Pseudouridine in the Anticodon of *Escherichia coli* $tRNA^{Tyr(Q\Psi A)}$ is Catalyzed by the Dual Specificity Enzyme RluF

Balasubrahmanyam Addepalli and Patrick A Limbach

Department of Chemistry, Rieveschl Laboratories for Mass Spectrometry, University of Cincinnati, Cincinnati OH 45221

Running Title: Pseudouridine in *E. coli* tRNA-Tyr Anticodon

To Whom Correspondence Should be Addressed: Prof. Balasubrahmanyam Addepalli, Department of Chemistry, PO Box 210172, University of Cincinnati, Cincinnati OH 45221 balasual@ucmail.uc.edu; Prof. Patrick A Limbach, Department of Chemistry, PO Box 210172, University of Cincinnati, Cincinnati OH 45221 Pat.Limbach@uc.edu

Supplementary Table S1: Reconstruction of ACU[Q]UA[ms^2i^6A]AΨCUG sequence (with or without carbodiimide or cyanoethylation units) by scoring the product ions following collision-induced dissociation (CID) of corresponding precursor oligonucleotide anion. Average mass-to-charge (*m/z)* values of product ions from the triply charged precursor ion, ACU[Q]UA[ms^2i^6A]AΨCUG, that correspond to underivatized and the carbodiimide or cyanoethylation-tagged species are shown. The observed *m/z* values of product ion series c1-c5 and y1-y5 correspond to singly, c6-c10 and y6-y10 correspond to doubly, and c11 and y11 correspond to the triply charged (deprotonated) condition.

c-ion series					
	underivatized	**1 CMC**	**2 CMC**	**1CE**	**2CE**
c_{11}-ACU[Q]UA[ms^2i^6A]AΨCU	1244.4	1328.6	1412.3	1261.85	1279.5
c_{10}-ACU[Q]UA[ms^2i^6A]AΨC	1714.6	1840.3	1965.9	1740.26	1766.7
c_9-ACU[Q]UA[ms^2i^6A]AΨ	1562.0	1687.7	1813.3	1587.74	1614.2
c_8-ACU[Q]UA[ms^2i^6A]A	1409.0	1534.6		1435.7	
c_7-ACU[Q]UA[ms^2i^6A]	1244.8	1370.0		1271.5	
c_6-ACU[Q]UA	1022.6	1148.3		1049.1	
c_5-ACU[Q]U	1717.1	1967.6		1769.8	
c_4-ACU[Q]	1410.9				
c_3-ACU	939.6				
c_2-AC	633.4				
c_1-A	328.2				
y-ion series					
	underivatized	**1 CMC**	**2 CMC**	**1CE**	**2CE**
y_1-G	282.2				
y_2-UG	588.4				
y_3-CUG	893.6				
y_4-ΨCUG	1199.8	1451.0		1252.5	
y_5-AΨCUG	1529.0	1780.2		1582.1	
y_6-[ms^2i^6A]AΨCUG	985.7	1111.3		1011.7	
y_7-A[ms^2i^6A]AΨCUG	1150.3	1275.9		1176.7	
y_8-UA[ms^2i^6A]AΨCUG	1303.4	1429.0	1554.6	1329.8	1356.5
y_9-[Q]UA[ms^2i^6A]AΨCUG	1539.1	1664.7	1790.3	1565.5	1592.8
y_{10}-U[Q]UA[ms^2i^6A]AΨCUG	1692.1	1817.8	1943.4	1717.9	1745.3
y_{11}-CU[Q]UA[ms^2i^6A]AΨCUG	1229.5	1313.2	1397.0	1247.3	1265.6

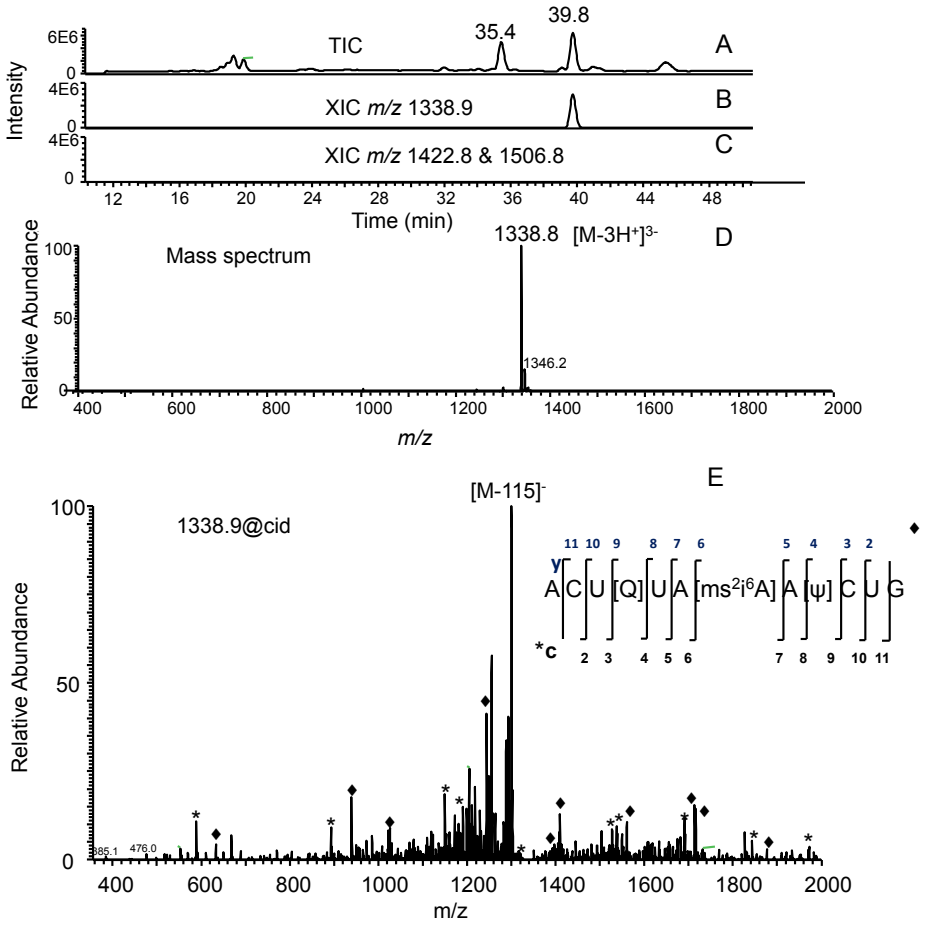
Supplementary Table S2: List of *E. coli* strains procured from Yale Coli Genetic Stock Center (CGSC, http://cgsc.biology.yale.edu/)

