Saturation of recognition elements blocks evolution of new tRNA identities

Adélaïde Saint-Léger, Carla Bello, Pablo D. Dans, Adrian Gabriel Torres, Eva Maria Novoa, Noelia Camacho, Modesto Orozco, Fyodor A. Kondrashov, Lluís Ribas de Pouplana

Understanding the principles that led to the current complexity of the genetic code is a central question in evolution. Expansion of the genetic code required the selection of new transfer RNAs (tRNAs) with specific recognition signals that allowed them to be matured, modified, aminoacylated, and processed by the ribosome without compromising the fidelity or efficiency of protein synthesis. We show that saturation of recognition signals blocks the emergence of new tRNA identities and that the rate of nucleotide substitutions in tRNAs is higher in species with fewer tRNA genes. We propose that the growth of the genetic code stalled because a limit was reached in the number of identity elements that can be effectively used in the tRNA structure.

INTRODUCTION

The expansion of the genetic code requires the evolution of new transfer RNA (tRNA) identities cognate to previously nonproteogenic amino acids. A major challenge in synthetic biology is the development of orthogonal aminoacylation systems that efficiently integrate new tRNAs into established translation machinery. Similarly, the expansion of the genetic code during early steps in the origin of life required the emergence of new tRNA isoacceptors that allowed the system to evolve from a simple error-prone protein synthesis to a 20–amino acid code that transcribes genes with an error rate of approximately $1 \times 10^{-3}$ (1–3). An important question in this regard is why the universal genetic code did not grow beyond 20 amino acids.

We have investigated the emergence of a family of eukaryote-specific tRNA identity elements required for the modification of tRNAs by a tRNA modification enzyme. Heterodimeric adenosine deaminases acting on tRNAs (ADAT) are eukaryotic-specific enzymes that catalyze the deamination of adenosine 34 (A34) to inosine 34 (I34) at the first position of tRNA anticodons. The presence of I34 enables a single tRNA anticodon to recognize codons ending with C, U, and A. In bacteria, the activity of the ortholog of ADAT [tRNA-specific adenosine deaminase (Tada)] is limited to tRNAArg. In eukaryotes, TadaA underwent a process of gene duplication and divergence to give rise to ADAT, whose substrate range is increased to include all A34-containing tRNAs acting in three-, four-, and six-box codon sets, with the sole exception of tRNA^Gly^GCC (4).

RESULTS

We have previously shown that the duplication and subsequent divergence that turned the homodimeric bacterial TadaA into the heterodimeric eukaryotic ADAT contributed to the evolution of eukaryotic genomes in terms of both tRNA gene content and codon usage bias (5). In particular, the emergence of ADAT in eukaryotes correlates with a marked enrichment of previously nonexistent A34-containing tRNAs and the loss of their synonymous G34-containing tRNA isoacceptors (Fig. 1) (4, 5). Surprisingly, tRNA^Gly^ACC did not follow this evolutionary trend. In contrast to ADAT substrates, tRNA^Gly^ACC is absent in eukaryotic genomes, and tRNA^Gly^GCC remains the most abundant tRNA^Gly^ isoacceptor (Fig. 1).

We investigated the reasons for the intriguing lack of tRNA^Gly^ACC in eukaryotic genomes. First, we asked whether a fundamental functional defect could prevent tRNA^Gly^ACC from being used in eukaryotes. Børén et al. (6) had previously shown that a mutated Escherichia coli tRNA^Gly^ACC is active in elongation. To extend this analysis to eukaryotes, we modified the anticodon of the genomically predominant human tRNA^Gly^ isoacceptor (tRNA^Gly^GCC) to ACC and found that it can be effectively charged by glycyll-tRNA synthetase (GlyRS) (Fig. 2A). Further, we established that the I34 modification does not impair the aminoacylation of tRNA^GCC (Fig. 2A). These results are in agreement with the observation that GlyRS interactions at position 34 are not sequence-specific (7). Thus, the absence of tRNA^Gly^ACC in eukaryotes is not due to a fundamental structural or functional defect hampering its function in translation.

