

Genetic Code Expansion for Protein Activity Control and Protein Conjugation

by

Jihe Liu

B.S. Bioinformatics, Huazhong University of Science and Technology, 2011

Submitted to the Graduate Faculty of the
Dietrich School of Arts and Sciences in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH
Dietrich School of Arts and Sciences

This thesis was presented

by

Jihe Liu

It was defended on

July 17th, 2017

and approved by

Seth Horne, PhD, Department of Chemistry

Seth Childers, PhD, Department of Chemistry

Michael Tsang, PhD, Department of Developmental Biology

Committee Chair: Alexander Deiters, PhD, Department of Chemistry

Copyright © by Jihe Liu
2017

GENETIC CODE EXPANSION FOR PROTEIN ACTIVITY CONTROL AND PROTEIN CONJUGATION

Jihe Liu, PhD

University of Pittsburgh, 2017

Expanding the genetic code by unnatural amino acid (UAA) incorporation allows for control of protein structure and function in cells and animals. Genetic incorporations of several UAAs, including dithiolane lysine, azidoethyl and azidomethyl lysine, photocaged tyrosine/thiotyrosine/azatorysine, alkene lysine, and bipyridine lysine, were achieved through the pyrrolysyl-tRNA synthetase (PylRS). The UAA mutagenesis system was further studied, including optimization of the incorporation efficiency in mammalian cells, library selection of UAA in *E. coli*, and genetic incorporation of photocaged cysteine through *E. coli* Leu synthetase. UAA mutagenesis provides a versatile method for protein activity control and protein conjugation. As one application, optical control of protein activities (intein, TEV protease, Src kinase, HRP) was studied, with the genetic incorporation of photocaged or photoisomerizable UAAs. Optical deactivation of protein function was studied with degron or genetic incorporation of nitrophenylalanine. Besides, small molecule control of protein activity was studied with the genetic incorporation of a propargyl carbamate or an allyloxy carbonyl functional group. As another application, an aminooxy functional group was genetically incorporated for protein labeling. Next, the

zebrafish with the expanded genetic code was developed, and was applied to the optical control of cell signaling (MEK/ERK pathway). Optical control of other proteins (Alk5 kinase, Cas9, Cre recombination) and small molecule control of luciferase function in zebrafish were studied. Efforts were also made to generate stable fishlines for genetic code expansion. Lastly, optical control of other biomolecules (morpholino, peptide nucleic acid) was pursued in zebrafish.

Table of Contents

List of Tables	ix
List of Figures	x
Preface	xvii
1.0 Introduction	1
1.1 Expanding the Genetic Code through Unnatural Amino Acid Incorporation.....	1
1.2 Pyrrolysyl-tRNA synthetase as a tool to expand the genetic code	6
2.0 Incorporation of New Unnatural Amino Acids	10
2.1 Pyrrolysyl-tRNA Synthetase Screening Using an sfGFP Reporter	10
2.1.1 Incorporation of dithiolane lysine for gold nanoparticle conjugation	16
2.1.2 Incorporation of azidoethyl lysine	25
2.1.3 Incorporation of 2 nd generation caged tyrosines	27
2.1.4 Incorporation of photocaged thiotyrosine and photocaged azatyrosine	31
2.1.5 Incorporation of alkene lysine	39
2.1.6 Incorporation of bipyridine lysine	42
2.2 Optimizing Incorporation Efficiency in Mammalian Cells	52
2.3 Library Selection for Unnatural Amino Acid.....	65
2.3.1 Construction of a Pyrrolysyl-tRNA Synthetase Library and a Selection Marker	65

2.3.2 Selection of o-nitrobenzyl-O-tyrosine (ONBY) incorporation in E. coli..	72
2.4 Incorporation of Caged Cysteine Using an E. coli Leu Synthetase	86
3.0 Conditional Control of Protein Activity	96
3.1 Incorporation of Caged Cysteine into Intein, TEV Protease and Src Kinase.....	96
3.2 Incorporation of a Photoisomerizable Amino Acid into Horseradish Peroxidase	120
3.3 Photo-deactivation of Protein Function through Degron or Incorporation of Nitrophenylalanine	135
3.4 Small molecule Activation of Protein Function using Proc and Alloc Groups.....	149
4.0 Installation of an Aminooxy Functional Group for Protein Labeling	157
5.0 Expanding the Genetic Code in Zebrafish Embryos	175
5.1 Incorporation of Unnatural Amino Acids into Proteins in Zebrafish	175
5.2 Optical Control of Alk5 Function in Mammalian Cells and Zebrafish.....	200
5.3 Optical Control of Cre Recombinase Function in Zebrafish	216
5.4 Small Molecule Control of Protein Function in Zebrafish.....	220
5.5 Optical Control of CRISPR/Cas9 Gene Editing in Zebrafish.....	223
5.6 Generation of Stable Fishline for Genetic Code Expansion.....	231
Appendix A. Optical control of other biomolecules in zebrafish	236
A1. Optical control of circular morpholino function in zebrafish	236
A2. Optical control of peptide nucleic acid (PNA) in zebrafish.....	239

Appendix B. ESI-MS spectrum, plasmid map, general protocol, buffer composition	
.....	246
Bibliography	268

List of Tables

Table 2.1.....	13
Table 2.2.....	15
Table 2.3.....	46
Table 2.4.....	63
Table 2.5.....	73
Table 2.6.....	76
Table 2.7.....	77
Table 2.8.....	78
Table 5.1.....	193

List of Figures

Figure 1.1	3
Figure 1.2	5
Figure 1.3	8
Figure 2.1	11
Figure 2.2	18
Figure 2.3	19
Figure 2.4	20
Figure 2.5	21
Figure 2.6	22
Figure 2.7	23
Figure 2.8	23
Figure 2.9	24
Figure 2.10	27
Figure 2.11	29
Figure 2.12	31
Figure 2.13	33
Figure 2.14	34
Figure 2.15	35
Figure 2.16	37
Figure 2.17	38
Figure 2.18	40

Figure 2.19	41
Figure 2.20	42
Figure 2.21	44
Figure 2.22	54
Figure 2.23	55
Figure 2.24	56
Figure 2.25	57
Figure 2.26	59
Figure 2.27	67
Figure 2.28	69
Figure 2.29	70
Figure 2.30	71
Figure 2.31	74
Figure 2.32	75
Figure 2.33	88
Figure 2.34	89
Figure 2.35	90
Figure 2.36	91
Figure 2.37	92
Figure 2.38	93
Figure 3.1	97
Figure 3.2	98
Figure 3.3	100

Figure 3.4	101
Figure 3.5	101
Figure 3.6	102
Figure 3.7	103
Figure 3.8	104
Figure 3.9	105
Figure 3.10	106
Figure 3.11	107
Figure 3.12	108
Figure 3.13	109
Figure 3.14	110
Figure 3.15	111
Figure 3.16	112
Figure 3.17	120
Figure 3.18	121
Figure 3.19	122
Figure 3.20	123
Figure 3.21	123
Figure 3.22	124
Figure 3.23	125
Figure 3.24	126
Figure 3.25	127
Figure 3.26	128

Figure 3.27	129
Figure 3.28	130
Figure 3.29	131
Figure 3.30	136
Figure 3.31	137
Figure 3.32	138
Figure 3.33	139
Figure 3.34	141
Figure 3.35	142
Figure 3.36	143
Figure 3.37	144
Figure 3.38	150
Figure 3.39	151
Figure 3.40	152
Figure 3.41	153
Figure 3.42	154
Figure 4.1	158
Figure 4.2	161
Figure 4.3	162
Figure 4.4	163
Figure 4.5	164
Figure 4.6	165
Figure 4.7	166

Figure 4.8	167
Figure 4.9	168
Figure 5.1	176
Figure 5.2	177
Figure 5.3	178
Figure 5.4	179
Figure 5.5	180
Figure 5.6	181
Figure 5.7	182
Figure 5.8	183
Figure 5.9	184
Figure 5.10	185
Figure 5.11	187
Figure 5.12	189
Figure 5.13	190
Figure 5.14	201
Figure 5.15	202
Figure 5.16	203
Figure 5.17	204
Figure 5.18	205
Figure 5.19	206
Figure 5.20	206
Figure 5.21	207

Figure 5.22	208
Figure 5.23	210
Figure 5.24	212
Figure 5.25	217
Figure 5.26	218
Figure 5.27	221
Figure 5.28	225
Figure 5.29	226
Figure 5.30	227
Figure 5.31	228
Figure 5.32	232
Figure A1	237
Figure A2.....	238
Figure A3.....	240
Figure A4.....	241
Figure A5.....	242
Figure A6.....	246
Figure A7.....	247
Figure A8.....	248
Figure A9.....	249
Figure A10.....	250
Figure A11.....	251
Figure A12.....	251

Figure A13.....	252
Figure A14.....	252
Figure A15.....	253
Figure A16.....	254
Figure A17.....	255
Figure A18.....	255
Figure A19.....	256
Figure A20.....	257
Figure A21.....	257
Figure A22.....	258

Preface

I would like to thank my advisor, Dr. Alexander Deiters, for his tremendous support and guidance during my study in graduate school. I am lucky to obtain scientific training in his lab. I would also like to thank Dr. Seth Horne, Dr. Seth Childers, and Dr. Michael Tsang, for serving as my committee members and providing suggestions on my thesis work.

I want to thank current and past members of Deiters lab, for their support and help on my projects. I also want to acknowledge all my friends in Pittsburgh and Raleigh, for bringing in fun and memorable moments during my study in the United States. Finally, I would like to thank my parents, Jun Liu and Xiaoyun He. The love and courage that you give me keep me moving on this journey and beyond.

1.0 Introduction

1.1 Expanding the Genetic Code through Unnatural Amino Acid Incorporation

Proteins, consisting of 20 canonical amino acids, are fundamental biomolecules for cellular activity and physiology. While the number of natural amino acids is small, various post-translational modifications, like phosphorylation, methylation, and acetylation, clearly suggest that the biological activities of proteins exceed the limit of canonical amino acids.¹ In fact, many proteins are functional only when post-translationally modified with other functional groups, and this lays the basis for a dynamic and robust biological system. Therefore, the development of a methodology to expand the genetic code to other, so called unnatural amino acids (UAAs) is very appealing.¹⁻² Such a methodology will allow us to directly generate proteins with desired UAAs, which will greatly enhance our ability to control protein structure and function. Additionally, it will be a powerful tool to study protein function in cells, thus facilitating our understanding of life.

Several approaches have been developed to incorporate UAAs into proteins, including chemical modifications, chemical synthesis, and *in vitro* and *in vivo* biosynthetic approaches.¹ Chemical modifications utilize a chemical reaction between an amino acid and a specific reagent. Commonly used amino acids include lysine and cysteine, due to their nucleophilic side chains. However, there are two major drawbacks of this method. First, it allows little selectivity over the modification sites if more than one reactive amino acids are present in a protein. Second, the general efficiency of the reaction is limited.

Chemical synthesis, on the other hand, seeks to generate a protein containing an UAA *de novo*. Stepwise solid-phase peptide synthesis (SPPS) is the most frequently used method to achieve this goal.³ However, the synthesis is generally limited to proteins with the size less than

100 amino acids. Two methods can be used to remedy this disadvantage. The native chemical ligation (NCL) ligates two peptide fragments into an intact protein.⁴ Yet, this method is expensive and time-consuming when a large protein is assembled. Another evolutionary method is the expressed protein ligation (EPL).⁵ In this method, the major, unmodified part of a protein, which contains a C-terminal α -thioester through intein-based cleavage, is expressed within living cells. It is then ligated to a second synthetic peptide with the UAA. Although this method greatly facilitates the generation of a large protein containing an UAA, it is only suitable when the UAA is incorporated near the N- or C-terminus.

An *in vitro* method for site-specific incorporation of an UAA into proteins was first reported in 1989.⁶ A chemically acylated suppressor tRNA was used in an *in vitro* transcription-translation system to incorporate an unnatural amino acid in place of the UAG amber stop codon in a protein of interest. Although this work initiated the concept to completely express unnaturally mutagenized proteins at the translational level, the low efficiency of *in vitro* systems, the difficulty in applying it to living cells, and the demanding chemical synthesis of the aminoacylated tRNA impeded its further application.

An *in vivo* method for UAA incorporation was therefore highly desired, due to the expected high yield of expression and the applications in live cells. In initial experiments, an auxotrophic *E. coli* strain was used for the incorporation of an analogue amino acid in place of its corresponding canonical amino acid in the whole proteome.⁷ However, this method is limited to analogues that are structurally similar to the canonical amino acids, and the incorporation is not site-specific.

In 2001, an UAA was site-specifically incorporated into a protein in *E. coli* using the cellular translational machinery (Fig 1.1).⁸ To achieve this goal, a tRNA that uniquely recognizes the amber stop codon, UAG, needed to be generated. The amber stop codon was chosen to encode the UAA, because it is the least frequently used among the three stop codons (9% in *E. coli*, 20% in human).⁹ The incorporation of an UAA across the proteome is therefore expected to bring minimal effect on cellular physiology. This tRNA should be orthogonal to any other aminoacyl-

tRNA synthetases (aaRSs) in the host cell. To aminoacylate this tRNA with an UAA, a novel aaRS was introduced. This non-endogenous aaRS must also be orthogonal to any other endogenous tRNAs or canonical amino acids. Likewise, the UAA must be the substrate only for this aaRS, but not any other endogenous aaRSs. Besides, the unnatural amino acid should not be toxic to the cells, and needs to be transported to the cytoplasm when added to the growth medium at an appropriate concentration.

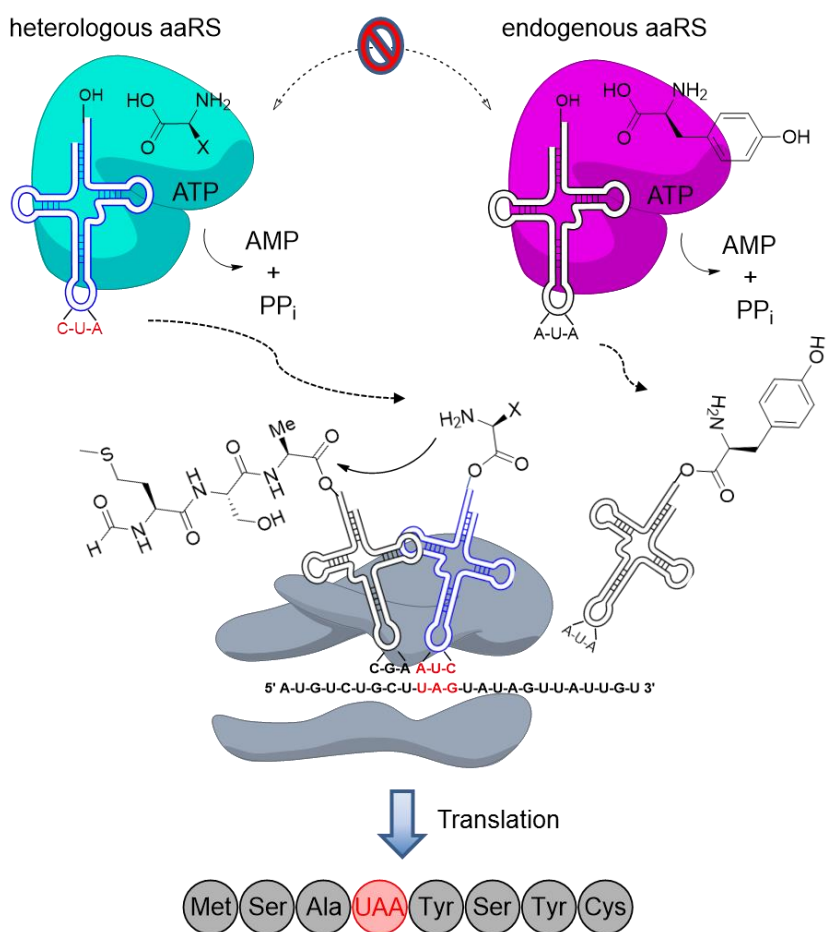


Figure 1.1. Orthogonal aaRS/tRNA pair for UAA mutagenesis. The orthogonal aaRS (cyan) aminoacylates the orthogonal tRNA (with CUA labeled in red) with an unnatural amino acid and does not cross-react with the endogenous aaRS/tRNA pairs (purple). Aminoacyl-tRNA is transported to the ribosome (brown), where it incorporates the unnatural amino acid in response to the UAG codon. Adapted from *J. Biol. Chem.* **2010**, 285, 11039-11044.

Because the initial attempt to directly evolve an orthogonal aminoacyl-tRNA synthetase/tRNA pair (aaRS/tRNA) from the existing *E. coli* aaRS/tRNA pairs was not successful, a heterologous tyrosyl-tRNA synthetase (TyrRS)/tRNA^{Tyr} pair from *M. jannaschii*, a thermophilic methanogenic archaean, was chosen.⁸ This pair perfectly meets the above requirements: both the *Mj*tRNA^{Tyr} (with several modifications)⁸ and *Mj*TyrRS are orthogonal to their *E. coli* counterparts, but are still functional in *E. coli*. To evolve an *Mj*TyrRS/tRNA^{Tyr} pair for a desired UAA, a two-step evolution process was developed.¹ First, an *Mj*tRNA^{Tyr} library was constructed, randomly mutating eleven nucleotides that do not interact with the *Mj*TyrRS. The library was subjected to directed evolution, for the selection of mutant tRNA^{Tyr} that will only be aminoacylated by the cognate *Mj*TyrRS, but not by any endogenous aaRSs. As a result, a mutant *Mj*tRNA^{Tyr} with five nucleotide substitutions was selected.⁸ Second, a library of *Mj*TyrRS mutants, with randomization of five residues (Y32, E107, D158, I159, L162) around the amino acid-binding site (within 6.5 Å of the para position of the aryl ring of bound tyrosine), were then constructed and subjected to directed evolution, for the selection of an aaRS that only recognizes an UAA but not any canonical amino acids. As a result, a mutant *Mj*TyrRS with four mutations was selected.⁸ The resulting *Mj*TyrRS/tRNA^{Tyr} pair thus allows site-specific incorporation of the desired UAA.

Using the heterologous *Mj*TyrRS/tRNA^{Tyr} pair described above, more than 70 UAAs have been genetically incorporated in *E. coli* (Fig 1.2). These UAAs represent distinct structural and functional properties not found in canonical amino acids, including chemically reactive groups for bioconjugation (**1**, **2**, **3**), biophysical probes (IR, NMR, fluorescent) for the study of protein structure and function (**4**, **5**), photocaged and photoreactive groups (**6**, **7**), and the product of post-translational modification (**8**). The site-specific incorporation of these UAAs in *E. coli* greatly expands our ability to investigate protein structure and function.

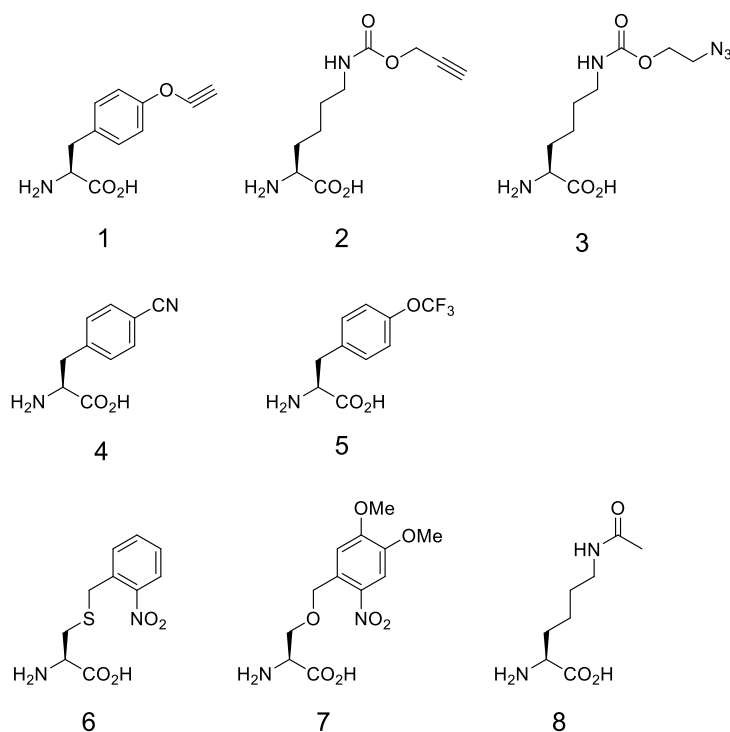


Figure 1.2. Chemical structures of genetically encoded unnatural amino acids. These UAAs represent distinct structural and functional properties, including chemically reactive groups for bioconjugation (**1**, **2**, **3**), biophysical probes (IR, NMR, fluorescent) for the study of protein structure and function (**4**, **5**), photocaged and photoreactive groups (**6**, **7**), and the product of post-translational modification (**8**). Adapted from *Annu. Rev. Biochem.* **2010**, 79, 413-444.

Although the *Mj*TyrRS/tRNA^{Tyr} pair is an effective platform for UAA incorporation in *E. coli*, it could not be readily applied to organisms like yeast and mammalian cells. This is because the *Mj*TyrRS/tRNA^{Tyr} pair is orthogonal in prokaryotes, but not in eukaryotes.¹⁰ In fact, the first incorporation of an UAA in eukaryotes (yeast) was achieved through the evolution of a prokaryotic *E. coli* aaRS/tRNA pair.¹¹ The evolved aaRS/tRNA pair from prokaryote could further be transferred to mammalian cells.¹² However, the requirement of separate selection procedures for incorporation of an UAA in prokaryotes and eukaryotes impedes its rapid application. Therefore, a time-efficient and facile approach is needed to facilitate the incorporation of UAAs in both kingdoms. Ideally, an aaRS/tRNA pair could be evolved for a desired UAA in the prokaryotes first,

and then be shuttled to the eukaryotes directly without the loss of orthogonality or functionality.

1.2 Pyrrolysyl-tRNA synthetase as a tool to expand the genetic code

In 2002, a 22nd amino acid, pyrrolysine, was found to be encoded by UAG in *Methanosarcina barkeri* (Fig 1.3A).¹³ The pyrrolysine is incorporated into the methylamine methyltransferase through an in-frame amber UAG codon, which was proposed to play key role in the activation of methylamine substrates.^{13b} The pyrrolysyl-tRNA synthetase (PylRS) and the pyrrolysyl-tRNA (PylT), which were proposed to be responsible for pyrrolysine incorporation, were identified in the *Methanosarcina barkeri* genome.^{13a} Although the complete biosynthetic pathway for pyrrolysine was identified a few years later,¹⁴ it was quickly realized that the components for the incorporation of pyrrolysine might be a powerful tool for genetic code expansion. Since the PylRS/PylT pair functions in a natural context where the canonical aaRS/tRNA pairs exist, it is speculated that this orthogonality will be retained when the PylRS/PylT pair is used in other organisms. Indeed, in *E. coli* or mammalian cells, the PylRS/PylT pair from either *M. barkeri* or *M. mazei* does not cross-react with endogenous aaRSs or tRNAs.¹⁵ This system is advantageous over the previously systems in several aspects. First, no directed evolution is needed for the tRNA, because the PylT is already a perfect substrate for its cognate PylRS, but not any endogenous aaRSs. Second, the anticodon region of the PylT recognizes the UAG codon, with no need for mutation. Third, the mutant PylRS is unlikely to incorporate any endogenous amino acids, which show distinct structural difference with the pyrrolysine. In contrast, for *Mj*TyrRS, a substantial evolutionary effort is needed to prevent its recognition of endogenous tyrosine. In summary, the PylRS/PylT pair has been established as a versatile and convenient tool for genetic code expansion in a wide range of organisms.¹⁶

The pyrrolysyl-tRNA synthetase (PylRS) belongs to the class II aaRS family. A core, seven-

stranded antiparallel β -sheet is surrounded by several long helices (Fig 1.3B).¹⁷ The PylRS differs from other members of its subfamily, with its deep hydrophobic pocket to harbor pyrrolysine, which is significantly larger than any 20 canonical amino acids. From the crystal structure of the *M. mazei* PylRS, the pyrrole ring is buried in this pocket, surrounded by the residues Tyr384, Trp417, Cys348 and Val401 (Fig 1.3B).¹⁷ The structural analysis reveals two important residues for pyrrolysine recognition: Asn346 and Tyr384. The Asn346 forms a hydrogen bond with the amide oxygen of pyrrolysine, while the Tyr384 hydrogen bonds with both the pyrrole nitrogen and the α -amino group through its hydroxyl group.¹⁷ Besides, Tyr384 also forms a π - π stacking with the pyrrole ring, further stabilizing the substrate. The sequence alignment of the PylRS from *M. barkeri* and *M. mazei* also reveals that these residues are highly conserved within the PylRS (Fig 1.3C).

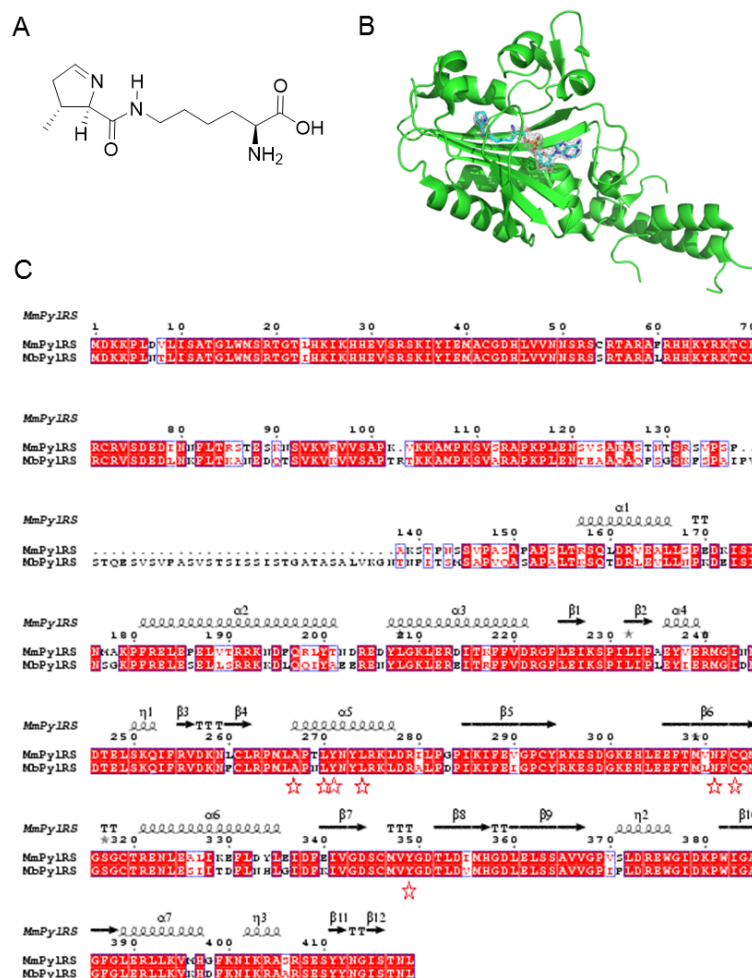


Figure 1.3. Structure of pyrrolysine and PylRS. (A) Structure of pyrrolysine. (B) Wild-type PylRS with pyrrolysyl-AMP, from *M. mazei* (PDB ID: 2Q7H). (C) Sequence alignment of the PylRS from *M. barkeri* (GenBank: AAL40867) and *M. mazei* (GenBank: NP_633469). Key residues for pyrrolysine recognition are marked with stars. The figure is generated by ClustalW and ESript.

It was speculated that UAA with similar structure as pyrrolysine could also be recognized by PylRS. Indeed, the wild-type PylRS exhibits extraordinarily broad substrate spectrum for many lysine analogues.¹⁸ These UAAs could be used for click chemistry,^{18b} photocrosslinking,^{18c} and fluorescence labeling.^{18d} For other UAAs that are not recognized by the wild-type PylRS, a library of PylRS mutants could be constructed and selected for the incorporation of a desired UAA

(discussed in Chapter 2.3). This strategy has allowed successful incorporation of many other lysine analogues,¹⁹ and further expands the application to the photocontrol of protein localization^{19d} and the site-specific post-translational modifications of proteins.^{19e} The substrate specificity could also be engineered for phenylalanine analogues,²⁰ histidine analogues,²¹ as well as photocaged tyrosine²² and photocaged cysteine.²³ For example, the N311A/C313A mutant allows incorporation of more than twenty phenylalanine analogues, with functional groups located at either *ortho*, *meta*, or *para* positions.²⁰ Altogether, these UAAs offer new probes for the study of protein function and structure, and expand the toolbox for the optogenetic control of protein function. Moreover, the PylRS/PyIT system has been successfully applied for the genetic code expansion of animals, including worms,²⁴ fruit flies,²⁵ mice,²⁶ and zebrafish (discussed in Chapter 5). In summary, the pyrrolysyl-tRNA synthetase is a great tool for genetic code expansion in *E. coli*, mammalian cells and animals.