S.No.	CGSC#	Name	Genotype	WT function
1.	8365	JW0057-1	F-, ΔrluA768::kan, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻,rph-1, Δ(rhaD-rhaB)568, hsdR514	Dual specificity (23S & tRNA) enzyme
2.	9136	JW1261-3	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻ , ΔrluB777::kan,rph-1, Δ(rhaD-rhaB)568, hsdR514	23S-2605 synthase
3.	9026	JW1072-2	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻ , ΔrluC753::kan,rph-1, Δ(rhaD-rhaB)568, hsdR514	23S-synthase for 955, 2504, 2580 positions
4.	10045	JW2576-3	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻ , ΔrluD759::kan,rph-1, Δ(rhaD-rhaB)568, hsdR514	23S-synthase for 1911, 1915, 1917 positions
5.	9050	JW1121-1	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻ , ΔrluE723::kan,rph-1, Δ(rhaD-rhaB)568, hsdR514	23S-2457 synthase
6.	10864	JW3982-1	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻, rph-1, Δ(rhaD-rhaB)568, ΔrluF789::kan, hsdR514	23S-2604 synthase
7.	9759	JW2171-1	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻ , ΔrsuA740::kan,rph-1, Δ(rhaD-rhaB)568, hsdR514	16S-516 synthase
8.	9857	JW2315-2	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻ , ΔtruA727::kan,rph-1, Δ(rhaD-rhaB)568, hsdR514	tRNA-synthase for 38-40 positions
9.	10386	JW3135-1	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻ , ΔtruB778::kan,rph-1, Δ(rhaD-rhaB)568, hsdR514	tRNA-55 synthase
10.	10165	JW2762-1	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻ , ΔtruC789::kan,rph-1, Δ(rhaD-rhaB)568, hsdR514	tRNA-65 synthase
11.	10137	JW2715-1	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻ , ΔtruD750::kan,rph-1, Δ(rhaD-rhaB)568, hsdR514	tRNA-13 synthase
12.	9745	JW2152-1	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻ , ΔyeiN721::kan,rph-1, Δ(rhaD-rhaB)568, hsdR514	Glycosidase for Ψ-5'-phosphate

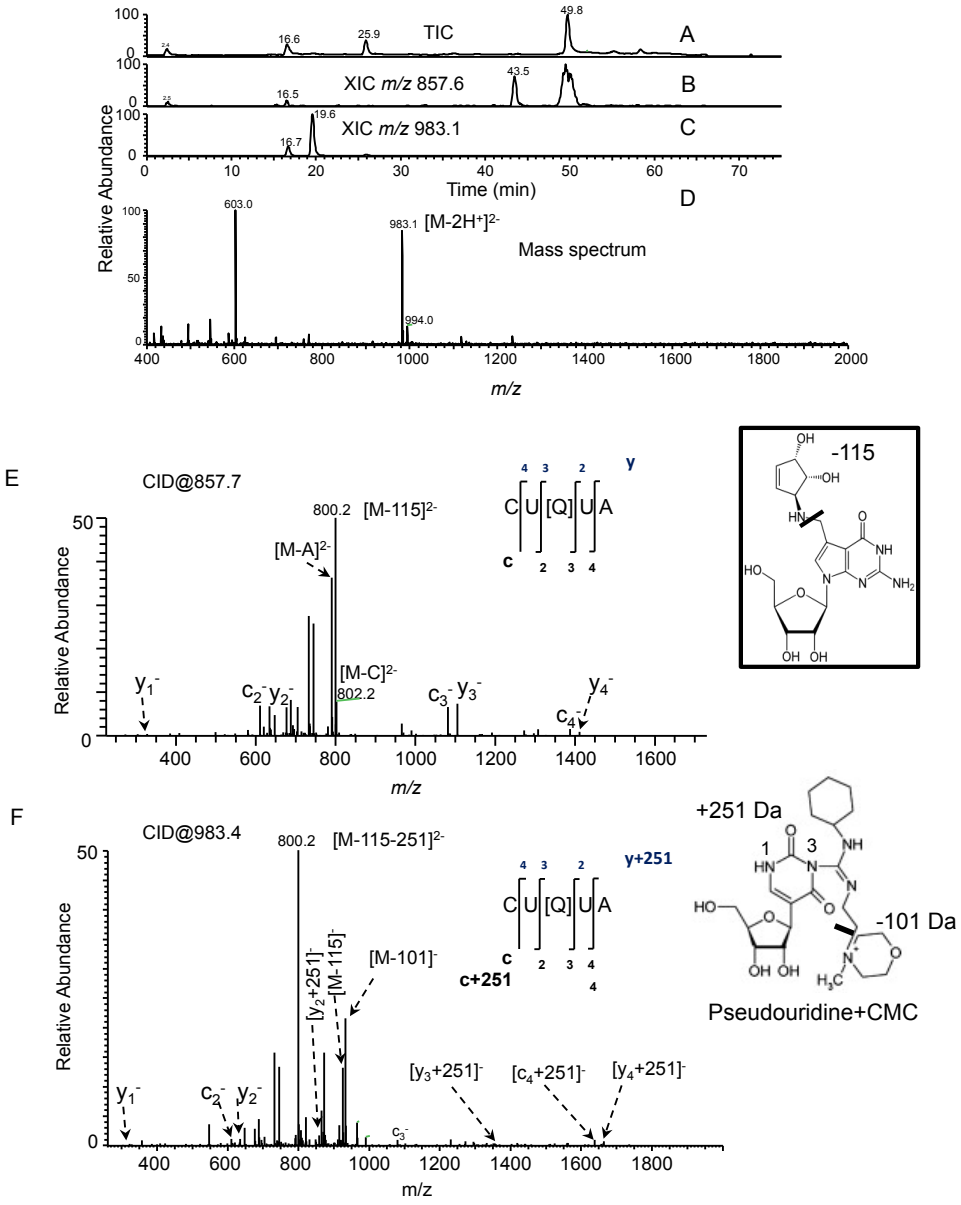
Supplementary Table S3: List of Primers used for genotyping, gene cloning and *in vitro* transcription

