JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

A Rationally Designed Pyrrolysyl-tRNA Synthetase Mutant with a Broad Substrate Spectrum

Yane-Shih Wang, Xinqiang Fang, Ashley L. Wallace, Bo Wu, and Wenshe R. Liu*

Department of Chemistry, Texas A&M University, College Station, Texas 77843, United States

Supporting Information

ABSTRACT: Together with $tRNA^{Pyl}_{CUA}$, a rationally designed pyrrolysyl-tRNA synthetase mutant N346A/C348A has been successfully used for the genetic incorporation of a variety of phenylalanine derivatives with large *para* substituents into superfolder green fluorescent protein at an amber mutation site in *Escherichia coli*. This discovery greatly expands the genetically encoded noncanonical amino acid inventory and opens the gate for the genetic incorporation of other phenylalanine derivatives using engineered pyrrolysyl-tRNA synthetase-$tRNA^{Pyl}_{CUA}$ pairs.

In the past decade, we have seen genetic incorporation of many noncanonical amino acids (NAAs) into proteins at nonsense mutation sites in *Escherichia coli*, *Saccharomyces cerevisiae*, and mammalian cells using unique aminoacyl-tRNA synthetase (aaRS)-nonsense suppressing tRNA pairs that do not cross-interact with endogenous aaRS-tRNA pairs.[1−7] Except for the wild-type (wt) pyrrolysyl-tRNA synthetase (PylRS) that has been directly used for genetic incorporation of >10 NAAs,[8−12] most of these unique aaRSs were evolved via complicated positive and negative selection systems.[1,2,13−16] So far, evolution is still the primary approach to identify NAA-specific aaRSs. The prerequisites of the referred aaRS evolution are the construction of a large mutant aaRS gene library and a readily available selection system.[1] Although several easily accessible selection systems have been developed for *E. coli* and *S. cerevisiae* cells, constructing a large mutant aaRS gene library that most of time needs to cover random variations of 5−6 aaRS active-site residues is not straightforward and needs an experienced molecular biologist to practice many times to achieve close to the coverage of the library variations.[17] Statistically, full coverage is not likely. One factor that also significantly influences the aaRS evolution is the selection pressure. Positive and negative selections used in the evolution are based on either antibiotic resistance or toxic gene expression. Varying concentrations of antibiotics and inducers that are used to trigger toxic gene expression could lead to dramatic different evolution results. In comparison to the evolution approach, rational design of aaRSs is relatively more straightforward and easier to be carried out by following standardized site-directed mutagenesis protocols. However, attempts to rationally design unique NAA-specific aaRSs often led to mutant aaRSs with nonexclusive recognition of endogenous canonical amino acids (CAAs).[18,19] In this work, we show that a rationally designed PylRS mutant N346A/N348A displays specific recognition toward NAAs and has a broad substrate spectrum.

We previously evolved two PylRS mutants that display specific recognition of phenylalanine.[21] In both mutants, N346 is mutated to alanine and C348 is mutated to a larger amino acid, leucine or lysine. No mutation at other sites was found. Figure 1 shows the structure of the PylRS complex with pyrrolysyl-AMP. Phenylalanyl-AMP was also modeled into the active site of the N346A mutant of PylRS and is shown as an overlay with pyrrolysyl-AMP in Figure 1. The structure in Figure 1 suggests how two mutations N346A and C348L (or C348K) induce the substrate specificity change from pyrrolysine (Pyl) to phenylalanine. In the wt PylRS, the side-chain amide nitrogen of N346 forms a hydrogen bond with the side-chain amide oxygen of Pyl, an interaction that anchors Pyl at the active site. The amide of N346 also shows a steric clash with the modeled phenylalanine in the active site, excluding its binding to the wt PylRS. The N346A mutation will not only significantly decrease the binding of PylRS to Pyl and also relieve the steric hindrance that prevents the binding of phenylalanine. Since both the aromatic side chain of Y384 and two backbone amides of residues 419−421 could form $\pi-\pi$ stacking interactions with the phenyl group of a bound phenylalanine, we think that dismissing the steric hindrance