2.0 Incorporation of New Unnatural Amino Acids

2.1 Pyrrolysyl-tRNA Synthetase Screening Using an sfGFP Reporter

To rapidly identify PylRS mutants that recognize newly synthesized UAAs, we applied a synthetase screening approach with an sfGFP reporter.²⁷ We undertook the screening based on the fact that some PylRS show broad substrate specificity for several UAAs, in particular if they are structurally similar.^{20a, 20c} The sfGFP shows enhanced folding capability over other GFP variants, and is suitable for *E. coli* expression.²⁸ A TAG amber stop codon was placed at the permissive Tyr151 position of the sfGFP. Tyr151 is located at the surface of the protein, and its side-chain is facing outward. Therefore, the incorporation of an UAA is not expected to disturb sfGFP maturation and fluorescence. Since sfGFP would only be fully translated and fluorescent when the amber stop codon is base-paired with PylT charged with an amino acid, the fluorescence of sfGFP is a robust readout for the UAA incorporation efficiency and thus PylRS function (Fig 2.1). To this end, a reporter plasmid, sfGFP-Y151TAG-pylT, was constructed. The expression of sfGFP-Y151TAG was under an arabinose promoter, and the expression of PylT was under a constitutive promoter.

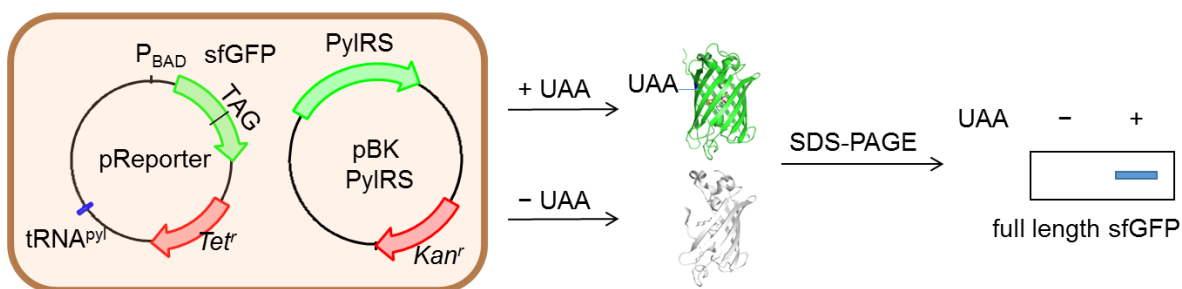


Figure 2.1. Pyrrolysyl-tRNA synthetase screening in *E. coli* using an sfGFP reporter. The newly constructed reporter plasmid (pReporter) contains an sfGFP gene sequence with TAG amber codon placed at Tyr 151 position, as well as PylT (from *M. Barkeri*) expression sequence. The synthetase plasmid (pBK-PylRS) contains a PylRS (from *M. Barkeri*) gene sequence. The sfGFP protein is expressed in *E. coli* in the presence of UAA. No protein is expressed in the absence of UAA. Protein expression is analyzed by SDS-PAGE.

Based on the mutants that we, and others, reported,¹⁶ we generated a panel of PylRS mutants for screening (Table 2.1). Structures of UAAs that are incorporated by these mutants are also presented (Table 2.2). Most of the synthetases were designed for either lysine derivatives (EV1 to EV13) or tyrosine derivatives (EV16 to EV20-1). In the following six sections, structurally different UAAs that were incorporated through synthetase screening will be discussed.

Code	Mutation sites											UAA
	M241	L266	A267	L270	Y271	L274	N311	C313	M315	Y349	W382	
EV1												AzMK, AzEK, NorK
EV1-2		M			L							
EV2	F		S		C	M						NPMK, PCC, PCHC
EV2-1	F		S		C	M				W		
EV2-2	F		S	G	C	M						
EV2-3	F		S		C	M		G				
EV2-4	F		S	G	C	M		G				
EV2-5	F		S		C	M				F		
EV3						V		V	Q			
EV8-1								V	Q			

EV9						V						
EV10					A							NPPK, NDBFK
EV12					A	M						HCK
EV13-1										F		DK
EV13					A					F		OABK, NorK, BPK
EV15							A	K				
EV16							A	A				
EV16-1					A		A	A				
EV16-2						V	A	A				
EV16-3							A	A		F		ONBAY
EV16-4					A		A	A		F		
EV16-5					M	A	A	A		F		ONBY, ONBSY, Azof
EV16-6					M	M	A	A		F		
EV16-7			T				A	A				
EV17					M	A				F		
EV18-1					M	A	G	G		F		
EV18-2				F	M	A	G	G		F		
EV19-1							G	G		F		
EV19-2				F			G	G		F		
EV20				F		M	G	G		F		ONBY, NPY, MNPY, NPEY
EV20-1				F		A	G	G		F		
EV21						A		A		F		
EV21-1						A				F		
EV22						A		V		F		
EV23						A		S		F		
EV24								V				
EV25					M	G		A		F		
EV26					A	M		A		F		
EV27					M	A		A		F		
EV28					M	A						
EV29										W		
EV30					A	M				F		
EV31			T				T	T				
EV33				I	F	G	G	G		F		
EV34				I	F	M	G	G		F		
EV35				F	F	G	G	G		F		
EV36								A		F		
EV37					M	V		A		F		
EV38			T			S	V	G		F		PSC
EV38-1							V	G		F		

EV39-1			T				A	A		F		
--------	--	--	---	--	--	--	---	---	--	---	--	--

Table 2.1. Panel of PylRS mutants (from M. Barkeri) used in screening. Wild-type amino acid positions and corresponding sequences are listed on the top. Mutations for individual PylRS mutant are listed. Blank indicates wild-type amino acid sequence at that position. EV1 represents wild-type PylRS. For PylRS from M. Mazei, the amino acid numbering is added by 35 (for example, Y349 in M. Barkeri corresponds to Y384 in M. Mazei).

Among this synthetase panel, EV1-2 was reported to increase the incorporation efficiency of propargyl lysine.²⁹ EV2 was identified through library selection to incorporate photocaged lysine,^{19d} and EV2-1 to EV2-5 were generated (mutations proposed by Yuta Naro from molecular docking studies) as an attempt to increase the incorporation efficiency of PCC; however, no increased incorporation efficiency was observed. The synthetases EV3 to EV13 were generated by Hank Chou (previous postdoc in the lab). EV15 was reported to incorporate phenylalanine.³⁰ EV16 was reported to incorporate several phenylalanine derivatives.²⁰ EV16-5 and EV17 were identified in the Cropp lab to incorporate ONBSY, and EV16-1 to EV16-4 were intermediate mutants toward the generation of EV16-5. EV20 was reported to incorporate ONBY,²² and EV18-1 to EV19-2 were the intermediate mutants during the generation of EV20. Both EV16-5 and EV20 could incorporate ONBY, however, EV16-5 showed some background incorporation of endogenous amino acids. We sought to explore if L274 played a role in background incorporation, as this position was mutated to structurally a distinct amino acid, such as A274. EV16-6 was therefore generated by mutating L274 to M (matching a mutation in EV20), and EV201 was generated by mutating L274 to A (matching a mutation in EV16-5). However, both synthetases could not incorporate ONBY. EV16-7 was generated, as A267T was proposed by Yuta Naro (from molecular docking studies; Deiters lab) to facilitate the incorporation of phenylalanine analogues. EV16, EV16-3, EV16-4, EV16-5 were the only four synthetases in the whole panel that showed background incorporation in the absence of unnatural amino acid.

More recently, the EV21 to EV39-1 synthetases were generated with assistance of Xinyu Chen (Deiters lab), to further expand the synthetase panel. Among them, the EV21 to EV31 mutants have been previously reported (for an overview see¹⁶). EV21 could incorporate Boc-lysine and an α -hydroxyacid analog of Boc-lysine.³¹ EV21-1 was an intermediate mutant toward the generation of EV21, EV22 and EV23. EV22 to EV25, EV27, and EV29 to EV31 were previously identified through genetic selections from a PylRS library. EV22 could facilitate the incorporation of N^ε-carbobenzyloxy-lysine.³² EV23 did incorporate an alkyl diazirine-containing lysine (DiZPK).³³ EV24 did incorporate thiapropyl-lysine.³⁴ EV25 could incorporate *p*-nitrocarbobenzyloxy-lysine.^{19c} EV26 could incorporate a *trans*-cyclooctene-containing lysine, identified through a focused synthetase panel screening.³⁵ EV27 could incorporate *o*-nitrobenzyloxycarbonyl-lysine.^{19b} EV28 was intermediate mutant toward the generation of EV27. EV29 (with single Y349W mutation) could incorporate propionyl-lysine.³⁶ EV30 could incorporate a spin-labeled lysine.³⁷ EV31 could incorporate *p*-benzoyl-L-phenylalanine.³⁸ EV33 to EV35 were generated based on the combinatorial mutations for ONBY incorporation (L270F, L274M, N311G, C313G, Y349F)²² and histidine analogue incorporation (L270I, Y271F, L274G, C313F, Y349F),²¹ as an attempt to incorporate photocaged histidine. Specifically, three mutations (N311G, C313G, Y349F) were retained in all three synthetases, while other mutations were shuffled and combined (L270I, Y271F, L274G for EV33; L270I, Y271F, L274M for EV34; L270F, Y271F, L274G for EV35). EV36 was an intermediate mutant toward the generation of EV25 and EV26. EV37 contains mutations for the incorporation of TEMPO-lysine (evolved by the Chin lab), with an extra Y349F mutation. EV38 contained mutations for the incorporation of a photoswitchable click amino acid (PSC),³⁹ with an extra Y349F mutation. EV38-1 was intermediate mutant toward the generation of EV38.

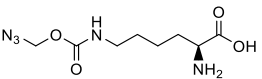
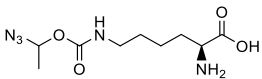
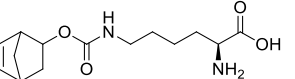
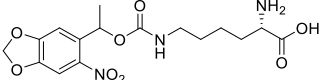
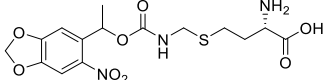
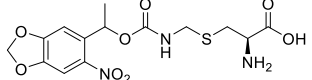
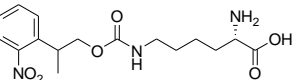
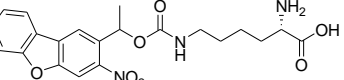
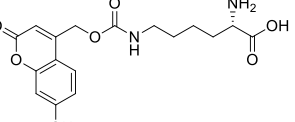
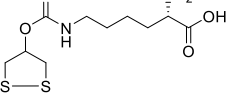
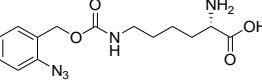
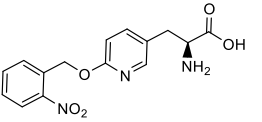
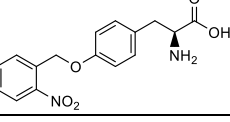
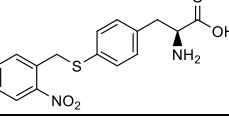
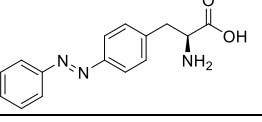
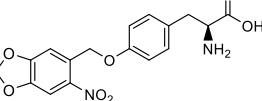
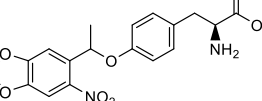
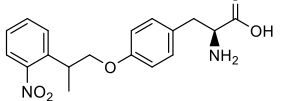
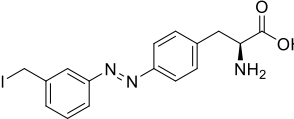
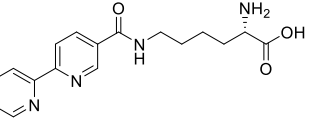
AzMK		AzEK		NorK	
NPMK		PCHC		PCC	
NPPK		NDBFK		HCK	
DK		OABK		ONBAY	
ONBY		ONBSY		AzoF	
NPY		MNPY		NPEY	
PSC		BPK			

Table 2.2. Structures of new UAAs that have been incorporated by the PylRS mutant shown above.

We applied a three phase partitioning (TPP) method for rapid extraction of sfGFP protein from cells.⁴⁰ Briefly, cells are first resuspended in solution with high ammonium sulfate concentration. t-Butanol is added to form a three-phase system. The ammonium sulfate concentration is well adjusted so that sfGFP stays at the bottom layer, while most cellular proteins, cell wall polysaccharides, and chromosomal DNA precipitate as a solid, middle phase. No sonication and column chromatography is needed in this process. Since the sfGFP expression and extraction could be done rapidly in a small scale (5 ml), medium-throughput screening of

multiple mutant PyIRS and UAAs in parallel becomes possible. We therefore used TPP method for the screening of the PyIRS panel.

2.1.1 Incorporation of dithiolane lysine for gold nanoparticle conjugation

Nanoparticles (NP) are microscopic particles with at least one dimension being less than 100 nm. The conjugation of NPs with a protein has been used in various applications, including biosensors, drug delivery, imaging and protein arrays.⁴¹ However, the inherent differences between inorganic materials and biological macromolecules, both in scale and character, make conjugation a challenging task. Importantly, a protein needs to maintain its structure and function after conjugation. This usually requires delicate conjugation methods, since the function of a macromolecule is easily impaired by surface effects at the nanoscale.⁴¹

Several approaches have been developed to conjugate proteins to NPs.⁴¹ The most widely used is the conjugation through direct electrostatic adsorption. This method is straightforward, with no chemical reactions required. However, there is little control over how many proteins are conjugated and how the proteins are orientated on the NP surface, since the conjugation is not residue-specific. It has been shown that the protein orientation on the NP is crucial for protein function.⁴¹ A second way is to covalently link a protein to a NP ligand. The conjugation reaction differs based on the choice of ligands and reactive residues on a protein. A clear advantage of this method is that the conjugate could be delicately controlled by adjusting the ratio of protein to NP. However, control of protein orientation on the NP surface is still not achieved with this method, since it relies on the position of reactive residues on a target protein. A third way is to directly conjugate a protein to the NP surface without a linker. This method usually utilizes cysteine residues on a target protein, which can covalently bond to the NP. Since this method results in no/short linker region between the protein and NP, it is favorable for the development of biosensors, where the output (fluorescence or luminescence) is sensitive to the distance between

biosensor and NP even on the Ångstrom scale. In addition, the conjugation orientation can be more easily controlled due to the rare frequency of cysteine in proteins. However, if the NP ligand is too densely packed on the surface, it will sterically block the protein from reaching the NP. Therefore, the conjugation efficiency would be a concern.⁴¹

Gold nanoparticles (Au NP) were chosen for conjugation due to their biocompatibility and outstanding biophysical properties.⁴² Currently, most proteins are conjugated to Au NPs in a non-specific way, usually through the electrostatic adsorption (first method described above). The conjugation through a defined chemical covalent bond was also reported (second method described above).⁴³ Both the Au NP and the protein target need to be modified with designated functional groups so that they can conjugate through chemical reactions. Usually the modification is time-consuming and results in low yield. The direct conjugation through a thiol group is thus an ideal route (third method described above). Previously, one or more cysteine residues on the surface of a target protein, either identified in a wild-type protein or genetically introduced, were used for conjugation.⁴⁴

Using UAA mutagenesis, we speculate that the conjugation of proteins to an Au NP could be achieved through the site-specific incorporation of an UAA that binds to the Au NP. This method displays several advantages over previous ones. First, no chemical modifications are required for either the Au NP or proteins. Second, the UAA can be designed to bind to the Au NP with higher affinity than the cysteine thiol group, for example a cyclic disulfide, thus generating a more stable conjugate compared to the cysteine-based strategy.⁴⁵

To this end, dithiolane lysine (DK), a lysine derivative with a dithiolane functional group, was synthesized by Rajendra Uprety in our lab (Table 2.2). The disulfide bond will break up to interact with Au, because the Au-S bond strength approximates the S-S bond strength; no independent, prior reduction of the disulfide to the dithiol is required.⁴⁶ Therefore, the site-specific incorporation of dithiolane lysine into a protein enables direct conjugation of that protein to the Au NP through an one-step incubation (Fig 2.2).

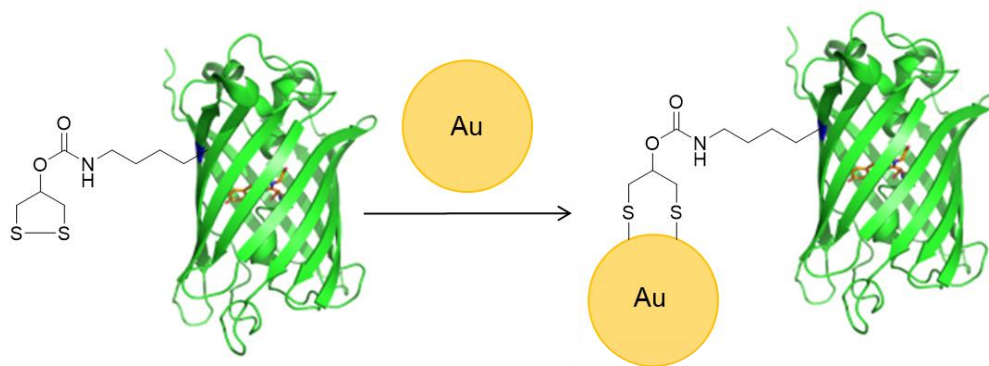


Figure 2.2. Conjugation of sfGFP-Y151TAG-DK to the Au NP. Dithiolane lysine is site-specifically incorporated into sfGFP at Tyr 151 position. The protein can be readily conjugated to the Au NP through the dithiolane group.

We first screened the small PyIRS panel for the incorporation of DK. Among eight synthetases that we tested, EV13-1 (with Y389F mutation) and EV13 (with Y271A and Y349F mutations) showed highest incorporation efficiency (Fig 2.3A). We then repeated the incorporation test in a large-scale (25 ml) expression using EV13-1 and EV13. EV13-1 showed a slightly higher incorporation efficiency (yield of 3 mg/L) based on the fluorescence of Ni-purified proteins (Fig 2.3B), which was further confirmed by SDS-PAGE analysis (Fig 2.3C). We therefore used EV13-1 for our future experiment.

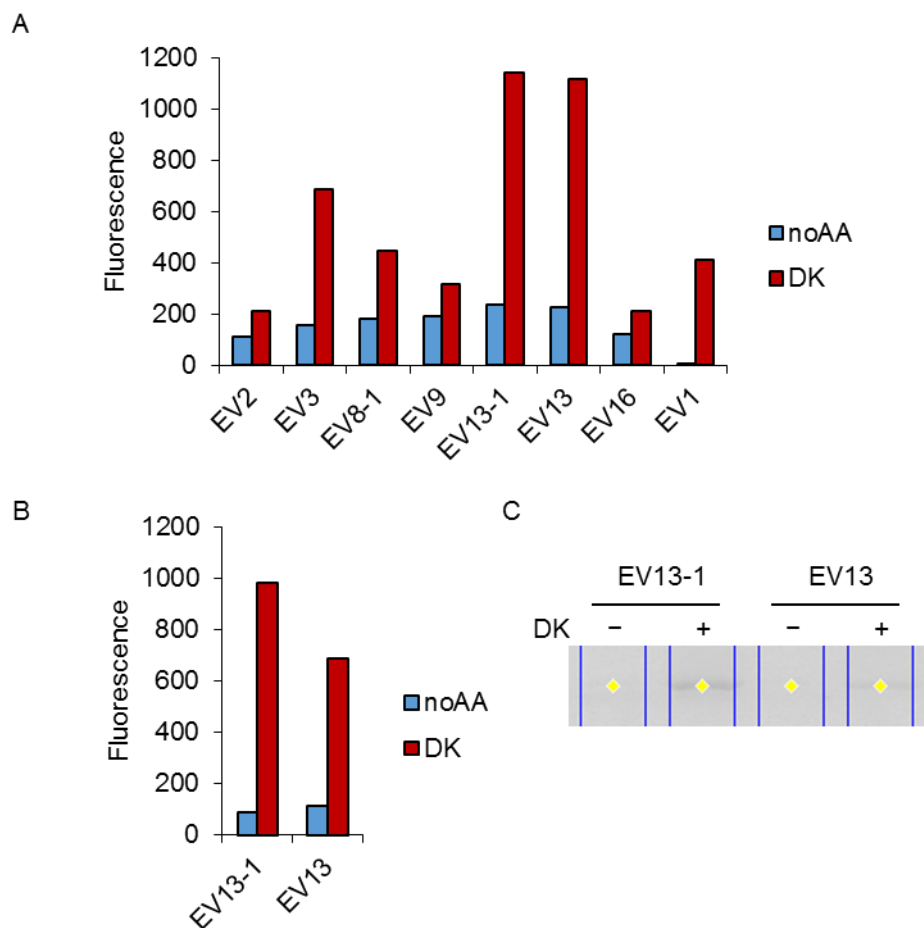


Figure 2.3. Screening of the PylRS panel for the incorporation of DK. (A) EV13-1 and EV13 showed highest incorporation in initial screening (5 ml) of the PylRS mutants. (B) EV13-1 performed better than EV13 in a large-scale expression experiment (25 ml). Fluorescence of the Ni-purified proteins was presented. (C) SDS-PAGE analysis for the incorporation of DK by EV13-1 and EV13.

EV13-1 contains a single Y349F mutation compared to wild-type PylRS, which is expected to increase the aminoacylation activity of PylRS.^{18a} For example, EV13-1 showed increased incorporation efficiency of Boc lysine and Alloc lysine.^{18a} We confirmed the expression of sfGFP-Y151TAG-DK, with the yield of 3 mg/L (Fig 2.4). The ESI-MS analysis of sfGFP-Y151TAG-DK suggests that the major peak agrees with the theoretical molecular weight (found: 28,375.51 Da, calculated: 28,375.95 Da). However, a minor mass peak at +16 Da was also detected (28,391.19

Da). We speculate that one of the sulfurs might be oxidized to S=O.

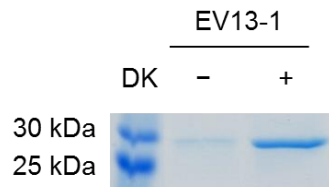


Figure 2.4. Incorporation of DK into sfGFP at the Tyr151 position. EV13-1 (with Y349F mutation) could incorporate DK, with the yield of 3 mg/L in *E. coli*.

sfGFP-Y151TAG-DK was then used for conjugation experiments. GFP was chosen because it has been previously conjugated to NPs,⁴⁷ and the incorporation of DK on GFP was confirmed by MS. The Au NPs with 40 nm diameter were used, since they are commonly applied for protein conjugation.⁴⁸ To this end, 40 nm standard Au NPs (150 nM, CytoDiagnostics) were mixed with sfGFP-Y151TAG-DK (1,000 nM). Following a standard protocol for adsorption of proteins onto Au NPs,⁴⁹ we incubated the mixture at 4 °C for 1 h. Unconjugated sfGFP was removed by centrifugation (13,200 rpm, 5 min), and the Au NP was resuspended in Tris buffer (300 µl, 20 mM, pH 7.0). This process was repeated twice. The NP before and after each centrifugation was subjected to a dynamic light scattering (DLS) measurement, a quantitative method for identifying the hydrodynamic size of NP.⁵⁰ Wild-type sfGFP (1,000 nM) was used as a negative control, as it was not expected to be conjugated to the Au NP in the absence of DK. Bovine serum albumin (BSA, 1,000 nM), a well-studied model protein for the NP conjugation,⁵¹ was used as a positive control. The result shows that the size of Au/sfGFP-DK conjugate is larger than that of Au/sfGFP-WT incubation, suggesting the successful conjugation of sfGFP-Y151TAG-DK to the Au NP (Fig 2.5).

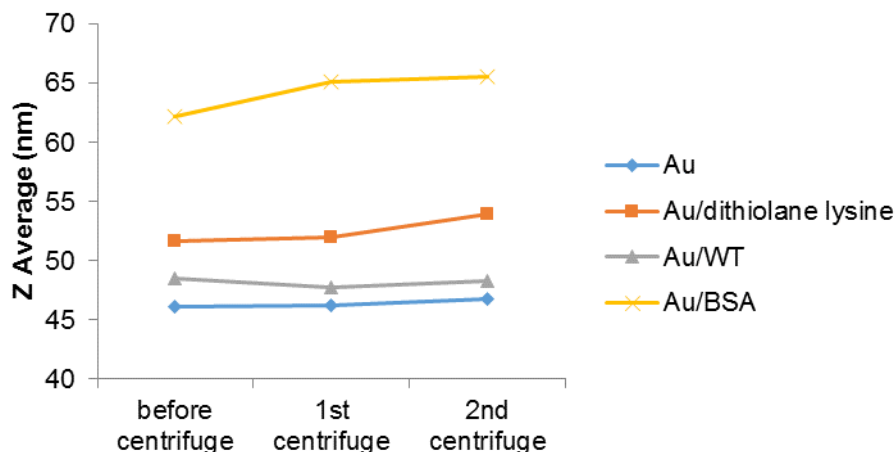


Figure 2.5. Dynamic light scattering (DLS) measurement of conjugates when using 40 nm standard Au NP. The experiment was performed in singlet. The Au NP (150 nM) and the sfGFP-Y151TAG-DK (1,000 nM) were incubated at 4 °C for 1 h. Wild-type sfGFP (1,000 nM) and BSA (1,000 nM) were used as negative control and positive control, respectively. The conjugation reaction mixture was centrifuged and resuspended in Tris buffer (300 µl, 20 mM, pH 7.0) to remove unconjugated sfGFP. This process was repeated twice. Measurements were taken before centrifugation, after first centrifugation, and after second centrifugation. The data was presented as Z-Average, which is the harmonic intensity-averaged particle diameter. Size increase compared to the Au: Au-BSA > Au-DK > Au-WT. No error bars are reported since the experiment was conducted in singlet.

The final conjugates were further analyzed on a 1.2% agarose gel (80 V for 45 min). The NP did not move on the gel (Fig 2.6). This is probably because the Au NP aggregates in the gel running buffer without the coating of protein (the 40 nm standard Au NP is not charged). It is also unexpected that the Au/sfGFP-DK presented similar mobility to that of the Au/sfGFP-WT, which probably suggests the non-specific binding of sfGFP-WT to the Au NP. The non-specific binding is due to the interaction between the negatively charged Au NP (in citrate buffer) and the positively charged residues on the wild-type protein (R29, K41, R73, K107, R122, PDB: 2B3P), which is a common phenomenon during NP conjugation.⁵² More effort is needed to understand the

contradictory results from DLS analysis and gel electrophoresis analysis.

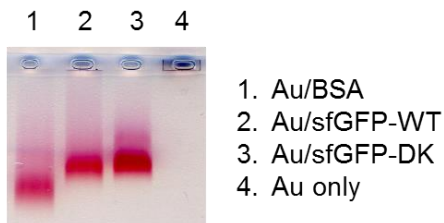


Figure 2.6. Gel electrophoresis analysis of conjugates when using 40 nm standard Au NP. Reaction samples were run on 1.2% agarose gel (run at 80 V, 45 min).

To reduce the non-specific binding of the wild-type sfGFP, we switched to 40 nm stabilized Au NP (Cytodiagnosics), which contains proprietary surfactant as stabilizer. Normal stabilizers used for Au NP include gelatin, polyvinylpyrrolidone, or polyvinyl alcohol.⁵³ This product is not ideal for the adsorption of protein, but the binding of thiolated ligands (for example, thiolated oligonucleotides; although literature reports are limited).⁵⁴ Stabilized Au NP and sfGFP-Y151TAG-DK were incubated for a longer conjugation time (44 hours), and the product was analyzed as previously performed. No size increase was observed for all samples in the DLS measurements (Fig 2.7). This suggested that no protein was conjugated to the Au NP. We speculated that the stabilizer on the Au NP might prevent the protein from binding.

