UNNATURAL AMINO ACID MUTAGENESIS: PROGRESS TOWARDS NON-NATURAL BIOCATALYSIS AND INVESTIGATIONS OF PROTEIN THERMODYNAMICS VIA ^{19}F NMR

Amy E. Rydeen

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirement for the degree of Doctor of Philosophy in the Department of Chemistry (Biological) in the College of Arts and Sciences.

Chapel Hill
2018

Approved by:

Eric Brustad

Gary Pielak

Bo Li

Dorothy Erie

Marcey Waters

ABSTRACT

Amy E. Rydeen: Unnatural amino acid mutagenesis: Progress towards non-natural biocatalysis and investigations of protein thermodynamics via ^{19}F NMR
(Under the direction of Eric Brustad)

Evolution created the complex and versatile protein landscape that underpins cellular function, through the combinatorial assembly of only 20 canonical amino acids. Contemporarily, protein engineers are expanding the capabilities of protein beyond the constraints of natural selection by tailoring them for applications in a 'man-made' world. For instance, proteins are finding widespread success as immuno-therapeutics and as industrial catalysts for the production small molecule pharmaceuticals. However, the pursuit of protein engineers is hindered by the limited chemical functionality imparted by the natural proteinogenic amino acids. *In vivo* unnatural amino acid mutagenesis technology, also known as genetic code expansion, has allowed the localized and/or widespread installment of non-canonical amino acids bearing non-natural functional groups. This simple yet powerful technology has supplemented the fundamental few with hundreds of unnatural amino acids (UAAs). Despite the numerous applications of UAAs that have been disclosed, there still exists significant potential for further exploration.

This dissertation advances the applications of UAA mutagenesis to include the installation of UAAs for novel biocatalysis and the biophysical investigation of protein and protein complex stability via ^{19}F NMR. We present the design, synthesis and *in vivo* incorporation of organocatalyst-inspired proline-modified UAAs. Free

proline-elaborated UAAs perform aqueous aldol chemistry, however fail to transfer this activity to a protein scaffold. The groundwork discussed herein should assist future efforts towards the creation of protein catalysts with UAA-based reactivity. In addition, we utilized a ^{19}F UAA to study the stability of a protein and weak protein complex in the presence of osmolytes and related biological small molecules via ^{19}F NMR. The data demonstrate that although osmolytes are primarily characterized as protein stabilizing agents, related non-osmolytes stabilize a protein fold within the range of the most and least powerful osmolytes. On the other hand, osmolytes and non-osmolytes differ in their effect on the stability of a protein complex. Osmolytes had negligible influences on a charge mediated protein-protein interaction and non-osmolytes were significantly perturbing. This result implies that osmolytes were selected by nature for their compatibility with the electrostatic interactions between protein surfaces that are attributed to the regulation of cellular function and structure.

This dissertation is dedicated to Beatrice Neffue, Mimi.

ACKNOWLEDGEMENTS

To my mentor, thank you Eric for your endless enthusiasm and support of all my research and career explorations. I truly appreciate being given the opportunity to make my phD, ‘my own’. To Gary, thank you for your hospitality, in your home and lab. You remind me to appreciate simplicity, whether it be a glass of champagne or a well-behaved protein, it’s the small things in life.

To my lab, all former and present members, and fellow red pod lab mates, thank you for always being there when I needed a hand, to discuss my projects, lend reagents, or an excuse to go for a walk.

I would like to thank the community of Franklin Street Yoga Center. I am constantly supported by the friendships and skills I developed through my practice. I owe my health, mentally and physically, to the studio.

Most graciously, I thank my family. I can’t understate how thankful I am to have had my sister and brothers close by and present during this journey. Our weekly meals together kept everything in perspective. To Austin, thank you for reminding me I am capable and for making me delicious salads for lunch, our relationship keeps me afloat. Thank you to my parents, you know I wouldn’t be what I am today, without you two. Lastly, to Mimi, you are my inspiration and my idol. Your humility, dedication to your family and tenacity towards life form the foundation of the character I strive to embody.

PREFACE

At the time this thesis was composed, the final chapter, ^{19}F NMR investigation of protein and protein complex stability via unnatural amino acid mutagenesis, was submitted as a research report to the Proceedings of the National Academy of Science (PNAS). The text of the chapter is verbatim as submitted, however the title was changed to reflect the narrative of this thesis. The title of the submitted research report, was “Osmolytes are compatible with protein surfaces”.

Prior to submitting the thesis for publication, we received notification that the submission to PNAS was rejected. The final chapter has since been rewritten and submitted to the Journal of the American Chemical Society as a communication titled, Osmolytes and protein-protein interactions.

TABLE OF CONTENTS

LIST OF MAIN TABLES

LIST OF MAIN FIGURES

LIST OF SUPPLEMENTARY TABLES

LIST OF SUPPLEMENTARY FIGURES

LIST OF ABBREVIATIONS

°C	Degrees Celsius
$\Delta G^{\circ\prime}_{D\rightarrow M}$	Free energy of dissociation
$\Delta G^{\circ\prime}_{F\rightarrow U}$	Free energy of unfolding
$\Delta\Delta G^{\circ\prime}_{D\rightarrow M}$	Change in the free energy of dissociation
$\Delta\Delta G^{\circ\prime}_{F\rightarrow U}$	Change in the free energy of unfolding
2XYT	Yeast extract tryptone
4-OT	4-oxalocrotonate tautomerase
9-BBN	9-Borabicyclo[3.3.1]nonane
aaRS-tRNA	Amino-acyl tRNA synthetase-tRNA
Ala	Alanine
Amp	Ampicillin
Arg, R	Arginine
ArM(s)	Artificial metalloenzyme(s)
Asn, N	Asparagine
BCN	Bicyclo[6.1.0]nonyne
BocK	N-Boc-L-lysine
Bpa	p-benzoyl-L-phenylalanine
BpyAla	(2,2'-bipyridin-5yl) alanine
Cat	Chloramphenicol acyltransferase
Cm	Chloramphenicol
Cys, C	Cysteine
Da	Dalton

dATP	Deoxyadenosine triphosphate
DCC	N,N'-Dicyclohexylcarbodiimide
DCM	Dichloromethane
dCTP	Deoxycytidine triphosphate
DERA	2-deoxyribose-5-phophate aldolase
dGTP	Deoxyguanosine triphosphate
DIPEA	N,N-Diisopropylethylamine
DMF	N,N-dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dTTP	Deoxythymidine triphosphate
E. *coli*	*Escherichia coli*
EDC•HCl	3-(3-Dimethylaminoproply)-1-ethyl-carbodimide hydrochloride
e.e.	Enantiomeric excess
EP	Error prone
FACS	Fluorescence assisted cell sorting
FPLC	Fast protein liquid chromatography
GB1	B1 domain of the streptococcal immunoglobulin binding protein G
GFP	Green fluorescent protein
Gln, Q	Glutamine
Glu, E	Glutamate
GLYT1	Na^+/Cl^--dependent glycine transporter

HCl	Hydrochloric acid
HPLC	High performance liquid chromatography
Hr	Hour (s)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOBt	Hydroxybenzotriazole
HSQC	^{15}N-^{1}H heteronuclear single quantum correlation
IGPS	Indole-3-glycerol phosphate synthase
IPTG	Isopropyl ß-D-1-thioglactopyranoside
Ile, I	Isoleucine
Kan	Kanamycin
K_D	Dissociation constant
kDA	Kilodaltons
KPi	Potassium phosphate
LB	Lysogeny broth
LCMS	Liquid chromatography mass spectrometry
Leu, L	Leucine
LmrR	Lactoccocal multidrug resistance regulator
Lys, K	Lysine
M	Molar
MbPylRS	*Methanosarcina barkeri* pyrrolysyl-tRNA synthetase
MHz	Megahertz
Min	Minutes (s)
MjTyrRS	*Methanococcua jannaschii* tyrosyl-tRNA synthetase

Mm	*Methanosarcina mazei*
mM	mili molar
Mol	Mole
MS	Mass spectrometry
NaCl	Sodium chloride
$NaCNBH_3$	Sodium cyanoborohydride
NaOH	Sodium hydroxide
NMR	Nuclear magnetic resonance
NNK	Randomized library site where N=any base and K=G or T
OD	Optical density
PBS	Phosphate buffered saline
pCNFRS	p-cyanophenylalanine amino acyl tRNA synthetase
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethyleneglycol
pH	Decimal logarithm of hydronium ion concentration
Phe, F	Phenylalanine
pK_a	Acid dissociation constant
POP	Prolyl oligopeptidase
Pro, P	Proline
Pro4AF	D-prolyl-4-amino-L-phenylalanine
ProK	D-prolyl-L-lysine
PyrrolK	N6-((pyrrolidin-3-yloxy)carbonyl)-L-lysine

Q-TOF	Quadrupole time-of-flight
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SH3	the SH3 domain of the *Drosophila* signal transduction protein drk
Ser, S	Serine
sfGFP	Superfolder green fluorescent protein
SPAAC	Strain-promoted azide-alkyne cycloaddition
Strep	Streptomycin
TAG	Amber stop codon
TBNS	trans-beta-nitro styrene
Tet	Tetracycline
TFA	Trifluoroacetic acid
TFMF	Trifluoromethyl-L-phenylalanine
tHisF	Histidine synthase F
Thr, T	Threonine
TIC	Total ion count
TMAO	Trimethylamine-N-oxide
tRNA	Transfer RNA
Tyr, Y	Tyrosine
UAA	Unnatural amino acid
UV	Ultraviolet
Val, V	Valine
WT	Wild type

Y2H	Yeast 2-hybrid

CHAPTER 1: *IN VIVO* UNNATURAL AMINO ACID MUTAGENESIS

1.1 An introduction to genetic code expansion technology

The *in vivo* incorporation of unnatural amino acids (UAAs) has proven to be an invaluable tool for the engineering of proteins and enzymes. UAAs can enhance protein stability, enzyme activity and selectivity, allow biophysical investigations of protein binding interactions, stability and dynamics, and create proteins with controlled and tunable photo- and chemical-reactive properties. Reviews of this technology, including its applications and the wealth of UAAs incorporated, continue to grow in size and number (1-8).

The installation of UAAs in protein sequences has been approached in several ways in multiple organisms. Overall, two general approaches have emerged: residue-specific and site-specific incorporation of UAAs. Residue-specific incorporation replaces each location of a specific residue with a non-natural structural mimetic. For example, fluorinated isosteres of tryptophan and leucine and been readily incorporated into proteins proteome-wide, as their similar size and shape when compared to their native counterparts permits promiscuous use by the translational machinery. This technique has been used to increase thermostabilities of proteins by universal incorporation of fluorinated amino acid derivatives (9). Residue-specific incorporation is hindered, however, by its reliance on the permissiveness of endogenous tRNA synthetases, which restricts UAA scope to structural analogs of the 20 canonical amino acids. It also necessitates the

deprivation of the canonical amino acid isostere from the growth media, which complicates and often contaminates scaled protein preparations. A major drawback to the residue specific method is the inability to select a single location for UAA incorporation.

The focus of this dissertation is the site-specific incorporation of UAAs via codon reassignment and suppression. This approach offers complete control over the location at which the UAA is installed and allows for multiple UAAs to be incorporated by reassigning multiple sense or non-sense codons. Herein, we employ *Escherichia coli* (*E. coli*) as a heterologous host for the engineering and production of proteins containing UAAs.

Site-specific incorporation of UAAs depends on the creation of orthogonal amino-acyl tRNA synthetase-tRNA (aaRS-tRNA) pairs unique for an UAA and reassigned codon. Orthogonality implies that the synthetase will not aminoacylate endogenous tRNAs and, in parallel, that the tRNA will not be recognized by native amino-acyl tRNA synthetases. The creation of orthogonal aaRS-tRNA pairs, in some cases, has proven to be straightforward since translational machinery from vastly different domains of life often do not cross-react with the *E. coli* translational apparatus. Two sets of aaRS-tRNA pairs have found widespread usage: the *Methanococcua jannaschii* tyrosyl-tRNA synthetase (MjTyrRS)-tRNA and the *Methanosarcina mazei* or *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase (Mm or MbPylRS)-tRNA.

To genetically encode an unnatural amino acid, a codon that does not encode a canonical amino acid, typically the amber stop codon (TAG), is reassigned to

encode for the unnatural amino acid. This is accomplished first mutating the anti-codon loop of the tRNA ($tRNA^{CUA}$); however, additional tRNA mutations are sometimes required to improve TAG-suppression efficiency. Next, the amino acid binding site of the synthetase is mutated and variants are screened or selected for the ability to specifically charge the $tRNA^{CUA}$ with the unnatural amino acid of interest, Figure 1.1. Several positive and negative selection have been reported and have found wide spread application for the isolation of an orthogonal synthetase from often-large number, > 10^6, of generated mutant synthetases (7).

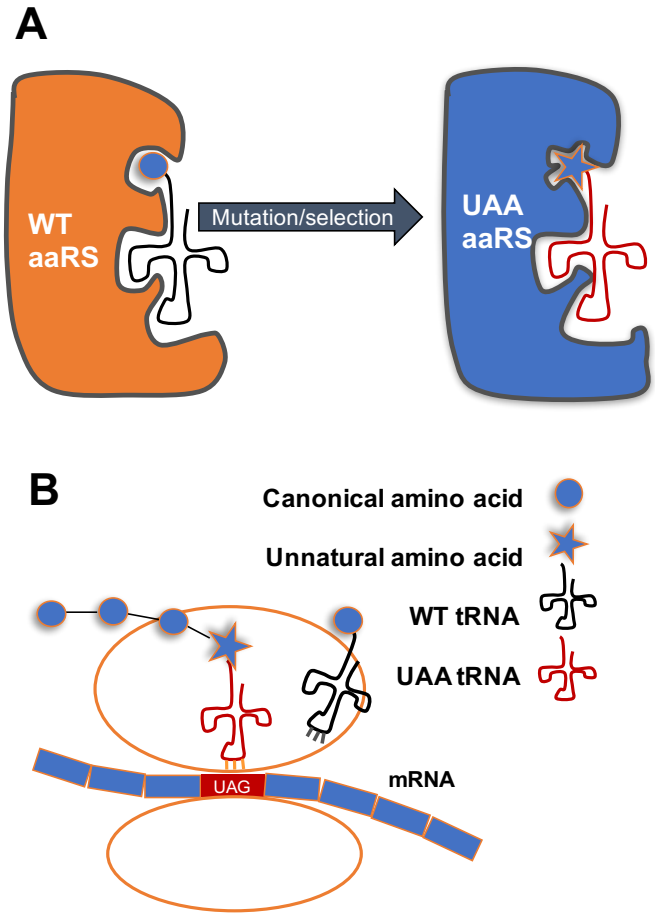
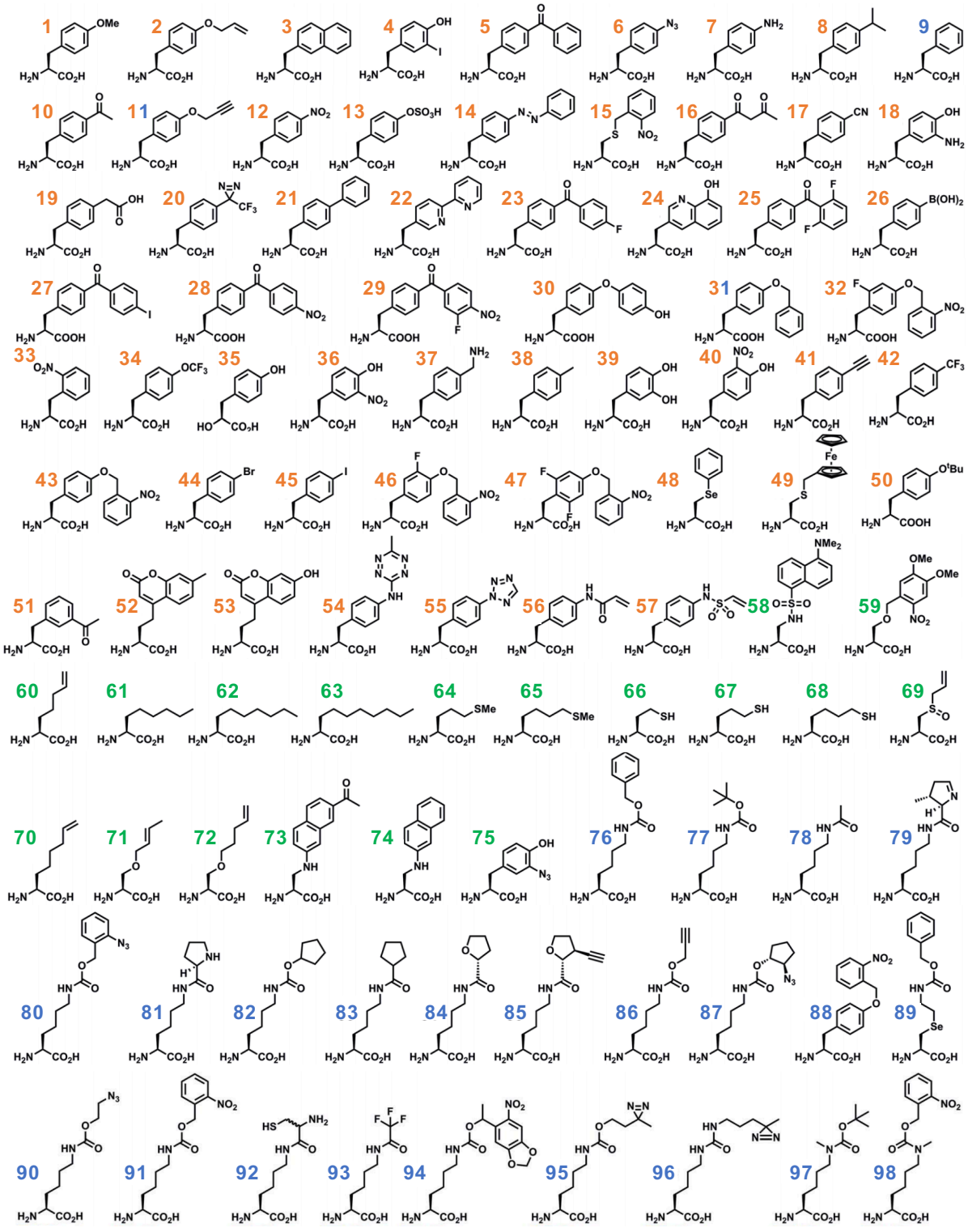


Figure 1.1 Site-specific UAA incorporation

(A) mutation and selection of an aaRS from different domain of life creates an orthogonal aaRS specific for the UAA. (B) UAG codons in mRNA encode for the UAA during translation on the ribosome.

Extensive mutational exploration of both *Mj*TyRS and *Mm*- or *Mb*PylRS has resulted in the creation of orthogonal aaRS-tRNA pairs specific for UAAs that span diverse structures and sizes, as shown in Figure 1.2. In summary, protein engineers, once limited by the diversity of the 20 canonical amino acids when making structural and chemical modifications to amino acid side chains, now have access to over 150 unnatural amino acids containing diverse structures and non-natural chemical functional groups (2). Functional diversity of unnatural amino acids that have been site specifically genetically encoded include: (1) biophysical, spectroscopic and imaging probes, (2) handles for bio-orthogonal reactions, (3) post translation modifications (PTM) or PTM mimetics, and (4) photo-activated motifs such as photo-caged, photo-isomerizable, or photo-crosslinkable amino acids. The reader is directed elsewhere for a review on the applications of this array of UAAs (10).


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98

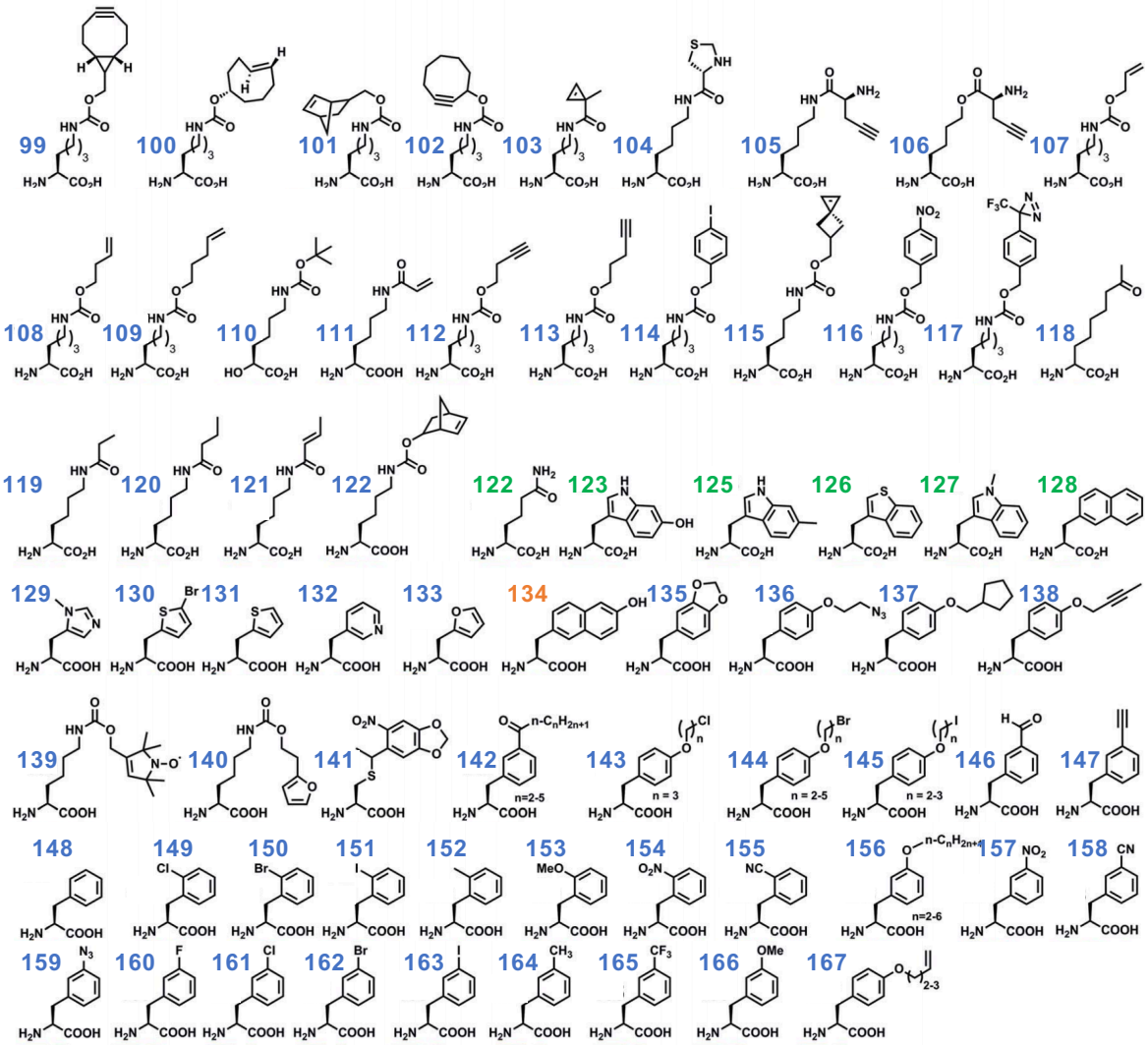


Figure 1.2 UAAs genetically encoded, *in vivo*, as of 2015

The color of the number indicates the aaRS/tRNA pair used, orange is *Mj*TyRS, blue is *Mm* or *Mb*PylRS and green is other. Figure adapted from [Chem. Sci., 2015, 6, 50.]-Published by The Royal Society of Chemistr

1.2 Selected applications of UAAs

To underscore the functional utility of genetic code expansion methodology, select, but representative, applications of UAA mutagenesis are described below.

1.2.1 UAAs install synthetic catalysts in protein scaffolds

Over billions of years, evolution has fine-tuned enzymes with catalytic capabilities near perfection. Enzymes demonstrate unrivaled selectivity, rate

enhancements and catalytic efficiencies for their native reactions (11, 12). Employing enzymes as compliments and/or supplements to conventional synthetic catalysts in industrial processes is a highly sought after goal. For example, enzymes operate under environmentally benign reaction conditions including in water at room temperature, neutral pH and without the need for organic solvents. However, it has proven challenging to engineer enzymes to catalyze industrial reactions on manmade substrates or to carry out chemical transformations that do not already occur in nature. The chemical capabilities of enzymes, resulting from the canonical 20 amino acids and cofactors, are uniquely suited for and thereby limited by the natural world in which they evolved.

Genetic code expansion technology merges the chemical ingenuity and creativity of man with that exquisite and powerful translational machinery already developed by nature. Thus, the UAA mutagenesis provides a unique mechanism to expand enzyme chemical functionality beyond the constraints of nature's design. For example, mounting conventional synthetic catalysts at a single site within protein active sites is possible through the use of UAAs displaying handles for covalent bioconjugation. Several classes of UAAs containing bio-orthogonal, or biocompatible, conjugation handles have been developed including reactive sites for Cu(I)-catalyzed or copper free alkyne-azide cycloaddtion (the so-called 'click' reaction), inverse-electron-demand Diels-Alder between tetrazines and strained alkenes, and the Staudinger ligation between alkyl azides and ester-functionalized triphenylphosphines, Figure 1.3 (13).

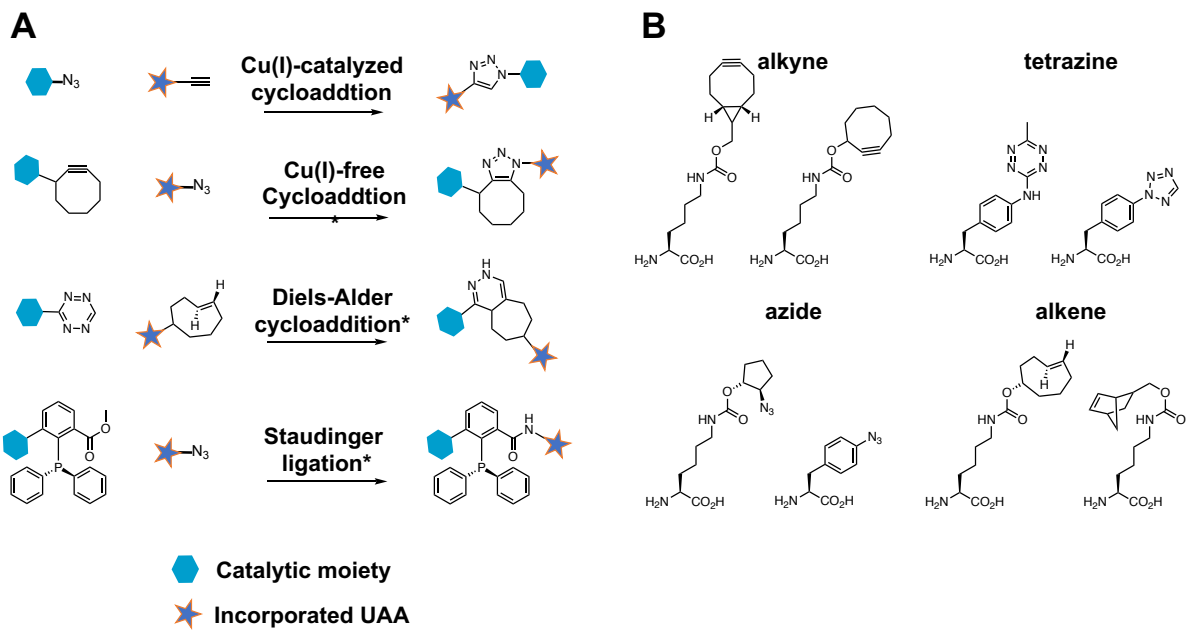


Figure 1.3 UAA-based bio-conjugation strategies

(A) Select bio-orthogonal reactions for conjugation of catalysts to protein scaffolds. *Reactive functionality can be switched between catalytic moiety and UAA. For example, in the Cu(I)-free cycloaddition, a cyclooctyne UAA can be coupled with a catalytic moiety displaying an azide. Figure is adapted from Nat Chem Biol., 2013, 8, 475. (B) Select UAAs containing reactive bio-orthogonal handles.

UAA-based bioconjugation strategies, in particular, the strain-promoted azide-alkyne cycloaddition (SPAAC, Figure 1.3A Cu(I)-free cycloaddition), have been exploited in the engineering of artificial metalloenzymes (ArMs) (14) . ArMs are hybrid catalysts composed of a covalently implanted conventional transition metal catalyst within the hydrophobic pocket of a protein scaffold. SPAAC between bicyclo[6.1.0]nonyne (BCN) and azide substituted reactive partners is a poplar conjugation strategy due to its rapid second order kinetics, bioorthogonality and specificity. Generally, an azide containing UAA is genetically encoded in a protein and subsequently reacted with a BCN-decorated transition metal catalyst. Because the conjugation will not proceed on naturally occurring amino acids, in theory, any protein could serve as a host, provided it can accommodate the UAA-catalyst pair.

Highly stable proteins with large active sites, such as *Pfu*-POP, a member of the prolyl oligopeptidase (POP) family from *Pyrococcus furiosus* and tHisF, a synthase subunit of the glutaminase synthase from *Thermatoga maritima* have successfully hosted bulky BCN substituted synthetic metal catalysts (14, 15). Figure 1.4 shows a schematic for the creation of an example ArM for cyclopropanation via bioconjugation of a BCN-substituted dirhodium catalyst with an azide UAA in the pocket of a scaffold protein.

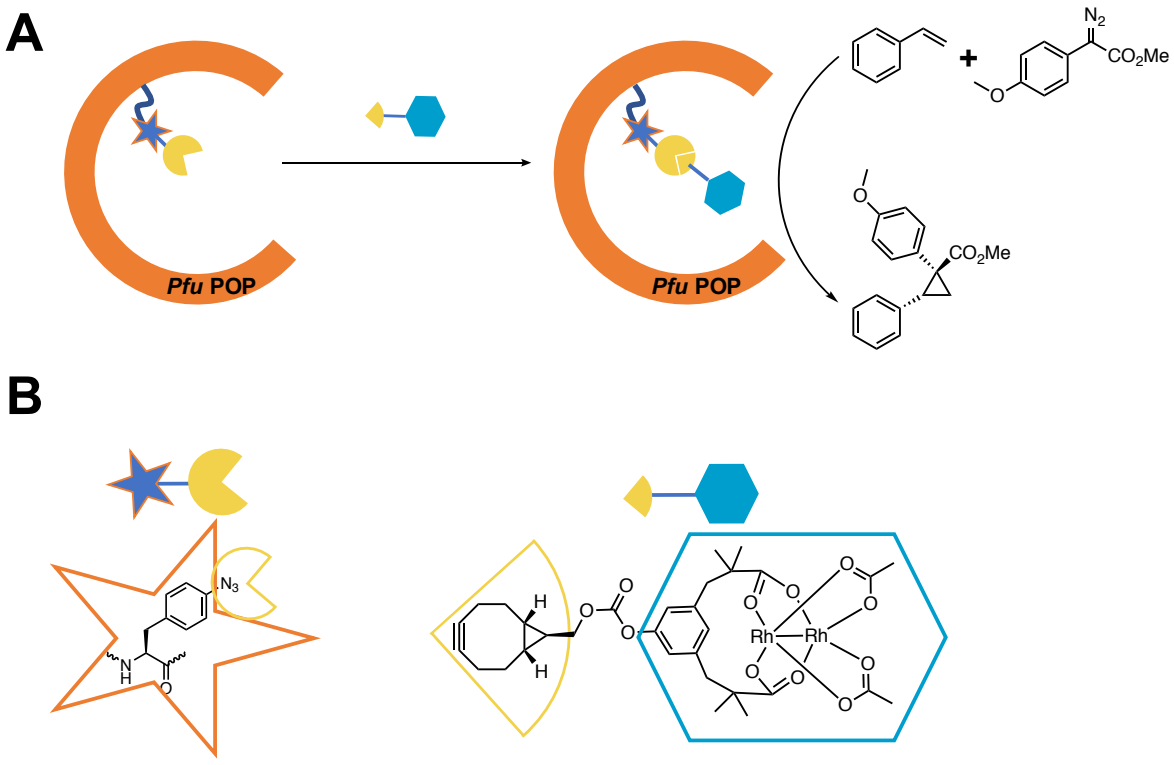


Figure 1.4 UAA-based SPAAC creates cyclopropanation ArM

(A) A prolyl oligopeptidase from *Pyrococcus furiosus* displaying genetically encoded p-azidophenylalanine within the active site is conjugated to bicyclo[6.1.0]nonyne substituted dirhodium complex and catalyzes the model cyclopropanation of styrene with donor-acceptor diazo compounds. (B) SPAAC bio-conjugation partners for the creation of an artificial metalloenzyme (ArM).

Primary ArM designs, constructed initially in a naïve protein scaffold by conjugation of the UAA and transition metal-catalyst, does not typically fabricate

catalysts with desirable yields or selectivities. Protein scaffolds often need to be engineered to complement and supplement the chemistry imparted by the catalyst. Rational mutagenesis of residues in proximity to the catalyst can enhance reactivity and/or selectivity but the methodology is time consuming and limited by the availability of a crystal structural for guidance (16). In addition, mutations outside of the active site can often contribute favorably to non-natural function, in unpredictable ways (17). However, mutations outside of the active site are difficult to predict, a priori, and require rigorous screening efforts to identify from random mutation libraries.

Directed evolution has been established as a key methodology for the non-rational engineering of enzymes for non-natural catalysis (18). Directed evolution mimics natural selection through iterative rounds of mutagenesis and screening the mutant library for improvements in a desired characteristic. Recently, Lewis reported the first directed evolution of an ArM via non-rational mutagenesis (17). An ArM was constructed via SPAAC between a BCN substituted dirhodium catalyst and an azide displaying UAA in the active site of *pfu* POP. The enzyme was evolved through iterative rounds of error prone PCR and screening the library for improvements in the selectivity of cyclopropanation between alkene and diazo compounds. The resulting variant, obtained after 3 rounds of evolution contained 4 mutations, both within and beyond the active site, catalyzed the cyclopropanation with improved selectivity (92% e.e.) and yield (~75%) compared to the parent ArM.

To date, all attempts to engineer non-natural chemistry in proteins using genetic code expansion technology have sought to incorporate non-natural transition

metal centers into proteins. However, there are a number of alternative catalytic strategies, including organocatalysis and photoredox catalysis, that have yet to be explored in this expanding field of protein design. Chapter two of this thesis describes our progress at engineering novel UAA structures that will open novel organocatalytic motifs into biocatalysts.

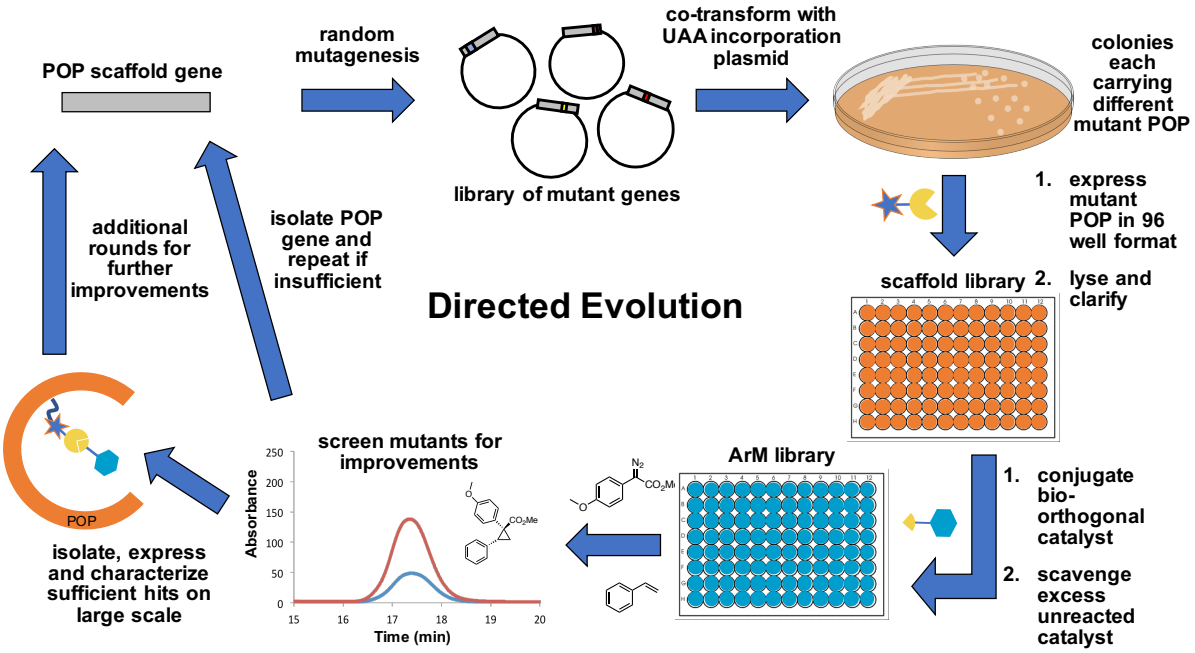


Figure 1.5 Directed evolution of bio-conjugation UAA-based ArM

The scaffold protein, POP, is non-rationally mutated via error prone PCR to create a library of protein variants that are co-transformed with a plasmid harboring the orthogonal aaRS-tRNA pair specific for the UAA. The scaffold variants are then expressed in the presence of UAA, lysed and clarified in a 96 well format. The ArM is constructed by addition of the catalyst equipped with the handle for bio-conjugation to the UAA. The resulting ArMs are screened for catalytic improvements, in this case, improved enantioselectivity in the cyclopropanation between styrene and diazo substrates. Hits are then verified on a large scale and subjected to further rounds of mutation and screening for further improvements. Figure adapted from Nature Chemistry, 2018,10, 318.