Gene	Primer	Sequence	Purpose
RluF	5'(-170 bp upstream of start)	ATCACGGGAGGCATGCTA	Genotyping
	3' (+137 bp downstream of stop)	GCTAATCCACAATCAACGCCT	
	5'BamH1+1	GGGTGGATCCGATGCTGCCCGACTCATCAGT	Gene cloning
	3'HindIII	TTTTTAAGCTTGCGCCCCTTCTTTTTACGC	
Kanamycin	K1-3'	CAGTCATAGCCGAATAGCCT	genotyping
	K2-5'	CGGTGCCCTGAATGAACTGC	
TruA	5'-(-240 bp upstream of Start)	AGAAGAAGCGCGTGATGC	genotyping
	3'(+237 bp downstream of stop	TCACCCGGTAAAAACGGC	
	5'BamH1+1	GGGTGGATCCGATGTCCGACCAGCAACAACCG	Gene cloning
	3' HindIII	TTTTTAAGCTTGTCCGCCAGAAATAGCGGGC	
$tDNA^{Tyr}$	5'-T7	TTTAGATCTATTAATACGACTCACTATAGGGGTGGGGTTCCCGAGCGGCCAAA	In vitro transcription
	3'	TGGTGGTGGGGGAAGGATT	
Firefly LUC	5'BglII+1	GGGGAGATCTCATGGAAGACGCCAAAAACATAAAGAAAGGC	Gene cloning
	3'XhoI	GGGGCTCGAGCACGGCGATCTTTCCGCCCTTCTT	
$Asn(AAY)_{10}$ 5' NdeI-BamHI 3'	+ strand	TTTTCATATGAATAATAATAATAATAACAACAACAACAACGCGGATCCGGGG	$(Asn)_{10}$-FF LUC
	- strand	CCCCGGATCCGCGTTGTTGTTGTTGTTATTATTATTATTATTCATATGAAAA	
Tyr $(TAY)_{10}$ 5' NdeI-BamHI 3'	+ strand	TTTTCATATGTATTATTATTATTATTACTACTACTACTACGCGGATCCGGGG	$(Tyr)_{10}$-FF LUC
	- strand	CCCCGGATCCGCGTAGTAGTAGTAGTAATAATAATAATAATACATATGAAAA	
Ren LUC	5'SacIRenLUC	GGGGAGCTCGATGACTTCGAAAGTTTATGATCCAGAAC	Gene cloning
	3'RenLUCHindIII	TTTTAAGCTTTTGTTCATTTTTGAGAACTCGCTCA	