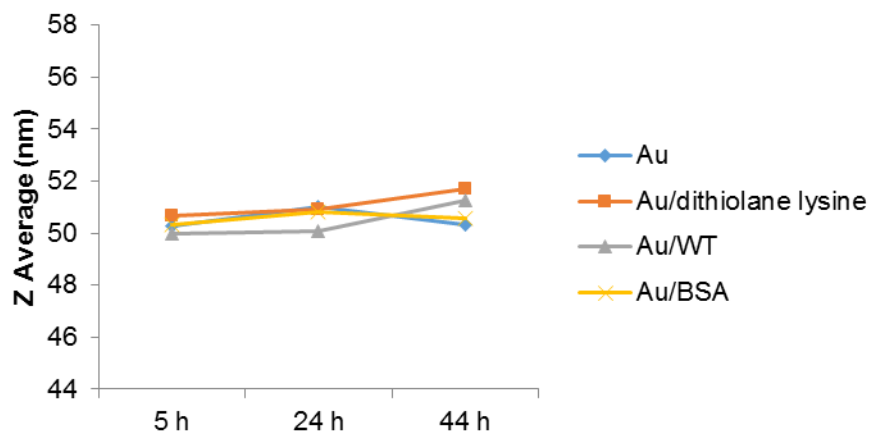


Figure 2.7. Dynamic light scattering (DLS) measurement of conjugates when using 40 nm stabilized Au NP. The Au NP (150 nM) and the sfGFP-Y151TAG-DK (1,000 nM) were incubated at 4 °C for designated time. Wild-type sfGFP (1000 nM) and BSA (1,000 nM) were used as negative control and positive control, respectively. Measurements were taken after 5 h, 24 h or 44 h of reaction. No size increase of conjugates was observed compared to the Au. Time point indicated the incubation time.

We further analyzed the samples on 1.2% agarose gel (run at 80 V, 45 min). Unexpectedly, Au/sfGFP-DK and Au/sfGFP-WT presented different mobility compared to that of Au/sfGFP-BSA and Au, although their DLS measurements were the same (Fig 2.8).

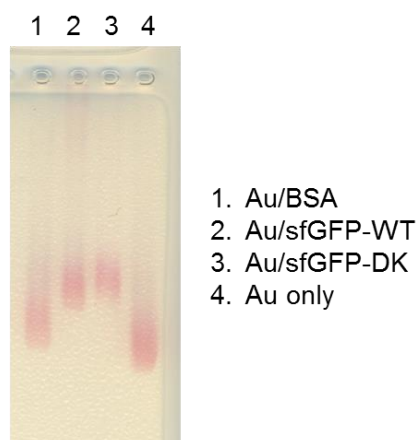


Figure 2.8. Gel electrophoresis analysis of conjugates when using 40 nm stabilized Au NP. Reaction samples were run on 1.2% agarose gel (run at 80 V, 45 min).

We then sent the protein to Dr. Vincent Rotello's lab for conjugation tests. They previously developed a nanoparticle-GFP conjugate-based method to measure protein concentration in human serum.⁵⁵ The GFP fluorescence was quenched in the nanoparticle-GFP conjugate, until the GFP was replaced by the serum protein and released from the particle. Different serum proteins showed different fluorescence response patterns, which were used to further calculate protein concentrations. We reasoned that the conjugation of GFP containing the dithiolane group could be monitored by the fluorescent quenching assay. To this end, Rubul Mout tested the fluorescent intensity in response to increasing concentration of nanoparticles (Fig 2.9). However, the sfGFP fluorescence remained the same, suggesting that no fluorescent quenching was observed. The result indicates that sfGFP-DK does not interact with their nanoparticles.

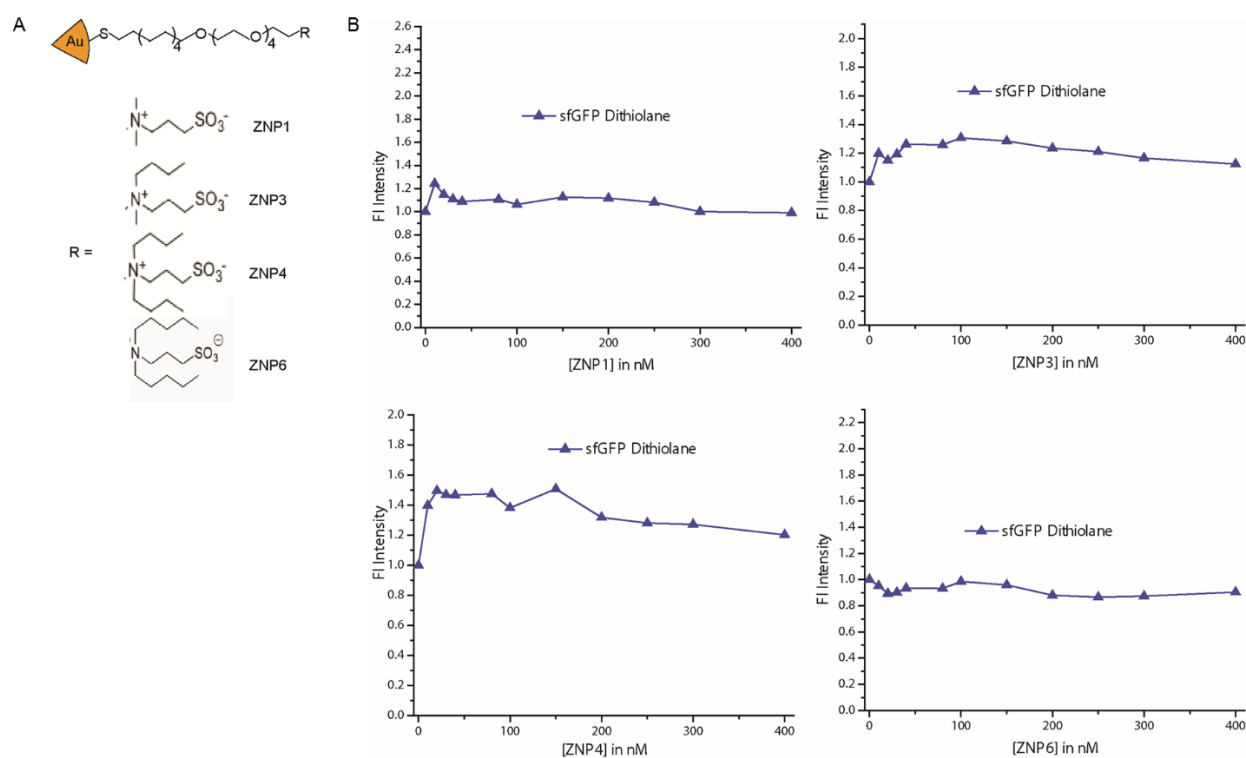


Figure 2.9. Nanoparticle quenching experiment suggested no interaction with sfGFP-DK. (A). Structures of the Au nanoparticles. (B) Fluorescent intensity with the increasing concentration of nanoparticles. Experiment performed by Rubul Mout (Rotello lab).

In conclusion, dithiolane lysine can be incorporated with a mutant PyIRS, EV13-1. sfGFP-Y151TAG-DK was expressed for Au NP conjugation. However, the 40 nm standard Au NP (Cytodiagnostics) displays non-specific binding with wild-type sfGFP, while the 40 nm stabilized Au NP is not able to conjugate to sfGFP. We also realized that the methodology to characterize sfGFP conjugation on Au NP is limited due to the quenching effect of Au NP on sfGFP fluorescence.⁵⁶ We therefore decided to work on the conjugation of an enzyme to Au NP through dithiolane lysine. We reasoned that the conjugation can be easily characterized by an enzymatic reaction. We also changed Au NP solutions to Au round wires, as they are easier to process during the wash step. Alaina McDonnell in our lab is now working on conjugation of bacterial alkaline phosphatase (BAP),⁵⁷ which activity can be reliably measured through p-nitrophenyl phosphate (pNPP)-based colorimetric method. Besides, similar to lipoic acid, the dithiolane could be used as a metal chelator.⁵⁸ The reduced dithiol group could also undergo As-based conjugation, allowing for in situ imaging of proteins in live cells.⁵⁹

2.1.2 Incorporation of azidoethyl lysine

Caged lysine was previously synthesized in our lab to control protein localization and function with light.^{19d, 60} Besides the photocontrol method, we are also interested in small molecule control of protein function. For example, with the genetic incorporation of an *ortho*-azidobenzoyloxycarbonyl amino acid by a mutant PyIRS, the protein function can be activated by a phosphine-mediated Staudinger reduction.⁶¹ We think similar strategy might be applied on an aliphatic azido lysine amino acid, which should be recognized by the PyIRS due to its small size. Such an amino acid may provide higher incorporation efficiency and faster deprotection rates; however, the small size may limit perturbation of protein function and structure. To this end, azidomethyl lysine (AzMK) and azidoethyl lysine (AzEK, more stable than AzMK⁶²) were synthesized by Qingyang Liu (Table 2.2). The azido group could be first reduced to an amine by triphenyl phosphine (PPh₃), and a lysine is generated following decarboxylation. The

incorporation of AzEK or AzMK at the active lysine residue of a protein will allow small molecule-triggered control of protein function.

We first performed the screening of five PylRS synthetase for AzEK. Among five synthetases, EV1 (wild-type PylRS) was found to incorporate AzEK (Fig 2.10A). Due to the structural similarity between AzEK and AzMK, we later tested incorporation of AzMK with EV1 without further screening.

We further confirmed the incorporation of AzEK (with the yield of 8 mg/L) and AzMK (with the yield of 4 mg/L) by SDS-PAGE (Fig 2.10B). However, the molecular weight of sfGFP-Y151TAG-AzEK determined by ESI-MS suggests that lysine, but not AzEK, is present at the incorporation site (found: 28,227.1 Da; calculated: 28,340.8 Da). The molecular weight of sfGFP-Y151TAG-AzMK suggests incorporation of lysine as well (found: 28227.9 Da; calculated: 28326.8 Da). Since it has been known that the wild-type PylRS does not recognize lysine as a substrate,⁶³ the azido group is probably reduced or cleaved otherwise either during the protein purification or during the ESI-MS analysis. We further tested incorporation of AzEK and AzMK in mammalian cells. However, both compounds were toxic to human embryonic kidney (HEK) 293T cells even at 0.5 mM concentration. Joshua Wesalo in our lab is now re-purifying the compounds and testing them in mammalian cells again.

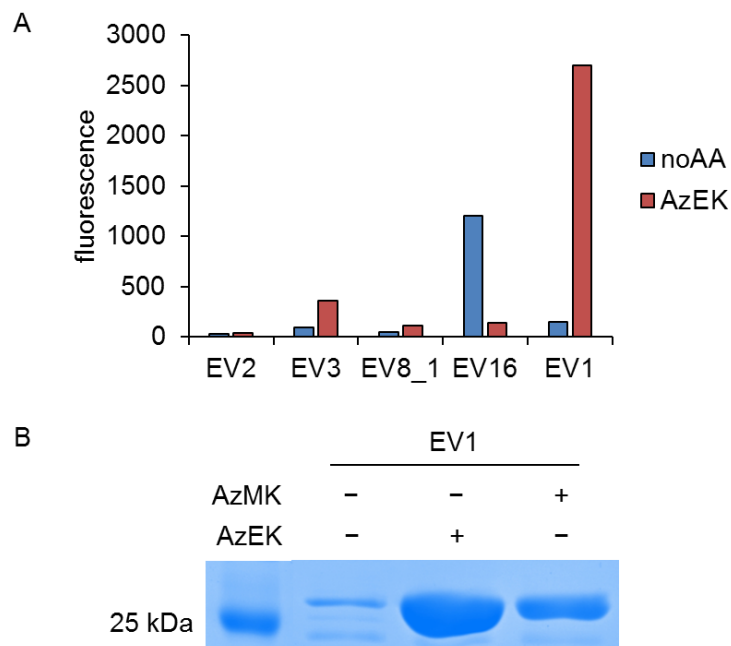


Figure 2.10. Screening of select PylRS mutants for the incorporation of AzEK and AzMK. (A) EV1 (wild-type PylRS) was the hit synthetase for the screening of AzEK. The same synthetase was tested with AzMK without further screening. (B) Incorporation of AzEK and AzMK into sfGFP at the Tyr151 position. Wild-type PylRS (EV1) can incorporate AzEK and AzMK, with the yield of 8 mg/L and 4 mg/L, respectively.

2.1.3 Incorporation of 2nd generation caged tyrosines

Although *o*-nitrobenzyl-*O*-tyrosine (ONBY) has been successfully applied to control tyrosine phosphorylation in mammalian cells by light,²² it has several drawbacks. The photolysis byproduct of ONBY is a reactive nitrosoaldehyde, which might be harmful to a biological sample.⁶⁴ Besides, ONBY has low solubility in media (up to 0.4 mM), which hampers its wider application in living cells. To overcome these issues, several other caged tyrosine analogs, MNPY, NPY and NPEY, were synthesized by Jessica Torres-Kolbus in our lab (Table 2.2). NPEY generates an *o*-nitrostyrene byproduct, which is expected to be less harmful in a biological context. NPY and MNPY can be decaged more efficiently than ONBY,⁶⁵ with an aldehyde and a ketone byproduct,

respectively.^{19d} NPY and NPEY also have better solubility in media (100 mM in 100% DMSO) compared to ONBY (40 mM in 100% DMSO).

We first performed the screening of the PylRS synthetase panel for these caged tyrosines. EV1 (wild-type PylRS) was found to incorporate both MNPY (Fig 2.11A) and NPY (Fig 2.11B). NPEY could be incorporated by multiple synthetases (EV16-1, EV16-4, EV16-5, EV17) (Fig 2.11C).

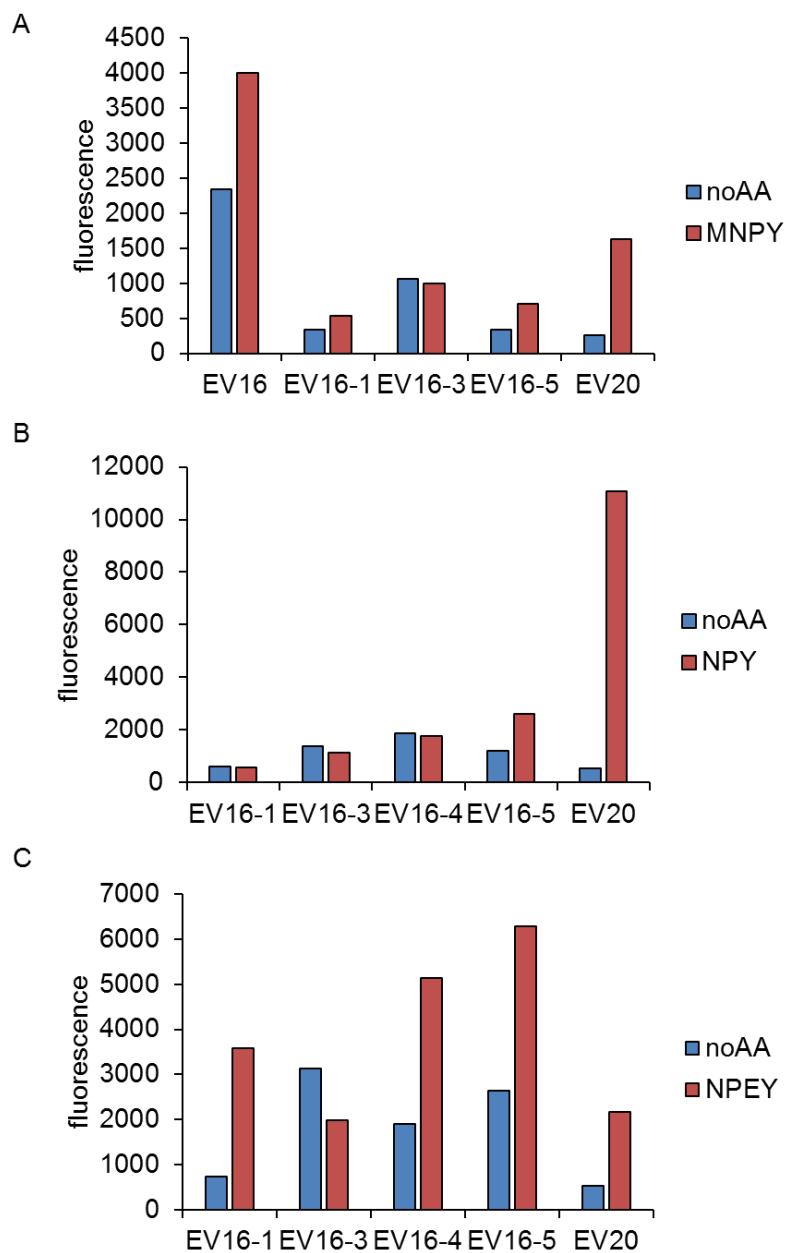


Figure 2.11. Screening of the PyIRS panel for the incorporation of MNPY, NPY and NPEY. EV20 is a hit synthetase for both MNPY and NPY. EV16-1, EV16-4, EV16-5 and EV20 are hit synthetases for NPEY.

We further confirmed the incorporation of MNPY by SDS-PAGE, with the yield of 15 mg/L (Fig 2.12A). EV20 can also incorporate NPY, with the yield of 18 mg/L (Fig 2.12B). The molecular

weight of sfGFP-Y151TAG-MNPY and sfGFP-Y151TAG-NPY, determined by ESI-MS, agrees well with their expected molecular weight (for MNPY, found: 28,455.89 Da; calculated: 28,455.90 Da; for NPY, found: 28,440.80 Da; calculated: 28,441.88 Da). EV20 is an excellent PylRS synthetase for the incorporation of caged tyrosine analogs, with minimal background expression. It contains an extra Y349F mutation compared to the previously reported synthetase that incorporates ONBY.²² The Y349F mutation is expected to increase the aminoacylation activity of PylRS.^{18a} Surprisingly, NPEY could be incorporated by multiple mutant PylRS (Fig 2.12C). Among them, EV16-1 and EV20 have medium incorporation with almost no background. EV16-4 and EV16-5 show better incorporation efficiency but also display a higher background. Compared to EV16-1, EV16-4 and EV16-5 harbor one or three extra mutations (Y349F for EV16-4, Y271M, L274A and Y349F for EV16-5). The Y349F mutation could facilitate the aminoacylation process, which might be responsible for the increased incorporation efficiency.^{18a} However, no reports so far demonstrated the background incorporation brought by this single mutation. Therefore, other mutations (N311A/C313A) might collectively change the binding pocket to the recognition of phenylalanine.^{20a} The molecular weight of sfGFP-Y151TAG-NPEY, when incorporated by EV20 synthetase, also agrees well with the expected molecular weight (found: 28,426.9 Da; calculated: 28425.9 Da). Ji Luo in our lab further encoded these new caged tyrosines in mammalian cells, and successfully controlled the activity of firefly luciferase and tobacco etch virus (TEV) protease with light.⁶⁶

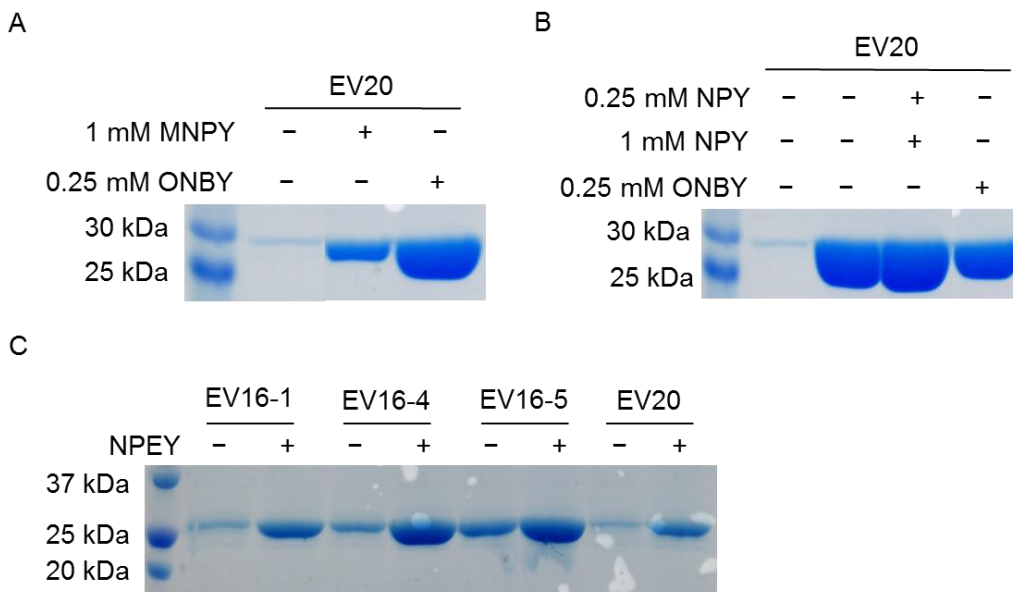


Figure 2.12. Incorporation of caged tyrosine into sfGFP at the Tyr151 position. ONBY was used as positive control for MNPY and ONBY. (A) Incorporation of MNPY by PylRS mutant EV20, with the yield of 15 mg/L. (B) Incorporation of NPY by PylRS mutant EV20, with the yield of 18 mg/L. (C) Incorporation of NPEY by multiple PylRS mutants (EV16-1, EV16-4, EV16-5, EV20).

2.1.4 Incorporation of photocaged thiotyrosine and photocaged azatyrosine

Thiotyrosine is a tyrosine analogue with a thiol group replacing the hydroxyl group. The thiol group on a protein has been widely applied to conjugate with thiol-reactive reagents.⁶⁷ Since an aryl thiol has a lower pKa compared to that of an aliphatic thiol, thiotyrosine is expected to be more reactive than cysteine in a bioconjugation reaction.⁶⁸ However, the evolution of an aaRS to incorporate thiotyrosine would be difficult due to its structural similarity to the endogenous tyrosine. A “photochemical disguise” strategy was previously used in our lab to overcome this hurdle.⁶⁹ *o*-Nitrobenzyl fluorotyrosine was first incorporated into a protein by an evolved aaRS. Upon UV irradiation, the caging group was removed and fluorotyrosine was present on the target protein. We therefore decided to site-specifically incorporate thiotyrosine into a protein by the same strategy. To this end, a photocaged thiotyrosine, *o*-nitrobenzyl thiotyrosine (ONBSY), was

synthesized by Rajendra Uprety in our lab (Table 2.2). Through a select PylRS mutant screening, we identified three PylRS mutants (EV16-2, EV16-5, EV17) that incorporate ONBYRS (Fig 2.13A). Among them, EV16-2 and EV16-5 bear higher background expression, due to N311A/C313A double mutations that likely lead to the recognition of phenylalanine.^{20a} EV17 showed less background expression, but also less incorporation efficiency. We therefore further analyzed EV16-5 and EV17 in a large-scale expression (25 ml).

Through SDS-PAGE analysis, we confirmed the expression of sfGFP-Y151-ONBSY, by both EV16-5 (with the yield of 7.7 mg/L) and EV17 (with the yield of 5.7 mg/L) (Fig 2.13B). An electrospray ionization spectrometry (ESI-MS) analysis of sfGFP-Y151TAG-ONBSY (using EV16-5) shows that the major peak of the spectrum agrees with its expected molecular weight (found: 28,413.18 Da; calculated: 28,413.93 Da). However, a minor peak that is 30 Da lower than expected was also detected (28,383.82 Da). This might be due to the reduction of the *o*-nitro group to an amino group in *E. coli*.⁷⁰ The sfGFP-Y151TAG-ONBSY was then irradiated under UV (365 nm) for 10 min and sent for ESI-MS analysis again. However, the expected molecular weight was not observed, and the two peaks before irradiation did not shift (found: 28,413.26 Da; calculated: 28,277.80 Da). This suggests that the caging group was not removed from sfGFP after 10 min of UV irradiation. The ESI-MS analysis of sfGFP-Y151TAG-ONBSY (using EV17) generated similar result. Previous studies suggested the UV irradiation (365 nm) of 1 h or 2 h for complete photolysis of ONB-protected thiols.⁷¹ Longer irradiation time might thus be needed for efficient decaging.

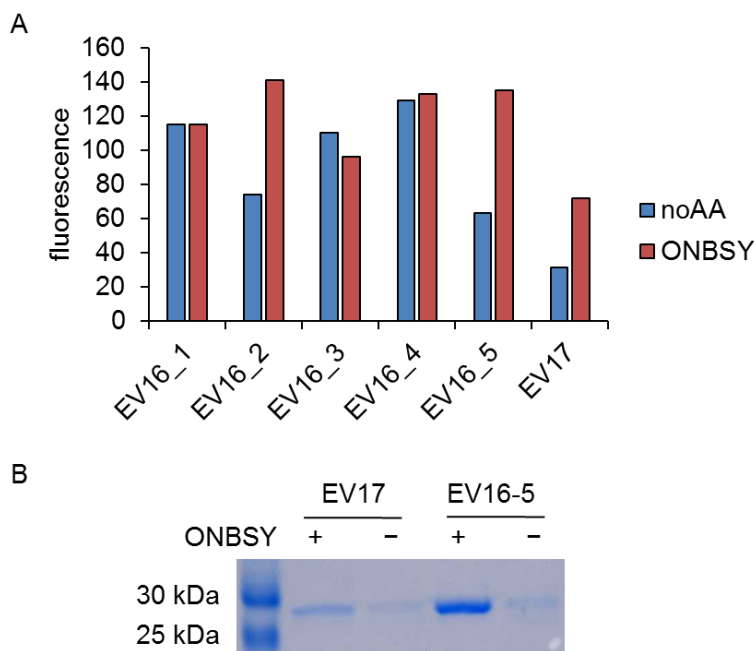


Figure 2.13. Screening of the PylRS panel for the incorporation of *o*-nitrobenzyl thiotyrosine (ONBSY). (A) EV16-2, EV16-5, and EV17 could incorporate ONBSY. (B) Incorporation of ONBSY into sfGFP at the Tyr151 position. EV16-5 and EV17 were further analyzed by SDS-PAGE.

We then incorporated ONBSY into sfGFP at the Tyr66 position (Fig 2.14A), as the fluorescence of sfGFP can serve as a reporter for tyrosine analogues at Tyr66 position.⁷² The fluorescence was measured before and after UV irradiation (from 1 min to 10 min). The same excitation (488 nm) and emission (510 nm) wavelengths were applied, as we did not expect much spectrum shift from tyrosine to thiotyrosine. The sfGFP-Y66TAG-ONBY, as a positive control, shows increased fluorescence after UV irradiation. However, the fluorescence of sfGFP-Y66TAG-ONBSY (5 μ M) stays constant before and after UV irradiation (Fig 2.14B). Although the decaging product slightly differs (thiotyrosine for ONBSY, tyrosine for ONBY), this experiment further supports that ONBSY is not able to decage under UV irradiation in a protein context. The decaging experiment of ONBSY compound in PBS buffer was later performed by Rajendra Uprety, but no decaged thiotyrosine was detected through HPLC analysis (data not shown).

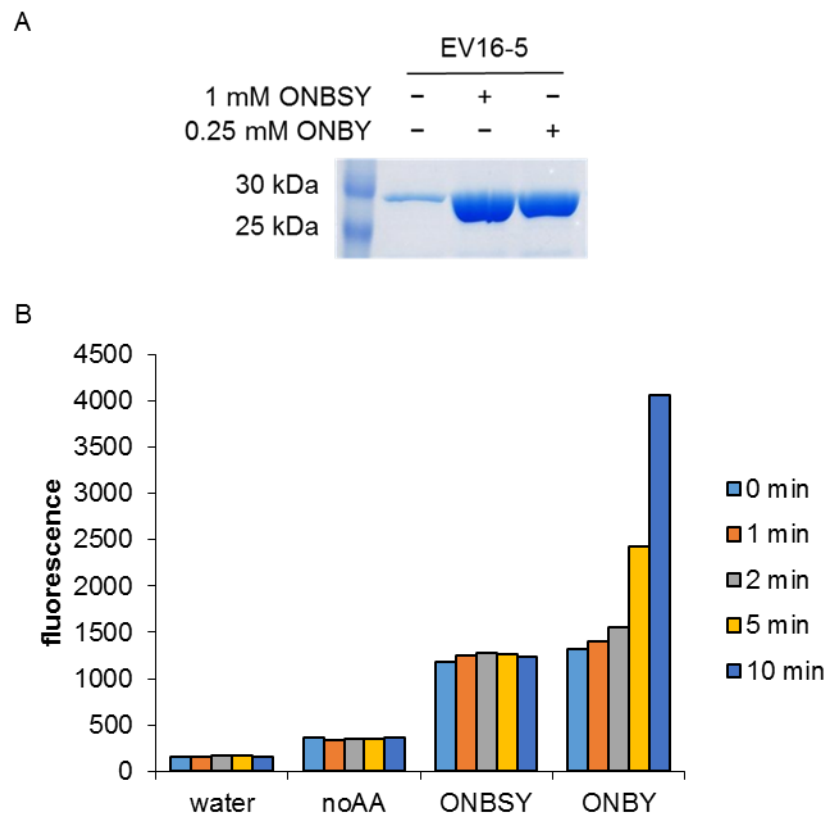


Figure 2.14. (A) Incorporation of ONBSY into sfGFP at the Tyr66 position. ONBY was used as a control for incorporation. (B) Fluorescence of sfGFP-Y66TAG-ONBSY (5 μ M) before and after UV irradiation. Different irradiation length was tested (1 min, 2 min, 5 min, 10 min). ONBY was used as a control. Excitation wavelength: 488 nm. Emission wavelength: 510 nm. Manual gain: 10.

