HHS Public Access

Author manuscript

*J Am Chem Soc*. Author manuscript; available in PMC 2017 March 17.

Published in final edited form as:

*J Am Chem Soc*. 2016 October 05; 138(39): 12948–12955. doi:10.1021/jacs.6b07121.

# Evolutionary gain of alanine mischarging to non-cognate tRNAs with a G4:U69 base pair

**Litao Sun**[1], **Ana Cristina Gomes**[2], **Weiwei He**[3], **Huihao Zhou**[1], **Xiaoyun Wang**[2], **David W. Pan**[2], **Paul Schimmel**[1], **Tao Pan**[2,*], and **Xiang-Lei Yang**[1,3,*]

[1]Department of Cell and Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

[2]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637, USA

[3]Department of Chemical Physiology, The Scripps Research Institute, La Jolla, CA 92037, USA

## Abstract

Fidelity of translation, which is predominately dictated by the accuracy of aminoacyl-tRNA synthetases in pairing amino acids with correct tRNAs, is of central importance in biology. Yet, deliberate modifications of translational fidelity can be beneficial. Here we found human and not *E. coli* AlaRS has an intrinsic capacity for mis-pairing alanine onto non-alanyl-tRNAs including $tRNA^{Cys}$. Consistently, a cysteine-to-alanine substitution was found in a reporter protein expressed in human cells. All human AlaRS-mischarged tRNAs have a G4:U69 base pair in the acceptor stem. The base pair is required for the mischarging. By solving the crystal structure of human AlaRS and comparing it to that of *E. coli* AlaRS, we identified a key sequence divergence between eukaryotes and bacteria that influences mischarging. Thus, the expanded tRNA specificity of AlaRS appears to be an evolutionary gain-of-function to provide posttranscriptional alanine substitutions in eukaryotic proteins for potential regulations.

## TOC image

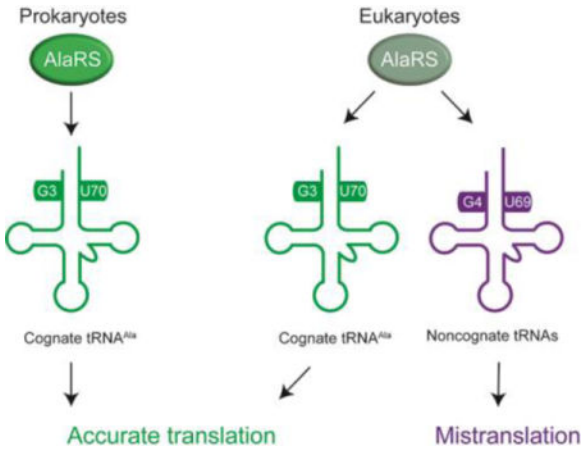
[*]Correspondence should be addressed to X.-L.Y. (xlyang@scripps.edu) and T.P. (taopan@uchicago.edu).

**Supporting Information**

Experimental methods, supplementary figures and supplementary tables.

**Accession codes**

Coordinates and structure factors have been deposited in the Protein Data Bank under accession code PDB 5KNN.




## Introduction

Aminoacyl-tRNA synthetases are responsible for establishing the genetic code, by pairing each amino acid with their corresponding tRNAs harboring the anticodon trinucleotides that ‘translate’ mRNAs into proteins. The accuracy of tRNA synthetases in charging the correct amino acids with their cognate tRNAs predominately dictates the fidelity of translation which, when compromised, could cause protein misfolding and/or loss-of-function and impose detrimental consequences[1]. However, deliberate modifications of translational fidelity also occur (with and without codon reassignment), especially when cells are stressed and undergo adaptation to environmental changes[2,3]. A prominent example is found in mammalian cells when they are exposed to innate immune and chemically triggered oxidative stress, which induces frequent mischarging of methionine onto non-methionyl-tRNAs and results in increased methionine incorporation into proteins[4]. Given the scavenging capacity of methionine for oxidative stress-induced reactive oxygen species, the mistranslation acts as a protective response against oxidative stress. Methionyl-tRNA synthetase (MetRS) appears responsible for the mischarging and, interestingly, its mischarging capacity is conserved throughout evolution[4–6] and is tunable to environmental cues[7,8].

We wondered whether the tRNA mischarging capacity is unique to MetRS. Discovery of the mischarging of MetRS was facilitated by the availability of [$^{35}$S]-Met for cell-based labeling of Met-charged tRNAs and the development of microarrays to detect all tRNA species within an organism[4]. Although mischarging was not detected with 5 other radiolabeled amino acids (i.e., Cys, Ile, Phe, Val, and Tyr), most amino acids are not available with sufficiently high specific radioactivity to be used for cell-based studies[4]. For instance, pulse labeling experiments with HeLa cells using [$^{3}$H]-Ala and [$^{3}$H]-Gly were unable to determine the absence or presence of mischarging due to insufficient signals obtained even for cognate tRNAs (data not shown).

We thus considered addressing this question from the angle of tRNA recognition. Perhaps an intrinsic capacity of a tRNA synthetase to mischarge is inversely related to the extent of specific interactions between the synthetase and its cognate tRNAs. Extensive studies

established tRNA identity elements for aminoacylation[9]. For the majority of tRNAs, both the trinucleotide anticodon and the acceptor stem are critical for synthetase recognition. For example, the single-stranded 'discriminator' nucleotide A73 and the C1:G72 base pair in the acceptor stem, and the GUA/AUA anticodons of human $tRNA^{Tyr}$ are used for recognition by the cognate tyrosyl-tRNA synthetase (TyrRS)[9,10] (Fig. 1a). Similarly, A73 and G1:C72 and C2:G71 in the acceptor stem, and the CCC/GCC/UCC anticodons are important for human $tRNA^{Gly}$ recognition[9,10] (Fig. 1a). With rare exception[11], the identity elements in $tRNA^{Tyr}$ and $tRNA^{Gly}$ involve both the acceptor stem and the anticodon throughout evolution. In contrast, the recognition of human alanyl-tRNA synthetase (AlaRS) only involves a single G3:U70 wobble base pair in the acceptor stem of $tRNA^{Ala}$, whereas the anticodon is not recognized at all[9,10] (Fig. 1a). The G3:U70 base pair also predominantly dictates the identity of $tRNA^{Ala}$ for aminoacylation throughout evolution[12]. If the tendency for mischarging is related to the amount of tRNA synthetase-tRNA recognition constraints, AlaRS may have a higher probability to mischarge than TyrRS and GlyRS.

To test this hypothesis, we used three recombinant human aminoacyl-tRNA synthetases to charge total human tRNA. Indeed, we found that AlaRS, but not TyrRS and GlyRS, mischarges non-cognate tRNAs. Notably, all Ala-mischarged non-cognate tRNAs contain a G4:U69 base pair in the acceptor stem. We demonstrated that the presence of this base pair is critical for mischarging. Unlike the situation for MetRS, the expanded tRNA specificity is unique to the human enzyme and not found in *E. coli* AlaRS. By solving the crystal structure of human AlaRS and comparing it to that of *E. coli* AlaRS, we identified a key sequence divergence between eukaryotes/archaea and bacteria that affects the mischarging. Thus, the mischarging of G4:U69-tRNAs by human AlaRS appears to be an evolutionary gain-of-function rather than an accidental mistake.