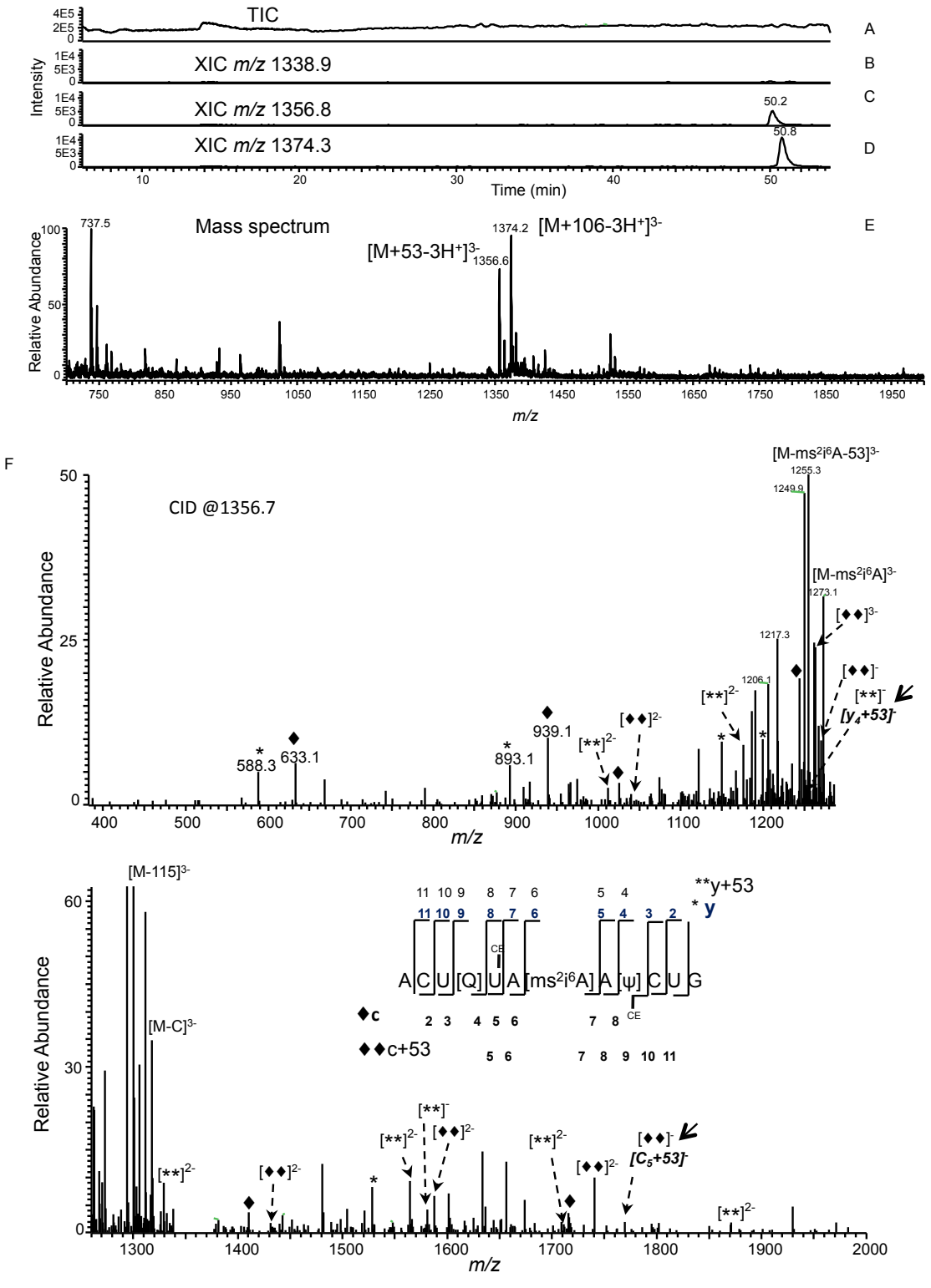
Supplemental Figures



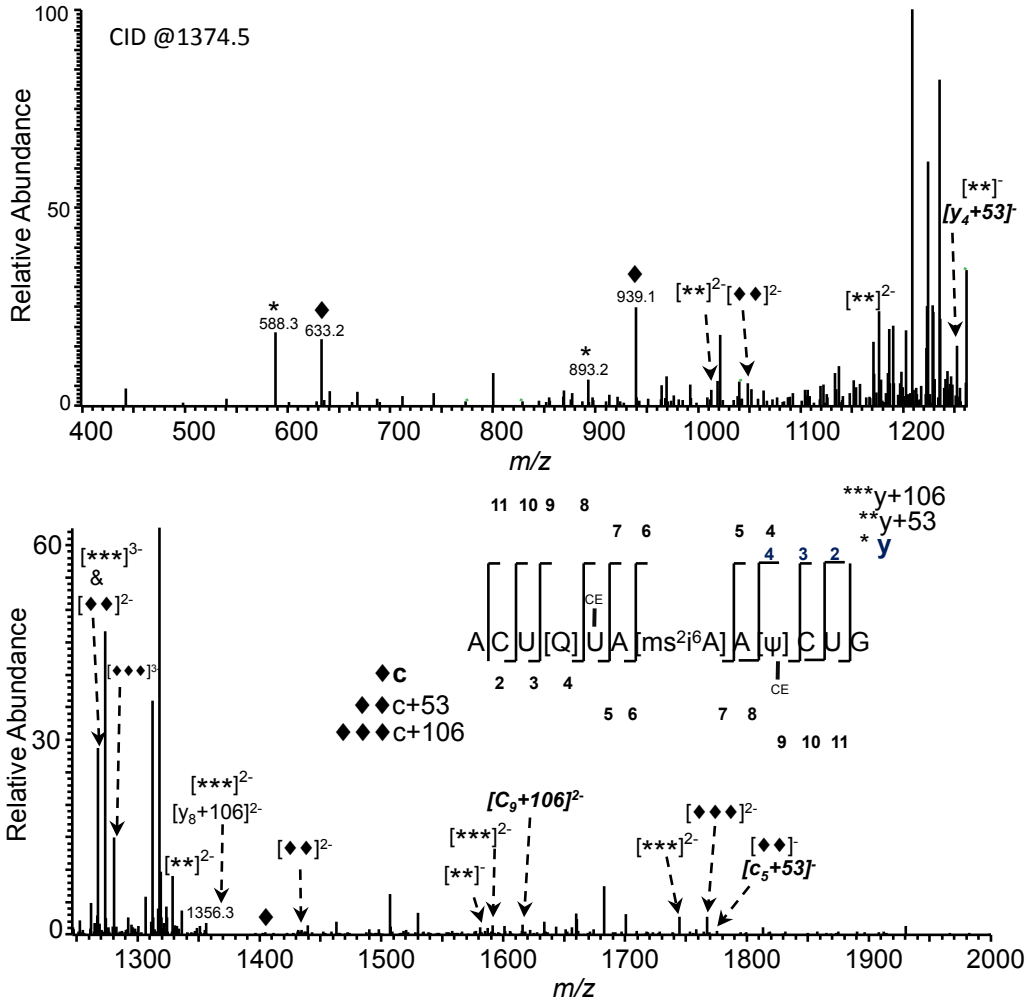
Supplementary Figure S1: LC-MS analysis of RNase T1 signature digestion product, ACU[Q]UA[ms^{2}i^{6}A]A[Ψ]CUG (*m/z* 1338.8), from anticodon loop of tRNA$^{Tyr(QUA)}$. **(A)** Total ion chromatogram (TIC) representing the elution pattern of oligonucleotide anions in the digestion mixture. **(B)** Extracted ion chromatogram (XIC) for *m/z* 1338.9 corresponding to the triply charged oligonucleotide, AACU[Q]UA[ms^{2}i^{6}A]A[Ψ]CUG. **(C)** XICs for *m/z* 1422.8 and 1506.8, which would correspond to expected CMCT-derivatized digestion products. No interfering ions from an underivatized digest are present. **(D)** Mass spectrum corresponding to the XIC peak eluting between 39.1-40.1 min is depicted. **(E)** Tandem (MS/MS) mass spectrum obtained by collision-induced dissociation (CID) of the precursor ion oligomer, ACU[Q]UA[ms^{2}i^{6}A]A[Ψ]CUG (*m/z* 1338.8), is shown. The 'c_n' product ion series that bear the common 5'-end are indicated by '*' and the 'y_n' product ion series that have the common 3'-end are indicated by '♦', respectively. The most abundant product ion in the spectrum correspond to the characteristic loss of (115 Da) due to internal cleavage of queuosine is illustrated.



