Distinct ways of G:U recognition by conserved tRNA binding motifs

Yeeting E. Chong, Min Guo, Xiang-Lei Yang, Bernhard Kuhle, Masahiro Naganuma, Shun-ichi Sekine, Shiheyuki Yokoyama, and Paul Schimmel

*The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037; ‡Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA 92037; †Department of Molecular Medicine, The Scripps Florida Research Institute, Jupiter, FL 33458; §Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Bunkyō-ku, 113-0033 Tokyo, Japan; ‡Laboratory of Structural Biology, Graduate School of Science, The University of Tokyo, Bunkyō-ku, 113-0033 Tokyo, Japan; and ¶RIKEN Systems and Structural Biology Center, Tsukuba-ku, 230-0045 Yokohama, Japan

Contributed by Paul Schimmel, June 5, 2018 (sent for review April 25, 2018; reviewed by Charles W. Carter Jr. and Osamu Nureki)

Throughout three domains of life, alanyl-tRNA synthetases (AlaRSs) recognize a G3:U70 base pair in the acceptor stem of tRNAAla as the major identity determinant of tRNAAla. The crystal structure of the archaeon Archaeoglobus fulgidus AlaRS in complex with tRNAAla provided the basis for G3:U70 recognition with residues (Asp and Asn) that are conserved in the three domains [Naganuma M, et al. (2014) Nature 510:507–511]. The recognition mode is unprecedented, with specific accommodation of the dyad symmetry of the G:U wobble pair and exclusion of the dyad symmetry of a Watson-Crick pair. With this conserved mode, specificity is based more on “fit” than on direct recognition of specific atomic groups. Here, we show that, in contrast to the archaeal complex, the Escherichia coli enzyme uses direct positive (energetically favorable) minor groove recognition of the unpaired 2-amino of G3 by Asp and repulsion of a competing base pair by Asn. Strikingly, mutations that disrupted positive recognition by the E. coli enzyme had little or no effect on G:U recognition by the human enzyme. Alternatively, Homo sapiens AlaRS selects G:U without positive recognition and uses Asp instead to repel a competitor. Thus, the widely conserved Asp-plus-Asn architecture of AlaRSs can select G:U in a straightforward (bacteria) or two different unconventional (eukarya/archaea) ways. The adoption of different modes for recognition of a widely conserved G:U pair in alanine tRNAs suggests an early and consistent role for G:U in the development of the genetic code.

Aminoacyl-tRNA synthetase | tRNA | evolution | specificity

second genetic code

Early work showed that a single G3:U70 base pair in the acceptor stem marked tRNAAla for aminoacylation with alanine (1–4). G3:U70 is conserved through all three kingdoms of life and is so robust that its transfer into nonalanine tRNAs converts them to alanine acceptors, both in vitro and in vivo (2, 3, 5). Using synthetic RNA substrates with both natural and nonnatural substitutions at 3:70, previous investigations also showed that the unpaired exocyclic 2-amino group of G3 in the RNA minor groove was critical for aminoacylation. (With the G:G pair, the 2-amino group is H-bonded, while it is free in the G:U wobble pair.) To establish this point, nonnatural pairs that placed a free amino group in the same position were shown to be active (Fig. L4). These included 2-amino adenosine and, separately, G paired with isocytidine (6). Ablation of the exocyclic amino group by creating an U:U pair abolished aminoacylation (7, 8). In addition, substitution of 2-aminopurine [paired with U (2-AP:U)] for G yielded an inactive substrate for aminoacylation (7, 8). The sensitivity of the system to subtle effects of the geometry of recognition was shown with the U:G substitution. Although the 2-amino group of the U:G wobble superimposes on that of G:U, its opposite orientation for projecting into the minor groove prevented aminoacylation (6). Although the role of O4 of U70 was not clear from these studies, O4 alone was clearly not sufficient to confer aminoacylation, because it is present in the inactive 2-AP:U and I:U substrates.

