FEBS Letters 584 (2010) 325–333

FEBS Letters

journal homepage: www.FEBSLetters.org

Review

Computational analysis of tRNA identity

David H. Ardell *

School of Natural Sciences and UC Merced Center for Computational Biology, University of California, Merced, CA 95343, USA

ARTICLE INFO

Article history:
Received 3 November 2009
Revised 20 November 2009
Accepted 20 November 2009
Available online 26 November 2009

Edited by Manuel Santos

Keywords:
Proteobacteria
tRNA
Determinant
Logo
Coevolution
Phylogeny

ABSTRACT

I review recent developments in computational analysis of tRNA identity. I suggest that the tRNA–protein interaction network is hierarchically organized, and coevolutionarily flexible. Its functional specificity of recognition and discrimination persists despite generic structural constraints and perturbative evolutionary forces. This flexibility comes from its arbitrary nature as a self-recognizing shape code. A revisualization of predicted Proteobacterial tRNA identity highlights open research problems. tRNA identity elements and their coevolution with proteins must be mapped structurally over the Tree of Life. These traits can also resolve deep roots in the Tree. I show that histidylation identity elements phylogenetically reposition *Pelagibacter ubique* within alpha-Proteobacteria.

1. tRNA identity: a model for evolutionary systems biology

There are more than 20 functional classes of tRNAs that mediate protein synthesis. Besides the canonical elongators, initiator tRNAs (charged with Methionine and then formylated, in Bacteria) initiate translation, while noncanonical elongators mediate expansion of the genetic code to the 21st and 22nd cotranslationally inserted amino acids, selenocysteine and pyrrolysine (reviewed in [1–3]). Within broad functional classes, different classes of isoacceptor tRNAs are targeted to interact with specific enzymes for covalent modification reactions or mediate programmed translational events at the ribosome [4,5]. Some isoacceptors participate directly in the biosynthesis of amino acids or other biochemical entities such as the cell wall [6–8].

This diverse portfolio of functions among tRNAs is achieved through specificity in RNA–protein (and RNA–RNA) interactions. The functional *identity* of a tRNA is achieved through its specific interactions with a unique, though not disjoint, clique of proteins, ribonucleoproteins and RNA complexes that catalyze specific biosynthesis, maturation, quality control, modification, amino acid charging, and targeting reactions (for recently updated reviews, see e.g. [9–11]). The tRNA–protein interaction network has a hierarchical structure. Some interactions, like with biosynthetic ribonucleases, EF-Tu, or the ribosome, are shared widely by many tRNA classes, while others are restricted to a few or only one, like in aminoacylation reactions.

This hierarchy sets up a conflict between interaction specificity, which favors structural divergence, and interaction uniformity with others, termed *conformity* [12], which may either constrain that divergence or be achieved through compensatory tuning, e.g. [13]. To accomodate this, tRNA identity is mediated not only by molecular recognition through *identity determinants*, but also molecular discrimination through *identity antideterminants*, structural features that block interactions with network members (Fig. 1).

My hypothesis is that these *tRNA identity elements*, along with the protein structural motifs they interact with, make up a theoretically arbitrary *molecular shape code*. Unlike metabolic enzymes selected to bind fixed small molecular substrates, tRNA–protein interactions achieve their specificities by dialectical convention. Because both the catalysts and their substrates are either genetically encoded or templated under conditions of constant evolutionary perturbation, the conventions that govern their interactions are not conserved. I expect that perturbations by mutation, selection for antibiotic resistance, horizontal gene transfer (HGT), genomic and stable RNA base content evolution, cross-talk with other network members, and other directional and randomizing forces orthogonal to function constantly buffet all players in the network. When purifying selection fails, because evolutionary reversals are rare, tRNAs and their binding proteins must maintain their interaction specificities against a slow but constant backbeat of coevolutionary divergence.

* Fax: +1 209228 2953.
E-mail address: dardell@ucmerced.edu

0014-5793/$36.00 © 2009 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/j.febslet.2009.11.084

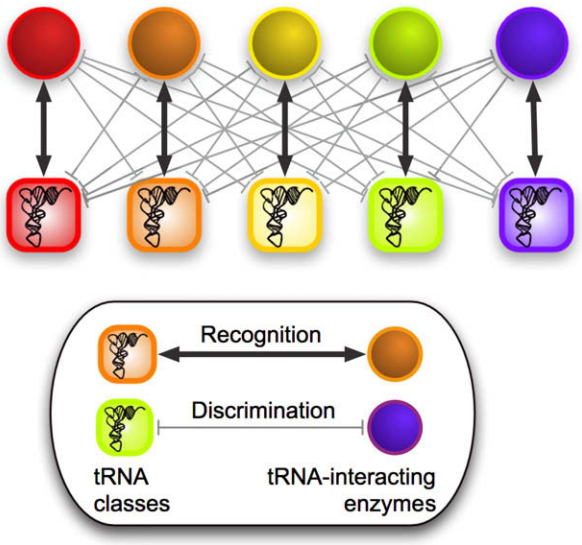

Fig. 1. A cartoon representation of part of a tRNA–protein interaction network. The L-shaped tRNA tertiary structure is derived from data in [14].

