J|A|C|S COMMUNICATIONS

Published on Web 02/01/2010

De Novo Generation of Mutually Orthogonal Aminoacyl-tRNA Synthetase/tRNA Pairs

Heinz Neumann, Adrian L. Slusarczyk, and Jason W. Chin*

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, U.K.

Received September 1, 2009; E-mail: chin@mrc-lmb.cam.ac.uk

The genetic code sets the correspondence between codons and the amino acids they encode in protein translation.[1] The code is enforced by two key processes: the aminoacylation of specific tRNAs with specific amino acids and the decoding of tRNA:mRNA (anticodon:codon) interactions by the ribosome.[2] To efficiently expand the genetic code of organisms we require new blank codons (since each of the 64 triplet codons is used in encoding proteome synthesis) and mutually orthogonal synthetase/tRNA pairs that recognize distinct amino acids and decode the blank codons.

The genetic code of *E. coli* has been expanded, allowing a range of functional groups to be site-specifically installed into proteins. This work has taken advantage of either evolved active site mutants of the *Methanococcus jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS)/$tRNA_{CUA}$ pair[3] or the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase (*Mb*PylRS)/$tRNA_{CUA}$ pair and its evolved active site mutants.[4−6] Using these pairs a single type of unnatural amino acid at a time can be incorporated, with limited efficiency,[7] in response to the amber stop codon.

We recently described the creation of an evolved orthogonal ribosome that efficiently decodes quadruplet codons, which are inefficiently decoded by the natural ribosome.[8,9] This provides a series of blank codons on the orthogonal mRNA, which is specifically translated by the orthogonal ribosome.[8] We have initially assigned one blank codon to an azide containing amino acid and another to an alkyne containing amino acid, using mutually orthogonal derivatives of the *Mj*TyrRS/tRNA and *Mb*PylRS/tRNA pairs.[9] This has allowed us to genetically direct the formation of specific, redox-insensitive, nanoscale protein cross-links, via bio-orthogonal cycloaddition reactions, and to explore the programming of emergent properties into proteins *via* combinations of unnatural amino acids.[9]

A central challenge in further reprogramming the genetic code is to discover additional mutually orthogonal aminoacyl-tRNA synthetase tRNA pairs. The orthogonality of the two useful synthetase/tRNA pairs identified so far results from the fortuitous evolutionary divergence of synthetase and tRNA sequence, over billions of years, between kingdoms of life.[3,4] However, the discovery of more naturally orthogonal pairs is likely to be fundamentally limited by the small number of orthogonal pairs that may have arisen by evolutionary drift and by the decreasing probability that arbitrarily combined pairs will display mutual orthogonality.

The evolutionary record suggests that a primitive genetic code expanded into the current genetic code, over billions of years, through duplication and specialization (neofunctionalization) of aminoacyl-tRNA synthetases and tRNAs from common ancestral synthetase/tRNA pairs.[10] This process produced the current set of mutually orthogonal aminoacyl-tRNA synthetases and tRNAs that direct natural protein synthesis.

In principle it should be possible to create new orthogonal synthetase/tRNA pairs from existing synthetase/tRNA pairs. This is a significant challenge that requires the *de novo* generation of orthogonality, with respect to the progenitor synthetase/tRNA module and all other synthetases and tRNAs present in the cell. The new tRNA must not recognize the progenitor synthetase, or any other synthetases in the cell, and must decode a new codon and not decode the original codon. The new synthetase must specifically aminoacylate the new tRNA with an amino acid and must not efficiently aminoacylate the progenitor tRNA or use the amino acid substrate of the original synthetase. At the molecular level this requires the *de novo* generation of orthogonality in RNA−RNA interactions, protein−RNA interactions, and small molecule substrate selection by protein catalysts.

While previous efforts to address the *de novo* creation of orthogonal synthetase tRNA pairs were unsuccessful[11] we now demonstrate that, by careful, structure guided design of combinatorial libraries and the implementation of a series of genetic selections on a duplicated pair, it is possible to evolve mutually orthogonal pairs from a single orthogonal pair. The mutually orthogonal pairs function in the same cell to specifically direct the incorporation of distinct amino acids in response to distinct codons.

