

REASSIGNMENT OF CODONS FROM ARG TO ALA BY MULTIPLE TRNA^{ALA} VARIANTS

Kazuaki Amikura^a and Daisuke Kiga^{a,b}

^aDepartment of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Midori-ku, Yokohama-shi, Kanagawa 226-8503, Japan; E-mail: kiga@sb.dis.titech.ac.jp

^bEarth-Life Science Institute, Tokyo Institute of Technology, 35 Meguro-ku, Tokyo 152-8551, Japan
(Received: April 4, 2013; Accepted: June 17, 2013)

Abstract

We have recently developed “simplified genetic codes” in which only 19 amino acids are assigned to one sense codon each. In one simplified code, for example, the UGG codon for tryptophan is reassigned to alanine (Ala) by a tRNA^{Ala} variant. We have previously generated such codes each of which was constructed using only one kind of tRNA variant. In this study, we describe a novel simplified genetic code in which the codons of Arg were reassigned to Ala using 3 kinds of tRNA^{Ala} variants. The usage of multiple tRNA variants will be important to construct more simplified genetic codes containing less than 19 amino acids. A Simplified genetic code will provide not only new insights into primordial genetic codes, but also an engineering tool for the assessment of protein evolution.

(Keywords)

simplified genetic code, amino acids, engineering, evolution, tRNA, aminoacyl-tRNA synthetase, anticodon, translation, cell-free, rewiring

Introduction

Transfer RNA achieves the correspondence between 20 amino acids and 61 codons in a genetic code in almost all known organisms[1]. Each amino acid is bonded to its cognate tRNA(s) by a distinct aminoacyl-tRNA synthetase (aaRS). These assignments are also achieved by functional interaction of aaRS, tRNAs, 20 amino acids, and other components of the translational apparatus. Each tRNA is bonded to its cognate amino acid by a distinct aaRS.

It is generally accepted that the number of amino acids in the genetic code has increased through evolution from fewer than 20 amino acids through evolution [1, 2]. Phylogenetic analysis of aminoacyl-tRNA synthetases suggests a way to add amino acid into the genetic code [3]. All known eukaryotes and a small number of bacteria have 20 canonical aaRSs to directly synthesize cognate aminoacyl-tRNAs. On the other hand, all known archaea and most bacteria do not have at least one of the 20 canonical aaRSs [4, 5]. Therefore, only 19 amino acids are directly connected to cognate tRNAs by cognate aaRS. Instead of the direct aminoacylation, the amino acid except the 19 amino acids is indirectly connected to cognate tRNA. In this case, first, the amino acid is connected to a non-cognate tRNA by a non-cognate aaRS. Second, the non-cognate amino acid moiety on the mischarged tRNA is transformed into the cognate

amino acid. This tRNA-dependent conversion, in the case of synthesizing asparagine, is the sole pathway for asparagine biosynthesis in *Deinococcus radiodurans* and *Thermus thermophilus* [6, 7].

When all proteins were synthesized from a primordial genetic code encoding fewer than 20 amino acids, was an organism able to survive using only these proteins? In other words, were these proteins able to construct a genetic code and to achieve all essential biological functions? The characterization of so-called “simplified proteins” supports the notion that not all the 20 canonical amino acids are required in order to construct a protein with biological function and structure.

We recently proposed that “simplified genetic codes” that encode less than 20 amino acids will be an effective tool to prepare a simplified protein [8]. Though addition of a single tRNA variant was enough to creation of a simplified genetic code with 19 amino acids, addition of multiple tRNA variants is required for creation of a more simplified genetic code encoding fewer than 19 amino acids requires.. In this work, we show that multiple tRNA^{Ala} variants reassign codons to amino acids to construct a simplified genetic code lacking Arg.

Material and Method

DNA constructs

For the expression of His-tagged *ras*, we used the pK7 plasmid containing the gene [9]. An expression plasmid for chloramphenicol acetyltransferase with an N-terminal His tag was constructed in our previous study [8]. Genes encoding tRNA variants were cloned into the pUC119 plasmid (Takara, Shiga, Japan). Each tRNA variant is created by site-directed mutagenesis from a plasmid encoding tRNA^{Ala}_{CCA} [8]. In the mutagenesis experiments, polymerase chain reaction (PCR) amplification was performed with KOD-plus(Toyobo, Tokyo, Japan) using the following PCR conditions: 2 min at 94°C for 1 cycle, followed by 20 cycle of 0.25 min at 94°C, 0.5 min at 65°C and 3.2 min at 68°C.

