

Generation of an Orthogonal Protein–Protein Interface with a Noncanonical Amino Acid

Minseob Koh,[†] Fariborz Nasertorabi,[‡] Gye Won Han,[‡] Raymond C. Stevens,^{*,‡} and Peter G. Schultz^{*,†}

[†]Department of Chemistry and Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 N Torrey Pines Road, La Jolla, California 92037, United States

[‡]Department of Biological Sciences, Bridge Institute, University of Southern California, 3430 S Vermont Avenue, Los Angeles, California 90089, United States

S Supporting Information

ABSTRACT: We have engineered the protein interface of the *Escherichia coli* chorismate mutase (EcCM) homodimer to be dependent on incorporation of a noncanonical amino acid (ncAA) at residue 72. The large hydrophobic amino acid *p*-benzoyl phenylalanine (pBzF) was substituted for Tyr72, which led to a catalytically inactive protein. A library of five residues (Leu25', Arg29', Leu76, Ile80' and Asp83') surrounding pBzF72 was generated and subjected to a growth based selection in a chorismate mutase deficient strain. An EcCM variant (Phe25', pBzF72, Thr76, Gly80' and Tyr83') forms a stable homodimer, has catalytic activity similar to the wild type enzyme, and unfolds with a T_m of 53 °C. The X-ray crystal structure reveals a pi–pi stacking and hydrogen bonding interactions that stabilize the new protein interface. The strategy described here should be useful for generating organisms that are dependent on the presence of a ncAA for growth.

In the past decade, we and others have shown that a large number of noncanonical amino acids (ncAAs) with unique chemical and physical properties can be added to the genetic codes of both prokaryotes and eukaryotes.^{1–9} The essential element is an orthogonal tRNA/aminoacyl-tRNA synthetase (aaRS) pair that incorporates an ncAA in response to unique “blank” codons without cross reacting with the translational machinery of the host.^{1,2} This orthogonal ncAA translational machinery can be used not only to confer an evolutionary advantage,^{3–6} but also to generate ncAA dependent organisms for use as vaccines or for biological containment.^{10–12} In particular, bacteria and viruses that are dependent on the presence of an ncAA for replication have been generated by incorporating ncAAs in multiple essential genes.^{10–14} Although computational approaches have proven very useful in the design of proteins requiring ncAAs for activity,¹⁰ it remains a challenge to engineer host strains that are strictly dependent on ncAA incorporation into a single gene product.

We hypothesized that one can disrupt a protein–protein interface by the introduction of a large amino acid side chain which is dissimilar to any of the 20 canonical amino acids. It should then be possible to reconfigure this “unnatural” interface by mutagenesis of the surrounding residues to accommodate the ncAA and restore function. Because the new interface

consists of multiple mutations, including the ncAA, it is unlikely that a single mutation could restore active protein in the absence of the ncAA. The net result is to create two mutually orthogonal protein interfaces, one of which strictly depends on the presence of an ncAA.

To test this notion, we chose to use a homodimeric chorismate mutase (CM) as the target protein. The ncAA *p*-benzoyl phenylalanine (pBzF) was chosen to create a new interface due to its large nonpolar surface area and lack of structural similarity to the canonical aromatic amino acids (Figure 1A), and the availability of an orthogonal amber

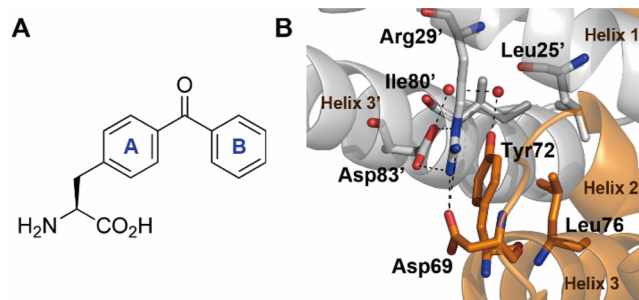


Figure 1. Chorismate mutase Tyr72 site. (A) Chemical structure of pBzF with A and B rings designated. (B) The wild type dimer is illustrated in orange and gray color. The salt bridge and hydrogen bonding network at the interface of helix 3, helix 1' and helix 3' are shown in black dashed lines (PDB ID code 1ECM).