We therefore decided to explore thiotyrosines with different caging groups. Nitropiperonyl methyl thiotyrosine (NPOMSY) and *o*-nitrophenylpropyl thiotyrosine (ONPPSY) were synthesized by Subhas Samanta in our lab (Fig 2.15A). Both of two compounds can be efficiently decaged in PBS buffer (tested by Subhas Samanta; data not shown). We then tried incorporation with EV16-5 and EV17 synthetases, using sfGFP-Y151TAG as a reporter. Protein expression was performed on a small-scale (200 μ l) in order to conserve the UAA, and overnight cell culture was directly used to measure fluorescence (Excitation: 488 nm. Emission: 510 nm). While the positive controls

worked, no incorporation was observed with NPOMSY or ONPPSY (Fig 2.15B). Screening of the rest of the PyIRS panel also did not reveal a hit synthetase (Fig 2.15C).

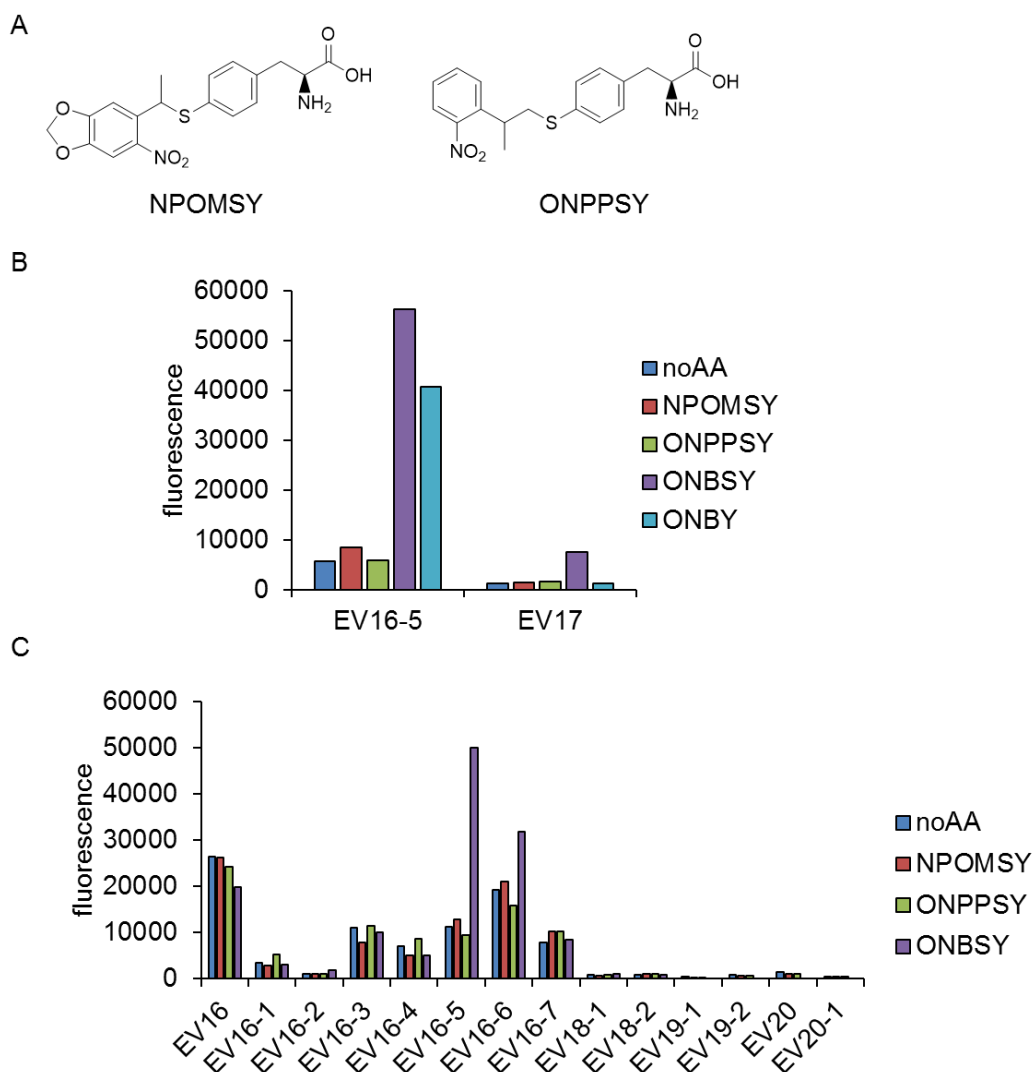


Figure 2.15. Incorporation test of NPOMSY and ONPPSY with synthetases panel. (A) Structures of NPOMSY and ONPPSY. (B) Fluorescence measurement of sfGFP with TAG at Tyr151 position. Protein expression was performed in a small-scale medium (200 μ l), and overnight cell culture was directly used to measure fluorescence (Excitation: 488 nm. Emission: 510 nm). Synthetases EV16-5 and EV17 (that were known to incorporate ONBSY) were first tested. ONBSY and ONBY were used as positive controls for synthetase EV16-5, and ONBSY was used as positive control for synthetase EV17. Manual

gain: 10. (C) Fluorescence measurement for other synthetases, performed under the same experimental conditions.

We further tried incorporation of another tyrosine analogue, azatyrosine, using the same strategy. Azatyrosine is a candidate drug for cancer treatment, since it converts a transformed cancerous phenotype back to the normal phenotype.⁷³ However, its mode of action remains elusive. It was proposed that azotyrosine could be misincorporated into the cellular proteome to replace tyrosine, which inhibits protein phosphorylation. Yet, the protein targets were not specifically identified.⁷³ Besides, the azatyrosine moiety undergoes tautomerization between pyridine and pyridone forms (Fig 2.16A).⁷⁴ The solvent polarity has been shown to play critical role in equilibrium state between lactam and lactim forms.⁷⁵ It can thus be used as a molecular probe for protein microenvironment (for example, polarity of the medium or local pH). A previous study discovered a mutant *E. coli* tyrosyl-tRNA synthetase that could incorporate azatyrosine more efficiently than tyrosine,⁷⁶ but no site-specific incorporation of azatyrosine has been reported up to date. To this end, a photocaged azatyrosine, O-caged 3-azatyrosine (ONBAY), was synthesized by Rajendra Uprety (Table 2.2). Through the screening of six PylRS synthetases (EV16-1, EV16-2, EV16-3, EV16-4, EV16-5, EV17) that were designed for the incorporation of phenylalanine derivatives), EV16-3 was found to incorporate ONBAY, with a yield of 1.5 mg/L (Fig 2.16B and C). EV16-3 contains three mutations compared to the wild-type PylRS: the N311A/C313A double mutations are beneficial for the incorporation of phenylalanine derivatives,^{20a} while the Y349F mutation is expected to increase the aminoacylation activity of PylRS. The ESI-MS analysis confirmed the incorporation of ONBAY (found: 28,398.56 Da; calculated: 28,398.86 Da). Upon UV irradiation (365 nm, 10 min), the caging group on sfGFP was fully removed (found: 28,262.90 Da; calculated: 28263.73 Da). Altogether, the data suggests successful incorporation and subsequent decaging of ONBAY on a protein.

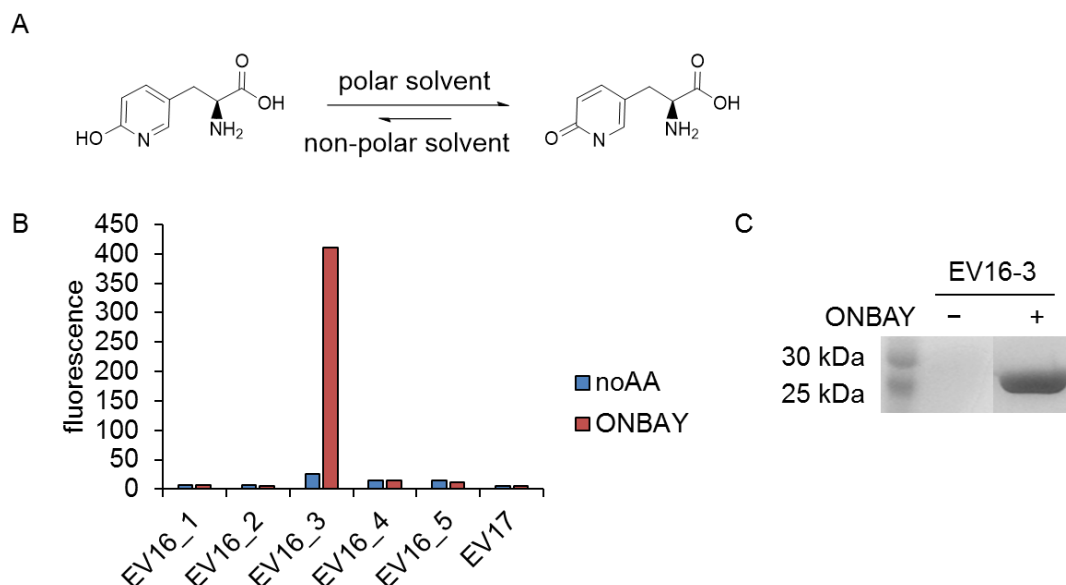


Figure 2.16. Incorporation of ONBAY into sfGFP at the Tyr151 position. (A) Azatyrosine undergoes tautomerization between pyridine and pyridone forms. (B) Screening of the PylRS synthetase panel for ONBAY reveals EV16-3 as a hit. (C) Incorporation of ONBAY by EV16-3, with a yield of 1.5 mg/L.

We then incorporated ONBAY into sfGFP at the Tyr 66 position, in order to potentially observe effects on fluorescence (Fig 2.17A). Unexpectedly, some background incorporation was observed even without supplying ONBAY. The protein was then irradiated with UV light (365 nm, 10 min), and the fluorescence was measured. No fluorescence of sfGFP-Y66TAG-ONBAY was observed after UV irradiation (Fig 2.17B). In contrast, UV irradiation increased the fluorescence of sfGFP-Y66TAG-ONBY. Through the ESI-MS analysis, we confirmed that the caging group is completely removed on azatyrosine moiety (found: 28,265.2 Da; calculated: 28,264.7 Da). We therefore speculated that the mature GFP chromophore with azatyrosine moiety might not be fluorescent. We reasoned that this is because the azatyrosine existed predominately in a lactam form in aqueous solution.⁷⁴

For both ONBSY and ONBAY decaging experiments, it should be noted that after UV irradiation, the samples were kept at 4 °C for chromophore maturation (see details in the

experimental section), while no literature support was found for chromophore maturation at this temperature. In the future, the samples should be kept at room temperature instead. Besides, we only performed the fluorescent measure at a defined wavelength (excitation at 488 nm, emission at 510 nm, see experimental), and did not take into consideration the possible spectra shift brought by UAA. Therefore, a full spectra scan may provide more information on the sfGFP maturation process.

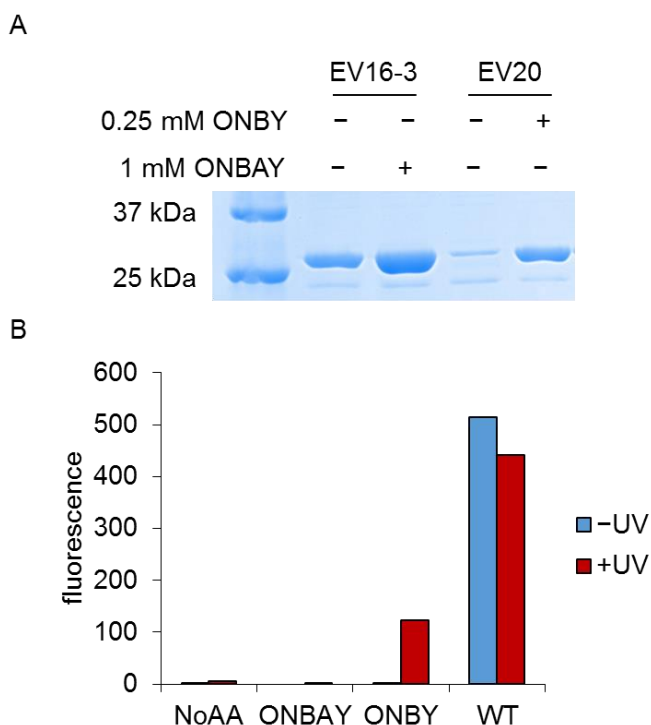


Figure 2.17. Incorporation and decaging test at Tyr 66 position on sfGFP. (A) ONBSY was incorporated into sfGFP at the Tyr 66 position. Background incorporation was observed. ONBY was used as a control. (B) Fluorescence of sfGFP-Y66TAG-ONBAY (5 μ M) before and after UV irradiation (365 nm, 10 min). sfGFP-Y66TAG-ONBY was used as a positive control. Excitation wavelength: 488 nm. Emission wavelength: 510 nm. Manual gain: 10.

In the future, we are planning to utilize azatyrosine as a biophysical probe in proteins. A previous study showed that azatyrosine undergoes pH dependent tautomerization.⁷⁴ We therefore

hypothesized that azatyrosine can be used to probe the local pH within a protein. Alaina McDonnell in our lab is working on incorporation of caged azatyrosine into different tyrosine positions in the ketosteroid isomerase (KSI) enzyme. KSI contains a triad of tyrosines (Y16, Y32, Y57) in its active site.⁷⁷ By generating KSI mutants with azatyrosine at different tyrosine positions, we will be able to probe the local pH effect on the activity of KSI enzyme. This provides a novel method to study hydrogen bonding network within KSI active site.⁷⁷⁻⁷⁸

2.1.5 Incorporation of alkene lysine

The thiol-ene reaction is a biorthogonal reaction that involves a radical-mediated addition of a thiol to an alkene upon UV irradiation.⁷⁹ Due to its specificity for alkenes and compatibility with aqueous solutions, the thiol-ene reaction has been widely applied in small molecule and polymer synthesis,⁸⁰ and glycoconjugation.⁸¹ Moreover, the reaction offers the possibility of using light to control both in space and time the formation of a stable thioether bond. Therefore, we are interested to expand the genetic code for the thiol-ene reaction. Jessica Torres-Kolbus in our lab synthesized a set of alkene lysines, and Chungjung Chou incorporated them in *E. coli* cells by the wild-type PylRS (Fig 2.18).⁸²

Chungjung Chou further validated the incorporation of 1-4 and 6 through ESI-MS analysis. However, the ESI-MS analysis for 7 and 9 were lacking. We therefore incorporated 7 and 9 into Myoglobin, and confirmed the incorporation of 7 (found: 18,479.80 Da, calculated: 18,480.09 Da) and 9 (found: 18,479.20 Da, calculated: 18,479.10 Da).

A

1 - 5 : $n = 1 - 5$

6

7

8 : $X = CH_2$
9 : $X = NH$

B

UAA	Yield (%)	Theoretical Mass (Da)	Experimental Mass (Da)
1	98	18480.09	18480.45
2	81	18494.10	18494.00
3	85	18508.27	18508.19
4	19	18522.13	18523.55
5	14	18536.15	-
6	79	18494.10	18494.03
7	100	18480.09	18479.80
8	9	18478.11	-
9	47	18479.10	18479.20

Figure 2.18. Genetic incorporation of alkene-lysine analogs into myoglobin by the wild-type PylRS. (A) Structures of alkenyl lysine derivatives bearing an ϵ -carbamate linkage (1–6), an inverted carbamate 7, an amide 8, and an urea 9. (B) Myoglobin comparative incorporation efficiencies (%) and ESI-MS results (theoretical and experimental mass). Adapted with permission from Torres-Kolbus, Jessica, et al. *PLOS ONE*, 2014, e105467.

Chungjung Chou further showed that, using this approach, alkene-modified sfGFP can either be labeled with fluorescent dansyl-thiol, or conjugated to lysozyme through its cysteine residue (Fig 2.19).⁸² Dansyl-thiol or lysozyme were reacted with alkenyl-sfGFP under UV irradiation (5 or 10 min). The protein labeling (with dansyl-thiol) and the protein conjugation (with lysozyme) were only observed in the presence of UV treatment, suggesting the necessity to trigger the thiol-ene reaction with light. Moreover, the successful conjugation between sfGFP and lysozyme demonstrated a novel method to generate protein heterodimer.

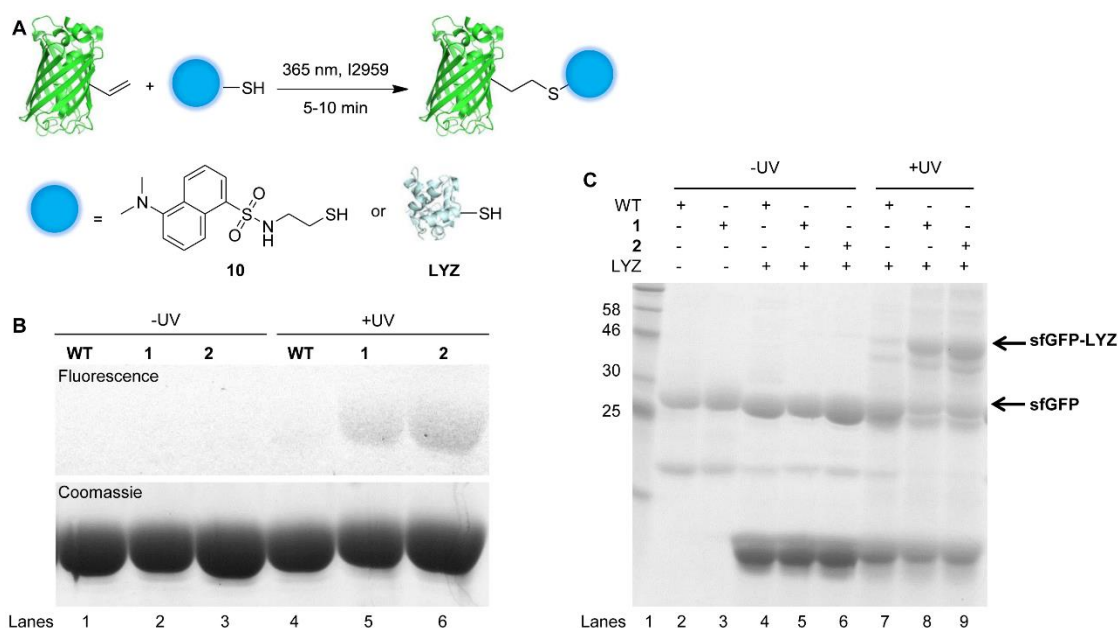


Figure 2.19. Alkenyl-sfGFP is fluorescently labeled with dansyl-thiol, and bioconjugated to lysozyme to assemble a non-linear protein dimer via the thiol-ene reaction. (A) sfGFP bearing an alkene functionality reacts photochemically with dansyl-thiol (10) or lysozyme (LYZ). (B) SDS-PAGE analysis demonstrates the labeling of alkenyl-sfGFP with 10 after 5 min of UV irradiation via thiol-ene ligation (lanes 5 and 6). Fluorescence (top) and Coomassie stain (bottom). (C) SDS-PAGE analysis shows mobility band shifts from 28 kD to 44 kD after samples were UV irradiated for 10 min (lanes 8 and 9), corresponding to the molecular weight of sfGFP-lysozyme conjugate. WT: wild-type sfGFP; 1 and 2: sfGFP carrying the corresponding UAA; LYZ: lysozyme. –UV: samples were not exposed to UV irradiation. +UV: samples were irradiated at 365 nm for 5 or 10 min. Adapted with permission from Torres-Kolbus, Jessica, et al. *PLOS ONE*, 2014, e105467.

In both bioconjugation strategies, the addition of sodium dodecyl sulfate (SDS) was found to be necessary for an efficient and specific conjugation reaction to alkene-labeled proteins within 5–10 min, in contrast to previously reported 1–2 h reaction times,⁸³ thus significantly reducing UV exposure. We performed circular dichroism (CD) spectroscopy of lysozyme (20 μ M) with and without SDS treatment (2%), and we observed increased intensity in the far-UV spectrum with

the addition of SDS (Fig 2.20). The result suggested that lysozyme is (at least partially) denatured under our experimental conditions.⁸⁴ This may result in more accessible cysteine residues and facilitate the thiol-ene bioconjugation reaction. Moreover, as a well-known surfactant, SDS has been proposed to form micelles in thiol-ene reactions for water-based polymerization reactions.⁸⁵ It is possible that the association of the proteins with micelles may increase their local concentration, thus further facilitating the reaction. This finding might suggest the use of SDS for more efficient thiol-ene reaction in future applications.

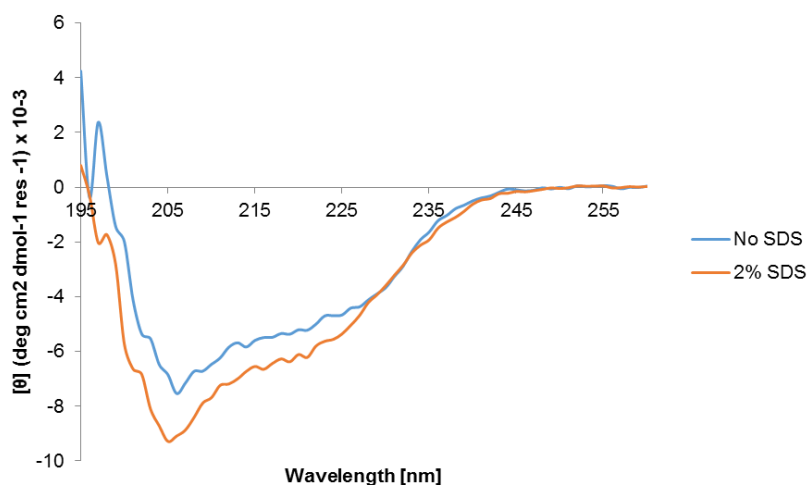


Figure 2.20. Circular dichroism (CD) spectrum of lysozyme (20 μ M) with and without SDS treatment (final concentration of 2%). CD spectra were collected from 195 to 260 nm with an integration time of 5 s and a bandwidth of 2 nm. Adapted with permission from Torres-Kolbus, Jessica, et al. *PLOS ONE*, 2014, e105467.

2.1.6 Incorporation of bipyridine lysine

Metals play an important role for the activity of many proteins.⁸⁶ Proteins that contains a metal ion cofactor (metalloproteins) are involved in biological processes such as photosynthesis, molecular oxygen reduction, water oxidation, respiration, and nitrogen fixation.⁸⁶ Due to their critical roles, there is a considerable interest to design proteins with novel metal-binding

functionalities.⁸⁷ UAA mutagenesis provides a versatile approach to install a metal-chelating amino acid, thus site-specifically introducing a metal binding site into proteins. For example, bipyridylalanine (BPA) was genetically incorporated into *E. coli* cells (Fig 2.21A).⁸⁸ The evolution with canonical amino acids and bipyridylalanine resulted in a new metal ion-binding motif.⁸⁹ However, the expression of the protein is only limited to *E. coli* cells. We therefore seek to genetically incorporate a metal-chelating amino acid using the PylRS/PylT pair, which is compatible in both bacterial cells and eukaryotic cells. To this end, Jessica Torres-Kolbus in our lab synthesized bipyridine lysine (BPK) (Table 2.2). The *N,N'*-bidentate group (2,2'-bipyridine) is known to strongly chelate transition-metal ions such as Fe²⁺, Cu²⁺, and Zn²⁺.^{88, 90}

Through the screening of four PylRS synthetases (EV3, EV10, EV13, EV13-1) that have been constructed at the time for the incorporation of lysine derivatives, EV13 (with Y271A and Y349F mutations) was found to incorporate BPK, with a yield of 2.4 mg/L (Fig 2.21B and C). EV13 shows wide substrate specificity, including *ortho*-azidobenzyloxycarbonyl lysine^{18a} and diazirinyl benzyloxycarbonyl lysine.⁹¹ However, the ESI-MS analysis showed that the molecular weight of sfGFP-Y151TAG-BPK is 48 Da less than expected (found: 28,361.2 Da, calculated: 28,409.9 Da). The ESI-MS analysis of Myo-4TAG-BPK suggests the same molecular weight difference (found: 18,531.3 Da, calculated: 18,578.4 Da). We confirmed that the molecular weight of BPK (through LC-MS) was correct (performed by Subhas Samanta). Strangely, previous publications did not report unexpected difference of molecular weight when incorporating a bipyridine group.⁸⁸ We later tested incorporation of BPK in mammalian cells, using mCherry-TAG-EGFP reporter. However, BPK (0.5 mM) is toxic to human embryonic kidney (HEK) 293T cells. In the future, LC-MS/MS analysis of the trypsin-digested protein might be pursued to understand the chemical composition at the desired position.

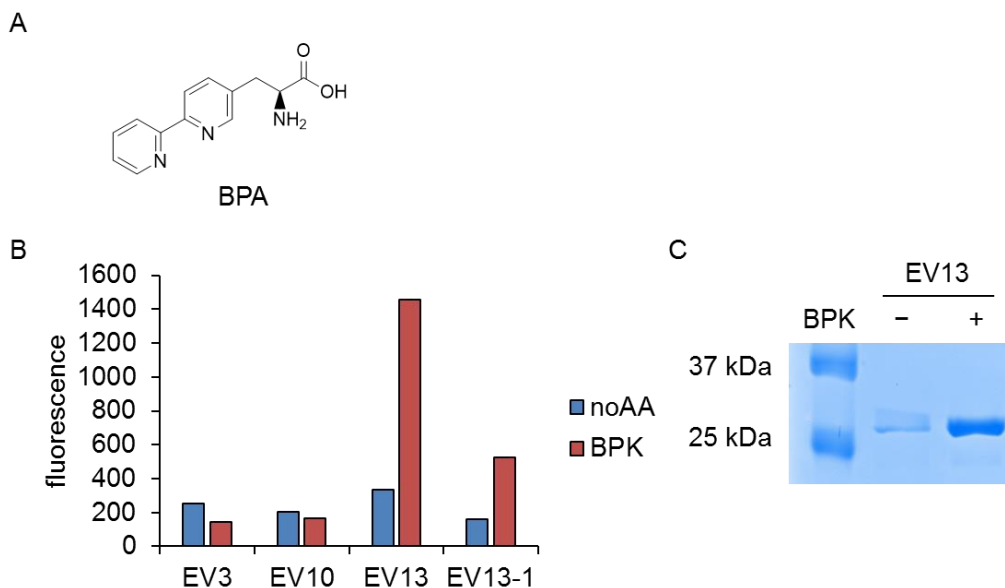


Figure 2.21. Incorporation of BPK into sfGFP at the Tyr151 position. (A) Structure of bipyrindylalanine (BPA) reported previously. (B) The screening of the PylRS panel for BPK revealed EV13 as a hit. (C) Incorporation of BPK by EV13 (with Y271A and Y349F mutations), with the yield of 2.4 mg/L.

In summary, we genetically incorporated several new unnatural amino acids through the screening of a synthetase panel. In the future, our focus will be to further characterize and apply these UAAs. For some UAAs (AzEK, AzMK, BPK), we will re-purify the compounds (to potentially reduce toxicity in cultured cells), and test their incorporations in mammalian cells. For UAA DK, we will conjugate an enzyme (bacterial alkaline phosphatase) to Au NP through dithiolane lysine. For UAA ONBAY, we will utilize azatyrosine as a biophysical probe in proteins, and study the local pH effect on the activity of KSI (ketosteroid isomerase) enzyme.

