Supplemental Data and Information for:

Mass Spectrometry Sequencing of Transfer Ribonucleic Acids by the Comparative Analysis of RNA Digests (CARD) Approach

Siwei Li and Patrick A. Limbach*

Rieveschl Laboratories for Mass Spectrometry, Department of Chemistry, PO Box 210172, University of Cincinnati, Cincinnati, OH 45221-0172

*To whom correspondence should be addressed.

Phone (513) 556-1871
Fax (513) 556-9239
email Pat.Limbach@uc.edu

Supplemental Table S1. Basic Local Alignment Search Tool (BLAST) results for searching 14 randomly chosen *E. coli* tRNA genes against the *C. koseri* genome. The value reported is the %sequence homology between the *E. coli* gene and that found in *C. koseri*.

***E. coli* query sequence:**	**tRNA ID**	***E. coli***	***C. koseri***
tRNA: Alanine 1 (Ala1) [A]	DA1660	100%	100%
tRNA: Cysteine (Cys) [C]	DC1660	100%	98%
tRNA: Glutamic Acid 1 (Glu1) [E]	DE1660	100%	100%
tRNA: Phenylalanine (Phe) [F]	DF1660	100%	100%
tRNA: Glycine 1 (Gly1) [G]	DG1660	100%	100%
tRNA: Histidine (His) [H]	DH1660	100%	100%
tRNA: Isoleucine 1 (Ile1) [I]	DI1660	100%	100%
tRNA: Leucine 1 (Leu1) [L]	DL1660	100%	100%
tRNA: Methionine (Met) [M]	DM1660	100%	100%
tRNA: Glutamine 1 (Gln1) [Q]	DQ1660	100%	100%
tRNA: Arginine 1 (Arg1) [R]	DR1660	100%	98%
tRNA: Serine 1 (Ser1) [S]	DS1664	100%	100%
tRNA: Threonine 1 (Thr1) [T]	DT1660	100%	97%
tRNA: Valine 1 (Val1) [V]	DV1662	100%	100%

Supplemental Table S2. Excel file containing tRNA sequences for *E. coli* obtained from the tRNAdb 2009 database (http://trnadb.bioinf.uni-leipzig.de/)[1] aligned with the appropriate *C. koseri* tRNA genes obtained from the Genomic tRNA Database (http://gtrnadb.ucsc.edu/)[2] or from BLAST searches against the *C. koseri* genome (NC_009792.1).[3] Aligned sequences are then compared as described in the text to yield anticipated singlets and doublets during comparative sequencing.

Supplemental Table S3. Excel file containing tRNA sequences for *E. coli* obtained from the tRNAdb 2009 database (http://trnadb.bioinf.uni-leipzig.de/)[1] aligned with the appropriate *S. enterica* tRNA genes obtained from the Genomic tRNA Database (http://gtrnadb.ucsc.edu/).[2] Aligned sequences are then compared as described in the text to yield anticipated singlets and doublets during comparative sequencing.

Supplemental Figure S1. Extracted ion chromatograms corresponding to the anticipated doubly- and triply-charged ions of AC[s^2C]U[mnm^{5}U]CU[t^6A]AGp, a singlet from tRNA-Arg(UCU), at 5 μg and 20 μg of *E. coli* total tRNA after RNase T1 digestion loaded on column. The XIC response at 31.5 min corresponds to a doubly-charged ion consistent with a base composition of $C_2U_2A_2Gp$ + methyl, which is not the anticipated singlet.

Supplemental Figure S2. Mass spectrum corresponding to the anticipated singlet [DUGp] from *C. koseri* tRNA-Asp(GUC). A number of interfering ions, likely originating from the ubiquitous digestion products CCGp (*m/z* 972.1), (CU)Gp (*m/z* 973.1), and UUGp (*m/z* 974.1), are present and can interfere with detection of the DUGp singlet.