suppressor tRNA/aaRS that efficiently incorporates pBzF into proteins in bacteria.^{8,9} The *Escherichia coli* (*E. coli*) CM (EcCM) domain of bifunctional CM-prephenate dehydratase forms a helical bundle homodimer.^{15,16} The CM is an essential enzyme that converts chorismate into prephenate for aromatic amino acid biosynthesis in bacteria, fungi, plants, and apicomplexan parasites.¹⁷ An excellent CM-selection system has been developed to identify mutants with enhanced activity based on use of the CM deficient *E. coli* strain KA12/pKIMP-UAUC and growth in the absence of tyrosine and phenylalanine. The pKIMP-UAUC plasmid expresses *Erwinia herbicola* prephenate dehydrogenase and *Pseudomonas aeruginosa* prephenate dehydratase, allowing it to utilize prephenate

Received: March 6, 2017

Published: April 17, 2017

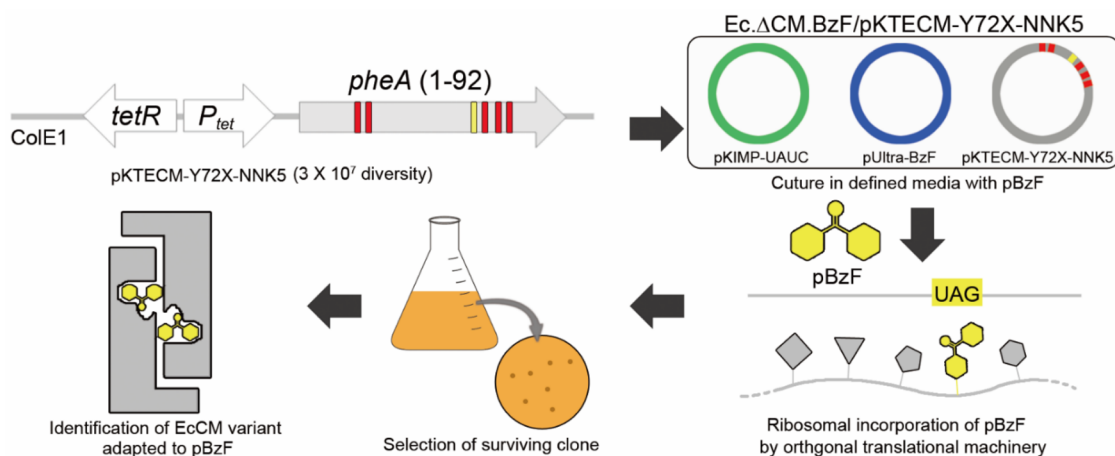


Figure 2. Schematic of selection strategy using the chorismate mutase deficient *E. coli* strain KA12/pKIMP-UAUC/pUltra-BzF (*Ec.ΔCM.BzF*) harboring pKTECM-Y72X-NNK5 plasmids in the presence of pBzF in defined media. The library contains NNK mutations at the selected residues (Leu25', Arg29', Leu76, Ile80' and Asp83'; red bars) with an amber codon (TAG) at position 72 (yellow bar). N is A, C, G or T; K is G or T.

in the shikimate pathway (Figure S1 in Supporting Information (SI)).¹⁸ A plasmid pKTECM was then constructed for expression of a CM which encodes the amino-terminal domain (Met1-Leu92) of bifunctional *E. coli* CM-prephenate dehydratase encoded by the *pheA* gene under a tetracycline inducible promoter (Figure 2).¹⁹ The KA12/pKIMP-UAUC strain survives when transformed with pKTECM in glucose minimal media (M9c) with tetracycline.

The subunit interface in the EcCM homodimer is composed largely of alpha helices with interhelical hydrophobic packing interactions and salt bridges. In particular, Tyr72 is located at the amino-terminus of helix 3 and interacts with helix1' and helix 3' through a salt bridge network and hydrogen bonding involving Arg29', Asp69, Asp83' and bridging water molecules (Figure 1B). In order to determine if Tyr72 is essential for catalytic activity, we substituted Tyr72 with alanine. The resulting Tyr72Ala mutant is catalytically inactive and failed to rescue the KA12/pKIMP-UAUC strain in M9c media (Figure S2 in SI). We therefore chose to substitute Tyr72 with pBzF (termed ECA) and found this variant also to be catalytically inactive (Figure S2 in SI).