1.2.2 Photocrosslinking UAAs investigate protein-protein interactions

Protein-protein interactions are universal and essential to the function of the cell. Mapping the cellular network of protein-protein interactions, which include both strong and weak contacts, furthers our understanding of disease (19). Traditional methods for studying these interactions, such as yeast 2-hybrid (Y2H) assays, are

performed under non-native conditions. Therefore, location specific and fleeting physiological interactions escape detection. Further, the Y2H methodology favors detection of strong interactions and many critical protein-protein interactions are weak and transient (20). Photocrosslinking UAAs offer many benefits over traditional non-covalent techniques of investigating the interaction of macromolecules *in vivo* and *in vitro*.

Photocrosslinking UAAs are activated by ultraviolet light to form radicals that undergo reaction with functional groups on a nearby associated proteins to covalently trap an otherwise weak interaction (3). Photocrosslinking UAAs detect specific protein-protein interactions and have minimal false identifications due to the high reactivity, short half-life and the ability of the activated state to relax or react with itself or water in the absence of a binding partner. The site-specific incorporation of three classes of photo-crosslinking UAAs, including benzophenones, aryl azides and alkyl and aryl diazarines has enabled the study of otherwise difficult to probe protein-protein interactions, such as membrane protein interactions (20-26).

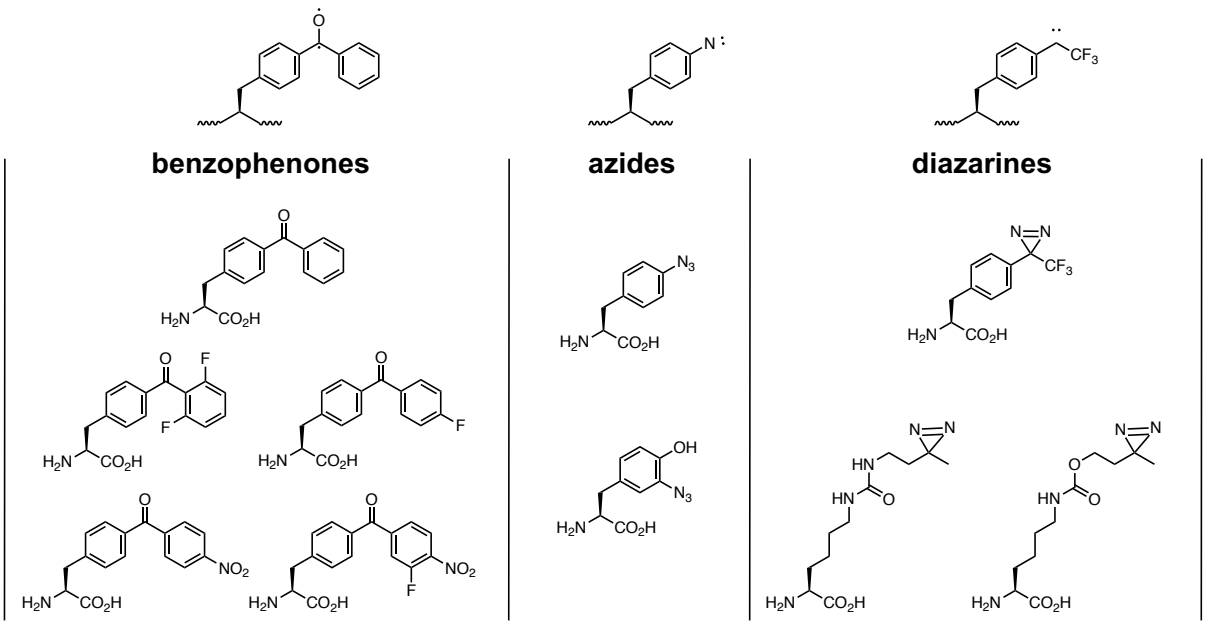


Figure 1.6 Photocrosslinking UAAs and corresponding reactive radical species

The site-specific incorporation of p-benzoyl-L-phenylalanine (Bpa, UAA #5 Figure 1.2) has found wide spread application in probing protein-protein interactions in *Escherichia coli (21)*. The interaction between transmembrane protein complex, SecYEG, and SecA was elucidated, *in vivo*, using this UAA, Figure 1.7 (27). The SecA – SecYEG interaction mediates post-translational translocation of outer membrane and periplasmic proteins across the inner membrane of *E. coli*. Bpa was site-specifically incorporated at 53 sites in the SecY protein of the SecYEG complex. This extensive scanning strategy identified previously unknown sites where SecA interacts with SecY in agreement with the crystal structure of the complex. Further, dynamics of the interaction were elucidated via *in vivo* experiments under stabilized conditions where the protein is locked in a 'membrane inserted' state and *in vitro* experiments where components (ATP, substrate, etc) were systematically omitted to probe how the interaction is dependent on the functional state of the both partners.

This example highlights the ability of site-specific UAA incorporation to study a difficult binding interaction under physiological conditions.

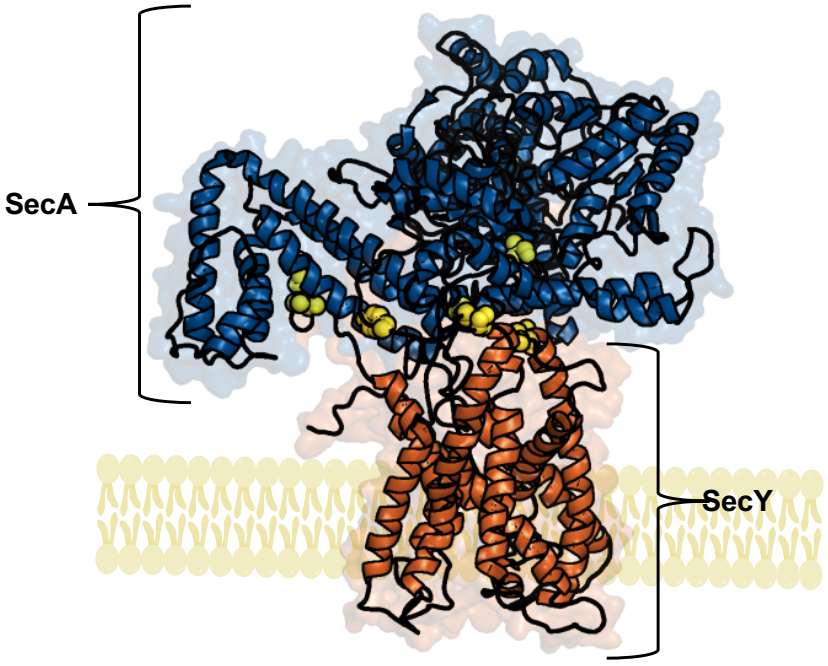


Figure 1.7 Bpa identifies residues of membrane protein-protein interaction

The structure of SecA, blue, and SecY, orange, complex, PDB ID: 3DIN. The residues where Bpa was incorporated and *in vivo* crosslinking was observed are shown as yellow spheres.

Photocrosslinking UAAs are superior to non-covalent affinity based methods for the identification of weak, transient interactions. They provide structural information about distinct interfaces via site-specific interrogation of single residues in a protein-protein, protein-nucleic acid and/or protein-ligand interaction. Lastly, the identification of unknown biologically relevant interactions with minimal nonspecific linking facilitates the discovery and correct mapping of macromolecular interactions.

1.2.3 UAAs as ^{19}F NMR probes

Nuclear magnetic resonance (NMR) spectroscopy is a fundamental tool in the investigation of macromolecular structures, interactions, dynamics and thermodynamics. Proteins can be studied *in vitro* and *in vivo* at a global and

residue–specific level via NMR. Fluorine is an ideal nucleus for the visualization of protein by NMR. ^{19}F is 100% naturally abundant, scarcely found in biology, hypersensitive to chemical environment when compared to ^{1}H yet similar in size and therefore tolerated with minimal structural rearrangements (28). Labeling with ^{19}F permits study of proteins otherwise too large for conventional ^{1}H- and ^{15}N-based NMR experiments. In addition, the spectra are significantly more simple to interpret (29). For these reasons, several UAAs containing fluorine have been developed and applied in the study of macromolecules via ^{19}F NMR, see UAA # 34, 42, 46, and 47 Figure 1.2.

The site-specific incorporation of ^{19}F UAAs overcomes drawbacks of residue specific technology for producing ^{19}F labeled proteins. Coaxing the native translational machinery to use fluorinated derivatives of natural amino acids, such as tryptophan, seldom reaches 95% incorporation at a single site and produces variability in labels throughout a protein sequence (30). The first report of the site-specific incorporation of a fluorine containing UAA, trifluoromethyl-L-phenylalanine (TFMF, UAA # 42 Figure 1.2), detailed the *in vivo* production of large quantities of uniform, homogenous, singly ^{19}F labeled protein (31). The UAA was incorporated at the interface active site of two dimeric proteins and reported structural changes of the complex resulting from inhibitor and substrate binding. Further, the signal of the labeled protein was sufficient to monitor chemical shift changes in cells given the UAA contains three equivalent ^{19}F atoms.

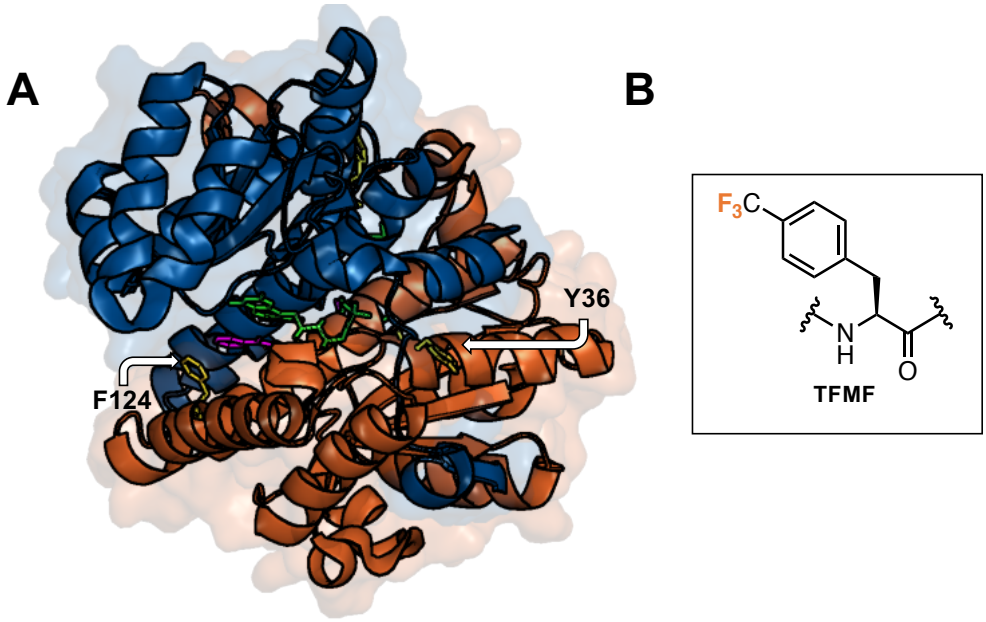


Figure 1.8 TFMF at dimer interface reports structural changes via ^{19}F NMR

(A) Dimer shown with subunits colored orange and blue, PDB code: 1ICR. Site for TFMF are colored yellow, Phe 124 of the active site and Try36 at the interface. Bound flavin mononucleotide, cofactor, is green and nicotinic acid, inhibitor, is pink. (B) Structure of TFMF.

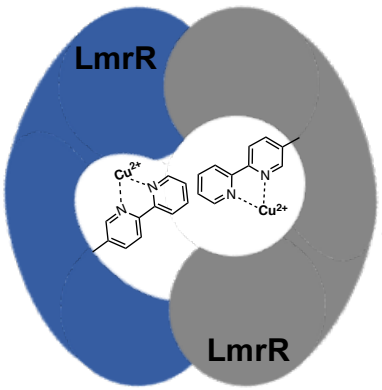
1.3 Furthering these applications of UAAs

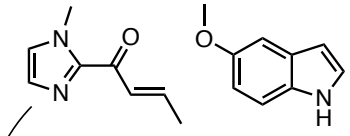
This dissertation, rooted in the previously described applications of UAAs, furthers a novel yet underexplored application of UAA mutagenesis and expands upon established applications. Firstly, inspired by the UAA-based conjugation of synthetic catalyst in protein scaffolds, we report progress towards the direct installation of catalytic moieties via the *in vivo* incorporation of catalytic UAAs employing orthogonal aaRS-tRNA pairs. We then discuss an ^{19}F NMR biophysical investigation of protein and protein complex stability utilizing the UAA, trifluoromethyl-L-phenylalanine (TFMF).

1.3.1 UAAs as synthetic catalysts in protein scaffolds

The genetic incorporation of catalytic unnatural amino acids as a mean to introduce non-natural chemistry to protein scaffolds represents an untapped

application of genetic code expansion technology. There is only one instance of *in vivo* unnatural amino acid incorporation for direct enantioselective catalysis (32). A metal binding amino acid, (2,2'-bipyridin-5yl) alanine (BpyAla), was installed in the hydrophobic cavity of the dimer interface of Lactoccocal multidrug resistance regulator (LmrR). The protein harboring the unnatural amino acid bound to copper(II) was able to catalyze a Friedel-Crafts acylation of 2-methylindole with 22% conversion and 66% enantiomeric excess after optimization of location of the UAA and mutation to residues proximal the Cu(II) catalyst, Figure 1.9.





catalyst	% conversion	% e.e.
$Cu(NO_3)_2$	98	-
LmrR-Bipy-Cu^{II}	27	49
LmrR-E107A-Bipy-Cu^{II}	22	66

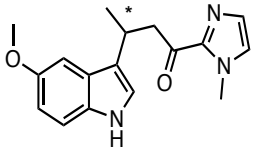


Figure 1.9 I*n vivo* incorporation of metal binding UAA for catalysis

Friedel-Crafts acylation of indoles by ArM created by genetically encoding BpyAla at the dimer interface of LmrR. Enantioselectivity was improved by rational mutagenesis of residues proximal to BpyAla.

In contrast to the singular report of a UAA acting as the catalytic moiety of protein catalyst, UAAs are frequently used to tether conventional catalysts in the hydrophobic pocket of protein scaffolds (33). Although this strategy has successfully fashioned ArMs for enantiocatalysis, it represents a significant engineering effort. We suggest that engineering non-natural protein catalysts via *in vivo* incorporation of

catalytic amino acids offers advantages over UAA- or canonical amino acid-based conjugation strategies. The most prominent benefit would be the direct simplification of the catalyst production methodology, Figure 1.10. When utilizing a catalytic UAA, the active protein scaffold is purified from a heterologous host and characterized in a single event. In contrast, for unnatural amino acid-based bio-conjugation of a catalyst, the protein scaffold must be purified from the expression host, conjugated to a catalyst and subsequently re-purified from non-conjugated catalyst. Consequently, iterative optimization, i.e. directed evolution, of the hybrid protein-catalyst conjugate is challenging and only one instance has been reported (17). Directed evolution, consisting of random mutagenesis followed by high throughput screening of the mutant library necessitated the removal of excess non-conjugated catalyst prior to screening to avoid skewing of yields and selectivities, Figure 1.5.

Selection of a protein scaffold to host the catalyst is a complicated endeavor of concern to both direct-UAA and bioconjugation-UAA based strategies for catalyst immobilization. In both cases, the scaffold should be highly stable in order to tolerate extensive mutations for engineering and optimization, and to facilitate high throughput purification. However, we highlight that catalytic UAAs may expand the viable scaffold structures because they can be designed in a more compact fashion (i.e. no linkers and bioconjugation handles are necessary) and introduced into proteins directly (not through post-translational bioconjugation). For example, only a select few scaffolds have been proven capable of accommodating bulky BCN-containing UAAs, #99 and 103 in Figure 1.2, which are necessary for the favored SPAAC conjugation strategy. Therefore, the UAA-based bioconjugation of catalysts

places challenging steric constraints on the proteins' hydrophobic core. Catalytic UAA's negate the inclusion of bio-conjugation handles that often contribute to the steric bulk of the conjugated catalyst. Catalytic UAAs may also alleviate the complexity and laborious nature of synthesizing both bio-conjugation partners, the UAA and the catalyst.

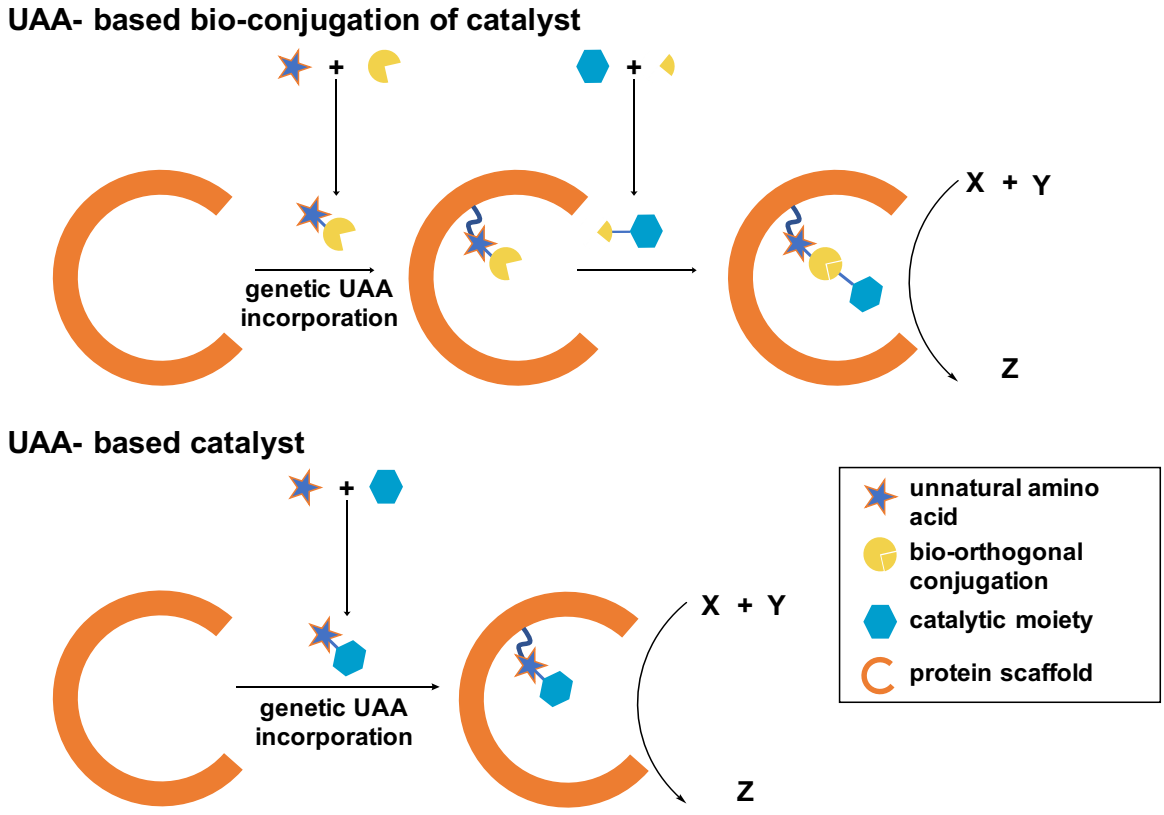


Figure 1.10 Advantage of installation of UAA-based catalyst

In vivo incorporation of catalytic UAAs may prove more attractive than alternative strategies for the creation of non-natural protein catalysts, including cysteine conjugation of catalysts and recruitment of cofactor-based catalysts in a protein scaffold. ArMs have been created via the introduction of a single cysteine residue and subsequent nucleophilic attack of the residue on a reactive site on a conventional catalyst (34). This strategy can require labor intensive mutational

efforts if many cysteines must be removed from the native protein scaffold (35). In contrast, UAA-based catalysts only require a single TAG mutation to site-selectively install the catalyst. Another popular strategy for the construction of ArM is the recruitment of cofactor-like catalysts to protein hydrophobic cores. For example, synthesizing a biotin containing catalytic complex and sequestering it within the streptavidin cavity has created ArM with novel non-natural catalytic functionality (36). However, this strategy suffers from the same labor intensive methodology as does the UAA-based bio-conjugation of catalysts within protein hosts. In addition, it is specific to the mutational and spatial restrictions of the streptavidin pocket. UAA-based installment of catalytic moieties would be possible in any given protein scaffold with the introduction a single TAG codon mutation.

1.3.2 Quantifying protein complex stability via ^{19}F NMR UAA probes

Non-natural chemical functionality contributed by UAAs has furthered the study of protein-protein interactions. Specifically, light activated photocrosslinking UAAs have been extensively exploited to probe protein interfaces in a site-specific manner (27). This technology has accelerated the identification of weak protein-protein interactions under physiological conditions (20). However, by nature of the covalent, irreversible, entrapment of protein binding partners, equilibrium thermodynamic quantification of the interaction is not possible. Therefore, other strategies must be undertaken to determine how manipulation of the environment under which the interaction occurs effects the stability of the complex.

In the final chapter in this dissertation, we employ a ^{19}F-containing UAA, TFMF, to quantity the stability of a weak protein-protein interaction and a protein

under various conditions. This project adds to the small subset of ^{19}F NMR protein investigations that utilize the site specific *in vivo* incorporation of unnatural amino acids (30, 31, 37-40). To date, the favored method to install ^{19}F probes is the residue specific manner exploiting native translational machinery (29, 41-43). However, replacing all locations of a residue (typically tryptophan or tyrosine) with fluorinated derivatives can destabilize the protein structure and complicate assigning resonances, especially for large proteins with many labeling sites. These impediments can limit ideal proteins for investigation to those with single sites for fluorine labeling and those that tolerate structural perturbations. In addition, protein expression must be done under deprived conditions or in auxotrophic strains to reduce canonical amino acid incorporation. Utilizing of the site-specific incorporation of UAAs provides a singly labeled protein expressed in rich media in standard heterologous hosts.

CHAPTER 2: GENETICALLY ENCODING ORGANO-CATALYTIC UNNATURAL AMINO ACIDS

2.1 Introduction

In nature, enzymes catalyze the synthesis of complex, enantiopure compounds with impressive rate enhancements (k_{cat}/k_{non}), highlighted by the highest observed k_{cat}/k_{non} reaching ~ 10^{21} (44). Moreover, enzymes operate under mild conditions, i.e. in water at physiological temperature, pressure and pH. Consequently, enzymes are increasingly being applied to the production of agrochemicals, pharmaceuticals, and fine and bulk chemical commodities (45). While enzymes are attractive alternatives and/or complements to synthetic catalysts, they are inherently limited by their chemical composition. This restricts the synthetic applications of enzymes to chemical transformations observed in nature. Expanding the catalytic behavior of enzymes beyond the ability of the canonical amino acids, native cofactors and bioavailable substrates is a contemporary goal.

Several tools for installing non-natural activity into enzymes have emerged. These include incorporating transition metals or metal complexes into proteins (46), *de novo* design using computational methods (47) and exploitation of enzyme's inherent promiscuity (that is, the ability of an active site to carry out a mechanistically similar reaction to its native chemistry) (48). Introducing non-natural transition metal based chemistry into enzymatic scaffolds represents one of the earliest and most prolific fields of biocatalyst engineering (49). The chiral secondary coordination

sphere provided by the biomolecular scaffold can be tuned to enhance the selectivity and activity of the metal-based chemistry. However, transition metals often require unfavorable reaction conditions such as the need for degassed buffer, absence of light and inert atmospheres (50-52). The *de novo* design of proteins with non-natural capabilities is appealing but computational efforts, to date, have yielded catalyst with catalytic efficiencies that are orders of magnitude lower than native enzymes (53). These activities may reflect current limitations in computational methods, which are unable to account for long-range, cooperative interactions and the complex and dynamic nature of protein fluctuations (47). Frances Arnold and colleagues engineer non-natural enzymatic catalysts by exploiting the promiscuity of metal cofactors in natural enzymes. Heme-enzymes including cytochrome P450s, cytochrome C and globins, have all been evolved to catalyze diverse reactions including, olefin cyclopropanation, C-H amination, sulfimidation and N-H insertions through mechanistic intermediates similar to the active oxidant of cytochrome P450 (54, 55). However, this strategy is limited to reactions that are mechanistically similar to existing enzyme catalysis.

Genetic code expansion beyond the canonical 20 amino acids using orthogonal aminoacyl-tRNA synthetase-tRNA (aaRS-tRNA) pairs represents an attractive method to install non-natural reactivity (4). This method is particularly compelling since unnatural amino acids (UAAs) can be incorporated into proteins via the native translational machinery, obviating the need for extensive post-translational manipulation of the protein of interest. In addition, >150 novel amino acid structures

have been incorporated into proteins via this approach, Figure 1.2, attesting to the diverse chemical structure that can be translated by the ribosome (2).

While the chemical functionality made available by UAA technology continues to grow, its applications for enantioselective catalysis remain confined to a select few approaches. With respect to biocatalysis, the use of UAAs has been markedly under-investigated. UAAs have predominately been employed as biorthogonal chemical handles for the covalent attachment of metal-based catalysts within a protein scaffold. This approach has created novel iron proteins that carry out the Fenton reaction (56) and Rh_2 catalysts competent for non-natural carbenoid insertion reactions including cyclopropanation of olefins (33). However, bulky conjugation handles place steric constraints on the choice of scaffold, limiting options. In addition, the methodology necessitates synthesis of both catalyst and UAA equipped with conjugation partners. Lastly, the protein must be purified and characterized from the heterologous host, followed by conjugation and subsequent purification and characterization.

We propose the creation of proteins capable of non-natural catalysis through the *in vivo* incorporation of catalytic unnatural amino acids. This represents an untapped application of the genetic code expansion technology. To date, there is only one report of *in vivo* incorporated UAA-based enantioselective catalysis. Located in the hydrophobic cavity of a protein dimer interface, the metal binding amino acid, BpyAla, bound to copper (II) catalyzed the Friedel-Crafts acylation, although with poor conversion (22%) and modest selectivity (66% e.e.), Figure 1.9

(32). Although modular exploration of chiral and steric aspects of the surrounding protein scaffold enhanced the UAA-based activity.

The majority of non-natural enzyme catalysis centers on transition metal-based chemistries. Herein, we draw inspiration from the marriage of metallo- and biocatalysis and seek to unite bio- and organocatalysis. Since the year 2000, organocatalysis, the use of small organic molecules to catalyze reactions has exploded in the synthetic community. Sparked by two back-to-back publications by Barbas and MacMilllan, the field of organocatalysis has grown rapidly (57, 58). Organocatalysts are typically cheap, available in enantiopure forms, operate under mild conditions and are less toxic than metal catalysts (59). Through the work of groups such as MacMillan, Barbas, Corey, Jørgensen and List, organocatalysis has been established as an invaluable tool, complementary to transition metal based approaches, for the enantioselective construction of molecules (59-67).

Therefore, we propose the site-specific *in vivo* incorporation of organocatalytic amino acids into protein scaffolds for the evolution of novel enzymes with the potential for stereo-selective catalysis. Enzymatic active sites provide an environment for engineered selectivity and activity while, in a complementary fashion, the organocatalyst motif should foster broadly applicable modes of action for various chemical transformations in a manner that is dependent on the UAA side chain.

Among organocatalysts, secondary amines are perhaps the most widely utilized for enantioselective synthesis (68, 69). The versatility of secondary amine catalysis stems from the capability and complementary of two activation modes,

enamine and iminium ion formation, Figure 2.1. This dual activation can be used for nucleophilic (enamine) or electrophilic (iminium ion) activation of carbonyl containing compounds.

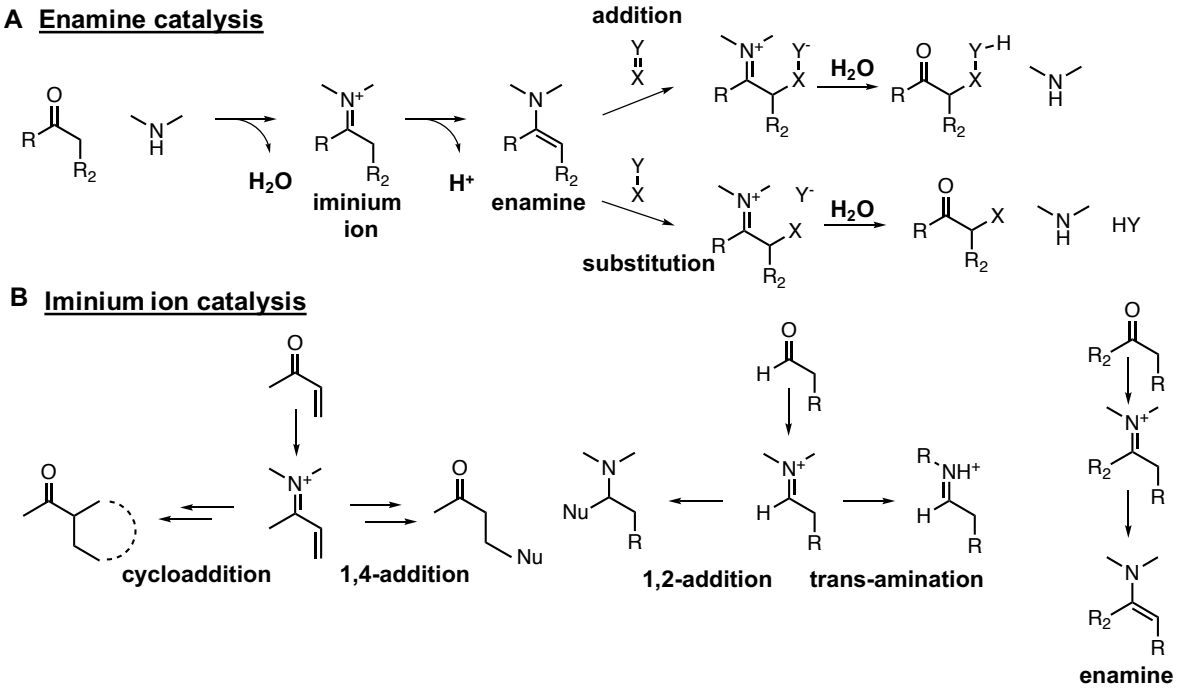


Figure 2.1 Enamine and iminium ion catalysis by secondary amines

(A) enamine mode of activation and potential nucleophilic reactions. (B) Iminium ion activation mode and potential electrophilic transformations.

Representative of this chemistry, the amino acid proline has served as a model and highly utilized secondary amine organocatalyst, Figure 2.2. Since the initial documentation of the proline catalyzed intermolecular aldol reaction, proline has been applied to many enamine-type reactions (57). In addition, proline has been utilized as a Lewis acid catalyst through iminium ion formation and employed in various cycloadditions as well as conjugate addition reactions functionalizing α, β unsaturated compounds. Furthermore, cyclopropanation and epoxidation are possible under a tandem mechanism in which the enamine formed after addition to

the iminium ion is trapped by an electrophilic species (68, 69). We hypothesize that the genetic incorporation of proline containing amino acids will enable the construction of evolvable protein catalysts for enantioselective enamine/iminium ion based via mechanisms that are comparable and complementary to reactions catalyzed by simple small molecule proline derivatives.

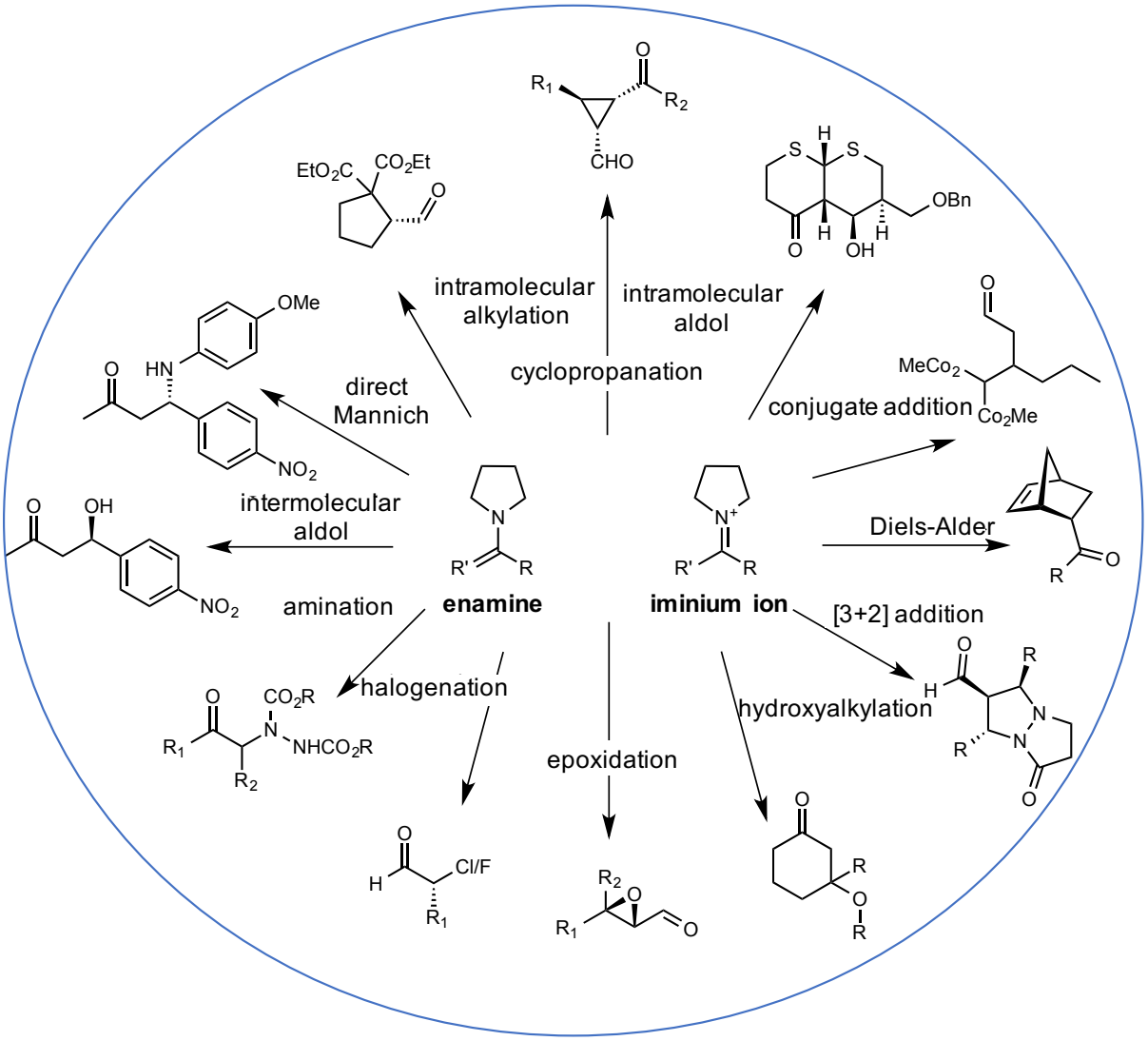


Figure 2.2 Versatility of proline as enamine and iminium ion organocatalyst

2.2 Results and discussion

2.2.1 Unnatural amino acid design and synthesis

We began our efforts with the design and synthesis of two unnatural amino acids equipped with a proline moiety. The structures of the amino acid were designed to target both sets of commonly used orthogonal aaRS/tRNA pairs: the *Methanococcua jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS)-tRNA and the pyrrolysyl-tRNA synthetase (PylRS)-tRNA from *Methanosarcina barkeri*, by creating a tyrosine and pyrrolysine based unnatural amino acid.

For the orthogonal *Mj*TryRS-tRNA pair, many of the first reports of UAA incorporation centered on para-substituted phenyl ring derivatives for use in bioconjugation (24, 70-73). In addition, bulky para-substituents such as tetrazine (74), tetrazole (75), and benzophenone (21) groups and conjugated aromatic systems such as 2-naphthol (76) have been incorporated by engineered *Mj*TyRS variants. Inspired by the successful engineering of *Mj*TyRS for the *in vivo* incorporation of diverse para-substituted UAAs, we designed proline-4-amino-L-phenylalanine (Pro4AF). The UAA provides a rigid scaffold to display the active proline moiety which may prove beneficial for enantioselective catalysis in addition to the potential for π interactions with substrates or active site side chains.

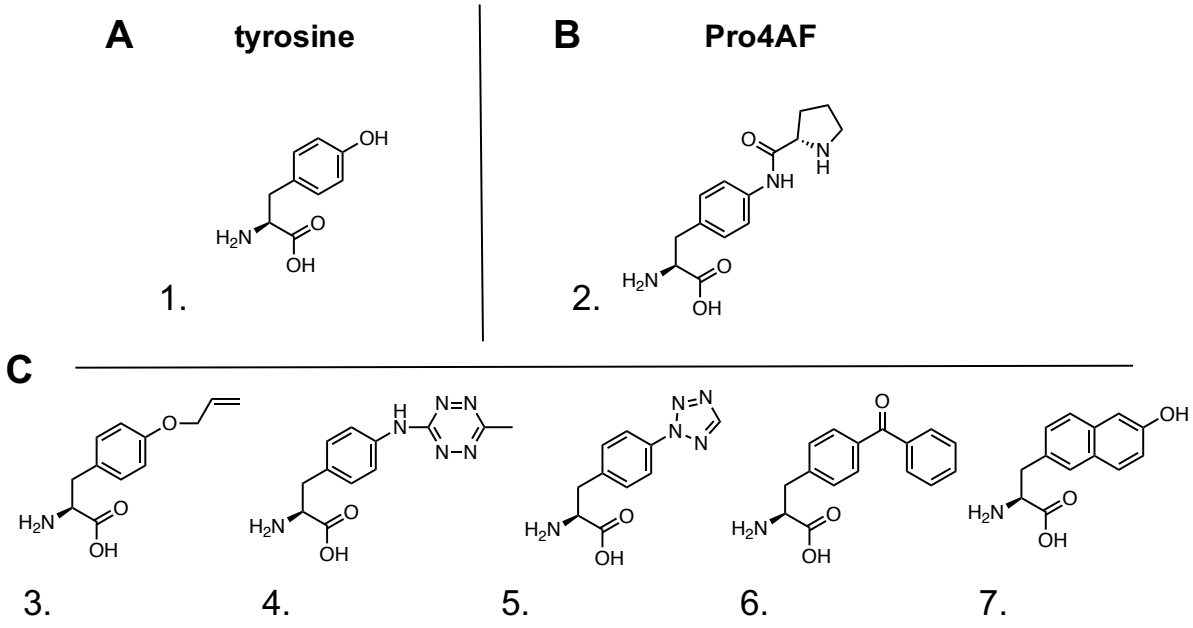


Figure 2.3 Substrates of engineered *Mj*TyRS

(A) The native substrate of *Mj*TyRS. (B) The proline containing tyrosine derived UAA utilized in this study. (C) Diverse UAAs *in vivo* incorporated by the orthogonal *Mj*TyRS/tRNA pair.

