Annotation of tRNA modifications genes in model organisms

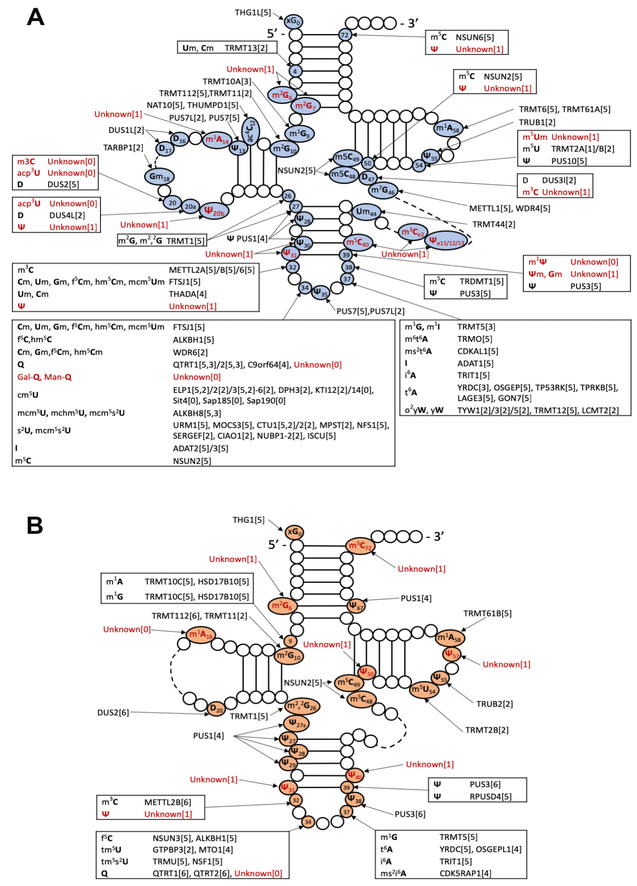
Valérie de Crécy-Lagard[1], Patrick Limbach[2], Ningxi Yu[2], Robert Ross[2], Pietro Boccaletto[3], Todd Lowe[4], Puneet Sharma[5], Sebastian Leidel[5], Janusz Bujnicki[3]

[1]Department of Microbiology and Cell Sciences, University of Florida, Gainesville, 32611, USA; [2]Rieveschl Laboratories for Mass Spectrometry, Department of Chemistry, University of Cincinnati, P.O. Box 210172, Cincinnati, Ohio 45221-0172, United States; [3]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, ul. Trojdena 4, 02-109 Warsaw, Poland; [4]Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA 95064, USA; [5]Research group for RNA Biochemistry, Institute of Chemistry and Biochemistry, University of Bern, 3012 Bern, Switzerland

Abstract

tRNAs are post-transcriptionally modified by chemical modifications that affect all aspects of translation. In any given genome, a range of 10 genes in intracellular bacteria or symbiont to over 130 genes in higher eukaryotes are dedicated to this process. The near complete sets of tRNA modification genes are currently available for only one organism per domain of life: *Saccharomyces cerevisiae* for Eukarya, *Escherichia coli* for Bacteria and *Haloferax volcanii* for Archaea. In this study, we set out to analyze to analyze the current status of annotation of these genes in Uniprot. We also predicted the tRNA modification genes in three other model organisms: the model gram-positive *Bacillus subtilis* 168, the model methanogen *Methanocaldococcus janaschii* and in *Homo sapiens*. In all cases, additional experimental analyses were required (LC/MS or tRNA-Seq) to confirm the identify and location of many modifications. Theses analyses revealed major differences in different organisms from the same kingdom, not only between the modifications found but also in the choice of enzymes catalyzing the same reactions. In addition, as an increasing number of mutations underlying human genetic diseases map to genes encoding for tRNA modification enzymes and our knowledge on human tRNA-modification genes remains fragmentary, this work provides a comprehensive up-to-date compilation of human tRNA modifications and their enzymes that can be used as a resource for further studies.