We then focused our attention on the potential deamination by ADAT of a tRNA^Gly^ACC, an activity that is dependent on tRNA structure (8). Through electrophoretic mobility shift assays (EMSA) (Fig. S1A), we first showed that human tRNA^Gly^ can be bound by ADAT independently of the sequence at position 34. We then used inosine-specific cleavage (9, 10) (Fig. S2A) and tRNA sequencing to detect the deamination of tRNA^GCC^ by ADAT (11, 12). Human tRNA^Ala^, a natural substrate of ADAT, was chosen as a control because of its high sequence similarity to tRNA^Gly^.

A control tRNA is completely modified by ADAT in 15 min, but tRNA^GCC^ is only partially deaminated after 60 min under the same conditions (Fig. 2 B and C). Similar results were obtained with tRNA^Gly^ derived from the other human tRNA^Gly^ isoacceptors or with in vivo–transcribed tRNA^Gly^ (fig. S2 B and C). Thus, tRNA^Gly^ is an extremely poor ADAT substrate.

We investigated the reasons for the low activity of ADAT with tRNA^Gly^ACC. Chimeric tRNAs based on combinations of human tRNA^Ala^ and tRNA^Gly^ were tested for deamination by ADAT and unequivocally
showed that the anticodon loop of tRNA^{Gly}_{ACC} prevents deamination, but not binding, by the enzyme (Fig. 2C and figs. S1B and S2D). Sequence alignments of tRNA substrates of ADAT with tRNA^{Gly}_{ACC} revealed that the anticodon loop of tRNA^{Gly}_{ACC} differs from tRNA^{Ala}_{AGC} only at positions 32, 35, and 38 (Fig. 2D). We tested the ability of these positions to block ADAT activity by individually mutating each base in the anticodon of tRNA^{Ala} to the sequence of the tRNA^{Gly}_{ACC} anticodon loop. Single point mutations did not significantly affect ADAT activity in tRNA^{Ala} (fig. S2D).

We then searched among human tRNAs for the ADAT substrate with the closest anticodon sequence to tRNA^{Gly}_{ACC}. We find that the tRNA^{Val}_{AAC} anticodon loop differs from tRNA^{Gly}_{ACC} only at position 35 (C^{35} in tRNA^{Gly}_{ACC} and A^{35} in tRNA^{Val}_{AAC}) (Fig. 2D). When an A35C mutation was introduced in the context of a tRNA^{Val}_{AAC}, this significantly reduced ADAT activity for this tRNA (fig. S2E). Thus, a cytidine in position 35 of tRNA^{Ala} does not significantly affect the activity of ADAT, but the same base prevents ADAT activity in the context of tRNA^{Val}. Similarly, the presence of A^{38} favors deamination of tRNA^{Gly} by ADAT, even though the sequence at position 38 is variable among ADAT substrates (Fig. 2D).

Our results indicate that the recognition of tRNA anticodons by ADAT is structure-dependent and that the inability of the enzyme to modify tRNA^{Gly} is not due to a specific sequence but rather to the structural context of this anticodon. To check this hypothesis, we performed state-of-the-art molecular dynamics (MD) simulations to analyze the relationship between sequence, structure, and motion in the anticodons of ADAT substrates and tRNA^{Gly}.

We first analyzed through 1 μs (10^9 integrations of Newton equations of motion) a control tRNA for which a crystallographic structure is available (tRNA^{Phe}_{CUC}) and found that the force field can reproduce the known structure. We then studied tRNA^{Gly}_{GCC} and tRNA^{Gly}_{GUC} (built from tRNA^{Phe}), which displayed full conservation of the structural signature of the anticodon loop obtained for tRNA^{Phe} (Fig. 3 and figs. S3 and S4).

The addition of A at position 34 in the two tRNA^{Gly} scaffolds previously shown to be stable caused a strong structural distortion of the anticodon (Fig. 3 and figs. S5 and S6), indicating that, in the context of the anticodon stem-loop of tRNA^{Gly}, the addition of A^{34} induces an important structural instability.