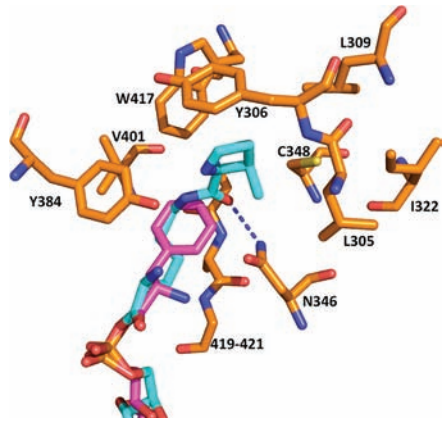

Figure 1. Structure of the PylRS complex with pyrrolysyl-AMP. Phenylalanyl-AMP that potentially binds the N346A mutant is shown as an overlay with pyrrolysyl-AMP. The structure is based on the PDB entry 2Q7H.[20]

Received: December 22, 2011
Published: January 30, 2012

   dx.doi.org/10.1021/ja211972x | *J. Am. Chem.Soc.* 2012, 134, 2950−2953

between a bound phenylalanine and N346 is the major contributing factor for the direct binding of two PylRS mutants to phenylalanine. The mutation of C348 to an amino acid with a larger side chain that apparently occupies the space of pyrrolidine of Pyl in the active site is believed to bring more interactions with the bound phenylalanine and therefore increase its binding potential. Since the N346A mutation relieves the steric hindrance that prevents the binding of phenylalanine and there is left a large empty space at the active site around the *para* position of phenylalanine after it binds to the N346A mutant of PylRS, we speculate that the N346A mutant could bind a phenylalanine derivate with a large *para* substituent better than phenylalanine itself. Given that mutations at C348 were prevalent in almost all evolved NAA-specific PylRS mutants and its mutation to alanine was observed in PylRS mutants specific for NAAs with large side chains,[22−25] we think an extra C348A mutation to the N346A mutant will generate a larger active site pocket that could provide a *para*-substituted phenylalanine with more structural flexibility to bind to the active site. To test this idea, we constructed a plasmid pBK-N346A/C348A that carries the gene coding the PylRS mutant N346A/C348A and used it together with pET-pylT-sfGFP2TAG to transform BL21 cells. The plasmid pET-pylT-sfGFP2TAG carries genes coding $tRNA^{Pyl}_{CUA}$ and an IPTG-inducible superfolder green fluorescent protein (sfGFP) with an amber mutation at the S2 position. The transformed BL21 cells were then used to examine the recognition of N346A/C348A toward 20 CAAs by expressing sfGFP in liquid glycerol minimal media (GMML) supplemented with 5 mM of a designated CAA. The sfGFP expression was induced by the addition of 1 mM IPTG. Since the sfGFP expression levels at all twenty conditions were low, we chose to detect the fluorescent emission of the expressed sfGFP at these conditions and show their relative intensities to represent the corresponding sfGFP expression levels in Figure 2. The condition that contained 5 mM phenylalanine displayed the highest sfGFP expression level, confirming our initial speculation that the N346A mutation can change the substrate specificity of PylRS from Pyl to phenylalanine. To see whether N346A/C348A can recognize a phenylalanine derivative with a large *para* substituent, the same BL21 cells were grown in GMML supplemented with 5 mM *p*-propargyloxy phenylalanine (**1**, Figure 3). In comparison to the condition with 5 mM phenylalanine, the expressed sfGFP at this condition displayed a 5-fold increase in the fluorescence emission intensity.

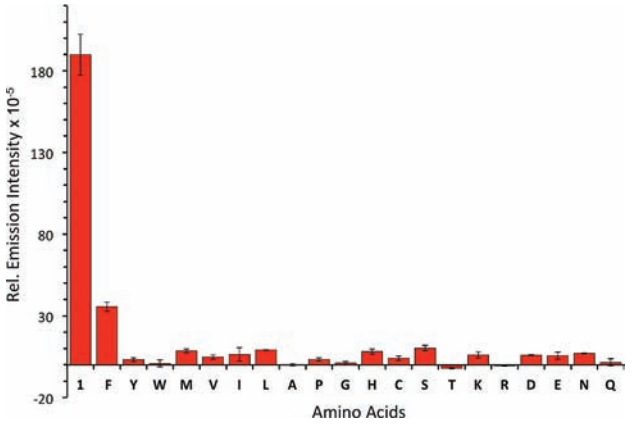

Figure 2. Relative fluorescence emission intensities of sfGFP expressed in BL21 cells transformed with pBK-N346A/C348A and pET-pylT-sfGFP2TAG and grown in GMML supplemented with different amino acids. Cell lysates were excited at 450 nm and fluorescence emission intensities were detected at 510 nm. Background emission from cell lysate of same cells grown in GMML and induced with the addition of 1 mM IPTG was subtracted from each data set. Twenty CAAs are shown as one-letter abbreviations in the *x*-axis labels.