Predicted Human tRNA modification enzymes



Cloverleaf representation of human tRNA, with modified positions indicated for (A) cytoplasmic and (B) mitochondrial tRNAs, respectively, indicating genes/ proteins experimentally validated in human, predicted with high confidence in other species, unknown with predictions, and unknown with no predictions) Evidence codes: [5] in vivo data in mammals; [4] in vivo data of the human or a related mammalian enzyme in a heterologous host; [3] in vitro data using the human enzyme; [2] similarity to an experimentally validated gene in a non-mammalian species; [1] candidates that have not been verified in any organism; [0] no clear candidate [6] enzymes that have been experimentally validated in the cytoplasm but not in mitochondria. We estimate that between ~135 genes are required to modify cytosolic and mitochondrial tRNAs and that 23% of these genes still need to be identified and that another 22% require further experimental validations. Approximately 50% of the human modification genes have been linked to a number of human diseases Searching for "tRNA modification" GO in *Homo sapiens* uniprot in uniport captured 84 correct genes and 2 incorrect. Approximately 50% of the human modification genes have been linked to a number of human diseases. Full analysis in Ref [5].

Background

The major locations of tRNA modifications are in the Anticodon Stem Loop (ASL) or in the structural core of the tRNA formed by the D-loop and and TΨC-loop. Their function can be structural but many modifications have key roles as determinants or anti-determinants for other components of the translation apparatus. Finally, ASL modifications have often key roles in decoding and the in translation efficiency and accuracy. See refs [1] and [2] for reviews.

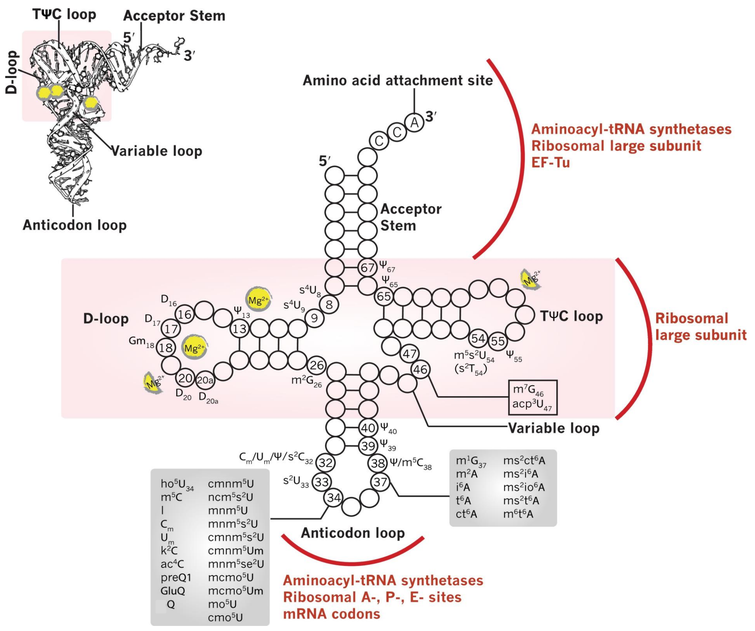
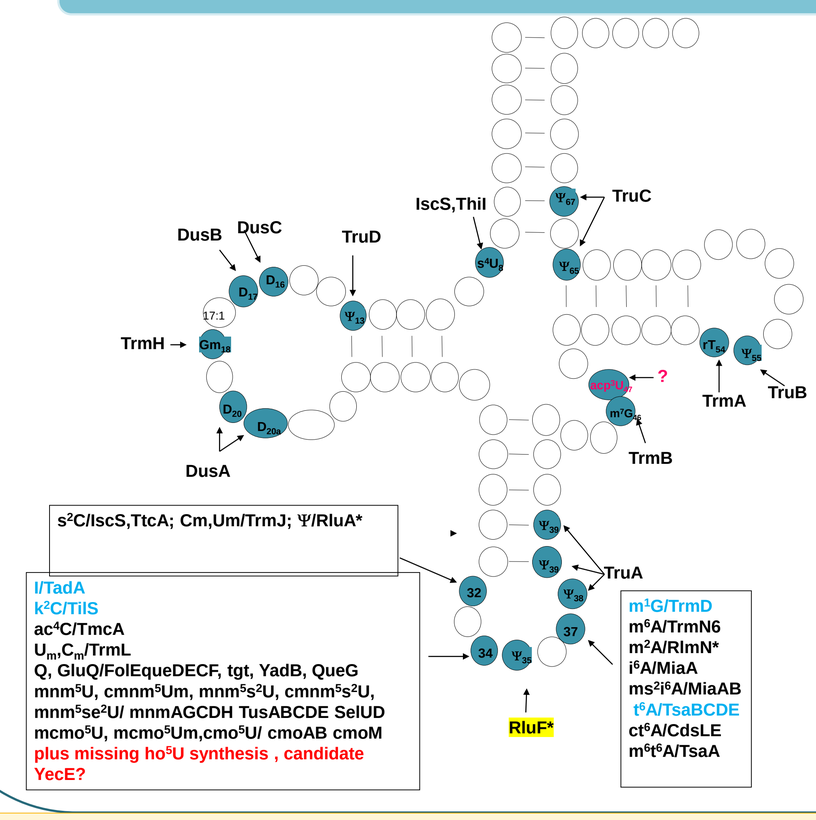