We then used the same approach to analyze the structural influence of A^{34} in the context of C^{35} or C^{36} (the ACC glycine anticodon) to determine whether the destabilizing effect of A^{34} seen in tRNA^{Gly} also appeared in other tRNAs. We in silico-constructed tRNA^{Gly}_{ACC} and showed that an A^{34}–C^{35}–containing anticodon remains stable during the entire microsecond of unbiased simulation (Fig. 3 and fig. S7). Similarly, simulations with tRNA^{Ala}_{AGC} showed that a fully stable structure was obtained with an A^{34}–C^{36}–containing anticodon (Fig. 3 and fig. S8).

Altogether, these results point to a sequence-independent role of positions 32, 35, and 38 in the rejection of tRNA^{Gly} by ADAT. Previous reports demonstrated the importance of positions 32, 35, 36, and 38 of tRNA^{Gly} for aminoacylation, modification, and accommodation in the ribosomal A-site (13–16). On the other hand, Auxilien et al. (8) demonstrated the importance of the anticodon conformation for ADAT activity. We find that the structural features of the tRNA^{Gly} anticodon loop are incompatible with the presence of an adenosine at position 34, explaining why an A^{34}–containing tRNA could not evolve and become enriched in eukaryotes after the emergence of ADAT in these species.

As shown, this distortion is not an impediment for aminoacylation. Additionally, the tRNA modification patterns in the anticodon stem-loop of ADAT substrates and of tRNA^{Gly} reveal that the former incorporate more modifications than the latter (fig. S9), again pointing to important structural differences between the structure of the tRNA^{Gly} anticodon and the same region in ADAT substrates.

We wondered whether the effect of signal saturation on tRNA evolution could be detected at the genomic level. We searched the genomic databases for pairs of related phylogenetic clades with different tRNA gene populations to test whether the rate of nucleotide substitution in tRNA sequences was higher in species with fewer tRNA genes. Indeed, we observed that bacterial species of the *Ehrlichia* genus (which contain 36 tRNA genes in their genome) have significantly more nucleotide differences between their orthologous tRNA pairs than species of...
the *Escherichia* and *Salmonella* genera, which have a larger number of tRNA genes (89 and 86 tRNA genes, respectively) than *Ehrlichia* (Fig. 4), a result fully consistent with our hypothesis.

Signal saturation may also limit the evolution of genomic tRNA populations in extant species. If preexisting recognition signals block the evolution of new tRNA identities, then one would expect that tRNA sequences would evolve faster in genomes with smaller numbers of tRNA genes. An example of this situation can be found in mitochondria, whose genomes have low numbers of tRNA genes (17). This reduction in tRNA genes correlates well with the frequent variations in the genetic code in these organelles and with the variability in identity elements and aberrant structures of tRNAs. For example, universally conserved identity elements in cytosolic tRNA\(^\text{Asp}\) (G\(^73\)), and tRNA\(^\text{Ala}\) (G\(^7\)/U\(^7\)) are lost in animal mitochondrial tRNAs (18, 19).
DISCUSSION

Our proposal predicts that systems with saturated tRNA signals would need to evolve different incorporation machineries to add new amino acids to their genetic code. The case of pyrrolysine would seem to contradict this prediction, but pyrrolysine displays a very limited distribution in the phylogenetic tree (20) in (almost exclusively) species that contain very simplified tRNA populations (5). Selenocysteine (Sec), on the other hand, is a proteogenic amino acid that is almost universally incorporated to proteins but not through direct aminoacylation of a cognate tRNA. Instead, Sec incorporation requires an alternative mechanism that depends on the use of a serylated tRNA and a complex set of dedicated components (21). Determining the exact functional constraints that limit the growth of the genetic code would be useful for the design of orthologous tRNA–aminoacyl-tRNA synthetase (ARS) pairs to expand the proteome composition in engineered organisms.

