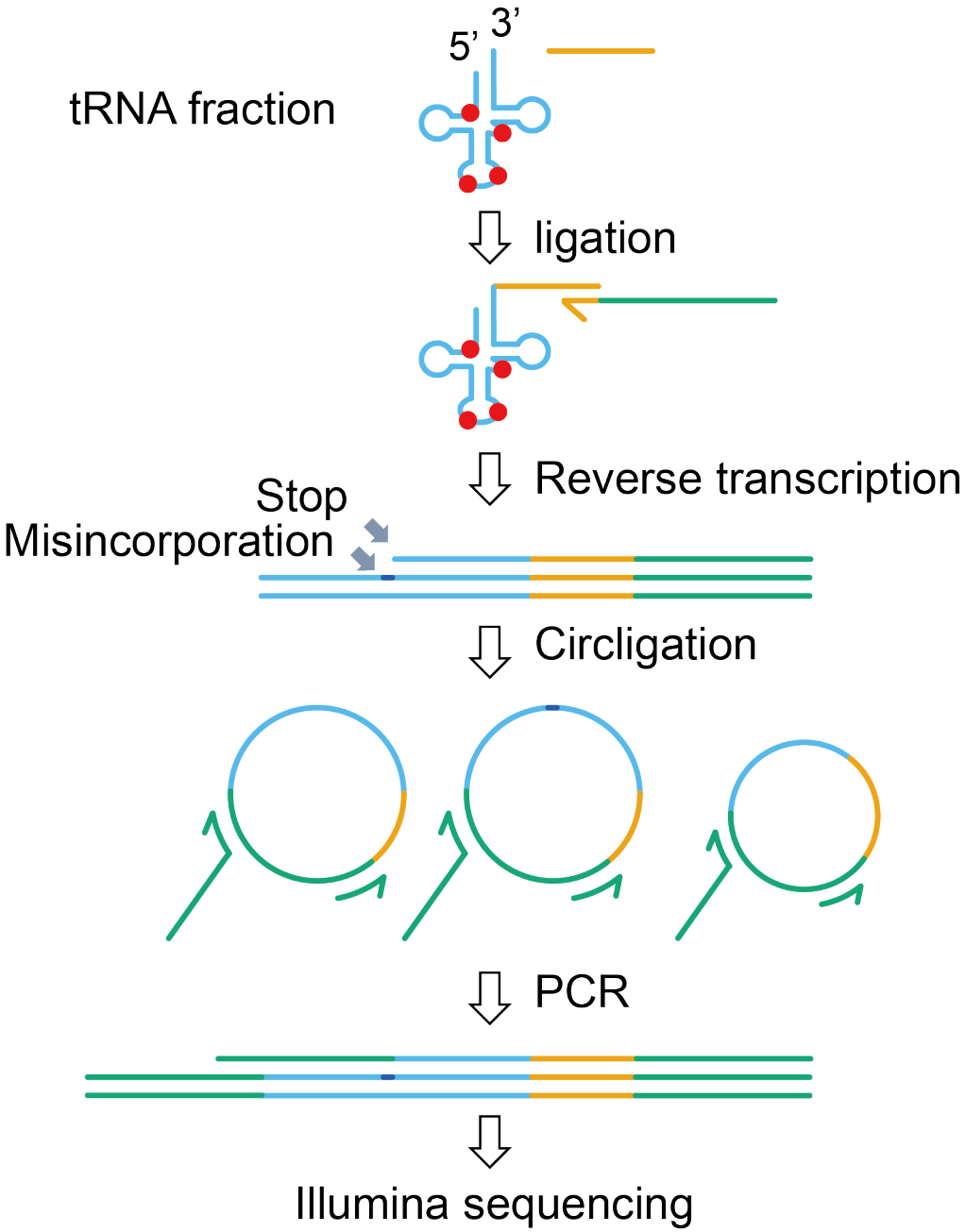
Supplementary Information

Comparative tRNA sequencing and RNA mass spectrometry for surveying tRNA modifications

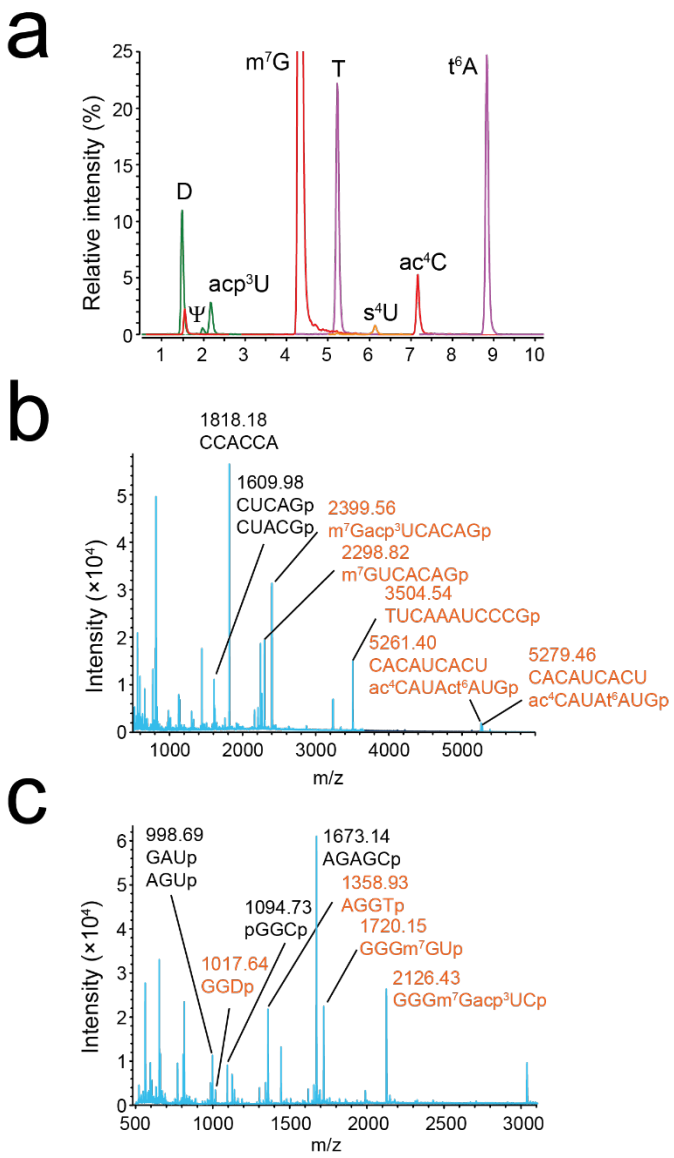
Satoshi Kimura[1,2,3], Peter C. Dedon[4,5] and Matthew K. Waldor[1,2,3]

[1]Division of Infectious Diseases, Brigham and Women's Hospital, Boston, Massachusetts, USA
[2]Department of Microbiology, Harvard Medical School, Boston, Massachusetts, USA
[3]Howard Hughes Medical Institute, Boston, Massachusetts, USA
[4]Department of Biological Engineering, Massachusetts Institution of Technology, Cambridge, Massachusetts, USA
[5]Singapore-MIT Alliance for Research and Technology Antimicrobial Resistance Interdisciplinary Research Group



Supplementary Figure 1. Scheme of tRNA-seq protocol.

tRNA was isolated via gel purification from total RNA samples. Following tRNA deacylation, linker DNA (yellow) was ligated to the 3’ end of tRNAs. Then, the TGIRT enzyme was used for reverse transcription (green and yellow primer). In this process, tRNA modifications (red circles) may be recorded in cDNAs as an interruption (stop) in reverse transcription and/or by incorporation of mismatched bases (misincorporation). cDNAs are then circularized using Circligase II, followed by PCR. Amplified cDNAs are then subjected to Illumina sequencing.