Indeed, many differences in tRNA identity elements have been experimentally characterized among the three domains descended from the Last Universal Common Ancestor, using model organisms. Differences in tRNA identity systems are thought to pose a barrier to interdomain horizontal gene transfer (HGT) of aminoacyl-tRNA synthetase genes [15,16]. Examples of HGT despite these barriers have been described, in at least one example probably in response to selection for antibiotic resistance (reviewed in [17]). Experimentally, tRNA identity elements have been "transplanted" or swapped between tRNA bodies, briefly reviewed in e.g. [18]. Complementary changes have been engineered in tRNAs and their cognate synthetases, e.g. [19,20]. However, to my knowledge, coupled swaps or rotation of both tRNA identity elements and the residues or domains that recognize those elements in proteins *among different cognate tRNA–protein pairs* has not been demonstrated, which would reinforce the arbitrary association of identity sets with specific functions.

The concept of tRNA identity as a code is an old one promulgated by several others as well. It nevertheless is a concept worth revisiting and perhaps recasting as follows: the tRNA–protein interaction network is an excellent model for the evolutionary systems biology of RNA–protein interactions. Because of the expense and difficulty of experimental characterization of non-model organisms, computational comparative genomics, structural biology, and data integration will continue to play an important role in investigating this model.

2. Computational analysis of tRNA sequence variation

By the very nature of its space limitations, this review must be incomplete. In particular, recent advances in the study of tRNA evolution or of organellar tRNAs are beyond the scope of this review. Developments in computational analysis of RNA gene-finding, structure, and function in general are available in [21–23].

2.1. tRNA gene-finding

Recent developments in computational analysis of RNA sequences is driven by the growth of genomics data. The first step is confidently annotating tRNA genes. Non-coding RNA gene-finding is more challenging computationally than protein-coding gene-finding [21], but both remain hot, active fields. Usually, specialists reannotate genomes for their genes of interest to use in specific studies. Room for improvements in gene-finding exists even for such relatively structurally conserved, compact and simple RNAs as tRNAs. Casey Bergman and I showed for annotation of tRNAs for the Drosophila Twelve Genomes Consortium [24] that using the union of predictions from two gene-finders, tRNAscan-SE [25] and ARAGORN [26] recovered more experimentally known *Drosophila melanogaster* tRNAs than either alone [27]. Recently, the creators of tRNAdb-CE [28] show independently that different tRNA gene-finders are complementary for prokaryotic genomic and metagenomic data, increasingly so with a third gene-finder called tRNAfinder [29]. Exhaustive analysis, for instance, in archaea, would include analysis of split tRNA genes arising in part from tRNA intronic variation [30,31].

My explanation for the advantage of using multiple tRNA gene-finders lies in two kinds of complementarity: different training data and different techniques and parameters for model generalization. To avoid risking loss in specificity, perhaps prediction overlaps should be large and the complementarity of differences should be mostly consistent and explicable.

There are also class-specific and taxonomic variation in tRNA start and end coordinates which needs more attention in genomic annotation. For instance, the 5′ end of bacterial Histidyl-tRNAs may not be correctly classified by some gene-finders currently in use. In addition, some gene-finders include bases 74–76 (in Sprinzl coordinates) regardless of whether they template CCA or not. Bacteria vary widely in their rates of genomically templating CCA [32].

2.2. Structural alignments of tRNAs

All comparative analysis of tRNA gene sequences necessitates their structural alignment, which creates a partial structural correspondence of bases and base-pairs over multiple tRNAs. Although structurally analogous parts of tRNAs need not be homologous (that is, derived from the same part in a common ancestral tRNA), structural alignments facilitate evolutionary and functional analyses. The classic work in this area is due to Sprinzl and co-workers who have also given us the standard tRNA coordinate system and recently updated genomic databases of aligned tRNAs [33]. Another resource for alignments of pre-defined tRNA sequences is the tRNAmart by Saks, Conery and collaborators, underlying their recent studies on functional sequence variation [34]. Their particularly careful attention to the alignment of type II tRNAs with longer variable arms merits special mention.

The most accessible way to create your own automated structural alignments of tRNA sequences is by using Covariance Models (CMs) [35,36], which align multiple RNA sequences to a probabilistic model that represents both primary and secondary structure. These models handle bulge, loop and stem insertions and deletions, but they do not mutually align different structural insertions relative to the model's consensus structure. Therefore, these models carry certain limitations in the structural alignment of larger-scale evolutionarily and functionally significant structural variation. For example, it is for this reason that tRNAscan-SE, which uses CMs, deploys a special model for selenocysteine-tRNAs [25]. It seems to me that improvements are called for even in modeling the universal structural variation present in variable arms, variable pockets and D-arms of standard elongators, over and above how they are currently modeled with CMs. Within the constraint of existing techniques, users who wish to make structural alignments of tRNAs could use Infernal [36] with a tRNA model from RFAM [23], or its predecessor, the COVE package, with the taxonomic domain-specific models distributed with tRNAscan-SE [25]. The