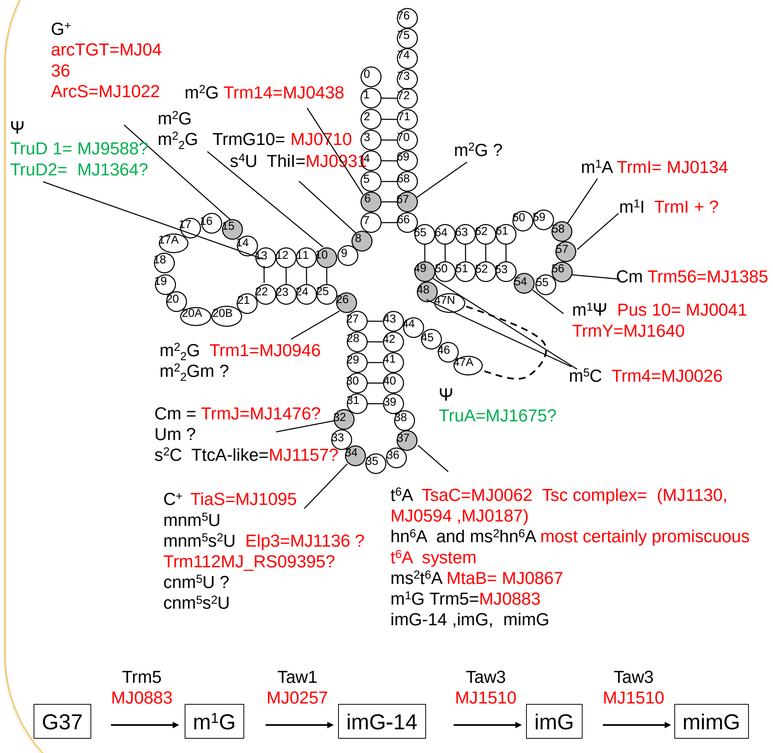
Figure from Reference [3]

Known *E. coli* tRNA modification genes

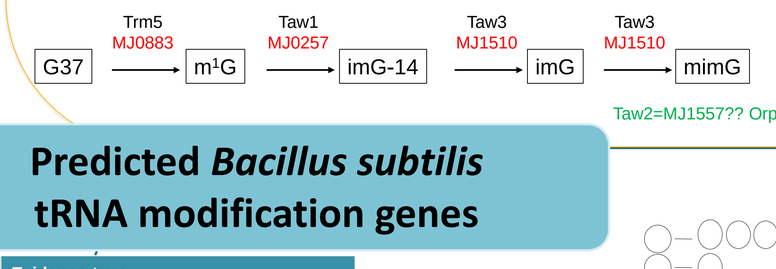


55 four genes have been shown to be involved in tRNA modification in the the gram-negative model *Escherichia coli* K12.The only missing ones are those involved in the synthesis of ho^{5}U34 and acp^{3}U47 (in red). Only a few are essential (in blue) and three (RluA, RluF and RlmN, starred) are dual specificity enzymes that also target rRNA molecules. 54 of these genes were correctly annotated in Uniprot at the time of the analysis (January 2019). The only one that was not was RluF, as the tRNA modification activity (ref [4]) was not captured. Conversely when the GO terms "tRNA modification were used to query the *E. coli* K12 proteome using query: goa:("tRNA modification [6400]") AND organism: "Escherichia coli (strain K12) [83333]" 53 proteins were retrieved however these were not all correct. All the preQ$_0$ synthesis proteins were missing (FolE, QueDCEF) as well as the last enzyme of Queuosine synthesis, QueG. In addition, AlaRS and RimO were incorrectly linked to tRNA modification. For the last two gene that did not overlap with our curated list: Fmt (Met-tRNA(fMet) formyltransferase) and LasT (Uncharacterized tRNA/rRNA methyltransferase LasT (EC 2.1.1.-)), one can understand the GO term linkage, because of the functional overlap.