Experimental

Plasmid construction

The backbone of the pBAD plasmid was generated by digesting the pBAD-Myo4TAG-pylT plasmid (8,000 ng, gift from Dr. Ashton Cropp) with NcoI and NdeI restriction enzymes, followed

by gel purification of the band corresponding to the correct length of the backbone (~5,500 bp). sfGFP-Y151TAG (or sfGFP-WT) was amplified from the pTRC-sfGFP plasmid (gift from Dr. Ashton Cropp, Fwd: 5'-gacccatggggggttctcatcatc-3', Rev: 5'-catatgtattatacagttcatccatg-3') and digested with NcoI and NdeI restriction enzymes (37 °C, 2 h). The pBAD-sfGFP-Y151TAG-pylT plasmid was constructed by ligating the sfGFP-Y151TAG fragment into the pBAD backbone using NcoI and NdeI restriction sites (T4 ligase, 16 °C, overnight). The pBAD-sfGFP-WT-pylT plasmid was constructed in a similar way. pBAD-sfGFPY66TAG-pylT was made through introduction of the Y66TAG mutation by site-directed mutagenesis (Fwd: 5'-ctgtcaccaccctgacgtagggcgtgcaatgcttagc-3', Rev: 5'-gctaaagcattgcacgccctacgtcaggggtggtgacaag-3', TAG mutation underlined). Wild-type pBK-PylRS (*M. barker*) was a gift from Dr. Jason Chin.^{19e} PylRS mutants were generated by site-directed mutagenesis. The parent plasmid (5 ng) was amplified with primers containing the desired mutations (Table 2.3). DpnI (0.4 µl) was used to remove parent plasmid (37 °C, 30 min). The PCR product (3 µl) was then transformed into chemical competent cells (procedures for generating chemical competent cells are described in Appendix B). The plasmid was extracted from resulting colony and was confirmed by sequencing (GeneWiz).

Mutations	Primers
Y271M_L274A	Fwd: 5'- gatgctggccccgaccctgATGaactatGCgcgtaaactggatcgattc-3'
	Rev: 5'- gaatac gatccagtttacgcGCatagttCATcagggtcggggccagcatc-3'
A311G_A313G	Fwd: 5'- gaagaattcaccatggttgGctttgGccaaatgggcagcggctgc-3'
	Rev: 5'- gcagccgctgcccatttggCcaaagCcaaccatggtgaattcttc-3'
L270F	Fwd: 5'-cgatgctggccccgaccTTTatgaactatgcgcgtaaac-3'
	Rev: 5'-gtttacgcgcatagttcatTTTggtcggggccagcatcg-3'
L274M	Fwd: 5'-cgacctttataactatATGcgtaaactggatcgattc-3'
	Rev: 5'-gaatac gatccagtttacgCATatagttataaaaggctcg-3'
M274A	Fwd: 5'-cgacctttataactatGCGcgtaaactggatcgattc-3'
	Rev: 5'-gaatac gatccagtttacgGCGatagttataaaaggctcg-3'
Y349W	Fwd: 5'-gcatagctgcatggtgTGGggcgataccctggatattatg-3'
	Rev: 5'-cataatatccagggatcgccCCAacccatgcagctatcgc-3'
C313K	Fwd: 5'-ggttggtttAAAc aaatgggcagcggctgcacc-3'
	Rev: 5'-tgcccatttgTTTaaaggcaaccatggtgaattc-3'
L270G	Fwd: 5'-ccgatgctgagtcgaccGGCtgtaactatgcgtaaac-3'
	Rev: 5'-gtttacgcataatagttacaGCCggtcggactcagcatcgg-3'
C313G	Fwd: 5'-caccatggttaactttGGCaaatgggcagcggctgc-3'
	Rev: 5'-gcagccgctgcccatttgGCCaaagtaaccatggtg-3'
A267T	Fwd: 5'-gtgcctgcgtccgatgctgACCCcgaccctgtataactatc-3'
	Rev: 5'-gatagttatacagggctcggGGTcagcatcggacgcaggcac-3'
A267S	Fwd: 5'-gtgcctgcgtccgatgctgAGCCcgaccctgtataactatc-3'
	Rev: 5'-gatagttatacagggctcggGCTcagcatcggacgcaggcac-3'
L274C	Fwd: 5'-ccgaccctgtgtaactatTGCcgtaaactggatcgatac-3'
	Rev: 5'-gtatac gatccagtttacgGCAatagttacacagggctcg-3'
Y349F	Fwd: 5'-ctgcatggtgTTTggcgataccctggatattatgc-3'
	Rev: 5'-ggtatcgccAAAacccatgcagctatcgccac-3'

Table 2.3. List of primers used for generating PylRS mutants. Mutations are capitalized.

Synthetase screening

To rapidly screen synthetase panel over a target UAA, we employed a protocol with small-scale expression (5 ml) followed by fast protein purification (three phase partitioning).⁴⁰ The reporter plasmid (pBAD-sfGFP-Y151TAG-pylT) and the mutant PylRS were co-transformed (50 ng each) into *E. coli* Top 10 competent cells. A single colony was grown overnight (37 °C, 250 rpm) in LB supplemented with Kan (50 µg/ml) and Tet (25 µg/ml), and the overnight culture (50 µl) was added to LB (5 ml) supplemented with the designated unnatural amino acid (1 mM unless specifically

noted), Tet (25 µg/ml), and Kan (50 µg/ml). The stock solution of UAAs (100 mM, dissolved in 100% DMSO) were used unless specifically noted. Cells were grown at 37 °C, 250 rpm, and protein expression was induced with arabinose (0.1 %) when OD₆₀₀ reached 0.4 (measured by Nanodrop). After overnight expression at 37 °C, cells (5 ml) were pelleted (5,000 rpm, 10 min) and resuspended in 300 µl of Tris buffer (50 mM, pH 7.6), 200 µl of ammonium sulfate (4 M), and 500 µl of *tert*-butanol. The mixture was vigorously shaken for 1 min, followed by centrifugation at 4,000 rpm for 10 min. The bottom layer (~400 µl) was collected for fluorescence measurement and SDS-PAGE analysis.

As the PyIRS panel was expanding, screening all synthetases in 5 ml expression scale required large amount of unnatural amino acids. Therefore, we later employed a cell-based fluorescent test to identify hit synthetases. Like the previous expression protocol, a single colony was first grown overnight (37 °C, 250 rpm) in 3 ml of LB supplemented with Kan (50 µg/ml) and Tet (25 µg/ml). The overnight culture (2 µl) was added to LB (200 µl) supplemented with the designated unnatural amino acid (1 mM unless specifically noted), Tet (25 µg/ml), and Kan (50 µg/ml) in 1.0 ml deep-well plate (Nunc, #260251). Cells were grown at 37 °C, 250 rpm, and protein expression was induced with arabinose (0.1 %) when OD₆₀₀ reached 0.4 (measured by Nanodrop). After overnight expression at 37 °C, cells (200 µl) were collected in a 96-well plate, and sfGFP fluorescence was measured on a Tecan M1000 plate reader (excitation wavelength: 488 nm, emission wavelength: 510 nm, manual gain: 10). The hit synthetase was discovered if the cells harboring corresponding synthetase showed sfGFP fluorescence in the presence of UAA, but no fluorescence without UAA. No normalization to OD₆₀₀ was conducted. For the cell-based fluorescent test, we have not tested day-to-day variations and determined Z' factors. The dynamic range and the cut-off to define synthetase hits have not been studied.

Protein expression

To further confirm the hit synthetase discovered from the screening, large-scale protein expression (25 ml) was performed. pBAD-sfGFP-Y151TAG-pyIT and mutant PylRS were co-transformed (50 ng for each) into *E. coli* Top10 cells. A single colony was grown overnight (37 °C, 250 rpm) in LB (3 ml) supplemented with Kan (50 µg/ml) and Tet (25 µg/ml), and the overnight culture (250 µl) was added to LB (25 ml) supplemented with the designated unnatural amino acid (1 mM unless specifically noted), Tet (25 µg/ml), and Kan (50 µg/ml). Cells were grown at 37 °C, 250 rpm, and protein expression was induced with arabinose (0.1 %) when OD₆₀₀ reached 0.4 (measured by Nanodrop). After overnight expression at 37 °C, cells were pelleted (5,000 g, 10 min) and resuspended in phosphate lysis buffer (pH 8.0, 50 mM of phosphate, 6 ml). Triton X-100 (60 µl, 10%) was added to the mixture. The lysate was incubated on ice for 1 h, sonicated (power level 5, pulse 'on' for 30 sec, pulse 'off' for 30 sec, with a total of 4 min, 550 sonic dismembrator), and then centrifuged (13,000 g) at 4 °C for 10 min. The supernatant was transferred to a 15 ml conical tube, and Ni-NTA resin (Qiagen, 100 µl) was added. The mixture was incubated at 4 °C for 2 h with mild shaking. The resin was then collected by centrifugation (1,000 g, 10 min) and washed with lysis buffer (300 µl), this was repeated twice, followed by two washes with wash buffer (300 µl) containing imidazole (20 mM). The protein was eluted with elution buffer (300 µl) containing imidazole (250 mM). The purified proteins were analyzed by SDS-PAGE (10%), and stained with Coomassie Brilliant Blue. For protein mass spectrometry, the purified protein (using dialysis tubing with a 6,000-8,000 molecular weight cut-off, Fisher) was dialyzed in 50 mM ammonia acetate solution (1 L) overnight at 4 °C. The concentration of the dialyzed sample was determined by the intensity of protein band on SDS-PAGE, using BSA as the standard (ImageJ). An aliquot of the dialyzed protein (10 µl, concentration adjusted to 5 µM with water) was analyzed by electrospray ionization mass spectrometry (Thermo Scientific Q-Exactive Orbitrap, connected to a Dionex Ultimate 3000 UHPLC system), via a ProSwift RP-10R, 1 mm by 5 cm column, flow rate of 200 µl/min and ACN gradient (0.1% formic acid) 26-80% for 30 min. The mass

spectrometer was operated in ESI positive-ion mode with a capillary voltage of 3.5 kV and resolution at 17,500. The Protein Deconvolution 3.0 software was used for the data analysis.

For alkene lysine, analogues 7 and 9 were incorporated into myoglobin. Plasmids pBAD-Myo4TAG-pylT and EV1 were co-transformed into *E. coli* Top10 cells and selected with 25 µg/ml of Tet and 50 µg/ml of Kan. A single colony was used to inoculate 2 ml of LB medium containing the same antibiotics and grown overnight. Next, 250 µl of culture was used to seed 25 ml of LB culture containing 1 mM of the corresponding UAAs and antibiotics. Cells were then cultivated to OD₆₀₀=0.4 and 50 µl of 20% arabinose solution was supplemented to induce arabinose promoter driven expression. The cells were cultivated at 37 °C shaker overnight and harvested by centrifugation at 3,000 g for 10 min. Lysis of the cell was conducted by re-suspending the cell pellets with phosphate lysis buffer (pH 8.0, 50 mM of phosphate, 6 ml) and Triton X-100 (60 µl, 10%). After 1 hour of incubation at 4°C, cells were sonicated on ice to release the soluble portion and debris was removed by centrifugation. The cleared lysates were incubated with 100 µl of Qiagen Ni-NTA agarose slurry at 4°C to bind His-tagged myoglobin. The resin was then collected by centrifugation (1,000 g, 10 min) and washed with lysis buffer (300 µl), this was repeated twice, followed by two washes with wash buffer (300 µl) containing imidazole (20 mM). The protein was eluted with elution buffer (300 µl) containing imidazole (250 mM). For protein mass spectrometry, the purified protein (in dialysis tubing, 6,000-8,000 MWCO, Fisher) was dialyzed in 50 mM ammonia acetate solution (1 L) overnight at 4 °C. The concentration of the dialyzed sample was determined by the intensity of protein band on SDS-PAGE, using BSA as the standard (ImageJ). The dialyzed sample (10 µl, concentration adjusted to 5 µM with water) was analyzed by electrospray ionization mass spectrometry (Thermo Scientific Q-Exactive Orbitrap, connected to a Dionex Ultimate 3000 UHPLC system) as described above.

Conjugation of sfGFP-DK onto Au NP

The expression of sfGFP-Y151TAG-DK protein was performed as described above. For conjugation experiment with 40 nm standard Au NP, sfGFP-Y151TAG-DK (1,000 nM) was mixed with 40 nm standard Au NP (150 nM, Cytodiagnostics) in Tris buffer (300 μ l, 20 mM, pH 7.0) and incubated at 4 °C for 1 h. The mixture was pelleted (13,200 rpm, 5 min) and resuspended in Tris buffer (300 μ l, 20 mM, pH 7.0). This was repeated once. The solution before and after each centrifugation was subjected to a dynamic light scattering (DLS) measurement (Zetasizer Nano S, Malvern Instruments). The measurements were performed at 25 °C, and the z-average diameter was obtained from Malvern software. For electrophoresis analysis, Au NP solution (16 μ l) was mixed with 50% glycerol (4 μ l, in TBE buffer), and was analyzed on a 1.2% agarose gel (80 V, 45 min). For conjugation with 40 nm stabilized Au NP, sfGFP-Y151TAG-DK (1000 nM) was mixed with 40 nm stabilized Au NP (150 nM, Cytodiagnostics), and incubated at 4 °C for 44 h. The DLS and electrophoresis analyses were performed as described above.

DLS measurements

The hydrodynamic diameters of the sample solution (300 μ l) were measured using a Zetasizer Nano S DLS system (Malvern Instruments). The data was analyzed using the Zetasizer Nano S software (version 6.01). All sizes were based on intensity average, which was obtained using the non-negative least squares (NNLS) method. A detection angle of 90° was used for the size measurement.

Fluorescence activation assay with ONBSY and ONBAY

The sfGFP-Y66TAG-pyIT plasmid was co-transformed with corresponding PyIRS mutant (EV16-5 for ONBSY, EV16-3 for ONBAY) into *E. coli* Top10 cells. Protein expression of sfGFP-Y66TAG-ONBSY and sfGFP-Y66TAG-ONBAY was performed as described above. The cells and proteins were protected from light by wrapping flasks and tubes in aluminum foil during the expression and

purification process, in order to avoid unintended decaging. Purified sfGFP-Y66TAG-ONBSY or sfGFP-Y66TAG-ONBAY protein sample (5 μ M) was placed in a PCR tube and was irradiated with a UV transilluminator for designated time (365 nm, VWR). The protein was placed in 4 °C for 2 h, which may have negatively impacted proper chromophore maturation. The fluorescence of protein with or without UV irradiation was then measured on a Tecan M1000 plate reader (excitation wavelength: 488 nm, emission wavelength: 510 nm, manual gain: 10), however, wavelengths may have shifted due to ONBS incorporation. sfGFP-Y66TAG-ONBY protein sample was used as positive control.

Circular Dichroism (CD) analysis of lysozyme

CD experiments were performed on an Olis Circular Dichroism Spectrophotometer using 0.1 cm quartz cuvettes. A solution containing 19 mg of lysozyme in 10 mM phosphate buffer pH 7.4 (1.32 ml) was prepared. SDS was added to the final concentration of 2%. The lysozyme concentration was diluted to 20 μ M for the CD experiment, and CD spectra were collected from 195 to 260 nm in 1 nm increments with an integration time of 5 s and a bandwidth of 2 nm.

2.2 Optimizing Incorporation Efficiency in Mammalian Cells

Although UAA mutagenesis has been widely applied in both bacterial and mammalian cells,² the expression level of the mutant protein containing UAA is often lower than that of the wild-type protein (conservative estimates range from 1% to 10%, depending on different protein targets; however, wild-type expression levels have been observed as well). Therefore, improving the incorporation efficiency is a crucial step for broad application of this technology. An enhanced system for the UAA mutagenesis in *E. coli* was developed by Schultz lab,⁹² allowing for the expression of mutant protein with the yield of 100 mg/L. An evolved sequence context (upstream and downstream of an amber codon) for highly efficient amber suppression efficiency was also discovered to further boost the expression level in *E. coli*.⁹³ Yet, it is more critical to improve the efficiency of UAA incorporation in mammalian cells, as many cell-based experiments require the target protein to be highly expressed in a specific cellular compartment. Moreover, in mammalian cells, the poor protein yield cannot be easily solved by bulk expression as in bacterial cells. To improve the delivery of the necessary components of UAA mutagenesis to mammalian cells, which was normally achieved through liposome-mediated transfection, a baculovirus-based delivery system was recently developed.⁹⁴ This system allows for incorporation of multiple UAAs into a protein in mammalian cells.⁹⁵ In this system, two different suppressor tRNAs (*D. hafniense* and *M. mazei*) were expressed from two different polymerase III promoters (U6 promoter for *D. hafniense* and H1 promoter for *M. mazei*). The PylTs from *D. hafniense* and *M. mazei* were orthogonal to *E. coli* and mammalian cells, thus suitable for genetic code expansion.² Both U6 and H1 promoters have been applied for the expression of short RNAs in mammalian cells.⁹⁶ The use of different tRNAs and promoters avoided possible loss of tRNA cassette through homologous recombination.⁹⁴

Previously, our lab used a two-plasmid system for the genetic code expansion in mammalian cells.^{19d} On the first plasmid, the expressions of the PylRS and the protein of interest were driven by two separate CMV promoters. On the second plasmid, four copies of the PylT (from *M. barkeri*) were encoded under control of four separate U6 promoters. In a typical experimental setting, the two plasmids (200 ng each) were transfected into mammalian cells in a 96-well plate with a proper transfection reagent (branched PEI, 1.5 μ l). The culture medium was supplemented with the designated UAA (1 mM), and the cells were analyzed after 24 or 48 h.

To improve the efficiency of our current two-plasmid system, we asked 1) whether an H1 promoter can substitute the U6 promoter for the efficient expression of PylT, and 2) what the minimal number of copies of the PylT cassette is for efficient incorporation (to avoid possible homologous recombination). We first designed a PylT (*M. barkeri*) expression cassette (driven by an H1 promoter), and cloned the H1-PylT cassette into a pcDNA3 backbone (Invitrogen). The resulting construct, termed pH1-PylT, is functional in mammalian cells (Fig 2.22, row 3). However, it is not as efficient as our previous PylT-expression plasmid (Fig 2.22, row 4). We then generated a previously reported PylT expression cassette, where one copy of *D. hafniense* PylT is driven by an U6 promoter, and one copy of *M. barkeri* PylT is driven by an H1 promoter.⁹⁴ This U6-PylT-H1-PylT cassette was inserted into a pcDNA3 backbone (Invitrogen). The resulting construct, termed pU6-PylT-H1-PylT, shows similar efficiency as our previous PylT-expression plasmid in mammalian cells (Fig 2.22, row 2). Inspired by this result, we inserted another copy of the U6-PylT-H1-PylT cassette into the pU6-PylT-H1-PylT construct, thus placing four copies of PylT in one plasmid. Yet, the resulting construct, termed pU6-PylT-H1-PylT-V2, did not significantly increase the incorporation efficiency (Fig 2.22, row 1).

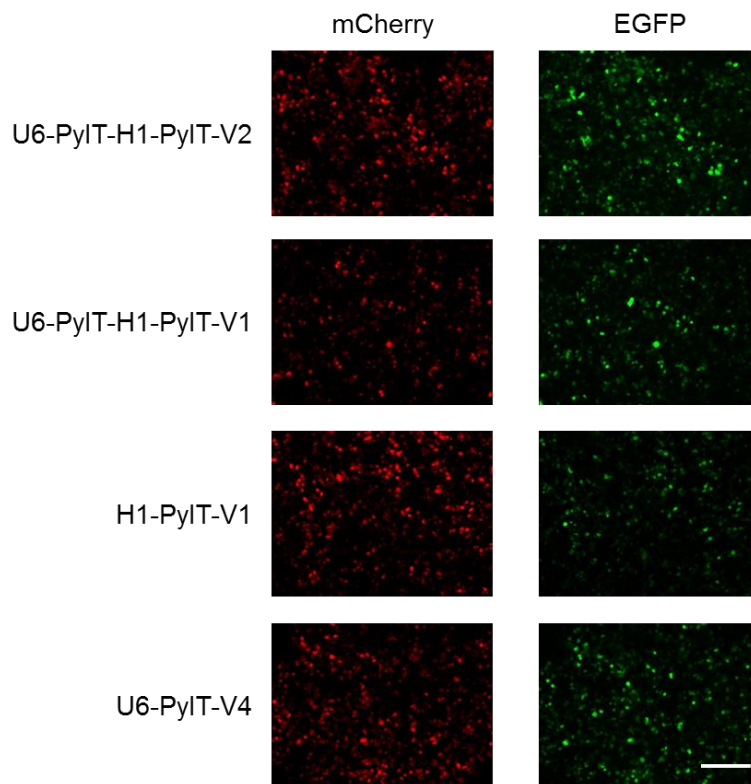


Figure 2.22. Incorporation efficiency of different PylT expression constructs. HEK 293T cells were transfected with PylT expression vector and pmCherry-TAG-EGFP-PCKRS (200 ng each). The PCK (1 mM) was supplemented in the culture medium. Cells were imaged after 48 h. Scale bar indicates 200 μ m. The PCK was incorporated into mCherry-TAG-EGFP. The EGFP expression indicated incorporation efficiency. The mCherry expression was a control for transfection efficiency.

As we worked on this project, Jason Chin's lab reported an optimized PylRS/PylT expression system.⁹⁷ To increase the expression level of the pylT, they placed four copies of the PylT (driven by four separate U6 promoters) on two separate plasmids, one expressing the protein of interest and one expressing the PylRS. They observed a significant increase in the PylT expression (based on Northern blot), as well as the incorporation efficiency. We tested their plasmid system using the wild-type PylRS, and confirmed that the incorporation efficiency was indeed significantly improved compared to the previous system where the PylT was placed only

on one plasmid (Fig 2.23, row 1 and row 3). In this test, Alloc-Lys (1 mM) was incorporated into the mCherry-TAG-EGFP (reporter used in our lab, row 1 and row 3) or sfGFP-150TAG (reporter used in Chin lab, row 2). Strong EGFP expressions were observed when the PylT was placed on both plasmids (row 2 and row 3). This result demonstrated that the amount of PylT was a limiting factor for efficient UAA incorporation in mammalian cells.

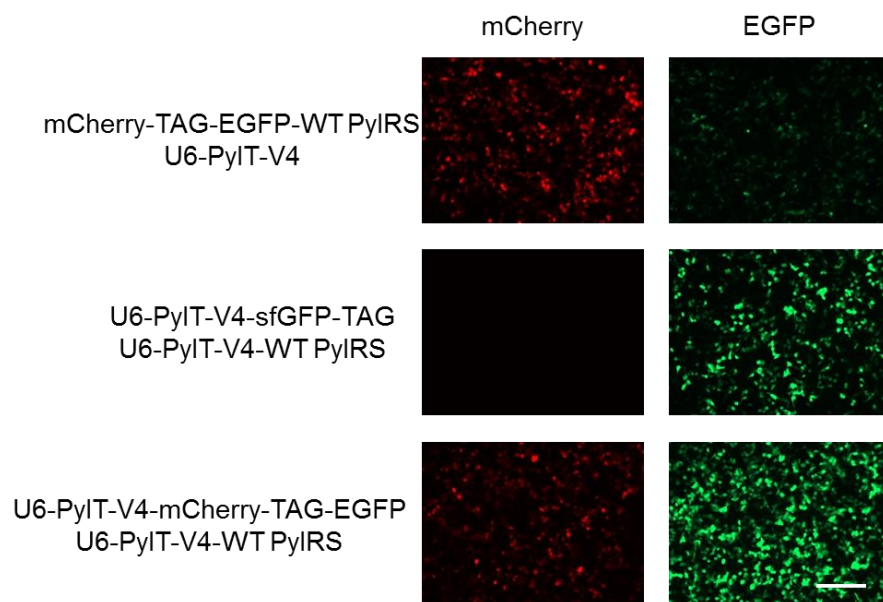


Figure 2.23. Incorporation efficiency was improved when the PylT was placed on both plasmids. HEK 293T cells were transfected with annotated plasmids (200 ng each), and Alloc Lys (1 mM) was supplemented in the culture medium. Cells were imaged after 48 h. Scale bar indicates 200 μ m. The Alloc Lys was incorporated into mCherry-TAG-EGFP or sfGFP-150TAG. For the mCherry-TAG-EGFP construct, the EGFP expression indicated incorporation efficiency. The mCherry expression was a control for transfection efficiency. For the sfGFP-150TAG construct, the sfGFP expression indicated incorporation efficiency. No control for transfection efficiency was used.

We then explored if the same strategy could be applied to our expression system design. To this end, a mutant PylRS (HCKRS, with Y271A and L274M mutations) and a reporter (mCherry-TAG-EGFP) were inserted into the pU6-PylT-H1-PylT construct separately. The

resulting constructs, pU6-PyIT-H1-PyIT-HCKRS and pU6-PyIT-H1-PyIT-mCherry-TAG-EGFP, were co-transfected to HEK 293T cells (200 ng each). HCK (1 mM) was supplemented in the culture medium. The incorporation efficiency of our new expression system (based on the EGFP expression) was significantly higher than that of the previous system where the PyIT cassette was placed only on one plasmid (Fig 2.24, row 1 and row 3). Gratifyingly, the efficiency of our system was similar to that of the system from Jason Chin's lab (Fig 2.24, row 2 and row 3).

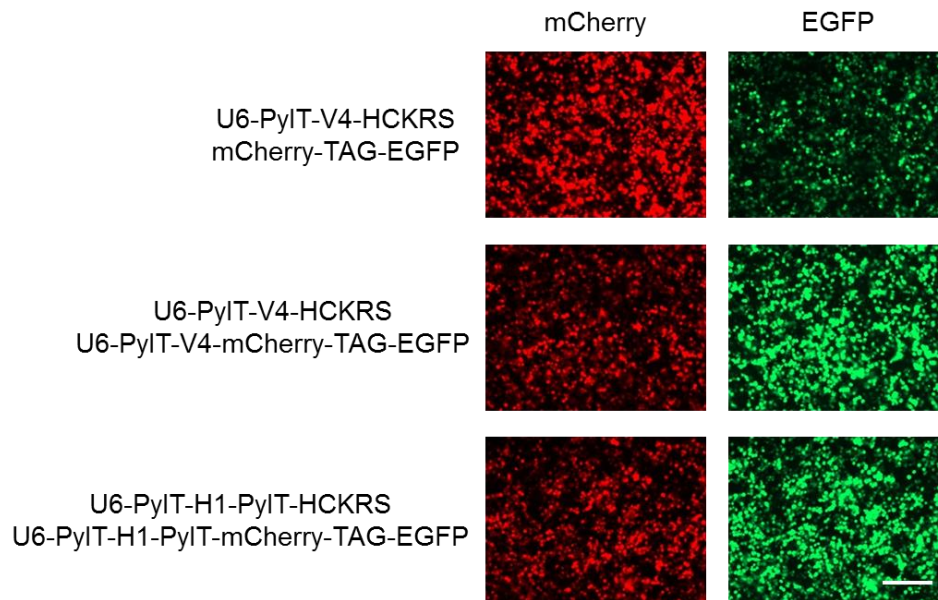


Figure 2.24. Incorporation efficiency of different systems comprising PyIT cassette on both plasmids. HEK 293T cells were transfected with the annotated plasmids (200 ng each). HCK (1 mM) was supplemented in the culture medium. Cells were imaged after 48 h. Scale bar indicates 200 μ m. HCK was incorporated into mCherry-TAG-EGFP. The EGFP expression indicated incorporation efficiency. The mCherry expression was a control for transfection.

To compare the protein expression from different systems, we further performed western blot analysis. HEK 293T cells were transfected with the designated plasmids (1,000 ng each). The culture medium was supplemented with or without HCK (1 mM). After 48 h, cell lysate was collected for western blot (with HA and GAPDH antibodies). Our new expression system (pU6-

PyIT-H1-PyIT-mCherry-TAG-EGFP-HA and pU6-PyIT-H1-PyIT-HCKRS, lane 2) presented highest expression of the full length mCherry-TAG-EGFP-HA (Fig 2.25), which is slightly better than the expression system from Jason Chin's lab (pU6-PyIT-V4-mCherry-TAG-EGFP-HA and pU6-PyIT-V4-HCKRS, lane 1). Not surprisingly, our previous expression system (pmCherry-TAG-EGFP-HA and pU6-PyIT-V4-HCKRS, lane 3) showed a lower expression level. We also tested the mixed use of plasmids from our lab and Chin lab (lane 4 and lane 5), and the use of pU6-PyIT-V4-mCherry-TAG-EGFP-HA (Chin lab) and pU6-PyIT-H1-PyIT-HCKRS (our lab) showed similar incorporation efficiency as our new expression system (lane 4). In summary, we generated a new expression system, where two copies of PyIT (driven by an U6 and an H1 promoter, respectively) were placed on two separate plasmids. The new system showed higher incorporation efficiency than the previous system, where the PyIT cassette was placed only on one plasmid.