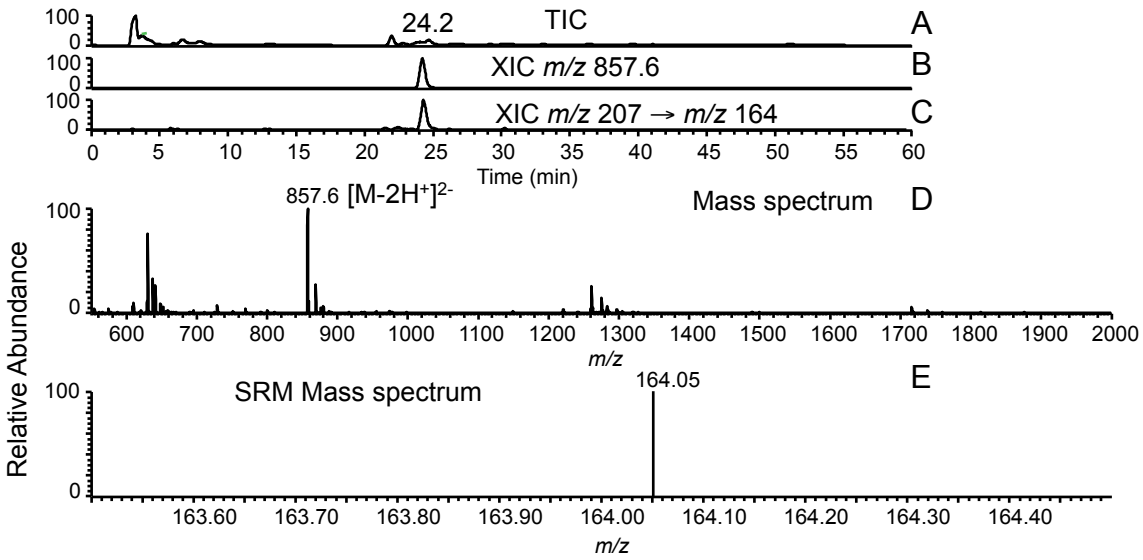
Supplemental Figure S2: **(A)** TIC of CMCT treated RNase U2 digest of $tRNA^{Tyr(QUA)}$ is shown. **(B)** XIC for *m/z* 857.6 corresponding to the doubly charged CUQUA anion. **(C)** XIC for *m/z* 983.1 corresponding to the carbodiimide-tagged oligonucleotide anion. **(D)** Mass spectrum of the XIC peak at 19.6 min depicting the presence of doubly charged anion with *m/z* 983.1 is shown. **(E)** MS/MS spectrum of the doubly charged oligonucleotide ion with no carbodiimide (*m/z* 857.7). **(F)** MS/MS spectrum of the doubly charged oligonucleotide ion with one carbodiimide (*m/z* 983.4). Addition of carbodiimide unit at specific product ion is indicated. The predominant product ion observed in both MS/MS spectra is identical even though precursor ions with different *m/z* values were targeted for dissociation indicating high lability of the carbodiimide group.


TIC
XIC m/z 1338.9
XIC m/z 1356.8
XIC m/z 1374.3
Intensity
Time (min)
A
B
C
D
50.2
50.8
Mass spectrum
737.5
1374.2
[M+106-3H+]3-
[M+53-3H+]3-
1356.6
Relative Abundance
m/z
E
F
CID @1356.7
[M-ms2i6A-53]3-
1255.3
1249.9
[M-ms2i6A]3-
1273.1
[♦♦]3-
1217.3
[♦♦]-
[**]-
[y4+53]-
[**]2-
1206.1
[♦♦]2-
939.1
893.1
633.1
588.3
[M-115]3-
[M-C]3-
**y+53
* y
ACU[Q]UA[ms2i6A]A[ψ]CUG
CE
♦c
♦♦c+53
[C5+53]-