## Results

### Human AlaRS but not TyrRS and GlyRS mischarge non-cognate human tRNAs

To identify potential tRNA mischarging, we purified recombinant human cytoplasmic AlaRS, TyrRS, and GlyRS and used them to aminoacylate total human tRNAs isolated from HeLa cells with [$^3$H]-alanine, [$^3$H]-tyrosine, [$^3$H]-glycine, respectively. Detection of charged cognate and non-cognate tRNAs was carried out using a tRNA microarray which contains probes for all human mitochondrial-encoded tRNAs and collectively all chromosomal-encoded human tRNAs[4]. Remarkably, for human AlaRS, in addition to the cognate $tRNA^{Ala}$ species, we readily detected $^3$H signals for cytoplasmic $tRNA^{Cys}$(GCA), $tRNA^{Thr}$(CGT/IGT) (but not $tRNA^{Thr}$(TGT)), and mitochondrial $tRNA^{Asp}$ (mt-$tRNA^{Asp}$) (Fig. 1b). No $^3$H-signals aside from cognate tRNAs were detected with human TyrRS and GlyRS, indicating that TyrRS and GlyRS do not misacylate under our conditions (Fig. 1d,e). For GlyRS, we also tested a truncated form GlyRS-ΔWHEP. Removal of the metazoan-specific WHEP domain induces a conformational opening[13] without affecting aminoacylation activity[14]. Like full-length GlyRS, the GlyRS-ΔWHEP did not mischarge non-cognate tRNAs (Fig. 1f). These results are consistent with the earlier report that tRNA mischarging is limited to certain amino acids[4] and with our hypothesis that the capacity for mischarging may be inversely correlated with the extent of tRNA recognition constraints.

We further investigated tRNA misacylation by human AlaRS. We performed crosshybridization control by adding a large excess of all $tRNA^{Ala}$ probes in array hybridization (Fig. 1c). The unblocked and blocked arrays were exposed on the same imaging plate for direct quantitative analysis. Although the total $tRNA^{Ala}$ signals dropped by almost 500-fold as expected (Fig. 1b,c), the $^{3}H$-signals for $tRNA^{Thr}$(CGT/IGT) and mt-$tRNA^{Asp}$ changed by <2-fold, indicating that the signals from these tRNAs are indeed from Ala-$tRNA^{Thr}$ and Ala-mt-$tRNA^{Asp}$ misacylated by the human AlaRS. (It should be noted that the mischarging of the mitochondrial $tRNA^{Asp}$ by the cytoplasmic AlaRS might not occur in cells. Mitochondrial tRNAs are specifically charged by another set of aminoacyl-tRNA synthetases that, in most cases are distinct from their cytoplasmic counterparts.)

Although the $^{3}H$-signal for cytoplasmic $tRNA^{Cys}$(GCA) is diminished by ~80-fold after the array is blocked by the $tRNA^{Ala}$ probes, it is still significantly less than the 500-fold decrease of the $tRNA^{Ala}$ signals (Fig. 1c). Both $tRNA^{Ala}$ and $tRNA^{Cys}$ gene families in the human genome are among the most complex for all tRNA families[15]. The $tRNA^{Ala}$ and $tRNA^{Cys}$ families consist of 34 and 30 genes with 21 and 24 different but very similar sequences, respectively (tRNAScan score >50) (Supplementary Fig. 1 and Supplementary Fig. 2a). At the resolution of ~8 nucleotide differences for our tRNA microarray[16], the entire chromosomal $tRNA^{Ala}$ family is represented by only 3, and the $tRNA^{Cys}$ family by only 2 probes. Despite that the diminished $^{3}H$-signal for $tRNA^{Cys}$(GCA) indicates crosshybridization of $tRNA^{Ala}$ to the Cys-tRNA probe on the array, the remaining signal should still represent a small amount of mischarged Ala-$tRNA^{Cys}$.

## G4:U69 base pair in tRNA is responsible for human AlaRS mischarging

Among the human $tRNA^{Cys}$ genes, one contains a G4:U69 base pair, whereas all others contain G4:C69 or C4:C69 in the acceptor stem (Supplementary Fig. 2a). Among the 20 human $tRNA^{Thr}$ genes, 4 of 9 AGT (modified to IGT) anticodon genes and 2 of 5 CGT anticodon genes contain G4:U69, whereas all others contain C4:G69, U4:G69, or U4:A69 (Supplementary Fig. 2b). None of the $tRNA^{Thr}$-TGT anticodon genes, which are not mischarged, contains G4:U69 (Supplementary Fig. 2b). Interestingly, the mischarged mt-$tRNA^{Asp}$ gene also has G4:U69 (Supplementary Fig. 2c). As expected, all cytoplasmic $tRNA^{Ala}$ genes have G3:U70 (Supplementary Fig. 1). In fact, except for $tRNA^{Cys}$ and $tRNA^{Thr}$, no other human cytoplasmic tRNAs has G4:U69 and not G3:U70, suggesting that the G4:U69 base pair, although shifted from the G3:U70 position in $tRNA^{Ala}$, may still be responsible for the mischarging. Therefore, we transcribed *in vitro* genes encoding the G4:U69-$tRNA^{Cys}$, a G4:C69-$tRNA^{Cys}$, and a G4:U69-$tRNA^{Thr}$(AGT) (noted as G4:U69-$tRNA^{Thr}$ for simplicity) to test for mischarging by AlaRS (Fig. 2a). We also transcribed the gene for a cognate $tRNA^{Ala}$ as the positive control (Fig. 2a). The aminoacylation result confirmed both G4:U69-$tRNA^{Cys}$ and G4:U69-$tRNA^{Thr}$, but not G4:C69-$tRNA^{Cys}$, can be mischarged (Fig. 2b). The mischarging activity of AlaRS on the G4:U69-$tRNA^{Cys}$ transcript was particularly robust (Fig. 2b).

To confirm that the G4:U69 base pair is responsible for mischarging, we mutated the G4:U69 base pair in the G4:U69-$RNA^{Cys}$ to G4:C69. The mischarging activity was completely lost (Fig. 2c). Reversely, we mutated the non-mischarged G4:C69-$tRNA^{Cys}$ to

G4:U69. It gained robust mischarging activity (Fig. 2c). These results clearly demonstrated that mischarging of non-cognate tRNAs by human AlaRS is dependent on the G4:U69 base pair.