Preparation of tRNA transcript

The tRNA variants were prepared by run-off transcription using T7 RNA polymerase and a PCR-amplified linear DNA template. The variants are purified using 8% urea polyacrylamide gel electrophoresis in a gel composed of TBE (Tris-EDTA-Borate Buffer; 5×) 8 ml, urea (19 g), and 40% acrylamide gel (29:1) 8ml. After denaturing PAGE, the band located by UV shadowing was cut out and was allowed to rotate for 12 hours with 1 ml

of 0.3 M NaCl.. Finally, the overnight product is sterilized by Millipore filtration (0.45 μM pore size).

Cell-free protein expression

The *Escherichia coli* S30 cell-free protein synthesis method was used in this study. The composition of the cell-free protein synthesis reaction has been previously described [10], except for the omission of a specific amino acid, and the addition of the tRNA variant and 5.0 μM a.a.-SA (5'-O-[N-(L-a.a.)sulfamoyl]adenosine aminoacyl adenylate analogs, Integrated DNA Technologies, Coralville, IA). The S30 extract was prepared from the *E.coli* BL21 (DE3) strain. The batch mode was employed, with 20-μl reaction volumes and a reaction time of 1h.

Detecting the radiolabeled products

Translation of CAT and Ras was performed using the 20 μl scale batch mode of synthesis at 37°C for 1 h with the components described above, except for the addition of [¹⁴C] Leu. The non-purified products were analyzed on 12% Bis-Tris gels with MES running buffer (50 mM MES, 50 mM Tris-base, 3.47 mM SDS, 1.0 mM EDTA, pH 7.3). Scanning was performed using an image analyzer, FLA-5000 (Fujifilm, Tokyo, Japan), and an imaging plate, BAS-IP MS 2040 (Fujifilm, Tokyo, Japan), to measure the radioactivity of the products.

Results

Construction of Arg-lacking simplified genetic code by a combination of tRNA^{Ala} variants.

Using multiple tRNA^{Ala} variants, we expanded our previous strategy for constructing a simplified genetic code. We designed 3 tRNA^{Ala} variants with the anticodon loop corresponding to Arg[8](Fig.1A). In the universal genetic code, 6 codons, with the sequences CGN and AGR, are assigned to Arg. We first unassigned those 6 codons by removing Arg from the *E. coli* S30 cell-free translation mixture and adding Arg-SA, a strong inhibitor of arginyl-tRNA synthetase (ArgRS). We then constructed 3 tRNA^{Ala} variants, which had the anticodons GCG, UCG, and UCU, respectively. We named these tRNA variants AR_GCG, AR_UCG and AR_UCU respectively. Ala is considered to be attached to these tRNA^{Ala} variants by alanyl-tRNA synthetase (AlaRS), because AlaRS does not recognize the anticodon loop [11-13]. Therefore, the tRNA^{Ala} variants can assign Ala to Arg codons (Fig.1B).

The tRNA variants is used in constructing a simplified genetic code are prepared by traditional biochemical methods. From a tRNA^{Ala}_{CCA}-expressing construct, each tRNA variant-expressing construct was generated by site-directed mutagenesis. The oligonucleotides necessary to create the plasmid by these methods are shown in Table1. The tRNA variants were prepared by run-off transcription using T7 RNA polymerase and a PCR-amplified linear DNA template. The transcription products were purified by urea polyacrylamide gel electrophoresis (see Materials and Methods). The AR_GCG (3.2