Structural insight into G:U70 recognition by alanyl-tRNA synthetase (AlaRS) was provided by the crystal structure of the archaeal Archaeoglobus fulgidus AlaRS in complex with tRNAAla. In this structure, the tRNA-recognition domain contacts the tRNA acceptor stem from both the minor and major grooves. Asn359 and Asp450 hereby make the key H-bonding contacts with the G3:U70 wobble pair. The amide nitrogen of the Asn359 side chain contacts O4 of U70 (major-groove side), and the carbonyl side chain as well as the backbone carbonyl of Asp450 H-bonds with the 2-amino group of G3 (minor-groove side) (Fig. 1C). Both archaeal AlaRS residues, Asn359 and Asp450, that recognize the identity determinant G3:U70 base pair are conserved throughout evolution (as Asn303 and Asp400 in Escherichia coli and as Asn317 and Asp416 in Homo sapiens AlaRS). In addition, the sequences that surround these residues are highly conserved between bacteria and human (Fig. 1B).

A superposition of the structures of the A. fulgidus and E. coli AlaRS catalytic fragments revealed that they share almost the same fold, except for the local difference in the orientation of the Mid2 subdomain from the tRNA-recognition domain (SI).

Significance

Aminoacyl-tRNA synthetases (aaRSs) establish the rules to express the universal genetic code. During aminoacylation, each of the 20 aaRSs associates 1 of 20 amino acids with a specific trinucleotide known as anticodon. Remarkably, for alanyl-tRNAs, the synthetase makes no contact with the anticodon. Instead, it uses a “second genetic code” by picking out a single G3:U70 base pair in the tRNA acceptor stem, which is close to the amino acid attachment site, but 76 Å away from the anticodon. Here, we show that, while in the three kingdoms of life, alanyl-tRNA synthetases use G3:U70 to identify alanyl-tRNAs, surprisingly, they use three different mechanisms to achieve this. We thus suggest that, in evolution, the genetic code had a powerful and persistent preference for associating G:U with alanine.
Appendix, Fig. S1). Based on sequence alignments (SI Appendix, Fig. S2) and the crystal structures of the E. coli AlaRS catalytic fragment and of the A. fulgidus AlaRS•tRNA•Ala complex, a homology model of E. coli AlaRS in complex with tRNA•Ala was constructed (Fig. 2A). The conformation of the E. coli AlaRS Mid2 subdomain was slightly adjusted for the tRNA binding. In this model, G3:U70 is recognized on the major- and minor-groove sides by the conserved amino acid residues, in the same manner as in the A. fulgidus AlaRS•tRNA•Ala complex (9) (Fig. 2B).

On the minor-groove side of G3:U70, the 2-amino group of G3 hydrogen bonds with the side-chain carboxyl and main-chain carbonyl groups of Asp400. In addition, the side-chain carboxyl group of Asp400 is near the 2′-hydroxyl group of C71. The 2′-hydroxyl group of the U70 ribose hydrogen bonds with the main-chain carbonyl groups of Asp400 and Tyr399, and the amide NH group of Gly403. Asp400, Tyr399, and Gly403 are strictly conserved for all three enzymes. This observation is consistent with atomic group mutagenesis studies that demonstrated the
importance of the ribose 2'-OH of U70 and other nearby atoms for aminoacylation activity (10, 11).

In *H. sapiens* AlaRS, the above-described Asp and Asn residues are conserved as Asp416 and Asn317. Using the coordinates of the *H. sapiens* AlaRS catalytic fragment, *H. sapiens* C-terminal domain, and *A. fulgidus* AlaRS•tRNA<sub>Ala</sub> complex (Fig. 2C and D and SI Appendix, Fig. S1), a homology model of the *H. sapiens* AlaRS•tRNA<sub>Ala</sub> complex was created. In contrast to the archaean and bacterial AlaRSs, which form a dimer in solution, *H. sapiens* AlaRS is presented as a monomer (12). To avoid steric clashes with the editing domain, the C-terminal domain (C-Ala) is reoriented relative to its archaean and bacterial counterparts and does not form the tight contacts with the tRNA elbow region that are observed in the *A. fulgidus* AlaRS•tRNA<sub>Ala</sub> complex, consistent with C-Ala having no effect on charging activity in the human enzyme (12). However, the interactions around the G3:U70 base pair are analogous to those in the homology model of the *E. coli* AlaRS•tRNA<sub>Ala</sub> complex (Fig. 2D).