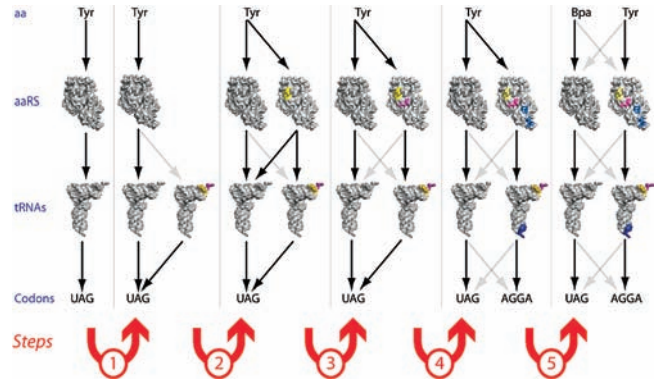

Figure 1. *De novo* generation of mutually orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA pairs. Black arrows indicate functional interactions; gray arrows indicate the absence of a functional interaction. Yellow, magenta, and blue patches indicate amino acids and nucleotides targeted for evolution in each step. Bpa (*p*-benzoyl-L-phenylalanine).

We began with the *Methanococcus jannaschii* (*Mj*) tyrosyl-tRNA synthetase (*Mj*TyrRS)/$tRNA_{CUA}$ pair, the active site of which has been altered to allow the incorporation of numerous

10.1021/ja9068722

unnatural amino acids in proteins in response to the amber codon.[3] In a first step (Figure 1, step 1) we created a version of *Mj*tRNA$_{CUA}$, which is neither a substrate for any of the endogenous aminoacyl-tRNA synthetases in *E. coli* nor a substrate for *Mj*TyrRS. The structure of *Mj*Tyr-tRNA synthetase in complex with its cognate tRNA and the characterized identity elements[12−14] suggest that the synthetase recognizes its cognate tRNA *via* interactions with the acceptor stem, discriminator base, and the anticodon of the tRNA (Supplementary Figures 1 and 2). We envisioned that combinations of mutations in these regions of the tRNA might provide orthogonal tRNAs. In this step we aimed to create a new amber suppressor tRNA and therefore focused on mutating the acceptor stem of the tRNA, rather than the anticodon. We created a library in which the C1-G72, C2-G71 base pairs were mutated to all other Watson−Crick pairs as well as the G-U, U-G wobble pairs to maintain the tRNA structure, and A73 was mutated to A, G, C, and U (Supplementary Figure 1). We screened this library of *Mj*-tRNA$_{CUA}$ variants for tRNAs which are not a substrate for any of the endogenous aminoacyl-tRNA synthetases or the wild-type *Mj*TyrRS (Figure 1, step 1). To do this we cotransformed *E. coli* with a vector containing *Mj*TyrRS and a library of pRep YC JYCUA vectors[15] (each vector in this library contains a chloramphenicol acetyl transferase (*cat*) gene containing an amber codon at a site permissive to many different amino acids and a tRNA which is a member of the library described above). We picked 400 individual colonies (oversampling the mutated sequence space 10-fold) and arrayed the resulting clones in the presence and absence of chloramphenicol (*Cm*). From this procedure we isolated 86 tRNA clones that do not survive on 10 μg mL^{-1} chloramphenicol, providing a pool of tRNAs that are functionally orthogonal to both the *E. coli* synthetases and *Mj*TyrRS.