Predicted *Methanocaldococcus janaschii* tRNA modification genes



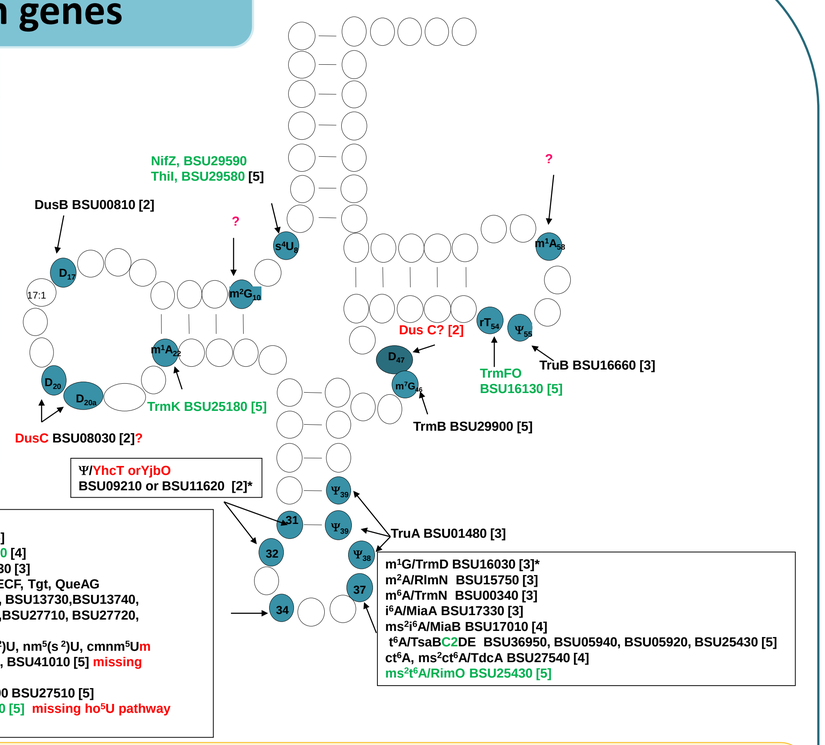
The consensus modification sequence for *M. jannaschii* tRNAs. The positions of modifications placed on the tRNA secondary structure were experimentally validated. The positions of modifications predicted or reported in other studies are listed below. The 28 predicted modifications enzymes are noted in red if the modification was experimentally mapped and in green if it was not, At least 5 genes are missing. The full analysis can be found in ref [6] .



Taw2=MJ1557?? Orphan did not detect predictedyW-86

Predicted *Bacillus subtilis* tRNA modification genes

Score	Evidence type
5	In vitro with *B. subtilis* enzymes, in vivo *B. subtilis* mutant
4	Orthologous group with validated gene in other species, genetic data in *B. subtilis*
3	Orthologous group with validated gene in other species
2	Similarity with experimentally validated in other species, but paralogs
1	Prediction only



The analysis is based on presence of genes as well as LC/MS analysis of bulk tRNA from *Bacillus subtilis* 168. We predicted 41 modifications genes with evidence codes. Starred are essential modifications, in green are those that use a different family than *E. coli* to perform the same reaction, in red a re uncertain or missing enzymes.

This work was funded by the National Institutes of Health GM70641.

References/ Acknowledgements

1. El Yacoubi B, Bailly M, de Crécy-Lagard V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. Annu Rev Genet. **2012,** 46:69-95. PMID: 22905870.
2. Pereira M, Francisco S, Varanda AS, Santos M, Santos MAS, Soares AR. Impact of tRNA Modifications and tRNA-modifying enzymes on proteostasis and human Disease. Int J Mol Sci. **2018** 19(12). pii: E3738. PMID: 30477220.
3. Koh CS, Sarin LP. Transfer RNA modification and infection - Implications for pathogenicity and host responses. Biochim Biophys Acta Gene Regul Mech. **2018**, 1861(4):419-432 PMID: 29378328.
4. Addepalli B, Limbach PA. Pseudouridine in the anticodon of *Escherichia coli* tRNATyr(QΨA) is catalyzed by the dual specificity enzyme RluF. J Biol Chem. **2016,** 291(42):22327-22337 PMID: 27551044.
5. de Crécy-Lagard V, Boccaletto P, Mangleburg CG, Sharma P, Lowe TM, Leidel SA, Bujnicki JM. Matching tRNA modifications in humans to their known and predicted enzymes. Nucleic Acids Res. **2019**, in press, PMID: 30698754..
6. Yu N, Jora M, Solivio B, Thakur P, Acevedo-Rocha CG, Randau L, de Crécy-Lagard V, Addepalli B, Limbach PA. Transfer RNA modification profiles and codon decoding strategies in *Methanocaldococcus jannaschii*. J Bacteriol**. 2019**, in press, PMID:30745370.