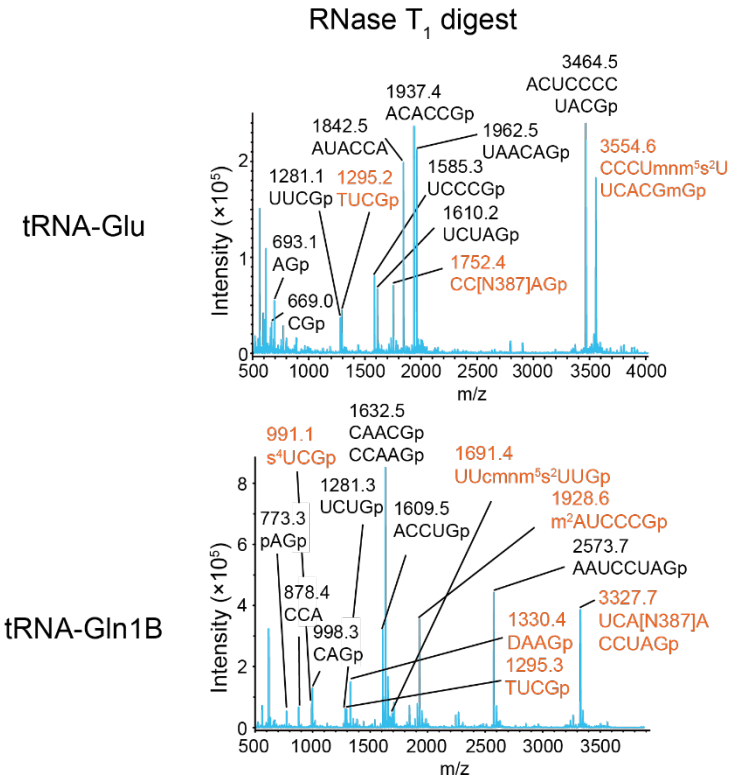


Supplementary Figure 2. RNA mass spectrometric analyses of *V. cholerae* tRNA-Met1 from stationary phase.

a, Nucleoside analysis detecting D, Ψ, acp^{3}U, m^{7}G, T, s^{4}U, ac^{4}C and t^{6}A. The peak heights between different nucleosides are not comparable. Representative data from two independent experiments with similar results is shown.

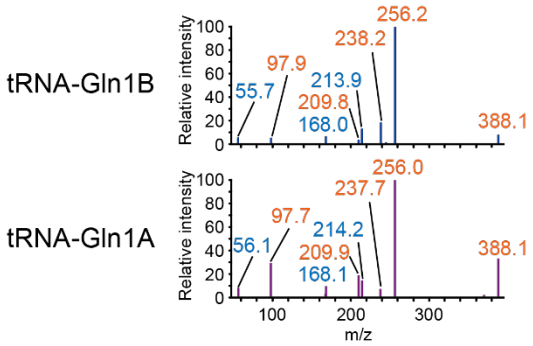
b, Fragment analysis of RNase T_1 digests. The fragments with or without modifications are shown in red and black, respectively. Measurement was conducted in the positive polarity mode. Representative data from two independent experiments with similar results is shown.

c, Fragment analysis of RNase A digests. The fragments with or without modifications are shown in red and black, respectively. Measurement was conducted in the positive polarity mode. Representative data from two independent experiments with similar results is shown.



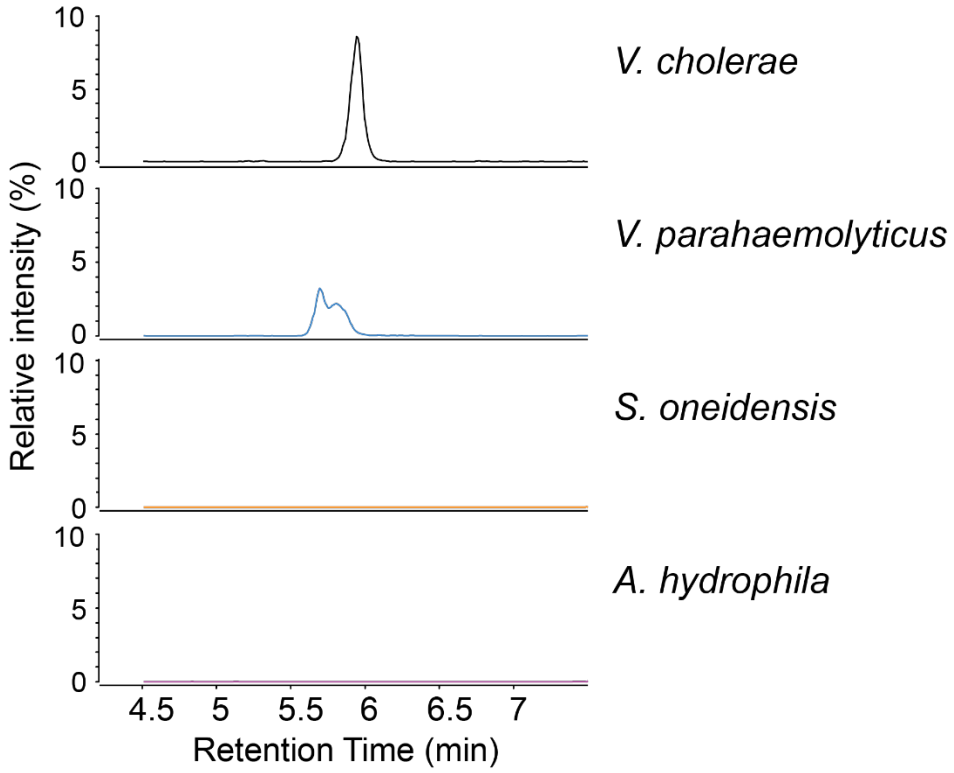
Supplementary Figure 3. Fragment analyses of RNase T_1 digests of tRNA-Glu (upper) and tRNA-Gln1B (lower).

The fragments with or without modifications are shown in red and black, respectively. Measurement was conducted in the positive polarity mode. Representative data from two independent experiments with similar results is shown.