## Cys-to-Ala mistranslation found in human cells

To determine that the mischarged Ala-tRNA$^{Cys}$ or Ala-tRNA$^{Thr}$ can be used for translation in cells, the mass spectrometry analysis of a flag-CaMKII protein expressed and purified from HEK293T cells was searched for Cys-to-Ala and Thr-to-Ala mistranslated peptides[17]. Although Thr is much more abundant than Cys in the sequence of the CaMKII protein, we did not find any Thr-to-Ala substitutions. However, one Cys-to-Ala substituted peptide was found (Fig. 2d, Supplementary Fig 3). This result suggests that the misacylated Ala-tRNA$^{Cys}$ is formed *in vivo* and used in translation.

## *E. coli* AlaRS does not mischarge non-cognate human tRNAs

G3:U70 as the major identity element for tRNA$^{Ala}$ is conserved in evolution. Single nucleotide changes of G3:U70 eliminate aminoacylation of *E. coli*, *Bombyx mori*, and human tRNA$^{Ala}$ by their respective AlaRS[12]. Among different members of the aminoacyl-tRNA synthetase family, AlaRS is the most conserved tRNA synthetase in evolution and cross-charges alanyl-tRNAs from different domains of life. In an attempt to understand how human AlaRS allows mischarging, we tested *E. coli* AlaRS. As expected, *E. coli* AlaRS charged human tRNA$^{Ala}$ at high efficiency, but in contrast to human AlaRS, did not mischarge either human G4:U69-tRNA$^{Cys}$ or G4:U69-tRNA$^{Thr}$ (Fig. 3a). This lack of mischarging by *E. coli* AlaRS provided an opportunity to identify the key element in human AlaRS that supports the mischarging activity.

AlaRS is constituted of four functional domains: active site-containing catalytic domain, tRNA binding domain, editing domain, and C-terminal domain (C-Ala) (Fig. 3b). Previous work showed that the N-terminal 385-aa fragment of *E. coli* AlaRS had severely diminished aminoacylation activity and more importantly, was unable to recognize G3:U70 distinctively and charged the I3:U70 or G3:C70 substituted tRNA$^{Ala}$ with similar efficiency[18]. In contrast, the N-terminal 461-aa fragment of *E. coli* AlaRS specifically recognizes the G3:U70 base pair[19]. These results suggest that the region between residues 385 and 461 in *E. coli* AlaRS, which lies in the tRNA binding domain, is critical for G3:U70 recognition. This region, as revealed in the crystal structure of the AlaRS/tRNA$^{Ala}$ complex from *A. fulgidus*[20], contains 3 helices (α14, α15, and α16), with α14 and α16 making direct contacts with the acceptor stem of tRNA$^{Ala}$ from the minor groove side (Fig. 3c). Particularly, Asp450 in *A. fulgidus* AlaRS (corresponding to Asp400 in *E. coli* AlaRS and Asp416 in human AlaRS) specifically recognizes the free N2 amino group of G3 (in a G:U but not in a G:C base pair) through its backbone and side chain carbonyls (Fig. 3c,d). The G3:U70 base pair is also recognized from the major groove side through hydrogen bonding of the free O4 carbonyl of base U70 with the carboxamide side chain of Asn359 in *A. fulgidus* AlaRS (corresponding to Asn303 in *E. coli* AlaRS and Asn317 in human AlaRS) also in the tRNA binding domain (Fig. 3c,d). Both the Asn and Asp are strictly conserved in AlaRS proteins throughout evolution (Fig. 3b).

## Crystal structure of human AlaRS

To understand the difference between human AlaRS and *E. coli* AlaRS, we determined the crystal structure of the N-terminal fragment of human AlaRS ($AlaRS^{N455}$) that contains the catalytic and the tRNA binding domains (Fig. 4a) and compared with that of *E. coli* AlaRS solved previously[21,22]. The structure was determined at 2.68 Å resolution by molecular replacement using the *E. coli* structure as the search model (PDB 3HXU). Data collection and structure refinement statistics are given in Table 1. Both structures were solved in complex with the non-labile 5′-O-[N-(L-alanyl) sulfamoyl]adenosine (AlaSA) analogue of the reaction intermediate Ala-AMP. Superposition of the two structures gave a root-mean-squared deviation (RMSD) of 0.787 Å for 359 Cα positions, indicating overall similarity (Fig. 4b). In particular, the RMSD is 0.59 Å and 0.99 Å for the catalytic domain and the tRNA binding domain, respectively, suggesting that the structure of the catalytic domain is more similar between human and *E. coli* AlaRS than that of the tRNA binding domain. Consistently, the residues that are involved in binding to AlaSA in the catalytic domain are highly conserved between the two orthologs (Supplementary Fig. 4).

## Identification of a key residue substitution in human AlaRS related to mischarging

Relative to the highly conserved catalytic domain, the tRNA binding domain is more divergent in structure (and also in sequence) between human and *E. coli* AlaRS. Asn359 in *A. fulgidus* AlaRS for G3:U70 recognition from the major groove side corresponds to Asn317 in human and Asn303 in *E. coli* AlaRS. Their positions are superimposable and each makes a conserved H-bonding interaction with a nearby Arg side chain (Arg325 of human and Arg310 of *E. coli* AlaRS) (Fig. 4c). In contrast, Asp450 in *A. fulgidus* AlaRS for G3:U70 recognition from the minor groove side corresponds to Asp416 in human and Asp400 in *E. coli* AlaRS, and their positions are shifted from each other by 2.25 Å at the carbonyl side chain (Fig. 4c). Notably, Asp400 in *E. coli* AlaRS makes a H-bonding interaction with Gln432 from α16 of the tRNA binding domain; however, this interaction is lost in human AlaRS, because the Gln432 residue in *E. coli* AlaRS is replaced by Ala448 in human AlaRS (Fig. 4c). Interestingly, Gln is strictly conserved in AlaRS in bacteria, but is replaced in archaea and eukaryotes by Ala or Ser (Fig. 3b), neither of which is able to interact with the critical Asp residue for G3:U70 recognition from the minor groove. We speculated that loss of this interaction in human AlaRS might be related the expanded tRNA selectivity and the mischarging of non-cognate tRNAs.

To test this idea, we first mutated Gln432 to Ala in *E.coli* AlaRS to see if Gln432 is important for the charging activity and if the mutation would cause mischarge. The initial rate of charging for the Q432A mutant was reduced by about 30% towards human $tRNA^{Ala}$ and no mischarging of human G4:U69-$tRNA^{Cys}$ was detected (Fig. 4d). To confirm that this observation is not unique to *E.coli* AlaRS, we also tested *Aquifex aeolicus* AlaRS with the corresponding mutation Q421A. The result is highly consistent with that from *E.coli* AlaRS (Supplementary Fig. 5a). Next, we made the opposite mutation A448Q in human and mouse AlaRS. In both cases, the mutation had no effect on charging the cognate human $tRNA^{Ala}$, however, it substantially reduced the mischarging activity towards G4:U69-$tRNA^{Cys}$ (Fig. 4e and Supplementary Fig. 5b), and thus confirmed the importance of the A448 position for affecting tRNA charging specificity. The result supports the idea that a Gln432-Asp400

interaction between α14 and α16 of the tRNA binding domain is present in *E. coli* AlaRS as well as in other bacteria AlaRSs to precisely position the Asp400 residue to recognize the free amino group of G3:U70. The loss of this interaction in human AlaRS provides more flexibility in the recognition and thus contributes to its mischarging activity toward non-cognate tRNAs having a G4:U69 base pair. The fact that this interaction is conserved in all bacteria and is lost in all archaea and eukaryotes suggests that AlaRS mischarging confers an evolutionary advantage.

