Title

Expanding the Library and Substrate Diversity of the Pyrrolysyl-tRNA Synthetase to Incorporate Unnatural Amino Acids Containing Conjugated Rings

Journal

ChemBioChem, 14(16)

ISSN

1439-4227

Authors

Lacey, Vanessa K
Louie, Gordon V
Noel, Joseph P
et al.

Publication Date

2013-11-04

DOI

10.1002/cbic.201300400

Peer reviewed

Published in final edited form as:
Chembiochem. 2013 November 4; 14(16): 2100–2105. doi:10.1002/cbic.201300400.

Expanding the Library and Substrate Diversity of the Pyrrolysyl-tRNA Synthetase to Incorporate Unnatural Amino Acids Containing Conjugated Rings

Vanessa K. Lacey[a], **Gordon V. Louie**[a], **Joseph P. Noel**[a,b], and **Lei Wang**[a]

Lei Wang: lwang@salk.edu

[a]Jack H. Skirball Center for Chemical Biology & Proteomics The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037 (USA)

[b]Howard Hughes Medical Institute

Abstract

Unnatural amino acids (Uaas) containing conjugated ring systems are of particular interest for their optical properties. Until now, such structurally bulky and planar Uaas could not be incorporated into proteins using the pyrrolsyl tRNA/synthetase shuttling system. By building a highly diverse synthetase library using the "small intelligent" approach, we evolved novel synthetases specific for two such Uaas and incorporated them into proteins in *E. coli* and mammalian cells.

Keywords

unnatural amino acid; pyrrolysine; genetic code expansion; small intelligent library; directed evolution

Genetically encoded unnatural amino-acids (Uaas), incorporated with orthogonal tRNA/aminoacyl-tRNA synthetase (aaRS) pairs, have greatly facilitated the study and manipulation of proteins and biological processes.[1–3] Uaas are available with a diversity far exceeding that of nature's canonical amino acids. Through site-directed incorporation of Uaas with specific physical or chemical characteristics, proteins can be engineered with novel properties.[3] A large number of distinct Uaas have been incorporated with tRNA/aaRS pairs evolved from the original tyrosyl amber suppressor tRNA ($\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$) and its cognate tyrosyl-tRNA synthetase (TyrRS).[4] In recent years, the pyrrolysyl amber suppressor tRNA and cognate synthetase ($\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$/PylRS) pair derived from the archaebacteria *Methanosarcina mazei* and *Methanosarcina barkeri* has provided a convenient alternative. This pair is orthogonal to endogenous tRNA/aaRS pairs in both *E. coli* and mammalian cells,[5–7] and so mutant synthetases specific for Uaas can be evolved in a bacterial host and subsequently imported into to eukaryotic cells for Uaa incorporation. A number of Uaas with structures similar to the native substrate pyrrolysine (Pyl) have been incorporated using the wild-type or engineered PylRS.[3, 6, 8] Interestingly, we previously demonstrated that through replacement of a residue serving as the "gate keeper" for Pyl binding, the PylRS can be engineered to incorporate Uaas with structures divergent from Pyl.[9] In these Uaas, the long aliphatic side chain and the core N^{ε}-carbonyl group of Pyl is

Correspondence to: Lei Wang, lwang@salk.edu.

Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author

replaced by an aromatic ring, thereby allowing the incorporation of tyrosine or phenylalanine derivatives.[9, 10]

To further expand the diversity of Uaas that can be incorporated by the $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$/PylRS pair, we sought to evolve the PylRS to incorporate Uaas containing multiple conjugated aromatic rings. Conjugated ring systems are of particular interest for protein engineering because they confer photo-responsive properties, such as amenability to photo-crosslinking, fluorescence, and photo-isomerization, all of which are invaluable for optical protein manipulation. To effect such a dramatic change in substrate specificity of the PylRS, we employed a new approach for synthetase library construction. In comparison to conventional methods, our strategy allowed a greater number of residues to be simultaneously randomized and consequently a more efficient search of a larger sequence space. New PylRS mutants specific for Uaas containing conjugated rings, L-3-(2-naphthyl)alanine (Nap) and *p*-benzoyl-L-phenylalanine (Bpa) (Figure 1A), were thereby identified, and these mutants were shown to enable Nap and Bpa incorporation in both *E. coli* and mammalian cells.

Here, *M. mazei* PylRS (MmPylRS) served as the basis for evolving synthetases that can accommodate the substrates Nap and Bpa, which bear a short, methylene linker between the alpha carbon (Cα) and the aromatic functionality (Figure 1A). Such Uaas have thus far been recalcitrant for $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$/PylRS pairs, as not surprisingly, most engineered PylRS have specificity for a Uaa that retains substantial structural similarity to Pyl and thus contains a side chain with a long flexible alkyl group.[3] For applications in which an incorporated Uaa introduces a key functional group into a target protein, a shorter side-chain linker may be desirable for minimizing both the separation between the functionality and the protein backbone and also potential impediments to protein folding. Prior work from this laboratory yielded a PylRS mutant, MmOmeRS, specific for *O*-methyl-L-tyrosine (Ome),[9] which, like Nap and Bpa, contains an aromatic functionality with a short linker to Cα. MmOmeRS presently represents the only PylRS evolved for a “short-linker” Uaa for which structural information is available. Thus, we sourced the published crystal structure of the MmOmeRS/Ome complex (PDB 3QTC) to identify residues to be varied in the construction of an expanded PylRs library for the evolution of Nap- and Bpa-specific variants. The Nap and Bpa Uaas were roughly modelled in the amino-acid binding pocket of MmOmeRS by superposition of the backbone atoms and the first aromatic ring onto the corresponding portions of Ome (Figure 1B). As expected, the much bulkier Nap and Bpa Uaas clashed sterically with a number of amino-acid residues adjoining the MmOmeRS binding pocket. Based on these preliminary analyses, positions 309, 346, 348, 401 and 417 were chosen for full randomization to any of the 20 canonical amino acids. Despite its presumed role as the Pyl “gate keeper”,[11] residue N346 revealed to be mutable; its central role in aminoacylation can be rescued through compensatory hydrogen bonding provided by an A302T substitution.[9] We therefore permitted Ala and Thr at site 302. In addition, we randomized G419 to permit Gly, Ala, and Ser. This position lies at the bottom of the active site’s β-sheet floor with its side chain pointing toward the aromatic ring of the Uaa, so we presumed that only small amino acids would be tolerated.