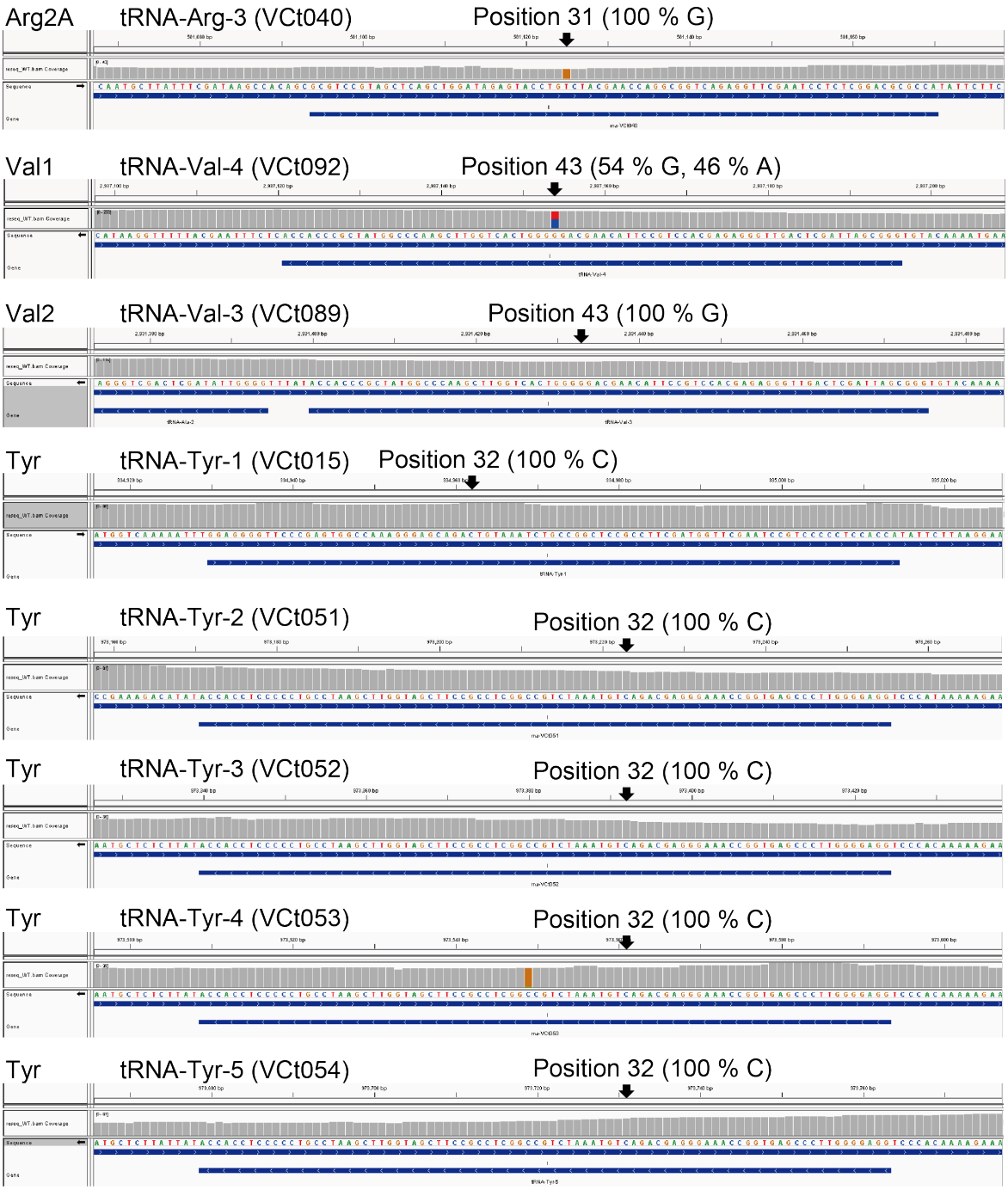


Supplementary Figure 4. MS/MS analyses of N387 in tRNA-Gln1B (upper) and tRNA-Gln1A (lower).

Fragment ions observed in acp^{3}U are colored in blue and N387 specific fragment ions are colored in red. Representative data from at least two independent MS/MS scans with similar results is shown.

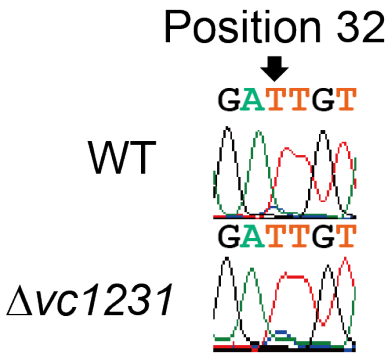


Supplementary Figure 5. Presence of acacp^{3}U in *V. cholerae* related organisms.
Nucleoside analyses of total tRNA fraction from the indicated organisms. Mass chromatograms detecting acacp^{3}U are shown. This experiment was done once.



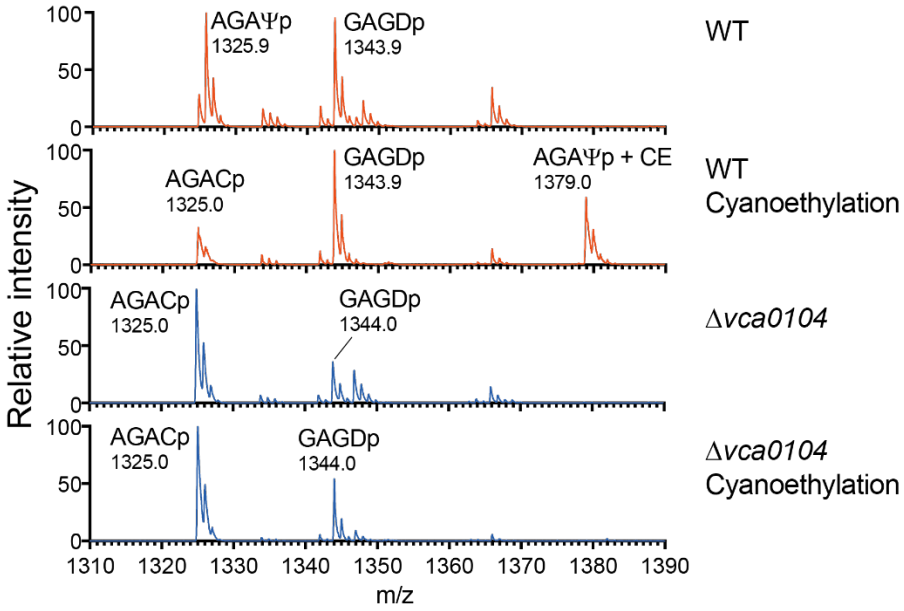
Supplementary Figure 6. Genomic sequences of *V. cholerae* tRNAs

V. cholerae genomic tRNA sequences are retrieved from whole genome sequencing data[1]. In contrast to the N16961 reference genome[2], in C6706, tRNA-Arg2 has a G (rather than a T) at position 31. Also, one of the two tRNA-Val1 genes has 46% A at position 43 (rather the G), while the other tRNA-Val1 is 100% G at this position, suggesting that a SNP is present in one of the tRNA-Val1 genes. All the tRNA-Tyr genes have 100 % C at position 32.



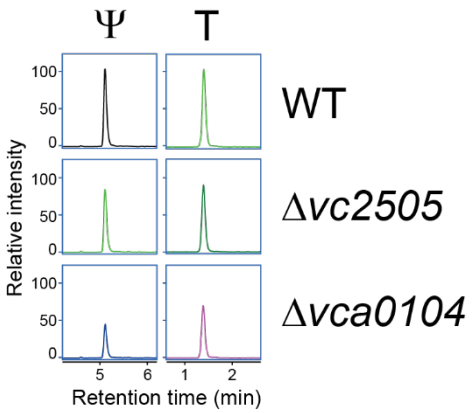
Supplementary Figure 7. Sanger sequencing of cDNA of tRNA-Tyr from the WT and Δ*vc1231* strains

Position 32 is indicated by the arrow.



Supplementary Figure 8. Fragment analysis of a portion of tRNA-Tyr from the WT and Δ*vca0104* strains.

MALDI analysis of RNase A digests of protected portion of tRNA-Tyr from the WT and Δ*vca0104* strains. In the second and fourth panels, RNAs were incubated with acrylonitrile by which pseudouridine is specifically modified with cyanoethylation (CE), which results in increase of the mass by 53 Da. *m/z* values of detected peaks with assigned fragment sequences are shown. The MALDI analyses were conducted in negative polarity mode. Representative data from two independent experiments with similar results is shown.



Supplementary Figure 9. Nucleoside analysis of tRNA-Tyr from the WT, Δ*vc2505*, and Δ*vca0104*. The height of the highest peak across samples is depicted at 100. The detected modified nucleosides (Ψ and T) are shown above panels and strains are shown on the right of the panels. Representative data from two independent experiments with similar results is shown.

Supplementary Table 1. Prediction of E. coli tRNA modifications using tRNA-seq data

