. (**c**) Microscopy images of eGFP-Nat10 construct confirming nuclear and nucleolar localization of ectopically expressed N-terminally tagged protein. Representative of 3 independent experiments with similar results. (**d**) RNA expression of Kre33 and Tan1 in WT yeast cells and in cells stably overexpressing Kre33 and either stably or inducibly overexpressing Tan1. Numbers above bars indicate fold increase from WT. (**e**) The number of sites displaying each of the 12 possible misincorporation patterns are displayed (barplot, Y axis) for sites found in Poly(A)-enriched RNA from both WT and Kre33/Tan1 overexpressing *S. cerevisiae* cells. The pie chart displays the proportion of sites harboring C->T misincorporations that were embedded within a C**C**G motif (73/74, 98.6%). (**f**) Schematic of the known $ac^4C$ sites in human tRNAs (Leu and Ser) and in helix 34 (C1337) and helix 45 (C1842) of human 18S rRNA. The acetylated cytidine residue (highlighted in blue) is embedded within a C**C**G motif in all known sites. (**g**) Fraction of $ac^4C$ sites found within the 5' UTR, CDS, and 3' UTR. Results are shown for the set of $ac^4C$ sites in mRNA of HEK-293T cells overexpressing Nat10/Thumpd1 (red bars, n=139), and - as controls - for all CCG motifs present within all genes within which any $ac^4C$ was found (blue bars, n=6,129). Error bars representing standard distribution of the binomial distribution. Data is based on 2 biologically independent samples. (**h**) Fraction of $ac^4C$ sites at the first, second and third position of each codon, shown for $ac^4C$ sites and controls as in (g). Error bars representing standard distribution of the binomial distribution. Data is based on 2 biologically independent samples. WT: wild-type; OE: overexpression; CDS: coding sequence; UTR: untranslated region.

**Extended Data Figure 4. Sequence and structure requirements for deposition of $ac^4C$.** (**a**) Oligonucleotides representing the WT sequence surrounding the acetylated site in BAZ2A mRNA, or variants with single mutations across the WT sequence, were synthesized as a pool and cloned into the 3'UTR of a reporter gene. The pool of plasmids was transfected into WT HEK-293T cells or cells transiently overexpressing Nat10 and Thumpd1. RNA extracted from cells was subjected to targeted ac4C-seq and ac4C levels were estimated on the basis of misincorporation rates. (**b**) Misincorporation rate of oligonucleotides described in (a), harboring the WT sequence of BAZ2A (green triangles) or a sequence mutated at the CCG motif and at its

surrounding bases (red and black, respectively). Center and boundaries of boxplots display median, 25th and 75th percentiles. Whiskers extend to ±1.5xIQR. Outliers are shown as individual dots. n=2 biologically independent samples (**c**) The difference in misincorporation rate of oligonucleotides with a single base mutation compared to the WT oligonucleotide is shown across all positions of the construct. (**d**) De novo construction of the motifs surrounding the modified cytidine were built based on the contribution of single-base mutations in the BAZ2A sequence to the reduction in misincorporation rate compared to WT BAZ2A sequence. (**e**) Secondary structure of the BAZ2A mRNA fragment as predicted by RNAfold. Bases are color-coded according to confidence level of the prediction. Regions highlighted by a blue and green line in panels **c-e** represent the CCG motif and a stem structure surrounding the modified cytidine, respectively.

**Extended Data Figure 5. Deletion of TkNat10 and TkThumpd1 in T. kodakarensis.** (**a**) Total RNA from *T. kodakarensis* was analyzed via ac$^4$C-seq. IGV browser traces display representative ac$^4$C sites in rRNA, ncRNA, mRNA and tRNA of *T. kodakarensis*, visualized as in Fig. 1c. Number in parentheses indicates the number of sites identified for each class of molecules. (**b**) Conserved domain architecture of human Nat10 and its homolog in *T. kodakarensis* – *TK0754* (referred to as TkNat10 in the text). (**c**) Expression of TkNat10 and TkThumpd1 (aka *TK2097*) in WT *T. kodakarensis* and the indicated deletion strains was quantified from ac$^4$C-seq data. Shown are mean TMM normalized values (n = 3 and 2 biological replicates in WT and deletion strains, respectively). (**d**) Quantitative LC-MS/MS proteomics analysis of *T. kodakarensis* in WT and ΔTK0754. Fold-change in protein abundance was based on comparison of distributed normalized spectral abundance factor (dNSAF) for individual proteins. Fold-change for proteins detectable exclusively in the WT or KO condition (fold-change = ∞) are graphed at 5.5 and 0.1, respectively, which represents the maximum and minimum of measured values. n=3 LC-MS/MS runs for each condition. (**e**) Anti-ac4C immuno-northern blot in- *T. kodakarensis* total RNA. Ethidium bromide staining is used to visualize total RNA. Results are representative of two biological replicates. For gel source data, see Supplementary Data 3. (**f**) Relative quantification of ac$^4$C in total RNA isolated WT and Δ*TK0754* (ΔTkNat10) *T. kodakarensis* strains as measured by LC-MS. Mean of n = 3 technical replicate. n.d. = not detectable. (**g**) Scatter-plot depicting misincorporation rate of ac$^4$C sites in WT *T. kodakarensis* is compared to TkThumpd1 deletion strain, showing no effect of the gene's deletion on ac$^4$C status. (**h**) Correlation between misincorporation rates in *T. kodakarensis* compared to *P. furiosus* and *T. AM4* for the different types of ncRNAs identified by ac$^4$C-seq. The Pearson's correlation coefficient is indicated at the bottom of each plot. n = 4 and 1 independent biological samples for *T. kodakarensis* and other archaea, respectively. Shading indicates 95% confidence level interval for predictions from a linear model.

**Extended Data Figure 6. Ac$^4$C accumulates in a temperature dependent manner across all RNA species in archaea.** (**a**) Relative quantification of ac$^4$C in total RNA isolated from WT *T. kodakarensis* as a function of temperature as measured by LC-MS. Mean is shown along with individual data points. n = 3 technical replicates. For 65, 75, 85 ˚C: representative of 2 independent experiments, for 95 ˚C: representative of 1 independent experiment. (**b-c**) Quantitative LC-MS/MS proteomics analysis of *T. kodakarensis* temperature-dependent protein

expression. Fold-change in *T. kodakarensis* protein abundance between 85 and 65 ˚C growth conditions was based on comparison of distributed normalized spectral abundance factor (dNSAF) for individual proteins. Fold-change for proteins detectable exclusively in the 85 or 65 condition (fold-change = ∞) were set at 7.8 and 0.1, respectively, which represents the maximum and minimum of measured values. n=3 LC-MS/MS runs for each condition. Student's t-test, paired, two tailed p=0.012. (**d**) Misincorporation rates of $ac^4C$ sites at distinct regions of *T. kodakarensis* tRNAs as a function of growth temperature (55-95 ˚C), segregated into distinct regions within the tRNA molecule. Only sites with a minimal stoichiometry of 5% in any sample are shown. Center and boundaries of boxplots display median, 25th and 75th percentiles. Whiskers extend to ±1.5xIQR. Outliers are shown as individual dots. n = 4 biologically independent samples for 85˚C, n=2 for 65 and 75 ˚C and n=1 for 55 and 95 ˚C. (**e**) Multiple alignment of 37 tRNA molecules, representing 19 distinct tRNAs in *T. kodakarensis*, plotted across three distinct temperatures. $ac^4C$ sites are indicated in color-code representing misincorporation rate (right color bar). Red-orange bar on the left segregates the aligned sequences into distinct tRNA molecules, identified by the single-letter abbreviated of their amino acid. Selected regions from the multiple alignment, where $ac^4C$ is particularly abundant, are shown and color-coded according to the bottom color bar. (**f**) Schematic representation of RNaseP RNA in *T. kodakarensis*. $ac^4C$ sites (all in C**C**G) are marked with circles color-coded by misincorporation rate measured in cells growing at 85 ˚C. Fine grey lines indicate regions which base pair in the folded structure of the molecule, according to the model in[31] (**g-h**) Distribution of 119 acetylated cytidine residues (in 86 mRNAs) in *T. kodakarensis* across different codons (g) and at specific position within codons (h) are shown, and compared to that of 2,245 control non-acetylated Cs, found at CCG motifs of the same mRNAs. Y axis presents the fraction of cytidines in each position. n=1 set of sites (comprised of 119 ac4Cs and 2,245 Cs) with error bars representing standard distribution of the binomial distribution. (**i**) Anti-$ac^4C$ immuno-northern blot in *P. furiosus* and *T. AM4* total RNA as a function of temperature. Ethidium bromide staining is used to visualize total RNA. Results are representative of two biological replicates. For gel source data, see Supplementary Data 3. (**j**) A heatmap showing misincorporation rates at conserved $ac^4C$ sites in 5S, 16S, 23S, RNase P RNA and SRP RNA of *T. kodakarensis, P. furiosus and T. sp. AM4* grown in various temperatures. Rows are ordered according to misincorporation rates quantified in *T. kodakarensis* grown at 95 ˚C. Arrowhead indicates the conserved $ac^4C$ site at helix45 (top site in heatmap).

**Extended Data Figure 7. Cryo-EM data processing and map reconstruction.** (**a**) Schematic representation of EM data processing for the *T. kodakarensis* ribosomes. Data processing has been performed in Relion 3 and included motion correction, CTF correction, particle picking and classification. Initial map reconstruction and post processing was performed by the 3D refinement algorithm implemented in relion on the complete 70S particle, indicating high residual mobility of the SSU head domain (top left panel, grey). Further implementation of multibody refinement with individual masks prepared for the LSU (blue), SSU body (green) and SSU head (orange) resulted in the complete reconstruction of the 70S particle. Final map consisting of all three ribosomal domains for the WT ribosome derived from cells grown to 85 ˚C is presented in the down left corner. FSC curves indicating overall (black) and per domain (color coded according to relevant masks) resolutions are presented in (**b**) for the WT strain grown at

85 ˚C (WT85), (**c**) for the WT strain grown at 65 ˚C (WT65) and (**d**) for the ΔTkNat10 ribosomes (mutant). FSC comparisons of full (gray) and half-maps (pink/cyan) to the final refined model are presented in (**e**), (**f**) and (**g**) for WT85, WT65 and mutant strains, respectively. The excellent agreement between cyan and pink curves indicates lack of overfitting.

**Extended Data Figure 8. Cryo-EM data quality and ac$^4$C visualization.** (**a**) Surface (top) and cross-section (bottom) representations of the cryo-EM density maps colored according to local resolution distribution. Growth conditions and *T. kodakarensis* strains used in the study along with PDB and EMDB accession codes are indicated. Resolution values are color coded according to the right index and are presented in Å. (**b**) Model in EM density for multiple ac$^4$C positions in WT *T. kodakaresis* grown at 85 ˚C (orange) and 65 ˚C (yellow) compared to an ac$^4$C ΔTkNat10 strain (mutant, blue) indicating the absence of acetate density (highlighted in light orange) in the mutant and in multiple positions of the strain grown at 65 ˚C. Positions highlighted with an asterisk are also acetylated in the 65 ˚C strain whereas positions that are unmarked are only acetylated in the archaea grown at 85 ˚C. These data are in good agreement with both the genomic-seq and MS approaches described in this manuscript, that similarly indicate that ac$^4$C distribution is highly dependent of growth temperature. A detailed list of ac$^4$C distribution and comparison with other methods is supplemented to the manuscript (**Supplementary Table 4**). 2D map with ac$^4$C distribution is in **Extended Data Figure 9d**.

**Extended Data Figure 9. RNA modifications of *T. kodakarensis* Ribosome and Thermostability.** (**a**) Misincorporation level as quantified by ac$^4$C-seq across all ac$^4$C sites identified in ribosomes of *T. kodakarensis* at 85 ˚C. Blue and red bars indicate sites which were and were not detected by cryo-EM, respectively. Dashed lines indicate median misincorporation of cryo-EM detected (upper, 13.7%) and not-detected (lower, 3.2%) sites. Acetylation detected by ac$^4$C-seq and also observed in the cryo-EM were generally of medium to high stoichiometry while the majority of acetylation sites detected by ac$^4$C-seq but not observed in the cryo-EM map density were of relatively low stoichiometry, rendering them invisible in the ensemble cryo-EM structure, which averages thousands of individual particles for map reconstruction. (**b**-**e**) Combined cryo-EM-MS analysis indicated six ac$^4$C residues that are also methylated at their 2'-O. Relative quantification of ac$^4$C and ac$^4$Cm detection in *T. kodakarensis* RNA via LC-MS is presented in (**b**). Mean and individual data points are shown. n = 3 technical replicates. An example of ac$^4$Cm in density is shown in (**c**) with acetate and methyl installations indicated by black arrows. A list of ac$^4$Cms is indicated in **Supplementary Table 4**. 2D (**d**) and 3D (**e**) visualization of ac$^4$C and ac$^4$Cm distribution in the *T. kodakarensis* ribosome with ac$^4$C highlighted orange and ac$^4$Cm green. Data is presented for the *T. kodakarensis* grown at optimal growth temperature (85 ˚C). Ac$^4$C positions highlighted in orange include genomic, MS and EM data. Ac$^4$Cm positions are a combination between cryo-EM and MS data. In (**e**) RNA and proteins are presented as grey ribbons, modified residues are highlighted as spheres. Protein exit tunnel (ET) is highlighted with a dashed black line, tRNA in yellow. The tRNA and mRNA coordinates are from PDB 4v5d. (**f**) A comparative view of RNA modification distribution in *E. coli*, yeast (*S. cerevisiae*), human (*H. sapiens*) and *T. kodakarensis*. Base modifications are colored blue, ac$^4$Cs in red, tRNA and mRNAs in yellow. Ribosome functional regions are designated in black with decoding center (DC), the peptidyl transferase center (PTC) and the protein exit tunnel (ET) highlighted in a dashed black line. PDB codes for the structures used for

comparison are 5AFI, 4V88, and 4UGO, for the *E. coli*, *S. cerevisiae* and human ribosome, respectively. (**g**) Misincorporation rate as quantified by ac$^4$C-seq for all ac$^4$C sites in the *T. kodakarensis* ribosome. Bar color indicates the lowest growth temperature at which the site was detected. (**h**) 3D representation of the *T. kodakarensis* ribosome with ac$^4$C sites detected at 55 ˚C and 85 ˚C shown and color-coded according to misincorporation rate in each temperature. (**i**) Ac$^4$Cs were shown to stabilize the *T. kodakarensis* ribosome via direct interactions with protein and RNA residues. An example of stabilization through RNA:protein interactions is presented in **Fig. 4g**. RNA:RNA interactions are presented in (**i**) and corresponds to ac$^4$C1434 interactions with OP2 of A1786 of LSU. (**j**) Temperature-dependent circular dichroism (CD) of synthetic RNAs containing cytidine (blue) or ac$^4$C (red). Thick and thin lines represent mean and individual measurements, respectively. n=3 independent experiments.

## Materials and methods

### Human cell culture

HeLa WT (ATCC) and Nat10-depleted cells[10] were maintained in Dulbecco's Modified Eagle's Medium (DMEM, Quality Biological, 112-013-101) supplemented with 10% fetal bovine serum (FBS, VWR, 89510-194), 25 mM *D*-glucose, 2 mM *L*-glutamine, and 1 mM sodium pyruvate. HEK293T cells (ATCC) were maintained in Dulbecco's Modified Eagle's Medium (DMEM, Quality Biological, 112-013-101) supplemented with 10% fetal bovine serum (FBS), 25 mM D-glucose, and 2 mM *L*-glutamine. All cells were maintained at 37 □C in the presence of 5% $CO_2$, and all cell culture reagents were purchased from Invitrogen Corp. (Carlsbad, CA) unless otherwise noted. Cells were found free of mycopalsma contamination and did not undergo authentication.