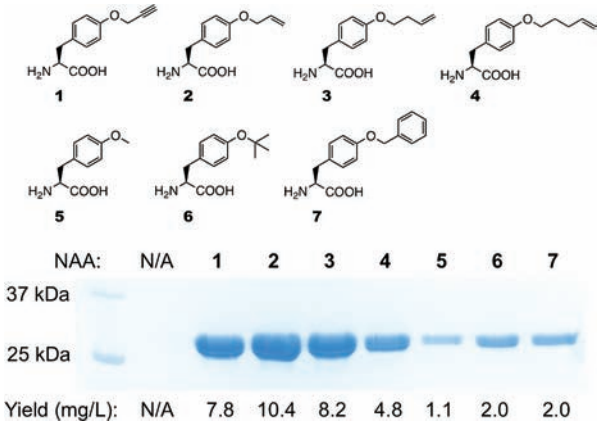

Figure 3. Structures of **1**−**7** and their specific incorporation into sfGFP at the S2 position.

With our initial success with **1**, we then either purchased or synthesized several other *para*-substituted phenylalanine derivatives shown as **2**−**7** in Figure 3 and tested their recognition by N46A/C348A. To better quantify the expression levels of sfGFP incorporated with these NAAs, another plasmid pEVOL-pylT-N346A/C348A was constructed. This plasmid carries genes coding both $tRNA^{Pyl}_{CUA}$ and N346A/C348A. Its $tRNA^{Pyl}_{CUA}$ is under control of a strong *proK* promoter that can boost up the expression level of $tRNA^{Pyl}_{CUA}$ in *E. coli*.[26] Together with pET-pylT-sfGFP2TAG, this plasmid was used to transform BL21 cells. The transformed cells were grown in GMML supplemented with 2 mM of a designated NAA (**7** can only be provided as 1 mM because of its low solubility). As shown in Figure 3, all seven NAAs promoted high levels of sfGFP expression 10 h after induction with 1 mM IPTG. The molecular weights of the expressed sfGFP variants determined by electrospray ionization spectrometry (ESI-MS) analysis agree well with their theoretical molecular weights (SI Table 1 and Figures 2−8). Without a NAA in the medium, no sfGFP was expressed. With the addition of 2 mM phenylalanine in the medium, an sfGFP expression level (1.5 mg/mL) close to that for **5** was obtained. Since **1** contains a terminal alkyne that undergoes the Cu(I)-catalyzed azide−alkyne cyclization reaction,[27] the expressed sfGFP incorporated with **1** (sfGFP-**1**) was also used separately to label with a fluorescein azide (**8** in Figure 4) in an optimized labeling condition that contained 0.1 mM Cu(I):tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine (TBTA) complex, 0.4 mM additional TBTA, 5 mM ascorbate, and 1 mM $NiCl_2$.[28] A 3-h incubation led to specific labeling of sfGFP-**1**. A parallel labeling reaction of the wt sfGFP in the same condition gave nondetectable labeling with **8**.

One intriguing question is whether N346A/C348A could also recognize a phenylalanine derivative with a small *para* substituent. Given that **5** has a relatively small *para* substituent and providing **5** in the growth medium led to a reasonable sfGFP expression level, one would expect N346A/C348A also recognizes a phenylalanine derivative with a small *para*

dx.doi.org/10.1021/ja211972x | *J. Am. Chem.Soc.* 2012, 134, 2950−2953

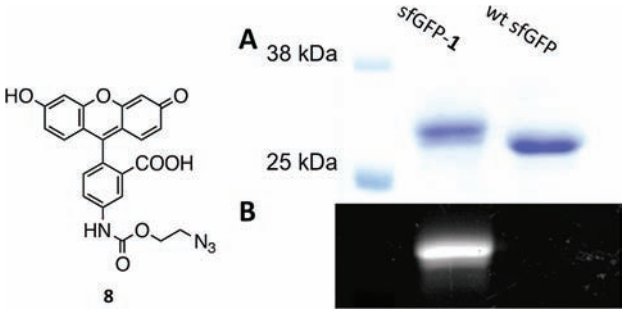

Figure 4. Site-selective labeling of sfGFP-**1** with **8**. (A) SDS-PAGE analysis of sfGFP-**1** and wt sfGFP after their reactions with **8**. The gel was stained with Coomassie blue. (B) Fluorescent imaging of the same gel under 365 nm UV irradiation.