To evolve ECA to be dependent on pBzF for growth, we hypothesized that mutation of the residues surrounding pBzF72 might reconfigure the subunit interface to restore favorable binding interactions. To this end, the codon at position 72 was mutated to the amber codon TAG and the surrounding residues (Leu25', Arg29', Leu76, Ile80' and Asp83') were randomized (NNK) (Figures 1B and 2). We used the pUltra-BzF plasmid (see Figure S3 in SI for details of pBzF incorporation by pUltra-BzF) which expresses the tRNA/aaRS pair specific for pBzF to afford the selection system KA12/pKIMP-UAUC/pUltra-BzF (*Ec.ΔCM.BzF*). We then grew *Ec.ΔCM.BzF* harboring the EcCM library in M9c media containing pBzF (1 mM), isopropyl 1-thio- β -D-galactopyranoside (1 mM) and tetracycline (100 ng/mL) at 30 °C until saturated (Figure 2). Sequencing of 20 colonies from this culture grown on LB-agar plates revealed that all clones contained identical mutations at 5 residues along with the amber mutation at position 72: Leu25'Phe, Tyr72pBzF, Leu76Thr, Ile80'Gly and Asp83'Tyr (termed ECB).

We next characterized the pBzF dependent growth of *Ec.ΔCM.BzF* transformed with the selected variant ECB encoded on the plasmid pKTECB. *Ec.ΔCM.BzF*/pKTECB

grew only in the presence of 1 mM pBzF; growth was undetectable in the absence of pBzF after 6 days (Figure 3A).

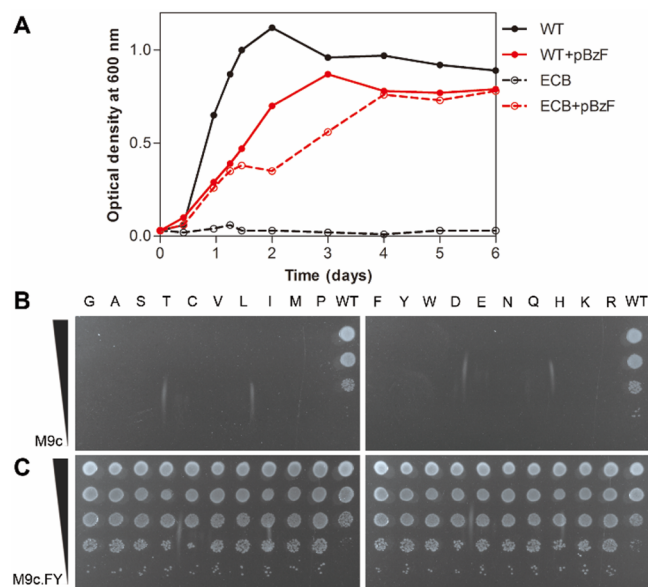


Figure 3. Characterization of pBzF dependent growth of chorismate mutase deficient *E. coli* strain. (A) Optical density was monitored by culturing *Ec.ΔCM.BzF* harboring pKTECM (solid line) or pKTECB (dashed line) in the presence (red) or absence (black) of pBzF. (B), (C) Survival of KA12/pKIMP-UAUC cells was monitored in selective (M9c) or nonselective (M9c.FY) media. The dilution series (10-fold dilutions from approximately 10^6 cells) of KA12/pKIMP-UAUC cells harboring pKTECM or pKTECB-Y72X₂₀ were spotted onto M9c and M9c.FY agar plates and incubated at 30 °C for 3 days.

By screening 8×10^{10} cells for survival in the M9c-agar plate, we determined the 10-day escape frequency to be less than 1.3×10^{-11} escapees per colony forming unit. This is well below the acceptable rate of maximum escape frequency (10^{-8}) suggested in the biosafety guidelines of the National Institutes of Health.²⁰ To determine whether any of the 20 canonical amino acids could substitute for pBzF at position 72, we constructed 20 individual plasmids which contain each of the 20 canonical mutations at position 72. None of these mutants were able to rescue KA12/pKIMP-UAUC in M9c media

supplemented with tetracycline (100 ng/mL) at 30 °C (Figure 3B and Table S1 in SI). This result suggests that single mutations at position 72 in ECB are unable to restore function.