Modification	Misincorporation				Termination				Total	tRNA species
	Frequency (%)	Position	Fraction >5%*[2]	Prediction*[3]	Frequency (%)	Position*[4]	Fraction >5%*[5]	Prediction*[7]	Prediction*[8]	
s^4U	4.5 - 50	8,9	28/29	+	0 - 42.9	10,11	9/29	-	+	Ala1B, Arg2, Asn, Asp, Cys, Gln2, Gln1, Gly1, His, Ile2, fMet1, fMet2, Leu5, Met, Phe, Pro1, Ser3, Ser5, Ser1, Trp, Tyr1, Tyr2, Val2A, Val2B, Val1
DD*[1]	0 - 2.4	16-17, 20-20A	0/20	-	0 - 51.5	17*[4], 20A*[4]	18/20	+	+	Gly3, His, Leu1A, Leu2, Lys, Met, Thr1, Thr2, Thr3, Trp, Val2A, Val2B, Asp, Ile1, Ile2, Ser1, Ser5
Gm	0 - 18.8	18	1/13	-	0 - 19.3	20	7/13	+	+	Gln2, Gln1, Ile2, Leu5, Leu1A, Leu2, Leu3, Met, Ser2, Ser5, Ser1, Tyr1, Tyr2
s^2C	1.7 - 7.3	32	2/4	+	1.5 - 26.4	34	3/4	+	+	Arg2, Arg3, Arg4, Ser3
Um	0.3 - 0.5	32	0/4	-	0.7 - 10.7	34	2/4	+	+	Pro1, Pro3, Gln2, Gln1
Q	0.8 - 11.6	34	1/4	-	4.5 - 62.9	36	3/4	+	+	Tyr1, Tyr2, His, Asn
Glu-Q	12.1	34	1/1	+	14.9	36	1/1	+	+	Asp
mnm^5s^2U	4.6, 4.9	34	0/2	-	15.1, 19.9	36	2/2	+	+	Glu, Lys
$cmnm^5s^2U$	6.5	34	1/1	+	5.3	36	1/1	+	+	Gln1
$cmnm^5Um$	11	34	1/1	+	62.6	36	1/1	+	+	Leu4
I	99.9	34	1/1	+	0.5	36	0/1	-	+	Arg2
k^2C	91.7	34	1/1	+	96.1	35*[4]	1/1	+	+	Ile2
m^1G	47.3 - 94.8	37	8/8	+	11.1 - 62.3	39	8/8	+	+	Leu1A, Leu1B, Leu2, Leu3, Arg3, Pro1, Pro2, Pro3
ms^2i^6A	5.6 - 56.2	37	9/9	+	71.8 - 91	39	9/9	+	+	Leu5, Phe, Leu4, Trp, Cys, Ser2, Ser1, Tyr1, Tyr2
m^6t^6A	1.9, 2.5	37	0/2	-	12.7, 13.1	37*[4]	2/2	+	+	Thr1, Thr3
acp^3U	28.8 - 67.5	47	8/8	+	23.7 - 55.9	48*[4]	8/8	+	+	Phe, Val2A, Val2B, Arg2, Met, Ile2, Ile1, Lys
Ψ	0 - 0.7	13,32,38, 39,40, 55,65	0/68	-	0 - 8.2	15,34,40, 41,42, 57,67	2/68	-	-	Ala1B, Ala2, Arg2, Arg3, Arg4, Asn, Asp, Cys, fMet1, fMet2, Gln1, Gln2, Glu, Gly1, Gly2, Gly3, His, Ile1, Ile2, Leu1A, Leu2, Leu4, Leu5, Lys, Met, Phe, Pro1, Ser1, Ser2, Ser3, Ser5, Thr1, Thr2, Thr3, Trp, Tyr1, Tyr2, Val1,
D"[1]	0 - 0.9	16,17, 20,20A	0/34	-	0 - 23.6	18,19, 21,22	3/34	-	-	Leu5, Leu4, Phe, Trp, Gly1, Gly3, Leu1A, Leu2, Pro1, Gln2, Gln1, His, fMet1, fMet2, Thr2, Thr1, Thr3, Lys, Asn,Val1, Ala2, Ala1B, Arg3, Arg2, Ile1,Leu5, Asp,Ser2, Ser3
Cm	0 - 0.3	32	0/5	-	0.1 - 10.7	34	1/5	-	-	Trp, Ser1, fMet1, fMet2, Thr4
ac^4C	1.9	34	0/1	-	0.6	36	0/1	-	-	Met
cmo^5U	0.3	34	0/2	-	0.9, 1.4	36	0/2	-	-	Val1, Leu3
$mcmo^5U$	0.1 - 0.5	34	0/4	-	1.2 - 8.1	36	1/4	-	-	Ser1, Ala1B, Pro3, Thr4
mnm^5U	0.2	34	0/2	-	0.3, 0.9	36	0/2	-	-	Gly2, Arg4
m^2A	0.2 - 1.5	37	0/6	-	0.2 - 0.9	39	0/6	-	-	Asp, Glu, Gln2, Gln1, His, Arg2
m^6A	0.2	37	0/1	-	0.2	39	0/1	-	-	Val1
ct^6A	0.3 - 2.9	37	0/8	-	0.4 - 5.4	39	1/8	-	-	Met, Ile1, Ile2, Ser3, Arg4, Lys, Asn, Thr2
m^7G	0.1 - 2.5	46	0/22	-	0.3 - 27.7	48	3/14*[6]	-	-	Ala2, Ala1B, Arg3, Arg2, Asn, Asp, Gly3, His, Ile1, Ile2, fMet1, Lys, Met, Phe, Pro1, Thr3, Thr1, Thr2, Trp, Val2B, Val2A, Val1
T	0 - 0.6	54	0/40	-	0 - 0.8	56	0/40	-	-	Ala1B, Ala2, Arg2, Arg3, Arg4, Asn, Asp, Cys, fMet1, fMet2, Gln1, Gln2, Glu, Gly1, Gly2, Gly3, His, Ile1, Ile2, Leu1A, Leu2, Leu4, Leu5, Lys, Met, Phe, Pro1, Ser1, Ser2, Ser3, Ser5, Thr1, Thr2, Thr3, Trp, Tyr1, Tyr2, Val1,

*1. Note that single dihydrouridine (D) and tandem D lead to distinct RT signatures.

*2. The denominators used for these calculations are the number of modified sites established in the literature[3-10] and/or the database[2] (Supplementary Data 1). The numerators are the number of known modified sites where the misincorporation frequency was >5 % in tRNA-seq data. For example, 29 sites are reported to have s^4U and the misincorporation frequency was > 5 % at 28 sites.

*3. Modifications were considered to be predictable (+) when the misincorporation fraction (previous column) was ≥ 50%.

*4. Positions of termination used for calculations were set empirically. Modifications led to a range of sites of termination. Most modifications led to the most pronounced termination 2 nucleotides downstream (3' side) from the modified site. For example, m^1G or ms^2i^6A at position 37 appear to lead to the termination of RT at position 39. However, at other sites, the highest termination occurs 1 nucleotide downstream from the modification; e.g., acp^3U and k^2C had peak termination at positions 48 and 35, respectively and tandem DD appears to induce termination at the latter D. In the case of m^6t^6A, termination appears at the same position as the modified site (position 37). The variation in the position of termination may be accounted for by variation in the mechanisms of inhibition of reverse transcription, e.g., based on the structure of modifications.

*5. The denominators used for these calculations are the number of modified sites established in the literature[3-10] and/or the database[2] (Supplementary Data 1). The numerators are the number of known modified sites where the termination frequency was >5 % in tRNA-seq data. For example, 4 sites are reported to have s^2C and the termination frequency was > 5 % at 3 sites.

*6. Eight out of 22 tRNA species have both m^7G and acp^3U at position 46 and 47, respectively. Since acp^3U has a strong effect on termination of RT, these sites were not considered in the assessment of m^7G's effect on termination.

*7. Modifications were considered to be predictable (+) when the termination fraction (previous column) was ≥ 50%.

*8. A modification is considered predictable when the modification is predictable by misincorporation and/or termination.

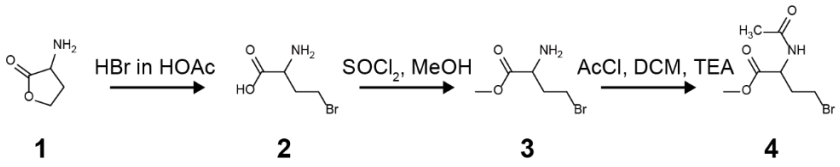
Supplementary Table 2. Strain list

Strains	Relevant genotype/description	Reference/source
Vibrio cholerae	C6706, wild-type El Tor clinical isolate (*Sm*^{R})	11
V. cholerae Δ*thiI*	C6706 Δ*vc0894*	1
V. cholerae Δ*miaA*	C6706 Δ*vc0346*	1
V. cholerae Δ*ttcA*	C6706 Δ*vc1432*	This work
V. cholerae Δ*trmK*	C6706 Δ*vca0634*	This work
V. cholerae Δ*acpA*	C6706 Δ*vc0317*	This work
V. cholerae Δ*vc1231*	C6706 Δ*vc1231*	This work
V. cholerae Δ*vc2505*	C6706 Δ*vc2505*	This work
V. cholerae Δ*trcP*	C6706 Δ*vca0104*	This work
V. cholerae vca0825::*Tn*	C6706 *vca0825*::*Tn*	12
V. cholerae vc1341::*Tn*	C6706 *vc1341*::*Tn*	12
V. cholerae vca0043::*Tn*	C6706 *vc0043*::*Tn*	12
V. cholerae vc0317::*Tn*	C6706 *vc0317*::*Tn*	12
V. cholerae vc0884::*Tn*	C6706 *vc0884*::*Tn*	12
Escherichia coli MG1655	wild-type MG1655	
Escherichia coli DH5α λ*pir*	Cloning strain	
Escherichia coli SM10 λ*pir*	Conjugation donor	
Vibrio parahaemolyticus	RIMD 2210633	13
Aeromonas hydrophila	ATCC7966	ATCC
Shewanella oneidensis	MR-1 ATCC 700550	ATCC

