ChemPubSoc Europe

DOI: 10.1002/cbic.201700039

CHEMBIOCHEM Full Papers

Two-Tier Screening Platform for Directed Evolution of Aminoacyl–tRNA Synthetases with Enhanced Stop Codon Suppression Efficiency

Andrew E. Owens, Katherine T. Grasso, Christine A. Ziegler, and Rudi Fasan*[a]

Genetic code expansion through amber stop codon suppression provides a powerful tool for introducing non-proteinogenic functionalities into proteins for a broad range of applications. However, ribosomal incorporation of noncanonical amino acids (ncAAs) by means of engineered aminoacyl–tRNA synthetases (aaRSs) often proceeds with significantly reduced efficiency compared to sense codon translation. Here, we report the implementation of a versatile platform for the development of engineered aaRSs with enhanced efficiency in mediating ncAA incorporation by amber stop codon suppression. This system integrates a white/blue colony screen with a plate-based colorimetric assay, thereby combining high-throughput capabilities with reliable and quantitative measurement of aaRS-dependent ncAA incorporation efficiency. This two-tier functional screening system was successfully applied to obtain a pyrrolysyl–tRNA synthetase (PylRS) variant (CrtK-RS(4.1)) with significantly improved efficiency (+250–370%) for mediating the incorporation of N^{ε}-crotonyl-lysine and other lysine analogues of relevance for the study of protein post-translational modifications into a target protein. Interestingly, the beneficial mutations accumulated by CrtK-RS(4.1) were found to localize within the noncatalytic N-terminal domain of the enzyme and could be transferred to another PylRS variant, improving the ability of the variant to incorporate its corresponding ncAA substrate. This work introduces an efficient platform for the improvement of aaRSs that could be readily extended to other members of this enzyme family and/or other target ncAAs.

Introduction

Methods for genetic code expansion provide valuable strategies for introducing non-proteinogenic functionalities into recombinant proteins.[1] A particularly versatile approach in this context is the amber stop codon suppression methodology, which relies upon the use of an engineered aminoacyl–tRNA synthetase (aaRS)/$tRNA_{CUA}$ pair for reassigning an amber stop codon (TAG) with a desired noncanonical amino acid (ncAA).[1a] This methodology has enabled the site-selective incorporation of ncAAs for a variety of applications, including protein conjugation and imaging,[2] controlling and modulating protein/enzyme function,[3] and generating post-translationally modified proteins,[4] among others.[5] Pyrrolysyl-tRNA synthetases (PylRSs) from *Methanosarcina* species[6] represent a particularly attractive choice for genetic code expansion, in large part due to the orthogonal reactivity of PylRSs and cognate $tRNA^{Pyl}$ molecules in both bacterial (*Escherichia coli*) and eukaryotic cells.[7] Initially identified for their ability to incorporate pyrrolysine in response to in-frame UAG codons,[6] PylRSs were found to be able to recognize a number of (pyrro)lysine derivatives.[4b,8]

A well-known challenge associated with the use of amber stop codon suppression is the competition between the ribosomal incorporation of the target ncAA with termination of translation in response to the stop codon. As a result, the ncAA incorporation efficiency of orthogonal aaRSs, in particular by those derived from PylRSs, is often reduced compared to sense codon translation, resulting in moderate expression yields for the target ncAA-containing protein. Over the past years, several strategies have been investigated to address this limitation, which include optimization of the promoters and copy numbers of aaRS and/or tRNA genes,[9] engineering of the tRNA molecule,[10] removal/manipulation of components of the translational machinery (e.g., EF-Tu, RF1),[11] optimization of sequence contexts,[12] and genome-wide engineering of the host cell.[13] These methods have proven effective toward improving the yield of ncAA-containing proteins, but they are not specifically aimed at improving the properties of the aaRSs. In fact, whereas major efforts have focused on altering the substrate specificity of aaRSs—a goal typically achieved through powerful selection-based methods[1a]—far fewer studies have been directed at improving the intrinsic properties of these enzymes as part of a stop codon suppressor system. Recently, Matsuura and co-workers reported a liposome-based in vitro evolution method for this purpose.[14] In spite of the high-throughput capability of this system, the *Methanosarcina mazei* PylRS variants isolated according this method showed either parent-like or only slightly improved (+70%) in vivo incorporation efficiency compared to the parent enzyme, which further highlighted the challenge of developing aaRSs with inherently improved in vivo activity for ncAA mutagenesis.

[a] *A. E. Owens, K. T. Grasso, C. A. Ziegler, Prof. R. Fasan*
Department of Chemistry, University of Rochester
Hutchinson Hall, Rochester, NY 14627 (USA)
E-mail: rfasan@ur.rochester.edu

Supporting information and the ORCID identification numbers for the authors of this article can be found under https://doi.org/10.1002/cbic.201700039.

Here, we report the development of a versatile platform useful for the engineering of aaRSs with enhanced efficiency for mediating ncAA incorporation through amber stop codon suppression. This strategy was applied to evolve variants of *Methanosarcina* sp. PylRSs with significantly improved efficiency toward the incorporation of N^{ε}-crotonyl-lysine (CrtK) and other lysine analogues relevant to the investigation of protein post-translational modifications into a target protein. These studies also demonstrate the value of targeting noncatalytic regions of PylRS enzymes to improve their activity in the context of unnatural mutagenesis by amber stop codon suppression.