**Fig. 3.** A$^{34}$ destabilizes tRNA$^{Gly}$ anticodon. (A) Initial structure used to build all the models analyzed by MD. The figure corresponds to the anticodon stem-loop of tRNA$^{Phe}$ [Protein Data Bank (PDB) ID 2TRA] (residues 27 to 43). (B) Final conformations of the studied anticodons after 1 μs of unbiased MD. The two top figures correspond to the final structures obtained for the tRNA$^{Gly}$ isoacceptors (with G$^{34}$ and C$^{34}$). The two middle structures illustrate the effect of introducing A at position 34 in tRNA$^{Gly}$ anticodons. Red arrows point to the more evident distortions. The two bottom structures show that presence of A at position 34 does not affect the stability of tRNA$^{Arg}$ and tRNA$^{Ala}$ anticodons. (C) Molecular superposition of the initial and final structures, obtained by aligning the phosphate groups. The superimposition above shows the similarity between the initial and final conformations of the simulation for the stable structures. Note the correct orientation of the anticodon loop triplet (residues 34 to 36). The superposition at the bottom highlights the magnitude of the structural distortion in the region of the loop, induced by the presence of A$^{34}$ in the context of tRNA$^{Gly}$ isoacceptors.

**Fig. 4.** tRNA sequence divergence is higher in the genome with reduced number of tRNA genes. *Ehrlichia ruminantium* str. Welgevonden and *Ehrlichia canis* str. Jake tRNA genes have significantly more nucleotide differences between their orthologous pairs than the orthologous tRNA pairs between *Salmonella enterica* subspecies I, serovar Typhimurium strain LT2 (*Salmonella typhimurium* str. LT2) versus *E. coli* CFT073. The divergence time of the two pairs of species was predicted to be similar based on 16S rRNA divergence (40- and 37-nt differences for each pair, respectively). See also table S1.
for processing, modification, transport, and other functions. On the basis of the biochemical, structural, and genomic data presented here, we propose that the accumulation of structural and functional constraints that operate on tRNAs acts as an operational limit that impedes the evolution of new tRNA identities and, as a result, the incorporation of new amino acids into the genetic code.

**MATERIALS AND METHODS**

**Experimental design**
To investigate the reasons for the absence of tRNA\textsubscript{Gly} \textsuperscript{ACC} in eukaryotic genomes, we first tested the hypothesis that functional defects could prevent this tRNA from being used. tRNA\textsubscript{Gly} \textsuperscript{ACC} and the other tRNAs mentioned in this paper were transcribed, in vivo or in vitro, and tested for aminoacylation by GlyRS and deamination by ADAT. We then conducted MD simulations to analyze the relationship between sequence, structure, and motion in the anticodons of ADAT substrates and tRNA\textsubscript{Gly}. Finally, we studied the rate of nucleotide substitution in tRNA sequences from phylogenetic clades with different tRNA gene populations (Escherichia, Salmonella, and Ehrlichia genera) to test whether this rate was higher in species with fewer tRNA genes.

**Enzyme cloning, expression, and purification**
Human cytoplasmic alanyl-tRNA synthetase (AlaRS) and GlyRS complementary DNAs (cDNAs) (NM_001605 and NM_002047, respectively) were cloned into a pQE-70 vector (Qiagen) for bacterial expression (BL21 (DE3)) cells (Life Technologies), transformed with pQE70-AlaRS or pQE70-GlyRS, were grown at 37°C up to an OD\textsubscript{600} (optical density at 600 nm) of 0.6. Protein expressions were assayed by aminoacylation assay. Human GlyRS enzyme was purified by affinity chromatography with a StrepTrap HP column (GE Healthcare). ADAT2 and ADAT3 were purified at an equimolar ratio. The purified heterodimer ADAT was stored in 50 mM Hepes (pH 8), 100 mM KCl, 1 mM MgCl\textsubscript{2}, 0.1 mM EDTA, and 2 mM dithiothreitol (DTT) with 20% glycerol.