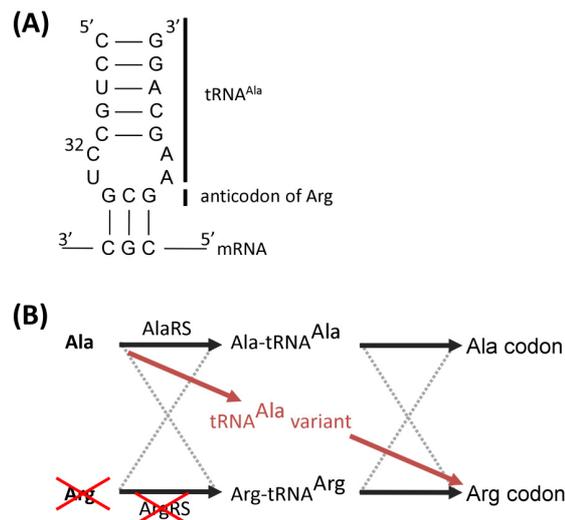


Fig.1. Strategy for constructing a simplified genetic code. (A) The anticodon stem loop sequence of the tRNA^{Ala} variant that was designed to translate the CGC codons for Arg on mRNA. (B) Schematic view of the simplified genetic code in the *E. coli* S30 cell-free translation mixture. The tRNA^{Ala} variant, with its anticodon corresponding to the CGC codon of Arg, assigns Ala to the codon. Arg is removed from the translation mixture. Arginyl-tRNA synthetase (ArgRS) is inhibited by Arg-SA.

Table 1. Oligonucleotides used to construct tRNA variants.

| Oligonucleotide name | Sequence(5'-3') |
|----------------------|------------------------------|
| FW primer | GCAGGAGGTCTGCGGTTGATC |
| AR_GCG_RV | CCGCAGACCTCTGCGTCGCAAGCAGGCG |
| AR_UCG_RV | CCGCAGACCTCTGCGTCGAAAGCAGGCG |
| AR_UCU_RV | CCGCAGACCTCTGCGTAGAAAGCAGGCG |

nmol), AR_UCG (2.9 nmol), and AR_UCU(2.9nmol) tRNA variants were produced from 150-μl reaction solution. These results showed that the efficiency of transcription by T7 RNA polymerase was not significantly affected by the anticodon sequence. In order to produce the 6 unassigned CGN and AGR codons that encode Arg in the universal code, we removed Arg from the mixture and added Arg-SA to the mixture. Arg-SA inhibits the activity of ArgRS that is required for protein synthesis. We confirmed that protein synthesis was reduced depending on the concentration of Arg-SA, with a range of 0.05 μM to 5 μM in the translation mixture (Fig.2A, lane 3-5). Finally, to reassign Ala to the unassigned codons, we added the 3 tRNA^{Ala} variants to the cell-free translation mixture with the 6 unassigned codons. Protein synthesis was increased depending on the addition of the tRNA variant mixture (1 μM each; Fig. 2B, lane 4). Based on this, we concluded that an Arg-lacking simplified genetic code was created by the 3 tRNA^{Ala} variants. Additionally, our results show that the Arg-lacking simplified genetic code can produce protein efficiently, on par with the universal genetic code (Fig. 2B, lane 1 and 4).

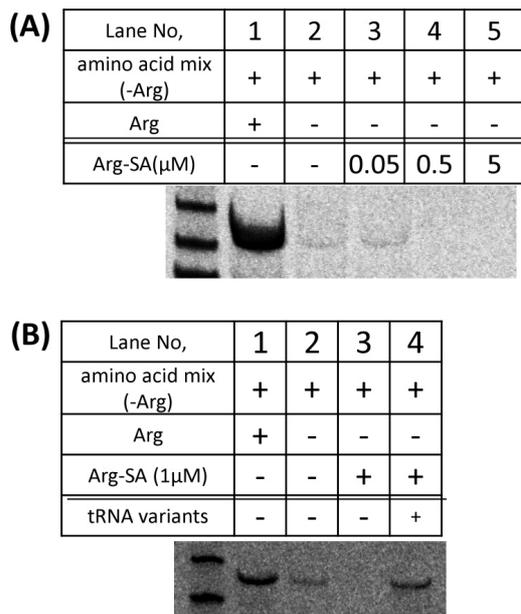


Fig.2. (A) Validation of the activity of Arg-SA. Ras was translated under the conditions noted at the top of each lane. (B) Reassignment of unassigned codons by 3 tRNA^{Ala} variants. CAT was translated under the conditions noted at the top of each lane. An autoradiogram of a polyacrylamide gel, with the products labeled with [¹⁴C] Leu, is shown.