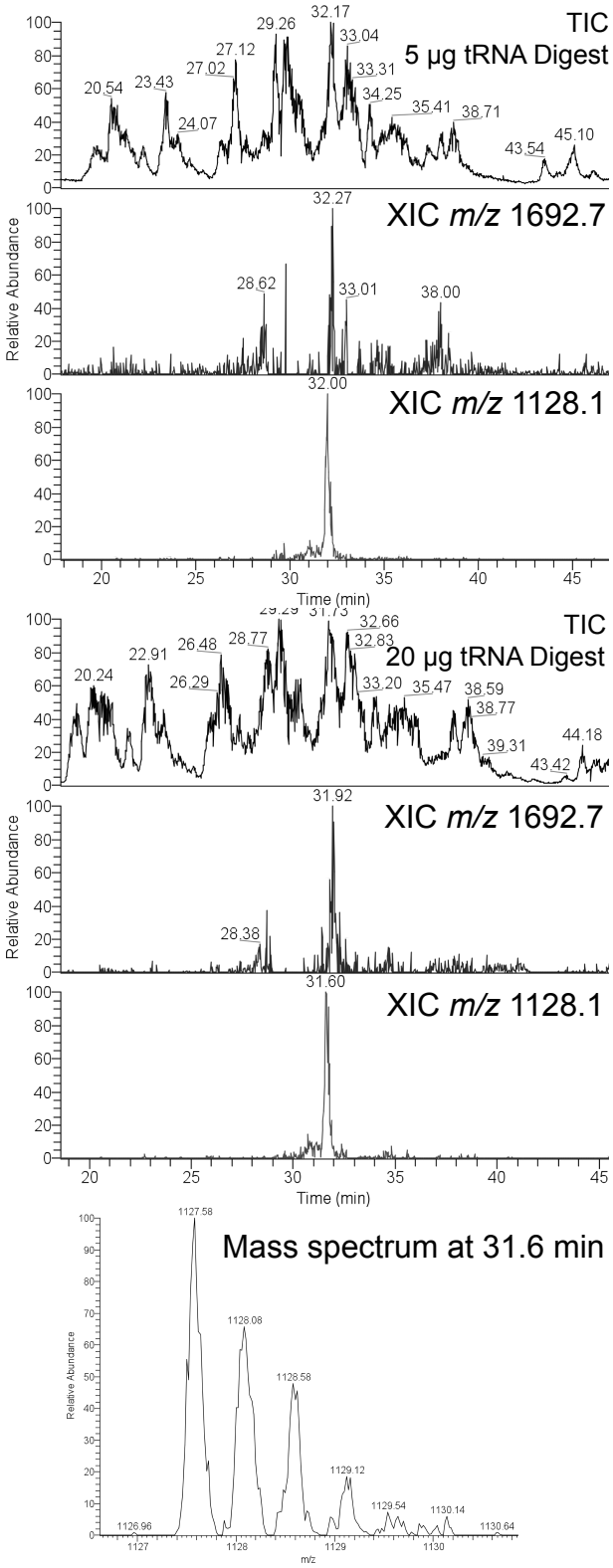
Supplemental Figure S3. Mass spectra corresponding to the anticipated singlet CAAAGp (*m/z* 827.5, 2- charge) when (a) *C. koseri* is labeled with ^{18}O and (b) *E. coli* is labeled with ^{18}O. Although the isotopic envelope shifts by -1 *m/z* unit when *C. koseri* is labeled with ^{16}O, interfering RNase T1 digestion products hinder unequivocal assignment of this singlet.

Supplemental Figure S4. (a) Collision-induced dissociation mass spectrum of the singlet U[s^4U]AACAAAGp from *E. coli* total tRNAs. (b) Collision-induced dissociation mass spectrum of the doublet [m^7G]UCCCCAGp from tRNA-Thr(GGU).