**tRNA sequences**
The original tRNA sequences were from the Genomic tRNA Database (25) (http://gtrnadb.ucsc.edu).

**In vitro tRNA preparation**
Wild-type or mutated tRNAs were assembled using six DNA oligonucleotides that were first annealed, and then ligated, between Hind III and Bam HI restriction sites of plasmid pUC19. Bst NI–linearized plasmids were used to perform the in vitro transcription with T7 RNA polymerase, according to standard protocols (26). Transcripts were separated by 8 M urea/10% polyacrylamide gel electrophoresis. The tRNA was identified by ultraviolet (UV) shadowing, electroeluted, and ethanol-precipitated. The tRNA pellet was resuspended in H\textsubscript{2}O and stored at −20°C.

**In vivo tRNA preparation**
*E. coli* BL21(DE3) cells expressing the plasmid pUC19-tRNA\textsubscript{Gly} or pUC19-tRNA\textsubscript{Val}, coding for human tRNA\textsubscript{Gly} \textsuperscript{ACC} or tRNA\textsubscript{Val} \textsuperscript{AAC}, were cultured in LB broth supplemented with ampicillin (100 mg/ml) at 37°C to a cell density of OD\textsubscript{600} = 0.6. Expression of tRNA was induced with
1 mM IPTG for 4 hours. Bacterial pellets were resuspended in 0.3 M NaOAc (pH 4.3) and 10 mM EDTA and extracted with water-saturated phenol (pH 4.8). The aqueous fractions were ethanol-preincubated, resuspended in FDM loading buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol), and resolved using 8 M urea/10% polyacrylamide gel electrophoresis. The tRNA was identified by UV shadowing, electroleuted, and ethanol-preincubated. The tRNA pellet was resuspended in H2O and stored at −20°C.

**Aminoacylation assays and tRNA binding experiment**

Aminoacylation of tRNA was performed at 37°C in 100 mM Hepes (pH 7.2), 25 μM amino acid, 30 mM MgCl2, 30 mM KCl, 0.5 mM DTT, 5 mM ATP, bovine serum albumin (0.1 mg/mL), [2-3H]-glycine (500 Ci/mol) or 1-[3-3H] alanine (500 Ci/mol) (PerkinElmer), and 5 μM tRNA transcripts (refolded at 85°C followed by gradual reduction of temperature). Reactions were initiated by the addition of a crude cell extract or 1 μM pure human GlyRS. Samples (22 μL) were spotted onto Whatman 3MM discs at varying time intervals. Radioactivity was measured by liquid scintillation.

For EMSA, dephosphorylated tRNAs were labeled with [γ-32P]ATP (PerkinElmer) using T4 polynucleotide kinase, refolded at 70°C followed by gradual reduction of temperature, and incubated for 15 min at 4°C with 50 mM Hepes (pH 7.5), 1 mM MgCl2, 50 mM KCl, 0.5 mM DTT, 2% (w/v) glycerol, 2.5 μM oligo(dt)25, and 0.05, 0.5, 1.5, or 3 μM ADAT. Reactions were separated by electrophoresis onto a 10% polyacrylamide gel (19:1) in 0.5× tris-borate EDTA. The signal was digitalized using a PhosphorImager from a gel-exposed storage phosphor screen.

**In vitro deamination assays**

For the deamination assays, dephosphorylated tRNAs were labeled with [γ-32P]ATP (PerkinElmer) using T4 polynucleotide kinase, refolded at 70°C following by gradual reduction of temperature in the presence of 2 mM MgCl2, and incubated for 1 hour at 37°C with 42 μM purified ADAT in 50 mM tris-HCl (pH 8.0), 25 mM KCl, 2.5 mM MgSO4, 0.1 mM EDTA (pH 8.0), and 2 mM DTT. Reactions were purified with the Viogene miTotal RNA mini kit (catalog number VTR1002), and tRNA was eluted from the purification column with DEPC-treated water. Cleavage of synthetic tRNA was monitored by electrophoresis on a 10% polyacrylamide gel. The inosine-specific cleavage reaction for the in vivo–transcribed tRNAs was performed as above except that the tRNAs were not 32P-labeled but were detected by Northern blot with a 32P-labeled specific probe. The gel images were analyzed and quantified with ImageJ software.