Results and Discussion

Two-tier functional screening platform for aaRS evolution

Selection-based methods have provided a valuable and widely adopted strategy for aaRS engineering.[1a] These methods rely on coupling amber stop codon suppression to cell survival and death, as resulting from the expression of an antibiotic resistance marker (e.g., chloramphenicol acetyl transferase) in the presence of a target ncAA and from the expression of a toxic protein (e.g., barnase) in the absence of the ncAA, respectively. Coupled with active site mutagenesis, this approach has proven highly effective for isolating aaRS variants with altered substrate specificity and orthogonal reactivity over natural amino acids, but various factors make it suboptimal for evolving aaRSs with improved ncAA incorporation efficiency. Among these factors is the difficulty of tuning stringency during the selection process without triggering survival mechanisms unrelated to the desired activity,[15] which can lead to a high frequency of false positives.[16] In addition, the use of an antibiotic resistance marker as a reporter of ncAA incorporation implies that cell survival ensues when the essential enzyme is produced in sufficient amounts to generate antibiotic resistance, hampering the selection of aaRS variants with further improved incorporation efficiency. In this regard, the use of reporter proteins such as green fluorescent protein (GFP) can provide a more reliable and quantitative means for monitoring aaRS-mediated ncAA incorporation,[17] but this approach has found application mainly in the screening of small collections of aaRS variants,[17–19] with a notable exception.[11d]

Based on these considerations, we aimed to implement a two-tier screening platform that could combine medium-/high-throughput capability with a quantitative readout of ncAA incorporation in a streamlined manner. As outlined in Figure 1, this system was designed to comprise an initial phenotypic screen based on a LacZ reporter gene containing two amber stop codons (TAG). Efficient suppression of both stop codons results in a functional β-galactosidase enzyme, whose expression can be detected based on a standard white/blue colony assay in the presence of the chromogenic substrate X-Gal (Figure 1, step A). This step offers high-throughput capability (typically, up to ten- to 50 000-member libraries), while uncoupling the phenotypic readout (blue/white colony) from cell survival. It also permits a first qualitative assessment of ncAA incorporation based on the blue color intensity of the colony. Positive colonies (blue) identified in this first-tier screen are then arrayed on multiwell plates, in which the levels of ncAA-containing galactosidase are measured in a quantitative manner by using a soluble, chromogenic substrate (*para*-nitrophenol-galactose or PNP-Gal; Figure 1, step B). As both experiments share the same reporter enzyme, transfer of the hits from the tier 1 screen to the tier 2 screen is both rapid and technically straightforward. Promising variants emerging from the tier 2 screen can be subjected to another round of directed evolution (Figure 1, step C) and/or evaluated by using a secondary functional assay (Figure 1, step D) in which the incorporation efficiency of the evolved aaRS is measured by using a yellow fluorescent protein (YFP) reporter protein containing an amber stop (TAG) codon at a permissive site. The orthogonal reactivity of the evolved aaRS enzyme toward the desired ncAA over the natural amino acids can be assessed by comparing β-galactosidase activity (Tier-2 screen) or YFP expression (secondary assay) in the presence and absence of the ncAA. A

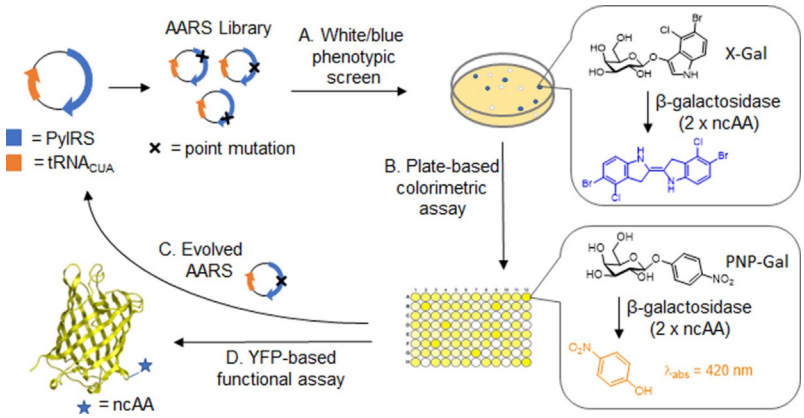

Figure 1. Schematic overview of the two-tier screening system for aaRS evolution. A) A library of engineered aaRS variants is first subjected to a phenotypic screen by using *E. coli* cells containing a LacZ(2xTAG) gene. aaRS-mediated amber stop codon suppression with the target ncAA leads to the expression of a functional β-galactosidase enzyme and processing of X-gal substrate. B) Positive clones (blue colonies) are then evaluated in a plate-based β-galactosidase activity assay with PNP-Gal, enabling quantification of ncAA suppression efficiency. C) Improved aaRSs are subjected to another round of directed evolution or, alternatively, D) evaluated for improved ncAA suppression efficiency by using the YFP(TAG) reporter protein.

dual vector system, in which one plasmid directs the expression of the reporter protein (β-galactosidase or YFP), and a second plasmid encodes for the aaRS/tRNA pair, facilitates the transfer of the aaRS-encoding plasmid from the tier 2 screen to the YFP-based functional assay. Of note, the LacZ gene was previously used as a reporter for amber stop codon suppression,[20] but it has never been applied to guide the engineering of aaRS enzymes.

Directed evolution of crotonyl-lysyl-tRNA synthetase

A previously reported *Methanosarcina barkeri* PylRS variant (CrtK-RS)[4e] capable of recognizing crotonyl lysine (CrtK, Figure 2A) was selected as the target aaRS for validating the system outlined above. Lysine crotonylation is a rare form of protein post-translational modification (PTM),[21] and CrtK-RS has proven useful for the preparation and functional investigation of histone proteins bearing this PTM at selected positions.[22] Random mutagenesis was applied as the evolutionary tool in order to identity mutations that could enhance PylRS-mediated ncAA incorporation while having minimal impact on the substrate recognition properties and selectivity of the enzyme. The latter result was anticipated, given that amino acid residues in direct contact with the substrate represent only a small fraction (<3%) of the entire amino acid sequence of PylRS.

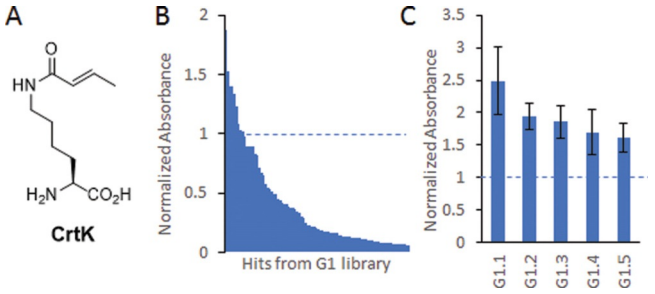

Figure 2. Hit evaluation in second-tier functional assay. A) Chemical structure of crotonyl lysine (CrtK). B) β-Galactosidase activity of top 90 hits (blue colonies) from first-generation aaRS library (G1), as determined by a PNP-Gal hydrolysis assay in a 96-well plate. C) β-Galactosidase activity of best-performing aaRS variants from the library based on rescreening ($n=3$). For (B) and (C), the absorbance values are normalized to that of cells expressing the parent enzyme (CrtK-RS).