### Microbial growth and media conditions

*T. kodakarensis* strains - TS559 and their derivatives thereof - were grown as previously described[32–34] in artificial seawater (ASW) medium supplemented with vitamins and trace minerals. ASW contains (per L) 20 g NaCl, 3 g $MgCl_2{\cdot}6H_2O$, 6 g $MgSO_4{\cdot}7H_2O$, 1 g $(NH_4)_2SO_4$, 200 mg $NaHCO_3$, 300 mg $CaCl_2{\cdot}2H_2O$, 0.5 g KCl, 420 mg $KH_2PO_4$, 50 mg NaBr, 20 mg $SrCl_2{\cdot}6H_2O$, and 10 mg $Fe(NH_4)_2(SO_4)2{\cdot}6H_2O$. The trace mineral solution (1,000x per L) contains 0.5 g $MnSO_4{\cdot}H_2O$, 0.1 g $CoCl_2{\cdot}6H_2O$, 0.1 g $ZnSO_4{\cdot}7H_2O$, 0.01 g $CuSO_4{\cdot}5H_2O$, 0.01 g $AlK(SO_4)_2{\cdot}12H_2O$, 0.01 g $H_3BO_3$ and 0.01 g $Na_2MoO_4{\cdot}2H_2O$. The vitamin mixture (200x per L) contains 0.2 g niacin, 0.08 g biotin, 0.2 g pantothenate, 0.2 g lipoic acid, 0.08 g folic acid, 0.2 g p-amionbenzoic acid, 0.2 g thiamine, 0.2 g riboflavin, 0.2 g pyridoxine and 0.2 g cobalamin. 5 g/L yeast extract (Y), 5 g/L tryptone (T), 5 g/L pyruvate (Pyr) and 2 g/L elemental sulfur (S˚) were added to rich medium (ASW-YT-Pyr-S˚). ASW-S˚ mixture supplemented with a combination of 20 amino acids formed minimal medium (ASW-aa-S˚). The amino acid mixture contains (20x per L) 5 g cysteine, 5 g glutamic acid, 5 g glycine, 2.5 g arginine, 2.5 g proline, 2 g

asparagine, 2 g histidine, 2 g isoleucine, 2 g leucine, 2 g lysine, 2 g threonine, 2 g tyrosine, 1.5 g alanine, 1.5 g methionine, 1.5 g phenylalanine, 1.5 g serine, 1.5 g tryptophan, 1 g aspartic acid, 1 g glutamine and 1 g valine.

All *T. kodakarensis* cultures were grown at 55˚-95 ˚C under strict anaerobic conditions in sealed vessels with a headspace gas composition of 95% $N_2$/5% $H_2$ at 1 atmosphere at 22 ˚C; 1 mM agmatine was provided when necessary. Solid medium was prepared by the addition of 1% gelzan, with polysulfides substituting for S˚[34,35]. Polysulfides were prepared (500x, per 15 mL) by dissolving 10 g $Na_2S$•$9H_2O$ and 3 g S˚ with heat to a deep red mixture. Colonies formed on solid medium were observed by lifting cells to PVDF membranes that were then flash frozen in liquid $N_2$ before being thawed and stained with coomassie brilliant blue.

*Pyrococcus furiosus* strain COM1 was cultured at 75-95 ˚C in an artificial seawater based medium supplemented with cellobiose, maltose, yeast extract, S˚, trace minerals, cysteine and sodium tungstate as previously described[36]. *Thermococcus sp. AM4*[37] was cultured under identical conditions to that for *T. kodakarensis*.

**Yeast growth and media conditions**

*Saccharomyces cerevisiae* strains were grown at 30 ˚C in standard YEP medium (1% yeast extract, 2% Bacto Peptone) supplemented with 2% dextrose (YPD). For induction of *Tan1* by galactose, cells were washed twice with water, resuspended in YEP medium (1% yeast extract, 2% Bacto Peptone) supplemented with 2% galactose (YPG) and grown at 30 ˚C for 21 hr prior to harvesting.

**Construction of *T. kodakarensis* strains markerlessly-deleted for *TK0754* or *TK2097***

Plasmids used to direct the markerless-deletion of genomic sequences from the parental strain TS559 were each individually constructed from the parental plasmid **pTS700**[33] and contain ~700 bp sequences complementary to both upstream and downstream regions of the respective locus under study[32]. Each vector also encodes expression cassettes for *TK0149* (provides agmatine autotrophy) and TK0664 (provides sensitivity to 6-methylpurine). Strains were constructed as previously described[32,33,38]. In brief, plasmids incapable of autonomous replication in *T. kodakarensis* were individually transformed into *T. kodakarensis* TS559 (ΔTK0149; ΔTK0664; ΔTK0254::TK2276; ΔTK2276)[32,33,35,38]. Plasmid integration at the desired locus was confirmed by several diagnostic PCR amplicons generated from genomic DNA purified from intermediate strains. Overnight growth in the presence of 1 mM agmatine permitted spontaneous plasmid excision, and colonies were selected on solid media containing 20 amino acids, 6-methylpurine and agmatine. DNA was extracted from 1 mL ASW-YT-Pyr-S˚-agmatine cultures grown from individual 6-MP resistant colonies for use in diagnostic PCRs to confirm the deletion of the desired locus. Final confirmation of each strain included whole genome sequencing[32] to confirm deletion endpoints and to ensure no unanticipated modifications were introduced into the genome at remote locations.

**Plasmids for Nat10, Tan1, and Thumpd1 overexpression**

Tan-1 was synthesized and cloned into pD1201 and pD1231 by ATUM.

The remaining plasmids were constructed using Gateway recombination cloning (ThermoFisher) as follows:

Nat10 was amplified from a cDNA plasmid (Dharmacon (accession# BC035558)) by PCR and cloned into **pDonr-255** with BP Clonase. The insert was sequence verified and subcloned with LR Clonase into a neomycin-resistant mammalian transfection backbone with CMV promoter and N-terminal 3xFLAG-eGFP fusion. The same strategy, NAT10 entry clone, and expression vector were used to generate 3xFLAG-Nat10.

ThumpD1 was amplified from a cDNA plasmid (Dharmacon (accession# BC000448)) by PCR and the entry clone was generated and verified in a similar fashion. This entry clone was then subcloned with LR Clonase into a neomycin-resistant mammalian transfection backbone with CMV promoter, and N-terminal myc tag.

Transfection-quality plasmid DNAs were prepared using ZymoPURE II Plasmid Maxiprep Kit (Zymo Research)

**Overexpression of eGFP-NAT10 in HEK-293T cells**

HEK-293T cells were plated in a 10 cm dish (2.5 x $10^6$ cells/dish in 10mL DMEM medium) and allowed to adhere and grow for 24 hr. eGFP-tagged NAT10 was overexpressed using FuGENE® 6 transfection reagent (Promega #E2691). Prior to transfection, 600 µL of OPTI-MEM (Gibco, #31985062) was incubated with 18 µL FuGENE® 6 for 5 min at room temperature before adding 6 µg of eGFP-NAT10 plasmid and incubating an additional 30 min. Transfection mixture was carefully added to cell monolayer without changing medium. Overexpression was carried out by incubating the cells for 24 hr at 37 °C under 5% $CO_2$ atmosphere, after which time cells were imaged using an EVOS FL fluorescence microscope at 10X and 40X magnifications.

**Co-overexpression of NAT10/THUMPD1 in HEK-293T cells**

HEK-293T cells were seeded into 20-10 cm dishes (2.5 x $10^6$ cells/dish in 10 mL DMEM medium) and allowed to adhere and grow for 24 hr. 3xFLAG-tagged NAT10 and myc-tagged THUMPD1 were overexpressed using FuGENE® 6 transfection reagent (Promega, #E2691). For each 10 cm dish, 600 µL of Opti-MEM™ I Reduced Serum Medium (Gibco, #31985062) was incubated with 18 µL FuGENE® 6 for 5 min at room temperature before adding 3 µg each of Nat10 and Thumpd1 plasmid and incubating an additional 30 min. Transfection mixtures were carefully added to cell monolayer without changing medium. Overexpression was carried out by incubating the cells for 24 hr at 37 ˚C under 5% $CO_2$ atmosphere, after which 19 plates were harvested by trypsinization and snap-frozen for total RNA extraction. The remaining plate was harvested using ice-cold PBS and pelleted for western blot analysis of overexpression. Cell pellet was resuspended in 500 µL of ice-cold PBS containing protease inhibitor cocktail (1X, EDTA-free, Cell Signaling Technology #5871S). Samples were then lysed by sonication using a

100 W QSonica XL2000 sonicator (3 × 1 s pulse, amplitude 1, 60 s resting on ice between pulses). Lysate was pelleted by centrifugation (20,817 r.c.f. x 30 min, 4 ˚C) and quantified using the Qubit 4.0 Fluorometer and Qubit Protein Assay Kit. Protein was run on SDS-PAGE alongside non-transfected control and immunoblotted with anti-FLAG-tag (Cell Signaling, #2044), anti-NAT10 (Bethyl Laboratories, #A304-385A), and anti-myc-tag (Cell Signaling, #5605) antibodies. For immunoblotting, SDS–PAGE gels were transferred to nitrocellulose membranes (Novex, Life Technologies # LC2001) by electroblotting at 30 V for 1 hr using a XCell II Blot Module (Novex). Membranes were blocked using StartingBlock (PBS) Blocking Buffer (Thermo Scientific) for 30 min and incubated overnight at 4 °C in primary antibody. The membranes were washed with TBST buffer and incubated with secondary HRP-conjugated antibody (Cell Signaling #7074) for 1h at room temperature. The membranes were again washed with TBST and treated with chemiluminescence reagents (Western Blot Detection System, Cell Signaling) for 1 min, and imaged for chemiluminescent signal using an ImageQuant Las4010 Digital Imaging System (GE Healthcare).

For targeted ac$^4$C-sequencing in cells overexpressing either Nat10, Thumpd1, neither, or both, HEK-293T cells were seeded in replicates in wells of a 6-well plate (0.5 x $10^6$ cells/well in 2 mL DMEM media) and allowed to adhere and grow for 24 hr. Cells were transfected using PolyJet (SignaGen Laboratories) according to the manufacturer's protocol, either with 0.5 µg NAT10, or with 0.5 µg Thumpd1, neither, or both. In all samples a total of 50 ng GFP plasmid was used to monitor transfection efficiency. Cells were grown for 24 hr before harvesting for RNA purification.

**Growth analysis of *T. kodakarensis***

Parental strain TS559 and TkNat10-deleted *T. kodakarensis* cells were grown as described above at 65-95 ˚C (11-12 replicates from each temperature). Growth of liquid cultures was monitored by measurements of optical density at 600 nm at hourly intervals for a total of 33 hr. Measurements were used to model cell growth using the 'locally estimated scatterplot smoothing' (loess) method[39].

**Total RNA isolation from yeast, human and archaea**

Total RNA from human cells was extracted using TRIZOL according to manufacturer's protocol. 1 mL TRIZOL was used for each 1x$10^7$ cells. RNA pellet was resuspended by briefly heating at 50 ˚C in 1.0 mL 1X TE buffer pH 8.0. Samples were quantified by UV absorbance and stored at -80 ˚C. Typical extractions were carried out with 4x$10^7$ cells and yielded 400 µg of total RNA.

For targeted ac$^4$C-sequencing, RNA was extracted using Nucleozol (Macherey Nagel) according to the manufacturer's instructions.

Total RNA was isolated from yeast using hot acidic phenol. Briefly, frozen yeast (S. cerevisiae) pellet suspended in 1.0 mL AES buffer (50 mM NaOAc, 10 mM EDTA pH 8.0, 1% SDS) per 0.5 mL pellet volume. To suspended pellet, 1.0 mL acid-buffered phenol per mL of AES buffer used was added. Sample was mixed by vortexing and incubated in a 65 ˚C water bath for 30 min,

vortexing every 2 min to mix. Samples were put on ice for 10 min and 1.0 mL chloroform:isoamyl alcohol (24:1) was added for each 1.0 mL phenol used. Sample was vortexed to mix and centrifuged 5000 rcf for 15 min. Aqueous layer (top) was transferred to a clean tube and extracted 3X with an equal volume of acid buffered phenol:chloroform:isoamyl alcohol (24:23:1). After each extraction sample was centrifuged at 5000 rcf for 10 min and aqueous layer transferred to a new tube. A final extraction with chloroform:isoamyl alcohol was carried out to remove residual phenol. Aqueous layer was transferred to a clean tube and RNA was precipitated by the addition of an equal volume of 100% isopropanol and 1/9th volume of 3 M sodium acetate. Samples were incubated -20 ˚C for 30 min and centrifuged 12,000 rcf at 4 ˚C for 15 min. Supernatant was decanted and the pellet washed with 4 mL ice-cold 70% ethanol. RNA pellet was resuspended by briefly heating at 50 ˚C in 1.0 mL 1X TE buffer pH 8.0. Samples were quantified by UV absorbance and stored at -80 ˚C. Typical extractions were carried out with 1.0 mL volume cell pellets and yielded 20 mg of total RNA. Total RNA was isolated from archaeal samples using TRIZOL according to manufacturer's protocol.

**Poly(A)-RNA isolation from yeast and human cells**

Poly(A) RNA from yeast and human total RNA was isolated by two rounds of purification using the GenElute mRNA miniprep kit (Sigma) according to manufacturer's protocol. 500 µg total RNA was used per purification column. Typical yield after two rounds of isolation was 1.2%.

For targeted ac$^4$C-sequencing, poly(A) RNA was isolated from total HEK-293T cells RNA by two rounds of purification using Dynabeads mRNA DIRECT Kit (Invitrogen), according to the manufacturer's protocol. 75 µg total RNA was taken from each sample, using 150 µL oligo dT beads. Typical yield after two rounds of isolation was 1.6%.

**Ribosome purification**

Purifications of *T. kodakarensis* ribosomes of the WT and TkNat10 deletion were conducted similar to previously documented procedures[40]. In brief, cell lysis was obtained through sonication in Buffer A (20 mM HEPES, pH 7.5, 10.5 mM $Mg(OAc)_2$, 100 mM $NH_4OAc$, 0.5 mM EDTA and 6 mM β-mercaptoethanol). Cell debris was discarded by centrifugation at 30,000 xg for 20 min at 4 ˚C, and the cytoplasmic fraction was loaded onto a 1.1 M sucrose cushion in buffer B (20 mM HEPES, pH 7.5, 10 mM $Mg(OAc)_2$, 150 mM KOAc 6 mM β-mercaptoethanol). Ribosome enriched pellet was achieved through overnight centrifugation at 220,000xg, 4℃. The pellet was resuspended in buffer B and ribosome particles were purified on a 10-40% sucrose gradient using a SW-28 rotor, at 43,000 xg for 17 hr at 4℃. Fractions containing 70S ribosomes were collected, combined and centrifuged at 230,000 xg overnight at 4℃. The pellet was resuspended in buffer B and an additional centrifugation step at 200,000 xg for 1.5 hr at 4℃, was designed to remove sucrose traces. The ribosomal pellet was resuspended in buffer C (20 mM HEPES pH 7.5, 10 mM $Mg(OAc)_2$, 100 mM KOAc, 100 mM $NH_4OAc$ and 1 mM DTT), diluted to a concentration of 1 mg $mL^{-1}$, aliquoted and stored at -80 ˚C until further use.