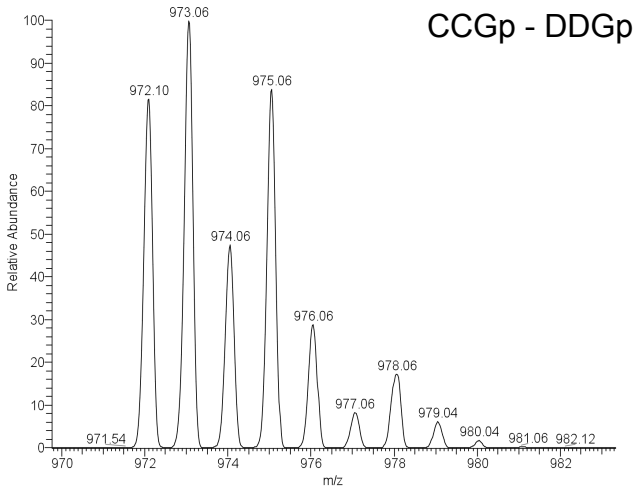
Supplemental Figure S5. Calculated codon usage frequencies for *E. coli, C. koseri* and *T. thermophilus*. Comparative sequencing is more effective when the codon usage frequencies of the reference and candidate organisms are similar.

References

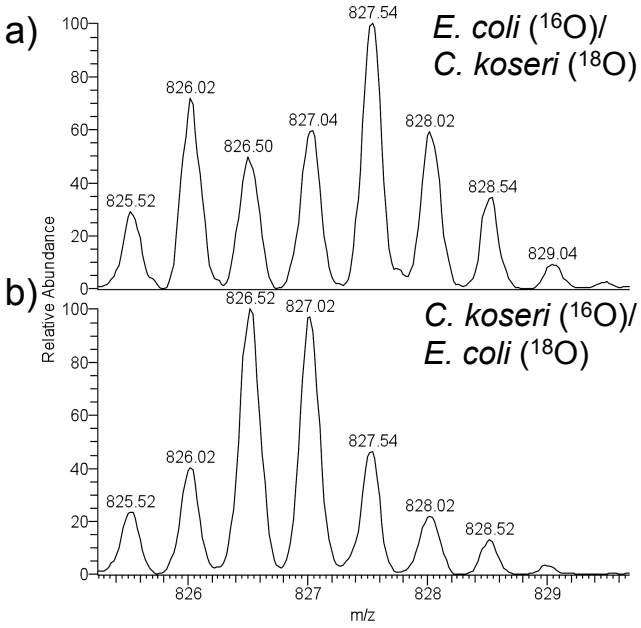
1. F. Juhling, M. Morl, R. K. Hartmann, M. Sprinzl, P. F. Stadler and J. Putz, *Nucleic Acids Res*, 2009, **37**, D159-162.
2. P. P. Chan and T. M. Lowe, *Nucleic Acids Res*, 2009, **37**, D93-97.
3. S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman, *J Mol Biol*, 1990, **215**, 403-410.