Codon reassignment in *E.coli* utilizing the orthogonal aaRS/tRNA machinery pyrrolysine (the so-called 22nd canonical amino acid), has been very successful due to the remarkably flexible substrate scope demonstrated by the wild type pyrrolysyl tRNA synthetase (PylRS) and its variants. The earliest investigation into the structural requirements for PylRS substrate recognition identified several proficient substrate analogs (2, 5, 6 in Figure 2.4) (77) followed by numerous and growing reports of the diverse chemical structures accommodated by PylRS and its variants, see Figure 1.2. Related to our efforts, this study reported the incorporation of a proline-modified pyrrolysine analog, D-prolyl-L-lysine (ProK), albeit with poor efficiency (77). Based on the known flexibility of PylRS, we also chose ProK as a starting point for our efforts.

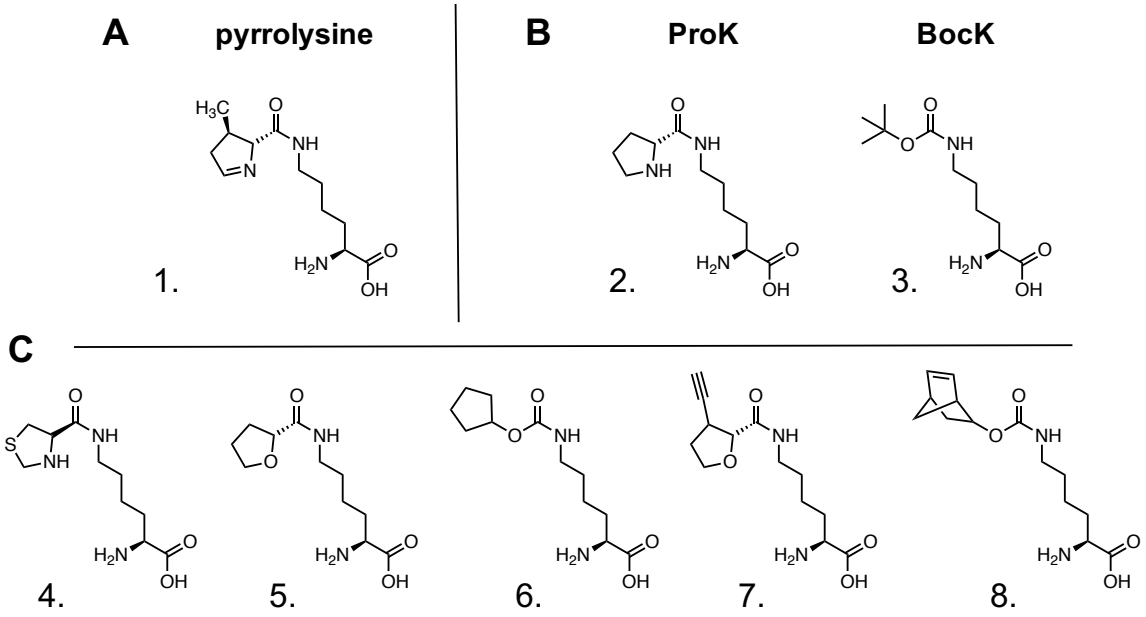


Figure 2.4 Substrates of the wild type pyrrolysine synthetase (WT-PylRS)

(A) The 22nd amino acid, pyrrolysine. (B) UAAs utilized in this study: ProK, the proline containing UAA and BocK, a positive control for WT-PylRS. (C) Diverse UAAs recognized by WT-PylRS that are structurally and chemical similar to proK.

The synthesis of Pro4AF and ProK are described in detail in Small molecule synthesis and characterization 2.5.3. Briefly, to construct Pro4AF, a 9-BBN protecting group strategy was employed followed by routine amine to amide dicylcohexylcarboiimide coupling. The 9-BBN protection of 4-amino-L-phenylalanine and subsequent coupling to Boc-D-proline proceeded in an overall 53% yield. The synthesis of ProK was modified from original reports in which yields were not conducive to scaled preparation. Our synthetic strategy first activated Boc-D-proline as a N-hydroxysuccinimide carbonate followed by coupling to N-α-Boc-protected L-lysine followed by dual deprotection. The three-step synthesis was carried out in an overall yield of 64%. Unnatural amino acid synthesis was verified by mass spectrometry and ^{1}H NMR analysis, see 2.5.4^{1}H NMR.

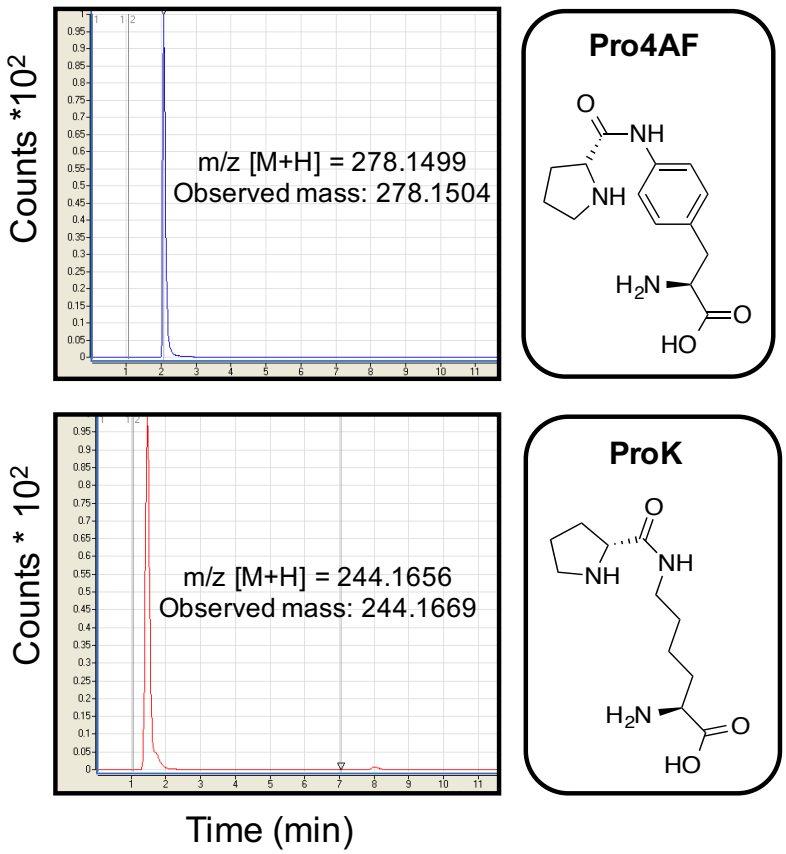


Figure 2.5 Proline-containing UAAs: Pro4F and ProK

2.2.2 The search for a Pro4AF synthetase

Initial efforts focused on the *in vivo* incorporation of Pro4AF. This work represents the first report of a proline containing *Mj*TyRS-targeted UAA. Therefore, a synthetase must be identified or evolved to recognize and incorporate Pro4AF into protein sequences. Our approach was to first screen a collection of synthetases available to us, in house, for successful incorporation of Pro4AF in response to the amber TAG codon. Variants of the *Mj*TyrRS and *Mb*PylRS were assessed given both systems have been engineered to accept *para*-substituted phenylalanine amino acids (4, 78). We selected a span of tyrosine-based synthetases that demonstrate broad substrate tolerance encompassing bulky hydrophobic *para*-substituents as

well as small polar substituents. Our screen also included alanine mutations at amino acid positions N246 and C248 in the *Mb*PylRS system, which have been shown to induce broad tolerance to phenylalanine derivatives, including bulky para-substituted derivatives (78), in PylRS. The complete list of synthetases evaluated is given in Table 2.1.

We utilized a green fluorescent protein (GFP) construct with a TAG codon at position 149 to screen potential tRNA synthetases. The detailed procedure is described in the methods ("GFP-based synthetase screen for Pro4AF incorporation"). Briefly, each individual tRNA synthetases was co-transformed with a GFP-149TAG construct. Expression was then carried out in the presence of Pro4AF. If the TAG codon was suppressed with the UAA, full length GFP would be produced and the extent of expression was qualitatively determined by fluorescence intensity of clarified lysates. Expression was simultaneously carried out in the absence of Pro4AF to judge the level of expression resulting from TAG suppression with canonical amino acids. Synthetases were deemed successful if fluorescence intensity was greater in the presence of Pro4AF than in the absence. Unfortunately, no synthetase was identified that met this qualification. As an alternative, it may be possible to construct and assess large, but non-trivial, libraries of *Mj*TyrRS variants for ProAF incorporation. However, with a second amino acid in hand and based on the known, broadened substrate specificity of PylRS, we chose to instead direct our efforts toward ProK incorporation. Further evolution of *Mj*TyrRS for ProAF remains a possible future target for this research.

Table 2.1 Aminoacyl-tRNA synthetases evaluated for Pro4AF incorporation

***Mj*TyrRS variants**	***Mb*PylRS variants**
p-acetyl-phenylalanine	pyrrolysine
p-azido-phenylalanine	pyrrolysine L274A C313A Y349F
p-benzoylphenylalanine	pyrrolysine L274A C313S Y349F
p-(2,2'-bipyridin-5-yl)alanine	pyrrolysine C313A Y349F
p-boronophenylalanine	pyrrolysine C313S Y349F
p-cyano-phenylalanine	pyrrolysine Y349F
p-carboxylmethyl-phenyalanine	pyrrolysine 246A 248A
p-(7-hydroxycoumarin-4-yl) ethylglycine	pyrrolysine 246A 248L
p-(8-hydroxy-quinolin-3-yl)alanine	pyrrolysine 246V 248W
p-iodo-phenylalanine	pyrrolysine 246S 248W
p-isopropyl-phenylalanine	pyrrolysine 246T 248T
p-hydroxy-L-phenyllactic acid	pyrrolysine HRS
p-proparglyoxyphenylalanine	
p-nitrophenylalnine	
3-(2-naphthyl)alanine	
o-nitrobenzyl-tyrosine	
sulfotyrosine	
o-methyl-tyrosine	
p-amino-phenylalanine	
tyrosine	

2.2.3 *Mb*PylRS accepts ProK

In light of unsuccessful attempts to incorporate a tyrosine-based proline UAA, *in vivo*, we turned our attention to the pyrrolysine based UAA, ProK. This UAA has been previously identified as a substrate analogue of the pyrrolysyl-tRNA synthetase from *Methanosarcina barkeri* (*Mb*PylRS), albeit with a reported 18% suppression efficiency (77). This observation was qualitatively verified in our hands using the GFP-based fluorescence reporting system. Herein, we report fluorescence intensity with respect to 1 mM BocK (an efficient substrate of *Mb*PylRS known to be incorporated at high yields and a standard positive control in our lab). Fluorescence intensity of clarified cellular lysate after expression of GFP-TAG in the presence of 2 mM ProK was found to be 9-fold less than in the presence of 1 mM BocK but seven

times greater than in the absence of ProK. A 2-fold improvement was observed with the addition of 5 mM ProK to the media, Figure 2.6, however further increases in amino acid concentration did not provided additional improvements in incorporation. GFP obtained from expressions in the presence of ProK was purified from cellular lysate and subjected to analysis via mass spectrometry, which revealed the successful incorporation of intact ProK at position 149 (Figure 2.6, expected mass, m/z = 27938.6, observed mass m/z = 27938.6).

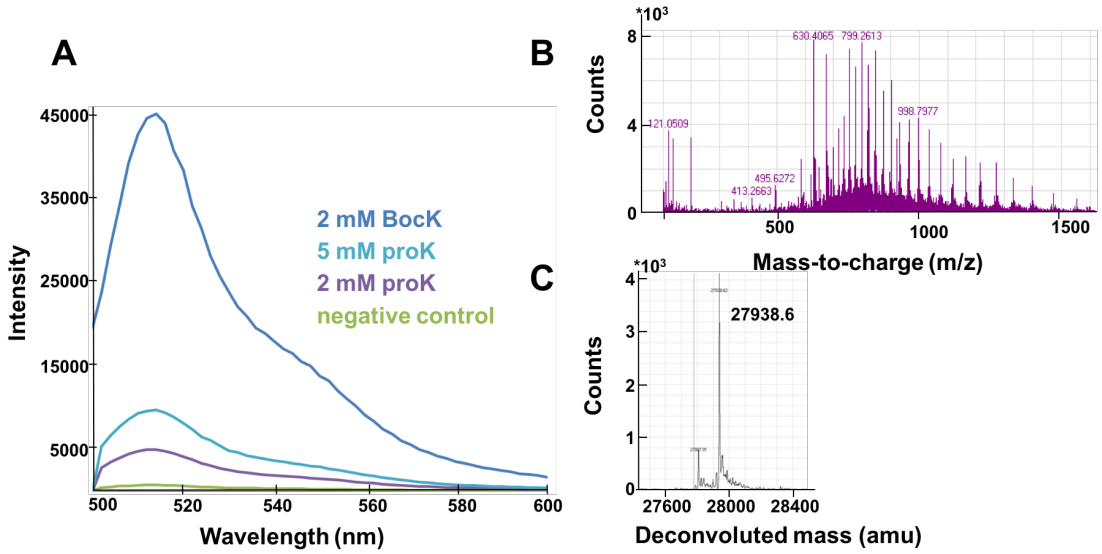


Figure 2.6 Incorporation of ProK into GFP by *Mb*PylRS

(A) GFP fluorescence intensity of clarified lysate after expression in the absence, negative control, and presence of UAAs. (B,C) Mass spectrometry analysis of GFP149ProK, expected mass, m/z = 27938.6, observed mass m/z = 27938.6. Truncation corresponding to cleavage of starting methionine residue is observed.

2.2.4 Free ProK aldol catalysis

After verification of incorporation of ProK into a protein sequence in response to TAG codon with *Mb*PylRS, we next sought to characterize the catalytic competencies of the UAA in the absence of protein. Given directed evolution can only improve upon activity and not create it, it is preferential that the unnatural amino acid be a capable catalyst prior to incorporation into a protein host. An

intermolecular aldol reaction, a hallmark reaction for amine catalyst development, was chosen to evaluate enamine reactivity of ProK. This reaction was selected because it is amenable to high-throughput screening as demonstrated by its utilization in the computational design of aldolase enzymes (53). In addition, there are several commercially available or synthetically straightforward retro-aldol and aldol chromogenic/fluorescence probes with divergent structures and sizes that may facilitate future screening and directed evolution (53, 79, 80). We anticipate protein hosts will require sterically compatible substrates such that diverse and commercially available substrates provide ideal model systems. In addition, after installation of the UAA in a protein host, improvements in catalysis will be necessary by mutational exploration of the protein host. Reactions that exhibits a characteristic absorbance or fluorescence signature provide convenient screens for assessing improvements in catalytic competency among protein variants.

To begin, acetone and 4-nitrobenzaldehyde were selected as aldol substrates. Product standards were synthesized consistent with previous reports (81). HPLC and comparison to an internal standard [trans-beta nitro styrene (TBNS)] was used to quantity conversion via peak integration, Figure 2.7. The optimal reaction buffer was chosen after screening HEPES, phosphate and Tris HCl. 200 mM potassium phosphate provided the largest conversion with minimal background reaction in the absence of catalyst, Figure 2.8. The reaction proceeded with 10% yield in the presence of 30 mol% ProK in 200 mM potassium phosphate buffer, pH 7.4. Notably, ProK proved to be more active as a catalyst when compared to proline under these reaction conditions. Proline is an efficient aldol catalyst for this particular

reaction, however the reaction is typically performed in organic solvents, such as DMSO (57). We suggest the improved catalysis by free ProK may be the result of a phase separation where catalysis is occurring in a concentrated organic layer whose formation is more favorable with ProK than proline (82, 83).

To summarize our initial findings, we have designed an UAA that is incorporated by the WT *Mb*PylRS *in vivo*, in response to a TAG codon in a protein sequence. This UAA is capable of aqueous aldol chemistry in a model reaction that has served as a means to screen protein catalysts in a high throughput fashion. Further, we have developed an analytical HPLC methodology to evaluate the catalytic conversion in the model reaction. The next step is to improve upon the incorporation efficacy of *Mb*PylRS for ProK. Traditional organocatalytic reaction necessitate high catalyst loading (10-30 mol%) and we anticipate similar catalyst loading in our initial experiments evaluating aldol chemistry of ProK installed in a protein host.

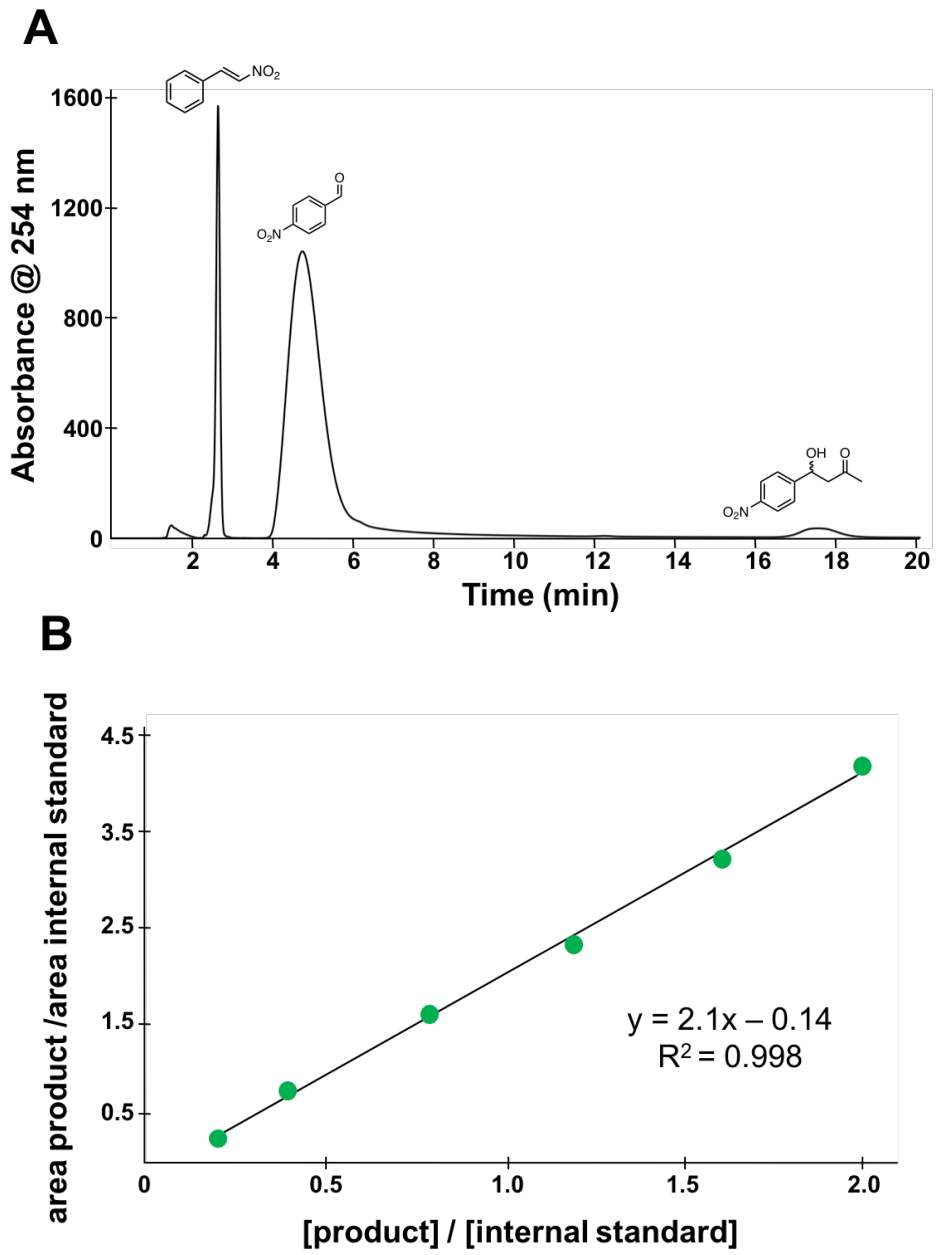


Figure 2.7 Quantification of model aldol reaction

(A) HPLC trace: an isocratic solvent system of 95:5 hexanes:ethanol was used at a flow rate of 1 ml/min. Detection was monitored at 254 nm. Enantiomeric excess was not determined. (B) aldol product standard curve: serial dilutions of a standard were performed. Internal standard at 5 mM was added to each sample and 5 µL was injected into Agilent LC 1260 Infinity II system.

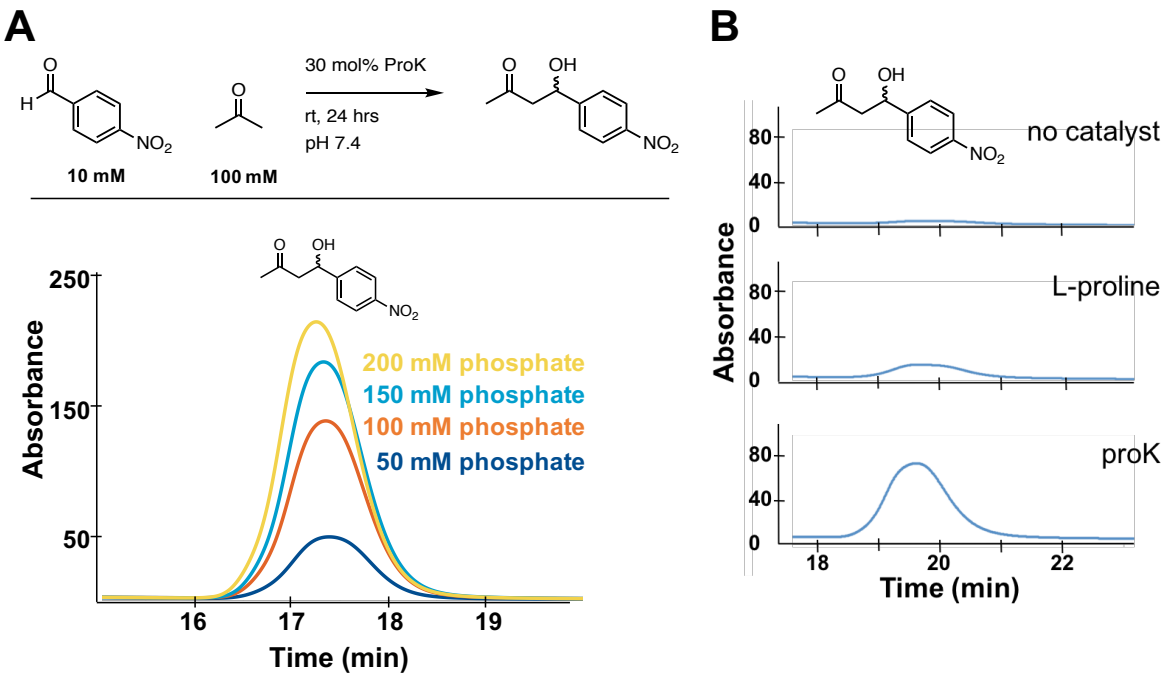


Figure 2.8 ProK catalyzed model aldol reaction

(A) Model aldol reaction with corresponding HPLC trace of aldol product obtained at various concentrations of phosphate buffer. The optimized reaction conditions were determined to be 30 mol% ProK, 200 mM phosphate, 400 mM Na^+, 10 mM aldehyde, 100 mM acetone, pH 7.4. (B) Comparison of 30 mol% proline and ProK catalyzed aldol reaction in the optimized buffered conditions.

2.2.5 Optimization of ProK incorporation by *Mb*PylRS

Previously, we determined that WT-*Mb*PylRS incorporated ProK into GFP, in response to a TAG codon, roughly 10-fold less efficient than BocK, a positive UAA control. For future catalytic evaluation of ProK-containing proteins, however, we anticipate high catalyst loadings will be necessary, reflecting standard conditions for organocatalysic reactions in the literature. Accordingly, we next focused our efforts on the improvement of *Mb*PylRS-mediated ProK incorporation. We first performed a series of experiments aimed at exploring vector-based effect on incorporation, in addition to improvements in ProK incorporation via mutation and evolution of the synthetase.

To begin, a comparative analysis between commonly utilized expressions systems was performed. A pET/pUltra and pBK/pBad dual plasmid systems have been developed and frequently used in our laboratory for unnatural amino acid mutagenesis (84, 85). These systems differ in location of tRNA and tRNA synthetase, as well as promoters for expression of all system components and the origins of replication. Notably, the fundamental principles of these two systems as it pertains to UAA expression are the same, however, in our experience, different UAA-containing proteins as well as different UAAs have shown disparate performances depending on the plasmid pair.

We compared a pET-GFP-2TAG/pUltra-*Mb*pylRS-*Mm*pylT system with a pBK-*Mb*pylRS/pBad-sfGFP149TAG.his-pylT system. By fluorescence intensity, expression of GFP was 2-fold greater with the pBK/pBad plasmid system than pET/pUltra. In addition, background incorporation of canonical amino acid incorporation in the absence of the unnatural amino acid was greater with the pET/pUltra system when compared to the pBK/pBad system, Figure 2.9. Moving forward with our optimized vector expression system for this particular protein and UAA, we transitioned to identifying improvements in incorporation efficiency by reaching out to mutations to the active site of *Mb*PylRS that have been shown to improve the incorporation of similar UAAs. One mutation, Y349F, is known to improve amino acylation rates in PylRS (86). The Y349F mutation improved GFP expression in the presence of ProK by an additional 2-fold as judged by fluorescence intensity. Therefore, pBK-*Mb*PylRS-Y349F served as our working plasmid for incorporating ProK and the target for further evolution efforts.

Quantification of ProK incorporation in GFP as determined by mass of purified protein obtained in the presence of BocK and ProK with *Mb*PylRS-Y349F demonstrated 12% of the incorporation of the positive control with BocK.

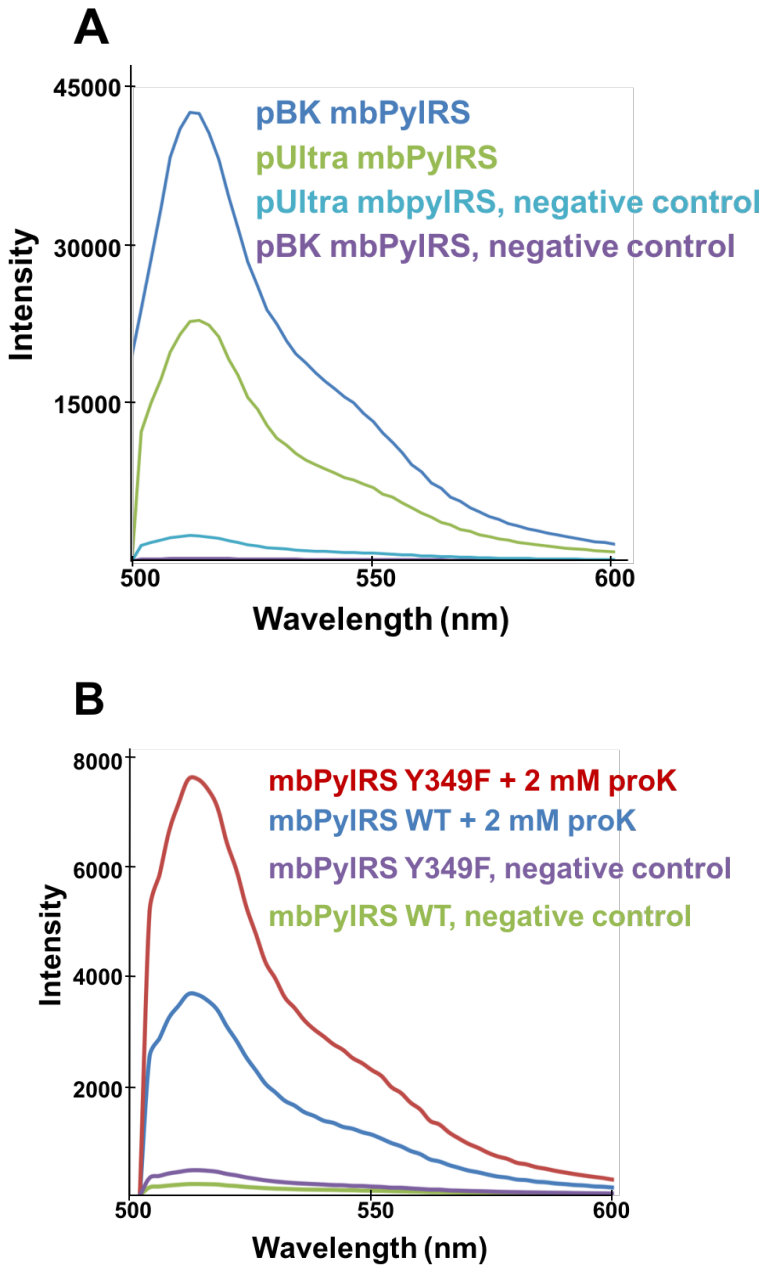


Figure 2.9 Optimization of UAA incorporation system for ProK

(A) GFP fluorescence intensity after incorporation in the presence and absence, negative control, of 2 mM ProK. (B) Improved florescence intensity observed with the mutation of tyrosine 349 to phenylalanine in *Mb*PylRS, red curve, as compared to *Mb*PylRS WT, blue curve.

2.2.6 Efforts towards the directed evolution of *Mb*PylRS

Our previous efforts to increase incorporation efficiency of ProK by *Mb*PylRS, which focused on the logistics of plasmid construction as well as a single known beneficial mutations, resulted in modest improvements. However, a single mutation, Y349F, provided improvements in ProK incorporation. This result suggested the possibility that other mutations to the PylRS scaffold may provide enhancement in incorporation efficiency. Therefore, we chose to carry out the direct evolution of the PylRS synthetase in an active site-directed manner. We employed a four site-saturation "NNK" randomized library of *Mb*PylRS that was created by a previous lab member, Joshua Gober. The sites chosen (see Figure 2.10) have all been shown to improve the incorporation of various N-ε-derivatized lysine UAAs (78). "NNK" codons (N = A,T,G,or C; K = T, or G), encoding all possible 20 canonical amino acids, were cloned onto the activating Y349F mutation to create pBK-*Mb*PylRS-Y349F/L305NNK/Y306NNK/L309NNK/C348NNK, Figure 2.10.

The library of mutant *Mb*PylRS variants was selected for the ability to specifically suppress a TAG codon with ProK even in the presence of canonical amino acids through iterative rounds of positive and negative selection. For positive selections, we made use of a plasmid (pREP-pylT, Figure 2.10) harboring an antibiotic resistance marker, the chloramphenicol acyltransferase gene, embedded with a TAG codon (87). Cells were evaluated for growth in the presence of ProK and varying concentrations of chloramphenicol (80-200 μg/ml, depending on stringency of selection). Variants that survived the positive selection were then subjected to negative selection using a toxic protein encoding a TAG codon (pBad-Barnase-TAG,

Figure 2.10) (88). In this selection, cells were grown in the absence of ProK and expression of the toxic protein was induced using arabinose. Cells that survive the negative selection should be specific for ProK since the synthetase did not suppress the TAG codon with canonical amino acids. The detailed protocol for the selection is described in Positive/negative selection system 2.4.8. Multiple rounds of increasing chloramphenicol stringency followed by screening of 200 isolated colonies for improvements in GFP-based fluorescence resulted in no improved variants when compared to the parent, *Mb*PylRS-Y349F. A reduced diversity library of Y271, L274 and L270 NNK was carried through the same procedure and no improvements were identified.

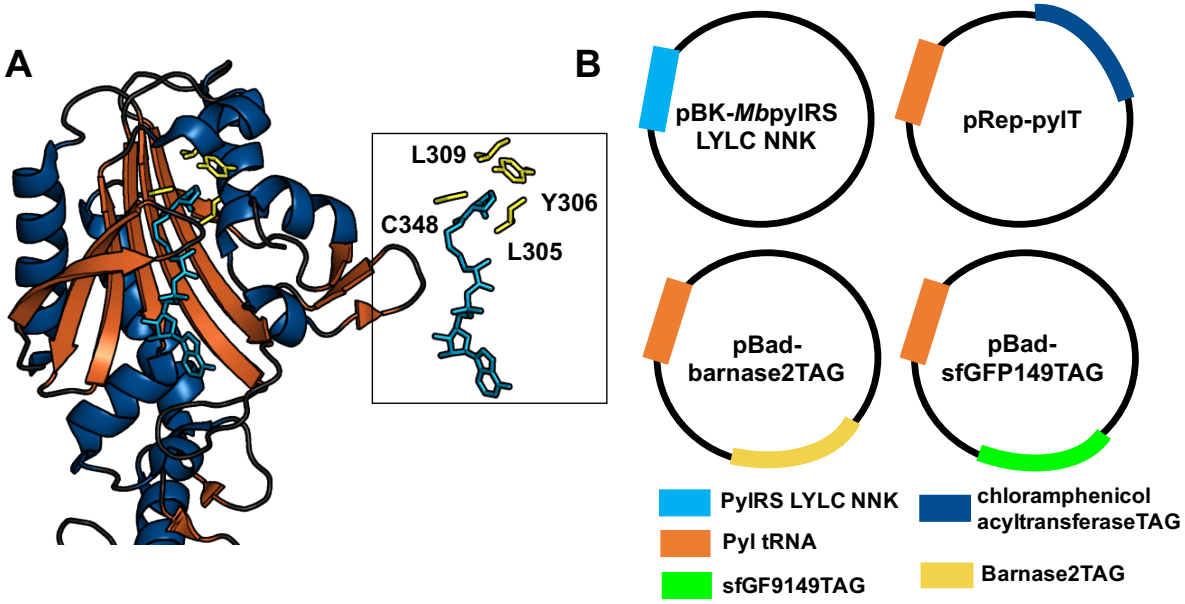


Figure 2.10 Direction evolution of *Mb*PylRS via a 4 site-saturation 'NNK' library

(A) *Mb*PylRS, PDB ID: 2ZIM, with adenylated pyrrolysine shown in blue and residues targeted for 'NNK' codons shown in yellow. (B) Plasmids used to select for improvements in ProK incorporation.

Failure to identify beneficial mutations for the improved incorporation of ProK from an active site-directed site-saturation library of *Mb*PylRS, prompted us to explore alternative mutation strategies to target the entire protein scaffold. Error

prone (EP) polymerase chain reaction (PCR) is an established means to introduce random nucleotide mutations in a gene sequence. We created EP libraries of the *Mb*PyLRS-Y349F gene. The detailed EP PCR methodology is described in Error prone *Mb*pylRS library and FAC sorting 2.4.9. Briefly, PCR reactions were performed with variations in the concentration of manganese chloride to decrease the fidelity of DNA polymerase in the presence of a thymine and cytosine biased nucleotide pool, to limit innate biases in nucleotide substitution patterns. The mutation rate of EP PCR reactions can be controlled by varying the $MnCl_2$ concentration. We found, libraries of *Mb*PylRS-Y349F could be obtained containing low mutation rates (1-2 per gene) or high mutation rates (up to 8 mutations per gene). We examined two error prone libraries bearing either low mutation rates (~2 per gene) to maximize the number of active synthetases in the library, or a high mutation rate (~6 per gene) to maximize mutational diversity. These libraries were co-transformed with pBad-sfGFP149TAG.his-pylT, to enable in cell screening of ProK incorporation via florescence assisted cell sorting, as described in Error prone *Mb*pylRS library and FAC sorting.

As an alternative to antibiotic selections, fluorescence assisted cell sorting (FACS) based on GFP fluorescence was employed to identify mutants with improved ProK incorporation. Selections often enable large diversity libraries to be screened in parallel, but are hampered by increased instances of false positives. In addition, stringencies in selections may demonstrate a limited dynamic range if mutants are only slightly more active than wild type synthetases. Screens, such as FACS, provide a direct readout of cellular activity and hence quantitative PylRS

activity. We first validated our FACS screening methodology by successfully discerning cells harboring pBK-*Mb*PylRS-Y349F from cells with pBK-*Mb*PYlRS-WT, Figure 2.11, after GFP expression in the presence of ProK. Subsequently, the error prone libraries of pBK-*Mb*PylRS-Y349F with high and low mutations per gene were co-transformed with pBad-sfGFP149TAG.his-pylT and subjected to florescence assisted cell sorting after expression overnight in the presence of 2 mM ProK. Unfortunately, evaluating 10^4 individual cells for GFP-based fluorescence did not identify further improvement in ProK incorporation, Figure 2.11.