To characterize the properties of the mutant protein, we purified EcCM and ECB both containing a carboxy-terminal hexa-histidine tag. The enzymes were purified using a Ni affinity column, followed by size-exclusion column chromatography (SEC) in a second step. The wild type chorismate mutase has multiple quaternary structures assessed by analytical SEC (Figure S4 in SI). Approximately 70% of the multimers were catalytically active and 25% was dimer, consistent with a previous report.¹⁶ In contrast, SEC showed ECB to be a homogeneous dimer (94% of total eluate) which is catalytically active. To determine the catalytic activity of ECB, the conversion of the chorismate to prephenate was spectroscopically monitored at 274 nm (Figure S6 in SI). The k_{cat} and K_m values of ECB ($k_{\text{cat}} = 10 \pm 0.3 \text{ s}^{-1}$; $K_m = 362 \pm 27 \mu\text{M}$; $k_{\text{cat}}/K_m = 28\,000 \text{ M}^{-1} \text{ s}^{-1}$) are comparable to those of wild type enzyme ($k_{\text{cat}} = 11 \pm 0.4 \text{ s}^{-1}$; $K_m = 645 \pm 57 \mu\text{M}$; $k_{\text{cat}}/K_m = 17\,000 \text{ M}^{-1} \text{ s}^{-1}$). The ECB dimer has high helical content, as expected, as judged by circular dichroism spectroscopy (Figure S7A in SI).²¹ The somewhat higher ratio ($[\theta]_{222/208}$ of 1.20 for ECB versus 1.09 for EcCM) may reflect that ECB has a more compact packing than the wild type protein. The unfolding transition temperature of ECB is 53 °C, whereas that of EcCM is 62 °C (Figure S7B in SI). The lower thermal stability is perhaps due to the loss of the extensive salt bridge and hydrogen bonding network involving Arg29', Asp69, Tyr72 and Asp83'.

The X-ray crystal structure of ECB was determined at 2.0 Å resolution (Figure 4; Figure S8 and Table S2 in SI). pBzF72 is surrounded by Phe25', Phe77' and Tyr83'. The A and B aryl rings of pBzF form parallel displaced pi-pi stacking interactions with Tyr83' and Phe25', respectively. The A ring appears to

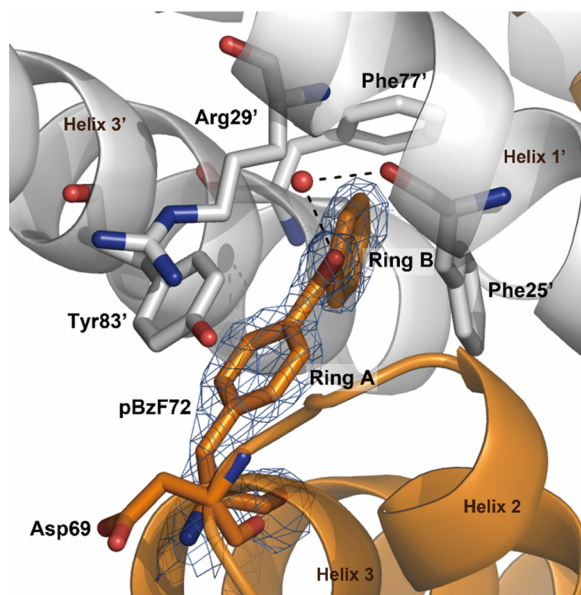


Figure 4. X-ray crystal structure of dimeric ECB is shown in orange and gray color. pBzF is located at the interface with the surrounding residues Phe25', Phe77' and Tyr83'. A water molecule (red sphere) bridges the carbonyl of pBzF and backbone carbonyl of Phe25' via hydrogen bonding interactions (black dashed lines). Electron density as a $2F_o - F_c$ contour map (2σ) is shown in meshed surface around pBzF (marine color); other residues are masked for clarity (PDB ID code 5VHT).

form a T-shaped pi-pi interaction with Phe25'.²² Phe77' interacts with the B ring and Phe25' by T-shaped pi-pi interaction. The carbonyl of pBzF72 interacts with the backbone carbonyl of Phe25' through a hydrogen bond with a bridging water molecule. Thus, the mutations that were selected to complement the Tyr72pBzF mutation lead to active enzyme with properties similar to the parent enzyme.

The strategy described above allowed us to rapidly create a homodimeric protein in which the protein–protein interface is dependent on the selective incorporation of a ncAA. None of the common amino acids were able to restore the interface when substituted for pBzF72, and the reversion rate is extremely low. Because catalytic activity depends on dimerization, the presence of pBzF is essential for the survival of Ec.ΔCM.BzF/pKTECB in defined media. Unfortunately, because CM deficiency can be complemented by aromatic amino acids in the host, this conserved enzyme is not a good candidate for the development of conditional live vaccines. Nonetheless, there are a large number of essential oligomeric proteins that are candidates for this simple approach to creating orthogonal protein–protein interfaces that are dependent on ncAAs. It may also be possible to apply a similar strategy to folding of the core of a monomeric protein.

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b02273.