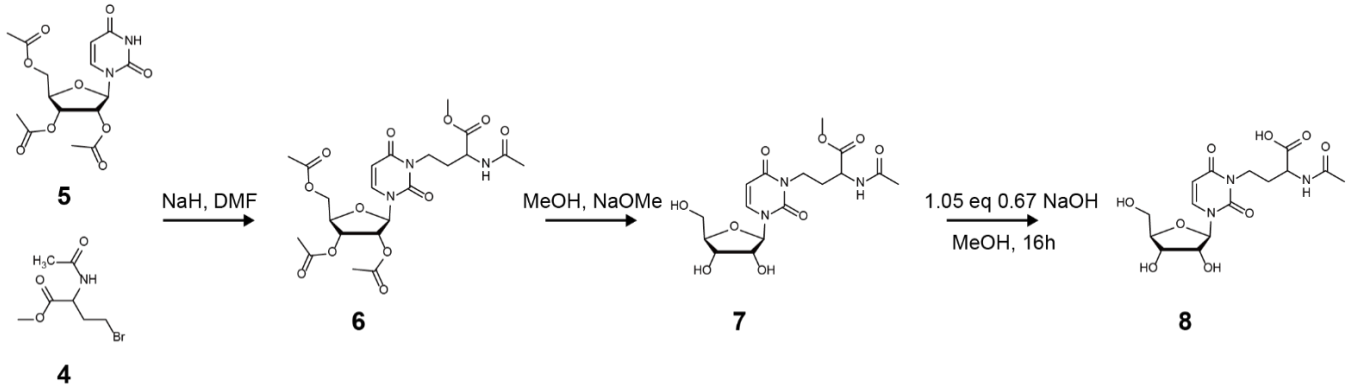
Synthetic procedures

Synthesis of acacp³U

3-(3-acetamido-3-carboxypropyl)-uridine **8** was purchased from Carbosynth LLC (Order number: NA171964). In brief, ring-opening of α-amino-γ-butyrolactone **1** by HBr yielded alkyl bromide **2** which was then methyl-protected and acetylated to compound **4** (Scheme 1). This compound **4** was conjugated with 2',3',5'-tris-*O*-acetyl uridine **5** to obtain **6** as previously described[14] (Scheme 2). Global deprotection of carboxylic and hydroxyl groups of **6** provided **8** (Scheme 2)**.**



Scheme 1



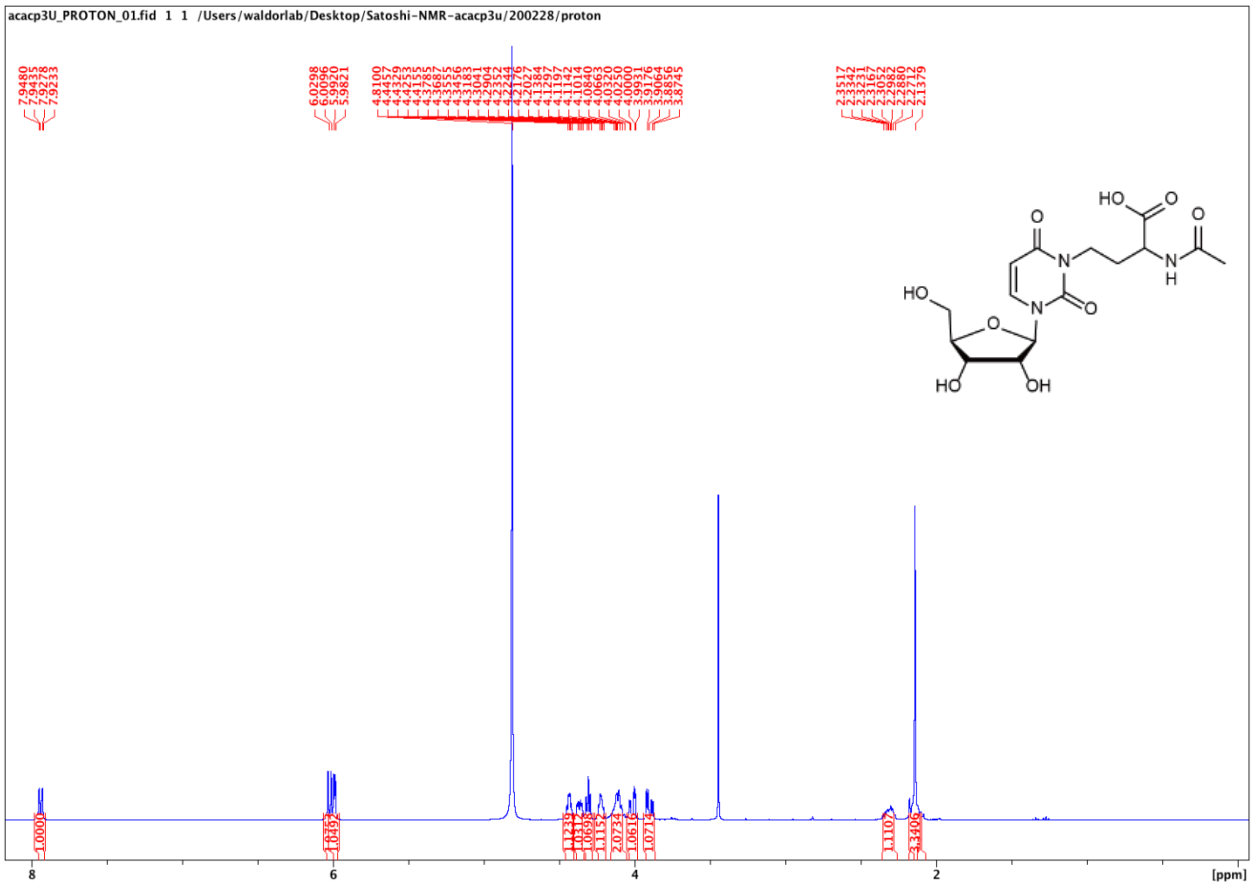
Scheme 2

Compound 8 (acacp³U)