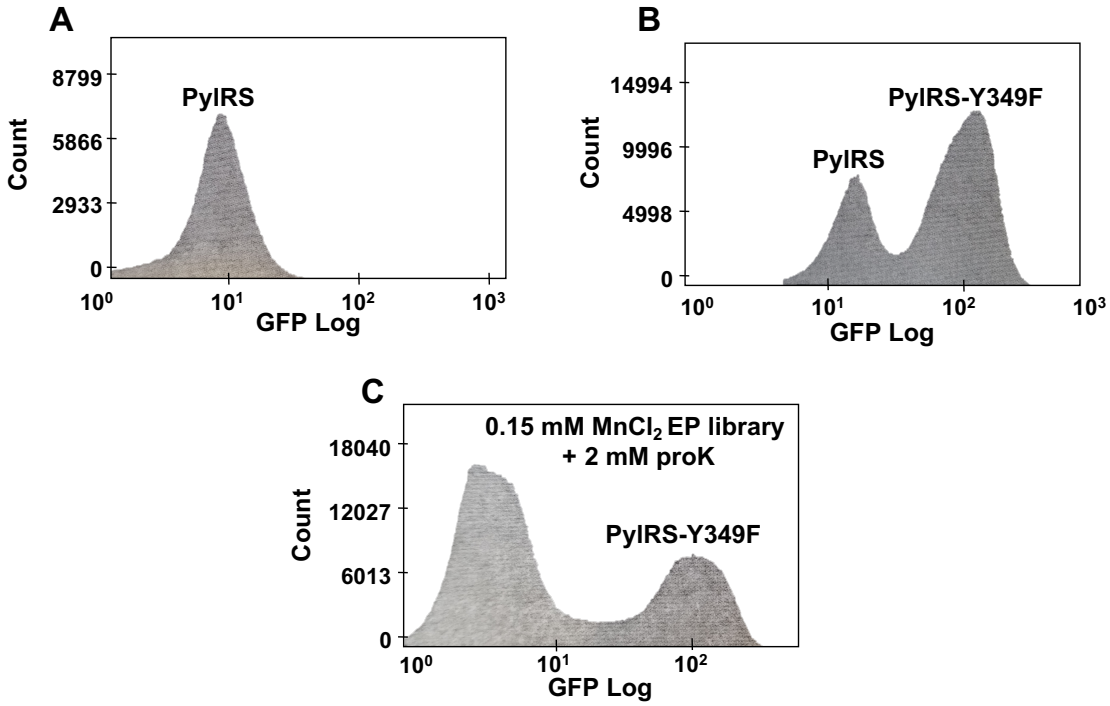


Figure 2.11 FAC sorting of error prone library fails to identify improvements

(A) Population of cells expressing GFP with the WT-*Mb*PylRS. (B) Mixture of cells expressing GFP with WT-*Mb*PylRS and *Mb*PylRS-Y349F. (C) Population of cells containing 6 mutations per *Mb*PylRS-Y349F gene created from error prone PCR with 0.15 mM $MnCl_2$. No improvements to *Mb*PylRS-Y349F were identified.

Efforts to evolve a synthetase for ProK with improved incorporation efficiency above *Mb*PylRS-Y349F were unsuccessful. We hypothesize that amino acid

degradation and/or inefficient cellular uptake are hindering incorporation yields. Experiments to access these and other possible limiting cause of ProK incorporation are underway. Despite the limited yields provided by *Mb*PylRS-Y349F, this synthetase remains the most efficient system for ProK incorporation. Accordingly, we moved forward with effort to identify ProK-mediate catalysis in proteins using *Mb*PylRS-Y349F for incorporation.

2.2.7 Protein host selection and catalytic evaluation

The next step towards the creation of an enzyme equipped with a catalytic amino acid is to identify potential biomolecular scaffolds. The scaffold chosen should meet several criteria: (1) An ideal scaffold will be highly stable. Evolution can often lead to reduced protein stability due to deleterious mutations. We therefore focused on thermophilic proteins which operate at temperatures between 50 and 112 °C. Starting with thermophilic proteins increases the mutation potential without compromising stability. In addition, we focused our scaffold choices to proteins that are known to (2) express well in *E. coli* and (3) that contain large enough pockets to spatially accommodate the unnatural amino acid. Finally, we refined our search criteria to include (4) proteins that have previously been used as scaffolds for engineering since they are often well characterized and robust to mutagenesis.

Scaffolds were chosen based on previous successes in the engineering of metalloenzymes and computationally designed enzymes. Scaffolds with large hydrophobic cores such as the α/β barrels and jellyrolls are popular choices (36, 46). Three enzymes that adopt an α/β barrel fold, 2-deoxyribose-5-phophate aldolase (DERA), indole-3-glycerol phosphate synthase (IGPS) and a synthase subunit of the

glutaminase synthase enzyme complex (tHisF) were selected. Mechanistically, DERA represents a privileged scaffold in aldolase chemistry. It is the only aldolase to catalyze the reaction between two aldehyde substrates. IGPS is an interesting scaffold that contains a wide opening on one side of the α/β barrel, termed the catalytic face. This face has been used in the design of a copper-binding site which was used in a catalytic Diels-Alder cycloaddition in the related imidazole-3-glycerol phosphate synthase. The authors were able to make seven mutations to the scaffold including four mutations of histidine to alanine without any negative effects in expression and folding (89). In the Baker lab, IGPS has been computationally remodeled as a retro-aldolase and DERA was redesign as a Kemp eliminase (53, 90). Both non-natural activities were improved upon using directed evolution (91). tHisF has been demonstrated as a robust host for the creation of artificial metalloenzymes and a suitable host for bulky organocatalysts. It is also very tolerant to mutations and simply purified from endogenous proteins by heat denaturation (35, 92). The genes encoding these proteins were obtained from LifeTechnologies as GeneArt Strings DNA Fragments and cloned into the pET21C(+) using the NdeI and XhoI restriction enzyme sites. The genes were amplified with primers containing NcoI and PmeI and subsequently digested and ligated into pBad_pylT backbone, see Primer Sequences 2.5.2.

Scaled preparation of the scaffold proteins was carried out followed by evaluation of aldol catalysis in the model reaction between acetone and 4-nitrobenzaldehyde. The protein host for the catalytic unnatural amino acid should not perform the reaction on its own. Both IGPS and DERA displayed significant product

formation in the reaction as judged by HPLC analysis of the reaction in the presence of 5 mol% protein, 10 mM aldehyde and 100 mM acetone in phosphate buffer, Figure 2.12. Therefore, ProK engineering efforts with these two scaffolds will require prior mutagenesis to knock out this background catalysis. Wild type tHisF did not catalyze the reaction above background in the absence of catalyst and could be moved forward to ProK incorporation directly.

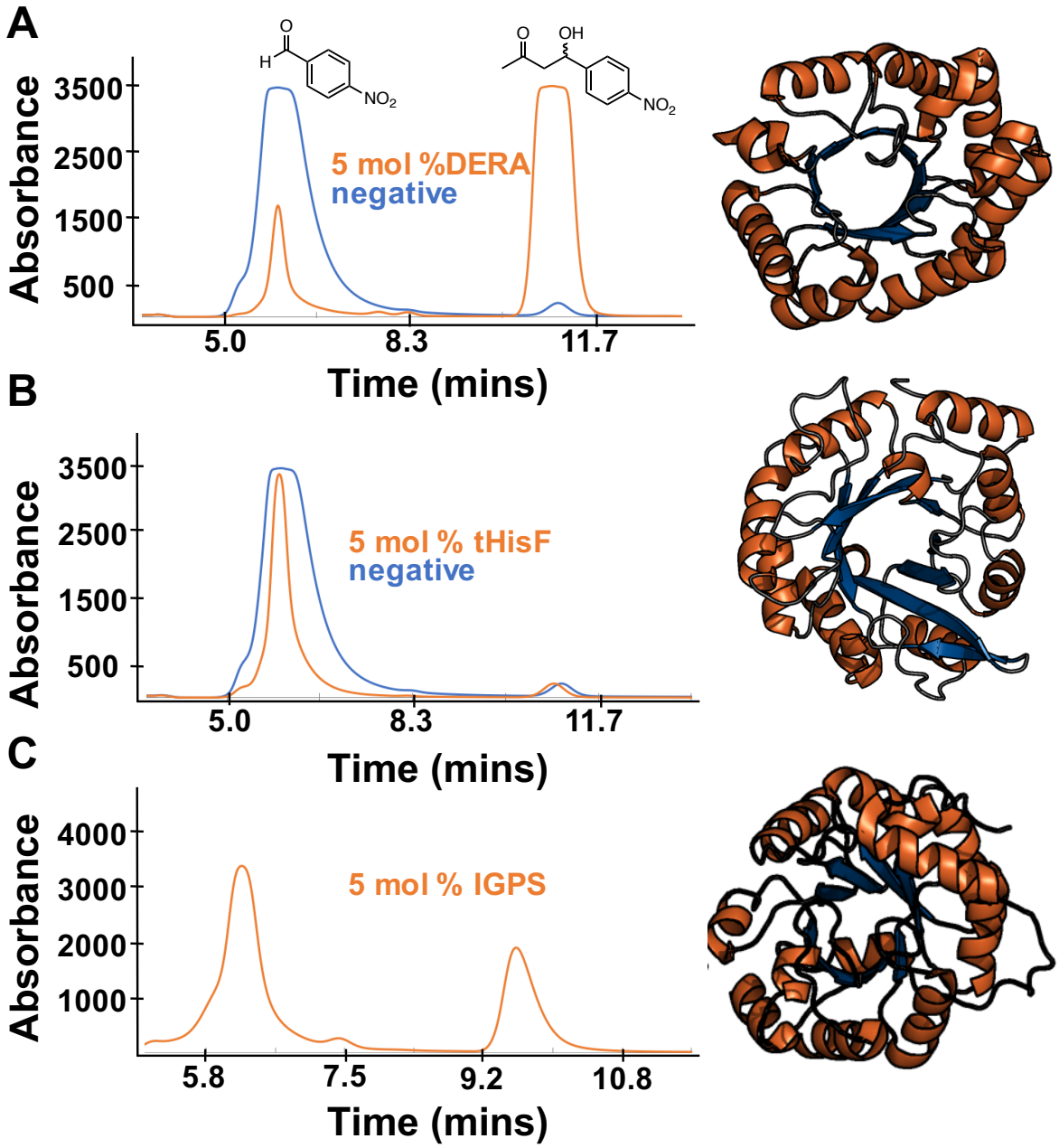


Figure 2.12 Evaluation of aldol activity in wild type scaffold proteins

HPLC traces of the reaction between 4-nitrobenzaldehyde and acetone with 5 mol% scaffold protein in optimized buffered conditions (A) DERA, PDB ID: 2UB3, demonstrates near complete conversion of the model aldol reaction. (B) tHisF, PDB ID: 1THF, does not catalyze the model aldol reaction. (C) IGPS, PDB ID: 1CV4, is efficient at the model aldol reaction.

The aldolase scaffold, DERA, is known to catalyze aldol reactions between various ketones and aldehydes, yet this is first example of catalysis with a

substituted aromatic acceptor substrate (93). DERA has two active site lysine residues (K151 and K180) that are crucial for its wild type chemistry, the conversion of 2-deoxy-D-ribose 5-phosphate to D-glyceraldehyde 3-phosphate and acetaldehyde. This reaction proceeds via a organocatalysis mechanism that requires the nucleophilic attack of lysine 151 on the aldehyde of the substrate. The nucleophilicity of K151 is made possible by the pK_a perturbation in response to the second lysine residue, K180, which sits adjacent to K151. We hypothesized that these two lysines are the primary source for the background catalysis in our model reaction. To test this hypothesis, either the catalytic lysine or the pK_a perturbing lysine was individually mutated to alanine and both alanine mutations eliminated aldol chemistry between acetone and 4-nitrobenzaldehyde, Figure 2.13.

A similar rational targeted mutagenesis effort to eliminate promiscuous aldol catalysis by IGPS was not undertaken. The native IGPS reaction, conversion of 1-(o-carboxyphenylamino)-1-deoxyribulose 5-phosphate into indole-3-glycerol phosphate, is catalyzed by an intricate interplay of several residues and dependent on conformation loop changes (94). Based on the complexity of this system, multiple mutations may be required to completely attenuate background catalysis and identifying rational mutations is not straight forward. Therefore, this protein was removed from future work. Since tHisF shows no aldol catalysis and we have identified point mutants of DERA that are also clear of background chemistry; these two protein scaffolds were selected for subsequent ProK-catalysis experiments.

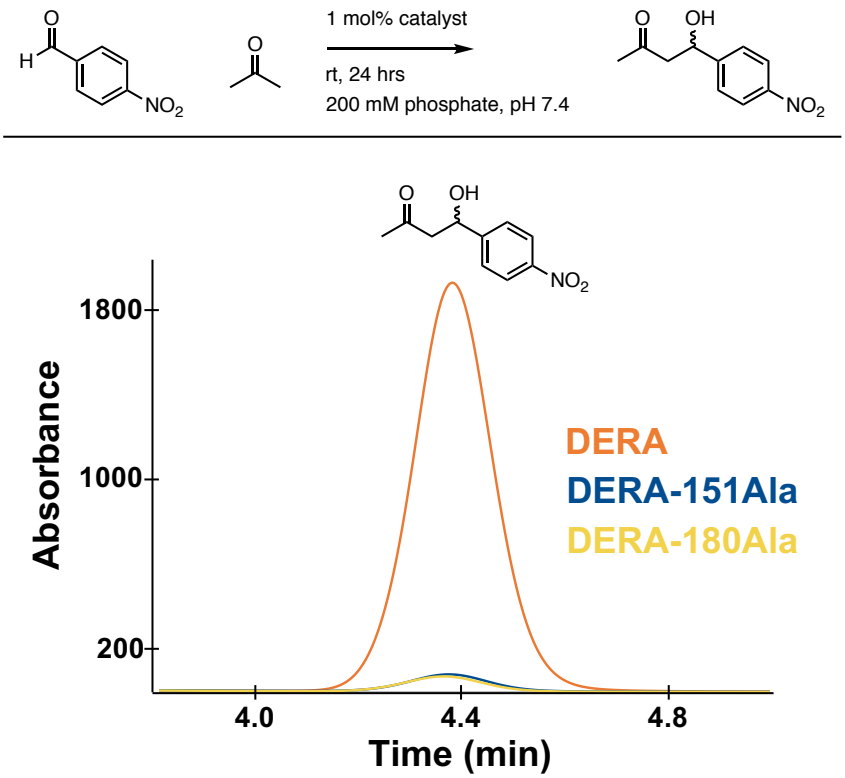


Figure 2.13 Elimination of WT-DERA aldol catalysis

2.2.8 Incorporation of ProK within protein scaffolds

We next sought to identify locations within the engineered DERA and tHisF to incorporate ProK, beginning with DERA as a target scaffold. In collaboration Dr. Chris Snow at Colorado State University, ProK was modeled in place of K151 or K180. Since ProK is a lysine analog, we hypothesized that these locations may be able to accept the non-natural lysine derivative (Figure 2.14). To further refine optimal location, the intermediate iminium ion as a result of acetone condensation on the amine was also modeled, Figure 2.14. This modeling identified two solutions with ProK installed at K180 that required 2 and 4 additional mutations to accommodate the increase bulk of the ProK side chain. These, in addition to K151TAG and K180TAG mutations were cloned and expressed on a small scale to evaluate UAA

incorporation. Protein expression was evaluated in the presence and absence of BocK and the clarified lysate was evaluated by SDS-PAGE, as described Small scale UAA incorporation 2.4.4. As shown in Figure 2.15, no location proved suitable for the generation of a ProK containing protein, as only small increases in DERA expression was obtained in the presence of amino acid when compared to no amino acid controls.

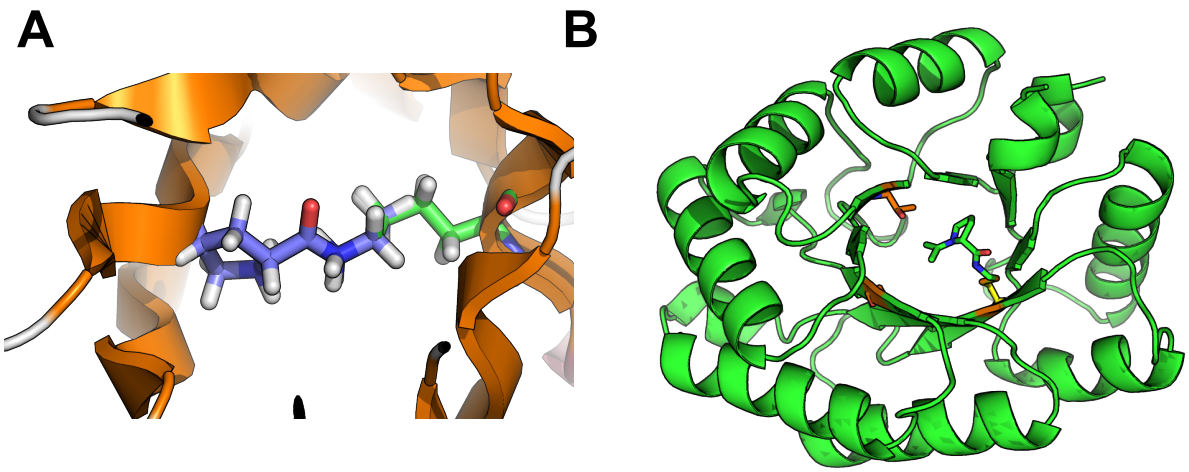


Figure 2.14 Computational selection of residues in DERA to host ProK

(A) Representation of proline amide modeled on the backbone of lysine in the active site of DERA, PDB ID: 1UB3. (B) Adduct after nucleophilic attack of ProK on acetone modeled in a hydrophobic pocket of the DERA scaffold. Alanine mutations to accommodate ProK are shown in orange.

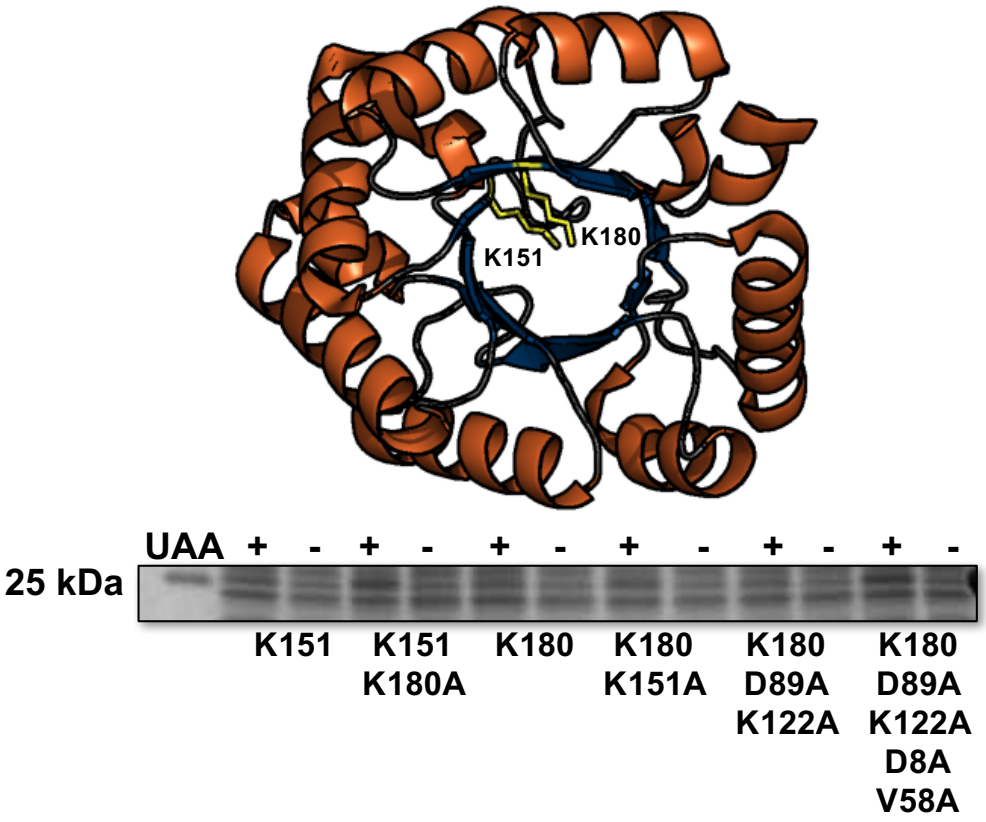


Figure 2.15 Searching locations within DERA to incorporate ProK

SDS-PAGE gel showing expression of DERA-TAG mutants in the presence and absence of 2 mM BocK. No location was found to host the UAA. Lysine residues tested for UAA incorporation are shown in yellow.

In light of the low expression yields obtained using DERA as a host protein, we next moved to tHisF, and examined this scaffold for locations suitable for ProK incorporation. Residues were chosen based on previous engineering efforts in addition to visual inspection of the protein crystal structure in PyMol. We sought to identify locations near a positively charged residue to mimic nature's strategy of lysine's pK_a reduction in aldolase enzymes. The locations were tested by SDS-PAGE analysis of clarified lysate from small-scale protein expressions in the presence and absence of BocK, a control amino acid that provides high yields and provides a diagnostic view of TAG suppression efficiency among a wide number of

variants. Of the TAG locations tested (L2, L50, T171, L169, L3, N22, N24, R27, I199, I7) L50, T78 and L169 resulted in significant production of BocK containing protein, Figure 2.16. Accordingly, we next examined incorporation of ProK at these three locations. While Bock provided robust yields at all three sites, scaled preparation of ProK-containing tHisF in LB media supplemented with 4 mM ProK yielded (at the best location, tHis-F50ProK), significantly lower yields (<0.5 mg of protein per 2-liter expression). Mass spectrometry was used to verify successful incorporation of ProK and BocK into tHisF at location 50, Figure 2.17. Auto-induction media using pET/pUltra system was also tested in an attempt to increase yields of tHisF50ProK. This plasmid system and media combination have been shown to improve protein yields for unnatural tyrosine analogs (95). However, previous experiments using this pET derived plasmid have shown increased background incorporation of canonical amino acid incorporation compared to the pBK/pBad system, Figure 2.9. This background incorporation was exacerbated when applied towards ProK in auto-induction media.

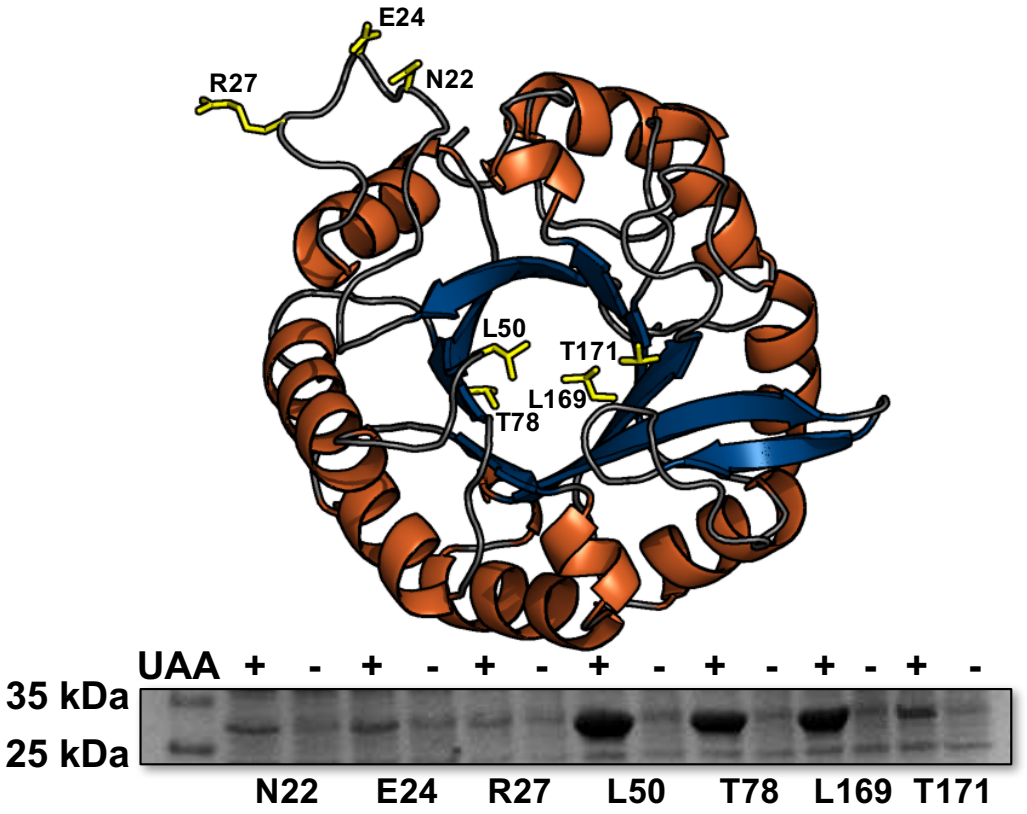


Figure 2.16 Searching tHisF for locations to install ProK

SDS-PAGE gel showing expression of tHisF dependent on the presence of 2 mM BocK at locations L50, T78 and L169. Residues chosen to host UAA are shown in yellow.

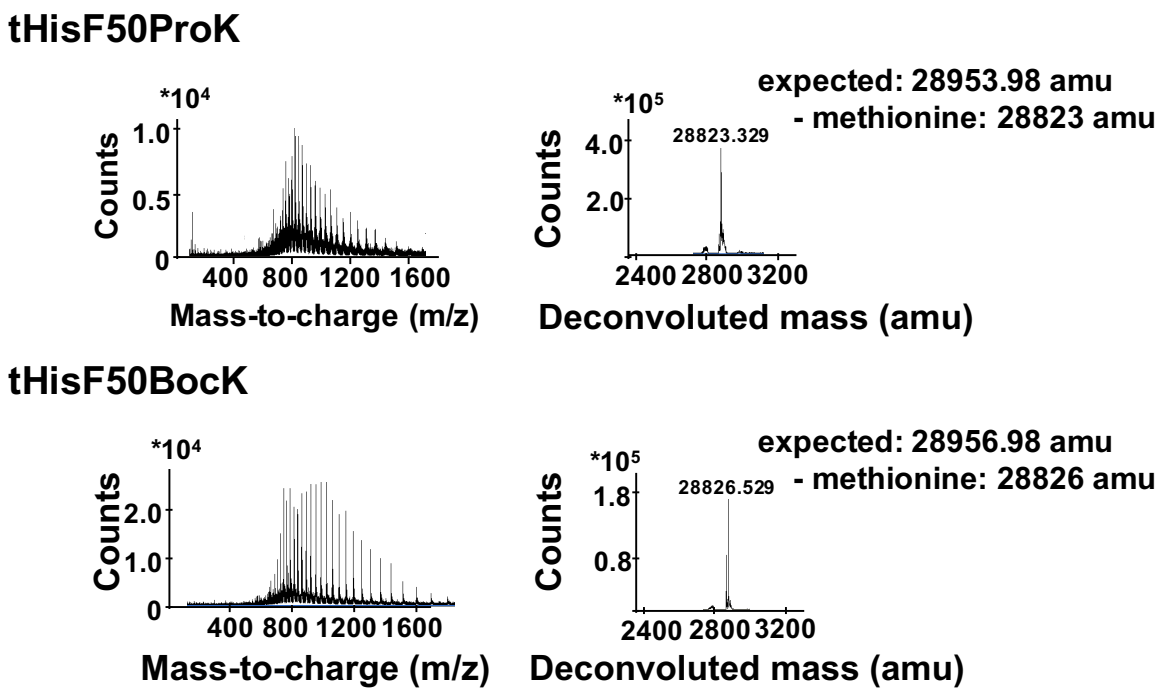


Figure 2.17 Mass spectrometry analysis of tHisF50ProK and tHisF50BocK

2.2.9 Evaluating the reactivity of a ProK containing protein scaffold

Despite its poor expression, tHisF50ProK was selected for further evaluation and we next sought to evaluate the nucleophilicity of the proline moiety within the hydrophobic cavity of tHisF. Enamine and/or iminium ion catalysis requires the amine of proline to be deprotonated in order to perform nucleophilic attack on the carbonyl of the electrophile in the reaction. A simple means to evaluate whether an amine is reactive as a nucleophile is reductive amination or imine trapping. In this experiment, the amine and electrophile are incubated in solution with a water-soluble reductant, in this case, sodium cyanoborohydride ($NaCNBH_3$). If condensation of the amine on the carbonyl occurs, the resulting iminium ion is reduced by $NaCNBH_3$ to form an amine, Figure 2.18. The result is a covalently modification that is observable by mass spectrometry. This method is used as a screen to identify potential

electrophilic substrates, such as screening various ketone and/or aldehyde containing compounds (96). It can also be used to evaluate the reactivity of the amine by screening different concentrations of the carbonyl compound and determining at which concentration the adduct is observed.

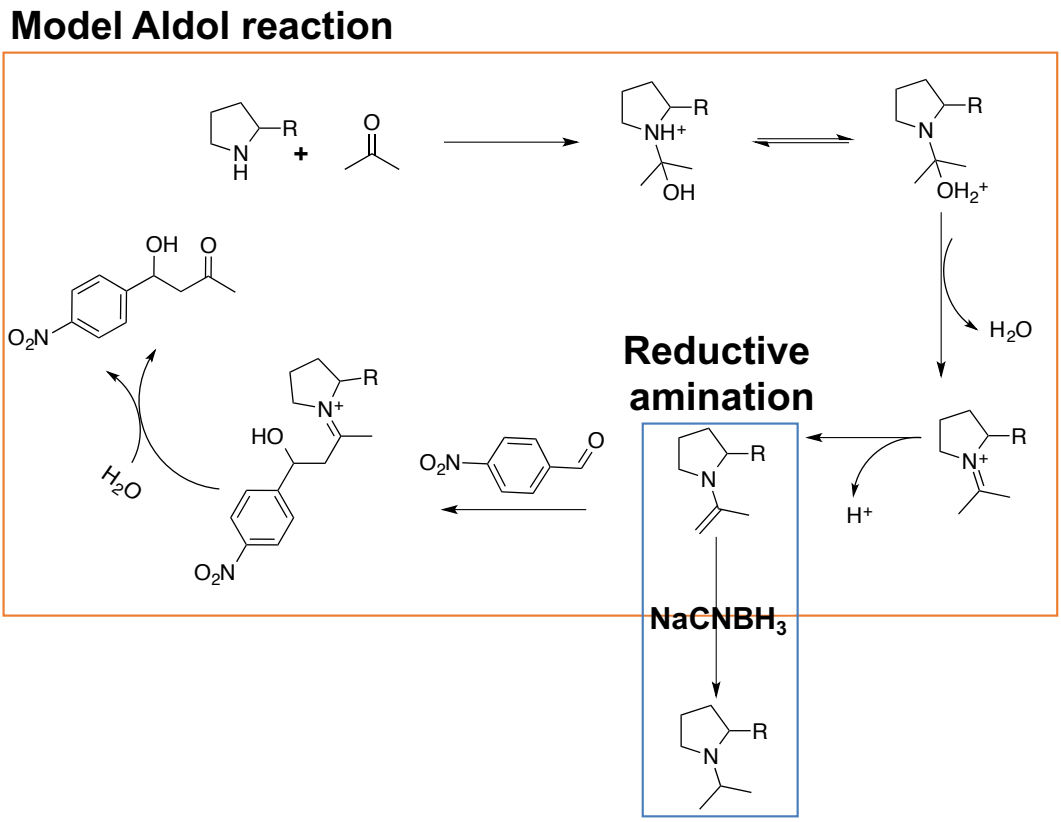


Figure 2.18 Mechanism of model aldol reaction and reductive amination.

Reductive amination was used to evaluate the potential for amine-based catalysis by tHisF50ProK. The amine of proline, if poised for nucleophilic attack, will form a covalent adduct with acetone indicated by a mass shift of 42 Daltons (Da). As positive controls, ProK as a free amino acid and the aldolase scaffold, DERA, were reacted with acetone and sodium cyanoborohydride. A detailed procedure is outlined in Reductive amination 2.4.11. The covalent adduct (42 Da) corresponding to reductive amination of acetone was observed for free ProK at 100 mM acetone and

at 1 mM acetone for DERA. However, no such adduct was observed for tHisF containing ProK in the presence of acetone (1, 10 and 100 mM) and reductant, Figure 2.19. This result implies that either the amine of ProK is protonated and therefore either solvent exposed or inclusion in the hydrophobic pocket does not drop the pK_a to a sufficient extent. Alternatively, the amine is deprotonated and acetone is unable to locate the reactive UAA, this outcome seems unlikely and is discussed further in the conclusions.

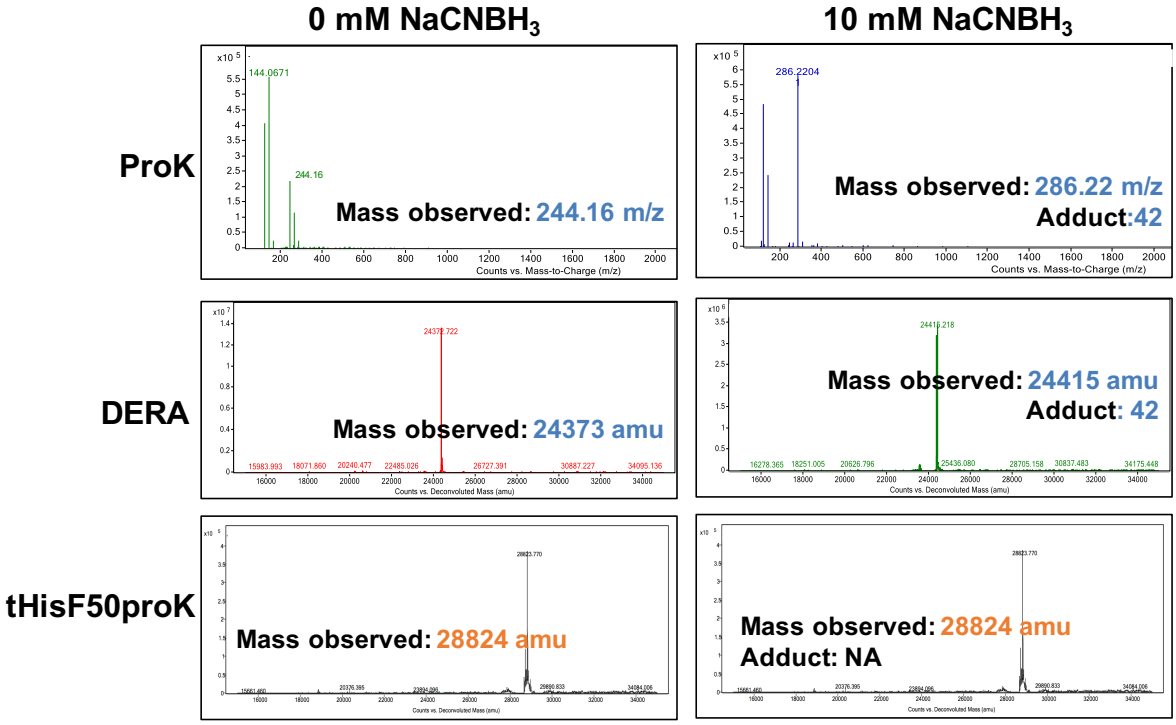


Figure 2.19 Reductive amination evaluates nucleophilicity of tHisF50ProK

Incubation of reactive amine in acetone with and without $NaCNBH_3$ demonstrates covalent modification of free ProK and DERA but not tHisF50ProK.

2.2.10 PyrrolK, the carbamate version of ProK

Thus far, the creation of a protein catalyst via the installation of a catalytic UAA has been limited by the yield of protein acquired due to poor incorporation efficacy of ProK by *Mb*PylRS. Efforts to improve incorporation efficiency via directed

evolution were unsuccessful. Therefore, we decided to alter the structure of the UAA in hopes to create a structure to host proline that is more amenable to the active site of *Mb*PylRS. It has been previously demonstrated that UAAs with carbamate linkages between lysine and an appendage are incorporated by PylRS with greater efficiency than the corresponding amide linkages (97, 98). This presents an opportunity to tweak the structure of our proline-lysine linked UAA.

We designed and synthesized, PyrrolK, a carbamate version of ProK. The UAA was designed to be structurally and chemically similar to ProK and was inspired from N-ε-cyclopentyloxycarbonyl-L-Lysine (UAA #6, Figure 2.4), a known substrate for the wild type MbPylRS. By taking this design approach, we hypothesized we could utilize the same 'NNK' and EP PCR *Mb*PylRS libraries that were created for ProK in the directed evolution of a synthetase for PyrrolK. In addition, knowing that *Mb*PylRS has stereo-specificity for D-prolyl-lysine over L-prolyl-lysine, we decided to synthesize both R and S enantiomerically pure forms of PyrrolK to evaluate the preference, if any, of the synthetase for enantiomers of PyrrolK. The synthetic route for this amino acid is described in detail in the Small molecule synthesis and characterization 2.5.3., section. Briefly, Boc protected enantiomerically pure R or S hydroxy-D-proline was activated with N,N'-disuccinimidyl carbonate and then coupled to N-α-boc-L-lysine. The amino acid synthesis was characterized by mass spectrometry.