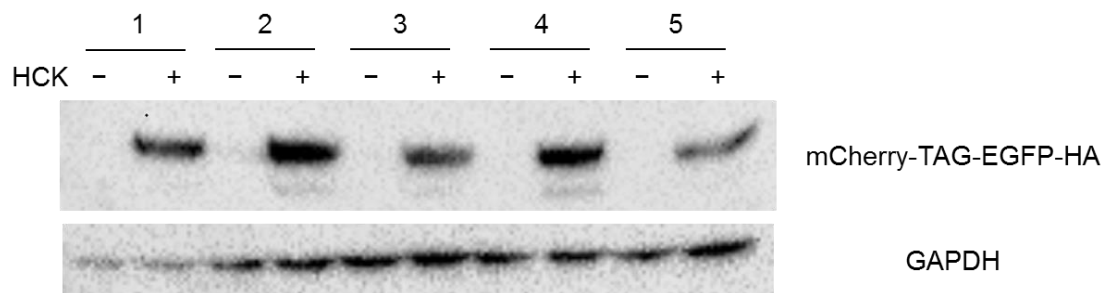


Figure 2.25. Western blot analysis for comparing the incorporation efficiency of different systems. HEK 293T cells were transfected with the designated plasmids (1,000 ng each). Cells were supplemented with or without HCK (1 mM). After 48 h, cell lysate was collected for western blot (with HA and GAPDH antibodies). The plasmids used for each group were as follows: 1. pU6-PyIT-V4-mCherry-TAG-EGFP-HA and pU6-PyIT-V4-HCKRS; 2. pU6-PyIT-H1-PyIT-mCherry-TAG-EGFP-HA and pU6-PyIT-H1-PyIT-HCKRS; 3. pmCherry-TAG-EGFP-HA and pU6-PyIT-V4-HCKRS; 4. pU6-PyIT-V4-mCherry-TAG-EGFP-HA and pU6-PyIT-H1-PyIT-HCKRS; 5. pU6-PyIT-H1-PyIT-mCherry-TAG-EGFP-HA and pU6-PyIT-V4-HCKRS.

As we sought to further improve the efficiency of two-plasmid system, the Arbely lab reported a single-plasmid system for efficient genetic code expansion in mammalian cells.⁹⁸ They combined all required genetic components for UAA mutagenesis into a single vector, and achieved high incorporation efficiency with just conventional transfection reagents. On their plasmid, two copies of PylT expression cassette (each cassette comprising a U6-driven *M. mazei* PylT and an H1-driven *D. hafniense* pylT) were placed on the same plasmid, but in opposite directions. The PylRS was driven by a CMV promoter, and the reporter was driven by an EF1 α promoter. Wenyuan Zhou in our lab tested their construct, and found that their single-plasmid system showed even higher incorporation efficiency than our two-plasmid system (Fig 2.26). This is most likely because the transfection efficiency for one plasmid is higher than that for two plasmids.

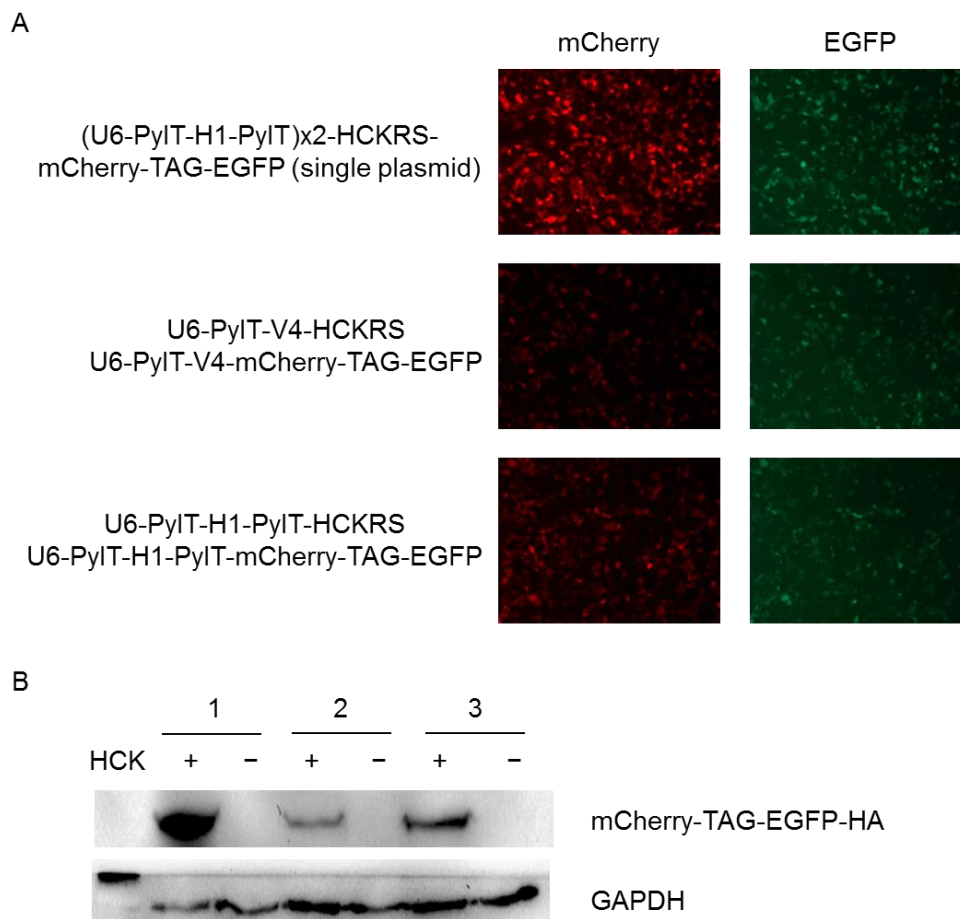


Figure 2.26. Comparison of the one-plasmid system and the two-plasmid system (data generated by Wenyuan Zhou). (A) HCK was incorporated into mCherry-TAG-EGFP. The EGFP expression indicated incorporation efficiency. The mCherry expression was a control for transfection. (B) western blot analysis for comparing the incorporation efficiency of different systems. The plasmids used for each group were as follows: 1, p(U6-PyIT-H1-PyIT)x2-HCKRS-mCherry-TAG-EGFP-HA (single plasmid). 2, pU6-PyIT-V4-mCherry-TAG-EGFP-HA and pU6-PyIT-V4-HCKRS. 3, pU6-PyIT-H1-PyIT-mCherry-TAG-EGFP-HA and pU6-PyIT-H1-PyIT-HCKRS.

In summary, we generated several new expression vectors for the incorporation of UAAs in mammalian cells. We utilized a U6 promoter for *D. hafniense* PyIT expression, and an H1 promoter for *M. mazei* PyIT expression. PyIT from different species were used to minimize the

possible loss of PyIT cassette via homologous recombination.⁹⁴ The genes for the PyIRS and the protein of interest were placed on two separate PyIT expression cassettes, and the incorporation efficiency was higher than our previous system where only one PyIT expression cassette was present, and was similar to the new expression system from the Chin lab (where four copies of PyIT were placed on both plasmids). The Arbely lab further found, and we confirmed, that the combination of all required genetic components into one single vector further boosted the incorporation efficiency.⁹⁸ In the future, we are going to apply this single-plasmid system to other protein targets.

Experimental

Plasmid construction

To construct the pH1-PyIT plasmid, the H1-PyIT cassette was ordered as a gBlock (200 ng, IDT, sequence in Talbe xx). The gBlock was dissolved in water (20 μ l), and digested with BglII (0.2 μ l) and BamHI (0.2 μ l) at 37 °C for 2 h. The pcDNA3 plasmid (8,000 ng, Invitrogen) was digested with BglII (2 μ l) and BamHI (2 μ l) at 37 °C for 2 h, and the backbone was extracted from a 0.8% agarose gel (run at 80 V for 45 min, expected band at ~ 5,400 bp). The H1-PyIT fragment was ligated into the pcDNA3 backbone by T4 ligase (16 °C, overnight), and the construct was confirmed by sequencing (GeneWiz).

To construct the pU6-PyIT-H1-PyIT plasmid, the U6-PyIT-H1-PyIT cassette was ordered as a gBlock (200 ng, IDT, sequence in Table 2.4). The gBlock was dissolved in water (20 μ l), and digested with BglII (0.2 μ l) and MfeI (0.2 μ l) at 37 °C for 2 h. The pcDNA3 plasmid (8,000 ng, Invitrogen) was digested with BglII (2 μ l) and MfeI (2 μ l) at 37 °C for 2 h, and the backbone was extracted from a 0.8% agarose gel (run at 80 V for 45 min, expected band at ~5,400 bp). The U6-PyIT-H1-PyIT fragment was ligated into the pcDNA3 backbone by T4 ligase (16 °C, overnight), and the construct was confirmed by sequencing (GeneWiz), using the primer 5'-gggttattgtctcatgagcgg-3'.

To construct the pU6-PyIT-H1-PyIT-V2 plasmid, the second U6-PyIT-H1-PyIT cassette was amplified from the dissolved U6-PyIT-H1-PyIT DNA (10 ng/μl), with primers containing MfeI and MluI restriction sites (Fwd: 5'-aagcatcaattgcgttcactgaaaggctcgggcaggaagagg-3', Rev: 5'-gtcactacgcgtagatccaaaaaacggaaaccc-3'). The PCR product was then digested with MfeI (0.2 μl) and MluI (0.2 μl) restriction enzymes at 37 °C for 2 h. The pU6-PyIT-H1-PyIT plasmid (8,000 ng) was digested with MfeI (2 μl) and MluI (2 μl) at 37 °C for 2 h, and the backbone was extracted from a 0.8% agarose gel (run at 80 V for 45 min, expected band at ~5,900 bp). The U6-PyIT-H1-PyIT fragment was ligated into the pU6-PyIT-H1-PyIT backbone by T4 ligase (16 °C, overnight), and the construct was confirmed by sequencing (GeneWiz), using the primer 5'-gtcagccaggcgggccatttac-3'.

To test the efficiency of the two-plasmid system from the Chin lab, we replaced their reporter (sfGFP-150TAG) with our reporter (mCherry-TAG-EGFP). The mCherry-TAG-EGFP fragment was first amplified from the pWTRS-mCherry-TAG-EGFP construct (Fwd: 5'-atagctagcgcaccatggtgagcaagggcgaggaggataacatggccatcatc-3', Rev: 5'-acggtaccttatcattaagcgtaatctggaacatcg). The PCR product (20 μl) was digested with NheI (0.2 μl) and KpnI (0.2 μl) restriction enzymes at 37 °C for 2 h. The pU6-PyIT-V4-sfGFP-150TAG (8,000 ng, gift from Dr. Jason Chin)⁹⁷ was digested with NheI (2 μl) and KpnI (2 μl) at 37 °C for 2 h, and the pU6-PyIT-V4 backbone was purified by a 0.8% agarose gel (run at 80 V for 45 min, expected band at ~7,500 bp). The mCherry-TAG-EGFP fragment was then ligated into the pU6-PyIT-V4 backbone by T4 ligase (16 °C, overnight), and the construct was confirmed by sequencing (GeneWiz).

To generate our own expression system with the PyIT cassette on both plasmids, we cloned HCKRS (*M. barkeri*, with Y271A and L274M mutations) and mCherry-TAG-EGFP into the pU6-PyIT-H1-PyIT construct. HCKRS was amplified from the pHCKRS-mCherry-TAG-EGFP plasmid (Fwd: 5'-aattggatccatggactacaaggacgacgacg-3', Rev: 5'-gcatgaattctcatcacaggttggtgctgatgc-3').⁶⁰ mCherry-TAG-EGFP was amplified from the pmCherry-

TAG-EGFP plasmid (Fwd: 5'-agctagaattcatggtgagcaagggcgaggagg-3', Rev: 5'-atctggatccttatcattaagcgtaatctgg-3').^{19d} The PCR products (20 µl) were digested with EcoRI (0.2 µl) and BamHI (0.2 µl) at 37 °C for 2h. The pU6-PylT-H1-PylT (8,000 ng) was digested with EcoRI (2 µl) and BamHI (2 µl) at 37 °C for 2 h, and the backbone was extracted from a 0.8% agarose gel (run at 80 V for 45 min, expected band at ~ 5,900 bp). The HCKRS (or mCherry-TAG-EGFP) fragment was then ligated into the pU6-PylT-H1-PylT backbone by T4 ligase (16 °C, overnight), and both constructs were confirmed by sequencing (GeneWiz), using the GeneWiz universal CMV-Forward primers.

To test the expression system from Chin lab by western blot, the wild-type PylRS on pU6-PylT-V4-WTRS (gift from Dr. Jason Chin) was mutated to HCKRS (with Y271A and L274M mutations) by site-directed mutagenesis (Fwd: 5'-cctatgctggccccaccctggccaactacatgcggaaactggacagaatcc-3', Rev: 5'-ggattctgtccagttccgcatgtagttggccaggggtgggggcccagcatagg-3').

	Sequence (from 5' to 3')
H1-PyIT	ggtaggAGATCTaattcgaacgctgacgtcatcaacccgctccaaggaatcgcgggccagtgctactaggcggaaca ccagcgcgcgctgcccctggcaggaagatggctgtgagggacaggggagtgccgcccgtcaatattgcatgtcgctatgtg ttctgggaaatcaccataaacgtgaaatgtctttggatttgggaatcttataagttccctatcagtgatagagacaccggaacctg atcatgtagatcgaacggactctaaatccgttcagccgggtagattcccggggtttccgtttttGGATCCttatcg
U6- PyIT- H1-PyIT	ggtaggCAATTGaaggtcgggcaggaagagggcctatttcccatgattcctcatattgcatatacgatacaaggctgtag agagataattagaattaattgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttctgggtagtt tgcagttttaaaattatgttttaaatggactatcatatgcttaccgtaactgaaagtatttcgatttctggctttatatacttgtggaaa ggacgggggggtggatcgaatagatcacacggactctaaatcgtgcaggcggggtgaaactcccgtactccccgtttttggatc tgacaagtgcggttttgctagtcaattcgaacgctgacgtcatcaacccgctccaaggaatcgcgggccagtgctactaggcg ggaacacccagcgcgctgcccctggcaggaagatggctgtgagggacaggggagtgccgcccgtcaatattgcatgtc gctatgtgttctgggaaatcaccataaacgtgaaatgtctttggatttgggaatcttataagttctgtatgagaccacagctctggga acctgatcatgtagatcgaacggactctaaatccgttcagccgggtagattcccggggtttccgtttttggatctAGATCTtat cg

Table 2.4. Sequence of the PyIT expression cassette ordered from IDT. Restriction sites were capitalized. The *D. hafniense* PyIT was labeled in blue. The *M. barkeri* PyIT was labeled in red.

Incorporation efficiency test using a fluorescent reporter

Human embryonic kidney (HEK) 293T cells (ATCC, #CRL-11268) were grown in DMEM (Dulbecco's Modified Eagle Medium) supplemented with 10% FBS, 1% Pen-Strep, and glutamine (2 mM) in 96-well plates in a humidified atmosphere with 5% CO₂ at 37 °C. HEK 293T cells were transiently transfected with the designated plasmids (200 ng each) at ~75% confluency in the presence of UAA (1 mM). Branched polyethylene imine (bPEI, 1.5 µl, 1 mg/ml) was used as the transfection reagent for each well. After an overnight incubation at 37 °C, cells were imaged on a Zeiss Observer Z1 microscope, with EGFP channel (filter set 38 HE; ex. BP470/40; em.

BP525/50; exposure time: 300 ms) and mCherry channel (filter set 43 HE; ex. BP575/25; em. BP605/70; exposure time: 50 ms).

Western blot

HEK 293T cells were co-transfected with the designated plasmids (2,000 ng each) in the presence or absence of HCK (1 mM) in a 6-well plate. Branched polyethylene imine (bPEI, 15 μ l, 1 mg/ml) was used as the transfection reagent for each well. After 48 h of incubation at 37 °C, the cells were washed with chilled phosphate-buffer saline (PBS, 1 ml), and lysed in mammalian protein extraction buffer (250 μ l, GE Healthcare). The cell lysates were separated by 10% SDS-PAGE (run at 60 V for 15 min, and 150 V for 45 min) and were transferred to a PVDF membrane (GE Healthcare). The membrane was blocked in tris-buffer saline (TBS) with 0.1% Tween 20 and 5% milk for 1 h. The blots were probed with the primary antibody (1:1,000, anti-HA (sc-805) or anti-GAPDH (sc-25778), Santa Cruz) overnight at 4 °C, followed by incubation with the secondary goat anti-rabbit IgG-HRP antibody (1:20,000, sc-2004, Santa Cruz) for 1 h at room temperature. The blots were further incubated with the SuperSignal West Pico working solution (mixture of the Stable Peroxide Solution and the Luminol/Enhancer Solution, 500 μ l each, Thermo Scientific) for 5 min at room temperature. The luminescence signal was detected by ChemiDoc (Chemi Hi Sensitivity setting, exposure time: 10 sec).

2.3 Library Selection for Unnatural Amino Acid

2.3.1 Construction of a Pyrrolysyl-tRNA Synthetase Library and a Selection Marker

Although it was demonstrated through the targeted panel screening that several UAAs could be successfully incorporated with in-house mutant PylRSs, their incorporation efficiency may not be fully optimized, compared to that of an evolved aminoacyl tRNA synthetase (aaRS) for a specific UAA. Moreover, several UAAs were not effectively incorporated with any in-house mutant PylRS. We therefore sought to discover a PylRS/PylT pair for an UAA through library selection.⁹⁹ To utilize this method, a library of PylRS mutants must be constructed, and the PylRS that incorporates a desired UAA needs to be selected through a series of positive and negative genetic selection rounds.

Selection markers are key components in library selection. Several selection markers were previously used to evolve an aaRS/tRNA pair for UAA incorporation.⁹⁹ The chloramphenicol acetyltransferase (CAT) gene is the most frequently used marker in positive selections. A permissive site in the CAT gene is mutated to TAG, and only the mutant aaRS that can incorporate the unnatural and/or natural amino acid(s) will pass the selection. The original method of a subsequent negative selection employed a replica plate containing chloramphenicol but in the absence of an UAA, and colonies were selected that grow on the original plate but not the replica plate,⁸ however, this method is cumbersome and difficult to apply to large libraries. Later, use of the barnase gene as a negative selection marker was developed. Multiple TAG mutations are introduced in the barnase gene, and the mutant aaRS that incorporates natural amino acid(s) in the absence of the UAA, which leads to the expression of barnase and to cell death, will be eliminated. Several rounds of alternating positive and negative selections will result in one or more evolved mutant aaRSs that specifically incorporate the desired UAA. The combined use of CAT/barnase selection markers became the predominant selection strategy due to its clear

phenotype and ease of operation. GFP is an alternative selection marker, where a TAG mutation is introduced at its permissive position.¹⁰⁰ Using fluorescence-activated cell sorting (FACS), fluorescent cells will be collected in the positive selection, while non-fluorescent cells will be collected in the negative selection. This method is time-efficient, as it avoids the transferring of surviving mutant synthetase libraries between positive and negative selection strains. It also allows for easy control of selection pressure during each round by adjusting the fluorescence threshold. However, it is more expensive than antibiotic-based selection, and thus not always applicable to large-scale selections.

Another selection marker, CATUPP, was developed to facilitate both time- and cost-efficient selection (Fig 2.27).¹⁰¹ CATUPP is a fusion gene consisting of CAT and UPP. The UPP gene encodes uracil phosphoribosyltransferase, which converts 5-fluorouracil (5-FU) to 5-fluoro-dUMP. Since 5-fluoro-dUMP inhibits thymidylate synthetase, the cells expressing uracil phosphoribosyltransferase in the presence of 5-FU will die.¹⁰¹ CATUPP has been previously used as a dual positive/negative selection marker.¹⁰¹⁻¹⁰² A TAG mutation is introduced at a permissive position of the CATUPP fusion gene. In the positive selection, mutant aaRSs that incorporate unnatural and/or natural amino acid(s) will be selected in the presence of chloramphenicol. In the negative selection, mutant aaRSs that incorporate natural amino acid(s) will be eliminated in the presence of 5-FU. Several rounds of alternating positive and negative selections will result in one or more evolved mutant aaRS that specifically incorporates the desired UAA. This method does not require library isolation and retransformation between each selection, and is cost-effective.

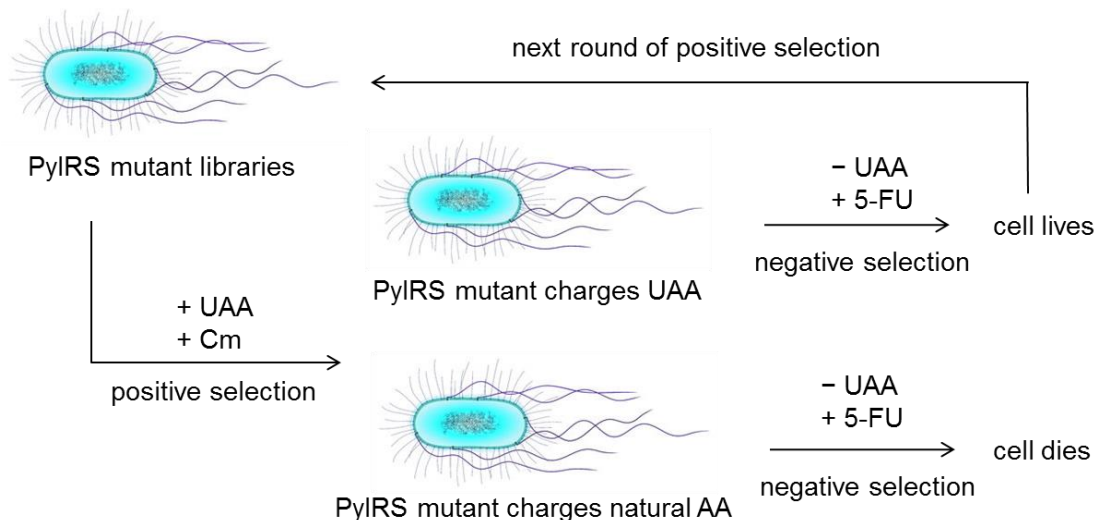


Figure 2.27. Workflow of library selection for PylRS, using the CATUPP selection marker. Positive selection (in the presence of chloramphenicol) selects mutant aaRSs that incorporate unnatural and/or natural amino acid(s). Negative selection (in the presence of 5-fluorouracil) removes mutant aaRSs that incorporate natural amino acids.

To this end, the CATUPP fusion gene was cloned into a pBAD backbone that also contains PyIT. A TAG mutation was introduced at the Q98 position of the CATUPP fusion gene as previously reported.¹⁰¹ The resulting plasmid, pBAD-CATUPP-Q98TAG, was used as the selection marker.

It should be noted that regular *E. coli* strains are sensitive to 5-FU, since the UPP gene is present and expressed in the *E. coli* genome.¹⁰³ Indeed, our phenotype test shows that TOP10 cells could not survive on an agar plate supplemented with 5-FU (50 µg/ml) (no images were recorded). To circumvent that effect, a UPP knockout *E. coli* strain (Δ UPP) was used.¹⁰¹ Though viable, Δ UPP cells still grow slower on an agar plate supplemented with 5-FU (50 µg/ml) compared to a plate without 5-FU. This suggests that Δ UPP cells are still sensitive to 5-FU, which will affect the phenotype in negative selections. In fact, several UAAs were subjected to negative selections, but cells grew slow regardless of the presence or absence of UAA. A subsequent phenotype test suggested that the negative selection did not have any effect on eliminating natural

amino acid incorporation. Realizing that CATUPP might not be an ideal negative selection marker, we decided to perform only positive selections. While the negative selection was always desired for the *Mj*TyrRS system,¹ we envisioned that the selection of PylRS mutants might succeed without it, because pyrrolysine is not present in bacterial cells, and PylRS shows high orthogonality toward natural amino acids.¹⁶

Another key component in library selection is the mutant library. To generate a mutant library, several residues in the binding pocket surrounding the pyrrolysine ring were selected based on the crystal structure of wild type PylRS with pyrrolysine,¹⁰⁴ and then randomized to all possible amino acids. To ensure the selection fidelity, it is important that the final mutant transformants cover the theoretical diversity of the library by at least 10-fold. For example, randomizing five residues would require a library of 3.3×10^7 mutants. However, the number of mutants is inherently limited by the *E. coli* transformation efficiency (10^{11} CFU/ μ g at most). To build a small library with good quality, we decided to start by randomizing only three residues in our library. To this end, two sub-libraries were constructed using inverse PCR (Fig 2.28).^{19e} Briefly, inverse PCR was performed using primers with NNK (N=A/C/G/T, K=G/T) codons at desired sites and BsaI restriction sites at the end.¹⁰⁵ If the desired sites were close to each other (with less than 6 bp in between), they were randomized at the same time. Otherwise, a second inverse PCR was performed using the first-round randomized library product (for more details, see primer design in the experimental). BsaI was used because it is a type IIS restriction endonuclease that produces 4 bp sticky ends away from its recognition site, thus not leaving any scar after cloning. The PCR product was treated with DpnI and BsaI, ligated, and transformed into electrocompetent Δ UPP cells. The resulting transformants were pooled and the mutant PylRS library DNA was extracted. The quality of the library was analyzed by the chromatogram of the sequencing reads (Sequence Scanner). A mixture of A/C/G/T peaks should be present at the randomized N position, and a mixture of G/T peaks should be present at the randomized K position. If necessary, the library could be used as a template for a next round of inverse PCR to introduce additional randomized

codons.

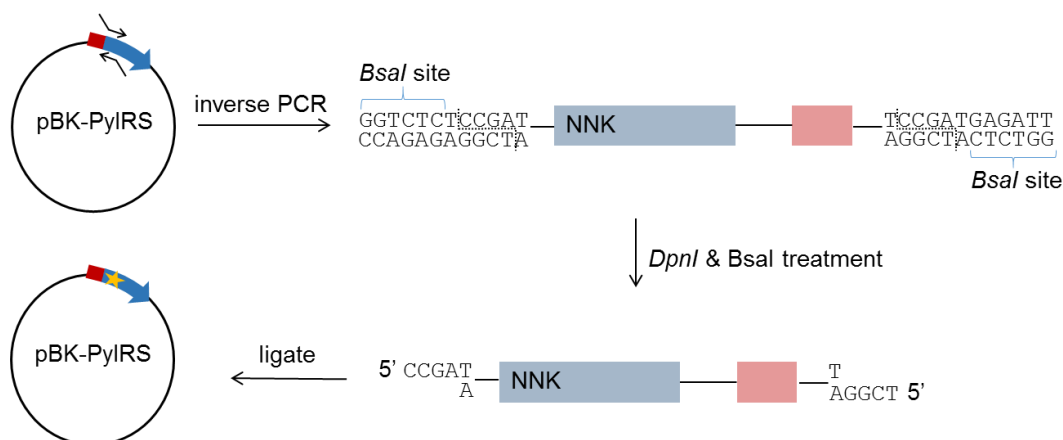


Figure 2.28. Generation of a mutant library. The original PylRS plasmid is amplified and linearized by inverse PCR, using primers containing NNK codons. The PCR product is treated with *DpnI* and *BsaI*, and ligated to a circular plasmid. The ligation product is transformed into *E. coli* cells, and the extracted DNA library is confirmed by sequencing. NNK (labeled with a yellow star) represents the degenerate codon (N=A/C/G/T, K=G/T). The red and blue boxes (with the randomized codon located at the N-terminus of the blue box) encode for the PylRS gene. Only one NNK codon is generated in this scheme, but additional randomized codons could be introduced in a similar way.

The first library was designed for the incorporation of lysine analogues (Fig 2.29A). A mutant PylRS (L274A/Y349F) was used as the starting gene. The L274A mutation is present in most PylRS mutants that are evolved for lysine analogues,¹⁶ and might enlarge the cavity for UAA binding. Y349F has been shown to increase the aminoacylation efficiency.^{18a} Three sites (A267, Y271, C313) were randomized to all possible amino acids. These three sites were picked because they were mutated in a previous library selection effort for the incorporation of photocaged lysine (PCK).^{19d} Two other sites (M241 and L274) were also randomized in the original study, but were not included in our design due to the small library size. In addition, mutations on Y271 and C313 were frequently observed in several other library selections.^{19c, 19e, 106} Transformants (8×10^4) were generated, covering the theoretical library size (8,000) by 10-fold. Sequencing of the extracted

DNA library confirmed the presence of the NNK codon at the expected positions (Fig 2.29B). However, some bias towards guanine was observed at all positions. We have identified the reason for the bias, and this had not been reported in previous library constructions.