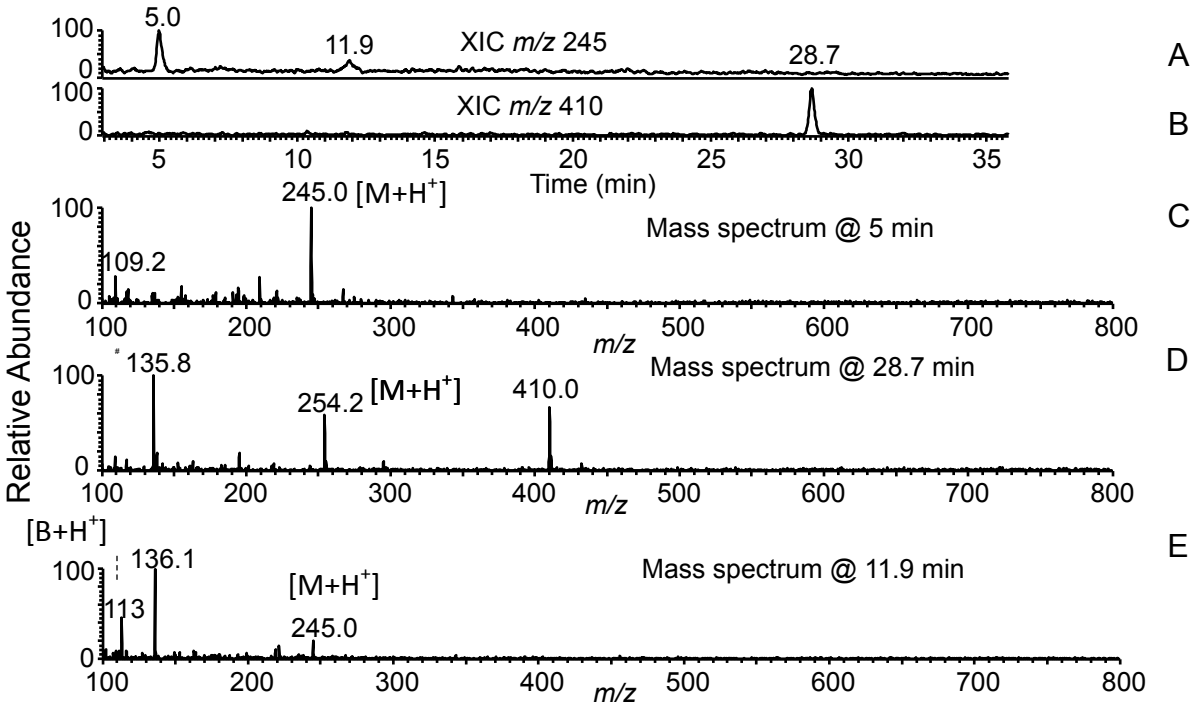
G



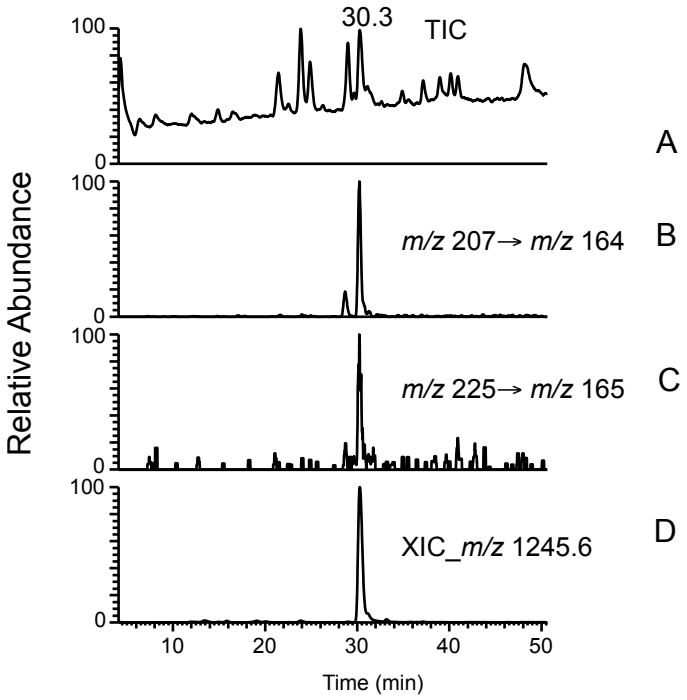
Supplemental Figure S3: **(A)** TIC of acrylonitrile treated T1 digest of $tRNA^{Tyr}$. **(B)** XIC for *m/z* 1338.9 corresponding to the triply charged underivatized oligonucleotide, AACU[Q]UA[ms^2i^6A]A[Ψ]CUG. Note the lower signal intensity for the underivatized digestion product. **(C)** XIC for *m/z* 1356.6 that corresponds to the triply charged ion with one cyanoethylation unit. **(D)** XIC for *m/z* 1374.2 that corresponds to the triply charged ion with two cyanoethylation units. **(E)** Mass spectrum of the XIC peak observed in **(C)** and **(D)**. The signal intensity for cyanoethylated oligonucleotides was significantly higher in the sample. **(F)** MS/MS spectrum of *m/z* 1356.7. The appearance of unique fragment ions with additional mass of cyanoethylation from c_5 (+53, m/z 1769.5) and y_4 (+53, m/z 1252.4) in the tandem mass spectrum that were not seen with the underivatized oligomer, are indicated by bold italics and arrows. **(G)** MS/MS spectrum of *m/z* 1374.3. The appearance of unique fragment ions with additional mass of cyanoethylation from c_5 (+53, m/z 1769.5) and y_4 (+53, m/z 1252.4) or two cyanoethylations from c_9 (+106 Da, m/z 1614.8) and y_8 (+106 Da, 1202.6) in the tandem mass spectrum that were not seen with the underivatized oligomer, are indicated in bold italics. Product ions with no cyanoethylation are shown by '* ' for y_n product ion series and '◆' for c_n product ion series, respectively. Product ions with one cyanoethylation unit are depicted as' ** ' for y_n+53 and '◆◆' for c_n+53. Product ions with an additional 106 Da were marked as '*** ' for y_n+106 and ' ◆◆◆ ' for c_n+106, respectively.