Accordingly, an initial random mutagenesis library of CrtK-RS variants (Lib-G1; $\approx 10^4$ members) was generated by using error-prone PCR under conditions that yielded an estimated mutational rate of 2.4 nucleotide mutations per gene. The enzyme library was expressed in *E. coli* TOP10F′ cells containing a low-copy-number plasmid encoding the LacZ alpha fragment with two amber stop codons within the N-terminal region. Upon growth on lysogeny broth (LB) agar medium supplemented with CrtK and X-gal, recombinant colonies expressing functional β-galactosidase as a result of CrtK incorporation were readily identified as blue colonies (≈20% of plated library). Based on color intensity, about 90 positive hits were then arrayed and rescreened in a 96-well plate format. Upon cell lysis, β-galactosidase activity was quantified in a colorimetric assay with *para*-nitrophenol-galactose, with the absorbance intensity thus providing a quantitative measure of the efficiency of CrtK suppression by action of the CrtK-RS mutant (Figure 2B). Upon rescreening, five improved CrtK-RS variants were identified from the initial Lib-G1 library, as evidenced by the 1.5- to 2.5-fold higher β-galactosidase activity measured in the assay compared to the parent enzyme (Figure 2C). Among them, the most promising variant, G1.1, was determined to contain a single mutation at position 36 (Table 1).

Table 1. Amino acid mutations in evolved aaRS enzymes.

Position	8	13	36	45	121	253	355
CrtK-RS	V	T	I	H	S	I	I
G1.1			V				
G2.1		I	V			V	
G3.1		I	V		R	V	
G4.1	E	I	V	L	R		T

Following a similar strategy, a second-generation random mutagenesis library (Lib-G2) was prepared and screened by using G1.1 as the parent enzyme. This process yielded a further improved PylRS variant, G2.1 (Figure S1B in the Supporting Information), which carries two mutations (T13I and I253V) in addition to the preexisting I36V substitution (Table 1). Starting from G2.1, a third round of random mutagenesis and screening led to identification of CrtK-RS variant G3.1 (Figure S1D), which contains an additional mutation at position 121 (S121R). At this point, a recombination library (Lib-G4) was prepared by shuffling the genes corresponding to G3.1 and to the 4–5 best performing variants identified after each round of directed evolution. Screening of this library resulted in the identification of four CrtK-RS variants with increased suppression efficiency compared to G3.1 based on the β-galactosidase assay (Figure S1F). Among them, the most promising variant (G4.1) was named CrtK-RS(4.1) and selected for further analysis.

Functional analysis and location of mutations in CrtK-RS(4.1)

Sequence analysis revealed that CrtK-RS(4.1) contains a total of six amino acid substitutions, namely V8E, T13I, I36V, H45L, S121R, and I355T. Three of these mutations were shared with G3.1 (Table 1), whereas the V8E mutation could be traced back to one of the improved variants isolated after the third round of directed evolution (G3.3; Table S2). The beneficial effect of each of these mutations was evident from side-by-side comparison of related variants identified throughout the process and with respect to their relative ncAA incorporation efficiency, as measured by the β-galactosidase assay (Figure 2 and Figure S1). For example, whereas two mutations were accumulated during the transition from G1.1 to G2.1, the functional importance of each of them became apparent from comparison of β-galactosidase activity observed with G2.5 compared to G2.1 (T13I mutation) and with G2.1 compared to G2.3 (I253V

mutation; Figure S1). Interestingly, during the DNA shuffling step (G3.1→G4.1), the T13 mutation was maintained, whereas the comparatively less important I253V mutation was not.

In contrast to the aforementioned mutations, the H45L and I355T substitutions were not present in the original pool of genes used for recombination and thus were most likely introduced during assembly and/or amplification of the shuffled DNA library. To examine the functional importance of these unexpected mutations, reversion variants were generated in which each of these substitutions was removed from CrtK-RS(4.1). Analysis of these CrtK-RS variants confirmed that both mutations provide a beneficial, albeit relatively modest, activity improvement (+20%; Figure S2). Based on the analyses outlined above, we concluded that all of the mutations accumulated in the best-performing variant, CrtK-RS(4.1), contribute to its improved activity for CrtK incorporation through amber stop codon suppression in bacterial cells.

Mb PylRS is a 47.5 KDa class II aminoacyl-tRNA synthetase comprising an N-terminal domain (residues 1–194) and a C-terminal domain (residues 210–419) connected through a linker region (16 residues).[23] The C-terminal domain encompasses the catalytic site involved in the activation of pyrrolysine as an aminoacyl-AMP adduct and its subsequent conjugation to the tRNA molecule. The N-terminal domain was previously determined to be important for tRNA binding[24] and essential for in vivo activity.[23] In light of this information, it is particularly interesting to note that the largest majority of the beneficial mutations (10/13 = 77%), accumulated in CrtK-RS(4.1) and identified throughout the course of the directed evolution experiment (Tables 1 and S2), was localized within the N-terminal, noncatalytic domain of the enzyme. The only exception was I355T, which resides within the C-terminal catalytic domain but at a rather distant location with respect to the bound substrate (≈12–15 Å), as derived by inspection of the homologous C-terminal domain of *M. mazei* PylRS (Figure S3).[25] The relative contribution of this particular mutation to the improved properties of CrtK-RS(4.1) is, however, modest compared to the other mutations (Figure S2). Thus, whereas prior PylRS engineering efforts have largely relied on mutating the active site of these enzymes,[7b] our results indicate that there are significant opportunities to enhance the amber stop codon efficiency of these enzymes through mutation of their noncatalytic N-terminal domain. Although the mechanisms of the beneficial mutations within this region are currently unclear, the native function of this domain suggests that they might be related to improving the interaction of the enzymes with their cognate tRNA molecules and/or other cellular components.