**rRNA depletion from total RNA of *T. kodakarensis***

To deplete abundant *T. kodakarensis* rRNAs prior to RNA-seq, we adapted a method originally reported by[41] using reagents provided in the NEBNext® rRNA Depletion Kit (NEB #E6310). The protocol in the manual for the kit was followed with the following changes. The NEBNext rRNA Depletion Solution provided in the kit was substituted for an equimolar mixture of 85 oligonucleotides complementary to *T. kodakarensis* rRNA sequences **(Supplementary Table 1b)**. The concentration of the oligo mix was 85 µM, such that each individual oligo was 1 µM in the mix. All volumes for the probe hybridization, RNase H treatment and DNase I treatment sections of the protocol were scaled up 2-fold and 24 µl of 62.5 ng/µl *T. kodakarensis* RNA was used as the starting material. Instead of bead purification as indicated in the manual, samples were purified using the Monarch® RNA Cleanup Kit (NEB# T2030) using the standard protocol. Sixteen depletion reactions were performed as described above for each *T. kodakarensis* total RNA sample and these were then concentrated into a single depleted RNA sample by pooling them and performing a second round of purification with the Monarch® RNA Cleanup Kit. Yield of RNA after depletion was measured using the Qubit™ RNA BR Assay Kit (Thermo Fisher).

**UV spectroscopic analysis of ac$^4$C reduction rates**

Model reactions to assess the rate of $NaBH_4$ and $NaCNBH_3$ reduction of ac$^4$C were performed using free N4-acetylcytidine nucleoside. For $NaBH_4$ reductions, stock solutions of $NaBH_4$ (100 mM) and N4-acetylcytidine (2 mM) were prepared fresh daily in water. Reactions (25 µL) consisted of N4-acetylcytidine (100 µM), $NaBH_4$ (20 mM) and reaction buffer (water, 100 mM sodium acetate [pH 4.5], or 100 mM potassium phosphate [pH 7.5]). At the indicated time point, reactions were adjusted to 50 µL using 100 mM HCl. To normalize pH, a further aliquot of 50 µL 100 mM sodium phosphate [pH 7.2] was added and reactions were transferred to Greiner-UV Star 96-well half-area microplates (#655801) for analysis. For $NaCNBH_3$ reductions, stock solutions of $NaCNBH_3$ (1 M) and N4-acetylcytidine (2.5 mM) were prepared fresh daily in water. Reactions (100 µL) consisted of N4-acetylcytidine (100 µM), $NaCNBH_3$ (100 mM), and HCl (100 mM). At the indicated time point, reactions were quenched with 30 µL of 1 M Tris-HCl [pH 8.0], and added to Greiner-UV Star 96-well microplates for analysis. Reduction of N4-acetylcytidine was analyzed on a Biotek Synergy plate reader by monitoring the absorbance of N4-acetylcytidine (λmax = 300 nm), and cytidine (λmax = 270 nm). For N4-acetylcytidine reactions, percent decrease in N4-acetylcytidine was calculated from 300 nm absorbance values using the formula: "% decrease = $(A_{ac4C(start)} - A_{ac4C(end)})/(A_{ac4C(untreated)} - A_{water(blank)})*100$."

**UV spectroscopic analysis of ac$^4$C deacetylation**

Model reactions to assess the rate of acid- and base-induced deacetylation of ac$^4$C were performed using free N4-acetylcytidine nucleoside. Stock solutions of N4-acetylcytidine (2.5 mM) were prepared fresh daily in water. Reactions (50 µL) consisted of N4-acetylcytidine (250 µM) and reaction buffer (KCl/HCl buffer [pH 1] or $NaHCO_3$ buffer [pH 10]) added to a Greiner-UV Star 96-well half-area microplate. Control reactions were set up similarly with cytidine (250 µM). Deacetylation of N4-acetylcytidine was analyzed on a Biotek Synergy plate reader by monitoring the absorbance of N4-acetylcytidine (pH 1 $\lambda_{max}$ = 310 nm; pH 10 $\lambda_{max}$ = 300 nm) over 18 hr. For N4-acetylcytidine reactions, percent decrease in N4-acetylcytidine was calculated

914 from $\lambda_{max}$ absorbance values using the formula: "% decrease = ($A_{ac4C(start)}$ – $A_{ac4C}$
915 $_{(end)}$)/($A_{ac4C(untreated)}$)*100."

916 ***In vitro* transcription of synthetic ac$^4$C-containing RNAs (spike-in controls)**

917 *In vitro* transcription was performed with the HiScribe T7 Kit (New England Biolabs), according
918 to the manufacturer's instructions using DNA templates containing a T7 promoter upstream of a
919 template sequence harboring a single cytidine within an ACA, GCA, ACG, or GCG sequence
920 context (**Supplementary Table 1a**). An exemplary construct is given below. For ac$^4$C-
921 containing transcripts, CTP was replaced in the reaction mixture with ac$^4$CTP (10 mM) as
922 described previously[12]. *In vitro* transcription reactions were analyzed by denaturing
923 polyacrylamide gel electrophoresis on 10% TBE-urea gels and visualized using SYBR Gold
924 staining. Synthetic RNA products were used in ac$^4$C-seq, LC-MS quantification, and RT stop
925 experiments, the latter of which was performed as previously described [11].

926 **Mass spectrometry analysis of ac$^4$C in synthetic spike-in controls**

927 Mass spectrometry analysis of ac$^4$C reduction in RNA probes was assessed following nuclease
928 digest as described previously[12]. Briefly, *in vitro* transcribed ac$^4$C RNA was treated with
929 nuclease P1 (2U/10 µg RNA, #N8630, Sigma) in 50 µL of buffer containing 100 mM ammonium
930 acetate [pH 5.5], 2.5 mM NaCl and 0.25 mM $ZnCl_2$ for 2 hr at 37 °C. Sample volumes were
931 adjusted to 60 µL by adding 3.5 µL of H2O, 6 µL of 10x Antarctic Phosphatase buffer (#
932 B0289S, NEB) and 0.5 µL of Antarctic Phosphatase (1U/10 µg RNA, #M0289S, NEB). Samples
933 were further incubated at 37 °C for 2 hr, adjusted to 150 µL with RNase-free water and filtered
934 via centrifugation to remove enzymatic constituents (Amicon Ultra 3K, #UFC500396). Following
935 lyophilization, samples were reconstituted in 10 µL RNase-free water and analyzed via LC-
936 MS/MS using reverse phase chromatography (Shimadzu LC-20AD) coupled to a triple-
937 quadrupole mass spectrometer (Thermo TSQ-ultra) operated in positive electrospray ionization
938 mode. Quantitation was accomplished by monitoring nucleoside-to-base ion transitions and
939 generating standard curves for each nucleoside using the stable isotope dilution internal
940 standardization method.

941 **Primer extension and RT stop analysis of ac$^4$C RNAs**

942 Primer extension assays were performed using PAGE-purified model RNAs containing a single
943 site of either ac$^4$C or C produced by in vitro transcription (sequence provided above). For each
944 reaction, RNA (2 µg) was treated in a final reaction volume of 100 µL. $NaBH_4$-treated samples:
945 1 M $NaBH_4$ was added to 2 µg RNA in nuclease-free $H_20$ to a final concentration of 100 mM and
946 samples incubated for 60 min at 37 °C, $NaBH_4$ was quenched with 1 M HCl (15 mL), and
947 neutralized by the addition of 1 M Tris-Cl [pH 8.0] buffer (15 mL). $NaCNBH_3$ treated samples: 1
948 M $NaCNBH_3$ was added to 2 µg RNA in nuclease free $H_2O$ to a final concentration of 100 mM.
949 Reactions were initiated by addition of 1 M HCl to a final concentration of 100 mM and samples
950 incubated 20 minutes at room temperature (20 °C). Reaction was stopped by neutralizing pH by
951 the addition of 30 µL 1 M Tris-HCl pH 8.0. Untreated control samples: 1 M HCl was added to 2
952 µg RNA in nuclease-free water to a final concentration of 100 mM and samples incubated 20

min at room temperature (20 °C). Reactions were stopped by neutralizing pH by the addition of 30 µL 1 M Tris-HCl pH 8.0. Reactions were adjusted to 200 mL with $H_2O$, ethanol precipitated, desalted with 70% ice-cold ethanol, briefly dried on Speedvac, resuspended in $H_2O$, and quantified by absorbance using a Nanodrop 2000 spectrophotometer. RNA from individual reactions (5 pmol) was incubated with 5'-Cy5 IVT primer (5'-/Cy5/ACTCATCACTTTTCTCCCTCTACACAATC-3'; 3.5 pmol) in a final volume of 50 µL. Individual reactions were heated to 65 °C for 5 min and cooled at a rate of 5 °C/min to a final temperature of 4 °C to facilitate annealing, with the following buffer conditions used for specific RTs: AMV: 1X AMV reaction buffer (New England Biolabs), 1.0 mM dNTPs; Superscript III: 500 mM dNTPs; TGIRT: 1X TGIRT reaction buffer (Ingex), 5 mM $MgCl_2$. After annealing, reverse transcriptions were performed as follows: 1) AMV reactions: 100 units RNaseOUT (Invitrogen), 25 U AMV RT, incubate 60 min, 48 °C; 2) Superscript III: 1x SSIII reaction buffer (from 10x stock; ThermoFisher), 5 mM $MgCl_2$, 10 mM DTT, 100 U RNaseOUT, 500 U Superscript III, incubate 60 min, 48 °C; 3) TGIRT reactions: first add 5 mM DTT, 500 U TGIRT RT, incubate 20 min room temperature, then add 500 mM dNTPs, incubate 1 hr, 57 °C. After the indicated incubation time, reactions were adjusted to 200 µL with $H_2O$, phenol:chloroform extracted, ethanol precipitated, desalted with 70% ice-cold ethanol, briefly dried on Speedvac, and resuspended in 20 mL of 1X RNA denaturing RNA loading buffer. Samples were heated at 95 °C for 4 min, cooled on ice, loaded onto a 10% denaturing polyacrylamide gel and run at 400 V (20 V/cm) for 5 hr. Gels were fluorescently visualized using an ImageQuant Las4010 (GE Healthcare) with red LED excitation ($\lambda_{max}$ 630 nm) and a R670 filter, with band intensities quantified by densitometry using Imagequant software. To calculate product/stop ratio, the fluorescence intensity of the bands observed at the $ac^4C$ site (-1, 0 or +1) were divided the total fluorescence intensity of all other primer extension products observed in each gel lane.

**Reverse transcription and misincorporation analysis of RNAs by Sanger sequencing**

For each reaction, RNA (1 µg) was incubated with either sodium cyanoborohydride (100 mM in $H_2O$ + 100 mM HCl) or untreated 'mock' control ($H_2O$ + 100 mM HCl) in a final reaction volume of 100 uL. Samples were incubated for 20 min at 20 °C. Reactions were stopped by neutralization of pH by the addition of 30 uL 1 M Tris-HCl pH 8.0. Reactions were adjusted to 200 µL with $H_2O$, Ethanol precipitated, desalted with 70% ice-cold ethanol, briefly dried on Speedvac, resuspended in $H_2O$, and quantified by absorbance using a Nanodrop 2000 spectrophotometer. RNA from individual reactions (200 pg) was incubated with 4.0 pmol RT primer. in a final volume of 20 µL. Individual reactions were heated to 65 °C for 5 min and transferred to ice for 3 min to facilitate annealing in 1X TGIRT reaction buffer (Ingex), 5 mM $MgCl_2$. After annealing, reverse transcriptions were performed as follows using TGIRT-III; DTT was added to 5 mM along with 100 U TGIRT RT and 25 U RNasin Plus (Promega). Reaction Incubated 20 min room temperature. RT reaction was initiated by addition of dATP, dTTP and dCTP to 500 mM and dGTP to 250 mM. Reactions were incubated 1 hr, 57 °C. 2 µL cDNA was used as template in 50 mL PCR reaction with Phusion Hot start flex (New England Biolabs). Reaction conditions: 1X supplied HF buffer, 2.5 pmol each forward and reverse primer, 200 mM each dNTPs, 2 U Phusion hot start enzyme, 2 µL template and the following specific conditions:

*In vitro* transcribed "single ac$^4$C": Primers: IVT rev (PCR primer), IVT forward (PCR primer). Thermocycling conditions: 71 °C annealing, 34 cycles.

Human 18S rRNA, helix 45 ac$^4$C site: Primers: human 18S helix 45 fwd, human 18S helix 45 rev. Thermocycling conditions: 67.4 °C annealing, 34 cycles.

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 (GeneWiz) using the forward PCR primer for 18S sites and reverse PCR primer for IVT "single ac$^4$C". Processed sequencing traces were viewed using 4Peaks software. Peak height for each base was measured and the percent misincorporation was determined using the equation: "Percent misincorporation = (Sum of non-cognate base peaks intensities)/(sum of total base peaks)*100%".

**Ac$^4$C-seq library preparation**

Strand-specific ac$^4$C-seq libraries were generated on the basis of previously described protocols[42,43]. Briefly, RNA was first subjected to FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific), followed by a 3′ ligation of an RNA adaptor using T4 ligase (New England Biolabs). Ligated RNA was reverse transcribed using TGIRT-III (InGex), and the cDNA was subjected to a 3′ ligation with a second adaptor using T4 ligase. The single-stranded cDNA product was then amplified for 9–12 cycles in a PCR reaction. Libraries were sequenced on Illumina NextSeq 500 or NovaSeq 6000 platforms generating short paired-end reads, ranging from 25 to 55 bp from each end.

**Samples used in ac$^4$C-seq analysis**

*Human*: Three experiments were conducted. In the first experiment, total RNA from WT HeLa cells or cells with reduced expression of Nat10[10] were $NaCNBH_3$-treated (with and without alkali pre-treatment) or mock-treated in three biological replicates. In the second, a set of 5 poly(A)-enriched HeLa samples (3 and 2 biological replicates for WT and Nat10 knock-down, respectively) were $NaCNBH_3$-treated or mock-treated. For the third experiment poly(A)-enriched HEK-293T cells co-overexpressing Nat10 and Thumpd1 (2 biological replicates) and a sample of WT cells were $NaCNBH_3$-treated (with and without alkali pre-treatment) or mock-treated.

*Yeast*: Two experiments were conducted. In the first, biological duplicates of WT S. cerevisiae cells and cells expressing a catalytic mutant of Kre33 [1] $NaCNBH_3$-treated (with and without alkali pre-treatment) or mock-treated. In the second, cells co-overexpressing Kre33 and Tan1 in a Kre33-catalytic mutant strain were analyzed in comparison to WT *S. cerevisiae* cells. One replicate of the co-overexpression cells expressed Tan1 under a constitutive GPD promoter, while the other under a GAL1-inducible promoter. These cells were grown in YPD and YPG, respectively, along with a matching WT sample grown at same conditions. These four samples were $NaCNBH_3$-treated or mock-treated. All libraries of yeast were prepared from poly(A)-enriched RNA.

*T. kodakarensis*: A total of 17 samples were analyzed, representing 25 treatment conditions. For all samples total RNA was analyzed from a single biological sample, unless stated otherwise. 1) TS559 cells grown at 55, 65, 75, 85 and 95 ˚C were $NaCNBH_3$-treated or mock-treated. For the 85 ˚C condition, four biological replicates were assessed, and one of them also underwent alkali pre-treatment. For 65 and 75 ˚C two biological replicates were assessed. 2) biological duplicates of cells deleted of TkNat10 (TK0754) or TkThumpd1 (*TK2097*) were $NaCNBH_3$-treated. ΔTkNat10 samples were also mock-treated. 3) rRNA-depleted RNA from TS559 cells grown at 85 and 95˚C $NaCNBH_3$-treated or $NaCNBH_3$-treated and mock-treated, respectively. 4) purified ribosomes from TS559 cells grown at 85 ˚C were $NaCNBH_3$-treated.

*Thermococcus sp. AM4*: total RNA from cells grown at 65, 75 and 85 ˚C was $NaCNBH_3$-treated or mock-treated.

*Pyrococcus furiosus*: total RNA from cells grown at 75, 85 and 95 ˚C was $NaCNBH_3$-treated or mock-treated.