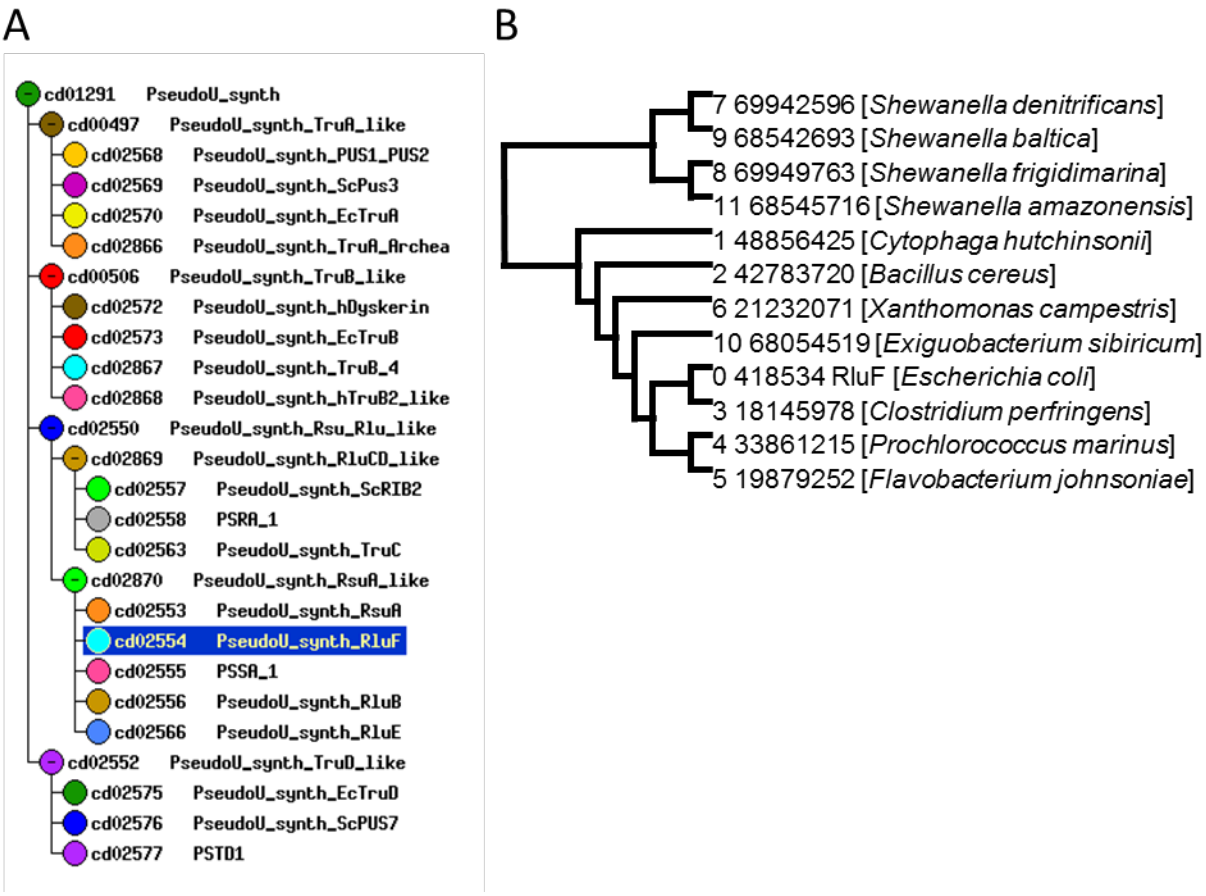
Supplemental Figure S4: LC-MS/MS of Rnase U2 oligonucleotide, CU[Q]UA from anticodon loop of E. coli $tRNA^{Tyr(QUA)}$. The tRNA was digested using RNase U2 and analyzed by LC-MS/MS. Presence of pseudouridine was detected by the Selected Reaction Monitoring (SRM) involving transition of precursor ion *m/z* 207 to product ion *m/z* 164. **(A)** Total ion chromatogram (TIC) of the RNase U2 digest of $tRNA^{Tyr(QUA)}$. **(B)** XIC of *m/z* 857.6, corresponding to the doubly charged anion of CUQUA. **(C)** SRM scan for pseudouridine-specific transition. **(D)** Mass spectrum of the XIC peak observed at 24.2 min in **(B)**. **(E)** Mass spectrum of the SRM peak at 24.2 min.