Improved CrtK incorporation efficiency

In *E. coli* cultures grown in flasks, CrtK-RS(4.1) was determined to exhibit a 370% improvement in CrtK incorporation efficiency based on the β-galactosidase assay (Figure 3A). To further assess the enhanced performance of the evolved PylRS variants toward mediating the ribosomal incorporation of CrtK into a protein of interest, a YFP variant containing an N-terminal amber stop codon was used. As shown in Figure 3B, progressive improvement in the expression levels of YFP(CrtK) was observed in the presence of the evolved aaRS enzymes, a trend consistent with that obtained with β-galactosidase as the reporter protein (Figure 3A). Based on the YFP reporter protein, CrtK-RS(4.1) showed a 250% higher CrtK incorporation efficiency compared to the initial enzyme. In addition, minimal levels of background expression were observed in both assays with cells grown in the absence of CrtK (Figure 3A and B), indicating that all evolved CrtK-RS variants maintained high specificity toward the ncAA over the pool of natural amino acids. This result was consistent with the observation that all of the beneficial mutations accumulated in these variants were located in positions remote from the enzyme active site. To confirm the selective incorporation of CrtK, YFP expressed in the presence of CrtK-RS(4.1) was purified by Ni-NTA affinity chromatography and characterized by SDS-PAGE and LC-MS (Figure 4). The observed mass of this protein was found to match the expected one, thus confirming efficient and specific incorporation of CrtK.

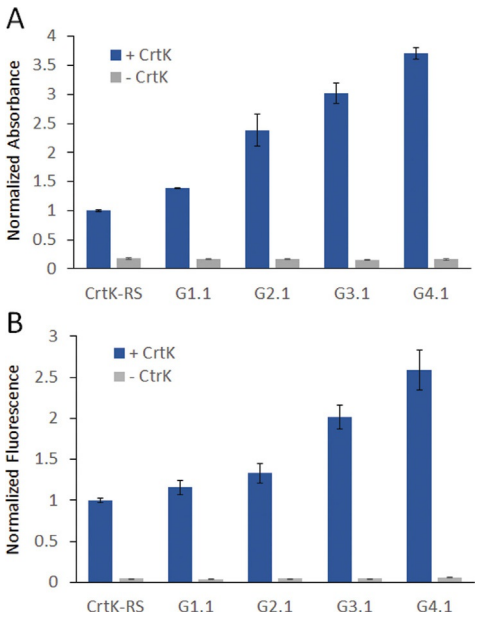

Figure 3. Suppression efficiency of evolved aaRS variants. Side-by-side comparison of CrtK incorporation efficiency for the best-performing aaRS from each round of directed evolution, as determined by A) β-galactosidase activity assay, and B) using YFP(TAG) as reporter protein. Background activity/fluorescence for cells grown in the absence of CrtK is shown. Values are normalized to those of the parent enzyme (CrtK-RS). Mean values and standard errors were obtained from three independent cell cultures.

Improved efficiency for incorporation of other N$^{\epsilon}$-acylated lysine analogues

In addition to *N*$^{\epsilon}$-crotonyl-lysine and *N*$^{\epsilon}$-acetyl-lysine (AcK), other N$^{\epsilon}$-acylated forms of lysine have been identified as a result of post-translational modification events in eukaryotic organisms, including *N*$^{\epsilon}$-propionyl (PrK) and *N*$^{\epsilon}$-butyryl lysine (BuK).[26] As these modified lysine analogues are poorly recognized by wild-type PylRS,[4f,22] PylRS variants have been engineered to incorporate these analogues into proteins with vary-

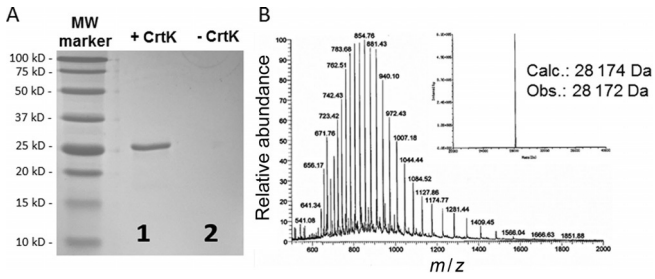

Figure 4. CrtK incorporation in YFP. A) SDS-PAGE gel of CrtK-containing YFP (28 kDa) after purification from *E. coli* in the presence CrtK-RS(4.1) and CrtK (lane 1). No background expression is observed in the absence of the ncAA (lane 2). B) ESI-MS spectra of purified CrtK-containing YFP.

ing degrees of substrate specificity.[22] In the interest of examining whether the improved properties of CrtK-RS(4.1) extended to other lysine analogues beyond CrtK, the incorporation of BuK, PrK, and AcK in the presence of the evolved aaRS was investigated (Figure 5 A). These experiments showed CrtK-RS(4.1) was able to provide improved expression yields for BuK-containing YFP (+220%) as well as for the PrK-containing counterpart, which could not be obtained by using the parent CrtK-RS enzyme. Negligible levels of AcK-containing YFP were obtained with both CrtK-RS and CrtK-RS(4.1), indicating that the enhanced properties of the evolved CrtK-RS variant were mostly confined to lysine analogues that contain a side-chain acyl group of similar length to that of CrtK.

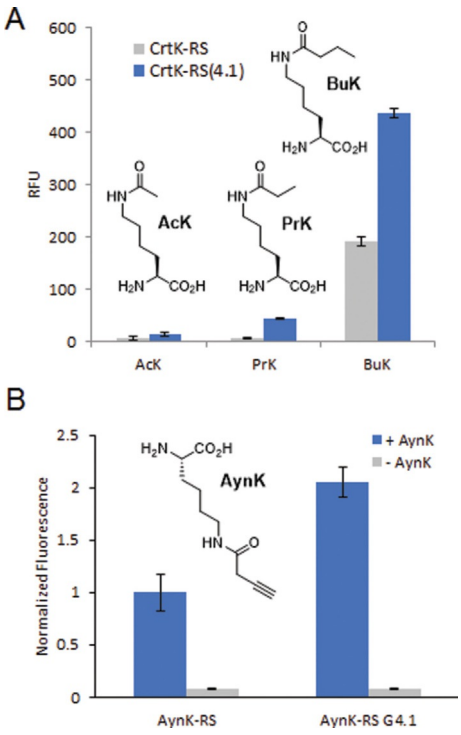

Figure 5. Incorporation efficiency with alternative lysine analogues. A) Fluorescence response corresponding to expression of YFP(TAG) reporter protein in the presence of CrtK-RS(G4.1) or CrtK-RS, and various N^{ε}-acylated lysine analogues. Reported values indicate fluorescence signal after background subtraction. B) Fluorescence response corresponding to expression of YFP(TAG) reporter protein in the presence and in the absence of AynK using cells containing AynK-RS or AynK-RS(G4.1). Values are normalized to that of AynK-RS. Mean values and standard errors were obtained from three independent cell cultures.