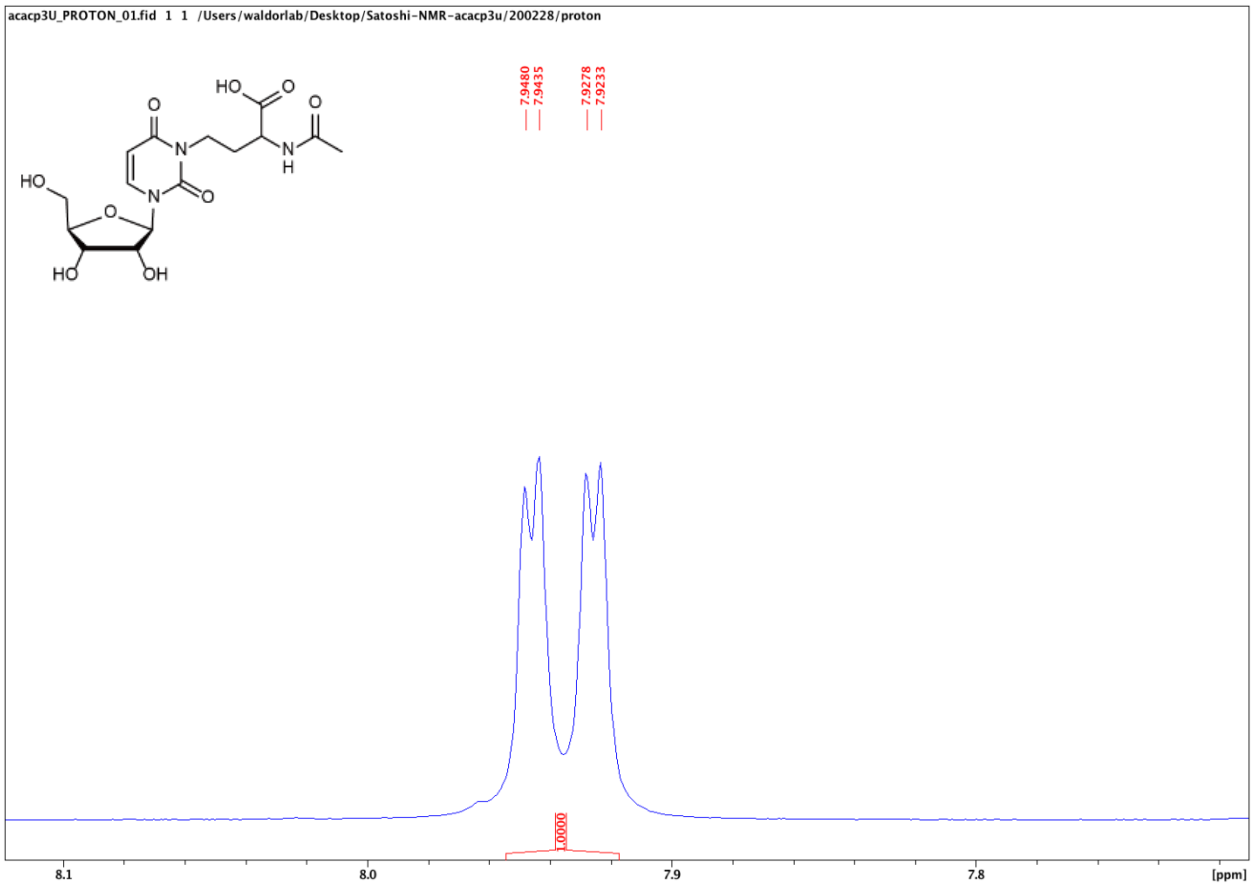
1H NMR (400 MHz, D2O): δ 7.94 (dd, J = 8.1, 1.8 Hz, 1H), 6.02 (d, J = 8.1 Hz, 1H), 5.99 (d, J = 4.0 Hz, 1H), 4.43 (m, 1H), 4.36 (dd, J = 9.5, 3.9 Hz, 1H), 4.30 (t, J = 5.6 Hz, 1H), 4.22 (m, 1H), 4.11 (m, 2H), 4.01 (dd, J = 12.8, 2.8 Hz, 1H), 3.90 (dd, J = 12.7, 4.5 Hz, 1H), 2.31 (m, 1H), 2.14 (s, 3H), 2.12 (m, 1H).

13C NMR (100 MHz, D2O): δ 176.6, 173.9, 165.0, 151.7, 139.8, 101.6, 90.4, 84.0, 73.7, 69.3, 60.7, 52.0, 38.3, 28.4, 21.9.

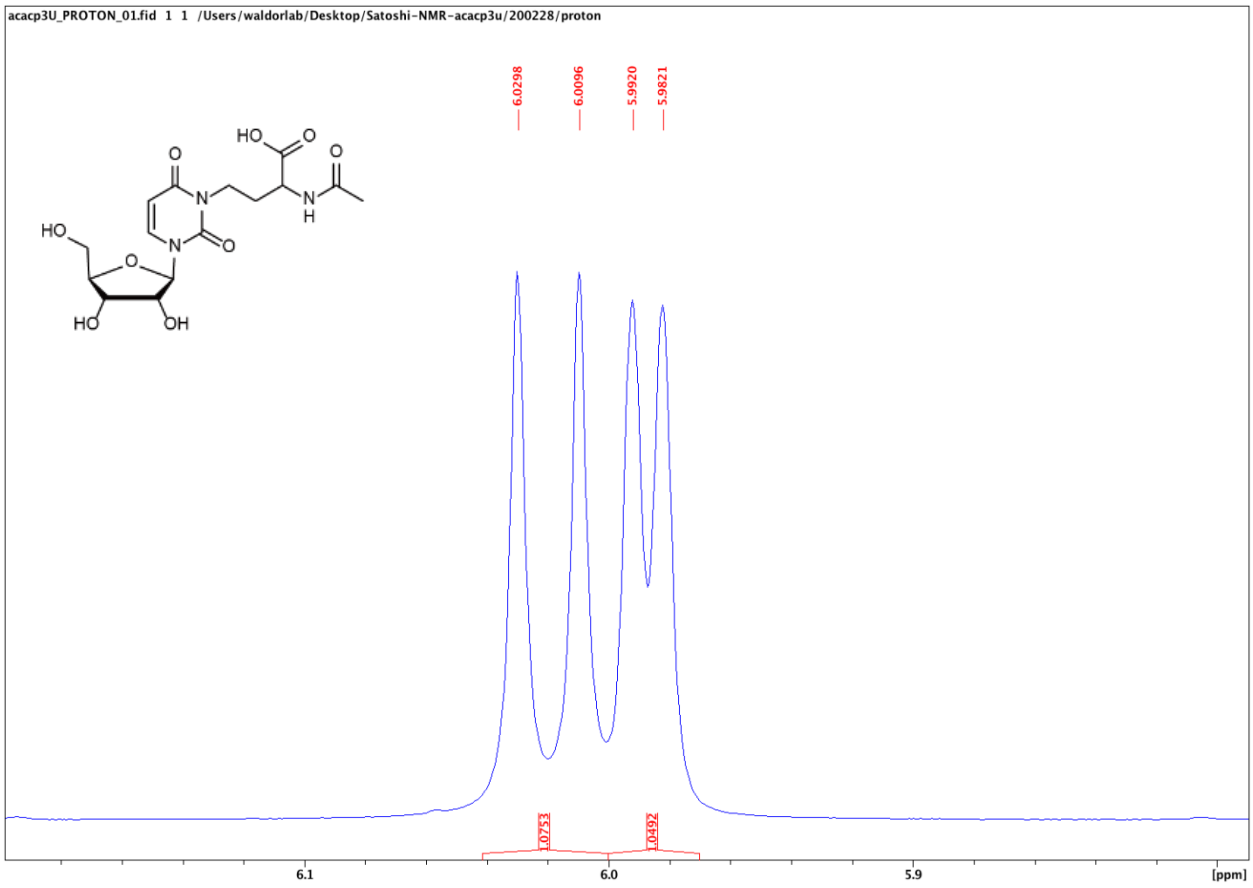
HRMS (m/z): calcd. For C15H21N3O9, 387.1278; found [M+H⁺] 388.1353, [M+Na⁺] 410.1167, [M+K⁺] 426.0911.