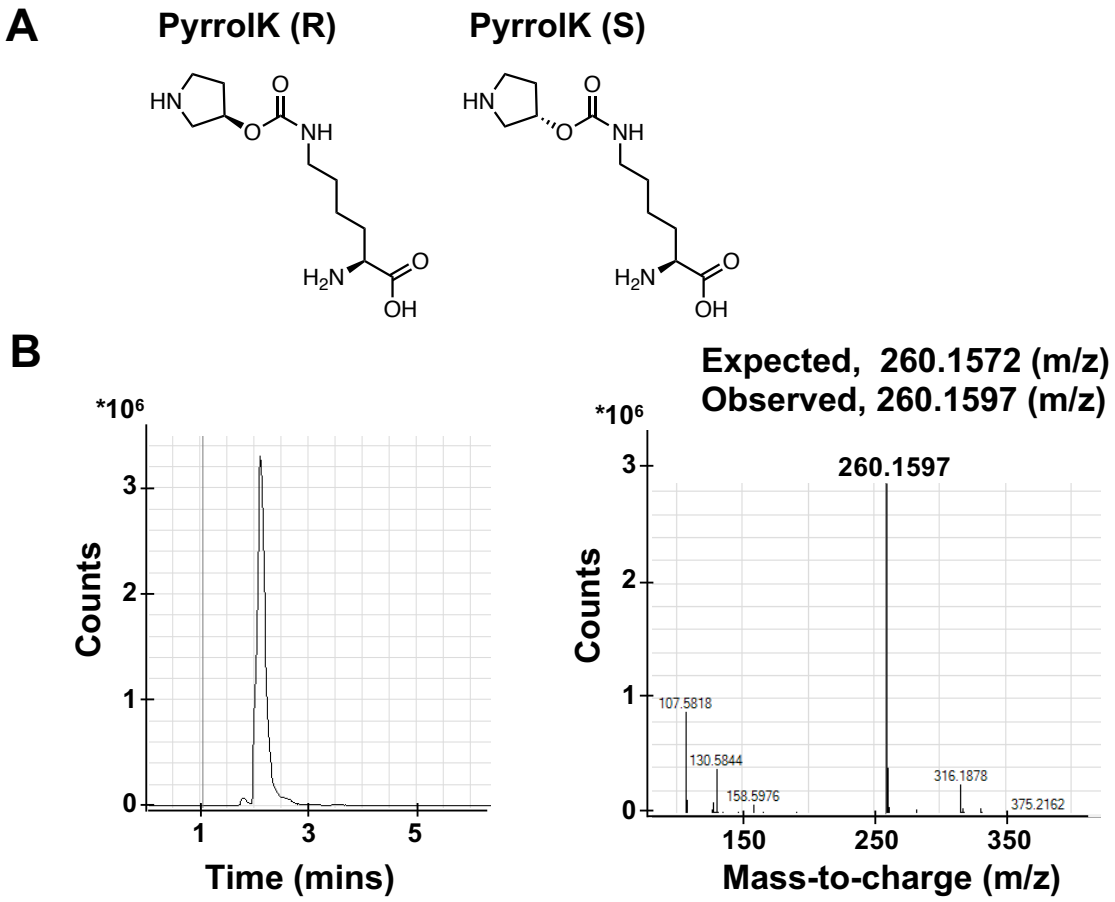


Figure 2.20 PyrrolK

(A) Structure of (R) and (S) PyrrolK. (B) Extracted ion chromatogram of mass corresponding to PyrrolK.

After construction of the two enantiomers of PyrrolK, we set out to test whether the UAA is a substrate for the WT-M*b*PylRS. Again, we employed a simple GFP-based experiment in which successful TAG suppression is indicated by full length gene expression and therefore, fluorescence, and the intensity gives a qualitative assessment of the efficiency. It was found that the UAA was weakly incorporated in both R and S forms by the WT-*Mb*PylRS at position 149 in GFP. Incorporation efficiency was determined to be 4 percent of the incorporation of the positive control with BocK.

Before tackling a directed evolution for PyrrolK, we first evaluated a few mutants of *Mb*PylRS as an initial investigation into whether the active site-directed 4 site 'NNK' library may provide improved incorporation. A small set of M*b*PylRS mutants, identified by previous lab member, Joshua Gober, that contain mutations favorable for incorporation of different carbamate UAAs were subjected to the GFP-based fluorescence analysis. A mutant, *Mb*PylRS-L274A/C313S/Y349F displayed greater fluorescence intensity than the WT-*Mb*PylRS in the presence of R and S PyrrolK, Figure 2.21A. This suggested that carrying out iterative rounds of positive and negative selections with the four-site 'NNK' library might provide additional improvements. The selection protocol, as described for ProK, was repeated in the presence of R and S PyrrolK, separately. Two rounds of positive selection with 80 µg/ml Cm followed by negative selection and screening 200 mutants for GFP florescence revealed a mutant with an improvement from WT-*Mb*PylRS for R PyrrolK, *Mb*PylRS-YL274A/C313V/Y349F. An additional round of positive selection with increased stringency at 200 µg/ml Cm did not provide further improvements. No further improvements were identified for S PyrrolK.

The incorporation efficiency of *Mb*PylRS-YL274A/C313V/Y349F, the mutant identified from directed evolution was quantitatively assessed. The mass of purified protein obtained after expression of GFP-TAG in the presence of 2 mM PyrrolK was compared to the mass of purified protein obtained from expression of GFP-TAG in the presence of Bock by the WT-*Mb*PylRS. Expressions were carried out on a preparative scale and full-length GFP was purified by His-tag affinity chromatography as described in Protein expression, purification and

characterization. It was determined that *Mb*PylRS-YL274A/C313V/Y349F produced an increase from 4 percent (WT-*Mb*PylRS) to 28 percent of the incorporation of the positive control of BocK, Figure 2.21B.

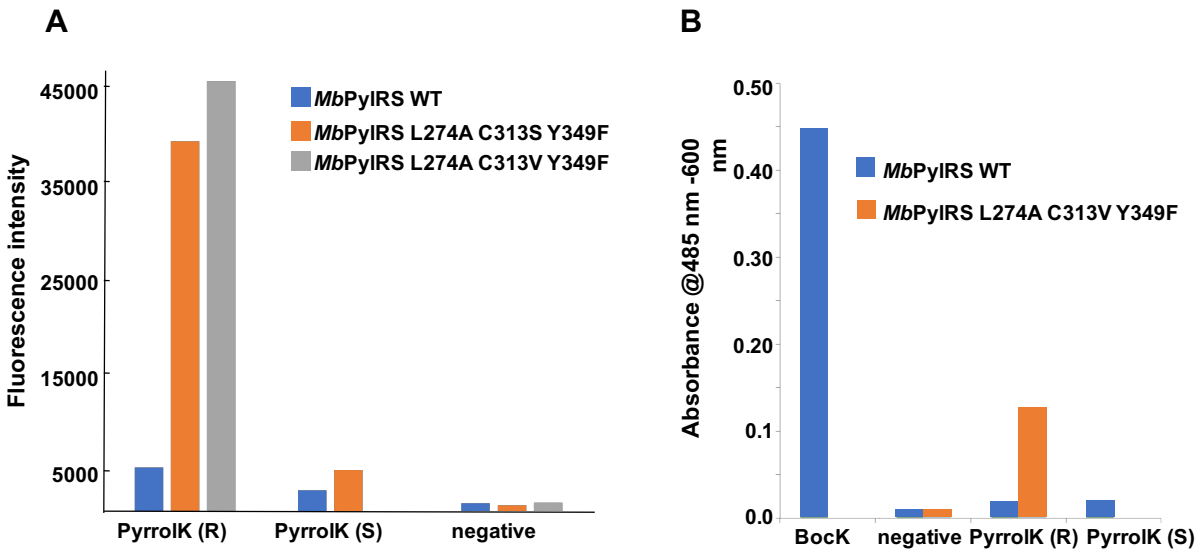


Figure 2.21 Evolution of *Mb*PylRS for PyrrolK

(A) Qualitative comparison of incorporation efficiency of PyrrolK by *Mb*PylRS variants. (B) Quantitative comparison of the incorporation efficiency of PyrrolK by the *Mb*PylRS variant identified from directed evolution to the incorporation of BocK by the WT-*Mb*PylRS.

As with ProK, creation of an optimized synthetase for incorporating the UAA into a protein sequence, *in vivo,* in response to TAG codons, called for the subsequent evaluation of the catalytic performance of PyrrolK. Therefore, PyrrolK was evaluated in the aldol model reaction between acetone and 4-nitrobenzaldehyde. The reaction conditions were optimized as before by screening buffers, HEPES, Tris HCl and phosphate at pH 7.4. Phosphate was found to be the preferred buffer resulting the highest yields with the smallest conversions in the absence of catalyst. PyrrolK was determined to catalyze the reaction to a similar extent as ProK, Figure 2.22.

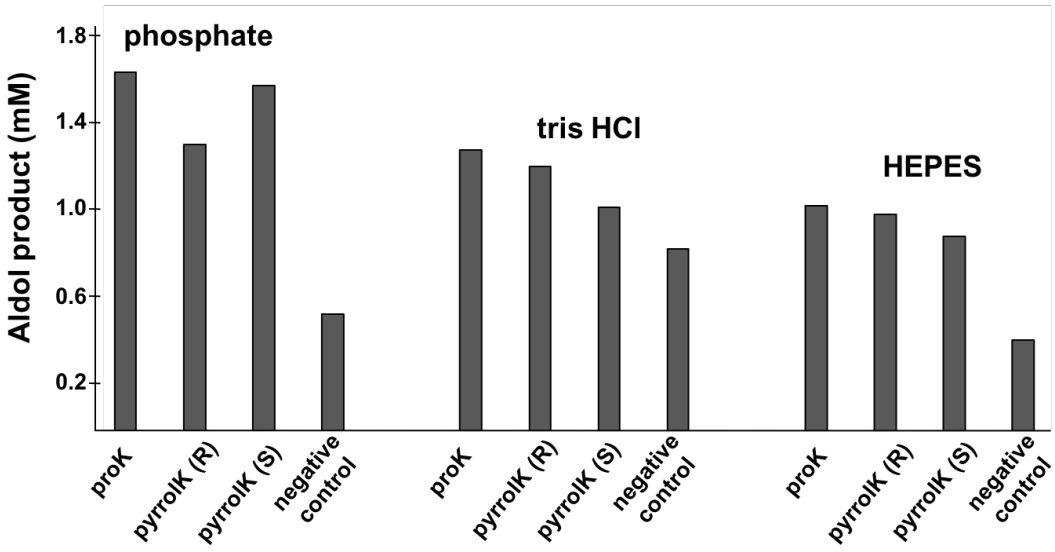


Figure 2.22 Buffer optimization of PyrrolK aldol reaction

The reaction conditions are as follows, 200 mM buffer, pH 7.4, 10 mM 4-nitrobenzaldehyde, 100 mM acetone, 10 mol% catalyst. aldol product concentration was determined via peak integration of HPLC chromatograph and method of internal standard using 5 mM TBNS.

Further efforts to press on with the installation of PyrrolK into protein scaffolds was impeded by complications in subsequent batches of PyrrolK that were synthesized. Incorporation by *Mb*PylRS-YL274A/C313V/Y349F was not reproducible with the second batch of synthesized PyrrolK and therefore efforts with PyrrolK as an organo-catalytic unnatural amino acid were halted. However, this UAA demonstrated the most promising incorporation efficiency, if repeated, by *Mb*PylRS of all three UAAs, Pro4AF, ProK and PyrrolK. Incorporation of PyrrolK into tHisF at locations found suitable for hosting ProK and evaluating the reactivity of the amine via reductive amination experiments are of future interest.

2.3 Conclusions

The production of enantioselective organocatalysts for enamine/iminium ion chemistry under aqueous conditions has been the focus of extensive research over the last 50 years (82, 99-102). Designing small molecules that can enrich for a single

enantiomer via the controlled orientation of substrates in a reaction's transition state, is not trivial. A commonality among successful catalysts is the proline or pyrolidine-catalytic centers (68, 81). If used as a single enantiomer, proline organizes the transition state of the reaction in a manner that transfers chirality to the formation of products (103). The transition state architecture of proline-catalyzed reactions was further promoted by the creation of more sterically complex and/or densely functionalized secondary- amine organocatalysts, Figure 2.23. These catalysts demonstrate improved enantioselectivities as compared to free proline (99). Enhanced catalysis was later attributed to the formation of emulsions and biphasic systems that sequester organic substrates and hydrophobic reactive centers within a concentration organic layer where the reaction occurs isolated from water (104), reminiscent of an enzymatic hydrophobic active site. However, even the best small molecule organocatalysts fail to produce synthetically useful enantio-inductions (104).

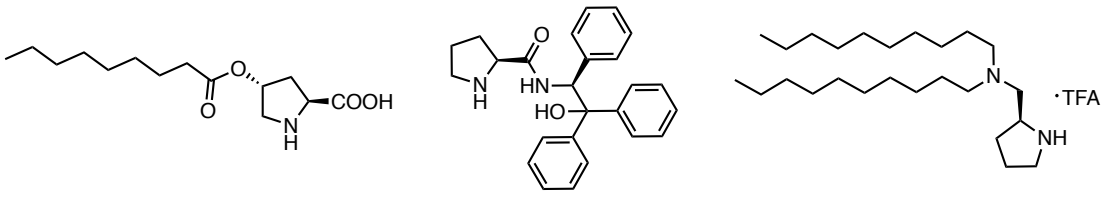


Figure 2.23 Example complex proline/ pyrolidine organocatalysts

In terms of a stereoselective "organocatalyst", few catalysts challenge the selectivities of biomolecular scaffolds for their native substrates (105). Enzymatic enantioselective enamine catalysis is carried out by type I aldolases that use a catalytic lysine residue for nucleophilic attack and enamine formation. The active site of aldolase enzymes is extremely complex. It is composed of several networks of

salt bridges and poised water molecules that orchestrate the reduction of the pK_a of lysine, binding and positioning the substrate and catalysis (106). Therefore, engineering aldolases for non-natural enamine/iminium catalysis is challenging as demonstrated by computationally designed retroaldolases with best designs having catalytic efficiencies (k_{cat}/K_M) of 1-100 $M^{-1}s^{-1}$, compared to natural enzyme's k_{cat}/K_M of 10^5-10^9 $M^{-1}s^{-1}$ (106, 107). Further, aldolases demonstrate limited substrate tolerance and must be engineered to accept non-phosphorlyated substrates (106). Although aldolases have proven difficult to adapt for non-natural aqueous organocatalysis, an exemplary success in the engineering of protein scaffolds is highlighted by the creation of catalytic antibodies (108-110). In fact, they are credited as the first, tailored, aqueous organocatalysts with high enantioselectivities (104).

In nature, enamine catalysis is confined to the reactivity of the amine in the lysine side chain. Proline catalysis does not exist. The lone pair of the nitrogen atom participates in amide bond formation with the sequential residue and therefore cannot act as nucleophile. However, there is a single example of a protein preforming non-natural proline-based enamine catalysis. The enzyme, 4-oxalocrotonate tautomerase (4-OT), contains an amino-terminal proline that, in its natural reaction, acts as a general base. The pK_a of the terminal proline, 6.4, is attenuated by the active site pocket in which it rests and therefore deprotonated at physiological pH (111). The enzyme was found to promiscuously catalyze the aldol reaction between acetaldehyde and benzaldehyde to form cinnamaldehdye and the Michael-type addition of acetaldehyde to TBNS. Both transformations proceed through proline catalyzed enamine formation of acetaldehyde, Figure 2.24 (112).

Amino-terminal tautomerases are an isolated class of protein-based proline catalysts. The enamine reactivity is the result of an evolved hydrophobic pocket specifically adapted to tune the pK_a of proline.

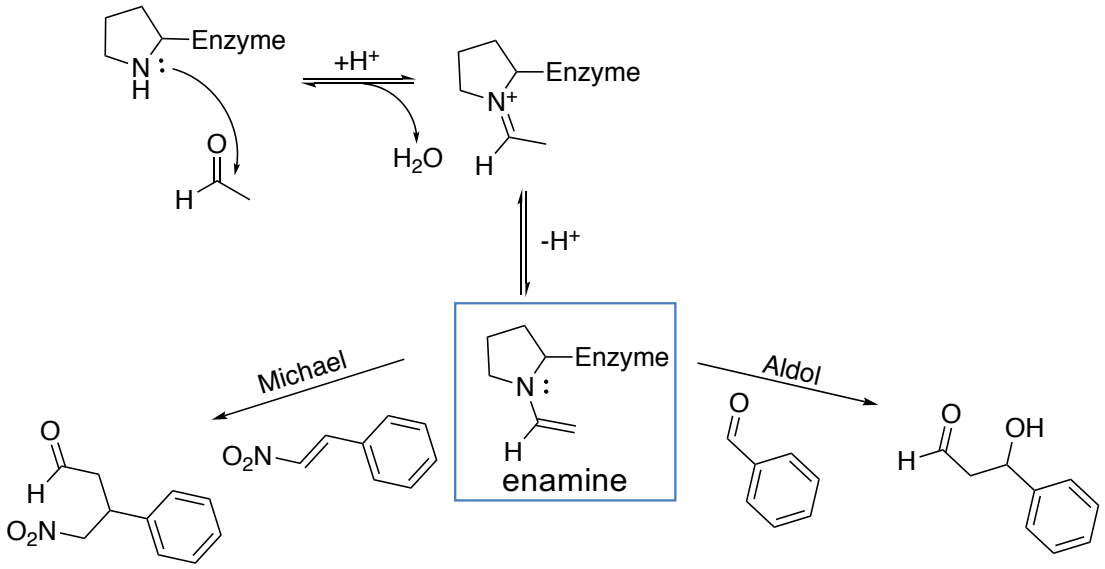


Figure 2.24 Promiscuous enamine chemistry of 4-OT

This chapter described our efforts towards aqueous organocatalysis by a proline-containing protein scaffold. Our strategy was to install a proline-catalyst, *in vivo*, via unnatural amino acid mutagenesis. We report the synthesis for two original UAAs, Pro4AF and PyrrolK, and both synthetic strategies are designed for scaled-preparation. The *in vivo* incorporation of a proline (ProK) and a pyrolidine (PyrrolK) containing unnatural amino acid by the WT-*Mb*PylRS was successful. Further, we describe the identification of a mutation that improves incorporation of ProK, *Mb*PylRS-Y349F and the evolution of a synthetase for PyrrolK, *Mb*PylRS-YL274A/C313V/Y349F. Pro4AF, a tyrosine-based proline UAA was synthesized however no synthetase was identified for its incorporation into protein sequences. The UAAs, ProK and PyrrolK, were demonstrated as capable catalysts, at high

catalyst loading, in a model aldol reaction between acetone and 4-nitrobenzladehyde. Lastly, a suitable protein scaffold to host the UAAs was identified, tHisF, and tHisF containing proK was subsequently evaluated for aldol catalysis.

Herein, the engineering of a catalytic-UAA containing protein scaffold was hindered by poor incorporation of proline/pyrrolidine-containing lysine-based UAAs by *Mb*PylRS. Efforts to evolve a synthetase for the incorporation of ProK, either by active-site directed site-saturation mutagenesis or by error prone PCR of the entire synthetase scaffold, were unsuccessful. Therefore, we were only able to produce minute amounts of UAA-containing protein (on average, <0.25 mg per liter of expression), hindering efforts to characterize the catalytic capabilities of the naïve UAA-containing protein scaffold. For free ProK aqueous aldol catalysis, standard organocatalyst loadings of 30 mol% were necessary. We were unable to recreate these reaction conditions with our proline-containing protein, and the reaction did not proceed at the catalyst loading we investigated, <1 mol%. Because no starting aldol activity was observed, we were unable to employ directed evolution mutagenesis and screening experiments.

Proline, with a pK_a of 10.6, is protonated at neutral pH. Nucleophilic attack of an amine can only occur by a free lone pair of electrons. On the basis of successes in the field of small molecule organocatalysis by sequestering the reaction to a concentrated organic layer, and the promiscuous catalysis by 4-OT, we hypothesized that confining a proline-based organocatalyst within a protein scaffold would provide initial, weak, activity, in an enamine-based reaction. We anticipated

that the confinement would also alleviate the high catalyst loading associated with free organocatalysts in solution, by protecting the catalyst from degradation and inactivation by electrophiles in the media. However, we discovered that this was not the case given our proline-UAA containing protein demonstrated no aldol activity at 1 mol% catalyst loading. Diagnostic experiments were performed to evaluate the ineffectiveness of tHisF50ProK at aldol chemistry. Reductive amination experiments suggest either the pK_a of proline within tHisF is poorly suited for nucleophilic attack, or acetone was unable to locate the amine of ProK, the later seems unlikely. The aldolase, DERA, has a similar fold and hydrophobic pocket and is very efficient at catalyzing the reaction via nucleophilic attack of lysine on acetone, as proved by reductive amination experiments. To further probe whether substrate access or pK_a is preventing catalysis, we planned to install ProK at a surface residue in tHisF and then perform reductive amination. However, all surface residues tested did not accept UAAs, Figure 2.16.

Despite the lack of catalysis by our proline-UAA-protein scaffold, this work presents ground work towards the creation and incorporation of amino-catalytic UAAs. Proline, the earliest and most extensively explored secondary amine catalyst fosters both enamine and iminium ion chemistry and has been proven capable of an impressive number of transformations (68). Even so, the stereo- and regio-selectivity is limited by the structurally simple nature of the small molecule catalyst and access to reactions such as aldehyde-aldehyde couplings are still confined to enzymatic systems (68, 93). Our goal of incorporating a reactive proline entity within a protein scaffold which can be engineered to accept various substrates through directed

evolution, is relevant. This approach will exploit the enantioselectivities inherent to protein scaffolds while circumventing the engineering difficulties associated with adapting aldolase enzymes and catalytic antibodies for applied chemical synthesis. However, the inability to recreate proline catalysis within a protein scaffold highlights the exquisite ability of nature to fine tune biomolecular scaffolds for a particular function in a manner specific to its unique substrate.

2.4 Materials and Methods

2.4.1 Protein expression, purification and characterization

Plasmids pET-GFP2TAG, pBad-sfGFP149TAG, pBK-*Mb*PylRS, pUltra-*Mb*PylRS, pREP-pylT, pBad-Barnase2TAG, pEvol-aaRS-tRNA^{CUA}*Mj*Tyr were acquired and used without manipulation from common lab supplies. Acetone, 4-nitrobenzaldehyde, trans-beta nitrostyrene, ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin were acquired from commercial sources and used without further purification. Antibiotics were used at the following concentrations, 100 μg/mL ampicillin, 25 μg/mL tetracycline, 40 μg/mL chloramphenicol, 50 μg/mL kanamycin and 50 μg/L streptomycin.

2.4.2 Amino acid toxicity test

A diagnostic toxicity test was performed for each unnatural amino acid. A single colony of DH10B competent cells (ThermoFisher) or Bl21-Gold(DE3) competent cells (Agilent Technologies) cells was used to inoculate 5 mL overnight cultures in LB and LB plus 1 mM unnatural amino acid. The cultures were grown to saturation overnight at 37 °C with shaking (225 rpm) and the optical density (OD) at 600 nm was determined. The OD in the presence of unnatural amino acid did not

differ from the culture grown in absence of unnatural amino acid, for Pro4AF, ProK, nor PyrrolK.

2.4.3 Plasmid construction and mutagenesis

Genes for *Thermus thermophilus* deoxyribrose-phosphate adolase (DERA), *Thermus thermophilus* Indole-3-glycerol phosphate synthase (IGPS) and *Thermotoga maritima* Histidine synthase F (tHisF) were ordered as GeneArt DNA strings with NdeI and XhoI cutsites from ThermoFisher Scientific. The DNA fragments and pET21C(+) were digested with NdeI and XhoI followed by ligation with NEB T4 DNA ligase at 16°C overnight. A 6X-C-terminal histidine tag was included. Amber stop codons were cloned using a BsaI-based cloning strategy. All TAG containing constructs were also cloned into the pBad-pylT backbone through amplification of the gene with NcoI and PmeI containing primers followed by restriction enzyme digest and ligation. All plasmids were verified by sequencing and restriction enzyme digests.

2.4.4 Small scale UAA incorporation

An individual colony was used to inoculate a starter culture in Luria-Bertani media (LB) supplemented with appropriate antibiotics. The culture was incubated at 37°C with shaking (225 rpm) overnight. 50 µL of this starter culture was used to inoculate three falcon tubes with 5 mLs LB media and antibiotics. To one culture, unnatural amino acid (typically 2 to 4 mM) was added from a stock 100 mM solution in water. To another, BocK was added to 1 mM final concentration. The third served as a negative control with no supplemented amino acid. The cultures were grown to an optical density at 600 nm of 0.6 after which expression was induced by

the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (pET, pUltra) or 0.2% arabinose (pBad, pBK). The cultures were incubated at 37°C with shaking overnight. The culture was pelleted by centrifugation at 1000g for 10 minutes. The supernatant was discarded and the pellet resuspended in 500 µL PBS with 0.25 mg/ml lysozyme. The slurry was transferred to an Eppendorf tube and incubated at 37°C for 30 minutes. The cells were then further lysed by sonication, 20% amplitude, 20 seconds at 50% duty cycle. The lysate was clarified by centrifugation at 15000 rpm for 10 minutes. 10 µL of the lysate was combined with 10 µL of 2X SDS load dye and 10 µL was ran on the gel. For GFP analysis, the clarified lysate was evaluated as described in GFP fluorescence measurements.

2.4.5 Large scale scaffold protein preparation

An individual colony was used to inoculate a starter culture in Luria-Bertani media (LB) supplemented with appropriate antibiotics. The culture was incubated at 37°C with shaking (225 rpm) overnight. This starter culture was used to inoculate a 1 L flask of LB supplemented with appropriate antibiotics and unnatural amino acid (2 - 4 mM). The culture was grown to an optical density at 600 nm of 0.6 after which expression was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (pET, pUltra) or 0.2% arabinose (pBad, pBK). The expression was carried out for 18 hours. The cells were then pelleted by centrifugation at 1000g for 30 minutes. The supernatant was discarded and the cells resuspended into 25 mL of buffer A (50 mM NaH_2PO_4, 30 mM imidazole, 150 mM NaCl, pH 8.0) supplemented with 0.5 mg/ml lysozyme. Cells were sonicated using a Fischer Scientific Sonic Dismembrator model 500 for 8 min at 20% amplitude at a 39% duty

cycle. Lysates were clarified by centrifugation at 27,000g for 1 hour, passed through a 0.22-μm filter and loaded on to a pre-equilibrated with buffer A Ni-NTA HisTrap HP column (5 ml HisTrap HP, GE Healthcare) using an AKTA Purifier (GE Healthcare). The column was washed with buffer A followed by a step elution to 10% buffer B (50 mM NaH2PO4, 300 mM imidazole, 150 mM NaCl, pH 8) to remove nonspecifically bounds proteins. The scaffold protein was eluted with 55% B. The protein was buffer exchanged into PBS using Amicon Ultra-10 centrifugal filters (EMD Milipore). A heat purification followed in which the protein was incubated at 70°C for 10 minutes to precipitate contaminating proteins. Concentrations were determined by nanodrop with extinction coefficient and molecular weight calculated from Scripps protein calculator. Purity of the protein was judged as >90% by SDS-PAGE and mass spectrometry on Agilent 6520 Accurate Mass QTOF LC-MS ESI positive in high-resolution mode.

2.4.6 GFP-based synthetase screen for Pro4AF incorporation

BL21 DE3 Gold cells (Agilent Technologies) harboring both pET22b(+)-sfGFP2TAG.his and pEvol carrying a unique synthetase and the cognizant $tRNA^{CUA}$*Mj*Tyr were inoculated into individual wells of a 96 well block containing 1 ml of LB media supplemented with 100 μg/mL ampicillin (amp) and 40 μg/mL chloramphenicol (cam). The block was incubated at 37 °C with shaking (980 rpm) overnight. This block was used to inoculate individual wells of a block containing LB amp/cam and 2 mM Pro4AF and a negative control block containing only LB amp/cam. The cultures were grown until at OD 0.6 when expression was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The blocks

continued shaking overnight after which the cells were pelleted at 1000g for 15 minutes. The supernatant was discarded and the cells resuspended into 500 µL of phosphate buffered saline with 0.25 mg/ml lysozyme added. The blocks were carried through 5 rounds of freeze thaw from -80°C to room temperature. The lysate was clarified by centrifugation at 1000g for 30 minutes. Fluorescent measurements were performed on clarified lysate.

2.4.7 GFP fluorescence measurements

Fluorescence measurements of clarified lysate or purified protein were carried out on Tecan M1000 PRO UV/Vis plate reader in 96 well clear plates by excitation at 488 nm and monitoring emission from 500 to 600 nm. Gain was set to the well of the positive control of BocK incorporation for each experiment.

2.4.8 Positive/negative selection system

200 to 500 ng of pBK-*Mb*PylRS-Y349F/L305NNK/Y306NNK/L309NNK/C348NNK was transformed into DH10B electrocompetent cells containing pREP pylT. The number of transformations was adjusted to cover the diversity of library, $1*10^7$ transformants. The rescue was used to inoculate 50 mLs of LB supplemented with kan and tet and grown to saturation. 1 mL of the saturated culture was used to inoculate 100 mL of LB-kan-tet containing 1 mM unnatural amino acid. The culture was incubated at 37 °C for 3 hours after which chloramphenicol was added to 80 µg/ml and then grown to saturation. A second positive selection was carried out by inoculation of 100 mL of LB-kan-tet containing 1 mM unnatural amino acid with 0.5 mL of positive selection round 1. This culture was grown to saturation followed by centrifugation of 20 mLs at 1000g

for 30 minutes. The culture was miniprepped and the pBK- *Mb*PylRS 4 site 'NNK' library was purified from the pREP-pylT by agarose gel electrophoresis followed by gel extraction.

A negative selection was carried out by co-transformation of 200 ng of positive selection round 2 library in pBK with 200 ng of pBad-barnase-Q2TAG/D44TAG. The rescue was used to inoculate 100 ml of LB-kan-tet containing 0.2% arabinose to induce expression of the barnase gene. The culture was incubated at 37 °C over night. 20 mLs of the negative selection was miniprepped and the pBK- *Mb*PylRS 4 site 'NNK' library was gel purified from the pBad negative selection plasmid.

The resulting library was then subjected to a GFP based screen to identify synthetase hits that are selective for the unnatural amino acid. 100 ng of the pBK-*Mb*PylRS 4 site 'NNK' library was co-transformed with pBAD-sfGFP149TAG-pylT. 1 mL of LB-kan-tet in individual wells in two 96 well blocks were inoculated with a single transformant. The block was grown to saturation overnight at 37 °C with shaking (990 rpm). 50 µL of each well from this block was used to inoculate individual wells of a fresh block containing 1 ml LB-kan-tet with 2 mM unnatural amino acid and a second block containing 1 mL LB-kan-tet. The blocks were grown for 2 hours at 37 °C with shaking after which arabinose was added to 0.2%. The blocks were incubated overnight followed by centrifugation at 1000g for 10 minutes. The supernatant was discarded and the blocks were lysed and clarified as previously described. 250 µL of the clarified supernatant were transfer to a black 96 well florescence plate and the emission spectra was recorded.

2.4.9 Error prone *Mb*pylRS library and FAC sorting

Error prone PCR was carried out on 10 ng of *Mb*PylRS-Y349F using the following primers, *Mb*PylRS -SacI-F (GATGTTGAGCTCATGATG GATAAAAAACCG CTGG) and *Mb*PylRS -SalI-R (GATGTTGTCGACTTACAGGTTCGTGCTAATGC) at 0.3 μM and Taq polymerase. Manganese chloride concentrations ranged from 0.15 mM to 0.05 mM. Magnesium chloride was held at 5.5 mM. 0.2 mM dGTP, 0.2 mM dATP, 1 mM dCTP and 1 mM dTTP were used. Ligations were carried out 16 °C overnight using a three to one ratio of SacI, SalI digested *Mb*PylRS-Y349F library to SacI, SalI digested pBK backbone. The ligation was transformed into DH10B cells to obtain a minimum of $1*10^6$ transformants. The rescue was used to inoculate 50 mLs of LB-kan-tet and grown to saturation overnight. 20 mL of the culture was miniprepped to isolate the library. The library was transformed with pBad-sfGFP149TAG-pylT in DH10B cells. The rescue was used to inoculate 50 mL LB-kan-tet and grown over night. 50 μLs of this culture was used to inoculate 5 mLs of LB-kan-tet with 2 mM unnatural amino acid and 5 mL of LB-kan-tet. The cultures were grown to an optical density of 0.6 at 600 nm followed by addition of 0.2% arabinose. After expression of GFP occurred overnight at 37 °C with shaking, cells were evaluated on a Beckman Coulter MoFlo. Only cells demonstrating florescence intensity above the control expression of DH10B cells harboring the pBK- *Mb*PylRS-Y349F parent and pBad-sfGFP149TAG-pylT were collected in LB media. The media was plated on LB-Kan-Tet agar and individual colonies were grown to saturation, miniprepped and sequenced.

2.4.10 Aldol reaction

Small scale protein reactions were performed in Agilent screw top clear glass vials equipped with a stir bar. Acetone was added to 200 mM from a 20X stock in water. 4-nitrobenzaldehyde was added to 10 mM from a 50 X stock in DMSO. Lastly, catalyst was added from a 10X stock, 25 µL of 10X PBS was added and the final volume was adjusted with water. Reaction volumes varied from 500 µL to 250 µL depending on protein yields. At completion, reactions were quenched with ammonium chloride and diluted with an equal volume of ethyl acetate. Trans-beta nitro styrene (internal standard) was added to 5 mM from a 100X stock in DMSO. The reaction was transferred to an Eppendorf tube and vortexed. The reaction was then centrifuged at 15000 rpm for 1 minute. The organic layer was injected into Agilent LC 1260 Infinity II system equipped with an Astec Chirobiotic V2 10cm * 4.6 mm, 5µM column. An isocratic 80:20 hexanes:isopropyl alcohol premixed solvent system at 1.00ml/min with detection at 254 nm was used. Peaks corresponding to product and internal standard were integrated to obtain yields.

2.4.11 Reductive amination

Protein, or control molecule was dilute to working concentration (10-100 µM) in 200 mM phosphate buffer, pH 7.4 in an Agilent screw top clear glass vial equipped with a stir bar. Ketone or aldehyde substrate was added from a 50X stock in acetonitrile to final concentration between 1 and 100 mM. The mixture was allowed to stir at room temperature for 30 minutes. Sodium cyanoborohydride was added to 25 mM from a 10X stock in water. The reaction was stirred for an additional hour. The reaction was then pipetted into an Amicon Ultra-10 centrifugal filters (EMD Milipore). The reaction was diluted with phosphate buffer to 15 ml and

centrifuged to 500 µL at 4500 rpm. This wash was repeated three times. The protein was then passed through a 0.22-µm filter and then carried on to mass spectrometry analysis.