*Saccharolobus solfataricus*: total RNA from cells grown at 85 ˚C was $NaCNBH_3$-treated (with and without alkali pre-treatment) or mock-treated. Total of three samples representing a single biological sample.

*Methanocaldococcus jannaschii*: a single sample was $NaCNBH_3$-treated (with and without alkali pre-treatment) or mock-treated.

**Identification of putative ac$^4$C sites**

Reference genomes were generated on the basis of the following genome assemblies: ASM996v1 for *Thermococcus kodakarensis*, ASM27560v1 for *Pyrococcus furiosus*, ASM15120v2 was used for *Thermococcus sp. AM4*, ASM700v1 for *Saccharolobus solfataricus* and ASM9166v1 for *Methanocaldococcus jannaschii*. For human poly(A)-enriched samples we used the GRCh37/hg19 with UCSC Genes annotations, supplemented with tRNA, rRNA, and snRNAs sequences, obtained from the Modomics database[44]. Samples from total RNA of human cells were aligned to a subset of the full reference containing only the tRNA, rRNA, and snRNA sequences. For *Saccharomyces cerevisiae* samples the sacCer3 assembly was used in experiments designed to detect modification in mRNA, whereas a limited reference containing only rRNAs and tRNAs (filtered to only retain non-redundant sequences) was used in experiments designed to detect only sites in these non-coding transcripts.

Samples were aligned to the genome, using STAR aligner[45]. For archaeal and *S. cerevisiae* samples intron size was limited to 500 bases ('alignIntronMax==500'). For poly(A)-enriched samples (applicable to some of the human and yeast samples, as indicated in the main text) duplicated reads and chimeric paires were filtered out by the dedup function of UMI-tools[46] (using '--chimeric-pairs=discard') followed by removal of overlapping reads by the clipOverlap function of bamUtil[47]. For human and yeast samples aligned to a limited reference containing only the ncRNA sequences mentioned above, multiple mapping was allowed ('multiMapping=200').

Single nucleotide variants were detected using the JACUSA software in pileup mode[48], which outputs a tabular format summarizing the abundance of each nucleotide (with minimal coverage of 5 reads) at each position. A custom script was used to extract the misincorporation rate at each position as well as to identify the most abundant nucleotide appearing instead of the WT nucleotide (aka, the "predominant base conversion").

For a position to be considered as putatively modified, it had to meet two sets of requirements. **First**, at the level of an individual $NaCNBH_3$-treated sample compared to a suitable control (whereby the control is in most cases a mock-treated sample, but in some cases also a chemically deacetylated sample or a Nat10 deficient genetic control) the fundamental requirement it had to meet was that the $\chi^2$ test based p-value comparing the misincorporation rates in the treated versus control samples was lower than 0.05. In experiments with multiple replicates, the $\chi^2$ test was conducted on 'pooled samples' combining misincorporation information from all replicates. **Second**, to reduce the computational load, we applied this statistical framework only to sites matching the minimal criteria below: 1) at least three reads with misincorporations in the $NaCNBH_3$-treated sample (or WT sample, when comparing to Nat10-deficient). 2) a misincorporation rate >MIN_RATE in the NaCNBH3-treated sample (for archaea we used a MIN_RATE_TREAT=2%, for human and yeast with larger genomes and consequently slightly reduced signal:noise ratios we used 3%). 3) a misincorporation rate lower than MAX_RATE_CONT in the control sample (MAX_RATE_CONT=5% in archaea, 1% in human and S. cerevisiae). 4) Misincorporation rates in the $NaCNBH_3$-treated sample were at least 2% higher than in their control counterparts 5) The predominant base conversion at the site in the $NaCNBH_3$-treated sample was from cytidine to thymidine (C->T). To eliminate redundancies, positions harbouring identical sequences in a 21-bp window (10 bp upstream + 10 bp downstream) surrounding the putative site were filtered to retain only one. Furthermore, when possible on the basis of the experimental design, we demanded that such a site be reproducibly identified across at least two distinct comparisons. The distinct experimental design for the different organisms (in some cases we monitored distinct temperatures, in others distinct genetic backgrounds, in others we obtained static snapshots under one condition) was taken into consideration, and the precise set of comparisons performed for each organism is detailed in **Supplementary Table 2**. This set of comparison was used to create a final 'catalog of $ac^4C$ sites' for each organism, which was used in downstream analyses. All catalogs, segregated by organism, appear in **Supplementary Table 2**.

**Motif analysis**

For each species, we extracted the 20 nt flanking the $ac^4C$ positions in its catalog of "significantly modified" sites. These 21-nt long sequences were used to generate sequence logos using the WebLogo software (available at https://weblogo.berkeley.edu/logo.cgi)[49], in which the height of each stack indicates the information content at that position (measured in bits), whereas the height of letters within the stack reflects the relative frequency of the corresponding nucleic acid at that position.

**Targeted $ac^4C$-sequencing**

mRNA samples treated with $NaCNBH_3$ were incubated with Turbo DNase (Invitrogen) for 30 min at 37 ˚C. 400 ng of the DNase-treated mRNA were reverse transcribed using TGIRT-III (InGex), with random primers (Applied Biosystems). Following cleanup of cDNA using Dynabeads MyOne SILANE beads (Life technologies), 10 cycles of PCR were carried out using Kapa HiFi HotStart Readymix PCR kit (Kapa Biosystems), and pairs of primers described in **Supplementary Table 1a**. 1 μl of the PCR reaction was used as template for a second PCR reaction (Kapa HiFi, 25 μl reaction volume, 20 cycles), in which barcoded Illumina adaptors were added. Amplicons were analyzed on 2% E-gel EX agarose gels (Invitrogen), and cleaned using two rounds of AMPure XP beads (Beckman Coulter). For targeted $ac^4C$-sequencing of over expressed sequences, total RNA samples were treated with $NaCNBH_3$ and incubated with Turbo DNase (Invitrogen) for 30 min at 37 ˚C. 600 ng of the DNase-treated total RNA were reverse transcribed using TGIRT-III (InGex), with random primers (Applied Biosystems). Following cleanup of cDNA using Dynabeads MyOne SILANE beads (Life technologies), 20 cycles of PCR were carried out using Kapa HiFi HotStart Readymix PCR kit (Kapa Biosystems), adding the barcoded Illumina adaptors.

**Construction of plasmids for over expression of WT (CCG) and mutated (CCA) $ac^4C$ sites**

The sequences described in **Supplementary Table 1a** were cloned using FastDigest SgsI (AscI) and BcuI (SpeI) restriction enzymes (Thermo Scientific) into **pZDonor FC** plasmid, as a 3' UTR of a reporter gene[50].

**Targeted $ac^4C$ -sequencing of a pool of sequence variants of BAZ2A mRNA**

Pool design:

A 91-bases long sequence surrounding the ac4C site identified in *BAZ2A* mRNA was used as a WT control fragment. Variants of the WT *BAZ2A* fragment were made by introducing a single point mutation at each base of the WT sequence, by replacing it with all possible bases. *BAZ2A* fragments were preceded by an 8 bases barcode, allowing each variant to be uniquely mapped, and flanked by 1) SpeI and AscI restriction sites to facilitate cloning, 2) Illumina adapter sequences to allow sequencing, and 3) primer sequences to allow amplification of the entire construct in the cloning stage.

Cloning of the oligonucleotide pool:

The pool of sequences was cloned as 3' UTR downstream of a reporter gene in the **pZDonor FC** plasmid, essentially as described in[51]. Specifically, the library was amplified in 5 different PCR reactions, each using 50 pg as a template and 14 cycles. The reactions were combined, cleaned by QIAquick PCR purification kit (Qiagen), and a total of 540ng was cut by SgsI (AscI) and BcuI (SpeI) restriction enzymes (FastDigest, Thermo Scientific). Following electro-elution from a gel using Midi GeBAflex tubes (GeBA, Kfar Hanagid, Israel), the library was ligated (in 1:1 ratio) to **pZDonor FC** plasmid digested by SgsI and BcuI, using CloneDirect Rapid Ligation kit (Lucigen Corporation) and transformed into E. cloni 10G electrocompetent cells (Lucigen) in a single cuvette. The bacteria were grown on four 14-cm plates, reaching in average about 1500

colonies per each sequence variant. Plasmids were purified directly from collected bacterial colonies.

Transfection, treatment and library preparation:

The plasmids pool was transfected to 10-cm plates of HEK293T cells in replicates using PolyJet reagent (SignaGen Laboratories), either by itself (2 µg) or together with both Nat10 and Thumpd1 (1.5 µg each). For targeted ac$^4$C-sequencing of the library variants, total RNA samples were treated with NaCNBH3 and incubated with Turbo DNase (Invitrogen) for 30 min at 37 ˚C. 1 µg of the DNase-treated total RNA was reverse transcribed using TGIRT-III (InGex), with random primers (Applied Biosystems). Following cleanup of cDNA using Dynabeads MyOne SILANE beads (Life technologies), half of the cleaned cDNA was used in a 25 cycles PCR reaction, using Kapa HiFi HotStart Readymix PCR kit (Kapa Biosystems), and Illumina adaptors as primers.

Analysis:

SAMtools mpileup was used to assess misincorporation rates at the ac$^4$C site of BAZ2A variants.

**mRNA expression analysis**

To estimate expression levels, reads were aligned against the human, yeast or *T. kodakarensis* genome using RSEM (version 1.2.31) in paired-end and strand-specific mode with default parameters[52]. For robust comparison between different samples, we used trimmed mean of M values (TMM) normalization[53] of the RSEM read counts as implemented by the NOISeq package[54] in R.

**Analysis of codon enrichment and distribution across transcript body**

Our analysis identified 146 and 119 putative ac$^4$C sites in mRNA of human and *T. kodakarensis*, respectively. For each site its relative position within the codon was identified on the basis of the genome annotation. As a control, the distribution of all remaining cytidines embedded in CCG sequences in the examined mRNAs was calculated. For *T. kodakarensis*, we further calculated the distribution of the putative ac$^4$C sites and the control cytidines between specific codons encoding for the different amino acids. For human sites we mapped the location of each ac$^4$C site and control cytidines (as described above) within the transcript body (i.e., 5' UTR, CDS or 3' UTR) and calculated the distribution across transcript regions.

**Multiple alignment of tRNAs**

All *T. kodakarensis* tRNA sequences were multiply aligned against each other using MAFFT v7.402 with default parameters[55]. Manual inspection of aligned sequences facilitated assignment of ac$^4$C sites into distinct regions within the tRNA structure and into specific positions within a canonical model of a tRNA.

## Conservation analysis between Archaea

Sequences of 16S, 23S, 5S, RNaseP RNA and SRP RNA were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/) from genome references NC_006624.1, NC_018092.1 and NC_016051.1 for *T. kodakarensis*, *P. furiosus* and *T. sp. AM4*, respectively. Multiple sequence alignment was conducted across all three archaea for each gene separately using the Clustal Omega software with default parameters (https://www.ebi.ac.uk/Tools/msa/clustalo/)[56]. A custom script was used to detect ac$^4$C at positions conserved between at least two species and assign it with the relevant misincorporation rates as calculated using ac$^4$C-seq across all samples. This data set was used for archaea conservation-related analysis presented in the main text.

## Phylogenetic tree

A phylogenetic tree for the archaea analyzed by ac$^4$C-seq was generated using the default parameters of phyloT tree generator (https://phylot.biobyte.de) based on the following NCBI taxonomy IDs: *T. kodakarensis* - 69014, T. AM4 - 246969, *P. furiosus* - 1185654, *S. solfataricus* - 555311 and *M. jannaschii* - 2190.

## Comparison between ac$^4$C sites in *T. kodakarensis* rRNAs as measured by ac$^4$C-seq and LC-MS

A total of 172 ac$^4$C sites at CCG motifs were identified in *T. kodakarensis* rRNA under the full set of comparisons detailed in **Supplementary Table 2**. While LC-MS identified a total of 146 potential ac$^4$C sites, only 25 of these could be uniquely assigned to specific positions within the ribosome, due to redundancies in the oligo sequences identified in the LC-MS. Thus all comparisons of ac$^4$C between the methods were conducted on a subset of these 25 sites.

## Northern blot analysis of ac$^4$C in archaeal total RNA

Immunonorthern blots were performed using Ambion NorthernMax reagents (Thermo Fisher Scientific). The amount of RNA used was dependent upon sample type, with 15 µg used for analysis of human and yeast total RNA, and 3 µg used hyperthermophilic archaea. Equal amounts of RNA were mixed together with 1 vol of NorthernMax-Gly Sample Loading Dye (Thermo Fisher Scientific), incubated at 65 ˚C for 30 min, and separated on a 1% agarose-1X Glyoxal Gel prepared using 10X NorthernMax-Gly Gel Prep/Running Buffer (Thermo Fisher Scientific). Gels were run at 75 V for approximately 70 min, or until the dye front had migrated about 3 inches/7.3 cm. Loading controls were analyzed by UV-imaging of ethidium bromide prior to transfer. RNA was transferred onto Amersham Hybond-N+ membranes (GE Healthcare) using a downward capillary method. After transfer, membranes were crosslinked three times with 150 mJ/cm$^2$ in a UV254nm Stratalinker 2400 (Stratagene). Membranes were then blocked in a solution of blocking buffer (5% non-fat milk in 0.1% Tween-20 TBS [TBST]) for 1 hr at room temperature and washed 3 times at 5 min each in 0.1% TBST. Membranes were then incubated overnight at 4 °C with the anti-ac$^4$C antibody (1:10,000 dilution, Abcam) in blocking buffer. Membranes were washed 3 x 5 minutes in 0.1% TBST and then incubated with HRP-

conjugated secondary anti-rabbit IgG in 5% non-fat milk in 0.1% TBST at room temperature for 2 hr. Membranes were washed 3 times at 10 minutes each in 0.1% TBST. SuperSignal ELISA Femto Maximum Sensitivity Substrate reagent (Thermo Fisher Scientific) was added directly to the membrane and signal was detected via chemiluminescent imaging. Typical exposure times ranged from 2 to 20 min depending on the concentration of individual RNA samples. We found that for hyperthermophilic archaea a 2-min exposure time was optimal but yeast and human RNA required a 15- to 20-min exposure time to yield optimal results.

**LC-MS analysis of ac$^4$C in total RNA**

For assessment of cellular ac$^4$C levels by LC-MS, total RNA was analyzed using a similar method as previously described[57]. Briefly, prior to UHPLC-MS analysis, 2000 ng of each oligonucleotide was treated with 0.5 pg/µl of internal standard (IS), isotopically labeled guanosine, [$^{13}$C][$^{15}$N]-G (Cambridge Isotope Laboratories Inc.). The enzymatic digestion was carried out using Nucleoside Digestion Mix (New England BioLabs) according to the manufacturer's instructions. Finally, the digested samples were lyophilized and reconstituted in 100 µl of RNAse-free water, 0.01% formic acid prior to UHPLC-MS/MS analysis. The UHPLC-MS analysis was accomplished on a Waters XEVO TQ-STM (Waters Corporation, USA) triple quadruple mass spectrometer equipped with an electrospray source (ESI) source maintained at 150 °C and a capillary voltage of 1 kV. Nitrogen was used as the nebulizer gas which was maintained at 7 bars pressure, flow rate of 500 l/h and at temperature of 500 °C. UHPLC-MS/MS analysis was performed in ESI positive-ion mode using multiple-reaction monitoring (MRM) from ion transitions (m/z 286.16 > 154.07 and m/z 286.16 > 112.06) previously determined for ac$^4$C[57]. A Waters ACQUITY UPLCTM HSS T3 guard column, 2.1x 5 mm, 1.8 µm, attached to a HSS T3 column, 2.1x50 mm, 1.7 µm were used for the separation. Mobile phases included RNAse-free water (18 MΩcm$^{-1}$) containing 0.01% formic acid (Buffer A) and 50:50 acetonitrile in Buffer A (Buffer B). The digested nucleotides were eluted at a flow rate of 0.5 mL/min with a gradient as follows: 0-2 min, 0-10%B; 2-3 min, 10-15% B; 3-4 min, 15-100% B; 4-4.5 min, 100 %B. The total run time was 7 min. The column oven temperature was kept at 35oC and sample injection volume was 10 µl. Three injections were performed for each sample. Data acquisition and analysis were performed with Waters software MassLynx V4.1 and TargetLynx. Calibration curves were plotted using linear regression with a weight factor of 1/x.