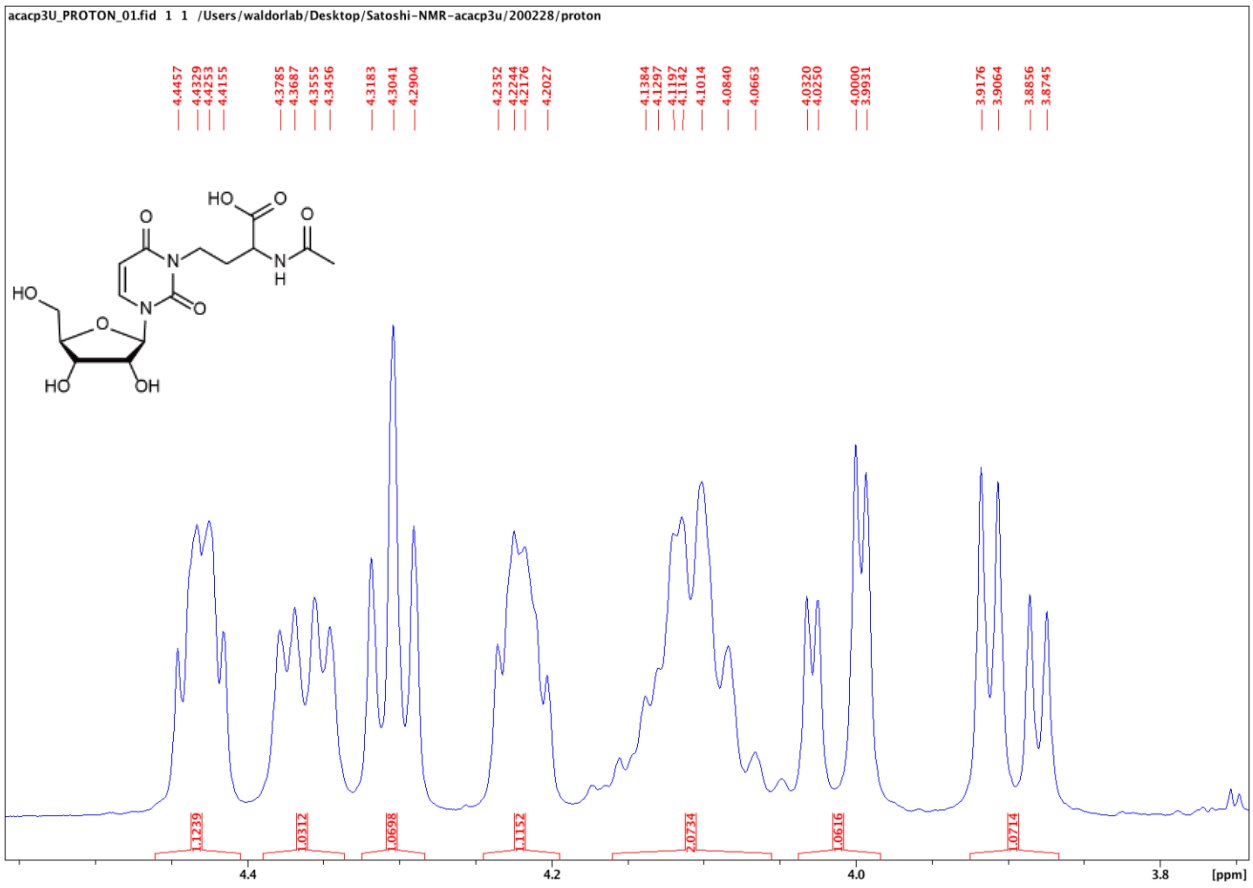
acacp^{3}U 1H NMR



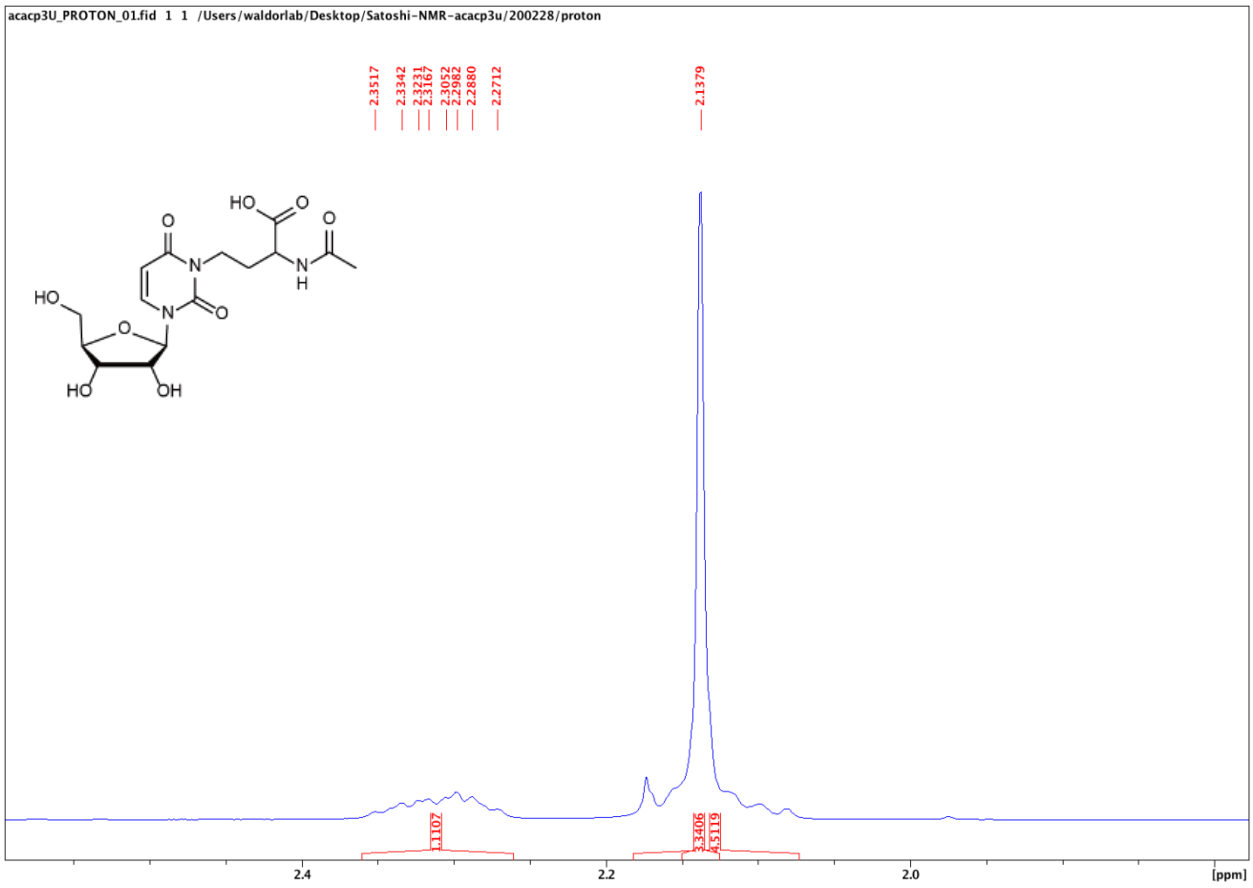
acacp^{3}U 1H NMR



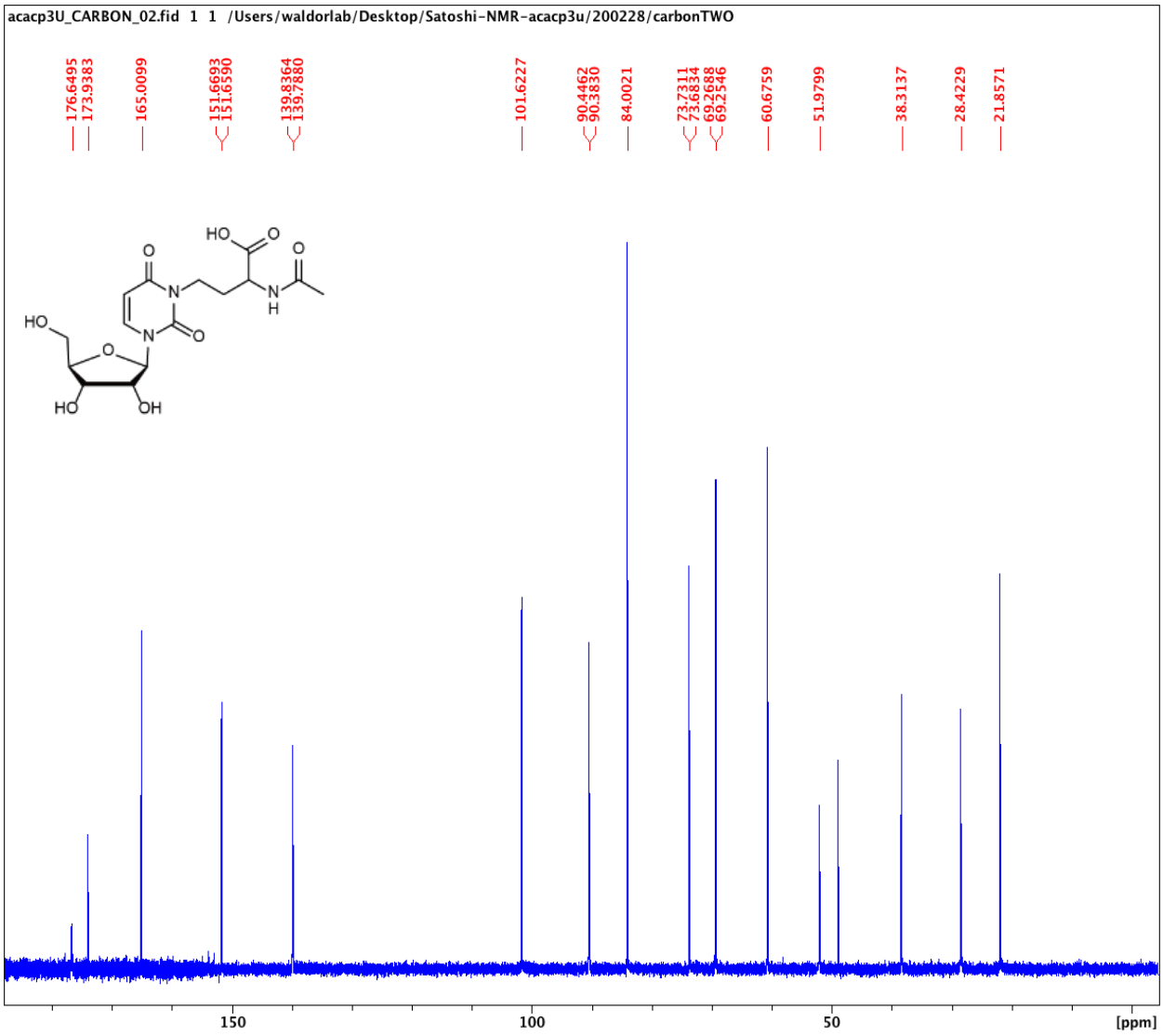
acacp^{3}U 1H NMR



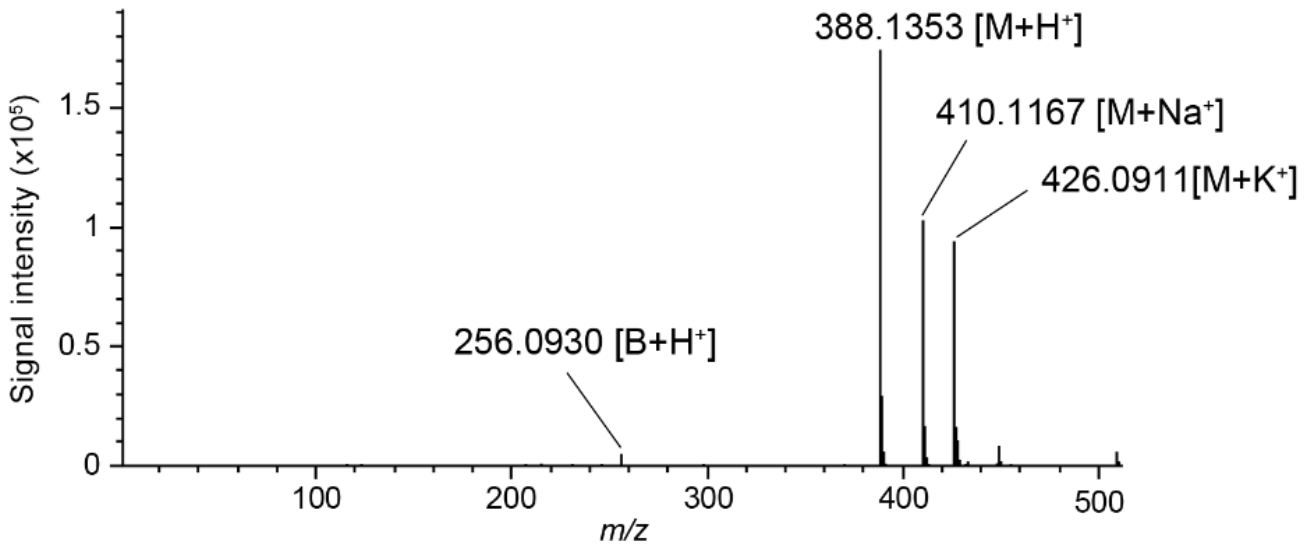
acacp[3]U 1H NMR



acacp^{3}U 1H NMR



acacp[3]U 13C NMR



High resolution Mass spectrometry, acacp^{3}U

Supplementary Data Titles and Legends

Supplementary Data 1. Reference *E. coli* tRNA sequences with modifications.
tRNA sequences are retrieved from tRNAdb[2] and partial or full sequences were changed or added based on the literature [3-10]. Bold letters indicate the changes and addition of sequences based on references.
Supplementary Data 2. Conservation of tRNA modification enzymes between *V. cholerae* and *E. coli*.
Supplementary Data 3. Primary sequences of *V. cholerae* tRNAs with modifications
The nucleosides that are detected in the RNase T_1 (Top rows), RNase A (Middle rows), and either RNase (Bottom rows) fragment analyses of digests are colored in black. The abbreviation of nucleosides is shown on the right.
Supplementary Data 4. Parameters of mass spectrometry for dynamic MRM analyses.
Supplementary Data 5. Comparative genomics for narrowing down candidate acetyltransferases required for acacp^{3}U biogenesis.
Putative acetyltransferases in *V. cholerae* are listed with E-values calculated by BLAST among homologs between *V. cholerae* and indicated organisms. n.d. means the E-value is higher than 1E-10 or no detectable homologs were found.
Supplementary Data 6. Primer list
Supplementary Data 7. Reference DNA sequences of tRNAs for mapping

Supplementary References

1 Kimura, S. & Waldor, M. K. The RNA degradosome promotes tRNA quality control through clearance of hypomodified tRNA. *Proc Natl Acad Sci U S A* **116**, 1394-1403, doi:10.1073/pnas.1814130116 (2019).

2 Juhling, F. *et al.* tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* **37**, D159-162, doi:10.1093/nar/gkn772 (2009).

3 Griffey, R. H. *et al.* 15N-labeled tRNA. Identification of 4-thiouridine in Escherichia coli tRNASer1 and tRNATyr2 by 1H-15N two-dimensional NMR spectroscopy. *J Biol Chem* **261**, 12074-12078 (1986).