To understand the mechanism of G3:U70 recognition by bacterial AlaRS, we used aminoacylation kinetics with transcripts of *E. coli* tRNA<sub>Ala</sub> and recombinant *E. coli* AlaRS purified from *E. coli*. Two mutations were introduced into *E. coli* AlaRS to test the role of Asp400 in minor groove recognition of the 2-amino group of G3 [Asp400:NH<sub>2</sub>(G3) interaction]. First,
the hydrogen bond between the carboxyl group of Asp400 and the 2-amino group of G3 was removed by introduction of a D400A substitution, which inserts a smaller residue that should be sterically accommodated but lacks the ability to make a positive, energetically favorable interaction. The D400A mutant enzyme was virtually inactive for aminoacylation (Fig. 1D and Table 1). Next, we created D400N AlaRS. Here, the Asp-to-Asn replacement is roughly isosteric. However, while model predictions that both carboxyl oxygens of the Asp side chain are involved in H-bonding to the 2-NH₂ of G3 and the 2-OH of C71 (Fig. 1C), the amide group of Asn cannot H-bond to G3. Like D400A AlaRS, D400N AlaRS was virtually inactive for aminoacylation (Table 1). These results mirror those in which the 2-amino of G3 was ablated (as in I1U) or shifted (as in G:C) (Fig. 1I).

For the recognition of O4 of U70 by Asn303 of E. coli AlaRS, two additional mutants of E. coli AlaRS were cloned and purified. The N303D substitution was selected as an approximately isosteric substitution to disrupt the interaction with O4 of U70. Consistent with the universal conservation of Asn303 and with an electrostatic repulsion between the introduced carboxyl of D303 and O4 of U70, N303D AlaRS was almost completely inactive for aminoacylation (Table 1). Next, we made an N303A replacement that substituted Asn with the smaller methyl group side chain of Ala. Remarkably, N303A AlaRS was twofold higher in activity than WT AlaRS (Fig. 1D and Table 1). Most of this change came from a change in κcat. Interestingly, the apparent dissociation constant Kₚ was the same [0.12 μM (WT) versus 0.14 μM (N303A)] when the H bond from Asn303 to O4 of U70 was removed. Thus, and surprisingly, the highly conserved H bond of Asn303 to U70 is a somewhat negative contribution to recognition of tRNAAla.

We then considered the possibility that the effects of N303A were dependent on, or coupled to, the presence of Asp400. To investigate this possibility, we asked whether the N303A substitution could rescue in part the deleterious effect of the D400A replacement. Indeed, the N303A substitution into D400A AlaRS (to give N303A/D400A AlaRS) promoted the activity over that of D400A AlaRS alone (Table 1 and SI Appendix, Fig. S3A). Remarkably, this enhancement amounted to an activity for N303A/D400A AlaRS that was only sixfold to sevenfold that of WT AlaRS. Thus, the effects of N303A are not dependent on the presence of Asp400. The result gave further motivation to understand the rationale for the presence of the highly conserved Asn303 H bond with O4 of U70. In particular, we considered the possibility that conservation of Asn303 throughout evolution was not solely for recognition of O4 of U70.

The most prominent possibility is that the universally conserved Asn303 is for repelling the major competitor 3:70 base pair in noncognate tRNA:AARS complexes. Indeed, the N303A substitution could rescue in part the deleterious effect of the D400A substitution. Thus, the effects of N303A are not dependent on the presence of Asp400. The result gave further motivation to understand the rationale for the presence of the highly conserved Asn303 H bond with O4 of U70. In particular, we considered the possibility that conservation of Asn303 throughout evolution was not solely for recognition of O4 of U70.

Table 1. Kinetics of aminoacylation by E. coli AlaRS mutants

| E. coli AlaRS | κcat, s⁻¹ | kₘᵣ, μM | kₜcat/Kₘᵣ, μM⁻¹s⁻¹ | Rel. kcat/Kₘᵣ |
|---------------|----------|---------|-----------------|-------------|
| WT            | 0.36 ± 0.02 | 0.12 ± 0.03 | 3.0 ± 0.9 | 1          |
| N303A         | 0.82 ± 0.05 | 0.14 ± 0.04 | 5.7 ± 1.9 | 1.9        |
| N303D         | 0.12 ± 0.01 | 3.4 ± 0.4  | 0.034 ± 0.006 | 0.012    |
| D400A         | 0.17 ± 0.01 | 1.4 ± 0.2   | 0.13 ± 0.03 | 0.042     |
| D400N         | 0.14 ± 0.01 | 1.2 ± 0.1   | 0.12 ± 0.02 | 0.040     |
| N303A/D400A   | 0.30 ± 0.02 | 0.67 ± 0.12 | 0.44 ± 0.11 | 0.15      |