## Discussion

AlaRS is unique among the aminoacyl-tRNA synthetase family members for its particularly simple tRNA recognition. Although several tRNA synthetases recognize their cognate tRNAs without involving the anticodon[9,10], only AlaRS uses a single base pair in the acceptor stem to determine the amino acid specificity of its cognate tRNA[10,23,24,25]. This feature prompted us to speculate that AlaRS may have a natural tendency for mischarging, if a similar specificity determinant can be found in a non-cognate tRNA. Indeed, when we tested three human cytoplasmic tRNA synthetases side-by-side, only AlaRS showed substantial mischarging activity (Fig. 1). All mischarged tRNAs have a G4:U69 base pair—translocated on the acceptor stem from the G3:U70 used for $tRNA^{Ala}$ recognition.

The AlaRS mischarged tRNAs include $tRNA^{Cys}$ and $tRNA^{Thr}$. That only 1 of 30 human $tRNA^{Cys}$ genes versus 6 of 20 human $tRNA^{Thr}$ genes contain the G4:U69 base pair correlates well with the stronger mischarging signal for $tRNA^{Thr}$ than for $tRNA^{Cys}$ detected after removing $tRNA^{Ala}$ for $tRNA^{Cys}$ crosshybridization signals (Fig. 1c). The observed mischarging would potentially produce Cys-to-Ala and Thr-to-Ala substitutions in a protein. However, using a reporter protein overexpressed in human HEK293T cells, we only detected a Cys-to-Ala but no Thr-to-Ala substitution (Fig. 2d). Possibly, mischarged Ala-$tRNA^{Thr}$ is deacylated by the editing function of ThrRS or by other trans-editing factors in the cell[26,27]. Unlike ThrRS, CysRS does not contain an editing domain to deacylate Ala-$tRNA^{Cys}$. Additionally, cell and tissue type specific expression of tRNA genes[16,28] may contribute to this apparent selection of a Cys-to-Ala substitution.

Although AlaRS is highly conserved throughout evolution and is capable of cross-charging its cognate tRNAs from different domains of life, *E. coli* AlaRS is unlike human AlaRS and unable to mischarge G4:U69-tRNAs (Fig. 3a). By solving the crystal structure of human AlaRS, and comparing with the structure of *E. coli* AlaRS, we identified a key residue divergence between eukaryotes and bacteria, which affects mischarging. Therefore, the capacity of AlaRS to mischarge G4:U69-tRNAs is likely to be conserved in eukaryotes but not in bacteria. Although AlaRS proteins from archaea resemble eukaryotic AlaRS at this key position, and thus potentially also have the mischarging capacity, we found only a small number of archaeal organisms (14 out of 80) annotated in the Genomic tRNA database[15] contain G4:U69-tRNA (Supplementary Table 1), indicating that AlaRS mischarging may not occur in most archaeal organisms. In contrast, G4:U69-tRNAs were found in 135 out of 150 eukaryotic genomes (Supplementary Table 2), including all plants, vertebrates, and mammals annotated in the Genomic tRNA database[15]. In plants, the most abundant G4:U69-containing tRNAs are $tRNA^{Asn}$ and $tRNA^{Ser}$, whereas in vertebrates and mammals

(including humans), G4:U69 is mostly contained in $tRNA^{Thr}$ genes (Fig. 5). It remains to be determined if these G4:U69-containing tRNAs can be mischarged by AlaRS and, if the mischarged tRNAs are used in translation.

Nevertheless, the apparent eukaryote-specific mischarging of AlaRS is in contrast to the broadly observed mischarging of MetRS across all three domains of life[4–6]. Mischarging of MetRS *in vivo* or *in vitro* so far has been demonstrated in human, mouse, *S. cerevisiae*, *A. pernix*, and *E. coli*. MetRS also has broad tRNA promiscuity. In human cells, 8 different cytoplasmic mis-methionylated tRNA families were found, and the number was increased to 26 upon oxidative stress[4]. The large number of different tRNA families mischarged by MetRS suggests that no clear tRNA identity element is required for the mischarging. At the level of protein mistranslation, at least 8 different gene-coded amino acids (i.e., Tyr, Glu, Asp, Val, Phe, Gly, Lys, Ala, Leu) have been identified to be replaced by Met, and the substitutions are generally located on the surface of proteins[7,17], consistent with the idea that the mis-incorporated Met residues act as ROS scavengers to provide cell protection under oxidative stress. In addition, Met mis-incorporation may create specific mutant proteins with distinct activities to enhance oxidative stress response[17].

A Cys-to-Ala substitution was found in a reporter human protein in human cells (Fig. 2d), suggesting that AlaRS mischarging and its resulting mistranslation occurs *in vivo*. Previous studies found remarkable consistency on mischarging between *in vitro* and *in vivo* studies, and that mischarging is often more readily detected *in vivo* than *in vitro*[4–6]. Therefore, AlaRS mischarging would likely occur *in vivo*, although our microarray technology is not suitable for specific application for alanine.

Cys is a unique amino acid with a highly reactive thiol group. Although it is one of the least abundant residues, Cys is frequently observed in structurally and functionally important sites of proteins, used to form inter- and intra-molecular disulfide bonds and metal-binding sites, as well as catalytic acceptors and donors[29]. The occurrence of Cys in proteins correlates positively with the complexity of the organism[30,31]. Cys occurs at a frequency of 2.3% in the mammalian proteomes, whereas the frequency in *E. coli* is only 1.1%[30]. Therefore, the increased use of Cys during evolution is potentially for enhancing protein functionality in eukaryotic organisms, which might necessitate a counteraction by a way of introducing Cys-to-Ala substitutions to regulate enhanced functionality in a tissue- or cell-specific manner. Interestingly, G4:U69-containing $tRNA^{Cys}$ genes are absent in plants, but relatively abundant in vertebrates and mammals (Fig. 5 and Supplementary Table 2). Ala differs from Cys only by the lack of a thiol group, so that a Cys-to-Ala substitution is highly conservative and unlikely to cause general protein misfolding. An important first step towards understanding the potential benefit of the Cys-to-Ala substitution enabled by AlaRS mischarging is to identify endogenous proteins with such substitutions. Elucidating the potential tissue or cell type specificity of the expression of the particular G4:U69-$tRNA^{Cys}$ gene might be a prerequisite for such identifications.