To produce a variant of *Mj*TyrRS that is able to aminoacylate a tRNA isolated in the previous step we designed and created a synthetase library that mutates amino acids responsible for recognizing the C1-G72, C2-G71 pairs in the *Mj*tRNA acceptor stem (Arg 132, Ser 133, Arg 174, Lys 175, and Met 178) to all 20 amino acids (Supplementary Figure 1). This library has a theoretical diversity of 10^7 and is more than 99% complete. We selected synthetase enzymes from this library that aminoacylate a new orthogonal tRNA by selection in *E. coli* bearing selected pRep YC JYCUA clones on 75 μg mL^{-1} *Cm*. This selection revealed a single synthetase mutant (6A3-TyrRS) that is able to aminoacylate a selected tRNA (6A3-tRNA$_{CUA}$). 6A3-TyrRS has the mutations Arg174Trp, Lys175Arg, Met178Ile, Ala130Val, Glu135Gly, while 6A3-tRNA$_{CUA}$ has the mutations C2A, G71U, and A73G. The 6A3-tRNA alone or in combination with *Mj*TyrRS confers a *Cm* resistance of less than 10 μg mL^{-1}, while the selected 6A3-TyrRS/6A3-tRNA$_{CUA}$ pair confers resistance to *Cm* to 250 μg mL^{-1}. These results demonstrate that the 6A3-tRNA$_{CUA}$ is orthogonal to both *E. coli* synthetases and *Mj*TyrRS and that the 6A3-TyrRS/6A3-tRNA$_{CUA}$ pair functions to efficiently decode amber codons. However, the selected 6A3-TyrRS in combination with *Mj*tRNA$_{CUA}$ efficiently suppresses amber codons in a *cat* gene, conferring resistance to *Cm* to 200 μg mL^{-1}. This indicates that we have isolated a dual specificity synthetase that recognizes both the progenitor *Mj*tRNA$_{CUA}$ and the selected 6A3-tRNA$_{CUA}$ (Figure 1, step 2).

To create a 6A3-TyrRS variant that efficiently aminoacylates 6A3-tRNA$_{CUA}$, but does not efficiently aminoacylate the progenitor *Mj*tRNA$_{CUA}$ (Figure 1, step 3), we designed a second-generation synthetase library and performed three rounds of selection (positive, negative positive). This library mutates residues Ser 12, Asn 193, Pro 194, Val 195 in 6A3-TyrRS that form a pocket around the discriminator base in the *Mj*TyrRS/tRNA complex structure to all 20 amino acids (Supplementary Figure 1). This library has a theoretical diversity of 10^6 and was more than 99% covered by 10^8 independent transformants. In the first round of positive selection we transformed this library in cells containing pRep 6A3-tRNA$_{CUA}$ (a variant of pRep YC JYCUA that contains the 6A3-tRNA$_{CUA}$ gene in place of the progenitor tRNA) and selected for synthetase/6A3-tRNA$_{CUA}$ pairs that can confer chloramphenicol resistance *via* amber suppression. For the negative selection the surviving synthetase clones were isolated and used to transform cells harboring a reporter plasmid that allows selection against an interaction of the synthetases with the progenitor *Mj*tRNA$_{CUA}$. This reporter plasmid codes for barnase.[16] Suppression of amber codons in the barnase gene leads to expression of the enzyme, killing the cell. After a third round of positive selection, 20 synthetases were screened for selectivity in amber suppression with 6A3-tRNA$_{CUA}$ versus *Mj*tRNA$_{CUA}$. Synthetase clones (10, Supplementary Tables 1 and 3) conferred chloramphenicol resistance up to 175 μg mL^{-1} on cells in the presence of 6A3-tRNA$_{CUA}$ but only 50−70 μg mL^{-1} in the presence of the *Mj*tRNA$_{CUA}$. This suggests that these selected synthetases have a reversed specificity with respect to the progenitor synthetase and specifically aminoacylate 6A3-tRNA$_{CUA}$ in preference to *Mj*tRNA$_{CUA}$.

While the selected synthetase/tRNA pair and progenitor pair are mutually orthogonal in their interaction, both *Mj*tRNA$_{CUA}$ and 6A3-tRNA$_{CUA}$ decode the same insertion signal on mRNAs, the amber codon (UAG). To differentiate the insertion codons used by the pairs (Figure 1, step 4), we altered the anticodon of 6A3-tRNA$_{CUA}$ from CUA to UCCU, which decodes the quadruplet codon AGGA, albeit inefficiently, to create XtRNA$_{UCCU}$. Expression of myoglobin-his6 from *myohis6(4AGGA)* (which contains an AGGA codon at position 4 of a myoglobin gene) using three 6A3-TyrRS derivatives (Supplementary Table 1) that are specific for 6A3-tRNA$_{CUA}$ produced little full length protein when paired with XtRNA$_{UCCU}$ (data not shown). This is consistent with XtRNA$_{UCCU}$ being a poor substrate for these synthetases, as expected based on the previous reports that the C-terminal domain of *Mj*TyrRS recognizes the anticodon and that alterations in the anticodon decrease aminoacylation efficiency 10^3-fold.[13] To discover synthetase variants that function with XtRNA$_{UCCU}$ we created a library in the C-terminal domain of the synthetase by mutating residues Tyr 230, Cys 231, Pro 232, Phe 261, His 283, Asp 286 (Supplementary Figure 2), using the pool of 10 6A3-tRNA$_{CUA}$ specific synthetases (Supplementary Table 1) as a template. We selected synthetase variants *via* three rounds of selection, positive - negative - positive, for the ability to function with XtRNA$_{UCCU}$ to decode the AGGA codon, while maintaining the ability to discriminate against *Mj*tRNA$_{CUA}$. The selected synthetases function in combination with XtRNA$_{UCCU}$ to decode an AGGA codon in chloramphenicol acetyl transferase and confer chloramphenicol resistance to cells up to 450 μg mL^{-1} (Supplementary Tables 2 and 4).