2.5 Supporting Information

2.5.1 Protein and DNA sequences

sfGFP14TAG DNA sequence:

ATGGTTAGCAAAGGTGAAGAACTGTTTACCGGCGTTGTGCCGATTCTGGTGGA
ACTGGATGGTGATGTGAATGGCCATAAATTTAGCGTTCGTGGCGAAGGCGAAG
GTGATGCGACCAACGGTAAACTGACCCTGAAATTTATTTGCACCACCGGTAAAC
TGCCGGTTCCGTGGCCGACCCTGGTGACCACCCTGACCTATGGCGTTCAGTGC
TTTAGCCGCTATCCGGATCATATGAAACGCCATGATTTCTTTAAAAGCGCGATG
CCGGAAGGCTATGTGCAGGAACGTACCATTAGCTTCAAAGATGATGGCACCTA
TAAAACCCGTGCGGAAGTTAAATTTGAAGGCGATACCCTGGTGAACCGCATTG
AACTGAAAGGTATTGATTTTAAAGAAGATGGCAACATTCTGGGTCATAAACTGG
AATATAATTTCAACAGCCATTAGGTGTATATTACCGCCGATAAACAGAAAAATGG
CATCAAAGCGAACTTTAAAATCCGTCACAACGTGGAAGATGGTAGCGTGCAGC
TGGCGGATCATTATCAGCAGAATACCCCGATTGGTGATGGCCCGGTGCTGCTG
CCGGATAATCATTATCTGAGCACCCAGAGCGTTCTGAGCAAAGATCCGAATGAA
AAACGTGATCATATGGTGCTGCTGGAATTTGTTACCGCCGCGGGCATTACCCA
CGGTATGGATGAACTGTATAAAGGCAGCCACCATCATCATCACCATTAA

sfGFP149TAG protein sequence:

M V S K G E E L F T G V V P I L V E L D G D V N G H K F S V R G E G E G D A T
N G K L T L K F I C T T G K L P V P W P T L V T T L T Y G V Q C F S R Y P D H M

K R H D F F K S A M P E G Y V Q E R T I S F K D D G T Y K T R A E V K F E G D

T L V N R I E L K G I D F K E D G N I L G H K L E Y N F N S H N V Y I T A D K Q K

N G I K A N F K I R H N V E D G S V Q L A D H Y Q Q N T P I G D G P V L L P D N

H Y L S T Q S V L S K D P N E K R D H M V L L E F V T A A G I T H G M D E L Y K

G S H H H H H H

DERA DNA sequence:

ATGGATCTGGCAGCCCATATTGATCATACCCTGCTGAAACCGACCGCAACCCT

GGAAGAAGTTGCAAAAGCAGCCGAAGAGGCACTGGAATATGGTTTTTATGGTC

TGTGTATTCCGCCTAGCTATGTTGCATGGGTTCGTGCACGTTATCCGCATGCAC

CGTTTCGTCTGGTTACCGTTGTTGGTTTTCCGCTGGGTTATCAAGAAAAAGAAG

TTAAAGCACTGGAAGCAGCACTGGCATGTGCACGTGGTGCAGATGAAGTTGAT

ATGGTTCTGCATCTGGGTCGTGCAAAAGCCGGTGATCTGGATTATCTGGAAGC

CGAAGTTCGTGCAGTTCGTGAAGCAGTTCCGCAGGCAGTTCTGAAAGTTATTCT

GGAAACCGGTTATTTTAGTCCGGAAGAAATTGCACGTCTGGCCGAAGCAGCAA

TTCGTGGTGGTGCCGATTTTCTGAAAACCAGCACCGGTTTTGGTCCGCGTGGT

GCCAGCCTGGAAGATGTGGCACTGCTGGTTCGTGTTGCACAGGGTCGTGCCC

AGGTTAAAGCAGCGGGTGGTATTCGTGATCGTGAAACCGCACTGCGTATGCTG

AAAGCGGGTGCAAGCCGTCTGGGCACCAGCAGCGGTGTTGCACTGGTTGCCG

GTGAAGGTGGCACCCTGGGCTATCTCGAGCACCACCACCACCACCACTGA

DERA protein sequence:

M D L A A H I D H T L L K P T A T L E E V A K A A E E A L E Y G F Y G L C I P P S

Y V A W V R A R Y P H A P F R L V T V V G F P L G Y Q E K E V K A L E A A L A

C A R G A D E V D M V L H L G R A K A G D L D Y L E A E V R A V R E A V P Q A

V L K V I L E T G Y F S P E E I A R L A E A A I R G G A D F L K T S T G F G P R G

A S L E D V A L L V R V A Q G R A Q V K A A G G I R D R E T A L R M L K A G A

S R L G T S S G V A L V A G E G G T L G Y L E H H H H H H

IGPS DNA sequence:

ATGCGTCCGGATCTGAGCCGTGTTCCGGGTGTTCTGGGTGAAATTGCACGTAA

ACGTGCCAGCGAAGTTGCACCGTATCCGCTGCCGGAACCGCCTAGCGTTCCG

AGCTTTAAAGAAGCACTGCTGCGTCCTGGTCTGAGTGTTATTGCAGAAGTTAAA

CGTCAGAGCCCGAGCGAAGGTCTGATTCGTGAAGTTGATCCGGTTGAAGCAGC

ACTGGCCTATGCACGTGGTGGTGCACGTGCCGTTAGCGTTCTGACCGAACCGC

ATCGTTTTGGTGGTAGCCTGCTGGATCTGAAACGTGTTCGTGAAGCAGTTGATC

TGCCGCTGCTGCGTAAAGATTTTGTTGTTGATCCGTTTATGCTGGAAGAGGCAC

GTGCGTTTGGTGCAAGCGCAGCCCTGCTGATTGTTGCACTGCTGGGTGAACTG

ACCGGTGCATATCTGGAAGAAGCCCGTCGTCTGGGTCTGGAAGCACTGGTTGA

AGTTCATACCGAACGTGAACTGGAAATTGCCCTGGAAGCCGGTGCCGAAGTTC

TGGGTATTAACAATCGTGATCTGGCAACCCTGCATATTAATCTGGAAACCGCAC

CTCGCCTGGGTCGTCTGGCACGCAAACGTGGTTTTGGCGGTGTGCTGGTTGCA

GAAAGCGGTTATAGCCGTAAAGAAGAACTGAAAGCACTGGAAGGTCTGTTTGA

TGCAGTTCTGATTGGCACCAGCCTGATGCGTGCACCGGATCTGGAAGCTGCAC

TGCGCGAACTGGTTGGCCTCGAGCACCACCACCACCACCACTGA

IGPS protein sequence:

M R P D L S R V P G V L G E I A R K R A S E V A P Y P L P E P P S V P S F K E A

L L R P G L S V I A E V K R Q S P S E G L I R E V D P V E A A L A Y A R G G A R

A V S V L T E P H R F G G S L L D L K R V R E A V D L P L L R K D F V V D P F M

L E E A R A F G A S A A L L I V A L L G E L T G A Y L E E A R R L G L E A L V E V

H T E R E L E I A L E A G A E V L G I N N R D L A T L H I N L E T A P R L G R L A

R K R G F G G V L V A E S G Y S R K E E L K A L E G L F D A V L I G T S L M R A

P D L E A A L R E L V G L E H H H H H H

*Mb*PylRS DNA sequence:

ATGATGGATAAAAAACCGCTGGATGTGCTGATTAGCGCGACCGGCCTGTGGAT

GAGCCGTACCGGCACCCTGCATAAAATCAAACATCATGAAGTGAGCCGCAGCA

AAATCTATATTGAAATGGCGTGCGGCGATCATCTGGTGGTGAACAACAGCCGT

AGCTGCCGTACCGCGCGTGCGTTTCGTCATCATAAATACCGCAAAACCTGCAA

ACGTTGCCGTGTGAGCGATGAAGATATCAACAACTTTCTGACCCGTAGCACCG

AAAGCAAAAACAGCGTGAAAGTGCGTGTGGTGAGCGCGCCGAAAGTGAAAAAA

GCGATGCCGAAAAGCGTGAGCCGTGCGCCGAAACCGCTGGAAAATAGCGTGA

GCGCGAAAGCGAGCACCAACACCAGCCGTAGCGTTCCGAGCCCGGCGAAAAG

CACCCCGAACAGCAGCGTTCCGGCGTCTGCGCCGGCACCGAGCCTGACCCGC

AGCCAGCTGGATCGTGTGGAAGCGCTGCTGTCTCCGGAAGATAAAATTAGCCT

GAACATGGCGAAACCGTTTCGTGAACTGGAACCGGAACTGGTGACCCGTCGTA

AAAACGATTTTCAGCGCCTGTATACCAACGATCGTGAAGATTATCTGGGCAAAC

TGGAACGTGATATCACCAAATTTTTTGTGGATCGCGGCTTTCTGGAAATTAAAA

GCCCGATTCTGATTCCGGCGGAATATGTGGAACGTATGGGCATTAACAACGAC

ACCGAACTGAGCAAACAAATTTTCCGCGTGGATAAAAACCTGTGCCTGCGTCC

GATGCTGGCCCCGACCCTGTATAACTATCTGCGTAAACTGGATCGTATTCTGCC

GGGTCCGATCAAAATTTTTGAAGTGGGCCCGTGCTATCGCAAAGAAAGCGATG

GCAAAGAACACCTGGAAGAATTCACCATGGTTAACTTTTGCCAAATGGGCAGC

GGCTGCACCCGTGAAAACCTGGAAGCGCTGATCAAAGAATTCCTGGATTATCT

GGAAATCGACTTCGAAATTGTGGGCGATAGCTGCATGGTGTATGGCGATACCC

TGGATATTATGCATGGCGATCTGGAACTGAGCAGCGCGGTGGTGGGTCCGGTT

AGCCTGGATCGTGAATGGGGCATTGATAAACCGTGGATTGGCGCGGGTTTTGG

CCTGGAACGTCTGCTGAAAGTGATGCATGGCTTCAAAAACATTAAACGTGCGA

GCCGTAGCGAAAGCTACTATAACGGCATTAGCACGAACCTGTAA

*Mb*PylRS protein sequence:

M M D K K P L D V L I S A T G L W M S R T G T L H K I K H H E V S R S K I Y I E

M A C G D H L V V N N S R S C R T A R A F R H H K Y R K T C K R C R V S D E

D I N N F L T R S T E S K N S V K V R V V S A P K V K K A MP K S V S R A P K P

L E N S V S A K A S T N T S R S V P S P A K S T P N S S V P A S A P A P S L T R

S Q L D R V E A L L S P E D K I S L N M A K P F R E L E P E L V T R R K N D F Q

R L Y T N D R E D Y L G K L E R D I T K F F V D R G F L E I K S P I L I P A E Y V

E R M G I N N D T E L S K Q I F R V D K N L C L R P M L A P T L Y N Y L R K L D

R I L P G P I K I F E V G P C Y R K E S D G K E H L E E F T M V N F C Q M G S G

C T R E N L E A L I K E F L D Y L E I D F E I V G D S C M V Y G D T L D I M H G D

L E L S S A V V G P V S L D R E W G I D K P W I G A G F G L E R L L K V M H G

F K N I K R A S R S E S Y Y N G I S T N L

2.5.2 Primer Sequences

DERA_Nde1_F:

GTGGTGCATATGGACCTGGCCGCCCACA

DERA_Xho1_R:

GTGGTGCTCGAGGTAGCCAAGGGTCCCC

IGPS_NdeI_F:

GCGGACCATATGCGTCCGGATCTGAGCCG

IGPS_XhoI_R:

GCGGACCTCGAGGCCAACCAGTTCGCGCAGTGC

DERA_NcoI_F:

GCGCACCCATGGATCTGGCAGCCCATATTG

HisTag_PmeI_R:

GTGGACGTTTAAACTCAGTGGTGGTGGTGGTGG

NcoI_tHisF_F:

GCGGACCCATGGGCCTGGCCAAACGTATTATTGC

DERA_K151_TAG_F:

GATGTTGGTCTCATCTGTAGACCAGCACCGGTTTTGGTCCGC

DERA_K151_TAG_R:

GATGTTGGTCTCACAGAAAATCGGCACCACCACGAATTGC

DERA_K151_Ala_F:

GATGTTGGTCTCATCTGGCTACCAGCACCGGTTTTGGTCCGC

DERA_K180_TAG_F:

GATGTTGGTCTCAGGTTTAGGCAGCGGGTGGTATTCGTGATCG

DERA_K180_TAG_R:

GATGTTGGTCTCAAACCTGGGCACGACCCTGTGCAACACG

DERA_K180_Ala_F:

GATGTTGGTCTCAGGTTGCTGCAGCGGGTGGTATTCGTGATCG

tHisF_N22_TAG_F:

GCAGCGGGTCTCACACCTAGTTTGAAAATCTGCGTG

tHisF_N22_TAG_R:

GCAGCGGGTCTCAGGTGCCTTTAACAACACG

tHisF_E24_TAG_F:

GCAGCGGGTCTCACTTTTAGAATCTGCGTGATAGC

tHisF_E24_TAG_R:

GCAGCGGGTCTCAAAAGTTGGTGCCTTTAAC

tHisF_R27_TAG_F:

GATGTTGGTCTCATCTGTAGGATAGCGGTATC

tHisF_R27_TAG_R:

GATGTTGGTCTCACAGATTTTCAAAGTTGGTGC

tHisF_L50_TAG_F:

GATGTTGGTCTCAGTTTTAGGATATCACCGCAAGCGTTG

tHisF_L50_TAG_R:

GATGTTGGTCTCAAAACACCAGCTCATCAATGCCAATTTC

tHisF_T78_TAG_F:

GATGTTGGTCTCTGTTTTAGGTTGGTGGTGGCATCCATG

tHisF_T78_TAG_R:

GATGTTGGTCTCTAAACGGAATATCAATCTGCTCTG

tHisF_L169_TAG_F:

GATGTTGGTCTCTAATTTAGCTGACCAGCATTGATCGTG

tHisF_L169_TAG_R:

GATGTTGGTCTCTAATTTCACCGGCACCACG

tHisF_T171_TAG_F:

GATGTTGGTCTCAGCTGTAGAGCATTGATCGTGATGGC

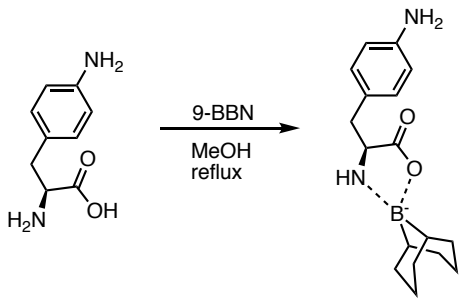
tHisF_T171_TAG_R:

GATGTTGGTCTCACAGCAGAATTTCACCGGCACCAC

2.5.3 Small molecule synthesis and characterization

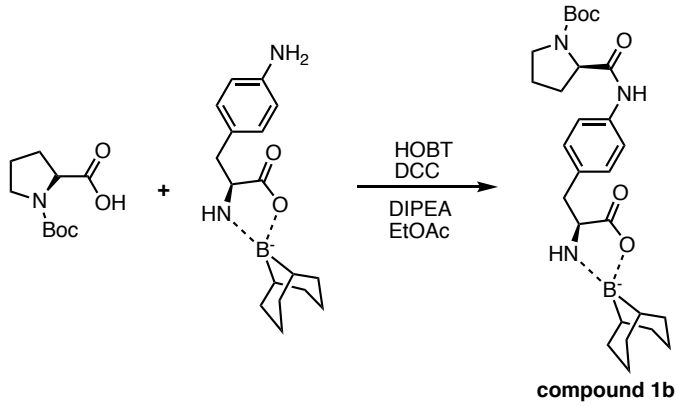
All reagents were obtained from commercial suppliers and used without further purification. N-Boc-D-proline, Nα-Boc-L-Lysine, (S)-N-Boc-3-hydroxypyrrolidine, (R)-N-Boc-3-hydroxyproline,3-(3-Dimethylaminoproply)-1-ethyl-carbodimide hydrochloride (EDC•HCl), N,N'-Disuccinimidyl carbonate, 4-Amino-L-phenylalanine, 1-Hydroxybenzotriazole hydrate (HOBt), N,N'-Dicyclohexylcarbodiimide (DCC) were obtained from Chem-Impex. 4-Dimethylaminopyridine (DMAP), anhydrous dichloromethane (DCM), and N-N-Dimethylformamide (DMF) "extra dry acro sealed" were obtained from Acros Organics. Lastly, N-hydroxysuccinimide, hydrogen chloride solution, 4 M in dioxane, anhydrous 1-4 dioxane and 9-Borabicyclo[3.3.1]nonane solution, 0.5 M in THF (9-BBN) and N,N-Diisopropylethylamine (DIPEA), 4-nitrobenzaldehyde, trans-beta-nitrostyrene and acetone were obtained from Sigma Aldrich. All anhydrous and extra dry reagents were sparged with N_2 or Ar prior to use.

N6-(D-prolyl)-4-Amino-L-phenylalanine, Pro4AF



compound 1a

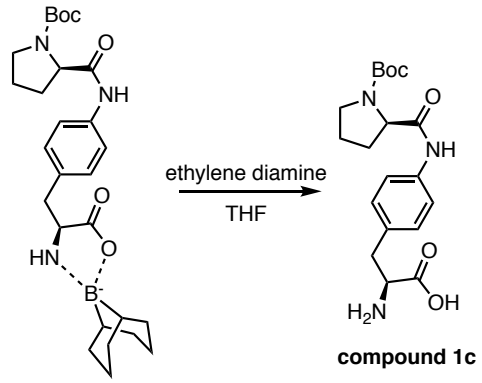
Compound 1a, 4-Amino-L-phenylalanine was dissolved in anhydrous methanol (1 mmol) then 9-BBN in THF (1.5 mmol) was added. The solution was stirred overnight. The reaction was diluted with ethyl acetate, workup with brine and the organic layer was dried over sodium sulfate. The crude reaction was concentrated and then dilute with minimal ethyl acetate to dissolve crude product. This was passed over silica and the product was eluted with a 20-80% gradient ethyl acetate/hexanes. (70%)



compound 1b

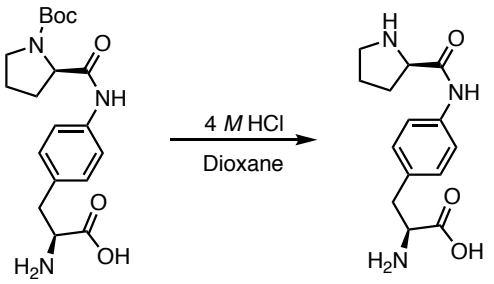
Compound 1b, Boc-D-proline (1mmol) and HOBt (1 mmol) was dissolved in ethyl acetate (0.2 M) in an ice bath. DCC (1.4 mmol) was added and the mixture was stirred until precipitation of dicyclohexylurea (DCU), ~30mins. The mixture was filtered to remove DCU. Compound 1a and DIPEA (1 mmol) was added and the reaction was stirred overnight. A white precipitant was collected by filtration. This

crude product, containing small amounts of DCU was carried out without further purification. (75%).



compound 1c

Compound 1c, Crude compound 1b was dissolved in THF (0.01 M). 14 equivalences of ethylene diamine was added dropwise. A solid precipitated. The reaction was stirred overnight and the precipitated solid was collected by filtration. This crude product was taken on without further purification.

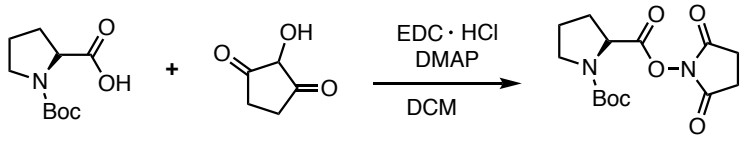


compound 1

Compound 1, Crude compound 1c was dissolved in 4 M HCL in dioxane (12 eq. of HCL) and stirred overnight. A white precipitant was collected by filtration. This was dissolved in water and purified by reverse phase chromatography by a 5 to 95% water/acetonitrile gradient. Pure amino acid eluted immediately from the column. (quantitative) Calculated mass: m/z [M+H] = 278.1499 observed mass: m/z [M+H] = 278.1504

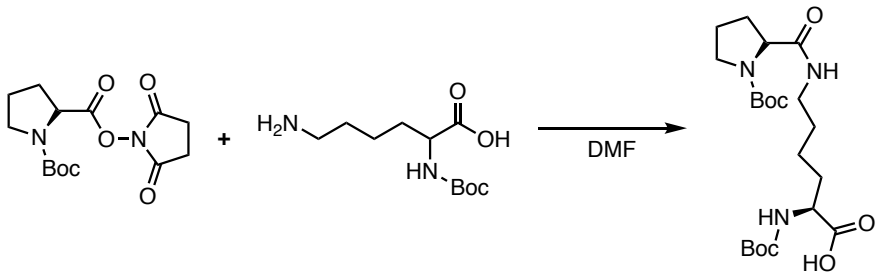
^{1}H NMR (400 MHz, D2O) δ 11.2 (s, 1H), 10.26 (s, 1H), 8.86 (s, 1H), 7.62 (d, 2H), 7.26 (d, 2H), 4.42 (t, 1H), 4.11 (q, 1H), 3.12 (t, 3H), 2.51 (m, 2H), 2.44 (m, 1H), 1.94 (m, 3H).

N6-(D-prolyl)-L-Lysine, ProK



compound 2a

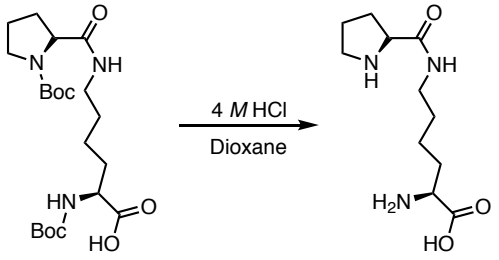
Compound 2a, N-Boc-D-proline, (1 mmol) was dissolved in anhydrous dichloromethane (0.2 M). EDC HCl (1.3 mmol) and DMAP (0.8mmol) were added slowly in an ice bath. The solution was stirred until a color change to yellow/orange was observed. N-hydroxysuccinimide (1.3 mmol) was then added and the water bath was removed. The reaction was stirred overnight. The orange solution with brine and dichloromethane. The DCM layer was dried using sodium sulfate and concentrated to give an orange solid. The product was then purified using flash column chromatography 20-80% ethyl acetate/hexanes gradient. The product was a while powder (80%).



compound 2b

Compound 2b, Nα-Boc-L-Lysine (1.3 mmol) was dissolved in DMF (0.2 M). Compound 2a was added to the solution (1 mmol). The reaction was stirred overnight and then diluted with water and ethyl acetate. Remaining Nα-Boc-L-Lysine

is soluble in the water layer. The ethyl acetate layer was concentrated in vacuo to give compound 2b as an oil (82%).

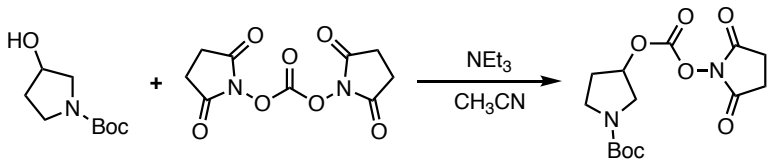


compound 2

Compound 2, D-ProK, Compound 2b was dissolved in 4 M HCl in dioxane (9 eq. HCl) and stirred overnight. The reaction was diluted to in water and ethylacetate. The product was soluble in water and compound 2b in ethylacetate. The water solution was then concentrated in vacuo yielding D-ProK, in the form of a yellow viscous oil (¬¬¬-quantitative). Calculated mass: m/z [M+H] = 244.1656, observed mass: m/z [M+H] = 244.1669

^{1}H NMR (400 MHz, DMSO- D6) δ 4.38-4.19 (m, 1H), 4.12-3.94 (m, 1H), 3.49-3.26 (m, 2H), 3.26-3.06 (m, 2H), 2.49-2.26 (m, 1H), 2.17-1.73 (m, 5H), 1.64-1.45 (m, 2H), 1.45-1.24 (m, 2H).

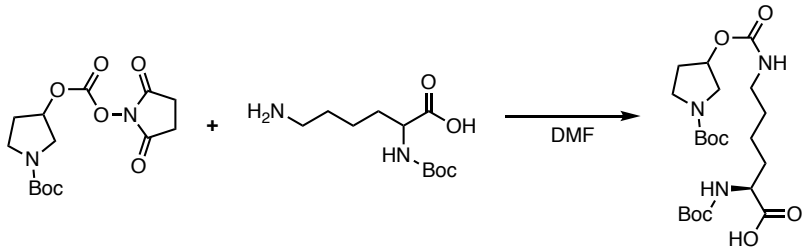
N6-((pyrrolidin-3-yloxy)carbonyl)-L-lysine, (R) and (S) PyrrolK



compound 3a

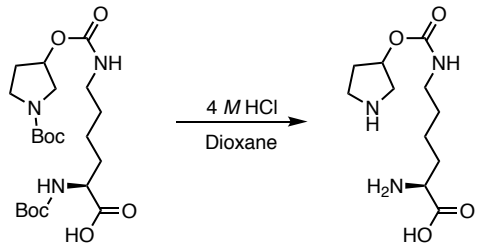
Compound 3a, (S)-N-Boc-3-hydroxypyrrolidine and (R)-N-Boc-3-hydroxyproline were carried through the synthesis separately to yield (R)-PyrrolK and (S)-PyrrolK. A solution of N-Boc-3-hydroxypyrrolidine (1 mmol), N,N'-Disuccinimidyl carbonate (1.5 mmol) and trimethylamine (2 mmol) in anhydrous acetonitrile (0.2 M) was stirred

at room temperature overnight. The reaction was diluted with saturated sodium bicarbonate and ethyl acetate. The organic layer was washed with brine and dried over magnesium sulfate. The organic layer concentrated to a yellow oil. The crude organic extract was purified by normal phase chromatography by a 20 to 80% gradient, ethyl acetate/hexanes. (20%) Calculated mass: m/z [M+Na] = 351.1192, observed mass: m/z [M+Na] = 351.1195



compound 3b

Compound 3b, Purified compounds 3a (1mmol) was dissolved in DMF (0.2 M). Nα-Boc-L-lysine (1.3 mmol) was added and the reaction was stirred overnight. The reaction was diluted with DCM and water. The DCM layer was washed with water and dried over sodium sulfate. The organic extract concentrated to a yellow/brown oil. The oil was purified using reverse phase chromatography with a gradient of 5 to 95 % water/acetonitrile. (quantitative) Calculated mass: m/z [M+Na] = 482.2492, observed mass: m/z [M+Na] = 482.2459



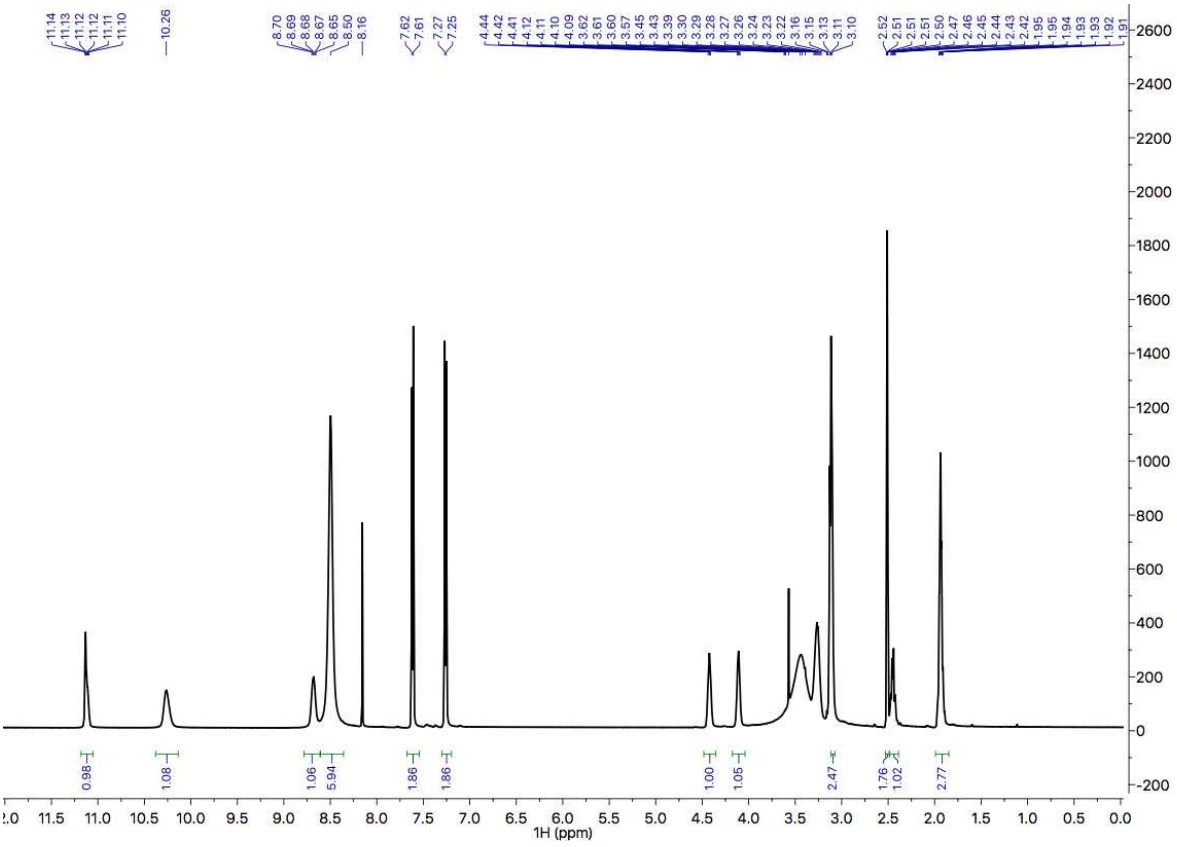
compound 3

Compound 3, (R)-PyrrolK and (S)-PyrrolK, Compound 3b was dissolved in 4 M HCL in dioxane (35 eq. of HCl). A solid precipitated as the reaction stirred overnight. The

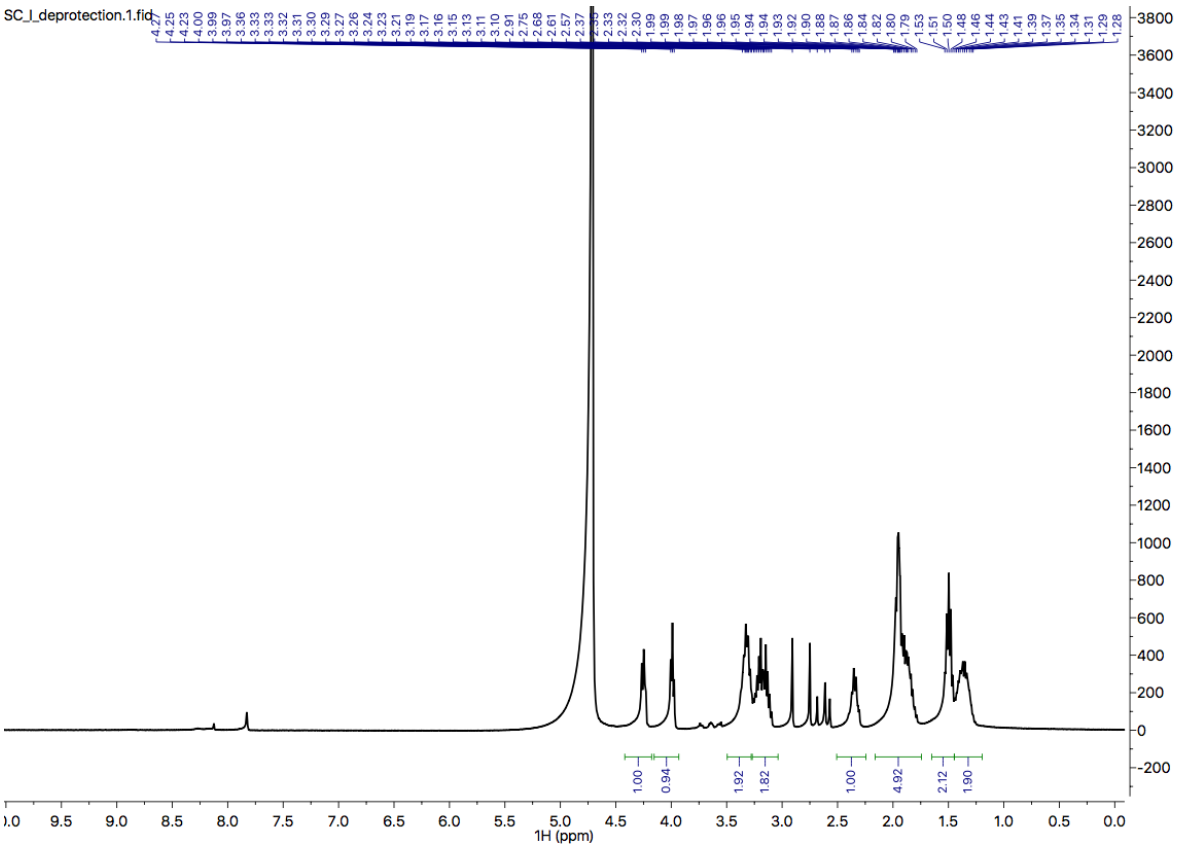
dioxane was removed by vacuum. The sticky solid was dissolved in water and purified by reverse phase chromatography by a 5 to 95% water/acetonitrile gradient. (quantitative) Pure amino acid eluted immediately from the column. Calculated mass: m/z [M+H] = 260.1573, observed mass: m/z [M+H] = 260.1611

2.5.4 ^{1}H NMR

Pro4AF



ProK



CHAPTER 3: ^{19}F NMR INVESTIGATION OF PROTEIN AND PROTEIN COMPLEX STABILITY VIA UNNATURAL AMINO ACID MUTAGENESIS[1]

3.1 Introduction

Life evolved in water. Accordingly, cellular survival required a mechanism to cope with fluctuations in external osmolality. Natural selection arrived at a simple solution; the accumulation or depletion of small molecules called osmolytes (113). Osmolytes are compatible solutes because they do not perturb macromolecular structure and function (113). The thermodynamic origin of compatibility as it applies to protein tertiary structure was elucidated through application of transfer free energies and the Tanford transfer model (114-116). This model suggests that osmolytes are preferentially excluded from protein backbone, and thus favor the folded state by decreasing the conformational entropy of the unfolded ensemble, although the idea remains controversial (117, 118). No matter the mechanism, osmolytes stabilize native protein folds and counteract the destabilizing forces arising from augmented intracellular crowding during osmotic stress (119) or the accumulation of urea (120). This stabilization has been the primary emphasis in discussions of osmolyte fitness. Proteins, however, do not function in isolation, and focusing solely on protein folding equilibria may provide an incomplete picture of the specialized ability of osmolytes to serve as compatible solutes.

[1] Figures and data described in this chapter have been submitted accompanying a similar narrative as a communication to the Journal of the American Chemical Society.

Protein-protein interactions are ubiquitous; for example, the interactome of human cells encompasses at least 5,400 proteins communicating via 28,500 interactions (121). Despite the fundamental role of protein surfaces, the influence of osmolytes on protein interfaces has garnered limited attention (122-129). In general, osmolytes that stabilize protein folds tend to stabilize protein complexes but there are exceptions (129). In addition, the transfer model, which accurately predicts osmolyte influence on folding equilibria, does not address effects on protein-protein interactions. This shortcoming has been rationalized in terms of differences in the interactions that drive folding and protein-complex assembly. Whereas protein folds are stabilized to a large extent by backbone hydrogen bonds localized in elements of secondary structure (130), protein interfaces are predominately composed of side chain interactions with less emphasis on backbone interactions (131, 132).

Recent studies have broadened our appreciation of the importance of protein surfaces. These studies emphasize that, in addition to supporting solubility and contact with water, electrostatic interactions of surface residues also contribute to protein stability (133). The physiological implications of these interactions, which are destabilizing if attractive and stabilizing if repulsive, have been demonstrated in the crowded interior of *Escherichia coli* cells (134-137) and under *in vitro* cell-like conditions (138, 139). During osmotic stress, electrostatic interactions between macromolecules are predicted to be intensified due to the efflux of cytosolic water and consequent increase in macromolecular crowding (119, 140, 141). In light of the known behavior of osmolytes as compatible solutes with respect to the stability of protein tertiary structure, we hypothesized that the molecules selected for

accumulation during stress must also be compatible with the interactions of protein surfaces.

Here, we evaluate the effect of a panel of cosolutes, including osmolytes and related non-osmolyte biomolecules, on the stability of a model protein (the SH3 domain of the *Drosophila* signal transduction protein drk) and a protein complex [the side-by-side symmetric homodimer of the B1 domain of the streptococcal immunoglobulin binding protein G, (GB1)]. Determination of protein and protein complex stability was facilitated by incorporating an unnatural amino acid bearing a trifluoromethyl probe that allowed simple quantification of protein populations via ^{19}F NMR. With the exception of urea, osmolytes and non-osmolytes stabilize the fold of SH3. Divergent behavior was observed for the protein-protein interaction; osmolytes had a benign influence on oligomerization, whereas non-osmolytes favored the dimer. These results suggest compatibility with both protein structure and protein-protein interactions are important features of osmolytes.

3.2 Results

3.2.1 Cosolute selection

We selected a panel of fourteen molecules (*SI Appendix*, Fig. S1) that have historically been categorized as osmolytes and encompass the diverse chemical structures utilized by various organisms. These molecules include amino acids and their derivatives (alanine, β-alanine, glutamate, glycine, proline and taurine), methylamines (betaine, sarcosine, trimethylamine-N-oxide), polyols (glycerol, trehalose, sorbitol, sucrose), and urea (a denaturing osmolyte). For comparison, we

chose eight molecules that are also critical for cellular function but are not osmolytes, including other amino acids (arginine, glutamine, and lysine), prominent intracellular cations (polyamines and sodium) and anions (acetate and citrate). Notably, the majority of these osmolytes and non-osmolytes are also employed as additives in formulations of protein therapeutics in order to extend shelf life and/or prevent aggregation (142).

3.2.2 Model Proteins

To investigate cosolute induced changes in protein stability, we chose the N-terminal SH3 domain of the *Drosophila* signal transduction protein drk, SH3. SH3 is meta stable with a dynamic equilibrium between the folded and unfolded populations and has served as a model system for protein folding investigations, including ^{19}F NMR experiments. Importantly, SH3 is destabilized in cells in response to osmotic stress, and the destabilization is alleviated by adding the osmolyte betaine (119, 137).

As a protein complex, we chose the B1 domain of the streptococcal immunoglobulin binding protein G (GB1). Wild type GB1 is monomeric, however, a single hydrophobic mutation, A34F, induces a side-by-side symmetric homodimer (143) that serves as a model system to examine protein oligomer stability (41). The stability of the dimer varies with weak attractive or repulsive surface electrostatic interactions, such as those that may accumulate during osmotic stress (144). Furthermore, the dissociation constant, K_D, of the GB1 dimer, >1 μM, is in the range of the transient protein-protein interactions that organize and regulate cells (134, 145). SH3 and GB1 are both acidic with pI values of 4.6 and 4.5, respectively, which

is representative of the majority of proteins in *Escherichia coli* and *Bacillus subtilis* (146, 147).

3.2.3 Quantifying Unfolding and Dimerization

^{19}F NMR was employed to measure the free energies of SH3 unfolding ($\Delta G^{\circ\prime}_{F\rightarrow U}$) and GB1 dimer dissociation ($\Delta G^{\circ\prime}_{D\rightarrow M}$). Fluorine is a convenient nucleus because^{19}F is 100% naturally abundant, rarely used in biology and ^{19}F NMR chemical shifts are sensitive to the local chemical environment (31). As the probe, we made use of a genetically encoded unnatural amino acid, 4-(trifluoromethyl)-L-Phenylalanine (TFMF, Fig. 1A Inset), which has been used for in-cell NMR (31, 148).

For SH3, TFMF was introduced in place of Tyr37 (*SI Appendix* Fig. S2), which is buried in the folded protein. The ^{19}F NMR spectra of SH3-Y37TFMF exhibits two resonances, one from the folded state at 61.96 ppm and the other from the unfolded ensemble at 61.80 ppm, (Fig. 1A). TFMF minimally impacts SH3 stability, reducing the midpoint temperature of denaturation in buffer from 36 °C to 32 °C (*SI Appendix*, Fig S3).

For GB1, we hypothesized that incorporation of TFMF (a Phe analog) in place of A34 (*SI Appendix* Fig. S2) would induce dimerization in a similar fashion to A34F and provide a probe of dimer formation (143). ^{15}N-^{1}H heteronuclear single quantum correlation (HSQC) spectra of the GB1-A34TFMF closely match spectra of GB1-A34F (*SI Appendix* Fig. S4) suggesting dimer formation is promoted and the overall structure is maintained. Small changes in shifts are observed for residues lining the hydrophobic dimer interface and are likely due to the increased size of the TFMF side chain with respect to Phe. These disruptions are reflected by a K_D (440

$\pm$ 10 μM) 7.5-times greater than the value for the A34F variant (41). The ^{19}F spectra of GB1-A34TFMF exhibits two concentration-dependent resonances (Figure 1B) indicative of the monomer (-61.90 ppm) and dimer (-62.05 ppm) populations.