In summary, our work uncovered that human AlaRS has an intrinsic capacity for mischarging alanine onto non-cognate-tRNAs with a $G_4$:$U_{69}$ base pair. The mischarging

activity is likely an evolutionary gain-of-function acting on selected tRNAs to provide potential regulation in eukaryotic organisms.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank the staff at beamline 11-1 of Stanford Synchrotron Radiation Lightsource (SSRL) for assistance in X-ray diffraction data collection. We thank K. Musier-Forsyth (The Ohio State University) for helpful discussions about this work. This research is supported by US National Institutes of Health grant R01 NS085092 (X.-L.Y.) and DP1 GM105386 (T.P.).

## References

1. Lee JW, Beebe K, Nangle LA, Jang J, Longo-Guess CM, Cook SA, Davisson MT, Sundberg JP, Schimmel P, Ackerman SL. Nature. 2006; 443:50. [PubMed: 16906134]
2. Pan T. Annu Rev Genet. 2013; 47:121. [PubMed: 23988117]
3. Sarkany Z, Silva A, Pereira PJ, Macedo-Ribeiro S. Front Mol Biosci. 2014; 1:27. [PubMed: 25988168]
4. Netzer N, Goodenbour JM, David A, Dittmar KA, Jones RB, Schneider JR, Boone D, Eves EM, Rosner MR, Gibbs JS, Embry A, Dolan B, Das S, Hickman HD, Berglund P, Bennink JR, Yewdell JW, Pan T. Nature. 2009; 462:522. [PubMed: 19940929]
5. Jones TE, Alexander RW, Pan T. Proc Natl Acad Sci U S A. 2011; 108:6933. [PubMed: 21482813]
6. Wiltrout E, Goodenbour JM, Frechin M, Pan T. Nucleic Acids Res. 2012; 40:10494. [PubMed: 22941646]
7. Lee JY, Kim DG, Kim BG, Yang WS, Hong J, Kang T, Oh YS, Kim KR, Han BW, Hwang BJ, Kang BS, Kang MS, Kim MH, Kwon NH, Kim S. J Cell Sci. 2014; 127:4234. [PubMed: 25097229]
8. Schwartz MH, Pan T. Nucleic Acids Res. 2016; 44:294. [PubMed: 26657639]
9. Giege R, Sissler M, Florentz C. Nucleic Acids Res. 1998; 26:5017. [PubMed: 9801296]
10. Beuning PJ, Musier-Forsyth K. Biopolymers. 1999; 52:1. [PubMed: 10737860]
11. Steer BA, Schimmel P. J Biol Chem. 1999; 274:35601. [PubMed: 10585437]
12. Hou YM, Schimmel P. Biochemistry. 1989; 28:6800. [PubMed: 2684266]
13. He W, Zhang HM, Chong YE, Guo M, Marshall AG, Yang XL. Proc Natl Acad Sci U S A. 2011; 108:12307. [PubMed: 21737751]
14. Xie W, Nangle LA, Zhang W, Schimmel P, Yang XL. Proc Natl Acad Sci U S A. 2007; 104:9976. [PubMed: 17545306]
15. Chan PP, Lowe TM. Nucleic Acids Res. 2016; 44:D184. [PubMed: 26673694]
16. Dittmar KA, Goodenbour JM, Pan T. PLoS Genet. 2006; 2:e221. [PubMed: 17194224]
17. Wang X, Pan T. PLoS Genet. 2015; 11:e1005745. [PubMed: 26709516]
18. Buechter DD, Schimmel P. Biochemistry. 1995; 34:6014. [PubMed: 7742303]
19. Jasin M, Regan L, Schimmel P. Nature. 1983; 306:441. [PubMed: 6358898]
20. Naganuma M, Sekine S, Chong YE, Guo M, Yang XL, Gamper H, Hou YM, Schimmel P, Yokoyama S. Nature. 2014; 510:507. [PubMed: 24919148]
21. Guo M, Chong YE, Shapiro R, Beebe K, Yang XL, Schimmel P. Nature. 2009; 462:808. [PubMed: 20010690]
22. Guo M, Shapiro R, Schimmel P, Yang XL. Acta Crystallogr D Biol Crystallogr. 2010; 66:243. [PubMed: 20179335]
23. Tian Q, Wang C, Liu Y, Xie W. Nucleic Acids Res. 2015; 43:2980. [PubMed: 25722375]
24. Palencia A, Crepin T, Vu MT, Lincecum TL Jr, Martinis SA, Cusack S. Nat Struct Mol Biol. 2012; 19:677. [PubMed: 22683997]

25. Saks ME, Sampson JR. EMBO J. 1996; 15:2843. [PubMed: 8654382]
26. Beebe K, Merriman E, Ribas De Pouplana L, Schimmel P. Proc Natl Acad Sci U S A. 2004; 101:5958. [PubMed: 15079065]
27. Liu Z, Vargas-Rodriguez O, Goto Y, Novoa EM, Ribas de Pouplana L, Suga H, Musier-Forsyth K. Proc Natl Acad Sci U S A. 2015; 112:6027. [PubMed: 25918376]
28. Ishimura R, Nagy G, Dotu I, Zhou H, Yang XL, Schimmel P, Senju S, Nishimura Y, Chuang JH, Ackerman SL. Science. 2014; 345:455. [PubMed: 25061210]
29. Marino SM, Gladyshev VN. J Biol Chem. 2012; 287:4419. [PubMed: 22157013]
30. Miseta A, Csutora P. Mol Biol Evol. 2000; 17:1232. [PubMed: 10908643]
31. Jordan IK, Kondrashov FA, Adzhubei IA, Wolf YI, Koonin EV, Kondrashov AS, Sunyaev S. Nature. 2005; 433:633. [PubMed: 15660107]