substituent such as Cl, Br, I, CN, etc. To test this possibility, we examined the genetic incorporation of **9**−**15** shown in Figure 5

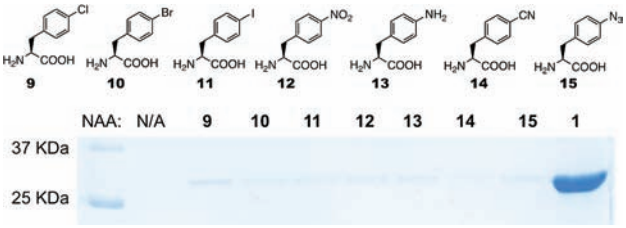

Figure 5. Structures of **9**−**14** and their genetic incorporation into sfGFP at S2. The protein expression yields are <1 mg/L for all NAAs.

into sfGFP at S2 using the N346A/C348A- $tRNA^{Pyl}_{CUA}$ pair. As shown in Figure 5, providing 2 mM of any of these NAAs in GMML led to sfGFP expression levels that are significantly lower than those for NAAs shown in Figure 3 and phenylalanine but still higher than the background expression level in GMML in which no NAA was provided. Therefore, N346A/C348A recognizes **9**−**15**. Since the expression levels of sfGFP incorporated with **9**−**15** are very low, we did not attempt to characterize these proteins by the ESI-MS analysis. Although the current analysis indicates that it is not applicable to use the N346A/C348A-$tRNA^{Pyl}_{CUA}$ pair to express proteins incorporated with **9**−**15**, it suggests that PylRS can be engineered to recognize phenylalanine derivatives with small *para* substituents. Since a PylRS mutant specific for **5** evolved by Wang and co-workers contains mutations at A302 and V401,[29] We are now introducing additional mutations at these two sites of N346A/N348A to search for PylRS mutants that allow efficient incorporation of phenylalanine derivatives with small *para* substituents.

In summary, we have rationally designed a PylRS mutant N346A/C348A that shows low recognition toward CAAs in minimal media but, together with $tRNA^{Pyl}_{CUA}$, mediates efficient incorporation of NAAs **1**−**7** into proteins at amber mutation sites in *E. coli*. These NAAs contain functional groups such as alkyne and alkene and can be applied to install different biochemical and biophysical probes to proteins for their structural and functional analysis. Since the PylRS-$tRNA^{Pyl}_{CUA}$ pair has been successfully introduced into *S. cerevisiae*, mammalian cells, and even multiple cellular organisms,[11,23,24,30,31] the N346A/C348A-$tRNA^{Pyl}_{CUA}$ pair could be potentially used in these systems to genetically encode **1**−**7**. Although many phenylalanine derivatives have been incorporated into proteins in *E. coli* using evolved *Methanococcus jannaschii* tyrosyl-tRNA synthetase (MjTyrRS)-$tRNA^{Tyr}_{CUA}$ pairs,[32−34] specifically evolved MjTyrRS variants for individual phenylalanine derivatives are usually required and the MjTyrRS-$tRNA^{Tyr}_{CUA}$ pair cannot be used in eukaryotic cells because of the recognition of $tRNA^{Tyr}_{CUA}$ by endogenous eukaryotic aaRSs (Liu and Schultz, unpublished data). Using the N346A/C348A-$tRNA^{Pyl}_{CUA}$ pair will resolve both issues. In addition, phenylalanine derivatives **3**, **4**, **6**, and **7** that are taken by N346A/C348A are also genetically encoded in *E. coli* for the first time. Given that N346A/C348A has a relatively deep and big binding pocket, the current study also opens a gate to test the recognition of this mutant toward other large phenylalanine derivatives. Another potential application of N346A/C348A is to couple its pair with $tRNA^{Pyl}_{UUA}$ together with evolved MjTyrRS-$tRNA^{Tyr}_{CUA}$ pairs for the genetic incorporation of two different phenylalanine derivatives into one protein.[35] This may find applications in enzyme mechanistic studies, protein FRET labeling, and phage-displayed unnatural peptide library construction.

■ ASSOCIATED CONTENT

Supporting Information

Plasmid constructions, NAA synthesis, and protein expression. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

wliu@chem.tamu.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by the National Institutes of Health (grant 1R01CA161158 to W.R.L.) and the Welch Foundation (grant A-1715 to W.R.L.). We thank Dr. Yohannes H. Rezenom from Laboratory for Biological Mass Spectrometry at Texas A&M University for characterizing our proteins using ESI-MS.