4 Horie, N. *et al.* Modified nucleosides in the first positions of the anticodons of tRNA(Leu)4 and tRNA(Leu)5 from Escherichia coli. *Biochemistry* **38**, 207-217, doi:10.1021/bi981865g (1999).

5 Salazar, J. C., Ambrogelly, A., Crain, P. F., McCloskey, J. A. & Soll, D. A truncated aminoacyl-tRNA synthetase modifies RNA. *Proc Natl Acad Sci U S A* **101**, 7536-7541, doi:10.1073/pnas.0401982101 (2004).

6 Bjork, G. R. & Hagervall, T. G. Transfer RNA Modification. *EcoSal Plus* **1**, doi:10.1128/ecosalplus.4.6.2 (2005).

7 Miyauchi, K., Kimura, S. & Suzuki, T. A cyclic form of N6-threonylcarbamoyladenosine as a widely distributed tRNA hypermodification. *Nat Chem Biol* **9**, 105-111, doi:10.1038/nchembio.1137 (2013).

8 Rodriguez-Hernandez, A. *et al.* Structural and mechanistic basis for enhanced translational efficiency by 2-thiouridine at the tRNA anticodon wobble position. *J Mol Biol* **425**, 3888-3906, doi:10.1016/j.jmb.2013.05.018 (2013).

9 Sakai, Y., Miyauchi, K., Kimura, S. & Suzuki, T. Biogenesis and growth phase-dependent alteration of 5-methoxycarbonylmethoxyuridine in tRNA anticodons. *Nucleic Acids Res* **44**, 509-523, doi:10.1093/nar/gkv1470 (2016).

10 Sakai, Y., Kimura, S. & Suzuki, T. Dual pathways of tRNA hydroxylation ensure efficient translation by expanding decoding capability. *Nat Commun* **10**, 2858, doi:10.1038/s41467-019-10750-8 (2019).

11 Millet, Y. A. *et al.* Insights into Vibrio cholerae intestinal colonization from monitoring fluorescently labeled bacteria. *PLoS Pathog* **10**, e1004405, doi:10.1371/journal.ppat.1004405 (2014).

12 Cameron, D. E., Urbach, J. M. & Mekalanos, J. J. A defined transposon mutant library and its use in identifying motility genes in Vibrio cholerae. *Proc Natl Acad Sci U S A* **105**, 8736-8741, doi:10.1073/pnas.0803281105 (2008).

13 Hubbard, T. P. *et al.* Genetic analysis of Vibrio parahaemolyticus intestinal colonization. *Proc Natl Acad Sci U S A* **113**, 6283-6288, doi:10.1073/pnas.1601718113 (2016).

14 Krog, J. S. *et al.* 3-(3-amino-3-carboxypropyl)-5,6-dihydrouridine is one of two novel post-transcriptional modifications in tRNALys(UUU) from Trypanosoma brucei. *FEBS J* **278**, 4782-4796, doi:10.1111/j.1742-4658.2011.08379.x (2011).