**RT-PCR and sequencing**

Two micrograms of tRNA (previously incubated with or without ADAT) were retrotranscribed using the High-Capacity cDNA Reverse Transcription Kit (part number 4368814) following the manufacturer’s protocol. The primer used for the RT is oligonucleotide number 6, which is used for the in vitro tRNA preparation of the corresponding tRNA. The thermal cycler conditions were as follows: 10 min at 25°C, 60 min at 37°C, and 5 min at 85°C. Following the retrotranscribed reaction, a standard PCR was performed using 2 μl of cDNA. The primers used were oligonucleotides numbers 1 and 6 of the corresponding tRNA. The PCR cycle was as follows: 94°C, 30 s at 55°C, and 90 s at 72°C (35 cycles); 10 min at 72°C (1 cycle). The PCRs were purified using the NucleoSpin Gel and PCR Clean-up kit (reference number 740609.250; Macherey-Nagel). Samples were sent for sequencing using the same oligos used for the PCR as sequencing primers.

**Selection of bacterial species for analysis (few versus many tRNAs)**

To evaluate the hypothesis that the constraints in tRNA variation depend on the total number of tRNA genes a species has, we selected bacterial species with either fewer or many tRNA genes. To evaluate the number of tRNA genes of all the species of bacteria, we made a distribution of their relative numbers using the Genomic tRNA Database (http://grtnadb.ucsc.edu).

16S ribosomal RNA species evolutionary divergence

The 16S ribosomal RNA (rRNA) gene for all candidate species of bacteria (with either fewer or many tRNA genes) was analyzed was downloaded from the National Center for Biotechnology Information (NCBI). We estimated the evolutionary divergence of the species by pairwise sequence alignment of the 16S rRNA gene using T-Coffee multiple sequence alignment package version 11.00.8cbe486 and selected those for which the predicted time of divergence was similar (40- and 37-nt differences for each respective pair of species).

We selected E. canis str. Jake and E. ruminantium str. Welgevonden as the representative pair of species with few tRNA genes and E. coli CFT073 and S. enterica subspecies I, serovar Typhimurium strain LT2 (S. typhimurium LT2) as the species with many tRNA genes.

**tRNA orthologs pairwise comparison**

The orthology of the tRNA genes for each pair of species was preliminarily assessed with the best reciprocal hit (BRH) approach by using
BLAST (Basic Local Alignment Search Tool) package version 2.2.28. In addition, to fully resolve the orthology between the pairs of tRNA genes, we constructed a synteny map between each pair of species using their complete genomes.

The complete genome of all species analyzed and their features were downloaded from NCBI. We performed a full-genome alignment between each pair of species (E. canis str. Jake versus E. ruminantium str. Welgevonden and E. coli CFT073 versus S. typhimurium LT2) using the progressive Mauve alignment of the Mauve genome alignment package version 2.3.1 with default parameters and inferred the positional orthologs of the tRNA genes with a minimum coverage of 90% aligned sequence and 90% identity. The flanking genes of each tRNA gene pair in the full-genome alignment of each pair of species had to be the same to be considered syntenic.

Once the orthology was resolved by BRH and synteny, we carried out pairwise alignments of each tRNA ortholog gene from each selected species group using a combination of T-Coffee multiple sequence aligner and its in-built R-Coffee secondary structural alignment package with additional manual curation and calculated the number of differences between the two (table S1).