Transferability of activity-enhancing mutations to other *Mb* PylS variants

Our finding that the beneficial mutations carried by CrtK-RS(4.1) were clustered within the N-terminal domain of the enzyme further suggested that they could be potentially transferred to other *Mb* PylRS variants. Indeed, *Mb* PylRS variants evolved for recognition of different lysine derivatives typically differ from each other by amino acid substitutions localized only within the substrate binding pocket. To examine this point, the six mutations identified in CrtK-RS(4.1) were introduced into a *Mb* PylRS variant (AynK-RS) previously reported for the incorporation of the alkyne-containing ncAA N^{ε}-[(2-propynyloxy)carbonyl]-L-lysine (AynK).[27] To our delight, a 2.2-fold improvement in expression levels for the model YFP protein was observed in the presence of the modified AynK-RS enzyme, AynK-RS(4.1), compared to AynK-RS under identical experimental conditions (Figure 5 B). Also, in this case, the improvement in incorporation efficiency for the corresponding ncAA did not come at the expense of specificity, as indicated by the comparably low levels of background fluorescence measured upon protein expression in the absence of AynK.

Conclusion

In conclusion, this work introduced and validated an efficient platform for the directed evolution and improvement of aaRS enzymes. Particularly advantageous features are the combination of high-throughput capability, as enabled by the phenotypic white/blue colony screen, with a quantitative readout of ncAA incorporation from a tandem functional assay. As such, this system provides a valuable complement to established selection-based methods for the discovery and improvement of orthogonal aaRS/tRNA pairs for genetic code expansion. By using the present approach, an improved PylRS variant was obtained for the incorporation of crotonyl-lysine and other N-acylated lysine analogues associated with human PTMs into a protein of interest. This evolved aaRS is expected to contribute a valuable and versatile tool for the production of proteins modified site-selectively with crotonyl, propionyl, or butyryl lysine, thus facilitating efforts toward elucidating the functional roles of these forms of PTMs.

Finally, these studies provide a first-time demonstration of the value of mutating the N-terminal, noncatalytic domain of PylRS enzymes to enhance their amber stop codon suppression efficiency in vivo. Importantly, these mutations did not affect the specificity of the aaRS and could be transferred to another *Mb* PylRS variant, leading to a significant improvement in suppression efficiency for its corresponding ncAA. These findings are expected to aid future efforts in PylRS engineering and suggest that domain engineering strategies[28] could be useful for the evolution of this class of enzymes.

Experimental Section

Plasmid construction: Custom-made oligonucleotides were purchased from IDT DNA Technologies, and their sequences are reported in Table S1. The vector encoding the LacZ(2xTAG) gene (pZS13_LacZ_2x_Stop) was derived from the low-copy-number plasmid pZS13 (Expressys), which contains a pSC101 origin of replication and an ampicillin marker (*bla*). pZS13_LacZ_2x_Stop was prepared by PCR amplification of the LacZ alpha gene by using primers LacZ2X_F and LacZ2X_R, followed by cloning of the resulting PCR product into the NcoI/XbaI cassette of pZS13. This process placed the LacZ(2xTAG) gene, which contains a TAG codon, in place of Met3 and Ser7, under an IPTG-inducible promoter ($P_{A1lacO-1}$). The preparation of the plasmid encoding for the YFP(TAG) protein (pET22_YFP_1X_Stop) was described previously.[5e] The CrtK–RS/tRNA pair was encoded by a modified form of the pEVOL[9a] vector—pEVOL(1x)—which contains a p15A origin of replication and chloramphenicol marker (*cat*). In this plasmid, one copy of the aaRS gene is under an arabinose-inducible (araBAD) promoter, whereas the tRNA gene is under a constitutive proK promoter. For library generation, the CrtK-RS gene was amplified by using primers CrtK_F and CrtK_R and inserted into the BglII and SalI cassette of the pEVOL(1x) vector. The ligation product was transformed into DH5α cells and plated on LB agar plates supplemented with chloramphenicol (34 μg mL^{-1}). AynK-RS G4.1 was generated by swapping the N-terminal domain of AynK-RS with that of G4.1, followed by introduction of the I355T mutation. The N-terminal domain of G4.1 was amplified with primers SOE_G4.1F and SOE_G4.1R and fused to the C-terminal region of the AynK-RS gene (amplified with primers SOE_AynKF and SOE_AynKR) by sequential overlap extension (SOE). With the SOE product as a template, mutation I355T was inserted by a second SOE reaction using gene fragments generated with primers SOE_G4.1F/SOE_I355TR and SOE_I355TF/SOE_AynKR. With CrtK-RS(4.1) as a template, CrtK-RS(4.1)-T355I was generated by SOE using gene fragments generated from amplification of SOE_G4.1F/ SOE_T355I_R and SOE_T355I_F/SOE_G4.1_R. CrtK-RS(4.1)-L45H was generated in the same way by using CrtK-RS(4.1) as a template and the gene products of SOE_G4.1F/SOE_L45H R and SOE_L45H_F/ SOE_G4.1_R in a SOE reaction. All plasmids and library clone sequences were confirmed by DNA sequencing.