The simultaneous mutation of seven residues requires a highly efficient method for library construction to achieve complete coverage. Iterative saturation mutagenesis,[12] an alternative strategy for creating multisite mutants, has yet to be demonstrated efficacious for the large-scale evolution of an aaRS. A practical limitation on the library size that can be managed in an *E. coli* host is imposed by the transformation efficiency of *E. coli*, which typically does not exceed ~10^{10} colony forming units (cfu) per μg of transformed DNA. Therefore, a paramount design consideration for maintaining maximal diversity in a mutant protein library expressed in *E. coli* is minimizing the size of the encoding gene library

without sacrificing any amino acid representation. In particular, with conventional methods for mutant synthetase library construction, each position to be fully randomized is specified by the degenerate codon NNK,[1] and thus 32 distinct codons are used to encode the 20 possible amino-acid identities. The redundant codons introduced by NNK coding not only include undesired stop and rare codons, and induce an inherent bias towards amino acids encoded by multiple codons, but most importantly unnecessarily increase the gene library size. Thus, the full randomization of 7 positions with NNK codons would require a theoretical gene library size of 3.4×10^{10}. With only single-fold coverage, a library of this size would be difficult to transform into *E. coli*, and achieving 10-fold coverage would be extremely challenging. To circumvent these limitations, we optimized the gene library diversity and theoretical size by choosing codon sequences for randomization based on the principle of "small intelligent" mutagenesis.[13] In this method, each of the twenty amino acids is represented uniquely by a single codon; the set of twenty distinct codons (which notably excludes stop and rare codons) is encoded by ATG, TGG, and the degenerate sequences NDT and VMA. In practice, for each target site, four mutagenic primers (encoding the codon sequences NDT, VMA, ATG, or TGG) are combined with the appropriate stoichiometry to ensure even distribution of each of the 20 amino acids (Figure 2A). In a proof-of-principle demonstration, this method was used to randomize one or two proximal sites.[13] Here, we expanded this strategy to randomize 7 sites dispersed throughout the PylRS primary structure. The mutant library was designed, as outlined above, to accommodate all 20 amino acids at positions 309, 346, 348, 401, and 417; Ala or Thr at position 302; and Ala, Ser, or Gly at position 419. These combinations produce a theoretical gene library size of only 1.9×10^7, which is readily manageable in *E. coli* with standard cloning.

We constructed the library using degenerate primers premixed in specific ratios. Four primers were used in the PCR for a single site (Figure 2B), and 16 primers were used for two proximal sites (e.g., sites 346 and 348) to ensure that every combination of amino acids was equally included (Figure 2C). The full-length synthetase gene was pieced together through overlapping PCRs and ligated into an expression vector. Transformation into *E. coli* resulted in a library of 2×10^9 cfu, indicating full saturation with a remarkable 100-fold coverage.

We next performed a positive selection and a negative screen to evolve Bpa- and Nap-specific PylRS mutants. Hits with high IC_{50}s (over 200 μg/mL) in the presence of Nap or Bpa and low IC_{50}s without any Uaa (20 μg/mL) were identified (Figure S1, Table S1), suggesting their specificity for the target Uaa. The two hits for Nap converged on the following amino-acid substitutions: 302T, 309L, 346G, 348T, 401I, 417Y, and 419G. The three hits for Bpa all contained the substitutions 302T, 309L, 346T, 348T, and 419G. Additionally, the two strongest hits (IC_{50} > 205 μg/mL) contained 401V and 417C, whereas the slightly weaker hit contained 401T and 417V (Table S2). Retrospectively, all hits identified here potentially occurred in a PylRS mutant library we previously generated by randomizing A302, L309, N346, C348, V401, and W417 using NNK primers.[9] However, they were not identified from the NNK library, possibly due to inherent bias of the NNK approach and only single-fold coverage of the gene library in *E. coli* (theoretical size of the gene library = 1.1×10^9; library made = 1×10^9 cfu). These results demonstrate that the high coverage (100×) and unbiased amino-acid encoding of the small-intelligent mutant gene library enabled more effective synthetase evolution.

To measure the *in vivo* translation efficiency and fidelity of the evolved synthetases, we expressed in *E. coli* a sperm whale myoglobin gene, encoding an C-terminal His-tag and a TAG codon at position 4, together with the genes for $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$ and the evolved MmBpaRS or MmNapRS synthetases. Protein expression in *E. coli* was assessed by purification of

affinity-tagged protein and analysis by SDS-PAGE (Figure 3A). With both the MmBpaRS or MmNapRS systems, a strong 18.5-kD band indicated that full-length myoglobin was produced by *E. coli* grown in the presence of the corresponding Uaa, and only a weak level of suppression was observed when the Uaa was excluded from LB growth medium. Using densitometry, we estimated that the MmBpaRS and MmNapRS systems produced full-length myoglobin at 12- and 16-fold, respectively, higher levels in the presence of the corresponding Uaa.