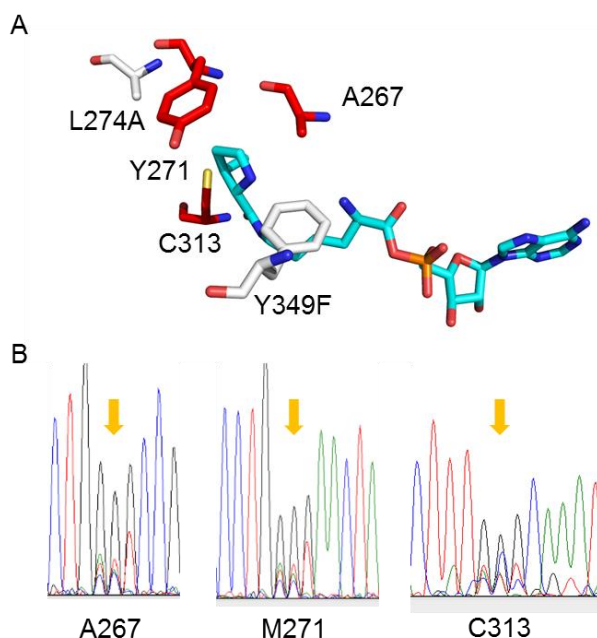


Figure 2.29. Generation of mutant library for lysine analogues. (A) Binding pocket of PylRS with selected residues. L274A and Y349F are fixed mutations (labeled in white). A267, Y271 and C313 are randomized to all possible amino acids (labeled in red). The randomized sites are in close proximity to the ligand, thus they are expected to modulate substrate specificity. (B) Sequencing result of the mutant library at positions that are randomized to NNK codon. NNK codon is indicated by a yellow arrow. NNK represents degenerate codon (N=A/C/G/T, K=G/T).

The second library was designed for the selection of phenylalanine analogues (Fig 2.30A). A mutant PylRS (N311A/C313A) was used as the starting gene. The double mutations N311A and C313A have been shown to enlarge the pocket size and facilitate the incorporation of aromatic rings.^{20a} The side chain of N311 also forms hydrogen bond with pyrrolysine, which is abolished with the N311A mutation.^{20a} As few libraries were constructed for the selection of phenylalanine

analogues at the time,^{22, 30} we chose to first randomize three sites (Y271, L274, Y349) to all possible amino acids. The Y271 and L274 sites were included in a previous selection for the photocaged tyrosine (ONBY), although Y271 was not mutated during the selection.²² The Y349 site has been included in several libraries, and mutations to either phenylalanine or tryptophan were frequently observed.^{19c, 107} Transformants (8×10^4) were generated, providing 10-fold coverage of the theoretical library size (8,000). Sequencing of the extracted DNA library showed the presence of NNK codons at expected positions (Fig 2.30B). But similar to the above library, some bias towards guanine were observed in this library at the Y271 and L274 sites. The two sub-libraries were then mixed to generate a “master” library (1:1 ratio, final concentration of 500 ng/μl), to select for both lysine and phenylalanine analogues.

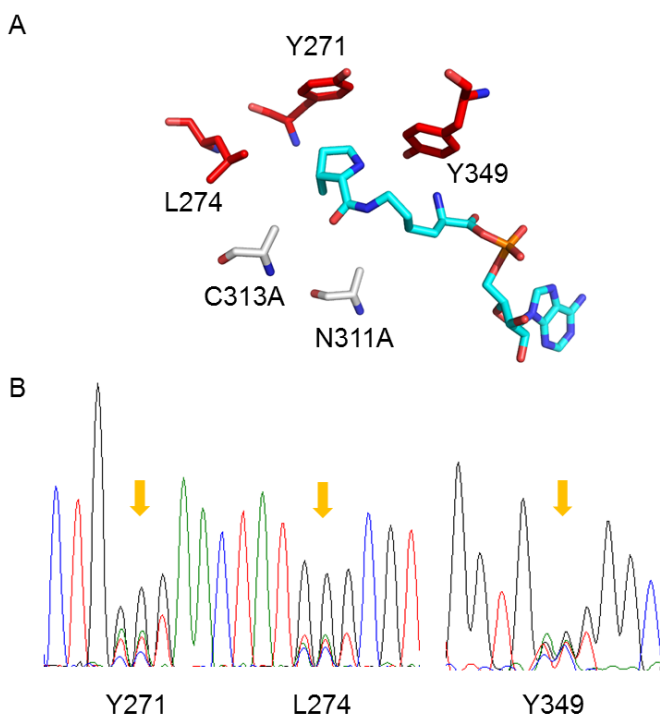


Figure 2.30. Generation of mutant library for tyrosine analogues. (A) Binding pocket of PylRS with selected residues. N311A and C313A are fixed mutations (labeled in white). Y271, L274 and Y349 are randomized to all possible amino acids (labeled in red). (B) Sequencing result of the mutant library at

positions that are randomized to NNK codon. NNK represents degenerate codon (N=A/C/G/T, K=G/T); indicated by a yellow arrow.

2.3.2 Selection of *o*-nitrobenzyl-O-tyrosine (ONBY) incorporation in *E. coli*

After the construction of the mutant libraries and selection plasmids were completed, the selection of a previously evolved UAA, *o*-nitrobenzyl-O-tyrosine (ONBY), was first performed as a proof-of-principle. The mutant library was transformed into Δ UPP electrocompetent cells that contained the CATUPP-Q98TAG selection plasmid. The colony forming units were higher than 10^6 , covering at least 10 times of the library size. Cells were plated on LB agar supplemented with chloramphenicol (60 μ g/ml) and ONBY (1 mM). Unexpectedly, 10 colonies grew on the selection plate, but 40 colonies grew on the control plate (containing 60 μ g/ml of chloramphenicol but no ONBY). Surviving colonies on the selection plate were grown in LB medium (3 ml) overnight, and 10 μ l of culture were plated on agar plates with/without ONBY. One out of ten mutants survived in the presence of ONBY, but died in the absence of ONBY, suggesting that it can successfully incorporate ONBY. Other mutants survived on both plates, suggesting that they incorporate natural amino acid(s). The PylRS was extracted from the hit mutant and named ONBYRS1. The selection was repeated and resulted in another hit mutant PylRS, named ONBYRS2. ONBYRS1 and ONBYRS2 were sequenced (Table 2.5). Both mutants harbor N311A and C313A mutations, suggesting that they were evolved from the sub-library for tyrosine analogues. A Y349W mutation has also been previously discovered for several other UAAs,^{19c, 19e} suggesting that it might also, in addition to Y349F, increase the aminoacylation efficiency.^{18a} Both synthetases shared three common mutations (Y271M, N311A, C313A) with EV16-5 (the synthetase that was originally identified in Dr. Ashton Cropp lab for the incorporation of ONBSY, and was later found to incorporate ONBY as well). The slight side-chain difference at the L274 position (Thr in ONBYRS1, Ser in ONBYRS2) could be responsible for their differences in incorporation efficiency, as well as the background incorporation for EV16-5 (Ala at this position). The EV20 synthetase harbors an

extra Y349F mutation compared to a synthetase (ONBYRS) previously identified for ONBY by Chin lab.²² EV20 was generated by us, and could incorporate ONBY, NPY, MNPY, and NPEY (details in Chapter 2.1). Although mutations on EV20 are very different than those on ONBYRS1/ONBYRS2, the presence of N311G and C313G mutations confirmed the hypothesis that small amino acids at the N311 and C313 sites facilitated the incorporation of phenylalanine analogues.

Code	Mutation sites					
	L270	Y271	L274	N311	C313	Y349
ONBYRS1	/	M	T	A	A	W
ONBYRS2	/	M	S	A	A	W
EV16-5	/	M	A	A	A	F
EV20	F	/	M	G	G	F

Table 2.5. Sequence of PylRS hits (ONBYRS1 and ONBYRS2) from library selection. Two PylRS that are known to incorporate ONBY (EV16-5 and EV20) are displayed for comparison.

The hit synthetases were further co-transformed with sfGFP-Y151TAG-pylT into TOP10 cells. The sfGFP expression in the presence of ONBY (1 mM) was performed to confirm the incorporation of ONBY by the hit synthetases. SDS-PAGE analysis suggests that ONBYRS1 shows highest incorporation efficiency, with a yield of 14 mg/L (Fig 2.31). It also performs better than either EV16-5 (some background incorporation without ONBY) or EV20 (less yield). ONBYRS2 shows less incorporation efficiency, with the yield of 8 mg/L (Fig 2.31), but little background incorporation is observed as well. In summary, these results suggest a functional selection system.

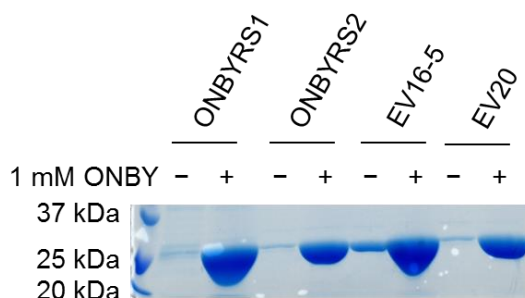


Figure 2.31. Incorporation of ONBY into sfGFP at the Tyr151 position. ONBYRS1 and ONBYRS2 were identified from library selection. EV16-5 and EV20 were used as positive controls. The sequences of these synthetases can be found in Table 2.5.

During the library construction, the parent plasmid might remain in the mutant library due to the incomplete DpnI treatment. This biased library will cause the survival of the parent plasmid that incorporates natural amino acid (phenylamine), since the N311A/C313A mutation was previously known to bear background incorporation.^{20a} To investigate the reason for the background colonies, four random mutant PylRS that survive on agar plate with chloramphenicol (200 µg/ml) but without ONBY were sequenced. Indeed, three out of four were the parent plasmid. Previously, for the selection of *Mj*TyrRS, randomized sites were normally mutated to alanine to create an inactive TyrRS.⁸ However, this step has not been performed for the selection of PylRS, due to its good orthogonality against the natural amino acids.

We therefore sought to decrease the background incorporation of the parent plasmid. The mutant PylS (N311G/C313G/Y349F) was used, as it bears no background incorporation in our previous test. Moreover, the N311G/C313G mutations were included in an evolved mutant that incorporates ONBY.²² It is expected to create similar space for an aromatic ring as the N311A/C313A mutation. Three other residues (L270, Y271, L274) that are near the pyrrolysine ring were randomized to all amino acids (Fig 2.32A). Transformants (8×10^4) were generated, covering the theoretical library size (8,000) by 10-fold. Sequencing of the extracted DNA library

confirmed NNK codon at expected positions, and no bias on guanine was observed (Fig 2.32B). The new library was used for the selection of ONBY. The surviving colonies after the positive selection (60 µg/ml of chloramphenicol) were incorporating natural amino acid(s). The sequencing of three random mutants showed a convergent mutation: L270R/Y271G/L274G/N311G/C313G/Y349F. However, the expression of sfGFP-Y151TAG with this mutant PyIRS did surprisingly not result in any fluorescence. We only run this expression once, and the result might not be accurate. In the future, experiment to analyze the protein expression on a SDS-PAGE will be further pursued.

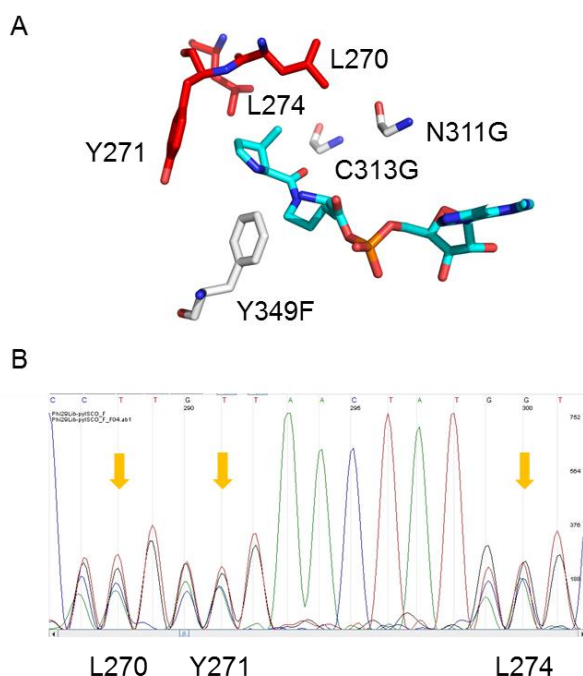


Figure 2.32. Second generation of mutant library for tyrosine analogues. (A) Binding pocket of PyIRS with selected residues. N311G, C313G and Y349F are fixed mutations (labeled in white). L270, Y271 and L274 are randomized to all possible amino acids (labeled in red). (B) Sequencing result of the mutant library at positions that are randomized to NNK codon. NNK represents a degenerate codon (N=A/C/G/T, K=G/T); the position is indicated by a yellow arrow.

Due to unsuccessful selection with different libraries, we sought to optimize the selection

plasmid. We first changed the TAG amber codon from position Q98 to position D111 or position D112. The Q98 position was originally applied in the one-plasmid selection system, since Q98 performed better than D181, another tested position.¹⁰¹ The D111 position was used by Dr. Ashton Cropp for library selection. The D112 position was previously used in CAT-based library selection.^{19c} A growth-based assay was performed to evaluate their performance. Briefly, the selection plasmid was co-transformed with wild-type PylRS into Δ UPP cells, and an overnight cell culture (10 μ l) was plated on LB agar supplemented with or without *N*^ε-allyloxycarbonyl-L-lysine (Alloc Lys), a known substrate for wild type PylRS. The plate was incubated at 37 °C for 16 h, and the area of bacterial growth was evaluated. Among three selection plasmids, CATUPP-D111TAG performs best (Table 2.6), and it was used in subsequent experiments.

	-UAA, -Cm	-UAA, +Cm (60 μ g/ml)	+Alloc Lys (1 mM), +Cm (60 μ g/ml)
CATUPP-WT	+++	+++	+++
CATUPP-Q98TAG	+++	–	+
CATUPP-D111TAG	+++	–	++
CATUPP-D112TAG	+++	–	+

Table 2.6. Evaluation of different selection plasmids. Wild type PylRS was co-transformed with selection plasmid, and the incorporation efficiency was tested by the area of bacterial growth (in the presence of Alloc Lys and Cm). The area was presented by the number of ‘+’.

We further explored if the selection plasmid can be used for negative selection, as previously described.¹⁰¹ The selection plasmid (CATUPP-D111TAG) and the synthetase (wild-type PylRS) were co-transformed into Δ UPP cells, and overnight culture (10 μ l) was plated on LB agar supplemented with different concentration of 5-FU (from 0.5 μ g/ml to 50 μ g/ml) and Alloc Lys (1 mM). The plate was incubated at 37 °C for 16 h, and the area of bacterial growth was evaluated. Toxicity of 5-FU (with different concentration) was even observed in Δ UPP cells (Table 2.7). We also can not differentiate the bacterial growth with or without supplementing Alloc Lys (Table 2.7). The result is not in agreement with previous finding that 100 μ g/ml is the optimal concentration

for selection.¹⁰¹ We therefore choose not to use CATUPP-D111TAG plasmid for negative selection.

5-FU concentration	0 µg/ml	0.5 µg/ml	5 µg/ml	10 µg/ml	50 µg/ml
-UAA	+++	++	+	+	-
+Alloc Lys (1 mM)	+++	++	+	+	-

Table 2.7. Evaluation of optimal 5-FU concentration for negative selection. Wild type PylRS was co-transformed with pCATUPP-D111TAG selection plasmid, and the toxicity was tested by the area of bacterial growth (in the presence of Alloc Lys and 5-FU). The area was presented by the number of '+'.

We then constructed pBarnase-Q2TAG-D44TAG-pylT as a negative selection plasmid. The barnase gene was amplified from pLWJ17B3 (negative selection plasmid for MjTyrRS, gift from Dr. Peter Schultz),⁸ and ligated into the pBAD backbone. The toxic barnase gene has two amber mutations at permissive sites (Gln2TAG, Asp44TAG). Cells expressing functional barnase in the absence of UAA indicates that the synthetase incorporates natural amino acid(s), and cell will not survive. The barnase gene has been used as a negative selection marker.⁸ We first tested the functionality of the pBarnase-Q2TAG-D44TAG-pylT plasmid using a known synthetase. The selection plasmid and the synthetase (EV20) were co-transformed into Top 10 cells, and overnight culture (10 µl) was plated on LB agar with or without ONBY (0.5 mM). The plate was incubated at 37 °C for 16 h, and the area of bacterial growth was evaluated. We observed that cells grew better when not supplying with ONBY (Table 2.8). The result suggests that functional barnase is expressed in the presence of UAA. We further used the barnase selection plasmid for negative selection. Electrocompetent cells harboring pBarnase-Q2TAG-D44TAG-pylT plasmid were transformed with the mutant library mixture. After incubation in LB plate (37 °C, 16 h), the PylRS from the surviving cells was extracted by gel electrophoresis, and was further transformed into electrocompetent cells harboring pBAD-CATUPP-Q98TAG-pylT plasmid (as we have not generated D111TAG selection plasmid at the time). However, no colonies survived in LB plate

supplemented with chloramphenicol (60 µg/ml) and ONBY (1 mM). Since CATUPP-Q98TAG is not the optimal positive selection plasmid, possible hits might be lost during the positive selection; in addition, previous genetic selections of tRNA synthetases required positive selections to be conducted first, followed by negative selections, and the experiment should be repeated accordingly. However, we speculated that the use of CATUPP-D111TAG might generate a better result.

	Barnase-TAG	GFP-TAG
-UAA	++	+++
+ONBY (0.5 mM)	+	+++

Table 2.8. Evaluation of barnase gene for negative selection. EV20 was co-transformed with the pBarnase-Q2TAG-D44TAG-pylT selection plasmid, and the toxicity was tested by the area of bacterial growth (in the presence of ONBY). The area was presented by the number of '+'. GFP-TAG was used as a control.

In summary, three PylRS mutant libraries (two tyrosine libraries and one lysine library) were generated through randomization of three select residues. Three CATUPP selection plasmids were generated and tested and the selection marker with the amber codon placed at D111 performed best in positive selection. However, it is not a good selective marker for negative selections due to the cellular toxicity of 5-FU, even in UPP knock-out strains. The established barnase gene may be a better negative selection marker. We successfully identified two PylRS mutants that incorporate ONBY, with one of them performing better than current known synthetases for ONBY incorporation. In the future, a suitable negative selection marker needs to be established and to rapidly identify PylRS hits, a EGFP fluorescent reporter can also be added to the positive selection plasmid.⁸ Once the library selection procedure is established, it will be applied to other UAAs (caged thiotyrosine, caged histidine, caged aminoalanine) where hit synthetases were not identified through our standard panel screening (discussed in Chapter 2.1

and Chapter 2.4).

Experimental

Construction of selection plasmid

The backbone of pBAD plasmid was generated by digesting pBAD-Myo4TAG-PyIT plasmid (8,000 ng, gift from Dr. Ashton Cropp) with NcoI and NdeI restriction enzymes, followed by gel extraction of the band corresponding to the correct length of the backbone (~5,500 bp). The CATUPP fragment was amplified from pCATUPP (gift from Dr. Ashton Cropp, Fwd: 5'-ccatggagaaaaaatcactgg-3', Rev: 5'-catatggtcgattatttcgtacc-3'), digested with NcoI and NdeI restriction enzymes, and ligated into the pBAD backbone (16 °C, overnight). The TAG mutation was introduced at position Q98 by site-directed mutagenesis (Fwd: 5'-gttacaccgtttccatgagtagactgaaacgtttcatcg-3', Rev: 5'-cgatgaaaacgtttcagtcactcatggaaaacggtgtaac-3', TAG mutation underlined). The same method was used to generate D111TAG (Fwd: 5'-gctctggagtgaaataccactaggattccggcagtttctac-3', Rev: 5'-gtagaaactgccggaatcctagtggtattcactccagagc-3', TAG mutation underlined) and D112TAG (Fwd: 5'-ctggagtgaaataccacgactagttccggcagtttctacac-3', Rev: 5'-gtgtagaaactgccggaactagtcgtggtattcactccag-3', TAG mutation underlined) mutations.

To generate the pBarnase-Q2TAG-D44TAG-pyIT plasmid, the Barnase-Q2TAG-D44TAG fragment was amplified from pLWJ17B3 (negative selection plasmid for *Mj*TyrRS, gift from Dr. Peter Schultz, Fwd: 5'-aattccatggcataggttatcaacacg-3', Rev: 5'-aattcatatgttatctgattttgtaaaggtctg-3'), digested with NcoI and NdeI restriction enzymes, and ligated into the pBAD backbone (16 °C, overnight).

Growth-based assay for evaluating selection plasmid

For evaluation of positive selection marker, the selection plasmid (pCATUPP-pyIT with TAG amber codon at Q98, D111, or D112 position) was co-transformed with wild type PylRS (50 ng for each

plasmid) into Δ UPP cells. A single colony was picked and grown overnight in LB medium (3 ml). Overnight cell culture (10 μ l) was plated on two LB agar plates supplemented with arabinose (0.2%), Cm (60 μ g/ml), Tet (25 μ g/ml), Kan (50 μ g/ml), with one plate also supplemented with Alloc Lys (1 mM in water, ChemImpex, Cas# 6298-03-9). The plates were incubated at 37 °C for 16 h, and the area of bacterial growth (from where the cell culture was spotted) was evaluated by naked eyes.

For evaluation of CATUPP as a negative selection marker, pCATUPP-D111TAG-pylT and wild type PylRS were co-transformed (50 ng for each) into Δ UPP cells. The overnight culture (10 μ l) was plated on LB agar supplemented with arabinose (0.2%), Tet (25 μ g/ml), Kan (50 μ g/ml), Alloc Lys (1 mM in water), and different concentration of 5-FU (Cas #: 51-21-8, 0.5 μ g/ml, 5 μ g/ml, 10 μ g/ml, or 50 μ g/ml). The plate was incubated at 37 °C for 16 h, and the area of bacterial growth was evaluated. For evaluation of barnase as a negative selection marker, pBarnase-Q2TAG-D44TAG-pylT and EV20 were co-transformed (50 ng for each) into Top 10 cells. The overnight culture (10 μ l) was plate on two LB agar plates supplemented with arabinose (0.2%), Tet (25 μ g/ml), Kan (50 μ g/ml), with one also supplemented with ONBY (0.5 mM). The plate was incubated at 37 °C for 16 h, and the area of bacterial growth was evaluated.

Library construction

To construct the “lysine library”, EV17 (*M. barkeri* PylRS with mutations Y271M, L274A and Y349F) was used as the starting template. Position C313 was first randomized by amplifying the plasmid using an inverse PCR (Fwd: 5'-gcgccaggaaggtctcaaaccttNNKcaaatgggcagcggctgcacccgtgaaaac-3', Rev: 5'-gcgccagagtaggtctcaagttaaccatggtgaattctccaggtgttcttg-3', randomized NNK is capitalized), with the NNK codon incorporated in the forward primer. The inverse PCR was performed as previously described.^{19e} The PCR reaction (50 μ l) was prepared as following: DNA template (50 ng), dNTP (200 μ M), primers (0.8 μ M for each), Phusion DNA polymerase (1 U, NEB). The PCR was run as

following: 95 °C for 2 min, 9 cycles of amplification (95 °C for 20 sec, 65 °C for 20 sec, with the annealing temperature decreased by 1 °C for each cycle, 68 °C for 4 min), 31 cycles of amplification (95 °C for 20 sec, 58 °C for 20 sec, 68 °C for 4 min), 68 °C for 9 min. The PCR product was confirmed by analysis on a 0.8% agarose gel (~ 2,000 bp), purified by QIAquick PCR Purification Kit (Qiagen), and eluted with 50 µl of water (normally with the yield of ~ 30 ng/µl). Purified product was digested by DpnI restriction enzyme (1 µl of 20,000 U/ml, NEB) for 2 h at 37 °C. The DpnI enzyme was heat inactivated (75 °C, 30 min), and the product was further digested with BsaI (1 µl of 10,000 U/ml, NEB) for 2 h at 37 °C. The BsaI enzyme was heat inactivated (75 °C, 30 min), and the product was purified by QIAquick PCR Purification Kit (Qiagen). The purified product (1,000 ng) was self-ligated with T4 ligase (5 µl of 400,000 U/ml, NEB) in a total of 100 µl reaction (16 °C for 12 h). The ligation product was purified with QIAquick PCR Purification Kit (Qiagen), and eluted with 30 µl of water. Purified ligation product (2 µl) was added to the Top 10 electrocompetent cells (50 µl). The cells were placed in a chilled 2 mm electroporation cuvette (VWR), and underwent a pulse of 2.5 KV in an electroporator (GenePulser Xcell, BioRad). Super Optimal broth with Catabolite repression (SOC) medium (1 ml) was immediately added to the cells. The cells were then incubated in 37 °C shaker for 1.5 h, and all the cells (1 ml) were then plated on one LB agar plate (150 × 15 mm, VWR) containing Kan (50 µg/ml). The plate was incubated at 37 °C for 16 h. The next day, cells were scraped with 1 ml of LB medium, and were grown in 100 ml of LB medium with Kan (50 µg/ml) overnight. Plasmid DNA was then extracted (Midi kit, Qiagen) and sent for sequencing. The plasmid randomized at C313 position was further used as the template for next round of randomization at A267 and Y271 positions. The plasmid was amplified using an inverse PCR (Fwd: 5'-gcgccagggtctcaatgctgNNKccgaccctgNNKaactatgcgcgtaaactggatcgatttc-3', Rev: 5'-gcgccagagtaggtctcagcatcggacgcaggcacagggttttatccacgcggaatttg-3', randomized NNK is capitalized), with the NNK codon incorporated in the forward primer. The PCR product was digested with DpnI, BsaI, self-ligated, and transformed to Top10 electrocompetent cells as

described above. The extracted mutant library was further sent for sequencing.

For the first generation of tyrosine library, EV16 (*M. barkeri* PylRS with mutations N311A and C313A) was used as the starting template. Position Y349 was first randomized to all natural amino acids. The plasmid was amplified using an inverse PCR (Fwd: 5'-ggaaaggtctcgtgcatggtgNNKggcgataccctggatattatg-3', Rev: 5'-cagtaggtctctgcagctatcgcccacaatttcgaagtcg-3', randomized NNK is capitalized), with the NNK codon incorporated in the forward primer. The PCR product was digested with DpnI, BsaI, self-ligated, and transformed to Top10 electrocompetent cells as described above. The extracted mutant panel was then sent for sequencing. The plasmid randomized at Y349 position was further used as the template for next round of randomization at Y271 and L274 positions. The plasmid was amplified using an inverse PCR (Fwd: 5'-gcgcaggtctcaaccctgNNKaactatNNKcgtaaactggatcgattctgccgggtc-3', Rev: 5'-gcgcagagtaggtctcagggtcggggccagcatcgagcgcaggcacaggtttttatc-3', randomized NNK is capitalized), with the NNK codon incorporated in the forward primer. The PCR product was digested with DpnI, BsaI, self-ligated, and transformed to Top10 electrocompetent cells as described above. The extracted mutant library was further sent for sequencing.

For the second generation of tyrosine library, EV19-1 (*M. barkeri* PylRS with mutations N311G, C313G and Y349F) was used as the starting template. Three positions (L270, Y271, L274) were randomized to all natural amino acids in one step. The plasmid was amplified using an inverse PCR (Fwd: 5'-gcgcaggtctcaccgatgctggccccgaccNNKNNKaactatNNKcgtaaactggatcgattctgccgggtc-3', Rev: 5'-gcgcagagtaggtctcatcgagcgcaggcacaggtttttatccacgcggaaaatttg-3', randomized NNK is capitalized), with the NNK codon incorporated in the forward primer. The PCR product was digested with DpnI, BsaI, self-ligated, and transformed to Top10 electrocompetent cells as described above. The extracted mutant library was then sent for sequencing.