Library preparation: The CrtK gene was mutagenized by error-prone PCR using *Taq* polymerase (NEB), oligonucleotides CrtK_F and CrtK_R as primers, and the plasmid encoding for CrtK-RS, G1.1, or G1.2 as template. The PCR reaction was performed by using a standard reaction mixture supplemented with dTTP (1.0 mM), dCTP (1.0 mM), $MgCl_2$ (3.5 mM) and $MnCl_2$ (0.2 mM). This error-prone PCR protocol was previously reported to provide minimal bias in nucleotide misincorporation.[29] The PCR product was amplified according to the method: 2 min at 95 °C (1 cycle); 30 s at 95 °C, 30 s at 60 °C, 1 min 30 s at 72 °C (20 cycles). The resulting PCR product (1.2 Kbp) was ligated into the BglII and SalI cassette of the pEVOL(1x) vector. The average error rate for the library was determined by sequencing ten randomly chosen recombinant colonies. The DNA shuffling library was generated by first pooling the PCR products from each top preforming aaRS variant (G1.1–G3.5) by using primers ePCR_F and ePCR_R and *Taq* polymerase (NEB). The genes were mixed in equal amounts to reach a final DNA concentration was 1 μg mL^{-1}. The DNA was digested with DNase I (NEB) for 4 min at room temperature in the presence of $MnCl_2$ (10 mM) and immediately purified by agarose gel electrophoresis. DNA fragments between 10–100 bp were collected and reassembled by using *Taq* polymerase according to the method: 2 min at 95 °C (1 cycle); 30 s at 95 °C, 30 s at 60 °C, 1 min at 72 °C (45 cycles). The resulting 1.2 kbp product was ligated into the BglII and SalI cassette of the pEVOL(1x) vector, transformed into DH5α cells and plated on LB agar plates supplemented with chloramphenicol (34 μg mL^{-1}).

Library screening: For library screening, each plasmid library was co-transformed with pZS13_LacZ_2x_Stop in TOP10F′ cells and plated on LB agar plates supplemented with ampicillin (100 μg mL^{-1}), chloramphenicol (34 μg mL^{-1}), arabinose (0.06%), IPTG (1 mM), X-Gal (20 μg mL^{-1}), and CrtK (2 mM). After incubation for 18 h at 37 °C, blue colonies were selected from the plates and regrown to saturation (overnight) in a master 96-deep-well plate containing LB medium (1 mL per well), supplemented with ampicillin (100 μg mL^{-1}) and chloramphenicol (34 μg mL^{-1}). A replicate plate with M9 medium (1 mL per well) containing ampicillin (100 μg mL^{-1}) and chloramphenicol (34 μg mL^{-1}) was generated from the master plate and grown to an OD_{600} of 0.5–0.6, followed by addition of arabinose (0.06%) and CrtK (2 mM). After 1 h at 37 °C, the wells were induced with IPTG (1 mM) and grown for 18 h at 27 °C. Cells were recovered by centrifugation at 4300 *g* in a swinging plate rotor, and decanted pellets were frozen at −80 °C. The cell pellets were then resuspended in lysis buffer (50 mM potassium phosphate, 150 mM NaCl, 10 mM $MgCl_2$, 0.8 μg mL^{-1} DNase, and 0.8 μg mL^{-1} lysozyme, pH 7.5) and incubated at 37 °C for 1 h and 15 min. Insoluble cellular debris was pelleted by centrifugation at 4300 *g* at 4 °C. The clarified lysate (130 μL) was transferred to a 96-well microtiter plate, and *para*-nitrophenol-galactose in Kpi buffer (20 μL of a 37.5 mM solution; final concentration of 5 mM) was added, followed by incubation at 37 °C in an orbital shaker (120 rpm) for 20 min. Absorbance at 420 nm was monitored by using a Tecan Infinity 1000 plate reader. Positive hits identified at this stage were rescreened in triplicate according to the same procedure as described above. Plasmids encoding the most promising aaRS variants were then isolated from the master plate by plasmid extraction from overnight culture, followed by digestion with EcoRI and separation of the pEVOL(1x)-based plasmid from digested pZS13_LacZ_2x_Stop plasmid by gel electrophoresis. The purified pEVOL(1x)-based plasmids were re-transformed into DH5α cells, and transformant cells were selected on LB agar plates containing chloramphenicol (34 μg mL^{-1}). Mutations in the aaRS gene were identified by DNA sequencing.

YFP assay: The plasmid encoding YFP(TAG) and the pEVOL(1x)-based vector encoding the desired aaRS were co-transformed in BL21(DE3) cells and grown to saturation in LB medium supplemented with ampicillin (100 μg mL^{-1}) and chloramphenicol (34 μg mL^{-1}). Overnight cultures were used to inoculate M9 medium (1 mL per well), supplemented with ampicillin (100 μg mL^{-1}) and chloramphenicol (34 μg mL^{-1}), in a 96-deep-well plate and were then allowed to reach an OD_{600} of 0.5–0.6. Cultures were then induced with arabinose (0.06%), ncAA (2 mM) was added, and cultures were grown at 37 °C for 1 h. Cultures were then induced with IPTG (1 mM) and grown for 18 h at 27 °C. Control cultures were prepared in a similar manner but without ncAA addition. Fluorescence measurements were obtained from cell cultures (150 μL) by using a Tecan Infinity 1000 plate reader (λ_{ex} = 514 nm, λ_{em} = 527 nm). Mean values and errors were obtained from three independent experiments.

β-Galactosidase assay: The plasmid encoding pZS13_LacZ_2x_Stop and the pEVOL(1x)-based vector encoding the desired aaRS were co-transformed in TOP10F′ cells and grown to saturation in LB medium supplemented with ampicillin (100 μg mL^{-1}) and chloramphenicol (34 μg mL^{-1}). Overnight cultures were used to in-

oculate M9 medium (5 mL) supplemented with ampicillin (100 μg mL^{-1}) and chloramphenicol (34 μg mL^{-1}), and cultures were allowed to reach an OD_{600} of 0.5–0.6. Cultures were then induced with arabinose (0.06%), CrtK (2 mM) was added, and the cultures were grown at 37 °C for 1 h. Cultures were then induced with IPTG (1 mM) and grown for 18 h at 27 °C. Control cultures were prepared in a similar manner but without addition of CrtK. Cells were recovered by centrifugation at 4300 *g*. The cells were then resuspended in lysis buffer (50 mM potassium phosphate, 150 mM NaCl, 10 mM $MgCl_2$, 0.8 μg mL^{-1} DNase, and 0.8 μg mL^{-1} lysozyme, pH 7.5) and incubated at 37 °C for 1 h and 15 min. Insoluble cellular debris was pelleted by centrifugation at 4300 *g*. The clarified lysate (130 μL) was transferred to a 96-well microtiter plate, and *para*-nitrophenol-galactose in Kpi buffer (20 μL of a 37.5 mM solution; final concentration 5 mM) was added, followed by incubation at 37 °C in an orbital shaker (120 rpm) for 20 min. Absorbance at 420 nm was monitored by using a Tecan Infinity 1000 plate reader. Mean values and errors were obtained from three independent experiments.