The myoglobin proteins produced in the presence of each Uaa were analysed by electrospray ionization Fourier transform ion trap mass spectrometry (ESI-FTMS) to verify incorporation of the correct Uaa and to assess specificity of the evolved synthetase. For myoglobin obtained with the $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$/MmBpaRS strain, a peak with a monoisotopic mass of 18,582.74 Da was observed (Figure 3B). This mass corresponds to intact myoglobin incorporating a single Bpa residue at position 4 (expected $[M + H]^+$ = 18,582.79 Da). A second peak measured corresponds to Bpa-incorporated myoglobin lacking the initiating Met (expected $[M - Met + H]^+$ = 18,451.75 Da, measured 18,451.75 Da). Similarly, for myoglobin obtained with the $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$/MmNapRS strain, monoisotopic masses confirming Nap incorporation were also observed (expected $[M + H]^+$ = 18,528.78 Da, measured 18,528.73 Da; expected $[M - Met + H]^+$ = 18,397.74 Da, measured 18,397.72 Da) (Figure 3C). Notably, no peaks corresponded to myoglobin resulting from the incorporation of any other endogenous amino acid in lieu of the Uaa in both samples. Assuming the amino acid incorporated at residue 4 does not significantly affect myoglobin ionization, the signal-to-noise ratio of >1,000 observed in the intact myoglobin mass spectra suggests a fidelity of 99.9% for the incorporation of Bpa and Nap. These results showed that although the evolved tRNA/synthetase pairs have low background suppression in the absence of the Uaa, they specifically incorporate the Uaa when the cognate Uaa is present. Similar observations have also been reported for other synthetases evolved from PylRS.[9]

Perhaps the most attractive feature of the $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$/PylRS pair is its dual orthogonality in *E. coli* and mammalian host systems. We therefore tested whether the MmBpaRS and MmNapRS evolved in *E. coli* could be used in mammalian cells for Uaa incorporation into proteins. Plasmids encoding $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$ with MmBpaRS or MmNapRS were transiently transfected into a HeLa reporter cell line that carries a stable, chromosomally integrated GFP-gene containing a TAG codon at the permissive Tyr182 site,[14] and the cells were supplemented with the appropriate Uaa (Bpa or Nap). As a positive control, $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$ and the wild-type MmPylRS were also transfected into the reporter cell line, and the cells were supplemented with N^{ε}-*tert*-butyloxycarbonyl-L-lysine (Boc-Lys), which is efficiently incorporated by the MmPylRS.[5] For all $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$/MmaaRS pairs, bright green fluorescence of GFP was observed by fluorescence microscopy only in cells fed the appropriate Uaa (Figure 4A). Flow cytometry was used to quantify the total fluorescence intensity of each sample with the same number of sorted cells (Figure 4B). In comparison to cells wherein Boc-Lys was incorporated by the wild-type PylRS, all cell lines transfected with $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$/MmaaRS pairs and grown in the absence of the appropriate Uaa exhibited less than 0.3% total fluorescence intensity, indicating a very low level of background suppression by the $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$/MmBpaRS and $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$/MmNapRS in HeLa cells. When the respective Uaa was added, the incorporation efficiency was 24.4% for Bpa and 5.6% for Nap.

To begin to understand how the evolved MmBpaRS and MmNapRS recognize their target Uaa, we structurally modelled each of synthetases and the binding of the respective Uaa,

using the MmOmeRS/Ome complex structure as the template (Figure 5).[9] The interactions formed to the backbone groups of Nap and Bpa are similar to those observed in the MmOmeRS/Ome complex, and in particular, a Thr at position 302 of the synthetase hydrogen bonds to the α-carboxylate group of the Uaa and compensates for the loss of Asn at 346.[9] Asn346 is thought to play a central role in aminoacylation of Pyl and analogues.[11] For the Ome-, Nap-, and Bpa-specific synthetases, N346 can be replaced by Val, Thr, or Gly in conjunction with the T302 substitution. Thus, whereas N346 serves as the gate keeper for Pyl analogues, T302 represents a key determinant for the recognition of Phe analogues. For all the three aromatic Uaas here, the side chain occupies a hydrophobic pocket and forms van der Waals contact with the side chains of residues 348, 401, and 417 and the main chain of residues 419 and 420. Notably, the observed amino-acid substitutions in evolved synthetases are entirely consistent with the accommodation of the shape and length of the side chain of the target Uaa.

Until now, the $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$/PylRS pair has not been amenable to evolution for incorporation of Uaas containing multiple conjugated rings, presumably due to incompatibility of the PylRS substrate binding site with the bulky, planar side chains carried by these Uaas. We have built a robust, highly saturated library with randomization at 7 dispersed positions using sets of "small intelligent" primers. This library yielded new synthetases for Nap, containing a condensed aromatic side chain, and Bpa, containing two cross conjugated rings. Both synthetases were evolved in *E. coli* and directly imported into mammalian cells for successful Uaa incorporation. We expect that this library will be useful in evolving synthetases specific for other Uaas containing conjugated ring systems. In addition, more residues in the PylRS or other aaRSs could be randomized simultaneously with the approach described here due to the significantly reduced size of the gene library; the resultant expansion of synthetase library diversity should enable a broader range of Uaas to be genetically incorporated in both *E. coli* and mammalian cells.