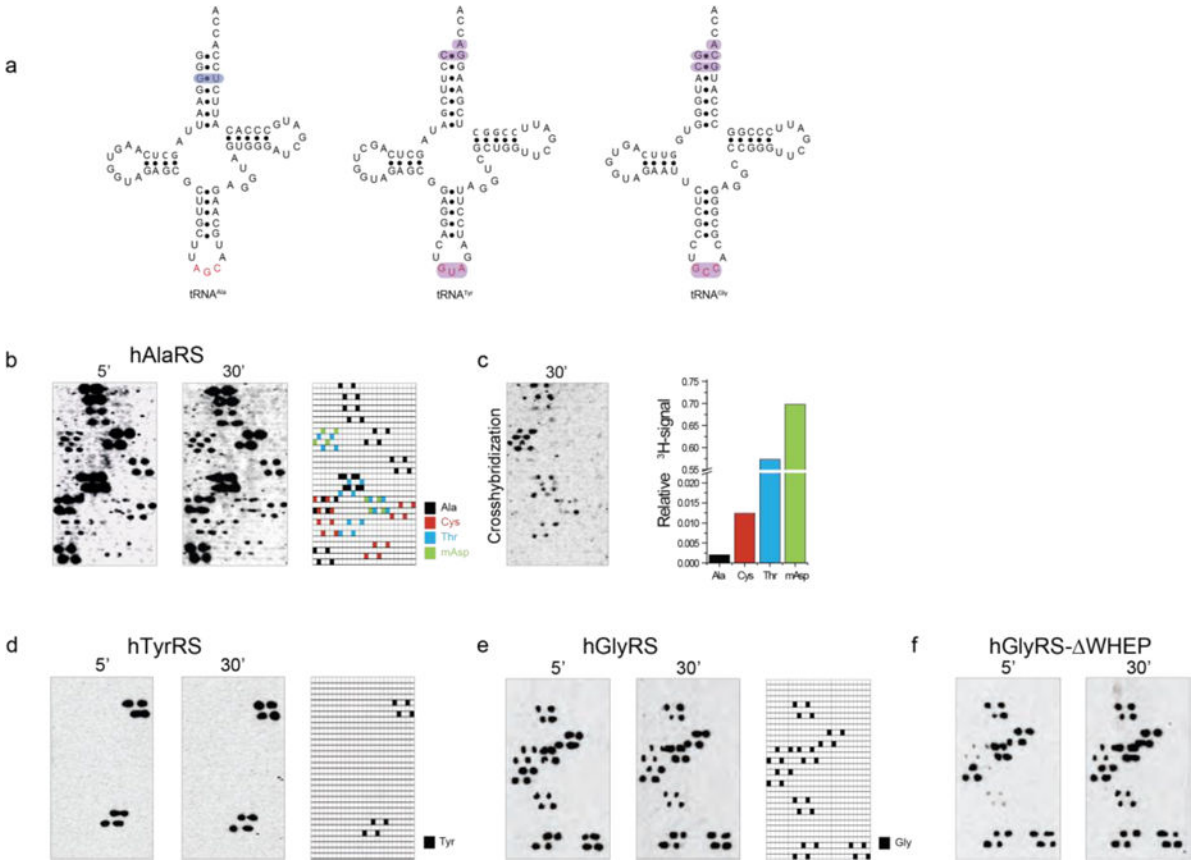



**Figure 1. Human AlaRS and not TyrRS and GlyRS mischarge non-cognate human tRNAs**

**(a)** Cloverleaf secondary structure of human $tRNA^{Ala}$, $tRNA^{Tyr}$, and $tRNA^{Gly}$. The identity elements are highlighted with blue shadow for $tRNA^{Ala}$, and with purple shadows for $tRNA^{Tyr}$ and $tRNA^{Gly}$. The anticodon nucleotides are in red. **(b)** Microarray showing misacylation of non-cognate tRNAs by human AlaRS after 5 or 30 min of reaction with total human tRNAs and [$^3$H]-alanine (black, Ala-$tRNA^{Ala}$; burgundy, Ala-$tRNA^{Cys}$; blue, Ala-$tRNA^{Thr}$; green, Ala-$tRNA^{mAsp}$). **(c)** Microarray analysis after blocking with a large excess of all $tRNA^{Ala}$ probes to exclude the potential signal from $^3$H-Ala-$tRNA^{Ala}$ cross-hybridizing to $tRNA^{Cys}$, $tRNA^{Thr}$ or mt-$tRNA^{Asp}$ probes. The relative remaining signal after crosshybridization is quantified separately for $tRNA^{Ala}$, $tRNA^{Cys}$, $tRNA^{Thr}$ and mt-$tRNA^{Asp}$. (**d, e, f**) No mischarging activity is detected with human TyrRS **(d)**, GlyRS **(e)**, and GlyRS without the metazoan-specific WHEP domain **(f)**.