To demonstrate that we have created an aminoacyl-tRNA synthetase/tRNA pair that is mutually orthogonal to the progenitor pair, we first examined the ability of *Mj*TyrRS or a selected XTyrRS (containing the mutations Ser12Gly, Asn193Ser, Pro194Ile, Val195Leu, Tyr230Arg, Cys231Lys, Pro232Gln, His283Ala, and Asp286Tyr with respect to 6A3-TyrRS) in combination with *Mj*tRNA$_{CUA}$ or XtRNA$_{UCCU}$ to read amber or AGGA codons in *myohis6* (4TAG) or *myohis6*(4AGGA) (Figure

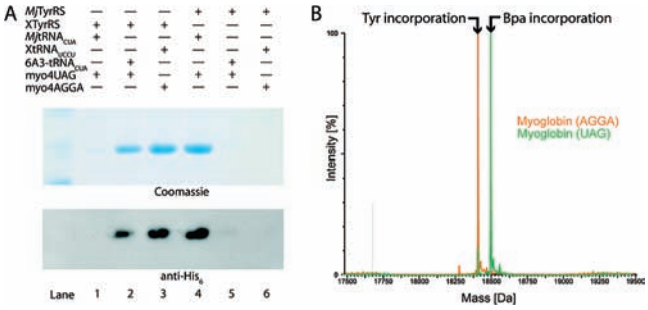

Figure 2. The XTyrRS/XtRNA$_{UCCU}$ pair is orthogonal to the *Mj*BPARS/*Mj*tRNA$_{CUA}$ pair. (A) Myoglobin-His$_6$ was purified from *E. coli* and stained with Coomassie or detected with an anti-His antibody. (B) Myoglobin-His$_6$ was expressed in the presence of 1 mM Bpa in cells containing *Mj*BPARS, XTyrRS, *Mj*tRNA$_{CUA}$, and XtRNA$_{UCCU}$ and a plasmid coding for myoglobin-His$_6$ with either an AGGA codon at position 4 (orange trace) or an amber codon at position 75 (green trace). The purified proteins were analyzed by ESI mass spectrometry. Tyr (tyrosine), Bpa (*p*-benzoyl-L-phenylalanine).