Discussion and Conclusion

We describe an Arg-lacking simplified genetic code in which multiple tRNA^{Ala} variants are used. This represents our strategy of constructing a simplified genetic code to have the versatility to reassign amino acids. Using a set of tRNA variants that correspond to anticodons of multiple amino acids, we will be able to construct a more simplified genetic code which encodes fewer than 19 kinds of amino acids. Furthermore, we can construct various simplified genetic codes using a combination of tRNA^{Ala} and tRNA^{Ser}. Our previous work showed that tRNA^{Ser} variants could be used to construct a simplified genetic code [8].

In addition to this work, previous work on a Trp-lacking code and a Cys-lacking code (personal communication) produced unassigned codons by the addition of a.a.-SA, in the range of 0.5 to 5 μ M. We will be able to produce unassigned codons for other amino acids by the addition of 5 μ M a.a.-SA corresponding to the amino acid that we require to remove from the genetic code. However, for various reasons, we might have to change the concentration depending on the kind of the amino acid. For example, the proportion of each amino acid removed from the translation mixture by dialysis might be different. A few amino acids are known to be produced in the metabolism of S30 cell-free extracts [14]. When the amino acid is only partly removed, with a significant quantity remaining in the translation mixture, more a.a.-SA corresponding to the amino acid is needed. Additionally, the affinity between the a.a.-AMP of a particular amino acid and its cognate aaRS could be different from the affinity

between the a.a.-SA of the amino acid and its cognate aaRS. If the inhibition constant of a.a.-SA is higher than the Michaelis constant of a.a.-AMP, we will have to add more of the a.a.-SA corresponding to the amino acid.

Our results with respect to constructing a simplified genetic code using multiple tRNA variants pave a new path to create a more simplified genetic code with less than 19 amino acids. By removal of multiple amino acids from the genetic code, we can design various simplified genetic codes each containing sets of amino acids which are considered as elements of a hypothetical primordial code. For example, 10 amino acids obtained from prebiotic biosynthesis were considered to be encoded in the primordial genetic code proposed in phase 1 of the coevolution hypothesis [15, 16]. From other viewpoints, the phylogenetic analysis of aminoacyl-tRNA synthetases suggested the kinds of amino acids that may be present in the primordial genetic code [17, 18]. Since there are many other hypothesis about the kinds of amino acids in the primordial genetic code in each stage of evolution before commonotes, we refer to the reports by Trifonov, which summarizes various studies of simplified genetic codes [19].

We suggest that the creation of a simplified protein, as well as the construction of a simplified genetic code, provides new insights to assess early life forms. Although we do not know whether these simplified proteins are able to achieve all essential biological functions, it is generally accepted that early life forms used only simplified proteins that have produced by a simplified genetic code [19, 20]. Previous studies showed a functional protein composed of 9 kinds of amino acids [21]. In other studies, many simplified proteins with greater numbers of amino acids have been created [22, 23]. However, these previous experiments are not enough to show that an organism can live with a primordial genetic code that employs less than 20 amino acids. Thus, we consider it is necessary to create sets of simplified proteins that can realize an essential biological function such as energy metabolism and assignments by aaRSs.

The simplified genetic code with the removal of multiple amino acids will be a more effective tool to create a simplified protein. Previously, the creation of a simplified protein by selection from a random mutagenesis library using the universal genetic code has been frustrated by the reappearance of codons of an amino acid that is required to be removed from the library. When removing many amino acids from a protein, the probability of such reappearance becomes higher. In contrast, with the use of a simplified genetic code, every protein sequence translated from any mRNA in the library does not include the eliminated amino acids. The tRNA composition versatility that we have demonstrated by the usage of multiple tRNA variants in this study will be crucial for the creation of a more simplified protein that encodes fewer than 19 kinds of amino acids.

Acknowledgments

We acknowledge funding from the KAKENHI programs [19680016, 21680026, 23119005, and 23650155 to D.K.]; Japan Society for the Promotion of Science (JSPS) and the Ministry of Education, Culture, Sports, Science and Technology (MEXT).

We also thank Takanori Kigawa and Shigeyuki Yokoyama for providing the Ras constructs.