Experimental Section

Library construction

The library VL5 was constructed using a combination of overlap PCR reactions with the primers listed in Supporting Information. These "small intelligent" primers were mixed in ratios calculated to ensure equal representation of each codon, as shown in Figure 2. The final PCR product was digested with restriction enzymes *BstX* I and *Eco0109* I, and ligated into the pre-cut vector pBK5.[9] The ligated DNA was transformed into electro competent MDS42 cells with a competency of 10^{10} cfu/μg for control DNA. The transformed cells were plated onto 8 × 15-cm dishes for overnight growth, after which the DNA was purified using a Maxi Prep kit (Qiagen). A test plate containing a diluted aliquot of the transformation revealed the library size to be 2×10^9 cfu. Twenty colonies were sequenced using primer MmSeq to verify the integrity of the library.

Selection and screen

Fifty nanograms of the purified library DNA was transformed into MDS42 cells harbouring the selection plasmid pREP/$\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$ [6, 15] and plated on to glycerol minimal medium plates containing Tet (10 μg/mL), Kan (50 μg/mL), Cm (40 μg/mL), and either Bpa or Nap (1 mM, purchased from Bachem). After 48 hours, green fluorescent colonies were picked and separately cultured into 1 mL of LB containing Tet and Kan but without any Uaa or Cm. Those cultures that remained non-fluorescent from lack of GFP after overnight growth were selected to assess the IC_{50} of Cm resistance in liquid format.

Determination of IC_{50} in chloramphenicol

A liquid well format was used for the IC_{50} screen rather than the original method using agar plates[16] due to the increased number of clones that could be screened and the reduced amount of Uaa required. Multiwell plates were prepared by dispensing 200 μL of LB medium containing Tet, Kan, and various concentrations of Cm between 0 μg/mL and 300 μg/mL. Starter cultures were normalized by the measured OD_{600} to contain the same number of cells, and were dispensed into the wells that either did or did not contain 1 mM Uaa. Cells were grown overnight, and the OD_{600} was measured on a plate reader to determine the relative cell survival. Figure S1 displays the OD_{600} as a function of Cm concentration from 5 clones. Table S1 reports the IC_{50} values determined using a variable slope sigmoidal curve fit of the data in Figure S1.

Hit identification

Clones that showed a high IC_{50} with Uaa and low IC_{50} without Uaa were sequenced using primer MmSeq to determine the identity of the mutations in the synthetase. The translated nucleotides at each mutated position are listed in Table S2.

Myoglobin expression

The pBK5 plasmids encoding MmBpa B1 and MmNap N1 synthetases were separately transformed into BL21 (DE3) cells containing the expression plasmid pTAKMyo4TAG.[9] Cultures (50 mL) were grown from single colonies with and without the appropriate Uaa (1 mM) in LB medium. Protein expression and purification were conducted as previously described.[9]

Mass spectrometry

The protein sample (approx. 10 μg) was desalted on a ZIP tip (Millipore) and analysed by high resolution Fourier-transform mass spectrometry on a Thermo LTQ-Orbitrap XL mass spectrometer (ThermoFisher, San Jose, CA). The sample was loaded onto a capillary column (75 μm diameter) with an integrated spray tip, which was filled with reversed phase material (Zorbax SB C-18, particle size 5 μm, bed length 5 cm). Protein was eluted with a gradient of 0.1% formic acid and an increasing proportion of acetonitrile at a flow rate of 300 nL/min. The eluate was electrosprayed directly into the mass spectrometer. Fourier-transform mass spectra were recorded at a resolution of 60,000 for a scan range of m/z = 400–1800 followed by a select ion scan of the most intense ion at resolution 60,000. Data were charge-deconvoluted using the Thermo Qualbrowser 2.0 Xtract program.

Uaa incorporation in mammalian cells

Genes for MmPylRS, MmNapRS and MmBpaRS were subcloned into plasmid pMPcua-OmeRS [9] between the *Xho*I and *Nhe*I sites using primers MmPylRS XhoI and MmPylRS NheI. The HeLa-GFP(182TAG) stable cells were transfected with the resultant pMPcua-MmPylRS, pMPcua-MmNapRS or pMPcua-MmBpaRS. Cells were grown in DMEM supplemented with 10% FBS; 1 mM of the corresponding Uaa (purchased from Bachem) was added to or withheld from the media. Cells were imaged 48 h after transfection with an Olympus IX8I microscope equipped with a Hamamatsu EM-CCD under the same image capture conditions for all samples ($\lambda ex = 480 \pm 20$ nm, $\lambda em = 535 \pm 40$ nm). Transfected HeLa cells were trypsinized and washed with PBS twice. After suspending in 1 mL PBS and 5 μL of propidium iodide, cells were analysed with a FACScan (Becton & Dickinson, excitation 488 nm, emission 530 ± 30 nm). For each sample the total fluorescence intensity of 10,000 cells was recorded and normalized to cells transfected with pMPcua-MmPylRS with Boc-Lys added.

Molecular docking

For MmNapRS and MmBpaRS, protein models bearing the selected amino-acid substitutions were generated from the 3QTC template. Initially, the replacement side-chains were introduced with a suitable preferred conformation in the program Coot [17] and the target Uaa ligand was positioned roughly as guided by the Ome in 3QTC, and these atomic groups were then manually adjusted to minimize steric conflicts. The initial synthetase-Uaa complexes were subsequently subjected to one round of simple dynamics with pdbtools in the program Phenix.[18]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Wolfgang Fischer and William Low for help with mass spectrometry and Caz O'Connor for help with flow cytometry. L.W. acknowledges support from the California Institute for Regenerative Medicine (RN1-00577-1) and US National Institutes of Health (1DP2OD004744-01 and P30CA014195).