**Preparation of RNase digests and Direct nanoflow LC-MS and tandem MS of rRNA fragments**

rRNAs was extracted from purified *T. kodakarensis* 70S ribosomes. An aliquot of the sample (100 µl, 1 mg/mL) was mixed with 800 µl of ISOGEN reagent (Nippon Gene) and passed 100 times through a 23-gauge needle. The sheared sample was mixed with 200 µl of chloroform and centrifuged at 10,000 xg for 15 min at 4 °C. The resulting upper phase (~500 µl) was mixed with a glycogen solution (0.5 µl, 20 mg/mL) and isopropanol (500 µl) and centrifuged to yield rRNAs as a precipitate. The precipitate was dissolved in RNase free water and stored -80 °C until further use. The three rRNA classes (5S, 16S and 23S) were separated by reversed-phase LC through a PLRP-S 4000Å column (4.6 × 150 mm, 10 µm, Agilent Technologies). After applying ~10 µg total RNA to the column, the rRNAs were eluted with a 60-min linear gradient of 12–14%

(v/v) acetonitrile in 100 mM TEAA, pH 7.0, 0.1 mM diammonium phosphate at a flow rate of 200 µl/min and 60°C while monitoring the eluate at A260[58].

RNA (~50 ng) was digested with RNase T1 (20 ng) in 100 mM triethylammonium acetate buffer (pH 7.0) at 37 °C for 1 hr. The RNA fragments were separated using a direct nanoflow LC-MS system as described[59,60]. Briefly, the digests were injected onto a reversed-phase Develosil C30-UG tip column (150 µm i.d. × 120 mm, 3-µm particle size; Nomura Chemical Co., Ltd.) equilibrated with solvent A (10 mM TEAA, pH 7, in water:methanol, 9:1). Samples were eluted at 100 nl/min with a 60-min 0–24.5% linear gradient of solvent B (10 mM TEAA, pH 7:acetonitrile, 60:40). The column was subsequently washed with 70% B for 10 min and re-equilibrated with A.

Each LC eluate was sprayed online at −1.4 kV with the aid of a spray-assisting device[60] into a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) operating in the negative ion mode and in the data-dependent mode to automatically switch between MS and tandem MS acquisition. Full-scan mass spectra (m/z=480–1980) were acquired at a mass resolution of 350 000. At most, the five most intense peaks, (>100,000 counts/s with a 60-ms maximum injection time), were isolated within a 3-m/z window for fragmentation. Precursors were fragmented by switching to a higher energy CID mode with a normalized collision energy of 20 or 50%. To retain mass resolution and to increase spectral quality, three tandem mass spectral micro-scans were acquired for each sample. A fixed starting value of m/z=100 was set for each tandem mass spectrum.

**Interpretation of the tandem mass spectra and quantification of modifications**

Ariadne[61] (http://ariadne.riken.jp/) was used for assignment of the tandem mass spectral peaks in conjunction with the sequence of rRNAs of *T. kodakarensis* (Gene ID: 3253116, 3253120 and 3253121). The Ariadne search parameters were: the maximum number of missed cleavages was one; two methylations per RNA fragment at any residue position were allowed; an RNA mass tolerance of ± 20 ppm and a tandem spectral tolerance of ± 50 ppm were allowed.

The quantification of post-transcriptional modification (PTM) was performed by the peak area-based method. The target oligonucleotide peaks were obtained from extracted-ion chromatograms with their theoretical mass values (± 5 ppm). Each peak area was measured using the Xcalibur software (Thermo Fisher Scientific) including a manual determination of the start-end of peak. The stoichiometry of PTM was calculated from the peak areas obtained by MS with the following equation where P and N refer peak areas of the oligonucleotide with or without PTM, respectively. Stoichiometry (%) = 100 x P/(P+N)

**Proteomic analysis of *T. kodakarensis***

Proteins isolated from cultures of *T. kodakarensis* were precipitated by trichloroacetic acid (TCA) and washed twice with cold acetone prior to digestion for MS analysis. Briefly, TCA precipitated proteins were resuspended in 100 mM Tris pH 8.5 containing 8 M urea. Cysteine residues were reduced by 5 mM TCEP for 30 minutes at room temperature and further modified

by 2-chloroacetamide for 30 min in the dark at room temperature. Proteins were first digested by recombinant Lys-C (Promega) overnight at 37 °C with shaking. The urea was diluted to 2 M before additional digestion overnight at 37 °C by the addition of trypsin at 1:100 enzyme to substrate (Promega). The digestion reaction was quenched with the addition of formic acid to 5% final. Peptides were quantitated by the Pierce Colorimetric Peptide Assay (Thermo Scientific) and diluted in buffer A (5% acetonitrile (ACN), 0.1% formic acid (FA)) such that 1µg was analyzed per technical replicate. Each sample was trapped on an Acclaim PepMap 100 C18 column (5µm particles 0.3 mm x 5 mm, Thermo Scientific) using the Ultimate 3000 autosampler (Dionex). Using chromatography conditions previously optimized[62], peptides were separated on an in-house packed reverse phase chromatography column (1.9 µm particles (ReproSil, Dr. Maish), 75µm x 20cm), directly interfaced to a QExactive Plus (QE+) mass spectrometer (Thermo Scientific). Peptides were eluted over a quick gradient from 2-7% buffer B (80% ACN, 0.1% formic acid) in 10 minutes before the gradient was gradually increased to 40% buffer B over 6 hours before ramping to 95% B in 15 minutes. The flow was kept at 95% B for 15 minutes before 20 minutes of re-equilibration at 2% B prior to the next injection. Flow rate was 180 nL/min. The application of a 2.5 kV distal voltage electrosprayed the eluting peptides directly into the QE+ mass spectrometer equipped with the Nanospray Flex source (Thermo Scientific). Full MS spectra were recorded on the eluting peptides at a resolving power of 70,000 over a 400 to 1600 m/z range, followed by higher energy dissociation (HCD) fragmentation at 30% normalized collision energy on the 15 most intense ions selected from the full MS spectrum. MS2 spectra were collected in the Orbitrap at a resolving power of 17,500. Dynamic exclusion was enabled for 30 s[63]. Mass spectrometer scan functions and HPLC solvent gradients were controlled by the XCalibur data system (Thermo Scientific).

RAW files were extracted into .ms2 file format[64,65] (McDonald et al., 2004) using RawDistiller v. 1.0, in-house developed software[64]. RawDistiller D(g, 6) settings were used to abstract MS1 scan profiles by Gaussian fitting and to implement dynamic offline lock mass using six background polydimethylcyclosiloxane ions as internal calibrants[64]. MS/MS spectra were first searched using ProLuCID[66] with a 10 ppm mass tolerance for peptide and 25 ppm tolerance for fragment ions. Trypsin specificity was imposed on both ends of candidate peptides during the search against a protein database containing 2301 *T. kodakarensis* proteins (NCBI 2018-11-09 release), as well as 386 usual contaminants such as human keratins, IgGs, and proteolytic enzymes. To estimate false discovery rates (FDR), each protein sequence was randomized (keeping the same amino acid composition and length) and the resulting "shuffled" sequences were added to the database, for a total search space of 5440 amino acid sequences. A mass of 15.9949 Da was differentially added to methionine residues.

DTASelect v.1.9[67] was used to select and sort peptide/spectrum matches (PSMs) passing the following criteria set: PSMs were only retained if they had a DeltCn of at least 0.08; minimum XCorr values of 1.0 for singly-, 1.4 for doubly-, and 2.1 for triply-charged spectra; peptides had to be at least 7 amino acids long. Results from each sample were merged and compared using CONTRAST[67]. Combining all replicate injections, proteins had to be detected by at least 2 peptides and/or 2 spectral counts. Proteins that were subsets of others were removed using the parsimony option in DTASelect on the proteins detected after merging all runs. Proteins that were identified by the same set of peptides (including at least one peptide unique to such

protein group to distinguish between isoforms) were grouped together, and one accession number was arbitrarily considered as representative of each protein group.

NSAF7[68] was used to create the final reports on all detected peptides and non-redundant proteins identified across the different runs. Spectral and protein level FDRs were, on average, 0.17 ± 0.05% and 0.18 ± 0.05%, respectively.

**Cryo-EM data acquisition and analysis**

A 3.5 µl of 70S ribosome sample (0.25 mg $mL^{-1}$ for the WT strains grown at 85 ˚C and 65 ˚C, and 0.4 mg $mL^{-1}$ for the $ac^4C$-deficient strains) was applied on glow-discharged holey carbon grids (Quantifoil R2/2) coated with a thin layer of continuous carbon film. Grids were blotted (3 s) and plunge-frozen using a Vitrobot Mark IV (FEI, Thermofischer scientific). Micrographs were recorded at liquid nitrogen temperature on a Titan Krios electron microscope (FEI, Thermofischer scientific) operating at 300 kV and equipped with a Falcon 3 direct electron detector (FEI, Thermofischer Scientific). Nominal magnification used was 96K and corresponded to a calibrated pixel size of 0.85 Å/pixel, with a dose rate of ~1.16 e-/$Å^2$/s and defocus values ranging from -0.5 to -1.5 µm. Automatic data acquisition was done using EPU (FEI, Thermofischer scientific) and yielded a total of 2,509 micrographs for the WT85, 3115 for the WT65 and 4,211 for the mutant. Micrographs were processed using Motioncor2[69] to correct for patched frame motion and dose-weighting and contrast transfer function parameters were estimated by CTFFIND 4.1[70,71]. Particle picking, extraction and classifications were performed using Relion 3.0[72]. The 60-Å lowpass filtered cryo-EM map of the *P. furiosus* ribosome (EMD-2009) was used as an initial reference and has been utilized for further particle classification in 3D. Final maps reconstructed from 53,737, 283,424 and 116,586 particles for the WT85, WT65 and mutant strains, respectively, were obtained through multibody refinement with the LSU, the SSU body and SSU head masked individually as demonstrated in **Extended Data Fig. 7a**[72]. Density maps were corrected for the modulation transfer function of the detector, and then sharpened by applying a negative B-factor that was estimated using automated procedures in Relion3[73]. Averaged map resolutions were 2.95 Å, 2.55 Å and 2.65 Å for the WT85, WT65 and TkNat10, respectively and were determined using the gold-standard FSC = 0.143 criterion as implemented in Relion3 and M-triage as implemented in Phenix[74] (**Extended Data Fig. 7b-d**). Local resolutions were estimated using Resmap[75] (**Extended Data Fig. 8a**).

**Model building and refinement**

The initial template of *T. kodakarensis* ribosome was derived from the cryo-EM model of *P. furiosus* ribosome (PDB 4v6u). The model was docked into the EM density maps using UCSF Chimera[76], followed by iterative manual building in Coot[77]. Coordinates and library files for the modified residues were generated through phenix.elbow[78] and were manually docked into the relevant positions using COOT following by real-space refinement. The final model was subjected to global refinement and minimization in real space using phenix.real_space_refine in Phenix[74]. Molprobity[79] was used to evaluate model geometry. The final refinement parameters are provided in **Supplementary Table 6**, and map *vs.* model diagrams are in **Extended Data Fig. 7e-g**. Examples of $ac^4C$ model in density are at **Extended Data Fig. 8b**.

## Biophysical characterization of ac$^4$C containing hairpins

*In vitro* transcription was performed using the NEB Highscribe T7 highyield RNA synthesis kit according to the manufacturer's instructions using DNA templates containing a T7 promoter upstream of a template sequence (**Supplementary Table 1a**). For ac$^4$C-containing transcripts, CTP was replaced in the reaction mixture with ac$^4$CTP (50 mM). Crude in vitro transcription reactions were purified by denaturing polyacrylamide gel electrophoresis (PAGE). Full length product bands were visualized by UV shadowing and excised with a razor blade. RNA was extracted by crushing the gel slices and shaking in 500 mM ammonium acetate with 0.2 mM EDTA pH 8.0. RNA was desalted by four sequential rounds of dilution and concentration in a 1K MWCO centrifugal ultrafiltration device. Prior to use in DSC and CD experiments, purified RNAs were analyzed for purity by denaturing PAGE and visualized using SYBR™ Gold Nucleic Acid Gel Stain from Invitrogen (Carlsbad, CA). DSC experiments were carried out on a VP-DSC instrument (Microcal). Desalted PAGE purified helix 45 oligos were diluted to 18 μM in 1X Oligo DSC buffer (10 mM phosphate buffer, 50 mM NaCl) and folded by heating to 95 ˚C for 10 min and rapidly cooled by placing on ice for 10 minutes. Samples were vacuum degassed with stirring for 8 min at 35 ˚C. DSC was equilibrated with 550 μL freshly degassed 1x Oligo DSC buffer in sample and reference cells through multiple scan cycles until a stable and flat differential heat flow curve was established. During downscanning, samples cell was emptied, and 550 μL freshly degassed helix 45 hairpins were loaded between 40 °C and 35 °C. Samples were equilibrated at 35 °C for 15 minutes and calorimetric data was collected from 35 °C to 120 °C at a scan rate of 1 °C/min. Raw DSC data from each scan was processed by linear baseline subtraction and the absolute value of each baseline was adjusted to allow curves to be observed on a single plot. Melting temperatures were calculated as the mean value of the local maxima of the major transition on each scan (n=3) and errors were calculated as standard deviation. CD analyses were performed on a JASCO J-1500 CD Spectrometer (Easton, MD, USA) using a 1 mm pathlength quartz cuvette. Briefly, desalted helix 45 oligos were diluted to 5 μM in 1X melting buffer (100 mM NaCl, 1.97 mM KCl, 0.1 mM EDTA and 8.7 mM sodium phosphate [pH 7.4]) and folding by fast cooling. Denaturation curves were recorded by monitoring the change in ellipticity at 260 nm while the temperature was increased from 30 °C to 95 °C at a rate of 2 °C/min. The minimum points in the first derivative curves of CD melting spectra were recorded (n=3) and errors were calculated as standard deviation.

## Code availability

Code for the analyses described in this paper is available from the corresponding author upon request.


## Funding

S.S. is funded by the Israel Science Foundation (543165), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 714023), and the Estate of Emile Mimran. S.S. is the incumbent of the Robert Edward and Roselyn Rich Manson Career Development Chair in Perpetuity. M.S.B. is supported by the Zuckerman STEM Leadership Program, by Ilse Katz Institute for Material

Sciences and Magnetic Resonance Research and by the Helen & Milton A. Kimmelman Center for Biomolecular Structure & Assembly. S.S. and M.S.B. are jointly supported by the Weizmann-Krenter-Katz Interdisciplinary Research grant. J.L.M. is supported by the Intramural Research Program of NIH, the National Cancer Institute, The Center for Cancer Research (ZIA BC011488-06). M.P.W. is supported by the Stowers Institute for Medical Research and the National Institute of General Medical Sciences of the NIH (R01GM112639). T.J.S. is supported by the National Institute of General Medical Sciences of the NIH (R01GM100329) and the Department of Energy, Basic Energy Sciences Division (DE-SC0014597).