■ REFERENCES

(1) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498.
(2) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. *Science* **2003**, *301*, 964.
(3) Liu, W.; Brock, A.; Chen, S.; Schultz, P. G. *Nat Methods* **2007**, *4*, 239.
(4) Wang, L.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2004**, *44*, 34.
(5) Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G. *J. Am. Chem. Soc.* **2004**, *126*, 14306.
(6) Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7566.
(7) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Chem. Biol.* **2008**, *4*, 232.
(8) Srinivasan, G.; James, C. M.; Krzycki, J. A. *Science* **2002**, *296*, 1459.
(9) Blight, S. K.; Larue, R. C.; Mahapatra, A.; Longstaff, D. G.; Chang, E.; Zhao, G.; Kang, P. T.; Green-Church, K. B.; Chan, M. K.; Krzycki, J. A. *Nature* **2004**, *431*, 333.
(10) Polycarpo, C. R.; Herring, S.; Berube, A.; Wood, J. L.; Soll, D.; Ambrogelly, A. *FEBS Lett.* **2006**, *580*, 6695.
(11) Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818.

(12) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. *Chem. Biol.* **2008**, *15*, 1187.
(13) Melancon, C. E. III; Schultz, P. G. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3845.
(14) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nat. Biotechnol.* **2002**, *20*, 1044.
(15) Wang, L.; Schultz, P. G. *Chem. Commun.* **2002**, *1*, 1.
(16) Umehara, T.; Kim, J.; Lee, S.; Guo, L.-T.; Soll, D.; Park, H.-S. *FEBS Lett.* DOI:10.1016/j.febslet.2012.01.024.
(17) Xie, J.; Liu, W.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2007**, *46*, 9239.
(18) Zhang, D.; Vaidehi, N.; Goddard, W. A. III; Danzer, J. F.; Debe, D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6579.
(19) Tanrikulu, I. C.; Schmitt, E.; Mechulam, Y.; Goddard, W. A. III; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15285.
(20) Kavran, J. M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Soll, D.; Steitz, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11268.
(21) Wang, Y. S.; Russell, W. K.; Wang, Z.; Wan, W.; Dodd, L. E.; Pai, P. J.; Russell, D. H.; Liu, W. R. *Mol. Biosyst.* **2011**, *7*, 714.
(22) Wang, Y. S.; Wu, B.; Wang, Z.; Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Moe, Y. N.; Russell, D. H.; Liu, W. R. *Mol. Biosyst.* **2010**, *6*, 1557.
(23) Gautier, A.; Nguyen, D. P.; Lusic, H.; An, W.; Deiters, A.; Chin, J. W. *J. Am. Chem. Soc.* **2010**, *132*, 4086.
(24) Chen, P. R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Geierstanger, B. H.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2009**, *48*, 4052.
(25) Zhang, M.; Lin, S.; Song, X.; Liu, J.; Fu, Y.; Ge, X.; Fu, X.; Chang, Z.; Chen, P. R. *Nat. Chem. Biol.* **2011**, *7*, 671.
(26) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. *J. Mol. Biol.* **2010**, *395*, 361.
(27) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2001**, *40*, 2004.
(28) Wu, B.; Wang, Z.; Huang, Y.; Liu, W. R. Manuscript in preparation.
(29) Takimoto, J. K.; Dellas, N.; Noel, J. P.; Wang, L. *ACS Chem. Biol.* **2011**, *6*, 733.
(30) Greiss, S.; Chin, J. W. *J. Am. Chem. Soc.* **2011**, *133*, 14196.
(31) Hancock, S. M.; Uprety, R.; Deiters, A.; Chin, J. W. *J. Am. Chem. Soc.* **2010**, *132*, 14819.
(32) Liu, C. C.; Schultz, P. G. *Annu. Rev. Biochem.* **2010**, *79*, 413.
(33) Xie, J.; Schultz, P. G. *Nat. Rev. Mol. Cell. Biol.* **2006**, *7*, 775.
(34) Wang, L.; Xie, J.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225.
(35) Wan, W.; Huang, Y.; Wang, Z.; Russell, W. K.; Pai, P. J.; Russell, D. H.; Liu, W. R. *Angew. Chem., Int. Ed.* **2010**, *49*, 3211.