References

1. Wang L, Brock A, Herberich B, Schultz PG. Science. 2001; 292:498–500. [PubMed: 11313494]
2. Wang Q, Parrish AR, Wang L. Chem. Biol. 2009; 16:323–336. [PubMed: 19318213]
3. Liu CC, Schultz PG. Annu. Rev. Biochem. 2010; 79:413–444. [PubMed: 20307192]
4. a) Wang L, Schultz PG. Chem. Biol. 2001; 8:883–890. [PubMed: 11564556] b) Wang L, Schultz PG. Angew. Chem. Int. Ed. Engl. 2005; 44:34–66. [PubMed: 15599909]
5. Mukai T, Kobayashi T, Hino N, Yanagisawa T, Sakamoto K, Yokoyama S. Biochem. Biophys. Res. Commun. 2008; 371:818–822. [PubMed: 18471995]
6. Chen PR, Groff D, Guo J, Ou W, Cellitti S, Geierstanger BH, Schultz PG. Angew. Chem. Int. Ed. Engl. 2009; 48:4052–4055. [PubMed: 19378306]
7. a) Blight SK, Larue RC, Mahapatra A, Longstaff DG, Chang E, Zhao G, Kang PT, Green-Church KB, Chan MK, Krzycki JA. Nature. 2004; 431:333–335. [PubMed: 15329732] b) Ambrogelly A, Gundllapalli S, Herring S, Polycarpo C, Frauer C, Soll D. Proc. Natl. Acad. Sci. U. S. A. 2007; 104:3141–3146. [PubMed: 17360621]
8. a) Yanagisawa T, Ishii R, Fukunaga R, Kobayashi T, Sakamoto K, Yokoyama S. Chem. Biol. 2008; 15:1187–1197. [PubMed: 19022179] b) Li X, Fekner T, Ottesen JJ, Chan MK. Angew. Chem. Int. Ed. Engl. 2009; 48:9184–9187. [PubMed: 19882608] c) Lin S, Zhang Z, Xu H, Li L, Chen S, Li J, Hao Z, Chen PR. J. Am. Chem. Soc. 2011; 133:20581–20587. [PubMed: 22084898] d) Borrmann A, Milles S, Plass T, Dommerholt J, Verkade JM, Wiessler M, Schultz C, van Hest JC, van Delft FL, Lemke EA. ChemBioChem. 2012; 13:2094–2099. [PubMed: 22945333] e) Schmidt MJ, Summerer D. Angew. Chem. Int. Ed. Engl. 2013; 52:4690–4693. [PubMed: 23512703]
9. Takimoto JK, Dellas N, Noel JP, Wang L. ACS Chem. Biol. 2011; 6:733–743. [PubMed: 21545173]
10. a) Wang YS, Fang X, Wallace AL, Wu B, Liu WR. J. Am. Chem. Soc. 2012; 134:2950–2953. [PubMed: 22289053] b) Wang YS, Fang X, Chen HY, Wu B, Wang ZU, Hilty C, Liu WR. ACS Chem. Biol. 2013; 8:405–415. [PubMed: 23138887]
11. a) Kavran JM, Gundllapalli S, O'Donoghue P, Englert M, Soll D, Steitz TA. Proc. Natl. Acad. Sci. U. S. A. 2007; 104:11268–11273. [PubMed: 17592110] b) Yanagisawa T, Ishii R, Fukunaga R, Kobayashi T, Sakamoto K, Yokoyama S. J. Mol. Biol. 2008; 378:634–652. [PubMed: 18387634]
12. Reetz MT, Wang LW, Bocola M. Angew. Chem. Int. Ed. Engl. 2006; 45:1236–1241. [PubMed: 16411254]
13. Tang L, Gao H, Zhu X, Wang X, Zhou M, Jiang R. BioTechniques. 2012; 52:149–158. [PubMed: 22401547]

14. Wang W, Takimoto JK, Louie GV, Baiga TJ, Noel JP, Lee KF, Slesinger PA, Wang L. Nat. Neurosci. 2007; 10:1063–1072. [PubMed: 17603477]
15. Santoro SW, Wang L, Herberich B, King DS, Schultz PG. Nat. Biotechnol. 2002; 20:1044–1048. [PubMed: 12244330]
16. Wang L, Magliery TJ, Liu DR, Schultz PG. J. Am. Chem. Soc. 2000; 122:5010–5011.
17. Emsley P, Cowtan K. Acta Crystallogr. D. Biol. Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]
18. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. Acta Crystallogr. D. Biol. Crystallogr. 2010; 66:213–221. [PubMed: 20124702]