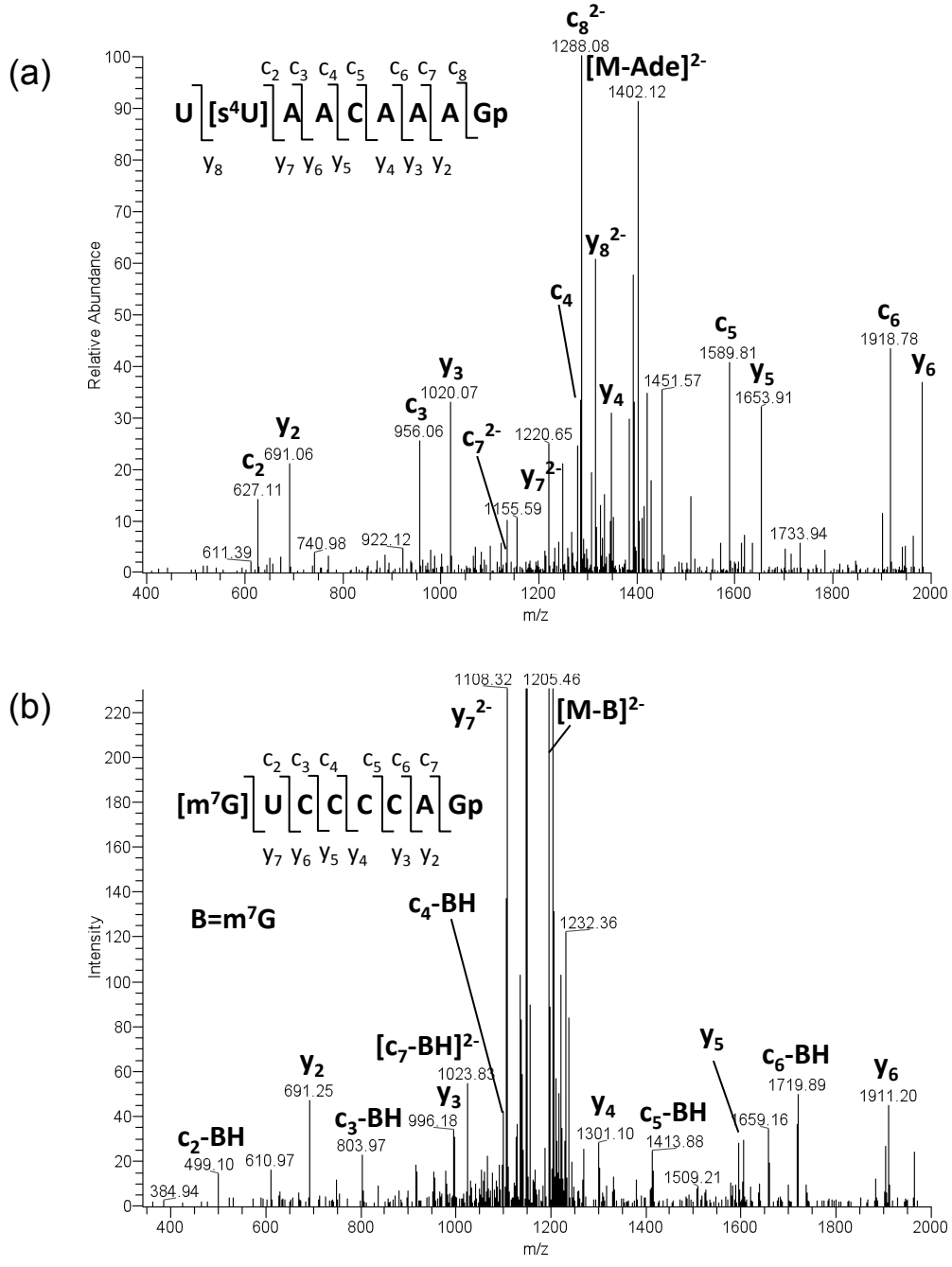
Supplemental Figure S1. Extracted ion chromatograms corresponding to the anticipated doubly- and triply-charged ions of AC[s^2C]U[mnm^{5}U]CU[t^6A]AGp, a singlet from tRNA-Arg(UCU), at 5 µg and 20 µg of *E. coli* total tRNA after RNase T1 digestion loaded on column. The XIC response at 31.5 min corresponds to a doubly-charged ion consistent with a base composition of $C_2U_2A_2Gp$ + methyl, which is not the anticipated singlet.