2A). We find that full-length myoglobin is produced by the *Mj*TyrRS/*Mj*tRNA$_{CUA}$ pair and the XTyrRS/XtRNA$_{UCCU}$ pair, but not by the *Mj*TyrRS/XtRNA$_{UCCU}$ pair or the XTyrRS/*Mj*tRNA$_{CUA}$ pair. These experiments demonstrate that each synthetase specifically directs amino acid incorporation when provided with its cognate, but not noncognate, tRNA. However this does not demonstrate that each synthetase pair is functionally independent when both pairs are placed within the same cell. This is a nontrivial distinction since the wild-type synthetase forms a dimer which binds two tRNA molecules and there are therefore are 10 possible multimeric complexes that could potentially assemble in cells that contain the two different synthetases and the two different tRNAs (Supplementary Figure 3). To assess the net functional consequences of these potential assemblies on amino acid incorporation, we first differentiated the amino acid specificity of the evolved and progenitor aminoacyl-tRNA synthetase (Figure 1, step 5) by replacing *Mj*TyrRS with *Mj*BpaRS (an active site variant of *Mj*TyrRS that aminoacylates *Mj*tRNA$_{CUA}$ with *p*-benzoyl-L-phenylalanine (Bpa)[16]). We combined *Mj*BpaRS, *Mj*tRNA$_{CUA}$, XTyrRS, XtRNA$_{UCCU}$ in the same cell and read out the amino acids encoded in response to amber or AGGA codons using myoglobin-his6 genes containing either a UAG or AGGA codon (Figure 2B). Electrospray ionization mass spectrometry (ESI-MS) of myoglobin purified from cells containing a myoglobin-his6 gene with an AGGA codon demonstrates the incorporation of tyrosine in response to this codon and gives no peak corresponding to the incorporation of Bpa. Similarly ESI-MS of myoglobin purified from cells containing a myoglobin-his6 gene with a TAG codon demonstrates the specific (~90%) incorporation of Bpa in response to this codon. These experiments demonstrate that we have created mutually orthogonal synthetase/tRNA pairs from a single orthogonal pair. The pairs function independently in the same cell; the synthetases use distinct amino acids to aminoacylate distinct tRNAs, and these tRNAs decode distinct codons. Overall the progenitor and evolved synthetases differ at 14 positions and are 95% identical, while the tRNAs differ at 6 positions and are 92% identical. In future work it will be interesting to determine the structural, thermodynamic, and kinetic basis for the *in vivo* specificities we have reported.

We have shown that it is possible to create new orthogonal aminoacyl-tRNA synthetase/tRNA pairs, which are mutually orthogonal with existing orthogonal pairs, *de novo*, by a logical series of steps implemented in the laboratory. This provides experimental evidence for duplication and specialization as a plausible route to the current set of synthetases and tRNAs in natural evolution.[10] Moreover our experiments extend billions of years of natural evolution and demonstrate that the small number of naturally occurring orthogonal aminoacyl-tRNA synthetase/tRNA pairs do not place an intrinsic limitation on the scope of synthetic genetic code expansion for the incorporation of multiple distinct unnatural amino acids into proteins or the synthesis and evolution of unnatural polymers in cells.

Acknowledgment. This work was supported by the Medical Research Council and the European Research Council.

Supporting Information Available: Experimental protocols and supplementary figures. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Crick, F. H.; Barnett, L.; Brenner, S.; Watts-Tobin, R. J. *Nature* **1961**, *192*, 1227–32.
(2) Schmeing, T. M.; Ramakrishnan, V. *Nature* **2009**, *461*, 1234–42.
(3) Xie, J.; Schultz, P. G. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775–82.
(4) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Chem. Biol.* **2008**, *4*, 232–4.
(5) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. *Chem. Biol.* **2008**, *15*, 1187–97.
(6) Nguyen, D. P.; Lusic, H.; Neumann, H.; Kapadnis, P. B.; Deiters, A.; Chin, J. W. *J. Am. Chem. Soc.* **2009**, *131*, 8720–1.
(7) Wang, K.; Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Biotechnol.* **2007**, *25*, 770–7.
(8) Rackham, O.; Chin, J. W. *Nat. Chem. Biol.* **2005**, *1*, 159–66.
(9) Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. *Nature* **2010**, in press (doi: 10.1038/nature08817).
(10) Ribas de Pouplana, L.; Schimmel, P. *Trends Biochem. Sci.* **2001**, *26*, 591–6.
(11) Liu, D. R.; Magliery, T. J.; Pastrnak, M.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10092–7.
(12) Fechter, P.; Rudinger-Thirion, J.; Tukalo, M.; Giege, R. *Eur. J. Biochem.* **2001**, *268*, 761–7.
(13) Kobayashi, T.; Nureki, O.; Ishitani, R.; Yaremchuk, A.; Tukalo, M.; Cusack, S.; Sakamoto, K.; Yokoyama, S. *Nat. Struct. Biol.* **2003**, *10*, 425–32.
(14) Steer, B. A.; Schimmel, P. *J. Biol. Chem.* **1999**, *274*, 35601–6.
(15) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nat. Biotechnol.* **2002**, *20*, 1044–8.
(16) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11020–4.

JA9068722