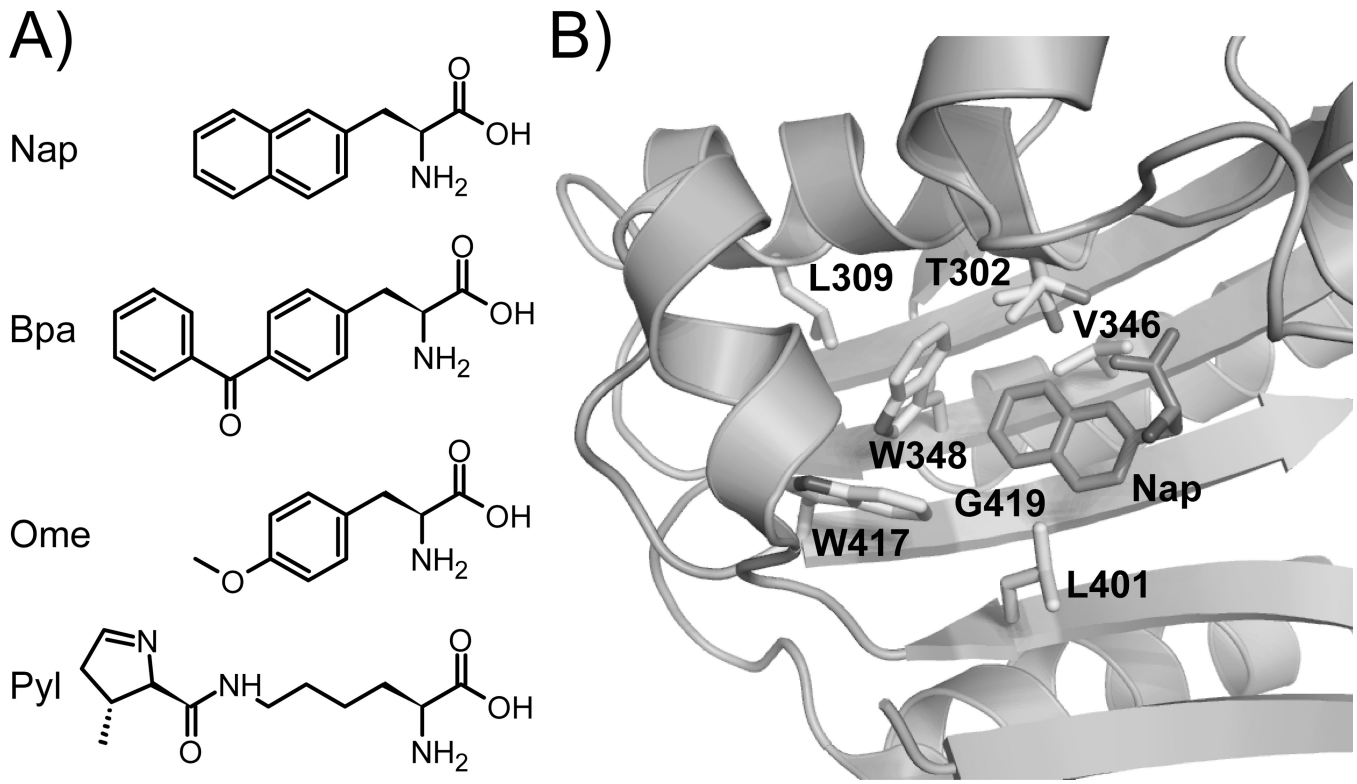

Figure 1.

Design of a library of MmPylRS mutants for incorporating Uaas with multiple conjugated rings. A) Chemical structures of the two Uaas with conjugated ring systems used in this study, Nap and Bpa. Ome and Pyl are shown for comparison. B) Superposition of Nap into the crystal structure of MmOmeRS (PDB 3QTC). The positions selected for mutagenesis are shown in stick and labelled: T302, L309, V346, W348, L401, W417, and G419.

A)

Codon	Amino acid encoded
NDT	N,S,I,H,R,L,Y,C,F,D,G,V
VMA	E,A,Q,P,K,T
ATG	M
TGG	W

Degenerate base	Base coded
N	A,T,G,C
D	A,G,T
V	A,C,G
M	A,C
K	G,T

B)

Codon	Ratio
NDT	12
VMA	6
ATG	1
TGG	1

C)

Codon pair		Ratio
NDT	NDT	144
NDT	VMA	72
NDT	ATG	12
NDT	TGG	12
VMA	NDT	72
VMA	VMA	36
VMA	ATG	6
VMA	TGG	6
ATG	NDT	12
ATG	VMA	6
ATG	ATG	1
ATG	TGG	1
TGG	NDT	12
TGG	VMA	6
TGG	ATG	1
TGG	TGG	1

Figure 2.

Small intelligent approach for constructing a synthetase mutant library. A) Codon set used to equally code each of the 20 amino acids without stop or rare codons. B) Premixed primers in the indicated ratio for randomizing a single site. C) Premixed primers in the indicated ratio for randomizing two adjacent sites.