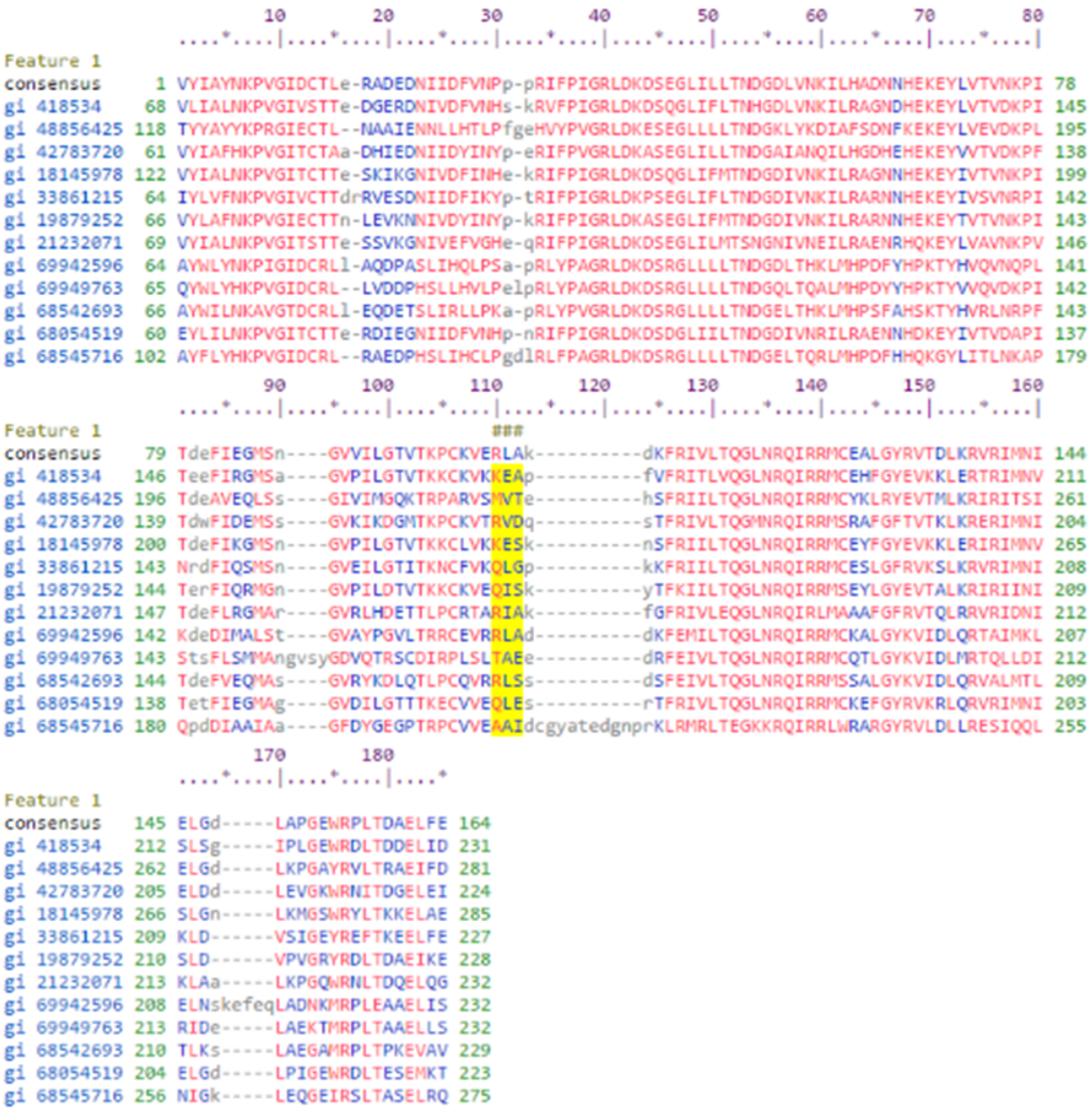
Supplemental Figure S5: LC-MS analysis of nucleoside hydrolysate of the oligonucleotide, CU[Q]UA. The tRNA(Tyr) was digested with RNase U2 and fractionated to collect the eluent corresponding to *m/z* 857.7 anion. This fraction was then hydrolyzed to nucleosides using snake venom phosphodiesterase and alkaline phosphatase. The hydrolysate was subjected to LC-MS on synergy hydro RP C-18 column to detect the nucleosides present in the sample. In this case the mass spectra are acquired in positive ion mode **(A)** XIC for *m/z* 245 corresponding to pseudouridine/uridine. Note the presence of two peaks with retention times of 5 min and 11.9 min, corresponding to pseudouridine and uridine. **(B)** XIC for *m/z* 410 corresponding to queuosine [Q]. **(C)** Mass spectrum of the XIC peak at 5 min. Note the absence of uracil nitrogen base ion with *m/z* 113. **(D)** Mass spectrum of the XIC peak at 28.7 min depicting the presence of *m/z* 410 corresponding to queuosine. **(E)** Mass spectrum of the XIC peak at 11.9 min depicting the presence of uracil nitrogen base ion (*m/z* 113) and nucleoside ion (*m/z* 245). Pseudouridine with its extra hydrogen bond donor capability is more hydrophilic than uridine, thereby eluting early (5.0 min) in the void volume compared to the uridine eluting at 11.9 min. Under the given MS-conditions which enable base loss from nucleosides, the modified nucleoside yielded only molecular ion (*m/z* 245) without exhibiting the base loss (due to the stable C-C glycosidic bond), while the uridine eluting late (11.9 min) in the gradient chromatography exhibited both nucleobase (*m/z* 113) and nucleoside (*m/z* 245) ions when analyzed in the positive ion mode.



Supplemental Figure S6: Selected Reaction Monitoring (SRM) pseudouridine analysis of RNase T1 oligonucleotide derived from *E. coli* $tDNA^{Tyr}$ *in vitro* transcript. The *in vitro* transcript was treated with purified RluF protein and digested with RNase T1 and BAP before analysis. Pseudouridine is measured by SRM, where the oligonucleotides are dissociated to generate the pseudouridine-specific precursor ions (*m/z* 207/225) at the ionization source and are subsequently fragmented to detect the product ions (*m/z* 164/165). **(A)** Total ion chromatogram (TIC) of the LC-MS analysis of RNase U2 digest of $tRNA^{Tyr(QUA)}$. **(B)** Pseudouridine-specific SRM transition. **(C)** 5'end Pseudouridine-specific SRM transition. **(D)** XIC of T1 oligonucleotide, UAAAUCUG (*m/z* 1245.6). Note the alignment of pseudouridine-specific MS/MS scans with XIC of m/z 1245.6 corresponding to oligonucleotide anion, UAAAUCUG.



Supplemental Figure S7: **(A)** Phylogenetic classification of pseudouridine synthases based on constraint-based multiple sequence alignment (http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?hslf=1&uid=cd01291&#seqhrch). There are four families that are named after RsuA/RluD, TruA, TruB and TruD proteins found in *E. coli*, which are arranged based on the conserved domain arrangement. The proteins that share the conserved domain models are arranged into a sequence cluster. The RluF cluster is designated as cd02554. **(B)** The Cd02554 cluster belonging to RluF like proteins comprises of bacterial proteins similar to *E coli* RluF. The 'gi' number and the designated bacterial name are shown.



Supplemental Figure S8: Alignment of the conserved regions of bacterial pseudouridine synthase proteins similar to *E. coli* RluF. *E.coli* RluF makes psi2604 in 23S RNA and psi35 in tRNATyr. Pseudouridine at 2604 has only been detected in *E. coli*.