Protein expression and isolation: BL21(DE3) cells were transformed with the plasmid pET22_YFP_1X_Stop and pEVOL(1x) plasmids and grown in LB medium with ampicillin (100 μg mL^{-1}) and chloramphenicol (34 μg mL^{-1}) overnight. YFP(CrtK) expression was carried out in M9 medium (250 mL) as described above. Cells were harvested by centrifugation at 4300 *g*, and cell pellets were resuspended in imidazole-Tris buffer (10 mL; 40 mM Tris·HCl, 150 mM NaCl, 20 mM imidazole, pH 7.5) and purified by Ni-NTA chromatography (Qiagen) as per the manufacturer's instructions. After elution from the Ni-NTA column, the protein was concentrated and the buffer exchanged with potassium phosphate buffer (50 mM potassium phosphate, 150 mM NaCl, pH 7.5). The purified protein was analyzed by LC-(ESI)-MS using a ThermoFisher Accela HPLC coupled to a ThermoFisher ESI-IT Velos Mass Spectrometer.

Acknowledgements

This work was supported by the U.S. National Institutes of Health (R21 CA187502). MS instrumentation was supported by the U.S. National Science Foundation (CHE-0840410 and CHE-0946653).

Conflict of Interest

The authors declare no conflict of interest.

Keywords: aminoacyl–tRNA synthetases · crotonyl lysine · directed evolution · pyrrolysyl–tRNA synthetases · stop codon suppression

[1] a) C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413–444; b) D. A. Dougherty, *Curr. Opin. Chem. Biol.* **2000**, *4*, 645–652; c) A. J. Link, M. L. Mock, D. A. Tirrell, *Curr. Opin. Biotechnol.* **2003**, *14*, 603–609; d) N. Budisa, *Angew. Chem. Int. Ed.* **2004**, *43*, 6426–6463; *Angew. Chem.* **2004**, *116*, 6586–6624.

[2] a) A. Deiters, T. A. Cropp, M. Mukherji, J. W. Chin, J. C. Anderson, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 11782–11783; b) A. Tuley, Y. J. Lee, B. Wu, Z. U. Wang, W. R. Liu, *Chem. Commun.* **2014**, *50*, 7424–7426; c) A. Chatterjee, J. T. Guo, H. S. Lee, P. G. Schultz, *J. Am. Chem. Soc.* **2013**, *135*, 12540–12543; d) Z. P. Yu, Q. Lin, *J. Am. Chem. Soc.* **2014**, *136*, 4153–4156; e) K. H. Wang, A. Sachdeva, D. J. Cox, N. W. Wilf, K. Lang, S. Wallace, R. A. Mehl, J. W. Chin, *Nat. Chem.* **2014**, *6*, 393–403; f) G. Knorr, E. Kozma, A. Herner, E. A. Lemke, P. Kele, *Chem. Eur. J.* **2016**, *22*, 8972–8979.

[3] a) J. C. Jackson, S. P. Duffy, K. R. Hess, R. A. Mehl, *J. Am. Chem. Soc.* **2006**, *128*, 11124–11127; b) K. R. Mehta, C. Y. Yang, J. K. Montclare, *Mol. BioSyst.* **2011**, *7*, 3050–3055; c) J. Y. Kang, D. Kawaguchi, I. Coin, Z. Xiang, D. D. M. O'Leary, P. A. Slesinger, L. Wang, *Neuron* **2013**, *80*, 358–370; d) J. N. Kolev, J. M. Zaengle, R. Ravikumar, R. Fasan, *ChemBioChem* **2014**, *15*, 1001–1010; e) H. Xiao, F. B. Peters, P. Y. Yang, S. Reed, J. R. Chittuluru, P. G. Schultz, *ACS Chem. Biol.* **2014**, *9*, 1092–1096.

[4] a) H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232–234; b) X. Li, T. Fekner, J. J. Ottesen, M. K. Chan, *Angew. Chem. Int. Ed.* **2009**, *48*, 9184–9187; *Angew. Chem.* **2009**, *121*, 9348–9351; c) Y. S. Wang, B. Wu, Z. Y. Wang, Y. Huang, W. Wan, W. K. Russell, P. J. Pai, Y. N. Moe, D. H. Russell, W. R. Liu, *Mol. BioSyst.* **2010**, *6*, 1575–1578; d) Y. M. Li, M. Y. Yang, Y. C. Huang, X. D. Song, L. Liu, P. R. Chen, *Chem. Sci.* **2012**, *3*, 2766–2770; e) C. H. Kim, M. Kang, H. J. Kim, A. Chatterjee, P. G. Schultz, *Angew. Chem. Int. Ed.* **2012**, *51*, 7246–7249; *Angew. Chem.* **2012**, *124*, 7358–7361; f) M. J. Gattner, M. Vrabel, T. Carell, *Chem. Commun.* **2013**, *49*, 379–381.

[5] a) M. R. Seyedsayamdost, J. Xie, C. T. Y. Chan, P. G. Schultz, J. Stubbe, *J. Am. Chem. Soc.* **2007**, *129*, 15060–15071; b) J. C. Jackson, J. T. Hammill, R. A. Mehl, *J. Am. Chem. Soc.* **2007**, *129*, 1160–1166; c) E. M. Brustad, E. A. Lemke, P. G. Schultz, A. A. Deniz, *J. Am. Chem. Soc.* **2008**, *130*, 17664–17665; d) J. M. Smith, F. Vitali, S. A. Archer, R. Fasan, *Angew. Chem. Int. Ed.* **2011**, *50*, 5075–5080; *Angew. Chem.* **2011**, *123*, 5181–5186; e) N. Bionda, A. L. Cryan, R. Fasan, *ACS Chem. Biol.* **2014**, *9*, 2008–2013; f) J. L. Furman, M. C. Kang, S. Choi, Y. Cao, E. D. Wold, S. B. Sun, V. V. Smider, P. G. Schultz, C. H. Kim, *J. Am. Chem. Soc.* **2014**, *136*, 8411–8417.