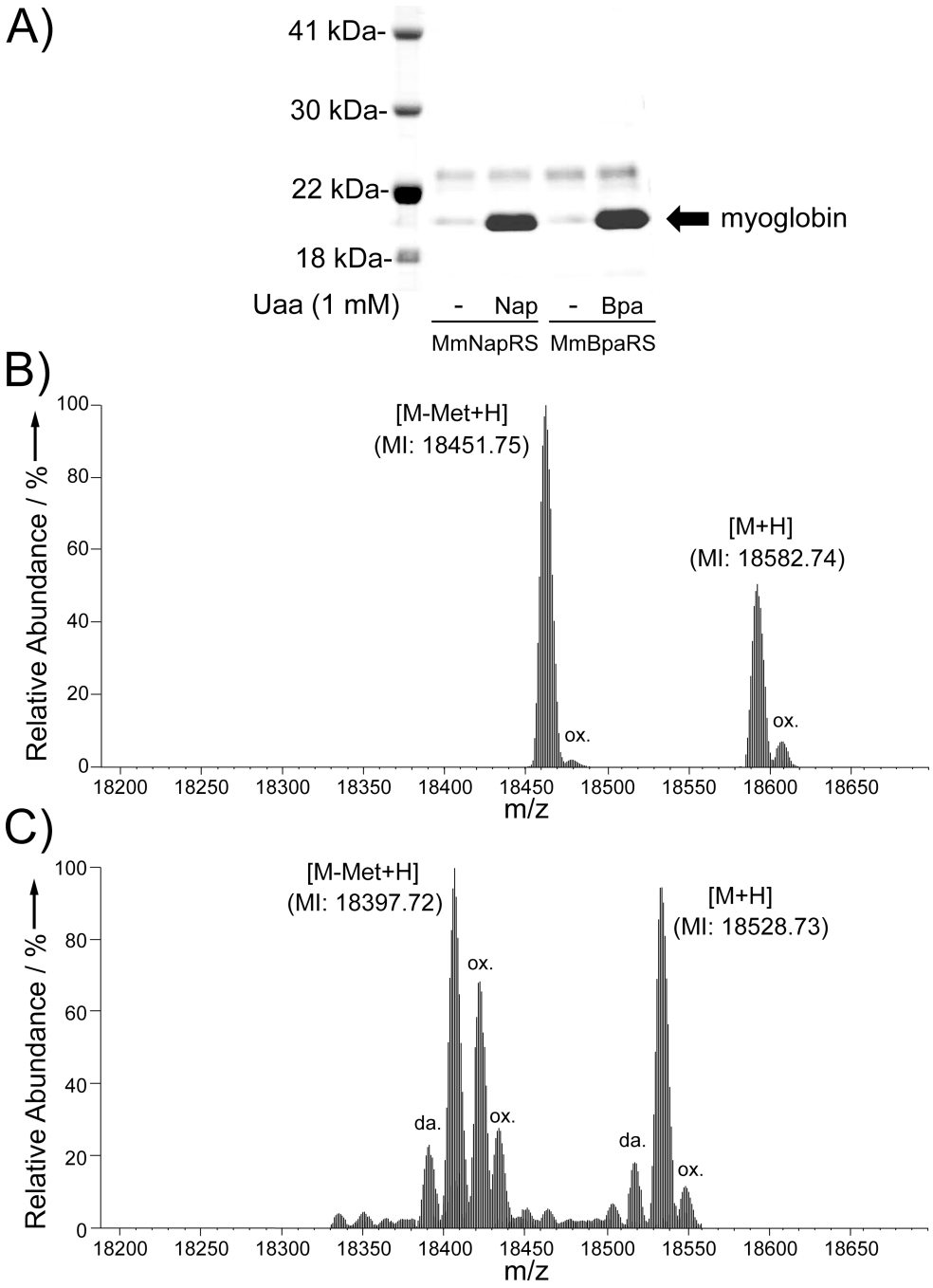

Figure 3.

The evolved MmBpaRS and MmNapRS specifically incorporate their cognate Uaa into myoglobin in *E. coli*. A) Suppression of the amber codon in myoglobin at position 4 by $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$ and MmBpaRS or MmNapRS with and without the cognate Uaa, respectively. Myoglobin proteins were purified using Ni-NTA resin, separated by SDS-PAGE, and stained with Commassie. Samples were normalized by culture density for each lane. B, C) High resolution FT ion trap MS of intact myoglobin proteins expressed in the presence of Bpa (B) or Nap (C). Monoisotopic masses (indicated by MI) are indicated. Proteins

produced using either synthetase were found lacking the initiating Met (M-Met). Oxidized (ox.) or deamidated (da.) products were also detected.