Materials and Methods, Figures S1–S8, Tables S1–S3 (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*stevens@usc.edu

*schultz@scripps.edu

ORCID

Peter G. Schultz: 0000-0003-3188-1202

Notes

The authors declare no competing financial interest.

X-ray structure of ECB is available in the RCSB protein data bank (PDB ID code 5VHT).

■ ACKNOWLEDGMENTS

We thank Dr. Donald Hilvert for providing KA12/pKIMP-UAUC. We acknowledge Dr. Angad P. Mehta, Dr. Weimin Xuan and Jack Li for helpful discussions and Kristen Williams for her assistance in manuscript preparation. We would like to thank the staff at both SSRL (BL 12-2, BL 9-2) and APS (BL 23-ID-B, BL 23-ID-D) for their excellent support during data collections. This work was supported by U.S. Department of Energy Grant DE-SC0011787 and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Education (NRF-2016R1A6A3A03009203).

■ REFERENCES

- (1) Wang, L.; Xie, J.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225.
- (2) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. *Chem. Sci.* **2015**, *6*, 50.

- (3) Liu, C. C.; Mack, A. V.; Tsao, M. L.; Mills, J. H.; Lee, H. S.; Choe, H.; Farzan, M.; Schultz, P. G.; Smider, V. V. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 17688.
- (4) Xiao, H.; Nasertorabi, F.; Choi, S. H.; Han, G. W.; Reed, S. A.; Stevens, R. C.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 6961.
- (5) Hammerling, M. J.; Ellefson, J. W.; Boutz, D. R.; Marcotte, E. M.; Ellington, A. D.; Barrick, J. E. *Nat. Chem. Biol.* **2014**, *10*, 178.
- (6) Young, T. S.; Young, D. D.; Ahmad, I.; Louis, J. M.; Benkovic, S. J.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 11052.
- (7) Ai, H. W.; Shen, W.; Brustad, E.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2010**, *49*, 935.
- (8) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11020.
- (9) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. *J. Mol. Biol.* **2010**, *395*, 361.
- (10) Mandell, D. J.; Lajoie, M. J.; Mee, M. T.; Takeuchi, R.; Kuznetsov, G.; Norville, J. E.; Gregg, C. J.; Stoddard, B. L.; Church, G. M. *Nature* **2015**, *518*, 55.
- (11) Rovner, A. J.; Haimovich, A. D.; Katz, S. R.; Li, Z.; Grome, M. W.; Gassaway, B. M.; Amiram, M.; Patel, J. R.; Gallagher, R. R.; Rinehart, J.; Isaacs, F. J. *Nature* **2015**, *518*, 89.
- (12) Tack, D. S.; Ellefson, J. W.; Thyer, R.; Wang, B.; Gollihar, J.; Forster, M. T.; Ellington, A. D. *Nat. Chem. Biol.* **2016**, *12*, 138.
- (13) Si, L.; Xu, H.; Zhou, X.; Zhang, Z.; Tian, Z.; Wang, Y.; Wu, Y.; Zhang, B.; Niu, Z.; Zhang, C.; Fu, G.; Xiao, S.; Xia, Q.; Zhang, L.; Zhou, D. *Science* **2016**, *354*, 1170.
- (14) Wang, N.; Li, Y.; Niu, W.; Sun, M.; Cerny, R.; Li, Q.; Guo, J. *Angew. Chem., Int. Ed.* **2014**, *53*, 4867.
- (15) Lee, A. Y.; Karplus, P. A.; Ganem, B.; Clardy, J. *J. Am. Chem. Soc.* **1995**, *117*, 3627.
- (16) MacBeath, G.; Kast, P.; Hilvert, D. *Biochemistry* **1998**, *37*, 10062.
- (17) Bentley, R. *Crit. Rev. Biochem. Mol. Biol.* **1990**, *25*, 307.
- (18) Kast, P.; AsifUllah, M.; Jiang, N.; Hilvert, D. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 5043.
- (19) Neuenschwander, M.; Butz, M.; Heintz, C.; Kast, P.; Hilvert, D. *Nat. Biotechnol.* **2007**, *25*, 1145.
- (20) *NIH Guidelines for Research Involving Recombinant or Synthetic Nucleic Acid Molecules*; Office of Science Policy, National Institutes of Health: Bethesda, MD, 2016; pp 87–88.
- (21) Andersen, N. H.; Liu, Z.; Prickett, K. S. *FEBS Lett.* **1996**, *399*, 47.
- (22) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200.