MD simulations: Limitations and system setup
Predicting the structure and folding of RNA has been particularly challenging using all-atom MD simulations for a number of reasons. (i) Instead of the few hydrogen bond (hbond) pairings observed in DNA, 38 different types of canonical and noncanonical pairing were observed in RNA molecules (27). (ii) This wide variety of distinct possible pairings between RNA bases often involves long-range interactions (even thousands of nucleotides away) (28). (iii) Structural transitions leading to the final folded tertiary structures of RNA are slow (seconds to hours), whereas base pair melting, reshuffling, and tertiary pairing occur in the microsecond to second time scale (29). (iv) Only in tRNA molecules, 107 naturally occurring chemical modifications were identified (30), providing new functional and structural characteristics of canonical nucleobases. For these reasons, we focused our simulations only on the anticodon loop that is composed of a single-stranded RNA of 17 bp, which folds into a stem-like A-RNA, formed by 5 bp stabilized with Watson-Crick (WC) hbinds. The remaining seven nucleobases form a loop stabilized by stacking and hbind interactions, some specific to tRNA anticodon loops (31). Using this reduced system, only a few modified nucleobases are needed, for which the specific force field parameters are known. Moreover, small structural rearrangements occur in the nanosecond time scale (32) where modern force fields have already proven to correctly predict the folded state of small RNA, including loops (33, 34).

All systems were built using, as starting geometry, the high-resolution x-ray structure of tRNA<sup>Phe</sup> with PDB code 2TRA (35). Each specific sequence studied was obtained by mutating residues 27 to 43 in tRNA<sup>Phe</sup> (corresponding to the anticodon loop) using ModelRNA (36). The systems were then solvated by adding TIP3P waters in a truncated octahedral box (10 Å around the solute) and neutralized by adding 16 K+.

Once the orthology was resolved by BRH and synteny, we carried out pairwise alignments of each tRNA ortholog gene from each selected species group using a combination of T-Coffee multiple sequence aligner and its in-built R-Coffee secondary structural alignment package with additional manual curation and calculated the number of differences between the two (table S1).

**MD simulation details**
All systems were simulated in the isothermal-isobaric ensemble (P = 1 atm, T = 298 K) using the Berendsen algorithm (39) to control the temperature and the pressure, with a coupling constant of 5 ps. Center-of-mass motion was removed every picosecond to limit the translational kinetic energy of the solute. SHAKE (40) was used to keep all bonds involving hydrogen at their equilibrium values, which allowed us to use a 2-fs step for integration of Newton equations of motion. Long-range effects were accounted for by using the particle mesh Ewald method (41) with standard defaults and a real-space cutoff of 10 Å. The parmbsc0-OL3 force field (42, 43) was used to represent RNA interactions. All simulations were done using the pmemd CUDA code (44) of AMBER 14.

**Structural signatures of “active” tRNA loops**
The crystal structure of the fully unmodified yeast tRNA<sup>Phe</sup> revealed the structural signatures of what are now considered to be common RNA structural motifs of functional tRNA anticodon loops. In this sense, our structural and dynamic analysis was centered on three different structural signatures thought to be present in functional anticodon loops. (i) A marked backbone turn on the loop region, stabilized by one or more intramolecular hbinds (45). Some of these hbinds have been well characterized (U-turn and A-turn) and are also present in other types of RNA molecules (46). This is particularly true for the U-turn motif observed in the crystal forms of several functional tRNAs, such as tRNA<sup>Phe</sup><sub>GAA</sub> (47), tRNA<sup>Phe</sup><sub>GCA</sub> (48), and tRNA<sup>Phe</sup><sub>UUU</sub> (49). (ii) Y<sup>32</sup> with the χ angle in the anti conformation and correctly stacked with the next base (5′-3′ direction). (iii) A correct relative of the anticodon triplet, pointing toward the wide groove (bases 34 to 36) (35). Apart from these specific features, the correct pairing of the bases in the stem (mostly WC and wobble hbinds) and the fluctuation of the stem around the canonical A-form were also considered as structural signatures required for active anticodon loops.