References

1. G. R. Moura, J. A. Paredes, M. A. S. Santos, Development of the genetic code: Insights from a fungal codon reassignment, *FEBS Letters*. 584, 334-341 (2010).
2. S. Osawa, *Evolution of the genetic code.* (Oxford University Press, New York, 1995).
3. Nureki *et al.*, Structure of an archaeal non-discriminating glutamyl-tRNA synthetase: a missing link in the evolution of Gln-tRNA^{Gln} formation, *Nucleic Acids Res.* 38, 7286-7297 (2010).
4. D. L. Tumbula, H. D. Becker, W. Z. Chang, D. Soll, Domain-specific recruitment of amide amino acids for protein synthesis, *Nature*. 407, 106-110 (2000).
5. K. Sheppard *et al.*, From one amino acid to another: tRNA-dependent amino acid biosynthesis, *Nucleic Acids Res.* 36, 1813-1825 (2008).
6. B. Min, J. T. Pelaschier, D. E. Graham, D. Tumbula-Hansen, D. Soll, Transfer RNA-dependent amino acid biosynthesis: an essential route to asparagine formation, *Proc Natl Acad Sci U S A.* 99, 2678-2683 (2002).
7. H. D. Becker, D. Kern, *Thermus thermophilus*: a link in evolution of the tRNA-dependent amino acid amidation pathways, *Proc Natl Acad Sci U S A.* 95, 12832-12837 (1998).
8. Kawahara-Kobayashi *et al.*, Simplification of the genetic code: restricted diversity of genetically encoded amino acids, *Nucleic Acids Res.* 40, 10576-10584 (2012).
9. T. Kigawa, Y. Muto, S. Yokoyama, Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis, *J Biomol NMR.* 6, 129-134 (1995).
10. T. Kigawa *et al.*, Preparation of *Escherichia coli* cell extract for highly productive cell-free protein expression, *J Struct Funct Genomics.* 5, 63-68 (2004).
11. Y. M. Hou, P. Schimmel, A simple structural feature is a major determinant of the identity of a transfer RNA, *Nature*. 333, 140-145 (1988).
12. C. Francklyn, P. Schimmel, Aminoacylation of RNA minihelices with alanine, *Nature*. 337, 478-481 (1989).
13. M. Guo *et al.*, The C-Ala domain brings together editing and aminoacylation functions on one tRNA, *Science.* 325, 744-747 (2009).
14. T. Kigawa *et al.*, Cell-free production and stable-isotope labeling of milligram quantities of proteins, *FEBS Lett.* 442, 15-19 (1999).
15. J. T. Wong, Coevolution theory of the genetic code at age thirty, *BioEssays.* 27, 416-425 (2005).
16. J. T. Wong, Coevolution of Genetic-Code and Amino-Acid Biosynthesis, *Trends in Biochemical Sciences.* 6, 33-36 (1981).
17. L. Ribas de Pouplana, P. Schimmel, Aminoacyl-tRNA synthetases: potential markers of genetic code development, *Trends Biochem Sci.* 26, 591-596 (2001).
18. L. Ribas de Pouplana, R. J. Turner, B. A. Steer, P. Schimmel, Genetic code origins: tRNAs older than their synthetases?, *Proc Natl Acad Sci U S A.* 95, 11295-11300 (1998).
19. E. N. Trifonov, Consensus temporal order of amino acids and evolution of the triplet code, *Gene.* 261, 139-151 (2000).
20. K. Miura *et al.*, Synthesis and expression of a synthetic gene for the activated human c-Ha-ras protein, *Jpn J Cancer Res.* 77, 45-51 (1986).
21. K. U. Walter, K. Vamvaca, D. Hilvert, An active enzyme constructed from a 9-amino acid alphabet, *J Biol Chem.* 280, 37742-37746 (2005).
22. S. Akanuma, T. Kigawa, S. Yokoyama, Combinatorial mutagenesis to restrict amino acid usage in an enzyme to a reduced set, *Proc Natl Acad Sci U S A.* 99, 13549-13553

- (2002).
23. C. E. Schafmeister, S. L. LaPorte, L. J. Miercke, R. M. Stroud, A designed four helix bundle protein with native-like structure, *Nat Struct Biol.* 4, 1039-1046 (1997).