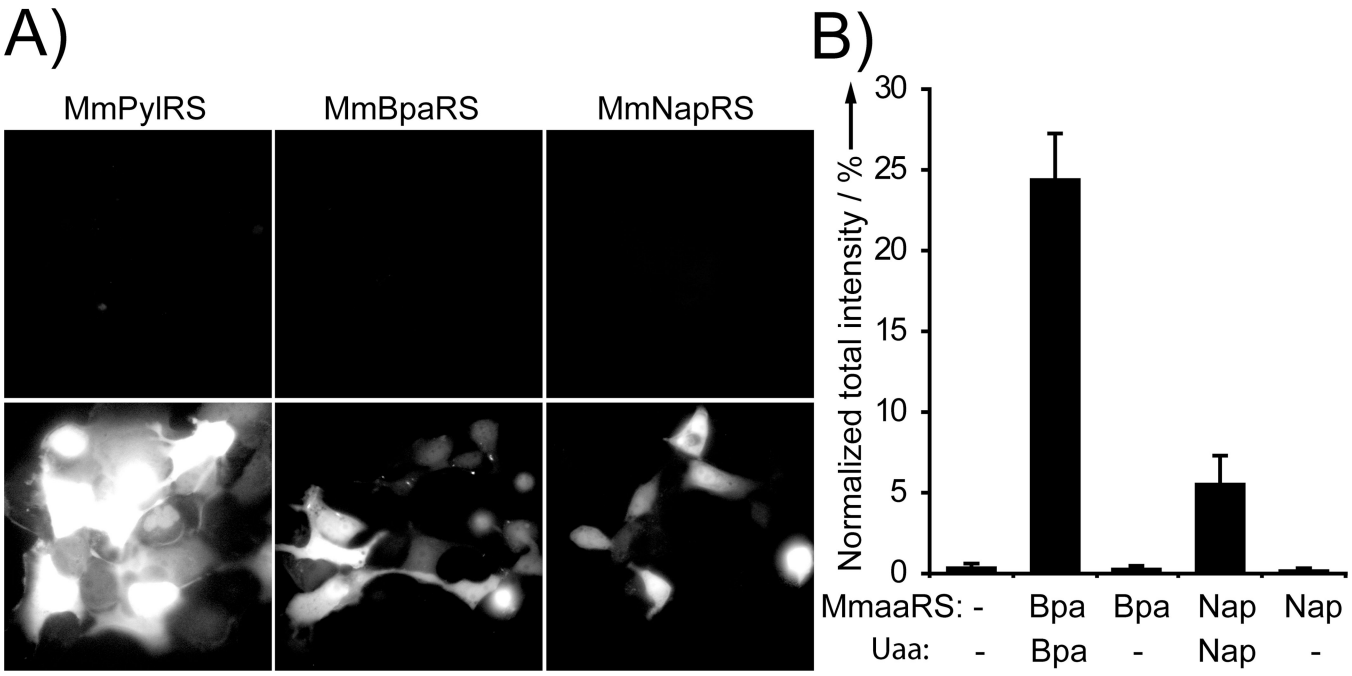

Figure 4.

The MmBpaRS and MmNapRS incorporate their cognate Uaa into GFP in mammalian cells. A) Fluorescent imaging of HeLa cells containing a chromosomal integrated GFP(182TAG) reporter and transfected with $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$ and the indicated synthetase gene. In all cases, bright GFP fluorescence was observed in cells cultured in medium containing 1 mM of the respective Uaa (bottom row). Boc-Lys was used as the substrate for the wild-type MmPylRS. No GFP fluorescence was detected in cells cultured in medium lacking the Uaa (top row). B) Uaa incorporation efficiency measured with flow cytometry. Cellular GFP fluorescence intensities were measured by flow cytometry and normalized to cells transfected with the $\mathrm{tRNA}^{\mathrm{Pyl}}_{\mathrm{CUA}}$/MmPylRS pair and supplemented with Boc-Lys. Error bars represent standard error of mean (s.e.m.). The values (mean ± s.e.m.) are 24.4 ± 2.7% (n = 5) for Bpa and 5.6 ± 1.6% (n = 5) for Nap.

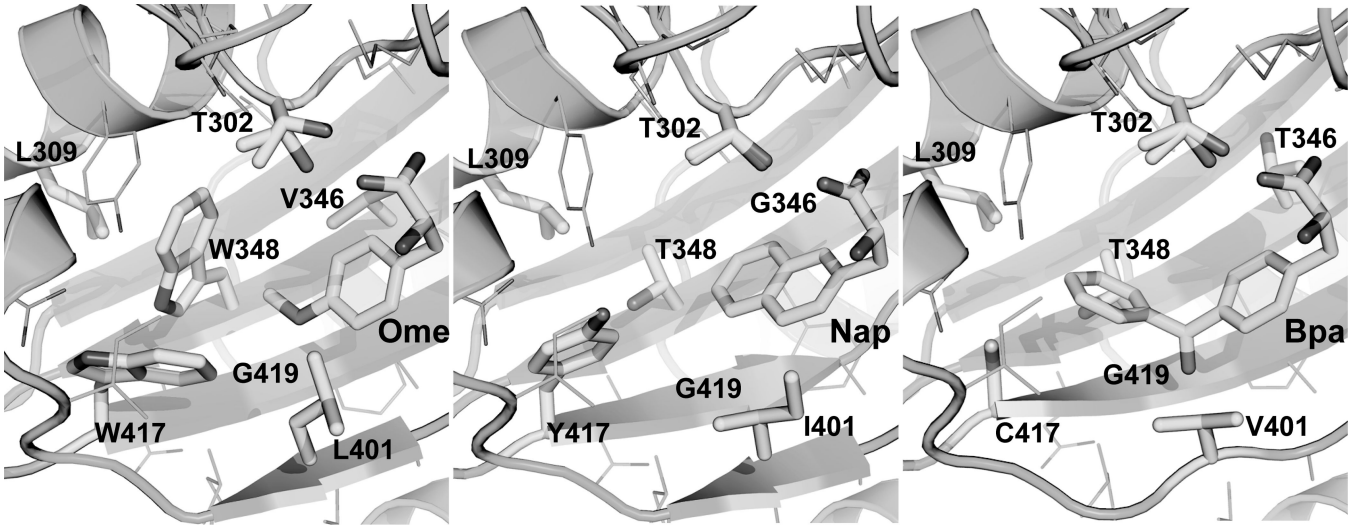

Figure 5.

Comparison of Ome (left, PDB 3QTC), Nap (middle) and Bpa (right) binding to the evolved synthetases. Nap and Bpa were docked into models of 3QTC bearing the appropriate amino-acid substitutions. W348 in MmOmeRS forms the inner wall of the binding pocket for the Ome side chain, whereas the shorter T348 in both MmBpaRS and MmNapRS extends the binding pocket for Nap and Bpa. Likewise, W417 in MmOmeRS contacts the *o*-methyl group of Ome; whereas the smaller Y417 in MmNapRS interacts with the 2nd phenyl ring of Nap, and the smallest C417 in MmBpaRS accommodates the 2nd phenyl ring of the bulkiest Bpa. In the central portion of the binding pocket, V346 and T346 interact with the first phenyl ring of Ome and Bpa, respectively. For MmNapRS, substitution at 346 by the smallest Gly serves to accommodate the conjugated 2nd ring of Nap, which also uniquely contacts the main chain of the neighbouring F347. The keto group of Bpa potentially forms H-bonds with the main-chain amide nitrogen of V401 and G419, but also has dipole-dipole repulsion with the main-chain carbonyl group of these two residues.