**Tools for the analysis of MD simulations**
During production runs, data were collected every 1 ps (leading to a collection of 1 million structures per sequence), which allowed us to study infrequent but fast movements. All the trajectories were postprocessed with the CPPTRAJ module of the AMBERTOOLS 15 package (50), the NAFlex server (51), and local tools developed in the group (http://mmmb.pcb.ub.es/www/tools). The subsates of the torsional angles of the backbone (α and ζ) were categorized following the standard definition: gauche positive (g<sup>+</sup>) = 60° ± 40°, trans (t) = 180° ± 40°, and gauche negative (g<sup>−</sup>) = 300° ± 40°. All the hbind distances (calculated with CPPTRAJ) were measured from the donor and acceptor heavy atoms and can be considered to be formed below 3.5 Å. The backbone turn, the stacking interactions at position 32, and the relative orientation of the anticodon trimer were addressed by visual inspection of the trajectories and by creating molecular graphics using VMD (52). All the trajectories were clustered using the average linkage clustering algorithm (53) based on the root mean square deviation of the backbone-heavy atoms to quantify the time “spent” in each conformational substrate. The conformation of the stem was analyzed in the helical space along simulations using Curves+ (54).
Acknowledgments: We thank D. Söll, H. Grosjean, and L. Filonava for comments and suggestions. M.O. and P.D.D. thank the Barcelona Supercomputing Center for CPU/GPU time on MareNostrum/MinoTauro. P.D.D. is a PEDECIBA (Programa de Desarrollo de las Ciencias Básicas) and an SNI (Sistema Nacional de Investigadores) (ANII, Uruguay) researcher. Funding: This work was supported in part by the Spanish Ministry of Economy and Competitiveness (grants BIO2012-32200, Sev-2012-0208, and BIO2012-32868 to L.R.d.P., F.A.K., and M.O., respectively) and by the Catalan Government (grants 2014-SGR-0771, 2014-SGR-0974, and 2014-SGR-0134 to L.R.d.P., F.A.K., and M.O., respectively). This work was also supported by the Howard Hughes Medical Institute International Early Career Scientist Program (55007424), by a European Research Council (ERC) Starting Grant (335980_EinME to F.K.), and by a grant from the ERC (ERC_SimDNA to M.O.). A.G.T. and C.B. are funded by the Spanish Ministry of Economy and Competitiveness (FPDI-2013-17742 and BES-2013-064004, respectively). Author contributions: A.S.-L., E.M.N., and N.C. cloned the constructs, purified the tRNAs and the recombinant GlyRS, and performed all biochemical assays. E.M.N. helped in the analysis of the tRNA gene frequencies. A.S.-L. collected and analyzed data with L.R.d.P. C.B. collected, processed, and (with F.A.K.) analyzed orthologous tRNA pairs. P.D.D. performed and (with M.O.) analyzed the MD simulations. A.G.T. analyzed the posttranscriptional modification patterns. L.R.d.P. designed the study. L.R.d.P. and A.S.-L. wrote the paper with contribution from the rest of the authors. All authors discussed the results and commented on the manuscript. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All the data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 18 December 2015
Accepted 31 March 2016
Published 29 April 2016
10.1126/sciadv.1501860

Saturation of recognition elements blocks evolution of new tRNA identities
Adélaïde Saint-Léger, Carla Bello, Pablo D. Dans, Adrian Gabriel Torres, Eva Maria Novoa, Noelia Camacho, Modesto Orozco, Fyodor A. Kondrashov and Lluís Ribas de Pouplana (April 29, 2016)
doi: 10.1126/sciadv.1501860

This article is publisher under a Creative Commons license. The specific license under which this article is published is noted on the first page.

For articles published under CC BY licenses, you may freely distribute, adapt, or reuse the article, including for commercial purposes, provided you give proper attribution.

For articles published under CC BY-NC licenses, you may distribute, adapt, or reuse the article for non-commercial purposes. Commercial use requires prior permission from the American Association for the Advancement of Science (AAAS). You may request permission by clicking here.

The following resources related to this article are available online at http://advances.sciencemag.org. (This information is current as of May 12, 2016):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://advances.sciencemag.org/content/2/4/e1501860.full

Supporting Online Material can be found at:
http://advances.sciencemag.org/content/suppl/2016/04/26/2.4.e1501860.DC1

This article cites 52 articles, 16 of which you can be accessed free:
http://advances.sciencemag.org/content/2/4/e1501860#BIBL