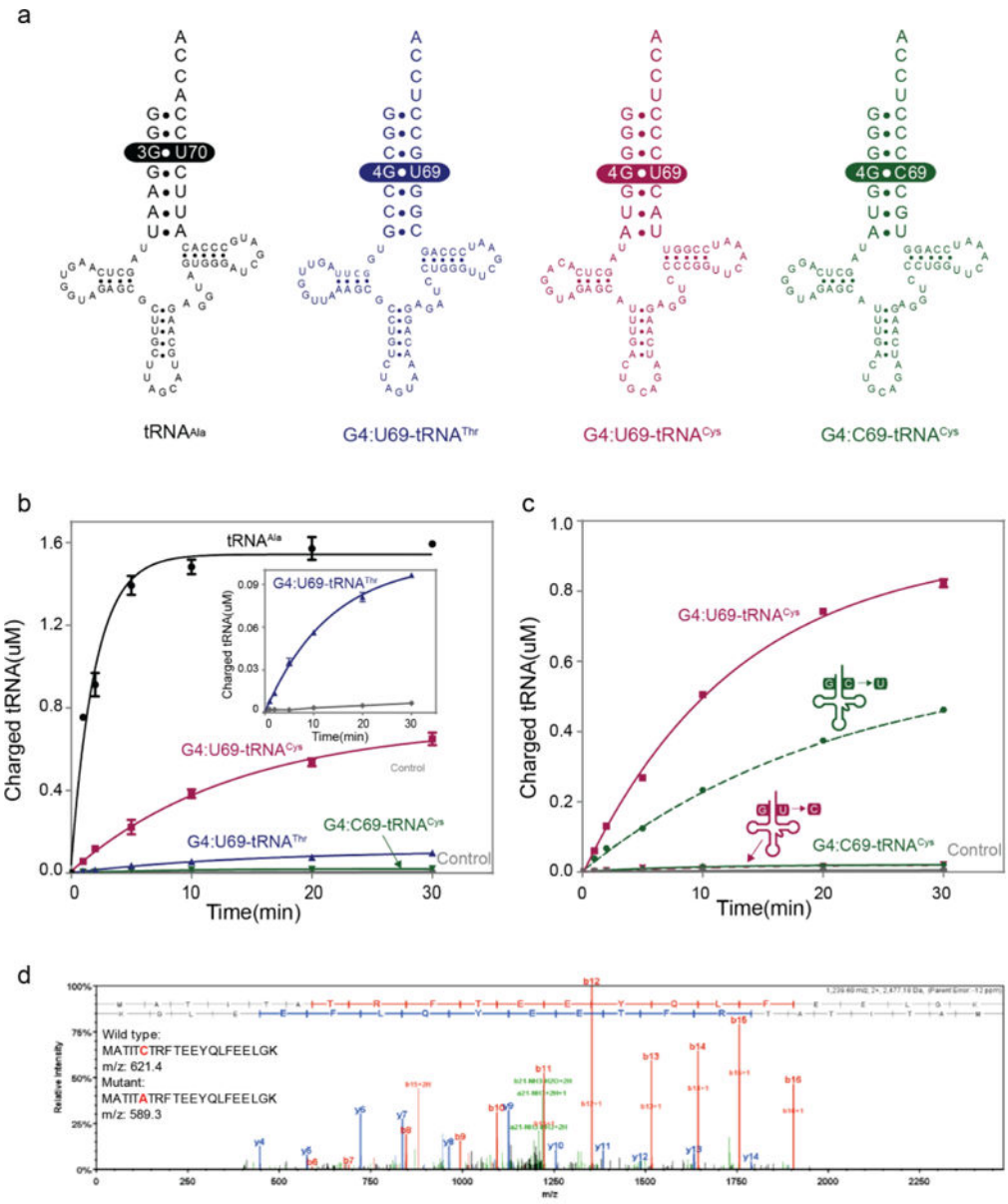



**Figure 2. G4:U69 base pair in tRNA is responsible for human AlaRS mischarging**

**(a)** G:U base pair location indicated on the cloverleaf secondary structure of different tRNAs. **(b)** *In vitro* aminoacylation assay showing that human AlaRS has mischarging activity with G4:U69-containing tRNAs. Assays were carried out with 200 nM AlaRS and 4 μM cognate $tRNA^{Ala}$ or 25 μM non-cognate tRNAs. Inset shows the zoom-in view of the mischarging activity of AlaRS with G4:U69-$tRNA^{Thr}$. Error bars indicate standard deviations. **(c)** Mutational analysis on $tRNA^{Cys}$ in both gain-of-function and loss-of-function manners confirming that the mischarging activity of AlaRS is dependent on the G4:U69 base pair. **(d)** Mass spectrometry analysis of a flag-CaMKII protein expressed and purified from HEK293T cells identified a Cys-to-Ala substitution.

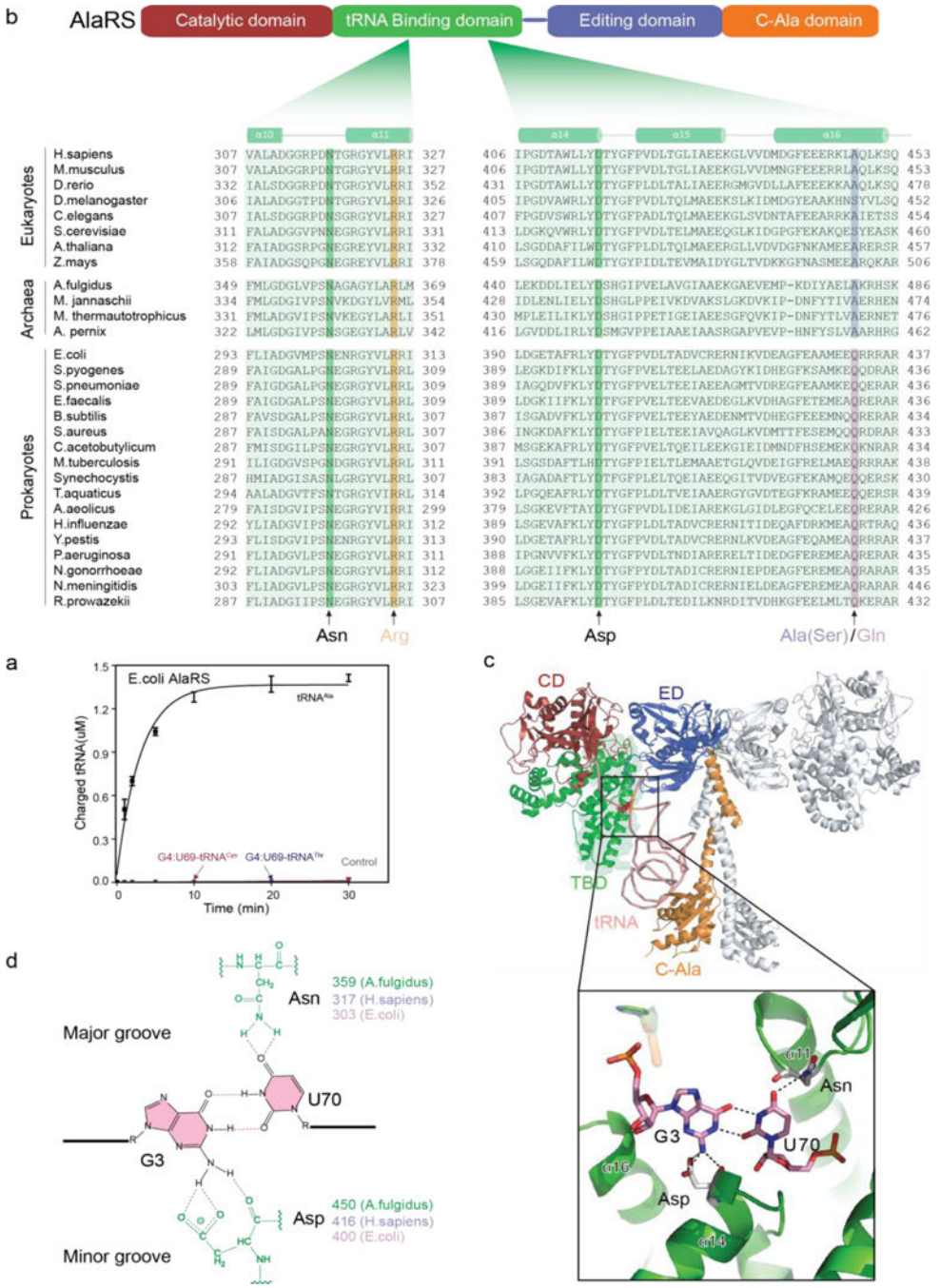



**Figure 3. *E. coli* AlaRS does not mischarge non-cognate human tRNAs**

**(a)** *In vitro* aminoacylation assay showing that *E.coli* AlaRS has no mischarging activity towards non-cognate tRNAs. Assays were carried out with 200 nM *E.coli* AlaRS and 4 μM cognate human $tRNA^{Ala}$ or 25 μM non-cognate human tRNAs. **(b)** Conservation analysis of AlaRS sequences across bacteria, archaea, and eukaryotes. **(c,d)** Specific recognitions of the G3:U70 identity element in $tRNA^{Ala}$ by AlaRS. The recognition is adapted from the crystal structure of *A. fulgidus* AlaRS and $tRNA^{Ala}$ complex (PDB 3WQY) (CD, catalytic domain; TBD, tRNA binding domain; ED, editing domain; C-Ala, C-Ala domain).