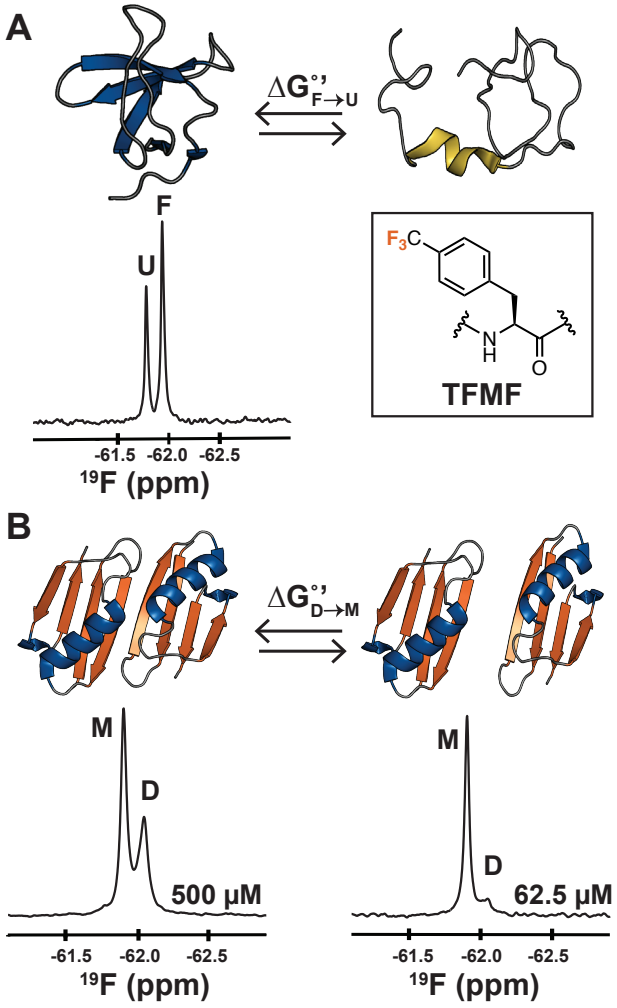


Figure 3.1 Quantification of protein- and dimer-stability by ^{19}F NMR

(*A*) TFMF-labeled SH3 displays resonances corresponding to the folded state (F) and the unfolded ensemble (U) at 298 K, pH 7.4. Inset: Structure of 4-(trifluoromethyl)-L-Phenylalanine (TFMF). (*B*) TFMF-labeled GB1 shows concentration-dependent dimer (D) and monomer (M) resonances at 298 K, pH 7.4.

3.2.4 Cosolute effects on unfolding and dissociation.

SH3-Y37TFMF and GB1-A34TFMF stability in the presence of cosolutes were quantified as described in Materials and Methods. Changes in stability with respect to buffer, $\Delta\Delta G^{\circ\prime}_{F\rightarrow U}$ and $\Delta\Delta G^{\circ\prime}_{D\rightarrow M}$ (summarized in *SI Appendix* Table S1) are defined such that increased stability results in a positive value and decreased stability is a negative value. For GB1-A34TFMF, $\Delta\Delta G^{\circ\prime}_{D\rightarrow M}$ was determined at a single concentration, 125 μM. For a subset of cosolutes, $\Delta\Delta G^{\circ\prime}_{D\rightarrow M}$ values were determined from binding isotherms and the results are in agreement with single concentration measurements. (*SI Appendix* Fig. S5 and Table S2).

With the exception of urea (a known denaturing osmolyte), all cosolutes stabilized SH3 (Fig. 2A). No trend was observed that distinguished osmolytes from non-osmolytes and some non-osmolytes (citrate, lysine, and acetate) provided the largest enhancements in stability. $\Delta\Delta G^{\circ\prime}_{D\rightarrow M}$ values for GB1 showed more variation than $\Delta\Delta G^{\circ\prime}_{F\rightarrow U}$ values of SH3. In addition, differences are observed in the performance of osmolytes with respect to non-osmolyte solutes (Fig. 2B). With the exception of glycine and glutamate, osmolytes minimally perturbed dimer stability, typically increasing $\Delta\Delta G^{\circ\prime}_{D\rightarrow M}$ by less than 1 kcal/mol. On the other hand, all non-osmolytes under investigation showed a more pronounced (1 – 2 kcal/mol) change in $\Delta\Delta G^{\circ\prime}_{D\rightarrow M}$ in favor of dimer formation. Notably, stabilization of the GB1 dimer qualitatively trends with the absolute charge of the cosolute, and species with the highest overall charge (either positive or negative) induced the greatest stabilization of the dimer.

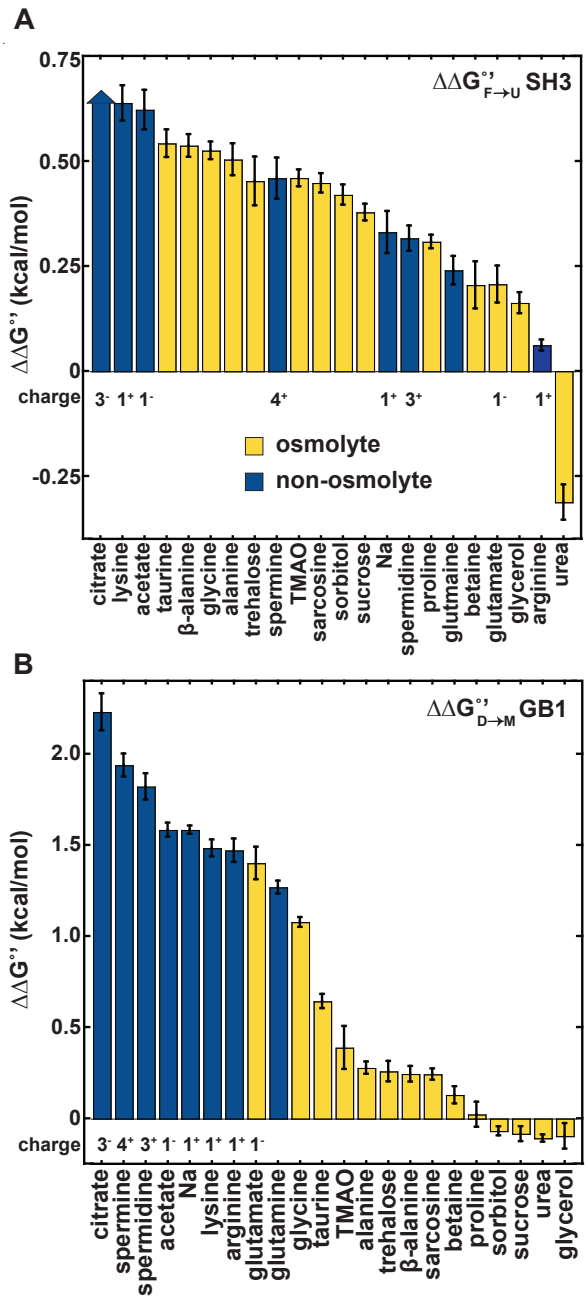


Figure 3.2 Osmolytes perturb neither unfolding nor dissociation

Changes in free energy ($\Delta\Delta G^{\circ\prime} = \Delta G^{\circ\prime}_{cosolute} - \Delta G^{\circ\prime}_{buffer}$) for SH3 unfolding ($\Delta\Delta G^{\circ\prime}_{F\rightarrow U}$, panel *A*) and GB1 dissociation ($\Delta\Delta G^{\circ\prime}_{D\rightarrow M}$, panel *B*) were determined in the presence or absence of 400 mM cosolute. Osmolytes are shown in yellow and non-osmolytes are shown in blue. Glutamine was used at 200 mM and glutamate at 60 mM because of solubility. Error bars represent the standard deviation of the mean from three independent protein preparations. Quantification of $\Delta\Delta G^{\circ\prime}_{F\rightarrow U}$ for SH3 in citrate was not possible because there was no resonance from the unfolded form, and is reported as greater than the most stabilizing quantifiable cosolute, lysine.

3.2.5 Electrostatic contributions to $\Delta\Delta G^{\circ'}_{D\rightarrow M}$

The stability of monomeric GB1 increases concomitant with increasing pH in living E. *coli* cells (136). This observation can be explained in terms of surface charge. GB1 has a net charge of −4 at pH 7.4. As the surface of GB1 transitions from anionic to cationic, there are more favorable electrostatic interactions between negatively-charged cytosolic E. *coli* macromolecules and GB1 (149). In addition, GB1 dimerization is driven by a hydrophobic interaction that overcomes the self-repulsion between two like-charged monomers. Given these observations and the stabilization of the GB1 dimer in the presence of charged cosolutes, we investigated the contribution of electrostatic interactions to dimer stability.

^{19}F spectra of the GB1 dimer were acquired in buffer between pH 7 and pH 4. A shift from predominately monomeric- to predominately dimeric- GB1 was observed with decreasing pH (*SI Appendix*, Fig. S6). In addition, a small amount of a higher-order aggregate, as characterized by dynamic light scattering (*SI Appendix*, Fig. S7), formed at pH values less than 6. Furthermore, the population of dimer correlates positively with the concentration of NaCl (*SI Appendix*, Fig S8). These data suggest that the stability of the GB1 dimer is strongly influenced by electrostatic interactions and that Coulombic screening of repulsive charges promotes association. We therefore evaluated the effect of a subset of cosolutes, both osmolytes and non-osmolytes, at a lower pH value where electrostatic repulsion would be minimized due to neutralization of acidic surface side chains (see Materials and Methods). At pH 4, the ^{19}F spectra indicated mostly (~ 75 %) dimer and higher-order species, consistent with the attenuation of repulsive charge-charge interactions. The

population of dimer at this pH value was unaffected by adding cosolutes, suggesting that these cosolutes have a negligible influence on the equilibrium between dimeric and monomeric GB1 (*SI Appendix*, Fig S9 and Table S3).

3.3 Discussion

Osmolytes are traditionally characterized as protein stabilizing agents (150). Our data suggest that this characterization is too narrow. Small molecules, related to but not utilized as osmolytes, stabilize SH3 within the range of the least and most powerful osmolytes. Stabilization by virtue of preferential exclusion does not preclude non-specific interactions with protein surfaces (151). Therefore, we suggest that nature selected osmolytes from those intracellular cosolutes that favorably impact protein stability, while avoiding interactions with protein surfaces. Comparing glycine to alanine and betaine elucidates some subtleties of the structural requirements for protein surface compatibility. Adding a methyl side chain to glycine to form alanine decreases dimer stabilization by 0.8 kcal/mol, and trimethylation of the glycine amine decreases stabilization to an even greater extent. The dipole of osmolytes such as proline and TMAO are sterically restricted from interaction with protein dipoles, and themselves, by hydrophobic substituents (152). We suggest that removing the hydrophobic bulk of alanine or betaine facilitates the interaction of glycine with protein surfaces.

Glycine and glutamate show disparate influences on GB1 dimer stability when compared to other osmolytes. Potassium glutamate is the predominate osmolyte of *E. coli* and *Salmonella typhimurium (153)*. These organisms rapidly accumulate potassium from their surroundings in response to osmotic stress and

subsequently synthesize glutamate to provide a counter ion. Potassium glutamate is subsequently replaced in the cell by other compatible solutes such as trehalose, proline or betaine (153). The definition of glycine as an osmolyte seems to arise from its high concentrations in the tissues of marine organisms (154, 155), which are specialized in their ability to survive in high concentrations of salt. To the best of our knowledge, the only other known occurrence of glycine accumulation in response to osmotic stress is in early-stage mouse embryos by action of a Na^+/Cl^--dependent glycine transporter, GLYT1 (156). These data suggest that glutamate and glycine have specialized roles in specific organisms and that it may not be advantageous to classify them as compatible osmolytes.

Osmolyte effects on enzyme tertiary structure and steady state kinetics (113, 120, 157) is well characterized and the molecular mechanism of osmolyte-induced protein stabilization is known (116, 158, 159). In stark contrast, the influence of osmolytes on protein-protein interactions and consequently, protein surfaces, has not been established. Electrostatic interactions are a fundamental component of biomolecular interactions. The charge of protein interfaces has been fine-tuned by evolution to enhance the orientation, stability, specificity, and rate of association of protein-partner recognition in a manner that is specific to the pH associated with their subcellular location (160-164). Disrupting these interactions, via pH, salt concentration or mutation, affects protein complex stability (165). Analogously, our data suggest that osmolytes are ‘compatible’ with protein-protein interactions because they minimally impact electrostatic interactions of protein surfaces.

Life inherently depends on the distribution and interaction of macromolecules, metabolites and ions. Spitzer, Poolman, McConkey and Srere attribute the organization of the cytoplasm to weak, transient electrostatic interactions between charged surfaces and metabolites (149, 166, 167). Moreover, the biological importance of protein surfaces has been recently underscored by reports showing that attractive electrostatic interactions are destabilizing to protein structure (135-137, 168). The transient interactions that organize the cellular interior must have factored into the selection of osmolytes by nature. Accordingly, in addition to their stabilizing influence on protein tertiary structure, the compatibility of osmolytes with quaternary and quinary interactions is a critical feature that enables osmolytes to maintain and restore cellular integrity, structure and function in response to environmental stress.

3.4 Materials and Methods

3.4.1 Materials

Antibiotics, osmolytes and other cosolutes were used without further purification. Preformulated Luria-Bertani and 2xYT media were acquired from Fisher Scientific. 4-(trifluoromethyl)-L-Phenylalanine (TFMF) was purchased from Chem-Impex International. Restriction enzymes were obtained from New England Biolabs. pUltra-CNFRS was obtained from Addgene. pH values are direct meter readings, uncorrected for the D_2O isotope effect.

3.4.2 Protein

Mutagenesis was performed *via* restriction ligation of *BsaI* digested PCR fragments generated from primers described in the *SI Appendix*, Supplemental Text.

Agilent BL21(DE3)-Gold cells were co-transformed with pET11a-SH3-Y37TAG or -GB1-T2Q-A34TAG and pUltra-pCNFRS, a plasmid harboring an orthogonal 4-cyano-*L*-phenylalanine tRNA synthetase and $tRNA_{CUA}^{MjTyr}$. The 4-cyano-*L*-phenylalanine tRNA synthetase is permissive for TFMF (95). The T2Q variant of GB1 was included to prevent proteolytic degradation (165). A single colony was used to inoculate 5 mL of LB media containing 100 μg/L ampicillin to maintain pET and 50 μg/L streptomycin to maintain pUltra. The culture was grown to saturation overnight at 37 °C with shaking (New Brunswick Scientific Innova I26, 225 rpm). The next morning, 1 L of 2xYT media, supplemented with 100 μg/L ampicillin and 50 μg/L streptomycin, was inoculated with the overnight culture. When the culture reached an optical density at 600 nm (OD_{600}) of 0.3, TFMF was added to final concentration of 1 mM. Expression was induced at OD_{600} 0.6 by adding isopropyl β-D-1-thiogalactopyranoside to a final concentration of 1 mM. After 5 h, the cells were pelleted for 30 mins at 1000*g*, the supernatant discard and the pellet stored at -20 °C.

The pellet was thawed at room temperature and resuspended in 25 mL of 20 mM tris, pH 7.5 for GB1 or 50 mM tris, pH 7.5 for SH3, supplemented with a Roche cOmplete protease inhibitor tablet. Cells were sonicated using a Fisher Scientific Sonic Dismembrator model 500 for 8 min at 20% amplitude at a 39% duty cycle. Lysates were clarified by centrifugation at 27,000*g* for 1 h and passed through a 0.22 μm filter. GB1 and SH3 were purified as described (41, 137) and their purity confirmed by SDS-PAGE and Q-TOF mass spectrometry. The purified proteins were

exchanged into H_2O by dialysis, divided into 500 µL aliquots of 500 µM GB1-A34TFMF or 50 µM SH3-Y37TFMF, lyophilized and stored at -20°C.

3.4.3 NMR

Stock cosolute solutions (1.0 M) were prepared in the corresponding buffers (*vide infra*) and the pH adjusted to 7.4 with HCl or NaOH. For SH3, an aliquot was thawed at room temperature and resuspended to 50 µM using 50 mM HEPES, 50 mM citrate, 50 mM bis tris propane, pH 7.4. NMR samples were prepared by combining this stock solution (250 µL) with 200 µL of stock cosolute solution in the same buffer and 50 µL of D_2O to given a final SH3 concentration of 25 µM. Experiments were performed in triplicate. Spectra were acquired at 298 K on a Bruker Avance III HD spectrometer at a ^{19}F Larmor frequency of 470 MHz with a cryogenic QCI probe and an H/F channel. The free energy of unfolding ($\Delta G^{o\prime}_{F\rightarrow U}$) was determined using the following equation,

$$\Delta G^{o\prime}_{F\rightarrow U} = -RTln(\frac{U}{F}).$$

where *U* and *F* are the areas of the corresponding resonances, *R* is the gas constant and *T* is the absolute temperature.

Aliquots of GB1 were thawed at room temperature and diluted to 250 µM with 20 mM Na_2HPO_4/NaH_2PO_4, pH 7.4. NMR samples were prepared by combinding this stock solution with 200 µL of the 1.0 M cosolute solution in the same buffer and 50 µL of D_2O added for a final GB1 concentration of 125 µM. $\Delta G^{\circ\prime}_{D\rightarrow M}$, was determined, in triplicate, by using the equation,

$$\Delta G^{\circ\prime}_{D\rightarrow M} = -RTln(\frac{2P_T F_M^2}{F_D})$$

where F_M and F_D, the fractions of monomer and dimer, respectively, were calculated from the areas of the corresponding resonances, and P_T, the total protein concentration, was 125 μM.

For a subset of co-solutes, the dissociation constant, K_{Dis}, was also obtained by serial dilution of a 500 μM sample at five concentrations with and without cosolute, in triplicate, and the data fit to the following equation

$$F_D = \frac{4P_T + K_{Dis} - \sqrt{K_D^2 + 8P_T K_{Dis}}}{4P_T}$$

using MATLAB version R2014b. $\Delta G^{\circ\prime}_{D\rightarrow M}$ for these experiments was calculated using the equation,

$$\Delta G^{\circ\prime}_{D\rightarrow M} = -RTln(K_{Dis}).$$

Samples for the pH dependence experiments were prepared as described above, but protein and cosolute were resuspended in 50 mM HEPES, 50 mM citrate, 50 mM bis-tris propane buffer with and without cosolute and adjusted to the desired pH using HCl. The fraction of dimer plus aggregate was determined by summing the areas of the dimer and aggregate resonances and dividing by the sum of the areas of the dimer, aggregate and monomer resonances.

3.5 Supporting Information

3.5.1 Supplementary Text

Primers for *BsaI* mutagenesis

GB1-T2Q-A34TAG F:

*GATGTT***GGTCTC**AGTACTAGAACGACAACGGTGTTG

GB1-T2Q-A34TAG R:

*GATGTT***GGTCTC**AGTACTGTTTGAAAACTTTTTCCGCG

GB1-T2Q-TAA STOP F:

*GATGTT***GGTCTC**ACGAATAAGATCCGGCTGCTAAC

GB1-T2Q-TAA STOP R:

*GATGTT***GGTCTC**ATTCGGTTACCGTGAAGG

Overhangs are italicized, *BsaI* cut sites are bold and codon mutations introduced are red.

3.5.2 Supplementary Tables

Table S1 Cosolute induced changes in SH3 and GB1 stability (298 K, pH 7.4)

cosolute (400 mM)[a]	$\Delta\Delta G^{\circ\prime}_{D\rightarrow M}$ **(kcal/mol) GB1-A34TFMF**	$\Delta\Delta G^{\circ\prime}_{F\rightarrow U}$ **(kcal/mol) SH3-Y37TFMF**
citrate	2.2±0.1	NA[d]
spermine	1.94 ± 0.06	0.46 ± 0.05
spermidine	1.82 ± 0.07	0.32 ± 0.03
acetate	1.58±0.04	0.62±0.05
NaCl	1.58 ± 0.02	0.33 ± 0.05
lysine	1.48 ± 0.05	0.64 ± 0.04
arginine	1.47 ± 0.06	0.06 ± 0.01
Glutamate[b]	1.40±0.09	0.21±0.04
Glutamine[c]	1.27 ± 0.04	0.24 ± 0.03
glycine	1.08 ± 0.03	0.52 ± 0.02
taurine	0.64 ± 0.04	0.54 ±0.03
TMAO	0.4 ± 0.1	0.46 ± 0.02
alanine	0.28 ± 0.03	0.50 ± 0.04
trehalose	0.26 ± 0.06	0.45 ± 0.05
β-alanine	0.24 ± 0.04	0.54 ± 0.03
sarcosine	0.24 ± 0.03	0.45 ± 0.02
betaine	0.13 ± 0.05	0.21 ± 0.06
proline	0.02 ± 0.07	0.31 ±0.02
sorbitol	-0.07 ± 0.03	0.42 ± 0.03
sucrose	-0.08 ± 0.04	0.38 ± 0.02
glycerol	-0.10 ± 0.07	0.16 ± 0.03
urea	-0.11 ± 0.02	-0.31 ± 0.04

Footnotes
[a] Abbreviations: TMAO, trimethylamine-N-oxide
[b]Glutamate was used at 60 mM
[c]Glutamine was used at 200 mM
[d]NA, not applicable

Table S2 GB1-A34TFMF stability from binding isotherms

cosolute[a,b] (400 mM)	$\Delta G^{\circ\prime}_{D\rightarrow M}$ **[c] (kcal/mol)**	K_{Dis} **[d] (μM)**	$\Delta\Delta G^{\circ\prime}_{D\rightarrow M}$ **(kcal/mol)**
buffer	4.57±0.01	442±11	NA[e]
lysine	6.21±0.04	28±2	1.63±0.04
TMAO	4.86±0.03	272±15	0.29±0.03
proline	4.61±0.04	416±29	0.04±0.04
urea	4.42±0.04	574±39	-0.15±0.04

Footnotes

[a] Abbreviations: TMAO, trimethylamine-N-oxide

[b] 298 K, 20 mM KH_2PO_4/K_2HPO_4, pH 7.4

[c] Error was determined using 1×10^6 iterations of a Monte Carlo analysis using the average and standard deviation of each condition

[d] Error was assessed using a thousand interactions of a Monte Carlo analysis

[e] NA, not applicable

Table S3 Cosolute effect on $\mathbf{F_{D+A}}$ of GB1-A34TFMF pH 4.0.

cosolute[a] (400 mM)	F_{D+A}
TMAO	0.79 ± 0.01
lysine	0.81 ± 0.02
proline	0.81 ± 0.02
urea	0.830 ± 0.005
trehalose	0.83 ± 0.02
glycine	0.86 ± 0.02
buffer (no cosolute)	0.801 ± 0.002

Footnotes
[a] Abbreviations: TMAO, trimethylamine-N-oxide

3.5.3 Supplementary Figures

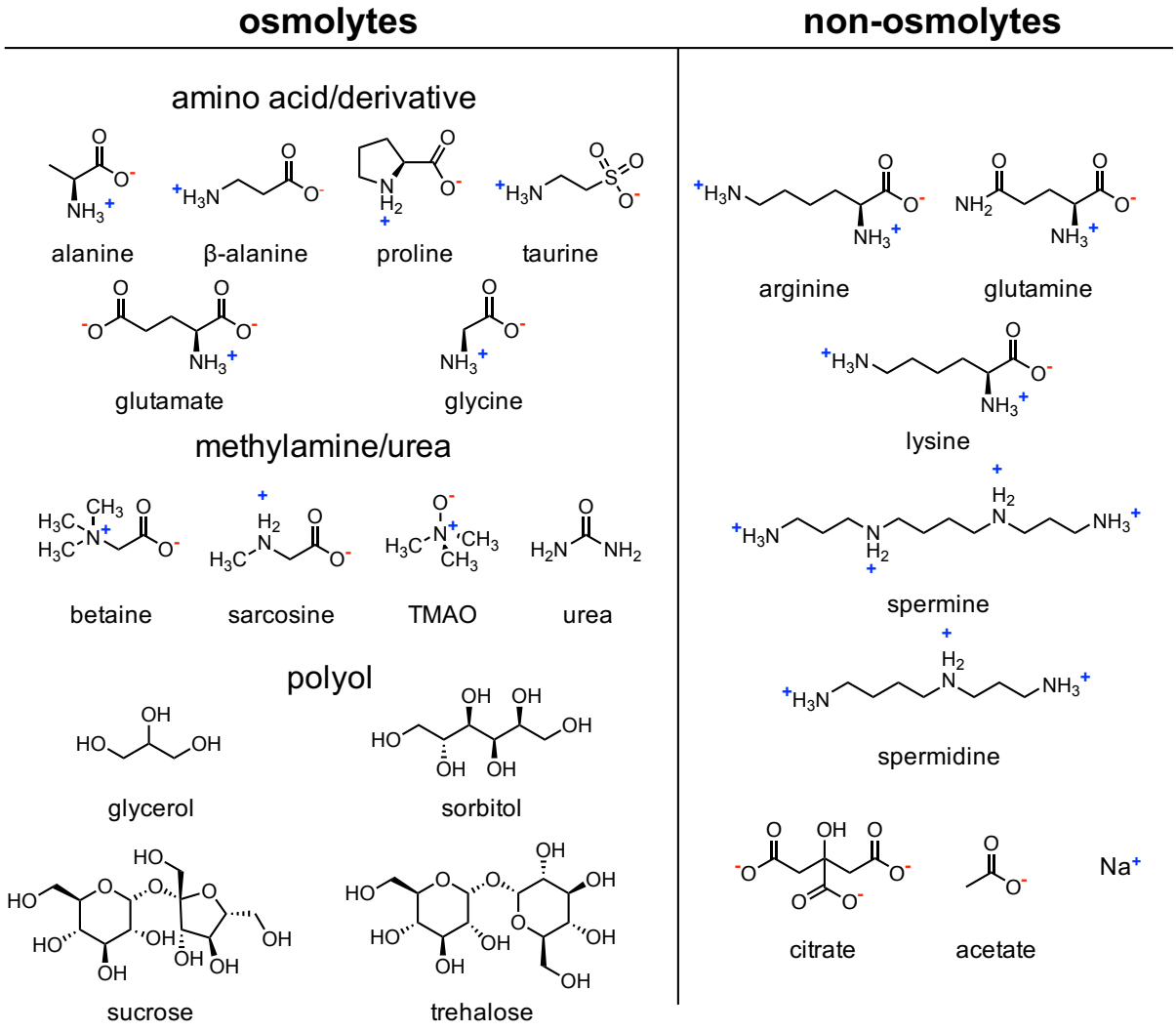


Figure S1 Cosolute structures

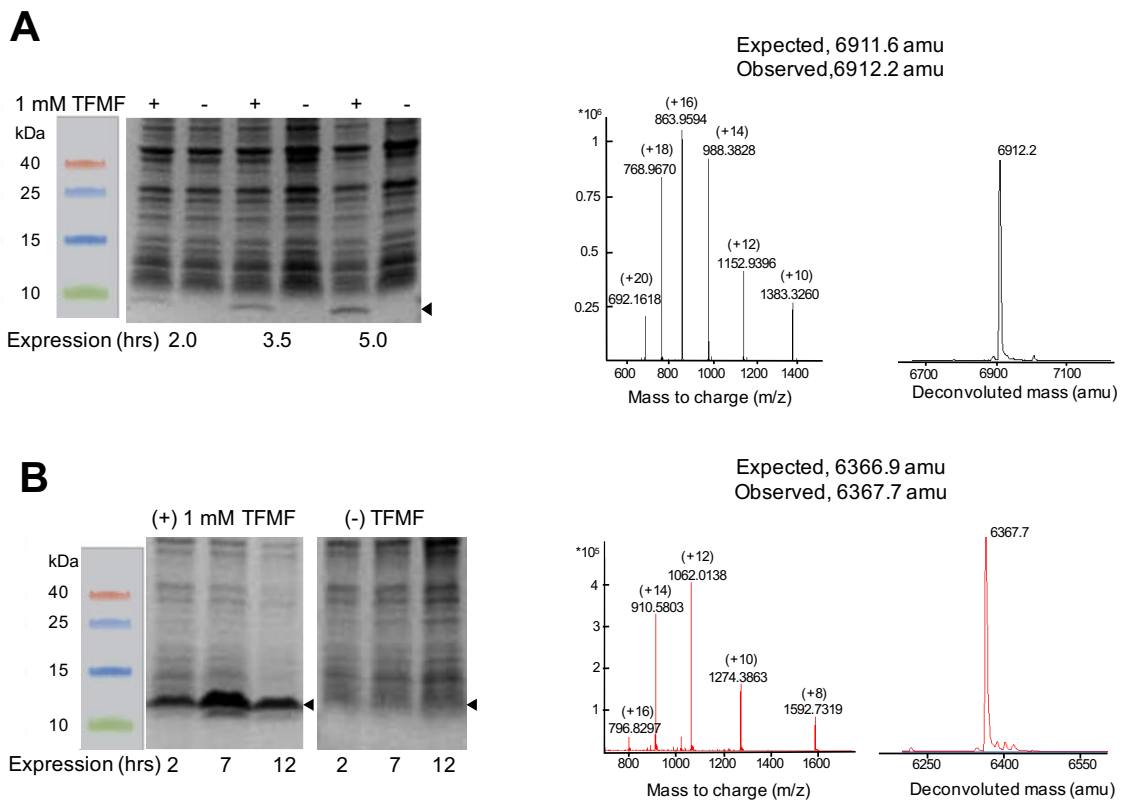


Figure S2 Selective incorporation of TFMF

Selective *in vivo* incorporation of TFMF by 4-cyano-L-phenylalanine synthetase. Expression of SH3-Y37TFMF (*A*) and GB1-A34TFMF (*B*) in the presence (+) and absence of (-) of TFMF over time verified by ESI-QToF mass spectrometry.

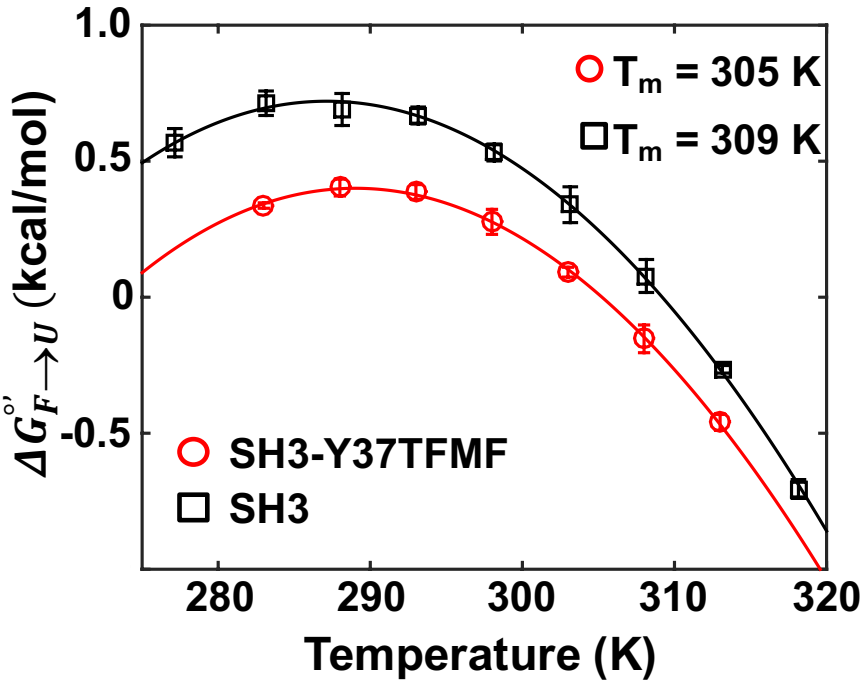


Figure S3 Thermal Stability of SH3 and SH3-Y37TFMF by ^{19}F NMR.

Stability was measured at 25 μM SH3 **(137)** and SH3-Y37TFMF in 50 mM HEPES, 50 mM citrate, 50 mM bis tris propane, pH 7.4.

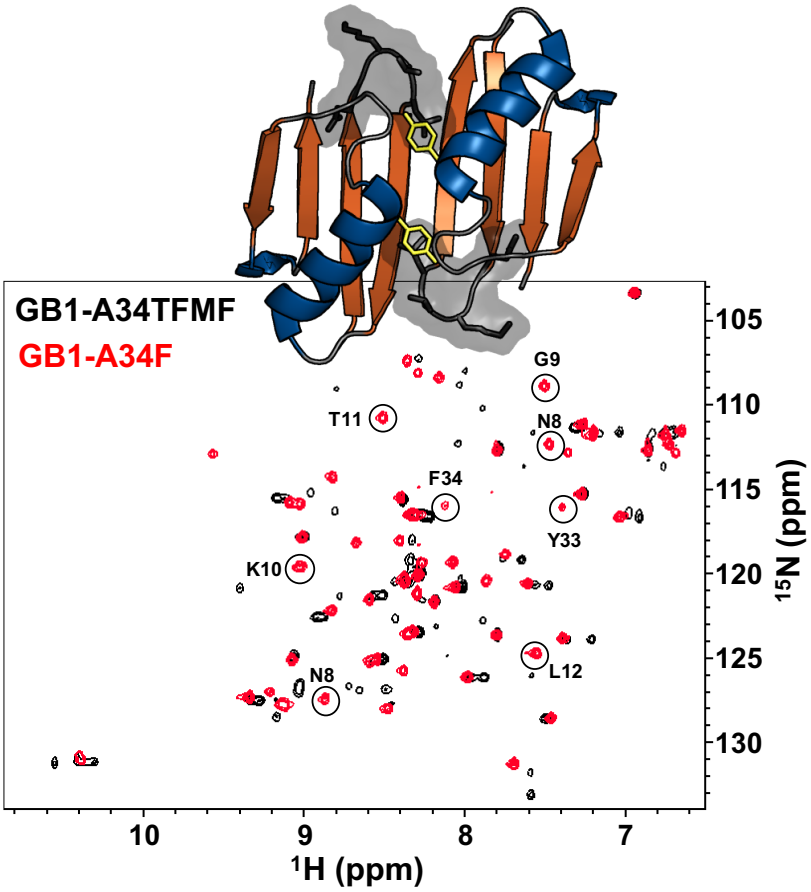


Figure S4 ^{1}H-^{15}N HSQC spectrum of GB1-A34TFMF

Overlay of ^{1}H-^{15}N HSQC spectrum 500 μM GB1-A34TFMF and 500 μM GB1-A34F in 20 mM KH_2PO_4/K_2HPO_4 buffer, pH 7 at 298 K. Spectra of GB1-A34TFMF, black, is in close agreement with spectra of GB1-A34F, red (41). Residues hypothesized to weaken dimer affinity, 8-12, Y33 and F34 are circled in the A34F spectra. The loop (residues 8-12) that participates in hydrophobic burial of Y33, yellow, is shown in gray on GB1 dimer structure.

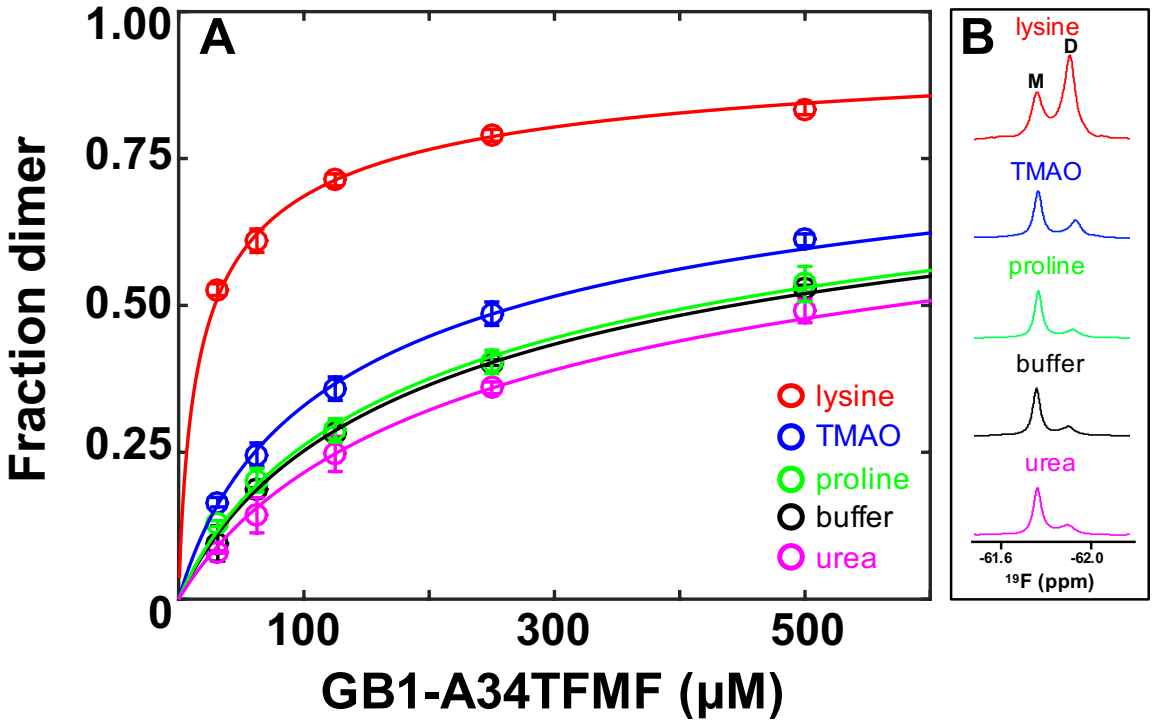


Figure S5 Binding isotherms of GB1-A34TFMF

GB1 binding isotherms in presence of representative cosolutes (298 K, 20 mM KH_2PO_4/K_2HPO_4, pH 7.4). (*A*) Fraction dimer is calculated from the area of monomer (M) and dimer (D) resonances shown panel B at five GB1-A34TFMF concentrations. Error bars indicate the standard deviation from three independent protein preparations. (*B*) ^{19}F spectra of 125 μM GB1-A34TFMF in buffer (20 mM KH_2PO_4/K_2HPO_4, pH 7.4) and buffer plus 400 mM cosolute.