## Additional references

31. Ueda, T. *et al.* Mutation of the gene encoding the ribonuclease P RNA in the hyperthermophilic archaeon Thermococcus kodakarensis causes decreased growth rate and impaired processing of tRNA precursors. *Biochem. Biophys. Res. Commun.* **468**, 660–665 (2015).

32. Gehring, A. M., Sanders, T. J. & Santangelo, T. J. Markerless Gene Editing in the Hyperthermophilic Archaeon Thermococcus kodakarensis. *Bio Protoc* **7**, (2017).

33. Hileman, T. H. & Santangelo, T. J. Genetics Techniques for Thermococcus kodakarensis. *Front. Microbiol.* **3**, 195 (2012).

34. Santangelo, T. J., Cubonová, L. 'ubomíra, James, C. L. & Reeve, J. N. TFB1 or TFB2 is sufficient for Thermococcus kodakaraensis viability and for basal transcription in vitro. *J. Mol. Biol.* **367**, 344–357 (2007).

35. Santangelo, T. J. & Reeve, J. N. Deletion of switch 3 results in an archaeal RNA polymerase that is defective in transcript elongation. *J. Biol. Chem.* **285**, 23908–23915 (2010).

36. Lipscomb, G. L. *et al.* Natural Competence in the Hyperthermophilic ArchaeonPyrococcus furiosusFacilitates Genetic Manipulation: Construction of Markerless Deletions of Genes Encoding the Two Cytoplasmic Hydrogenases. *Applied and Environmental Microbiology* vol. 77 2232–2238 (2011).

37. Oger, P. *et al.* Complete genome sequence of the hyperthermophilic archaeon Thermococcus sp. strain AM4, capable of organotrophic growth and growth at the expense of hydrogenogenic or sulfidogenic oxidation of carbon monoxide. *J. Bacteriol.* **193**, 7019–7020 (2011).

38. Farkas, J. A., Picking, J. W. & Santangelo, T. J. Genetic techniques for the archaea. *Annu. Rev. Genet.* **47**, 539–561 (2013).

39. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer, 2016).

40. Matzov, D. *et al.* The cryo-EM structure of hibernating 100S ribosome dimer from pathogenic Staphylococcus aureus. *Nat. Commun.* **8**, 723 (2017).

41. Morlan, J. D., Qu, K. & Sinicropi, D. V. Selective Depletion of rRNA Enables Whole Transcriptome Profiling of Archival Fixed Tissue. *PLoS ONE* vol. 7 e42882 (2012).

42. Shishkin, A. A. *et al.* Simultaneous generation of many RNA-seq libraries in a single reaction. *Nat. Methods* **12**, 323–325 (2015).

43. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).

44. Machnicka, M. A. *et al.* MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic Acids Res.* **41**, D262–7 (2013).

45. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

46. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).

47. Jun, G., Wing, M. K., Abecasis, G. R. & Kang, H. M. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* **25**, 918–925 (2015).

48. Piechotta, M., Wyler, E., Ohler, U., Landthaler, M. & Dieterich, C. JACUSA: site-specific identification of RNA editing events from replicate sequencing data. *BMC Bioinformatics* **18**, 7 (2017).

49. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

50. Slutskin, I. V., Weingarten-Gabbay, S., Nir, R., Weinberger, A. & Segal, E. Unraveling the determinants of microRNA mediated regulation using a massively parallel reporter assay. *Nat. Commun.* **9**, 1–18 (2018).

51. Weingarten-Gabbay, S. *et al.* Comparative genetics. Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* **351**, (2016).

52. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

53. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).

54. Tarazona, S., García, F., Ferrer, A., Dopazo, J. & Conesa, A. NOIseq: a RNA-seq differential expression method robust for sequencing depth biases. *EMBnet.journal* vol. 17 18 (2012).

55. Katoh, K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* vol. 30 3059–3066 (2002).

56. Sievers, F. *et al.* Fast, scalable generation of high□quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, (2011).

57. Basanta-Sanchez, M., Temple, S., Ansari, S. A., D'Amico, A. & Agris, P. F. Attomole quantification and global profile of RNA modifications: Epitranscriptome of human neural stem cells. *Nucleic Acids Res.* **44**, e26 (2016).

58. Yamauchi, Y. *et al.* Denaturing reversed phase liquid chromatographic separation of non-coding ribonucleic acids on macro-porous polystyrene-divinylbenzene resins. *J. Chromatogr. A* **1312**, 87–92 (2013).

59. Taoka, M. *et al.* An analytical platform for mass spectrometry-based identification and chemical analysis of RNA in ribonucleoprotein complexes. *Nucleic Acids Res.* **37**, e140 (2009).

60. Nakayama, H., Yamauchi, Y., Taoka, M. & Isobe, T. Direct Identification of Human Cellular MicroRNAs by Nanoflow Liquid Chromatography–High-Resolution Tandem Mass Spectrometry and Database Searching. *Anal. Chem.* **87**, 2884–2891 (2015).

61. Nakayama, H. *et al.* Ariadne: a database search engine for identification and chemical analysis of RNA using tandem mass spectrometry data. *Nucleic Acids Res.* **37**, e47 (2009).

62. Zhang, Y., Wen, Z., Washburn, M. P. & Florens, L. Evaluating Chromatographic Approaches for the Quantitative Analysis of a Human Proteome on Orbitrap-Based Mass Spectrometry Systems. *J. Proteome Res.* **18**, 1857–1869 (2019).

63. Zhang, Y., Wen, Z., Washburn, M. P. & Florens, L. Effect of dynamic exclusion duration on spectral count based quantitative proteomics. *Anal. Chem.* **81**, 6317–6326 (2009).

64. Zhang, Y., Wen, Z., Washburn, M. P. & Florens, L. Improving proteomics mass accuracy by dynamic offline lock mass. *Anal. Chem.* **83**, 9344–9351 (2011).

65. McDonald, W. H. *et al.* MS1, MS2, and SQT-three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. *Rapid Commun. Mass Spectrom.* **18**, 2162–2168 (2004).

66. Xu, T. *et al.* ProLuCID: An improved SEQUEST-like algorithm with enhanced sensitivity and specificity. *J. Proteomics* **129**, 16–24 (2015).

67. Tabb, D. L., McDonald, W. H. & Yates, J. R., 3rd. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1**, 21–26 (2002).

68. Zhang, Y., Wen, Z., Washburn, M. P. & Florens, L. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal. Chem.* **82**, 2272–2281 (2010).

69. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

70. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

71. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).

72. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, (2018).

73. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).

74. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531–544 (2018).

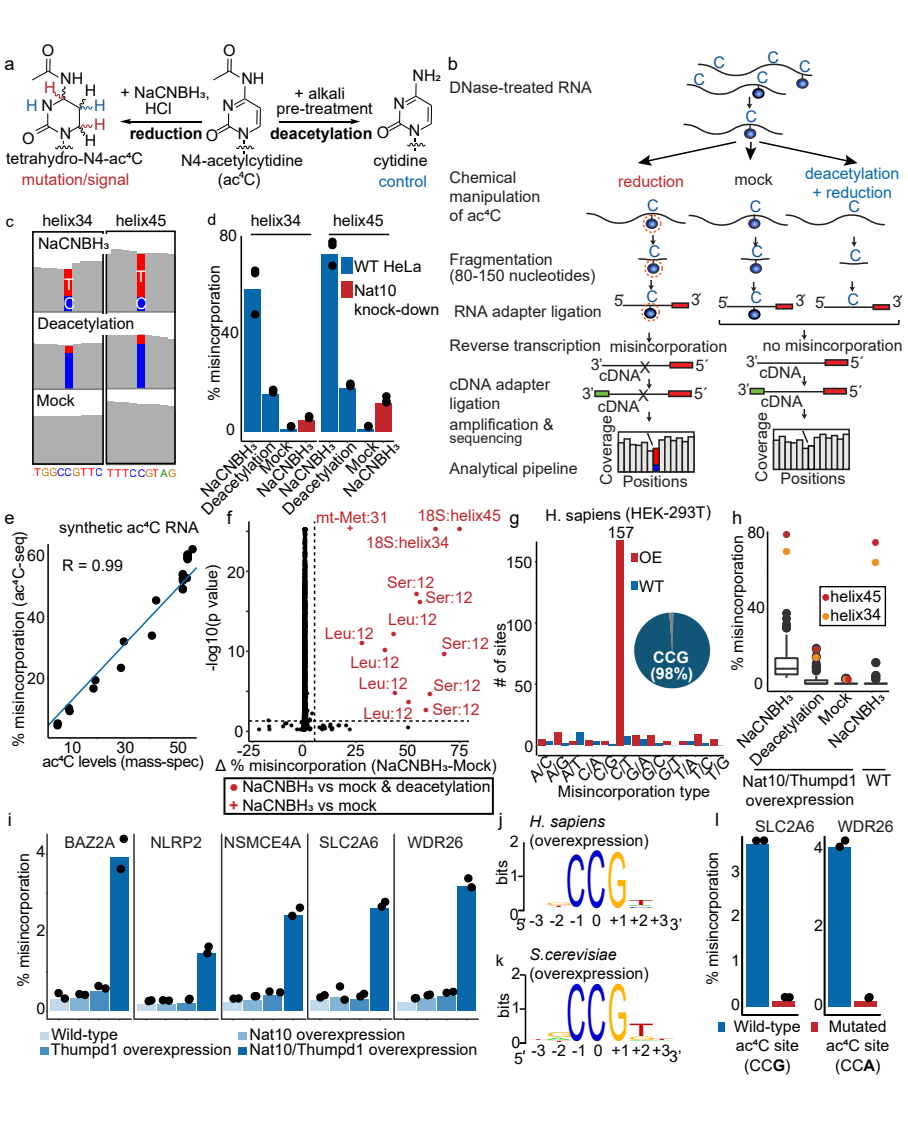
75. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).

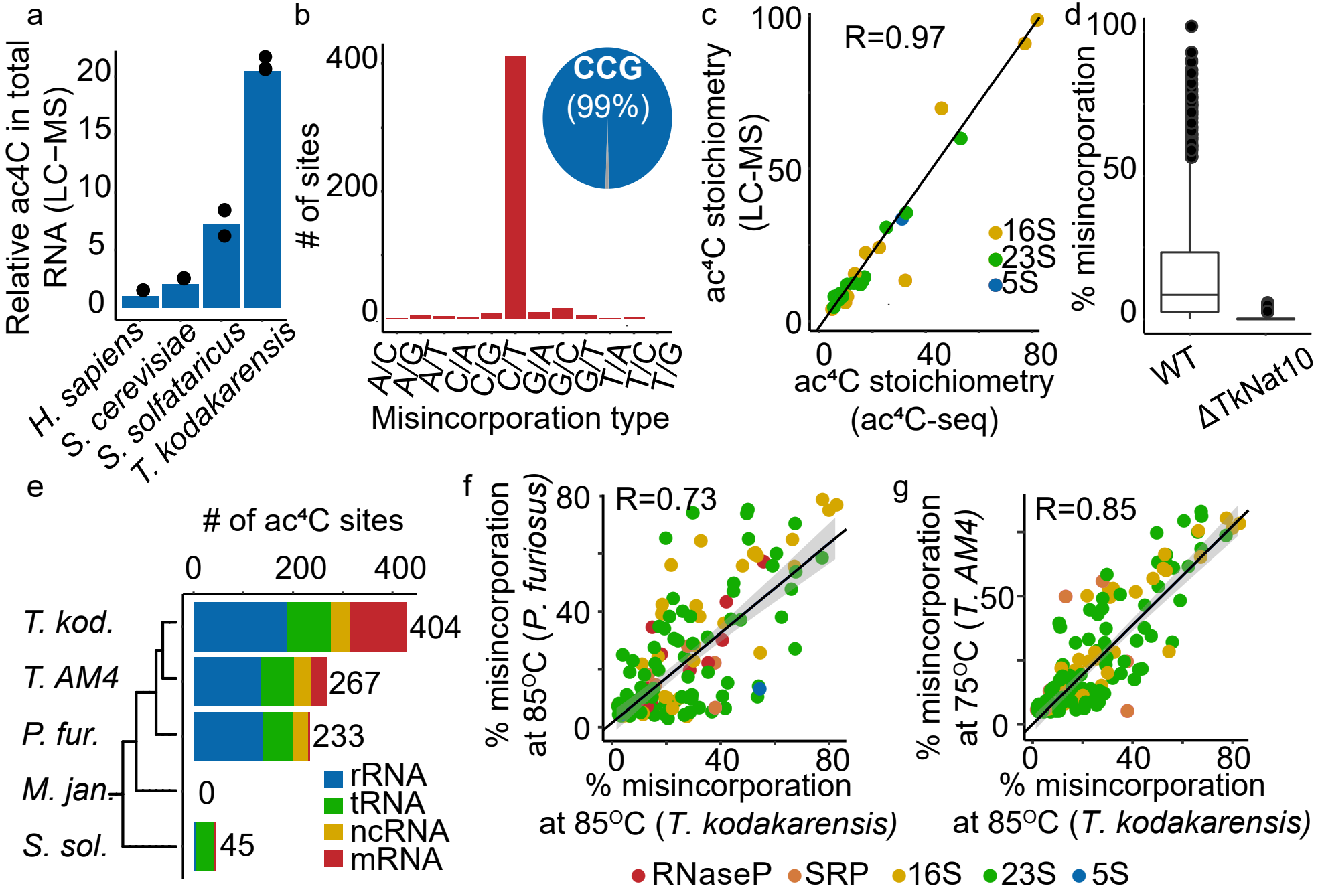
76. Pettersen, E. F., Goddard, T. D. & Huang, C. C. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of* (2004).

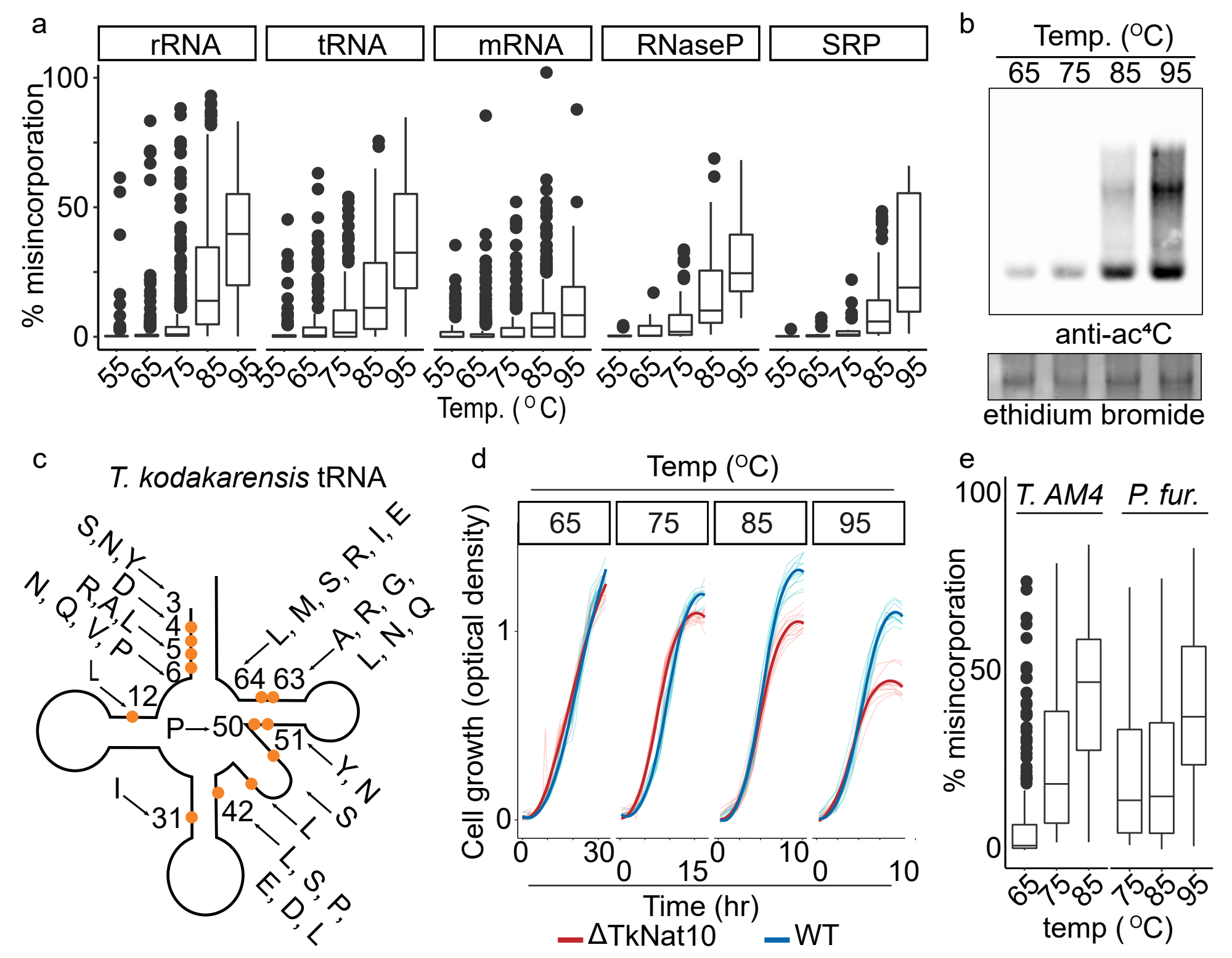
77. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

78. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080 (2009).