Assays were carried out at 25 °C and pH 7.5, with tRNAAla concentrations ranging from 62.5 nM to 24.7 μM. Enzyme concentrations were 10 nM (WT), 12.5 nM (N303A), 1 μM (N303D), 80 nM (D400A), 60 nM (D400N), and 20 nM (N303A/D400A). The data represent the mean of triplicate experiments ± SEM.
relative activities of the mutant versus the WT enzymes (Fig. 3A–C).
To compare the relative loss of discriminatory power of the mutant AlaRSs for G3:U70 tRNA\textsuperscript{Ala}, the relative activity for each mutant for G3:C70 tRNA\textsuperscript{Ala} (Fig. 3B) was divided by the relative activity for G3:U70 tRNA\textsuperscript{Ala} (Fig. 3A) (WT AlaRS is normalized to 1). Strikingly, the effect of the mutations on the loss of discriminatory power in every instance is more profound with the bacterial enzyme (Fig. 3C).

Table 2. Kinetics of aminoacylation by human AlaRS mutants

| Human AlaRS | $k_{\text{cat}}$, s\textsuperscript{-1} | $K_{\text{m}}$, μM | $k_{\text{cat}}/K_{\text{m}}$ | Rel $k_{\text{cat}}/K_{\text{m}}$ | $k_{\text{cat}}/K_{\text{m}}$ | $k_{\text{cat}}/K_{\text{m}}$ |
|-------------|----------------|----------------|----------------|------------------|----------------|----------------|
| WT          | 0.23 ± 0.01    | 0.26 ± 0.04    | 0.89 ± 0.18    | 1                | 0.23 ± 0.01    | 0.26 ± 0.04    |
| N317A       | 0.20 ± 0.01    | 0.29 ± 0.05    | 0.69 ± 0.15    | 0.78             | 0.20 ± 0.01    | 0.29 ± 0.05    |
| N317D       | 0.08 ± 0.001   | 0.41 ± 0.01    | 0.059 ± 0.018  | 0.065            | 0.08 ± 0.001   | 0.41 ± 0.01    |
| D416A       | 0.32 ± 0.02    | 0.35 ± 0.08    | 0.92 ± 0.26    | 1.0              | 0.32 ± 0.02    | 0.35 ± 0.08    |
| D416N       | 0.20 ± 0.01    | 0.47 ± 0.07    | 0.43 ± 0.08    | 0.48             | 0.20 ± 0.01    | 0.47 ± 0.07    |
| N317A/D416A | 0.21 ± 0.01    | 0.23 ± 0.06    | 0.90 ± 0.31    | 1.0              | 0.21 ± 0.01    | 0.23 ± 0.06    |

Assays were carried out at 25 °C and pH 7.5, with tRNA\textsuperscript{Ala} concentrations ranging from 62.5 nM to 24.7 μM. Enzyme concentrations were 20 nM (WT), 43 nM (N317A), 250 nM (N317D), 50 nM (D416A), 56 nM (D416N), and 25 nM (N317A/D416A). The data represent the mean of triplicate experiments ± SEM.

The G3:U70 recognition modes are distinct in E. coli, H. sapiens (Fig. 3D), and A. fulgidus AlaRSs (9). Particularly significant is the lack of any positive functional interaction in the eukaryal and archaeal systems, such as that provided by the favorable contact between the carboxyl of D416 and the 2-NH\textsubscript{2} of G3 in E. coli. The congruency of the conserved Asn-plus-Asp structure in ensuring G3:U70 recognition in all three domains of life, as exemplified in this work, support the early origin of the G:U base pair as an identity element but not with a specific mechanism of its recognition. The genetic code ostensibly had a powerful and persistent preference for associating G:U with alanine.

We speculate that the different recognition modes of the bacterial, archaeal, and human AlaRSs may be a consequence of the effects of domain–domain interactions, rather than of local interactions. E. coli AlaRS can be divided into an N-terminal catalytic fragment and another half (the editing domain and the C-terminal domain), which are both soluble (13–25). In contrast, A. fulgidus AlaRS cannot be similarly divided, probably because of the hydrophobic interface between the catalytic fragment and the editing domain (26). Consequently, the interaction between the catalytic fragment and the editing domain of E. coli AlaRS is probably more flexible than that of A. fulgidus AlaRS and possibly than that of H. sapiens AlaRS. The flexibility of this interface could provide the accommodation of a direct system of recognition for the bacterial enzyme, while its rigidity could enforce a mechanism based on steric fit.