[6] G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **2002**, *296*, 1459–1462.

[7] a) O. Namy, Y. Zhou, S. Gundllapalli, C. R. Polycarpo, A. Denise, J.-P. Rousset, D. Söll, A. Ambrogelly, *FEBS Lett.* **2007**, *581*, 5282–5288; b) W. Wan, J. M. Tharp, W. R. Liu, *Biochim. Biophys. Acta Proteins Proteomics* **2014**, *1844*, 1059–1070.

[8] a) C. R. Polycarpo, S. Herring, A. Bérubé, J. L. Wood, D. Söll, A. Ambrogelly, *FEBS Lett.* **2006**, *580*, 6695–6700; b) T. Fekner, X. Li, M. M. Lee, M. K. Chan, *Angew. Chem. Int. Ed.* **2009**, *48*, 1633–1635; *Angew. Chem.* **2009**, *121*, 1661–1663.

[9] a) T. S. Young, I. Ahmad, J. A. Yin, P. G. Schultz, *J. Mol. Biol.* **2010**, *395*, 361–374; b) A. Chatterjee, S. B. Sun, J. L. Furman, H. Xiao, P. G. Schultz, *Biochemistry* **2013**, *52*, 1828–1837.

[10] a) J. T. Guo, C. E. Melancon, H. S. Lee, D. Groff, P. G. Schultz, *Angew. Chem. Int. Ed.* **2009**, *48*, 9148–9151; *Angew. Chem.* **2009**, *121*, 9312–9315; b) N. X. Wang, T. Ju, W. Niu, J. T. Guo, *ACS Chem. Biol.* **2015**, *4*, 207–212.

[11] a) T. Mukai, A. Hayashi, F. Iraha, A. Sato, K. Ohtake, S. Yokoyama, K. Sakamoto, *Nucleic Acids Res.* **2010**, *38*, 8188–8195; b) D. B. F. Johnson, J. F. Xu, Z. X. Shen, J. K. Takimoto, M. D. Schultz, R. J. Schmitz, Z. Xiang, J. R. Ecker, S. P. Briggs, L. Wang, *Nat. Chem. Biol.* **2011**, *7*, 779–786; c) I. L. Wu, M. A. Patterson, H. E. C. Desai, R. A. Mehl, G. Giorgi, V. P. Conticello, *ChemBioChem* **2013**, *14*, 968–978; d) M. Amiram, A. D. Haimovich, C. G. Fan, Y. S. Wang, H. R. Aerni, I. Ntai, D. W. Moonan, N. J. Ma, A. J. Rovner, S. H. Hong, N. L. Kelleher, A. L. Goodman, et al., *Nat. Biotechnol.* **2015**, *33*, 1272–1279.

[12] M. Pott, M. J. Schmidt, D. Summerer, *ACS Chem. Biol.* **2014**, *9*, 2815–2822.

[13] M. J. Lajoie, A. J. Rovner, D. B. Goodman, H.-R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, et al., *Science* **2013**, *342*, 357–360.

[14] A. Uyeda, T. Watanabe, Y. Kato, H. Watanabe, T. Yomo, T. Hohsaka, T. Matsuura, *ChemBioChem* **2015**, *16*, 1797–1802.

[15] I. K. Nordgren, A. Tavassoli, *Mol. Biosyst.* **2014**, *10*, 485–490.

[16] R. B. Cooley, J. L. Feldman, C. M. Driggers, T. A. Bundy, A. L. Stokes, P. A. Karplus, R. A. Mehl, *Biochemistry* **2014**, *53*, 1916–1924.

[17] S. J. Miyake-Stoner, C. A. Refakis, J. T. Hammill, H. Lusic, J. L. Hazen, A. Deiters, R. A. Mehl, *Biochemistry* **2010**, *49*, 1667–1677.

[18] R. B. Cooley, P. A. Karplus, R. A. Mehl, *ChemBioChem* **2014**, *15*, 1810–1819.

[19] J. R. Frost, N. T. Jacob, L. J. Papa, A. E. Owens, R. Fasan, *ACS Chem. Biol.* **2015**, *10*, 1805–1816.

[20] a) Y. Ryu, P. G. Schultz, *Nat. Methods* **2006**, *3*, 263 – 265; b) H. Xu, Y. Wang, J. Lu, B. Zhang, Z. Zhang, L. Si, L. Wu, T. Yao, C. Zhang, S. Xiao, L. Zhang, Q. Xia, et al., *ChemBioChem* **2016**, *17*, 1250 – 1256.
[21] M. Tan, H. Luo, S. Lee, F. Jin, Jeong S. Yang, E. Montellier, T. Buchou, Z. Cheng, S. Rousseaux, N. Rajagopal, Z. Lu, Z. Ye, et al., *Cell* **2011**, *146*, 1016 – 1028.
[22] B. J. Wilkins, L. E. Hahn, S. Heitmüller, H. Frauendorf, O. Valerius, G. H. Braus, H. Neumann, *ACS Chem. Biol.* **2015**, *10*, 939 – 944.
[23] S. Herring, A. Ambrogelly, S. Gundllapalli, P. O'Donoghue, C. R. Polycarpo, D. Söll, *FEBS Lett.* **2007**, *581*, 3197 – 3203.
[24] R. Jiang, J. A. Krzycki, *J. Biol. Chem.* **2012**, *287*, 32738 – 32746.
[25] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *Chem. Biol.* **2008**, *15*, 1187 – 1197.
[26] Y. Chen, R. Sprung, Y. Tang, H. Ball, B. Sangras, S. C. Kim, J. R. Falck, J. M. Peng, W. Gu, Y. M. Zhao, *Mol. Cell. Proteomics* **2007**, *6*, 812 – 819.
[27] D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 8720 – 8721.
[28] R. Fasan, M. M. Chen, N. C. Crook, F. H. Arnold, *Angew. Chem. Int. Ed.* **2007**, *46*, 8414 – 8418; *Angew. Chem.* **2007**, *119*, 8566 – 8570.
[29] D. S. Wilson, A. D. Keefe, in *Curr. Protoc. Mol. Biol.*, John Wiley & Sons, Inc. **2001**.

Manuscript received: January 25, 2017
Accepted manuscript online: April 6, 2017
Version of record online: May 16, 2017