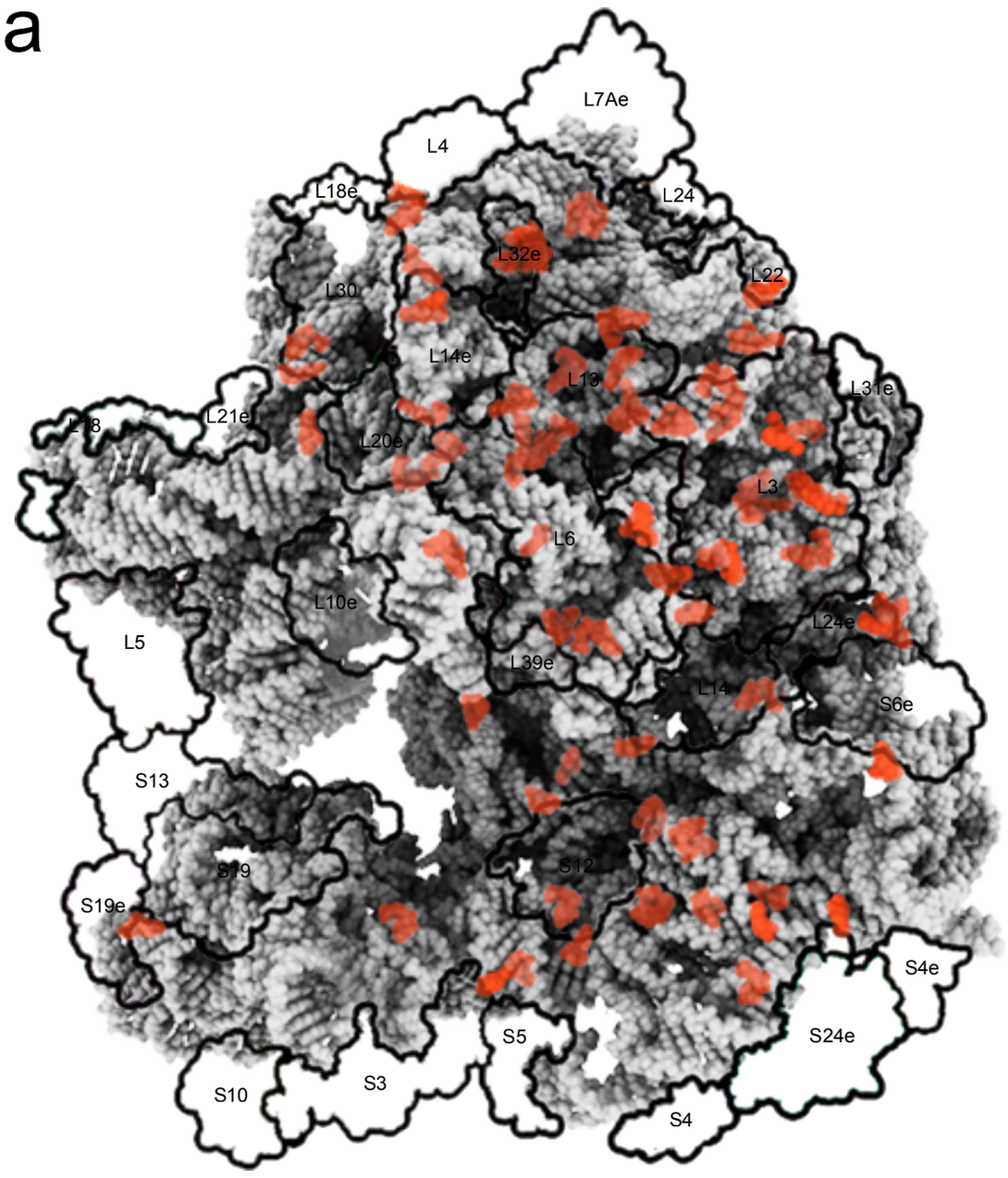
79. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).


a
tetrahydro-N4-ac4C
mutation/signal
+ NaCNBH3, HCl
reduction
N4-acetylcytidine (ac4C)
+ alkali pre-treatment
deacetylation
cytidine
control
b
DNase-treated RNA
Chemical manipulation of ac4C
reduction
mock
deacetylation + reduction
Fragmentation (80-150 nucleotides)
RNA adapter ligation
Reverse transcription
misincorporation
no misincorporation
cDNA adapter ligation
amplification & sequencing
Analytical pipeline
Coverage
Positions
c
helix34
helix45
NaCNBH3
Deacetylation
Mock
TGGCCGTTC
TTTCCGTAG
d
helix34
helix45
% misincorporation
WT HeLa
Nat10 knock-down
NaCNBH3
Deacetylation
Mock
e
synthetic ac4C RNA
R = 0.99
% misincorporation (ac4C-seq)
ac4C levels (mass-spec)
f
-log10(p value)
Δ % misincorporation (NaCNBH3-Mock)
mt-Met:31
18S:helix45
18S:helix34
Ser:12
Leu:12
NaCNBH3 vs mock & deacetylation
NaCNBH3 vs mock
g
H. sapiens (HEK-293T)
157
OE
WT
CCG (98%)
# of sites
Misincorporation type
h
% misincorporation
helix45
helix34
NaCNBH3
Deacetylation
Mock
i
BAZ2A
NLRP2
NSMCE4A
SLC2A6
WDR26
% misincorporation
Wild-type
Thumpd1 overexpression
Nat10 overexpression
Nat10/Thumpd1 overexpression
j
H. sapiens
(overexpression)
bits
k
S.cerevisiae
(overexpression)
l
Nat10/Thumpd1 overexpression
WT
SLC2A6
WDR26
% misincorporation
Wild-type ac4C site (CCG)
Mutated ac4C site (CCA)

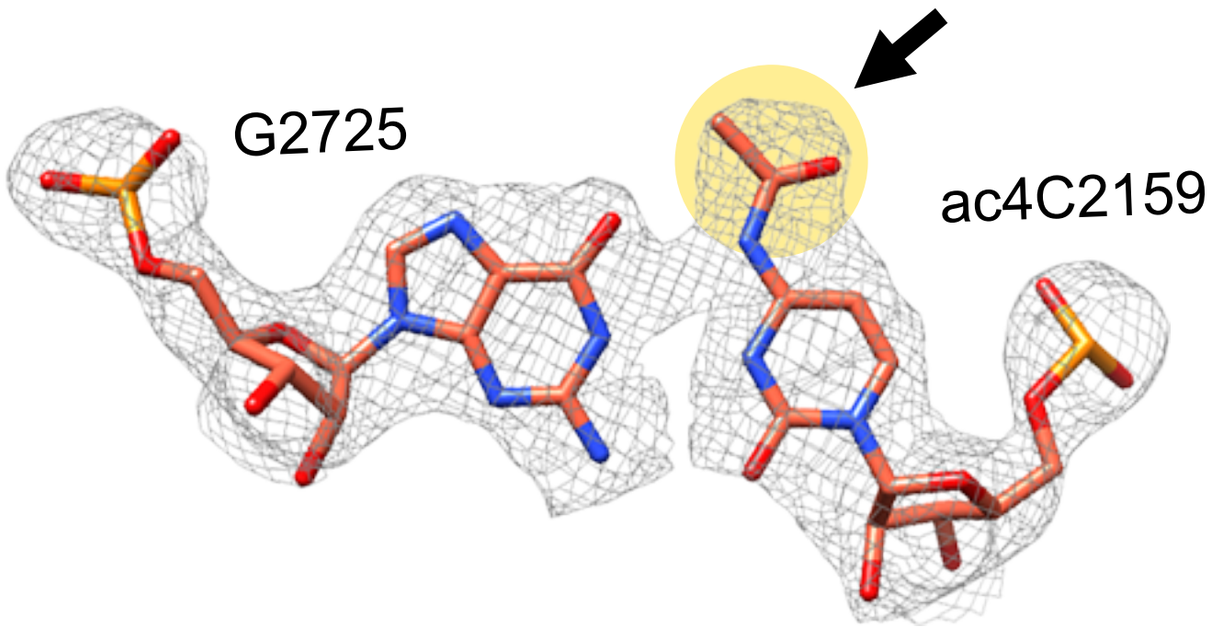

a
Relative ac4C in total RNA (LC–MS)
H. sapiens
S. cerevisiae
S. solfataricus
T. kodakarensis
b
# of sites
CCG (99%)
A/C
A/G
A/T
C/A
C/G
C/T
G/A
G/C
G/T
T/A
T/C
T/G
Misincorporation type
c
ac⁴C stoichiometry (LC-MS)
R=0.97
16S
23S
5S
ac⁴C stoichiometry (ac⁴C-seq)
d
% misincorporation
WT
ΔTkNat10
e
# of ac⁴C sites
T. kod.
T. AM4
P. fur.
M. jan.
S. sol.
404
267
233
0
45
rRNA
tRNA
ncRNA
mRNA
f
% misincorporation at 85°C (P. furiosus)
R=0.73
% misincorporation at 85°C (T. kodakarensis)
g
% misincorporation at 75°C (T. AM4)
R=0.85
% misincorporation at 85°C (T. kodakarensis)
RNaseP
SRP
16S
23S
5S


a
rRNA
tRNA
mRNA
RNaseP
SRP
% misincorporation
100
50
0
55 65 75 85 95
Temp. (°C)
b
Temp. (°C)
65 75 85 95
anti-ac⁴C
ethidium bromide
c
T. kodakarensis tRNA
S,N,Y
D
R,A,L
N, Q, V, P
L
3
4
5
6
12
64 63
L, M, S, R, I, E
A, R, G,
L, N, Q
P
50
51
Y, N
S
I
31
42
L
L, E, D, L
S, P,
d
Temp (°C)
65
75
85
95
Cell growth (optical density)
1
0
0 30 0 15 0 10 0 10
Time (hr)
ΔTkNat10
WT
e
T. AM4
P. fur.
100
50
0
% misincorporation
65 75 85 75 85 95
temp (°C)

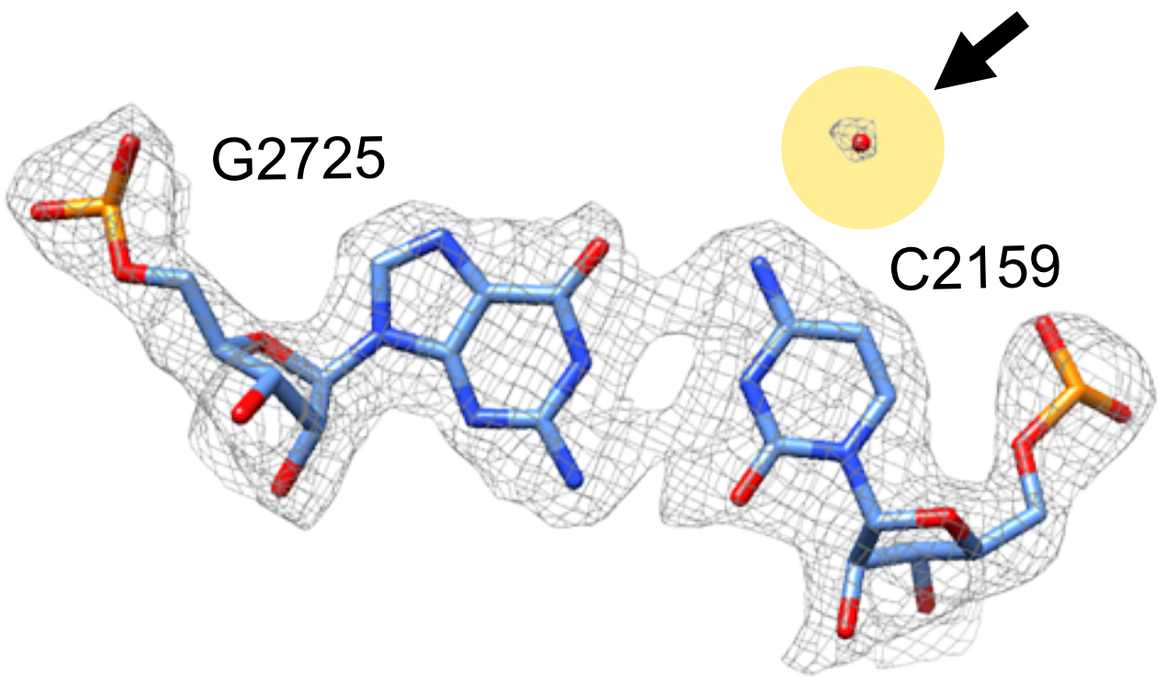
a


L7Ae
L4
L18e
L24
L32e
L30
L22
L14e
L13
L31e
L21e
L18
L20e
L3
L6
L10e
L5
L24e
L39e
L14
S6e
S13
S19
S12
S19e
S4e
S5
S24e
S10
S3
S4