In examining more than 2,500 deposited sequences of AlaRSs from the three kingdoms of life, we found that the common Asn-plus-Asp structure for recognition of G3:U70 is highly conserved (>95%) in all three domains of life (SI Appendix, Table S1). Interestingly, for both the bacterial and human enzyme, replacement of the residue that is proximal to O4 of U70—N303A (E. coli) or N317A (human)—yields an enzyme with robust activity on each respective native tRNA\textsuperscript{Ala} substrate (Fig. 3A). However, in both instances, specificity is lost (Fig. 3C).
observations are consistent with the idea that repulsion of the competitor tRNAs was of paramount importance in the selection and retention of the conserved Asn. Similarly, unlike the bacterial enzyme, in the human system the effect of substitution of Asp416 with either Ala or Asn is minimal, and yet those substitutions still result in a decline in specificity and selective pressure to retain the conserved Asp.

Methods Summary

E. coli and H. sapiens AlaRS enzymes were overexpressed in E. coli and purified as previously described (25). Concentrations of purified AlaRS enzymes were in vitro transcribed using T7 polymerase. Aminoacylation with WT E. coli tRNAAla was performed by incubating purified enzyme (5 nM) in reaction buffer (50 mM Hepes, pH 7.5, 20 mM KCl, 5 mM MgCl2, 1 mM DTT), with 4 mM ATP, 20 μM l-[35S]Ala, and 2 μM tRNAAla at 25 °C. For reactions with G3:C70 tRNAAla, 500 nM enzyme and 10 μM G3:C70 tRNAAla was used, at 37 °C. Incorporation of l-[35S]Ala was measured by scintillation counting in a MicroBeta plate reader (27).

ACKNOWLEDGMENTS. This work was supported by NIH Grants GM15539 and GM23562 and by a fellowship from the National Foundation for Cancer Research.

1. Rich A, RajBhandary UL (1976) Transfer RNA: Molecular structure, sequence, and properties. Annu Rev Biochem 45:805–860.
2. Hou YM, Schimmel P (1988) A simple structural feature is a major determinant of the identity of a transfer RNA. Nature 333:140–145.
3. McClain WH, Foss K (1988) Changing the identity of a tRNA by introducing a G-U wobble pair near the 3′ acceptor end. Science 240:793–796.
4. Giegé R (2008) Toward a more complete view of tRNA biology. Nat Struct Mol Biol 15:1007–1014.
5. Hou YM, Schimmel P (1989) Evidence that a major determinant for the identity of a transfer RNA is conserved in evolution. Biochemistry 28:6800–6804.
6. Park SJ, Hou YM, Schimmel P (1989) A single base pair affects binding and catalytic parameters in the molecular recognition of a transfer RNA. Biochemistry 28:2740–2746.
7. Musier-Forsyth K, et al. (1991) Specificity for aminoacylation of an RNA helix: An unpaired, exoosyclic amino group in the minor groove. Science 253:784–786.
8. Musier-Forsyth K, Scaringe S, Usman N, Schimmel P (1991) Enzymatic aminoacylation of single-stranded RNA with an RNA cofactor. Proc Natl Acad Sci USA 88:209–213.
9. Naganuma M, et al. (2014) The selective tRNA aminoacylation mechanism based on a single G=U pair. Nature 510:507–511.
10. Beuning PJ, Gulotta M, Musier-Forsyth K (1997) Atomic group “mutagenesis” reveals major groove fine interactions of a tRNA synthetase with an RNA helix. J Am Chem Soc 119:8397–8402.
11. Musier-Forsyth K, Schimmel P (1992) Functional contacts of a transfer RNA synthetase with 2′-hydroxyl groups in the RNA minor groove. Nature 357:513–515.
12. Sun L, Song Y, Blocquel D, Yang XL, Schimmel P (2016) Two crystal structures reveal design for repurposing the C-Ala domain of human AlaRS. Proc Natl Acad Sci USA 113:14300–14305.
13. Ibba M, Soll D (1999) Quality control mechanisms during translation. Science 286:1893–1897.
14. Ling J, et al. (2009) Resampling and editing of mischarged tRNA prior to translation elongation. Mol Cell 33:654–660.