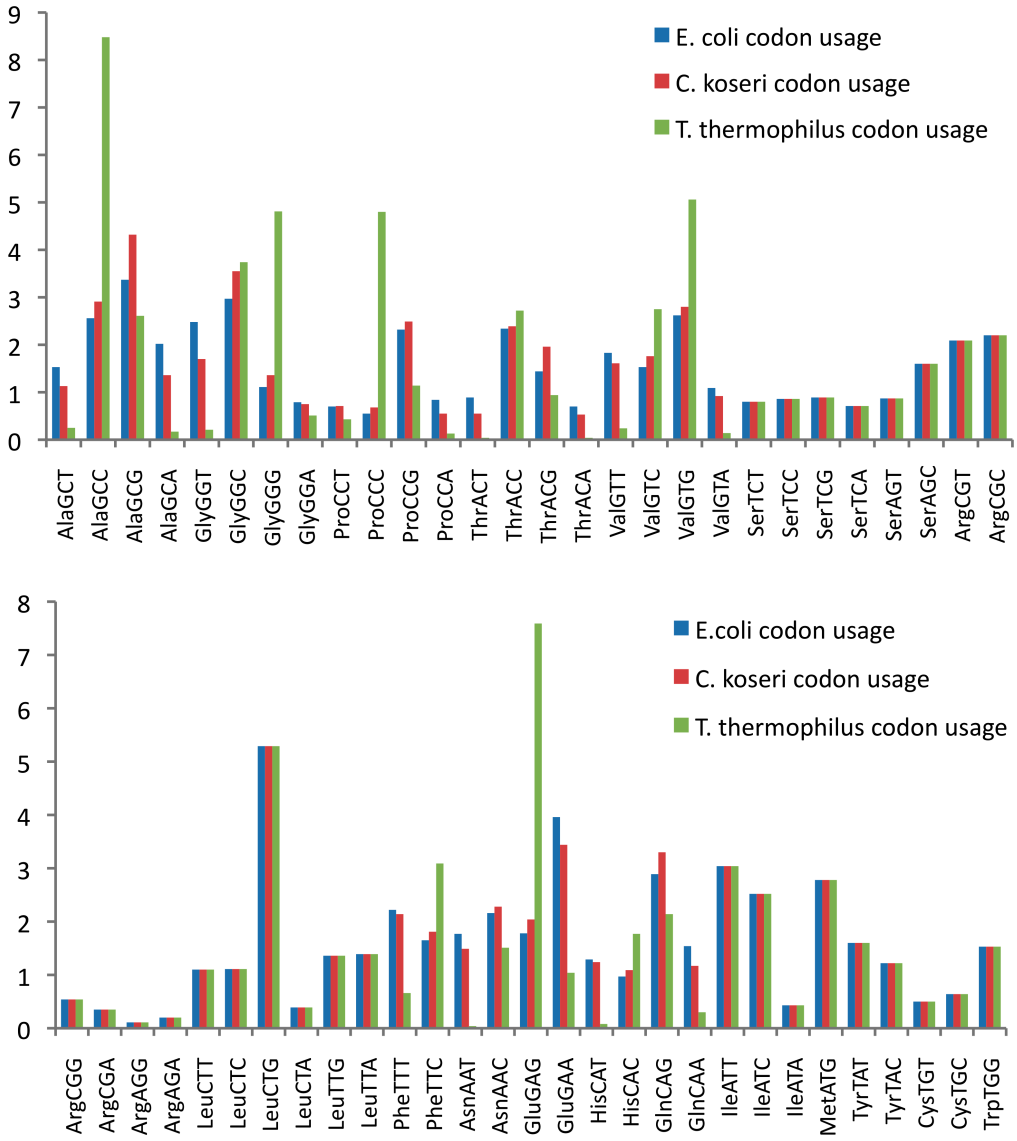
Supplemental Figure S2. Mass spectrum corresponding to the anticipated singlet [DUGp] from *C. koseri* tRNA-Asp(GUC). A number of interfering ions, likely originating from the ubiquitous digestion products CCGp (*m/z* 972.1), (CU)Gp (*m/z* 973.1), and UUGp (*m/z* 974.1), are present and can interfere with detection of the DUGp singlet.



Supplemental Figure S3. Mass spectra corresponding to the anticipated singlet CAAAGp (*m/z* 827.5, 2- charge) when (a) *C. koseri* is labeled with ^{18}O and (b) *E. coli* is labeled with ^{18}O. Although the isotopic envelope shifts by -1 *m/z* unit when *C. koseri* is labeled with ^{16}O, interfering RNase T1 digestion products hinder unequivocal assignment of this singlet.



Supplemental Figure S4. (a) Collision-induced dissociation mass spectrum of the singlet U[s^{4}U]AACAAAGp from *E. coli* total tRNAs. **(b)** Collision-induced dissociation mass spectrum of the doublet [m^{7}G]UCCCCAGp.



Supplemental Figure S5. Calculated codon usage frequencies for *E. coli, C. koseri* and *T. thermophilus*. Comparative sequencing is more effective when the codon usage frequencies of the reference and candidate organisms are similar.