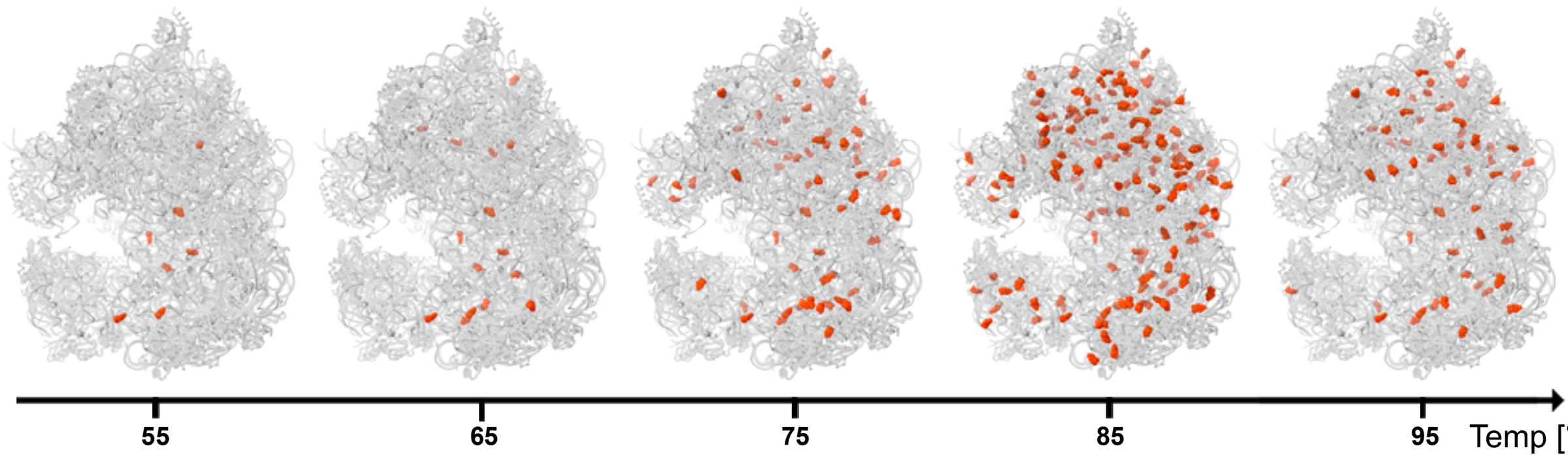
b


G2725
ac4C2159

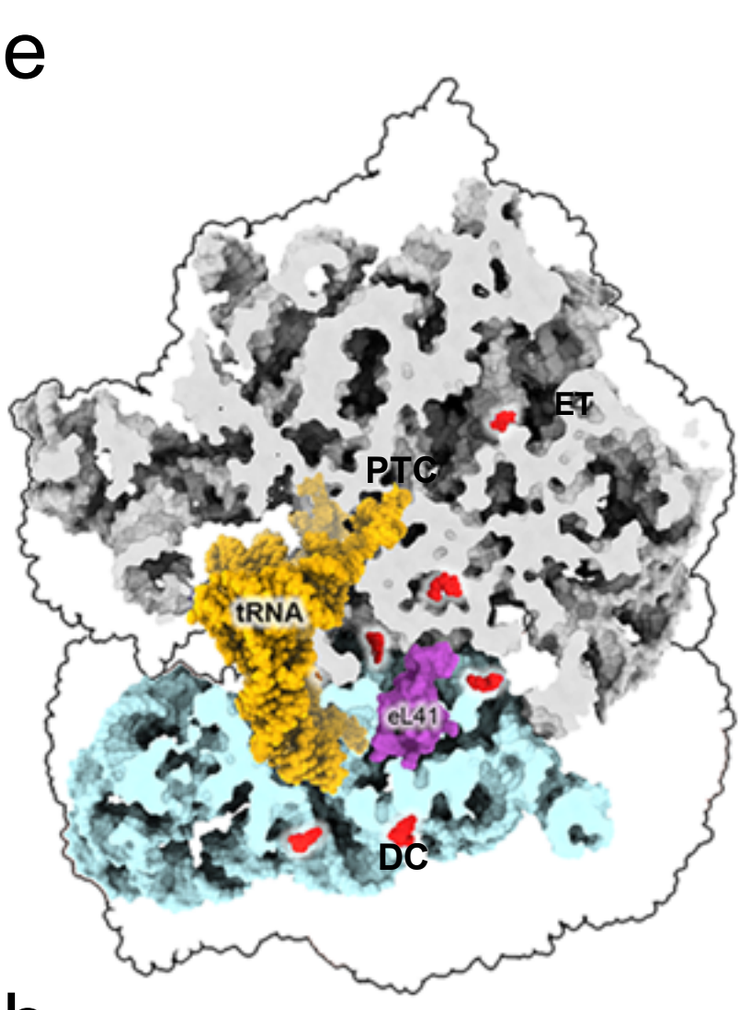
c


G2725
C2159

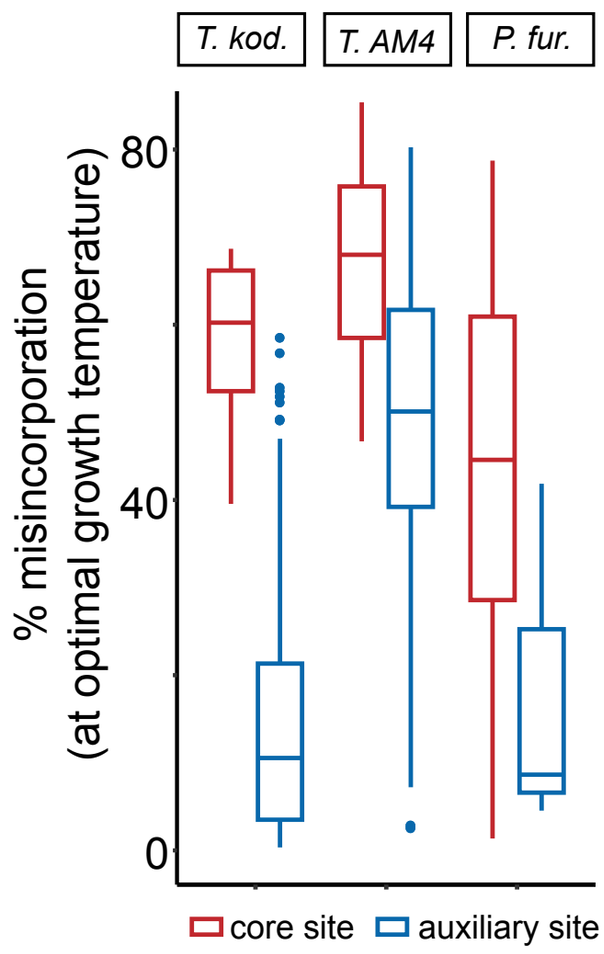
d


55
65
75
85
95
Temp [°C]

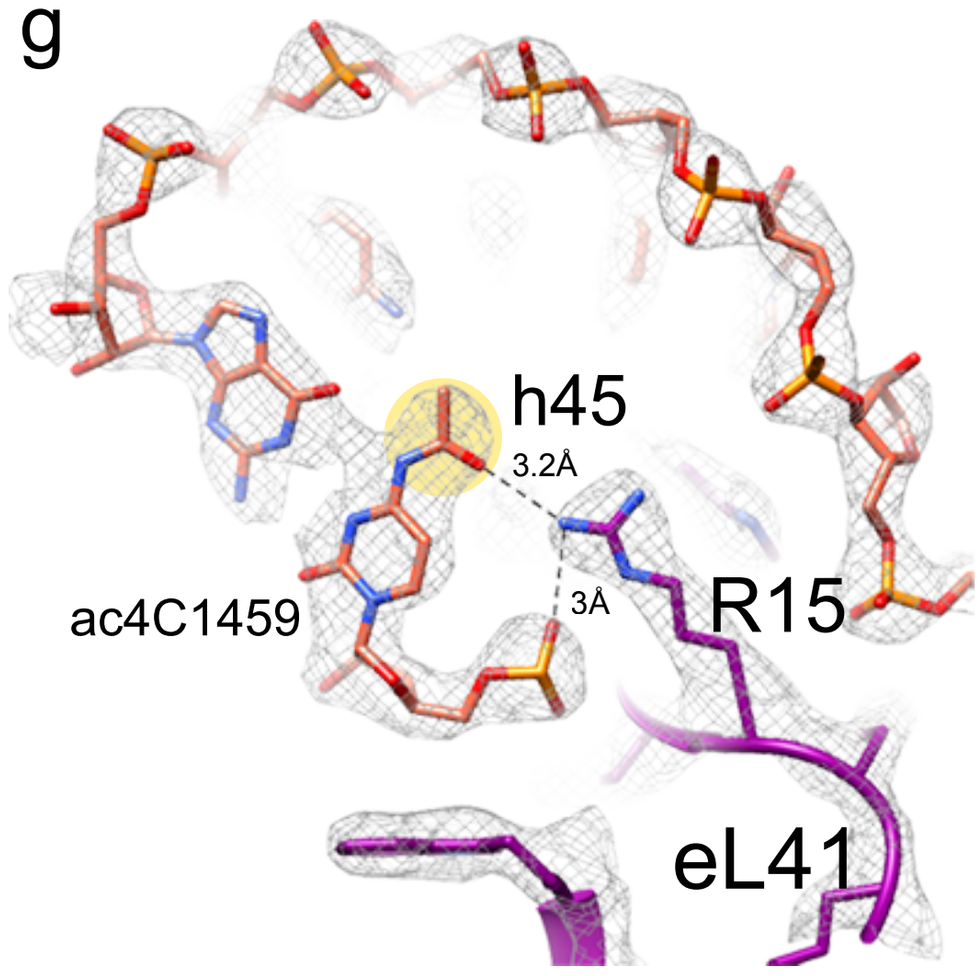
e


ET
PTC
tRNA
eL41
DC

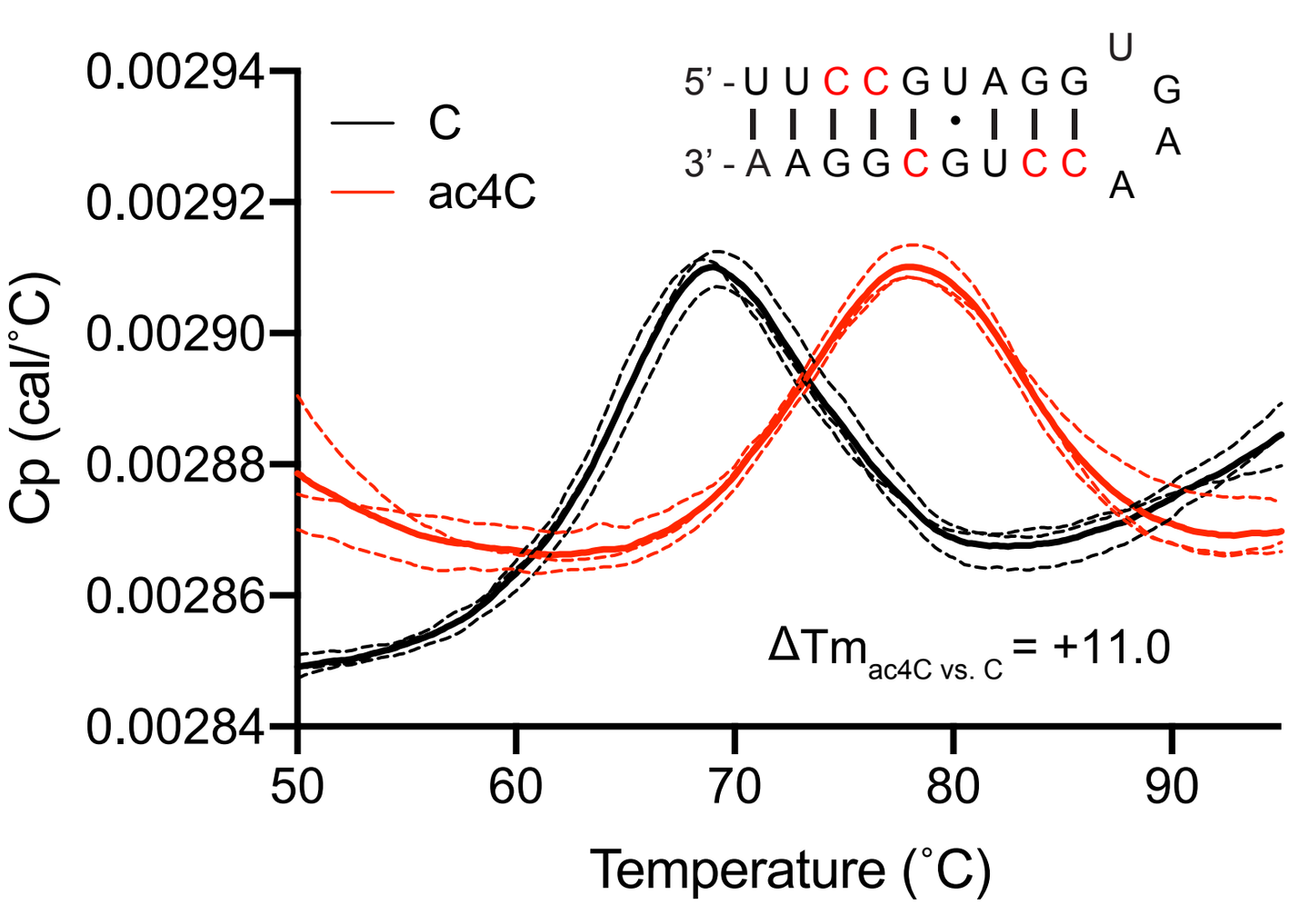
f

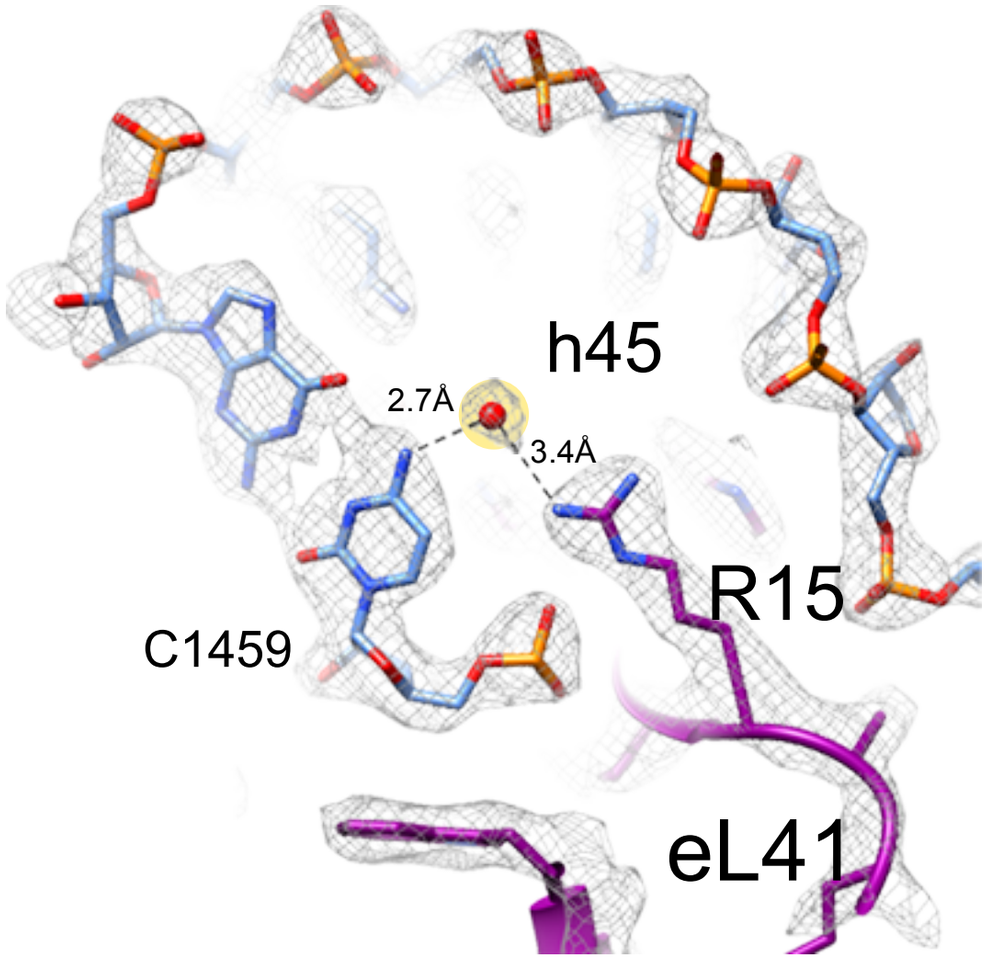

T. kod.
T. AM4
P. fur.
% misincorporation
(at optimal growth temperature)
80
40
0
core site
auxiliary site

g


h45
3.2Å
ac4C1459
3Å
R15
eL41

h


5’ - U U C C G U A G G U G
3’ - A A G G C G U C C A A
C
ac4C
Cp (cal/°C)
0.00294
0.00292
0.00290
0.00288
0.00286
0.00284
ΔTm ac4C vs. C = +11.0
50
60
70
80
90
Temperature (°C)


h45
2.7Å
3.4Å
C1459
R15
eL41