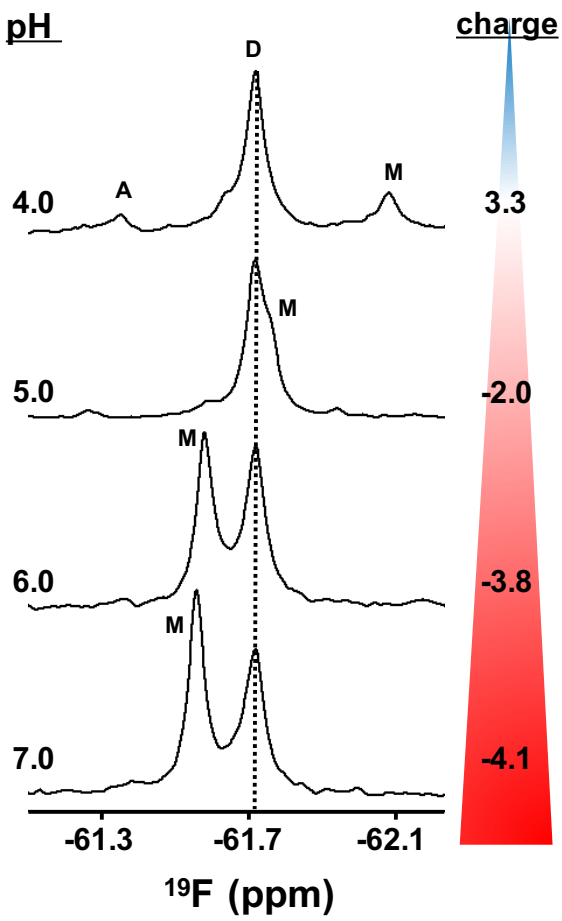
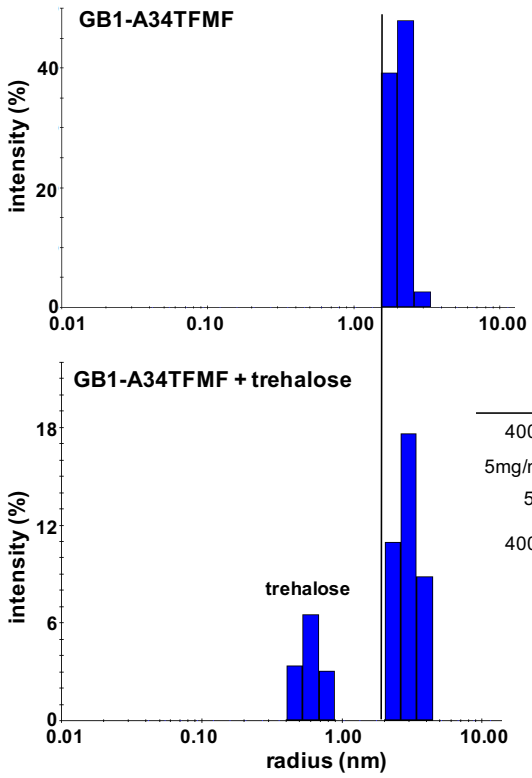


Figure S6 ^{19}F NMR spectra of GB1-A34TFMF at various pH values

^{19}F NMR spectra at 298 K of 125 μM GB1-A34TFMF in 50 mM HEPES, 50 mM citrate, 50 mM bis-tris propane at various pH values. Spectra are aligned by dimer resonance. Right hand scale indicates net charge of GB1-A34TFMF at each pH value. M = monomer; D = dimer; A = aggregate.



	Radius (nm)	% Polydispersity	Molecular weight (kDa)
400 mM trehalose	0.6	13.7	1
5mg/ml GB1-A34TFMF	2.1	14.8	19
5 mg/ml GB1-A34TFMF 400 mM trehalose	3.0	19.5	45

Figure S7 DLS characterization of GB1 aggregate at pH 4.0.

Dynamic light scattering characterization of GB1-A34TFMF aggregate in 50 mM HEPES, 50 mM citrate, 50 mM bis-tris propane buffer, pH 4.0. Vertical line visualizes increase in radius (nm) after addition of trehalose.

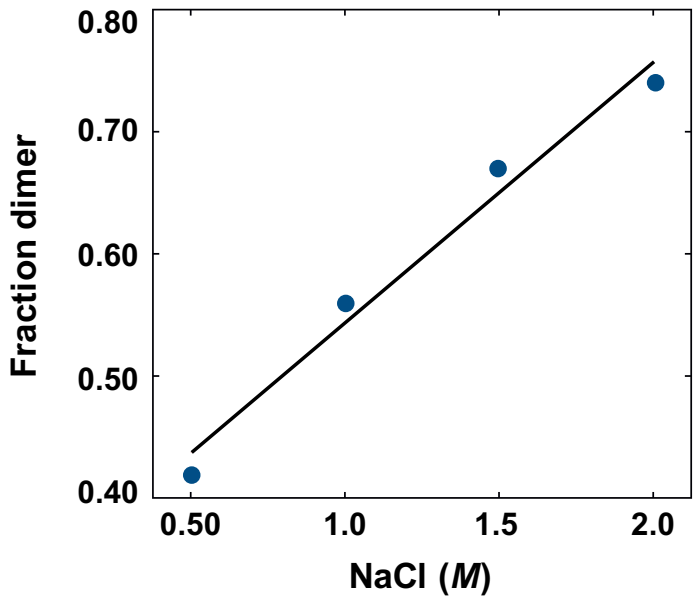


Figure S8 NaCl induced changes in fraction dimer

125 μM GB1-A34TFMF at 298 K in 20 mM KH_2PO_4/K_2HPO_4, pH 7.4 with increasing concentration of sodium chloride.

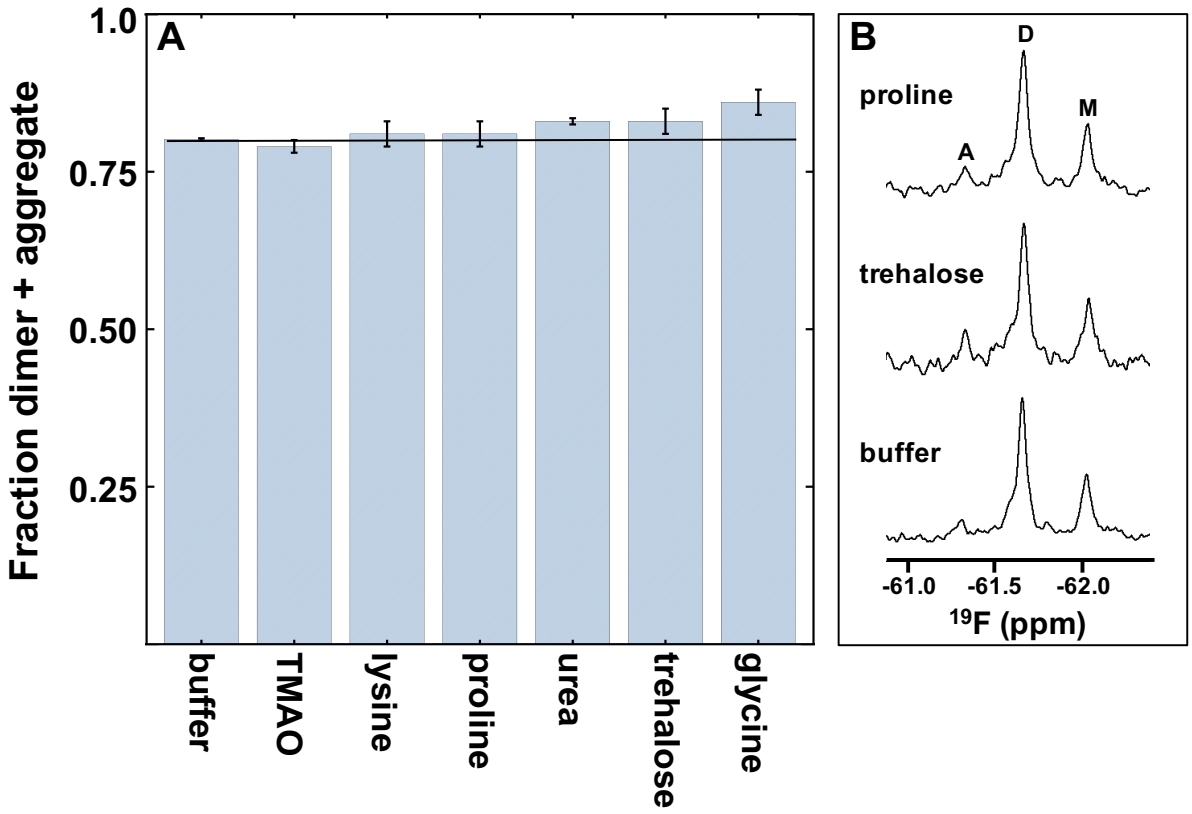


Figure S9 Cosolute induced changes in GB1 oligomerization at pH 4.0

Cosolute (400 mM) induced changes in oligomerization of 50 μM GB1-A34TFMF in 50 mM HEPES, 50 mM citrate, 50 mM bis-tris propane buffer, pH 4.0. (*A*) The effect of cosolutes on the fraction dimer + aggregate. Error bars represent the standard deviation from three independent protein preparations. A horizontal line is drawn to visualize deviations from buffer. (*B*) The ^{19}F spectra indicate three species at 50 μM. Resonances at -61.7 ppm and -62.0 ppm were assigned as the dimer and monomer, respectively, by dilution and monitoring the chemical shift versus pH. The third peak at -61.3 ppm is a higher order aggregate characterized by dynamic light scattering experiments, Figure S7. Addition of stabilizing cosolute, particularly trehalose promotes formation of aggregate. M = monomer; D = dimer; A = aggregate.

REFERENCES

1. Ravikumar Y, Nadarajan SP, Yoo TH, Lee CS, & Yun H (2015) Unnatural amino acid mutagenesis-based enzyme engineering. *Trends Biotechnol* 33(8):462-470.

2. Dumas A, Lercher L, Spicer CD, & Davis BG (2015) Designing logical codon reassignment - expanding the chemistry in biology. *Chem Sci* 6(1):50-69.

3. England PM (2004) Unnatural amino acid mutagenesis: A precise tool for probing protein structure and function. *Biochemistry* 43(37):11623-11629.

4. Wang L, Xie J, & Schultz PG (2006) Expanding the genetic code. *Annu Rev Biophys Biomol Struct* 35:225-249.

5. Chin JW (2014) Expanding and reprogramming the genetic code of cells and animals. *Annu Rev Biochem* 83:379-408.

6. Anthony-Cahill SJ, Griffith MC, Noren CJ, Suich DJ, & Schultz PG (1989) Site-specific mutagenesis with unnatural amino acids. *Trends Biochem Sci* 14(10):400-403.

7. Young TS & Schultz PG (2010) Beyond the canonical 20 amino acids: Expanding the genetic lexicon. *J Biol Chem* 285(15):11039-11044.

8. Zhang WH, Otting G, & Jackson CJ (2013) Protein engineering with unnatural amino acids. *Curr Opin Struct Biol* 23(4):581-587.

9. Budisa N, Wenger W, & Wiltschi B (2010) Residue-specific global fluorination of Candida antarctica lipase B in Pichia pastoris. *Mol Biosyst* 6(9):1630-1639.

10. Davis L & Chin JW (2012) Designer proteins: Applications of genetic code expansion in cell biology. *Nat Rev Mol Cell Biol* 13(3):168-182.

11. Bar-Even A*, et al.* (2011) The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50(21):4402-4410.

12. Radzicka A & Wolfenden R (1995) A proficient enzyme. *Science* 267(5194):90-93.

13. Grammel M & Hang HC (2013) Chemical reporters for biological discovery. *Nat Chem Biol* 9(8):475-484.

14. Yang H, Srivastava, P., Zhang, C., Lewis, J. C. (2014) A general method for artificial metalloenzyme formation through strain-promoted azide–alkyne cycloaddition. *ChemBiochem Communications* 15:223-227.

15. Srivastava P, Yang H, Ellis-Guardiola K, & Lewis JC (2015) Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. *Nat Commun* 6:7789.

16. Hyster TK, Knorr L, Ward TR, & Rovis T (2012) Biotinylated Rh(III) complexes in engineered streptavidin for accelerated asymmetric C-H activation. *Science* 338(6106):500-503.

17. Yang H*, et al.* (2018) Evolving artificial metalloenzymes *via* random mutagenesis. *Nat Chem* 10(3):318-324.

18. Bornscheuer UT*, et al.* (2012) Engineering the third wave of biocatalysis. *Nature* 485(7397):185-194.

19. Bonetta L (2010) Protein-protein interactions: Interactome under construction. *Nature* 468(7325):851-854.

20. Pham ND, Parker RB, & Kohler JJ (2013) Photocrosslinking approaches to interactome mapping. *Curr Opin Chem Biol* 17(1):90-101.

21. Chin JW, Martin AB, King DS, Wang L, & Schultz PG (2002) Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc Natl Acad Sci U S A* 99(17):11020-11024.

22. Kanamori T, Nishikawa S, Shin I, Schultz PG, & Endo T (1997) Probing the environment along the protein import pathways in yeast mitochondria by site-specific photocrosslinking. *Proc Natl Acad Sci U S A* 94(2):485-490.

23. Ye Z, Bair M, Desai H, & Williams GJ (2011) A photocrosslinking assay for reporting protein interactions in polyketide and fatty acid synthases. *Mol Biosyst* 7(11):3152-3156.

24. Chin JW*, et al.* (2002) Addition of p-azido-l-phenylalanine to the genetic code of *Escherichia coli*. *J Am Chem Soc* 124(31):9026-9027.

25. Farrell IS, Toroney R, Hazen JL, Mehl RA, & Chin JW (2005) Photo-cross-linking interacting proteins with a genetically encoded benzophenone. *Nat Methods* 2(5):377-384.

26. Tippmann EM, Liu W, Summerer D, Mack AV, & Schultz PG (2007) A genetically encoded diazirine photocrosslinker in *Escherichia coli*. *Chembiochem* 8(18):2210-2214.

27. Mori H & Ito K (2006) Different modes of SecY-SecA interactions revealed by site-directed *in vivo* photo-cross-linking. *Proc Natl Acad Sci U S A* 103(44):16159-16164.

28. Merkel L, Schauer M, Antranikian G, & Budisa N (2010) Parallel incorporation of different fluorinated amino acids: On the way to "teflon" proteins. *Chembiochem* 11(11):1505-1507.

29. Arntson KE & Pomerantz WC (2016) Protein-observed fluorine NMR: A bioorthogonal approach for small molecule discovery. *J Med Chem* 59(11):5158-5171.

30. Cellitti SE*, et al.* (2008) *In vivo* incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J Am Chem Soc* 130(29):9268-9281.

31. Jackson JC, Hammill JT, & Mehl RA (2007) Site-specific incorporation of a (19)F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J Am Chem Soc* 129(5):1160-1166.

32. Drienovska I, Rioz-Martinez A, Draksharapu A, & Roelfes G (2015) Novel artificial metalloenzymes by i*n vivo* incorporation of metal-binding unnatural amino acids. *Chem Sci* 6(1):770-776.

33. Lewis JC (2015) Metallopeptide catalysts and artificial metalloenzymes containing unnatural amino acids. *Curr Opin Chem Biol* 25:27-35.

34. Schwizer F*, et al.* (2018) Artificial metalloenzymes: Reaction scope and optimization strategies. *Chem Rev* 118(1):142-231.

35. Reetz MT*, et al.* (2008) A robust protein host for anchoring chelating ligands and organocatalysts. *Chembiochem* 9(4):552-564.

36. Bos J & Roelfes G (2014) Artificial metalloenzymes for enantioselective catalysis. *Curr Opin Chem Biol* 19:135-143.

37. Sharaf NG, Xi Z, Ishima R, & Gronenborn AM (2017) The HIV-1 p66 homodimeric RT exhibits different conformations in the binding-competent and -incompetent NNRTI site. *Proteins* 85(12):2191-2197.

38. Shi P*, et al.* (2012) *In situ* 19F NMR studies of an *E. coli* membrane protein. *Protein Sci* 21(4):596-600.

39. Shi P*, et al.* (2011) Site-specific (1)(9)f NMR chemical shift and side chain relaxation analysis of a membrane protein labeled with an unnatural amino acid. *Protein Sci* 20(1):224-228.

40. Shi P*, et al.* (2010) Site-specific protein backbone and side-chain NMR chemical shift and relaxation analysis of human vinexin SH3 domain using a genetically encoded 15N/19F-labeled unnatural amino acid. *Biochem Biophys Res Commun* 402(3):461-466.

41. Guseman AJ & Pielak GJ (2017) Cosolute and crowding effects on a side-by-side protein dimer. *Biochemistry* 56(7):971-976.

42. Li C*, et al.* (2009) 19F NMR studies of alpha-synuclein conformation and fibrillation. *Biochemistry* 48(36):8578-8584.

43. Marsh EN & Suzuki Y (2014) Using (19)f NMR to probe biological interactions of proteins and peptides. *ACS Chem Biol* 9(6):1242-1250.

44. Lad C, Williams NH, & Wolfenden R (2003) The rate of hydrolysis of phosphomonoester dianions and the exceptional catalytic proficiencies of protein and inositol phosphatases. *Proc Natl Acad Sci U S A* 100(10):5607-5610.

45. Yang G & Ding Y (2014) Recent advances in biocatalyst discovery, development and applications. *Bioorg Med Chem* 22(20):5604-5612.

46. Durrenberger M & Ward TR (2014) Recent achievments in the design and engineering of artificial metalloenzymes. *Curr Opin Chem Biol* 19:99-106.

47. Kries H, Blomberg R, & Hilvert D (2013) *De novo* enzymes by computational design. *Curr Opin Chem Biol* 17(2):221-228.

48. Renata H, Wang ZJ, & Arnold FH (2015) Expanding the enzyme universe: Accessing non-natural reactions by mechanism-guided directed evolution. *Angew Chem Int Ed Engl* 54(11):3351-3367.

49. Wilson MEW, George M. (1978) Conversion of a protein to a homogeneous asymmetric hydrogenation catalyst by site-specific modification with a diphosphinerhodium(i) moiety. *J Am Chem Soc* 100:306-307.

50. Oelerich J, Gerard Roelfes (2013) DNA-based asymmetric organometallic catalysis in water. *Chem Sci* 4(5):2013-2017.

51. Bos J (2012) Enantioselective artificial metalloenzymes by creation of a novel active site at the protein dimer interface. *Angew Chem Int Ed Engl* 51(30):7472-7475.

52. Coelho PS, Brustad EM, Kannan A, & Arnold FH (2013) Olefin cyclopropanation *via* carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* 339(6117):307-310.

53. Jiang L*, et al.* (2008) *De novo* computational design of retro-aldol enzymes. *Science* 319(5868):1387-1391.

54. McIntosh JA, Farwell CC, & Arnold FH (2014) Expanding P450 catalytic reaction space through evolution and engineering. *Curr Opin Chem Biol* 19:126-134.

55. Gober JG & Brustad EM (2016) Non-natural carbenoid and nitrenoid insertion reactions catalyzed by heme proteins. *Curr Opin Chem Biol* 35:124-132.

56. Lee HS & Schultz PG (2008) Biosynthesis of a site-specific DNA cleaving protein. *J Am Chem Soc* 130(40):13194-13195.

57. List BL, R. A. Barbas, C. F. (2000) Proline-catalyzed direct asymmetric aldol reactions. *J Am Chem Soc* 122:2395-2396.

58. Ahrendt KA, Borths, Christopher J. , MacMillan, David W. C. (2000) New strategies for organic catalysis: The first highly enantioselective organocatalytic Diels-Alder reaction. *J Am Chem Soc* 122:4243-4244.

59. MacMillan DW (2008) The advent and development of organocatalysis. *Nature* 455(7211):304-308.

60. List B (2012) Organocatalysis. *Beilstein J Org Chem* 8:1358-1359.

61. Pan SC & List B (2007) New concepts for organocatalysis. *Ernst Schering Found Symp Proc* (2):1-43.

62. Seayad J & List B (2005) Asymmetric organocatalysis. *Org Biomol Chem* 3(5):719-724.

63. Shih HW, Vander Wal MN, Grange RL, & MacMillan DW (2010) Enantioselective alpha-benzylation of aldehydes via photoredox organocatalysis. *J Am Chem Soc* 132(39):13600-13603.

64. Terrett JA, Clift MD, & MacMillan DW (2014) Direct beta-alkylation of aldehydes via photoredox organocatalysis. *J Am Chem Soc* 136(19):6858-6861.

65. Barbas CF, 3rd (2008) Organocatalysis lost: Modern chemistry, ancient chemistry, and an unseen biosynthetic apparatus. *Angew Chem Int Ed Engl* 47(1):42-47.

66. Mase N & Barbas CF, 3rd (2010) In water, on water, and by water: Mimicking nature's aldolases with organocatalysis and water. *Org Biomol Chem* 8(18):4043-4050.

67. Notz W, Tanaka F, & Barbas CF, 3rd (2004) Enamine-based organocatalysis with proline and diamines: The development of direct catalytic asymmetric aldol, mannich, michael, and diels-alder reactions. *Acc Chem Res* 37(8):580-591.

68. Mukherjee S, Yang JW, Hoffmann S, & List B (2007) Asymmetric enamine catalysis. *Chem Rev* 107(12):5471-5569.

69. Erkkila A, Majander I, & Pihko PM (2007) Iminium catalysis. *Chem Rev* 107(12):5416-5470.

70. Santoro SW, Wang L, Herberich B, King DS, & Schultz PG (2002) An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat Biotechnol* 20(10):1044-1048.

71. Deiters A & Schultz PG (2005) *In vivo* incorporation of an alkyne into proteins in *Escherichia coli*. *Bioorg Med Chem Lett* 15(5):1521-1524.

72. Miyake-Stoner SJ*, et al.* (2009) Probing protein folding using site-specifically encoded unnatural amino acids as FRET donors with tryptophan. *Biochemistry* 48(25):5953-5962.

73. Wang L, Zhang Z, Brock A, & Schultz PG (2003) Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proc Natl Acad Sci U S A* 100(1):56-61.

74. Seitchik JL*, et al.* (2012) Genetically encoded tetrazine amino acid directs rapid site-specific *in vivo* bioorthogonal ligation with trans-cyclooctenes. *J Am Chem Soc* 134(6):2898-2901.

75. Wang J*, et al.* (2010) A biosynthetic route to photoclick chemistry on proteins. *J Am Chem Soc* 132(42):14812-14818.

76. Chen S & Tsao ML (2013) Genetic incorporation of a 2-naphthol group into proteins for site-specific azo coupling. *Bioconjug Chem* 24(10):1645-1649.

77. Polycarpo CR*, et al.* (2006) Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett* 580(28-29):6695-6700.

78. Wan W, Tharp JM, & Liu WR (2014) Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim Biophys Acta* 1844(6):1059-1070.

79. Guo HM & Tanaka F (2009) A fluorogenic aldehyde bearing a 1,2,3-triazole moiety for monitoring the progress of aldol reactions. *J Org Chem* 74(6):2417-2424.

80. List B, Barbas CF, 3rd, & Lerner RA (1998) Aldol sensors for the rapid generation of tunable fluorescence by antibody catalysis. *Proc Natl Acad Sci U S A* 95(26):15351-15355.

81. Cordova A, Notz W, & Barbas CF, 3rd (2002) Direct organocatalytic aldol reactions in buffered aqueous media. *Chem Commun (Camb)* (24):3024-3025.

82. Mase N*, et al.* (2006) Organocatalytic direct asymmetric aldol reactions in water. *J Am Chem Soc* 128(3):734-735.

83. Tang Z*, et al.* (2005) A highly efficient organocatalyst for direct aldol reactions of ketones with aldehydes [corrected]. *J Am Chem Soc* 127(25):9285-9289.

84. Chatterjee A, Sun SB, Furman JL, Xiao H, & Schultz PG (2013) A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli*. *Biochemistry* 52(10):1828-1837.

85. Hammill JT, Miyake-Stoner S, Hazen JL, Jackson JC, & Mehl RA (2007) Preparation of site-specifically labeled fluorinated proteins for 19F-NMR structural characterization. *Nat Protoc* 2(10):2601-2607.

86. Yanagisawa T*, et al.* (2008) Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode n(epsilon)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem Biol* 15(11):1187-1197.

87. Rackham O & Chin JW (2005) A network of orthogonal ribosome x mRNA pairs. *Nat Chem Biol* 1(3):159-166.

88. Wang Q, Parrish AR, & Wang L (2009) Expanding the genetic code for biological studies. *Chem Biol* 16(3):323-336.

89. Podtetenieff J, Taglieber A, Bill E, Reijerse EJ, & Reetz MT (2010) An artificial metalloenzyme: Creation of a designed copper binding site in a thermostable protein. *Angew Chem Int Ed Engl* 49(30):5151-5155.

90. Rothlisberger D*, et al.* (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453(7192):190-195.

91. Khersonsky O, Roodveldt C, & Tawfik DS (2006) Enzyme promiscuity: Evolutionary and mechanistic aspects. *Curr Opin Chem Biol* 10(5):498-508.

92. Yang H, Srivastava P, Zhang C, & Lewis JC (2014) A general method for artificial metalloenzyme formation through strain-promoted azide-alkyne cycloaddition. *Chembiochem* 15(2):223-227.

93. Carlos F. Barbas Y-FW, Chi-Huey Wong (1990) Deoxyribose-5-phosphate aldolase as a synthetic catalyst. *J Am Chem Soc* 112:2013-2014.

94. Schlee S*, et al.* (2013) Kinetic mechanism of indole-3-glycerol phosphate synthase. *Biochemistry* 52(1):132-142.

95. Baril SA*, et al.* (2017) Investigation of trimethyllysine binding by the HP1 chromodomain via unnatural amino acid mutagenesis. *J Am Chem Soc* 139(48):17253-17256.

96. Zandvoort E, Baas BJ, Quax WJ, & Poelarends GJ (2011) Systematic screening for catalytic promiscuity in 4-oxalocrotonate tautomerase: Enamine formation and aldolase activity. *Chembiochem* 12(4):602-609.

97. Nguyen DP*, et al.* (2009) Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins *via* a pyrrolysyl-tRNA synthetase/trna(cua) pair and click chemistry. *J Am Chem Soc* 131(25):8720-8721.

98. Torres-Kolbus J, Chou C, Liu J, & Deiters A (2014) Synthesis of non-linear protein dimers through a genetically encoded thiol-ene reaction. *PLoS One* 9(9):e105467.

99. Raj M & Singh VK (2009) Organocatalytic reactions in water. *Chem Commun (Camb)* (44):6687-6703.

100. Palomo C*, et al.* (2007) Water-compatible iminium activation: Organocatalytic michael reactions of carbon-centered nucleophiles with enals. *Angew Chem Int Ed Engl* 46(44):8431-8435.

101. Hayashi Y*, et al.* (2006) Combined proline-surfactant organocatalyst for the highly diastereo- and enantioselective aqueous direct cross-aldol reaction of aldehydes. *Angew Chem Int Ed Engl* 45(33):5527-5529.

102. Raj M, Vishnumaya V, Ginotra SK, & Singh VK (2006) Highly enantioselective direct aldol reaction catalyzed by organic molecules. *Org Lett* 8(18):4097-4099.

103. Lam YH, Grayson MN, Holland MC, Simon A, & Houk KN (2016) Theory and modeling of asymmetric catalytic reactions. *Acc Chem Res* 49(4):750-762.

104. Brogan AP, Dickerson TJ, & Janda KD (2006) Enamine-based aldol organocatalysis in water: Are they really "all wet"? *Angew Chem Int Ed Engl* 45(48):8100-8102.

105. Reetz MT (2004) Controlling the enantioselectivity of enzymes by directed evolution: Practical and theoretical ramifications. *Proc Natl Acad Sci U S A* 101(16):5716-5722.

106. Windle CL, Muller M, Nelson A, & Berry A (2014) Engineering aldolases as biocatalysts. *Curr Opin Chem Biol* 19:25-33.

107. Lassila JK, Baker D, & Herschlag D (2010) Origins of catalysis by computationally designed retroaldolase enzymes. *Proc Natl Acad Sci U S A* 107(11):4937-4942.

108. Pollack SJ, Jacobs JW, & Schultz PG (1986) Selective chemical catalysis by an antibody. *Science* 234(4783):1570-1573.

109. Tramontano A, Janda KD, & Lerner RA (1986) Chemical reactivity at an antibody binding site elicited by mechanistic design of a synthetic antigen. *Proc Natl Acad Sci U S A* 83(18):6736-6740.

110. Tramontano A, Janda KD, & Lerner RA (1986) Catalytic antibodies. *Science* 234(4783):1566-1570.

111. Zandvoort E, Geertsema EM, Baas BJ, Quax WJ, & Poelarends GJ (2012) Bridging between organocatalysis and biocatalysis: Asymmetric addition of acetaldehyde to β-nitrostyrenes catalyzed by a promiscuous proline-based tautomerase. *Angew Chem Int Ed Engl* 51(5):1240-1243.

112. Baas BJ, Zandvoort E, Geertsema EM, & Poelarends GJ (2013) Recent advances in the study of enzyme promiscuity in the tautomerase superfamily. *Chembiochem* 14(8):917-926.

113. Yancey PH, Clark ME, Hand SC, Bowlus RD, & Somero GN (1982) Living with water stress: Evolution of osmolyte systems. *Science* 217(4566):1214-1222.

114. Tanford C (1970) Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* 24:1-95.

115. Auton M & Bolen DW (2004) Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale. *Biochemistry* 43(5):1329-1342.

116. Auton M & Bolen DW (2005) Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc Natl Acad Sci U S A* 102(42):15065-15068.

117. Bennion BJ & Daggett V (2004) Counteraction of urea-induced protein denaturation by trimethylamine N-oxide: A chemical chaperone at atomic resolution. *Proc Natl Acad Sci U S A* 101(17):6433-6438.

118. Liao YT, Manson AC, DeLyser MR, Noid WG, & Cremer PS (2017) Trimethylamine N-oxide stabilizes proteins via a distinct mechanism compared with betaine and glycine. *Proc Natl Acad Sci U S A* 114(10):2479-2484.

119. Stadmiller SS, Gorensek-Benitez AH, Guseman AJ, & Pielak GJ (2017) Osmotic shock induced protein destabilization in living cells and its reversal by glycine betaine. *J Mol Biol* 429(8):1155-1161.

120. Yancey PH & Somero GN (1979) Counteraction of urea destabilization of protein structure by methylamine osmoregulatory compounds of elasmobranch fishes. *Biochem J* 183(2):317-323.

121. Hein MY*, et al.* (2015) A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 163(3):712-723.

122. Hand SC & Somero GN (1982) Urea and methylamine effects on rabbit muscle phosphofructokinase. Catalytic stability and aggregation state as a function of pH and temperature. *J Biol Chem* 257(2):734-741.

123. Kornblatt JA, Kornblatt MJ, Hoa GH, & Mauk AG (1993) Responses of two protein-protein complexes to solvent stress: Does water play a role at the interface? *Biophys J* 65(3):1059-1065.

124. Xavier KA, Shick KA, Smith-Gil SJ, & Willson RC (1997) Involvement of water molecules in the association of monoclonal antibody HyHEL-5 with bobwhite quail lysozyme. *Biophys J* 73(4):2116-2125.

125. Morar AS, Wang X, & Pielak GJ (2001) Effects of crowding by mono-, di-, and tetrasaccharides on cytochrome *c*-cytochrome *c* peroxidase binding: Comparing experiment to theory. *Biochemistry* 40(1):281-285.

126. Patel CN*, et al.* (2002) Effects of molecular crowding by saccharides on α-chymotrypsin dimerization. *Protein Science* 11(5):997-1003.

127. Chebotareva NA, Andreeva IE, Makeeva VF, Livanova NB, & Kurganov BI (2004) Effect of molecular crowding on self-association of phosphorylase kinase and its interaction with phosphorylase *b* and glycogen. *J Mol Recognit* 17(5):426-432.

128. Ignatova Z & Gierasch LM (2007) Effects of osmolytes on protein folding and aggregation in cells. *Methods Enzymol* 428:355-372.

129. Silvers TR & Myers JK (2013) Osmolyte effects on the self-association of concanavalin A: Testing theoretical models. *Biochemistry* 52(51):9367-9374.

130. Rose GD, Fleming PJ, Banavar JR, & Maritan A (2006) A backbone-based theory of protein folding. *Proc Natl Acad Sci U S A* 103(45):16623-16633.

131. Xu D, Tsai CJ, & Nussinov R (1997) Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng* 10(9):999-1012.

132. Janin J, Miller S, & Chothia C (1988) Surface, subunit interfaces and interior of oligomeric proteins. *J Mol Biol* 204(1):155-164.

133. Strickler SS*, et al.* (2006) Protein stability and surface electrostatics: A charged relationship. *Biochemistry* 45(9):2761-2766.

134. Wang Q, Zhuravleva A, & Gierasch LM (2011) Exploring weak, transient protein--protein interactions in crowded *in vivo* environments by in-cell nuclear magnetic resonance spectroscopy. *Biochemistry* 50(43):9225-9236.

135. Monteith WB, Cohen RD, Smith AE, Guzman-Cisneros E, & Pielak GJ (2015) Quinary structure modulates protein stability in cells. *Proc Natl Acad Sci U S A* 112(6):1739-1742.

136. Cohen RD, Guseman AJ, & Pielak GJ (2015) Intracellular pH modulates quinary structure. *Protein Sci* 24(11):1748-1755.

137. Smith AE, Zhou LZ, Gorensek AH, Senske M, & Pielak GJ (2016) In-cell thermodynamics and a new role for protein surfaces. *Proc Natl Acad Sci U S A* 113(7):1725-1730.

138. Miklos AC, Sarkar M, Wang Y, & Pielak GJ (2011) Protein crowding tunes protein stability. *J Am Chem Soc* 133(18):7116-7120.

139. Crowley PB, Chow E, & Papkovskaia T (2011) Protein interactions in the *Escherichia coli* cytosol: An impediment to in-cell NMR spectroscopy. *Chembiochem* 12(7):1043-1048.

140. Cayley DS, Guttman HJ, & Record MT, Jr. (2000) Biophysical characterization of changes in amounts and activity of Escherichia coli cell and compartment water and turgor pressure in response to osmotic stress. *Biophys J* 78(4):1748-1764.

141. Sukenik S, Ren P, & Gruebele M (2017) Weak protein-protein interactions in live cells are quantified by cell-volume modulation. *Proc Natl Acad Sci U S A* 114(26):6776-6781.

142. Ohtake S, Kita Y, & Arakawa T (2011) Interactions of formulation excipients with proteins in solution and in the dried state. *Adv Drug Deliv Rev* 63(13):1053-1073.

143. Jee J, Byeon IJ, Louis JM, & Gronenborn AM (2008) The point mutation A34F causes dimerization of GB1. *Proteins* 71(3):1420-1431.

144. Guseman AJ, Speer SL, Perez Goncalves GM, & Pielak GJ (2018) Surface-charge modulates protein-protein interactions in physiologically-relevant environments. *Biochemistry*, in press.

145. Acuner Ozbabacan SE, Engin HB, Gursoy A, & Keskin O (2011) Transient protein-protein interactions. *Protein Eng Des Sel* 24(9):635-648.

146. Weiller GF, Caraux G, & Sylvester N (2004) The modal distribution of protein isoelectric points reflects amino acid properties rather than sequence evolution. *Proteomics* 4(4):943-949.

147. Eymann C*, et al.* (2004) A comprehensive proteome map of growing *Bacillus subtilis* cells. *Proteomics* 4(10):2849-2876.

148. Li C*, et al.* (2010) Protein (19)F NMR in *Escherichia coli*. *J Am Chem Soc* 132(1):321-327.

149. Spitzer JJ & Poolman B (2005) Electrochemical structure of the crowded cytoplasm. *Trends Biochem Sci* 30(10):536-541.

150. Arakawa T & Timasheff SN (1985) The stabilization of proteins by osmolytes. *Biophys J* 47(3):411-414.

151. Abriata LA, Spiga E, & Peraro MD (2016) Molecular effects of concentrated solutes on protein hydration, dynamics, and electrostatics. *Biophys J* 111(4):743-755.

152. Schneck E, Horinek D, & Netz RR (2013) Insight into the molecular mechanisms of protein stabilizing osmolytes from global force-field variations. *J Phys Chem B* 117(28):8310-8321.

153. Burg MB & Ferraris JD (2008) Intracellular organic osmolytes: Function and regulation. *J Biol Chem* 283(12):7309-7313.

154. Robertson JD (1965) Studies on the chemical composition of muscle tissue. 3. The mantle muscle of cephalopod molluscs. *J Exp Biol* 42:153-175.

155. Carr W, JC IN, Gleeson RA, & Derby CD (1996) Stimulants of feeding behavior in fish: Analyses of tissues of diverse marine organisms. *The Biological Bulletin* 190(2):149-160.

156. Steeves CL & Baltz JM (2005) Regulation of intracellular glycine as an organic osmolyte in early preimplantation mouse embryos. *J Cell Physiol* 204(1):273-279.

157. Bowlus RD & Somero GN (1979) Solute compatibility with enzyme function and structure: Rationales for the selection of osmotic agents and end-products of anaerobic metabolism in marine invertebrates. *J Exp Zool* 208(2):137-151.

158. Street TO, Bolen DW, & Rose GD (2006) A molecular mechanism for osmolyte-induced protein stability. *Proc Natl Acad Sci U S A* 103(38):13997-14002.

159. Auton M, Rosgen J, Sinev M, Holthauzen LM, & Bolen DW (2011) Osmolyte effects on protein stability and solubility: A balancing act between backbone and side-chains. *Biophys Chem* 159(1):90-99.

160. Schreiber G & Fersht AR (1996) Rapid, electrostatically assisted association of proteins. *Nat Struct Biol* 3(5):427-431.

161. Sheinerman FB, Norel R, & Honig B (2000) Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol* 10(2):153-159.

162. Selzer T, Albeck S, & Schreiber G (2000) Rational design of faster associating and tighter binding protein complexes. *Nat Struct Biol* 7(7):537-541.

163. Selzer T & Schreiber G (2001) New insights into the mechanism of protein-protein association. *Proteins* 45(3):190-198.

164. Zhang Z, Witham S, & Alexov E (2011) On the role of electrostatics in protein-protein interactions. *Phys Biol* 8(3):035001.

165. Lindman S*, et al.* (2006) Salting the charged surface: pH and salt dependence of protein G B1 stability. *Biophys J* 90(8):2911-2921.

166. McConkey EH (1982) Molecular evolution, intracellular organization, and the quinary structure of proteins. *Proc Natl Acad Sci U S A* 79(10):3236-3240.

167. Srere PA (1985) The metabolon. *Trends Biochem Sci* 10:109-110.

168. Cohen RD & Pielak GJ (2016) Electrostatic contributions to protein quinary structure. *Journal of the American Chemical Society* 138(40):13139-13142.