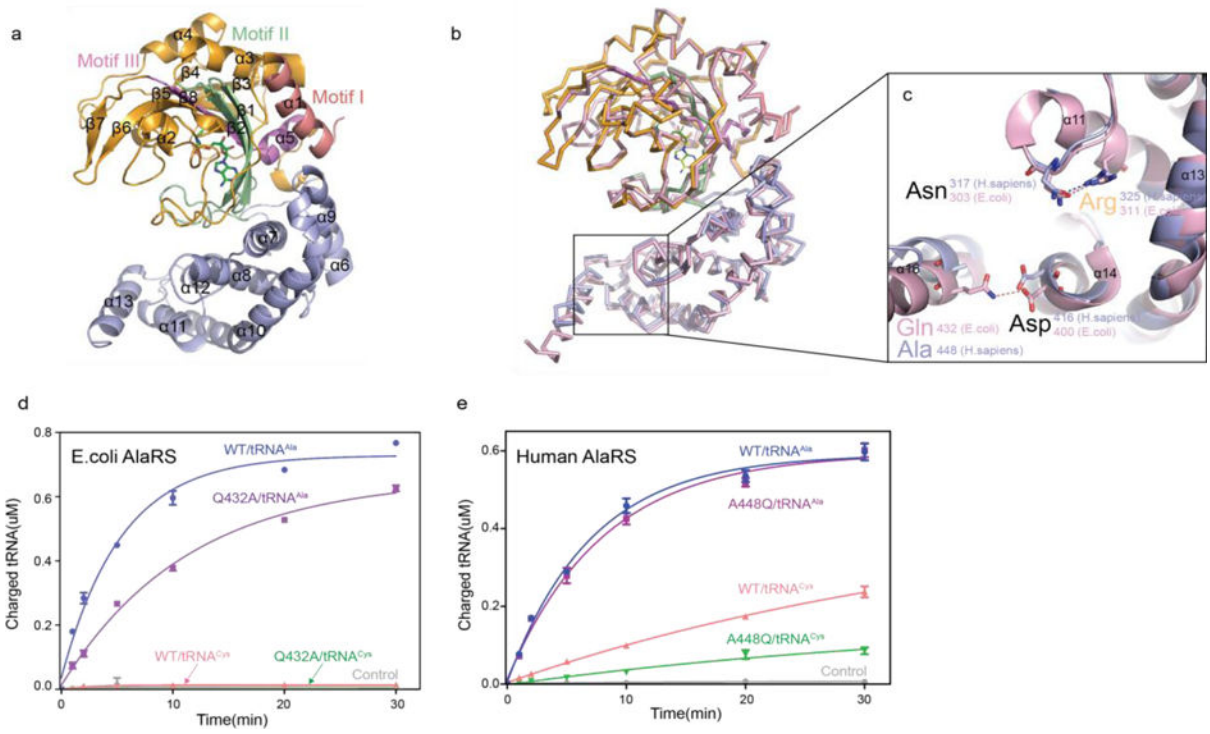



**Figure 4. Crystal structure of human AlaRS**

**(a)** The co-crystal structure of human AlaRS$^{N455}$ and AlaSA. Catalytic domain is colored in orange and tRNA binding domain is colored in blue. The three conserved sequence motifs (Motif I, II, and III) in the catalytic domain are highlighted with different colors. AlaSA is bound in the active site within the catalytic domain. **(b)** Superimposition of the crystal structure of human AlaRS$^{N455}$ and of *E. coli* AlaRS (PDB 3HXU). Human AlaRS$^{455}$ is colored as indicated above, while *E. coli* AlaRS is colored in pink. **(c)** A zoom-in view showing human AlaRS lost the Gln-Asp interaction found in *E. coli* AlaRS. **(d)** *In vitro* aminoacylation assay showing Gln-to-Ala (Q432A) substitution in *E. coli* AlaRS is unable to mischarge human G4:U69-tRNA$^{Cys}$. Assays were carried out with 200 nM of WT or Q432A AlaRS, and 2.5 μM tRNA$^{Ala}$ or 25 μM G4:U69-tRNA$^{Cys}$. **(e)** Ala-to-Gln (A448Q) substitution reduces the mischarging activity of human AlaRS towards G4:U69-tRNA$^{Cys}$. The aminoacylation assays were carried out with 200 nM of WT or A448Q AlaRS, and 2 μM tRNA$^{Ala}$ or 10 μM G4:U69-tRNA$^{Cys}$.

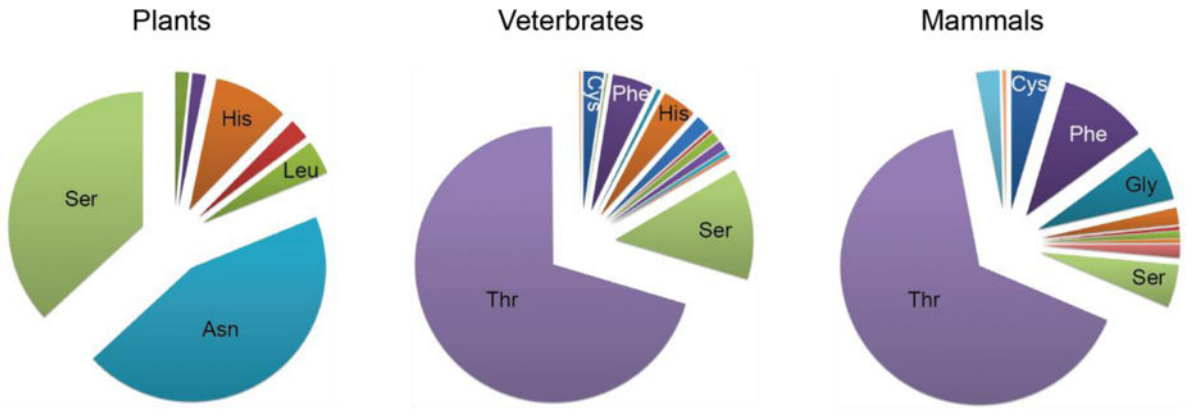



**Figure 5. Distribution of G4:U69-containing tRNA genes in plants, vertebrates and mammals**

Only tRNA genes with tRNAScan score of >50 are included here. These tRNAs have a high likelihood to form the cloverleaf secondary structure.

## Table 1

Data collection and refinement statistics

| **Parameter*** | **AlaRS$^{N455}$** |
|---|---|
| Data collection statistics | |
| Wavelength (Å) | 1.1271 |
| Resolution range (Å) | 50–2.68(2.77–2.68) |
| Space group | P 1 |
| Cell dimensions | |
| a,b,c (Å) | 52.17 98.26 201.39 |
| α,β,γ (°) | 90.07, 89.95, 90.11 |
| Unique reflections | 105564 (9918) |
| Multiplicity | 1.9(1.9) |
| Completeness (%) | 93.3(87.4) |
| Mean I/sigma(I) | 14.4(4.2) |
| Wilson B-factor | 60.10 |
| R-merge (%)† | 9.2(29.3) |
| Structure refinement | |
| Rwork/Rfree (%) | 21.6/25.4 |
| Number of non-hydrogen atoms | 28584 |
| macromolecules | 28360 |
| ligands | 224 |
| waters | 0 |
| Protein residues | 3585 |
| RMSD | |
| Bond length (Å) | 0.018 |
| Bond angles (°) | 2.01 |
| Average B-factor | 56.2 |
| Ramachandran plot (%) | |
| favored region | 96.9 |
| allowed region | 2.5 |
| Outliers region | 0.5 |

*Values in parentheses are for highest-resolution shell.

†Rmerge= Σhkl |I-<I>|/ΣhklI where I is the intensity of unique relfection hkl and <I> is the average over symmetry-related observations of unique reflection hkl.