Generation of electrocompetent cells for selection

The UPP knockout (Δ UPP) strain was a gift from Dr. Ashton Cropp, and electrocompetent Δ UPP cells containing the selection plasmid were made. The Δ UPP cells were chemically transformed with pBAD-CATUPP-Q98TAG plasmid (10 ng). A single colony was inoculated in LB medium (3 ml) with Tet (25 μ g/ml) and grown overnight (37 °C, 250 rpm). The next day, saturated overnight culture (2 ml) were added to a large LB medium (200 ml) without antibiotics. The cell culture was grown at 37 °C, 250 rpm until OD₆₀₀ reached 0.4 (normally takes around 3 h). The culture was chilled on ice for 20 min, and the cells were pelleted (4,000 rpm, 15 min, 4 °C). The cells were carefully resuspended in chilled water (50 ml) and pelleted again (4,000 rpm, 15 min, 4 °C). This process was repeated once. The cells were then carefully resuspended in chilled water containing 10% glycerol, and pelleted (4,000 rpm, 15 min, 4 °C). The cells were carefully resuspended in 500 μ l of 10% glycerol (in water), and were aliquoted into prechilled 1.7 ml eppendorf tube (50 μ l). Each tube was fast frozen in an isopropyl alcohol/dry ice bath, and was stored in a -80 °C freezer.

To test the competency of the cells, 1 μ l of pEGFP-N1 (Clontech, 1 ng/ μ l) was added to the electrocompetent cells harboring the pBAD-CATUPP-Q98TAG plasmid (50 μ l). The cells were placed in a chilled 2 mm electroporation cuvette (VWR), and underwent a pulse of 2.5 KV in an electroporator (GenePulser Xcell, BioRad). Super Optimal broth with Catabolite repression (SOC) medium (1 ml) was immediately added to the cells. The cells were incubated in 37 °C shaker for 1.5 h. The 100x diluted culture was made (1 μ l of culture and 99 μ l of water), and was plated on LB agar containing Kan (50 μ g/ml) and Tet (25 μ g/ml). After 16 h of incubation at 37 °C, the number of colonies was counted. The competency (CFU/ μ g) was calculated as: number of colonies $\times 10^6$. The competency was normally between 10^7 to 10^8 CFU/ μ g (similar to chemically competent cells).

Library selection

The mutant library (1 μ l of 500 ng/ μ l) was transformed into the Δ UPP electrocompetent cells that contained the CATUPP-Q98TAG selection plasmid, using the electroporation protocol described above. The transformants (1 ml) were grown in LB medium (100 ml) supplemented with Tet (25 μ g/ml) and Kan (50 μ g/ml) overnight (37 °C, 250 rpm). For making library stocks, 50 ml of the cells were centrifuged down (4,000 rpm, 10 min) the next day, and resuspended in 5 ml of 25% glycerol (in water). The concentrated cell stocks were aliquoted (1 ml in each tube) and were stored in -80 °C freezer. In the future, the library can be recovered by growing the concentrated cell stock overnight. For library selection, diluted overnight culture (300 μ l of OD₆₀₀ = 1.0, total of 8×10^8 cells) was plated on LB agar plate supplemented with Tet (25 μ g/ml), Kan (50 μ g/ml), Cm (60 μ g/ml), arabinose (0.2% arabinose) and ONBY (1 mM). A plate without ONBY was used as a control. After incubation at 37 °C for 16 h, the surviving colonies from the selection plate were grown in LB medium (3 ml, containing Tet (25 μ g/ml), Kan (50 μ g/ml)) overnight. The saturated cell culture (10 μ l) was plated on two agar plates, except one was supplemented with ONBY (1 mM) and one without, and grown overnight (37 °C). The hit colonies were the ones that grew on ONBY plate, but did not grow on control plate. The hit colonies were grown in LB medium (3 ml, containing Tet (25 μ g/ml), Kan (50 μ g/ml)) overnight. The plasmids (pPylRS and pCATUPP-Q98TAG-pylT) were extracted from the cell culture (miniprep kit, Qiagen), and were separated on a 0.8% agarose gel (run 80 V for 45 min). The pPylRS plasmid was gel-extracted (~ 2,000 bp on the gel) and sent for sequencing.

For the negative selection, the mutant library (1 μ l of 500 ng/ μ l) was transformed into the Δ UPP electrocompetent cells that contained the pBarnase-Q2TAG-D44TAG-pylT selection plasmid, using the electroporation protocol described above. After incubation in LB plate (37 °C, 16 h), the surviving cells were collected with LB medium (1 ml), and the plasmids (pPylRS and pBarnase-Q2TAG-D44TAG-pylT) were extracted (miniprep kit, Qiagen) and separated on a 0.8% agarose gel (run 80 V for 45 min). The mixed pPylRS mutant plasmids were gel-extracted (~

2,000 bp on the gel), and were further transformed into electrocompetent cells harboring pBAD-CATUPP-Q98TAG plasmid. The transformants (1 ml) were grown in LB medium (100 ml) supplemented with Tet (25 µg/ml) and Kan (50 µg/ml) overnight (37 °C, 250 rpm). Diluted overnight culture (300 µl of OD₆₀₀ = 1.0, total of 8×10^8 cells) was plated on LB agar plate supplemented with Tet (25 µg/ml), Kan (50 µg/ml), Cm (60 µg/ml), arabinose (0.2% arabinose) and ONBY (1 mM). A same plate without ONBY was used as a control.

Validation of the hit synthetase

To validate the functionality of the hit synthetase, we co-transformed the hit synthetase with sfGFP-Y151TAG-pyIT (50 ng for each plasmid) into Top10 cells. A single colony was grown overnight (37 °C, 250 rpm) in LB supplemented with Kan (50 µg/ml) and Tet (25 µg/ml), and the overnight culture (250 µl) was added to LB (25 ml) supplemented with ONBY (1 mM), Tet (25 µg/ml), and Kan (50 µg/ml). Cells were grown at 37 °C, 250 rpm, and the protein expression was induced with arabinose (0.1 %) when OD₆₀₀ reached 0.4 (measured by Nanodrop). After overnight expression at 37 °C, cells were pelleted (5,000 g, 10 min) and resuspended in phosphate lysis buffer (pH 8.0, 50 mM of phosphate, 6 mL). Triton X-100 (60 µl, 10%) was added to the mixture. The lysate was incubated on ice for 1 h, sonicated (power level 5, pulse 'on' for 30 sec, pulse 'off' for 30 sec, with a total of 4 min, 550 sonic dismembrator), and then centrifuged (13,000 g) at 4 °C for 10 min. The supernatant was transferred to a 15 ml conical tube, and Ni-NTA resin (Qiagen, 100 µl) was added. The mixture was incubated at 4 °C for 2 h with mild shaking. The resin was then collected by centrifugation (1,000 g, 10 min) and washed with lysis buffer (300 µl), this was repeated twice, followed by two washes with wash buffer (300 µl) containing imidazole (20 mM). The protein was eluted with elution buffer (300 µl) containing imidazole (250 mM). The purified proteins were analyzed by SDS-PAGE (10%), and stained with Coomassie Brilliant Blue.

2.4 Incorporation of Caged Cysteine Using an *E. coli* Leu Synthetase

Cysteine is a less abundant, yet important amino acid for structure and function of proteins.¹⁰⁸ The unique chemistry of its thiol group accounts for broad range of functional roles.¹⁰⁹ Cysteine is involved in multiple cellular events, including protein ubiquitylation,¹¹⁰ balance of redox potential,¹¹¹ cellular signaling pathways,¹¹² as well as protein trafficking and localization.¹¹³ Cysteine is also a key residue for several post-translational modifications, including alkylation, oxidation, and nitrosylation.¹¹⁴ In addition, two cysteine residues form a disulfide bond, resulting in intra- or intermolecular linkages on proteins. The substantial change in protein folding further affects protein structure and function.¹¹⁵

Given the importance of cysteine residues, we envisioned that genetic code expansion of a caged cysteine will greatly expand the toolbox for spatiotemporal control of protein function with light. Previously, we successfully incorporated a caged cysteine and a caged homocysteine through the PylRS/PylT system (discussed in detail in chapter 3.1),¹¹⁶ where a mutant PylRS (with M241F, A267S, Y271C, and L274M mutations) that incorporated a photocaged lysine (PCK) was used. However, the incorporation efficiency of the caged cysteine was only half of that of the caged lysine.¹¹⁶ The low incorporation efficiency hindered the expression of light-activatable intein and TEV protease (discussed in detail in chapter 3.1). Another caged cysteine (NPMC) was incorporated through the PylRS/PylT system by Jason Chin's lab, which was further applied for light-activation of the TEV protease in mammalian cells.²³ However, the incorporation efficiency of their caged cysteine (at the concentration of 1 mM) was lower than that of our caged cysteine (tested by Subhas Samanta in our lab). Therefore, a highly efficient system for the incorporation of caged cysteine is needed.

Previously, a caged cysteine, 4,5-dimethoxy-2-nitrobenzyl-cysteine (NVC), was genetically encoded in mammalian cells using the orthogonal *E. coli* leucyl-tRNA synthetase

(*EcLeuRS*)/tRNA^{Leu} pair.¹¹⁷ The *EcLeuRS* mutant (M40G, L41Q, Y499L, Y527G, H537F) used in their study was originally evolved for the incorporation of a caged serine, 4,5-dimethoxy-2-nitrobenzyl-serine (NVS), in *Saccharomyces cerevisiae*.¹¹⁸ Due to the structural similarity between serine and cysteine, the same *EcLeuRS* mutant incorporates NVC as well. Huiwang Ai's lab later discovered that a caged cysteine with an extra methyl group (NVMC), which is less toxic to cells as a result of the less reactive ketone byproduct, could also be incorporated by the same *EcLeuRS* mutant.¹¹⁹ They further applied this system for the control of intein splicing with light (discussed in detail in chapter 3.1). In their system, four copies of tRNA^{Leu} were expressed under four separate H1 promoters. The *EcLeuRS* mutant was expressed under a CMV promoter. The expression cassettes for tRNA^{Leu} and *EcLeuRS* were built in a single plasmid (pMAH). To identify the most efficient caged cysteine for UAA mutagenesis, we tested their expression system for the incorporation of several caged cysteines (Fig 2.33). Besides NVMC and NVC, we also tested the caged cysteine reported by Chin lab (NPMC),²³ and two other caged cysteines with *ortho*-nitrobenzyl group (NBMC and NBC). Previously, NBC was incorporated by another *EcLeuRS* mutant (with M40W, L41S, Y499I, Y527A, H537G mutations; not used/tested in this chapter), which was identified through library selection.¹²⁰ However, generating this *EcLeuRS* mutant will require either multiple rounds of site-directed mutagenesis or ordering of the full-length *EcLeuRS* with desired mutations. We therefore decided to just test the *EcLeuRS* mutant from Huiwang Ai's lab.

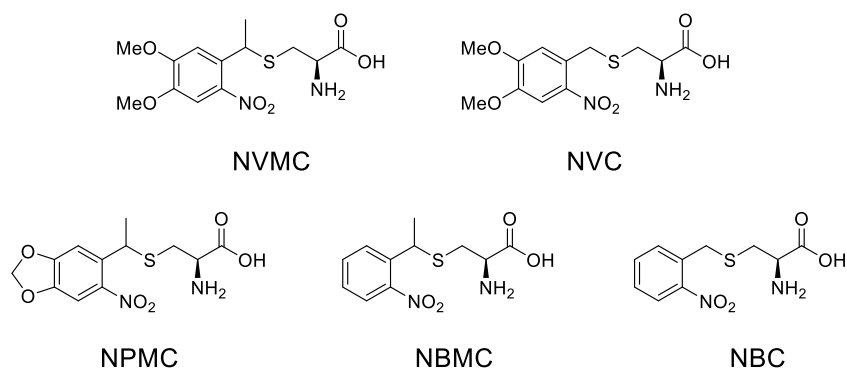


Figure 2.33. Structures of caged cysteines studied in this chapter.

We used mCherry-TAG-EGFP-HA as a reporter for incorporation efficiency, where a TAG codon was placed between mCherry and EGFP. The EGFP fluorescence indicated incorporation efficiency, and the mCherry fluorescence served as a control for transfection efficiency. HEK 293T cells were co-transfected with pMAH (encoding the *EcLeuRS* mutant (M40G, L41Q, Y499L, Y527G, H537F) and the tRNA^{Leu})¹¹⁹ and pmCherry-TAG-EGFP-HA (200 ng each), and the culture medium was supplemented with the designated UAA (at the concentration of 1 mM). After 48 h, cells were imaged for EGFP and mCherry fluorescence. We found that NVMC is toxic to the cells, even at the concentration of 0.5 mM. The chemical might be re-purified and re-tested. Among other caged cysteines, NVC presented highest incorporation efficiency (Fig 2.34), which is also higher than that of our caged cysteine using the PylRS/PylT system.¹¹⁶ NBC presented second-highest incorporation efficiency, while NPMC and NBMC were only moderately incorporated (Fig 2.34). This result suggested NVC as the optimal caged cysteine for genetic code expansion in mammalian cells.

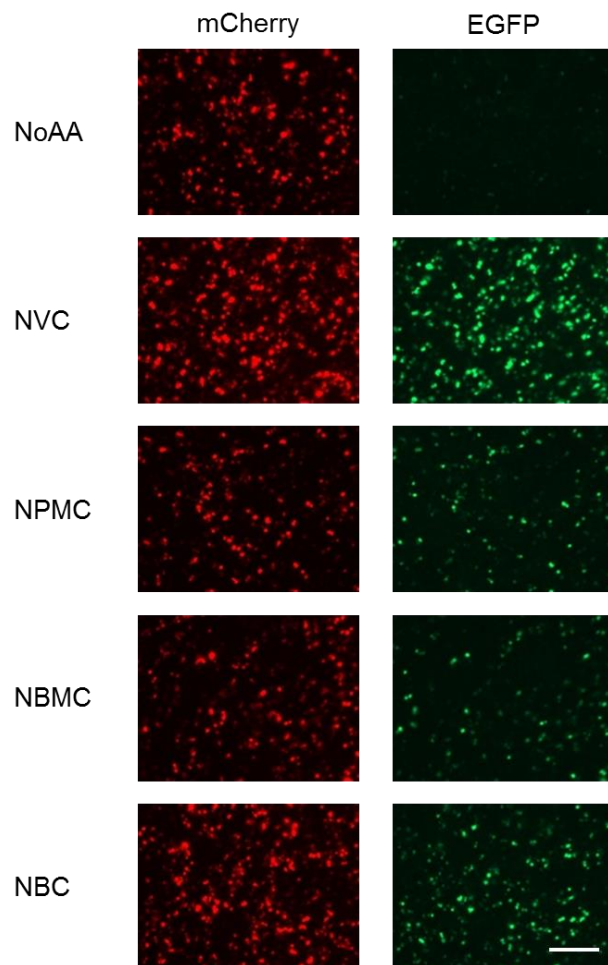


Figure 2.34. Incorporation of different caged cysteines by *EcLeuRS* mutant (M40G, L41Q, Y499L, Y527G, H537F). HEK 293T cells were transfected with the reporter plasmid (pmCherry-TAG-EGFP-HA) and the *EcLeuRS*-tRNA^{Leu} plasmid (200 ng each). The culture medium was supplemented with the designated caged cysteine (1 mM). Fluorescent images (mCherry and EGFP) were taken at 48 h. Scale bar indicates 200 μ m. Incorporation efficiency: NVC > NBC > NPMC = NBMC. NVMC was toxic to cells.

We further asked if the same *EcLeuRS* mutant incorporates caged serines that have similar structure as the caged cysteine. Previously, a caged serine, 4,5-dimethoxy-2-nitrobenzyl-serine (NVS), was genetically incorporated in *Saccharomyces cerevisiae* through *E. coli* leucyl-tRNA synthetase (*EcLeuRS*)/tRNA^{Leu} pair.¹¹⁸ However, applications of caged serine in mammalian cells have not been reported yet. We therefore prepared a few caged serines (NVS,

MNPOCS, MNPOCAMS) and tested their incorporation by the same *EcLeuRS* mutant (Fig 2.35). MNPOCAMS was designed based on a relay caging/decaging strategy as a previous caged cysteine.¹¹⁶

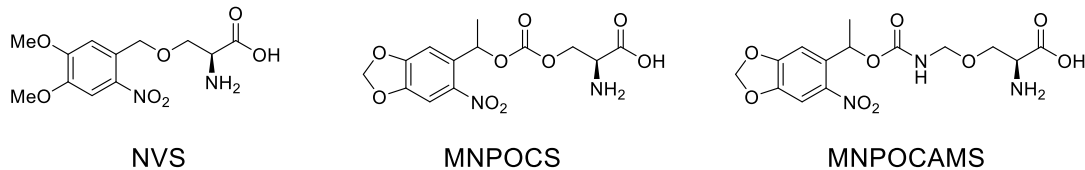


Figure 2.35. Structures of caged serines studied in this chapter.

We used mCherry-TAG-EGFP-HA as the reporter for incorporation, and performed fluorescent imaging experiments as described above. We observed similar incorporation efficiency among these three caged serines (Fig 2.36). The result is not surprising, as the *EcLeuRS* mutant (M40G, L41Q, Y499L, Y527G, H537F) was originally evolved for the incorporation of NVS in *Saccharomyces cerevisiae*.¹¹⁸ However, the incorporation efficiency is lower than that of NVC.

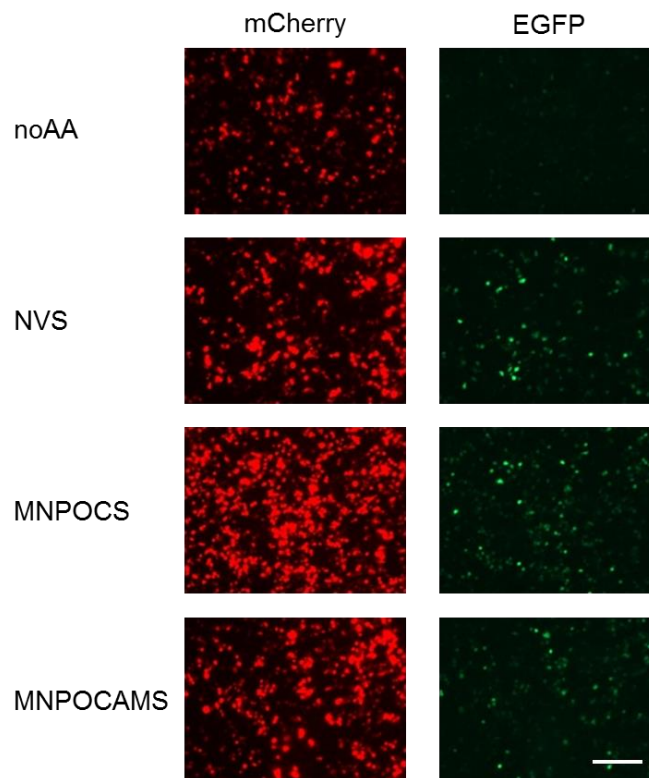


Figure 2.36. Incorporation of different caged serines by the *EcLeuRS* mutant (M40G, L41Q, Y499L, Y527G, H537F). HEK 293T cells were transfected with the reporter plasmid (pmCherry-TAG-EGFP-HA) and the *EcLeuRS*-tRNA plasmid (200 ng each). The culture medium was supplemented with the designated caged serine (1 mM). Scale bar indicates 200 μ m. Fluorescent images (for mCherry and EGFP) were taken at 48 h. Incorporation efficiency was similar among different caged serines.

To further evaluate the incorporation of caged cysteine and caged serine into mCherry-TAG-EGFP-HA in mammalian cells, we performed western blot analysis (Fig 2.37). For caged cysteines, we observed the incorporation for NVC, but not for NVMC and NPMC (lane 2 - 4). For caged serines, we only observed a faint incorporation for xxc, but not for NVS and MNPOCS (lane 5 - 7). Since our fluorescence-based test suggested incorporation of all these UAAs, further effort is needed to confirm the incorporation of other caged cysteines and caged serines by western blot.

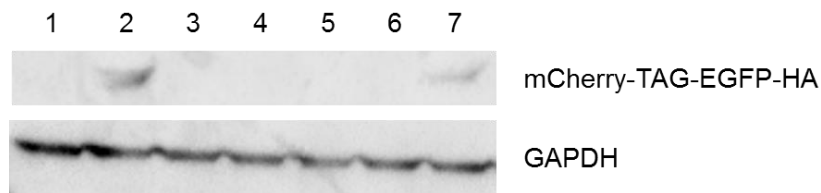


Figure 2.37. Incorporation of caged cysteines and caged serines by the *EcLeuRS* mutant (M40G, L41Q, Y499L, Y527G, H537F), analyzed by western blot. HEK 293T cells were transfected with the pmCherry-TAG-EGFP-HA plasmid and the *EcLeuRS*-tRNA plasmid (1,000 ng each) in a 6-well plate. The culture medium was supplemented with the designated UAA (1 mM). After 48 h, cell lysate was collected for western blot (with HA and GAPDH antibodies). 1, NoAA. 2, NVC. 3, NVMC. 4, NPMC. 5, NVS. 6, MNPOCS. 7, MNPOCAMS.

Since the *EcLeuRS* mutant (M40G, L41Q, Y499L, Y527G, H537F) showed broad substrate specificity, we further explored whether other UAAs with the similar structure could be incorporated. We first tested a caged aminoalanine (PCAA1), which generates an aminoalanine group after photo-decaging. Once installed at the cysteine position of a protease, the aminoalanine could be used to covalently trap the substrates of the protease allowing for their identification by mass spectrometry. The caged aminoalanine was successfully incorporated by this *EcLeuRS* mutant (Fig 2.38B). However, we later found that this compound cannot be efficiently decaged with the UV light (365 nm) in test tube (tested by Subhas Samanta in our lab). We therefore prepared another caged aminoalanine (PCAA2), which undergoes efficient decaging with the 365 nm UV light (tested by Subhas Samanta). However, this compound cannot be incorporated by this *EcLeuRS* mutant (Fig 2.38B).

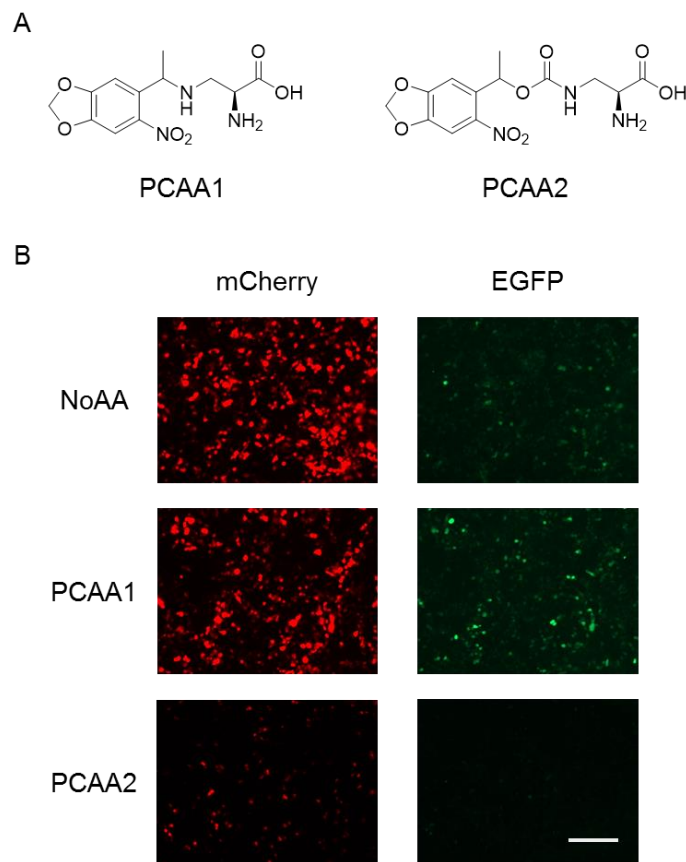


Figure 2.38. Incorporation of caged aminoalanines in mammalian cells by the *EcLeuRS* mutant (M40G, L41Q, Y499L, Y527G, H537F). (A) Structures of caged aminoalanines studied in this chapter. (B) Incorporation of PCAA1 and PCAA2 in mammalian cells. HEK 293T cells were transfected with the reporter plasmid (pmCherry-TAG-EGFP) and the *EcLeuRS*-tRNA plasmid (200 ng each). The culture medium was supplemented with PCAA1 or PCAA2 (1 mM). Fluorescent images (for mCherry and EGFP) were taken at 48 h. Scale bar indicates 200 μ m.

In summary, we incorporated four caged cysteines (with different efficiencies) and three caged serines (albeit with low efficiency) into proteins, using the *E. coli* leucyl-tRNA synthetase (*EcLeuRS*)/tRNA^{Leu} pair. In the future, we will use the most efficient caged cysteine (NVC) to generate light-activatable proteins where cysteine plays a critical role at the active site (for example, a phosphatase). At the same time, NPC will be synthesized and tested for incorporation

efficiency. We have attempted optical control of Src kinase through the incorporation of a caged cysteine by PylRS (details in Chapter 3.1), and the use of NVC could be re-tested here. We will also test light activation of proteins through the genetic incorporation of caged serine, using the *EcLeuRS*/tRNA^{Leu} pair. To further improve the incorporation efficiency, a one-plasmid system comprising all genetic components for UAA mutagenesis might be generated for the *EcLeuRS*/tRNA^{Leu} pair (similar to the PylRS/PylT system discussed in chapter 2.2). Besides, several other *EcLeuRS* mutants have been reported for the genetic incorporation of structurally diverse UAAs,¹²¹ A small panel of *EcLeuRS* mutants could be generated for rapid screening of new UAAs in the future. The test of a different *EcLeuRS* mutant evolved for NBC might also present different substrate specificities.¹²⁰ Due to their similarity to PCK, the amino acids MNPOCS and MNPOCAMS should also be tested for incorporation by PCKRS.

Experimental

Genetic incorporation of UAAs in mammalian cells

Human embryonic kidney (HEK) 293T cells (ATCC, #CRL-11268) were grown in DMEM (Dulbecco's Modified Eagle Medium) supplemented with 10% FBS, 1% Pen-Strep, and glutamine (2 mM) in 96-well plates in a humidified atmosphere with 5% CO₂ at 37 °C. The single plasmid expressing the *EcLeuRS* mutant (M40G, L41Q, Y499L, Y527G, H537F) and the tRNA^{Leu}, coined pMAH, was a gift from Dr. Huiwang Ai.¹¹⁹ HEK 293T cells were transiently transfected with pmCherry-TAG-EGFP-HA (200 ng) and pMAH (200 ng) at ~75% confluency in the presence of the designated UAAs (1 mM). Branched polyethylene imine (bPEI, 1.5 µl, 1 mg/ml) was used as the transfection reagent for each well. After an overnight incubation at 37 °C, cells were imaged on a Zeiss Observer Z1 microscope, with EGFP channel (filter set 38 HE; ex. BP470/40; em. BP525/50; exposure time: 300 ms) and mCherry channel (filter set 43 HE; ex. BP575/25; em. BP605/70; exposure time: 50 ms).

Western blot

HEK 293T cells were co-transfected with pmCherry-TAG-EGFP-HA (2,000 ng) and pMAH (2,000 ng) in the presence or absence of the designated UAAs (1 mM) in a 6-well plate. Branched polyethylene imine (bPEI, 15 μ l, 1 mg/ml) was used as the transfection reagent for each well. After 48 h of incubation, the cells were washed with chilled phosphate-buffer saline (PBS, 1 ml), and lysed in mammalian protein extraction buffer (250 μ l, GE Healthcare). The cell lysates were separated by 10% SDS-PAGE (run at 60V for 15 min, and 150 V for 45 min), and were transferred to a PVDF membrane (GE Healthcare). The membrane was blocked in tris-buffer saline (TBS) with 0.1% Tween 20 and 5% milk for 1 h. The blots were probed with the primary antibody (1:1,000, anti-HA (sc-805) or anti-GAPDH (sc-25778), Santa Cruz) overnight at 4 °C, followed by incubation with the secondary goat anti-rabbit IgG-HRP antibody (1:20,000, sc-2004, Santa Cruz) for 1 h at room temperature. The blots were further incubated with the SuperSignal West Pico working solution (mixture of the Stable Peroxide Solution and the Luminol/Enhancer Solution, 500 μ l each, Thermo Scientific) for 5 min at room temperature. The luminescence signal was detected by ChemiDoc (Chemi Hi Sensitivity setting, exposure time: 10 sec).

3.0 Conditional Control of Protein Activity

3.1 Incorporation of Caged Cysteine into Intein, TEV Protease and Src Kinase

Given the importance of cysteine residue in proteins (discussed in Chapter 2.4), we are interested in genetic incorporation of a caged cysteine residue for spatiotemporal control of protein function with light. To this end, a photocaged cysteine (PCC) and a photocaged homocysteine (PCHC) were synthesized by Rajendra Uprety in our lab (Fig 3.1). The caging group is removed upon UV irradiation and the resultant intermediate spontaneously hydrolyzes to its corresponding natural amino acid. Since PCC and homocysteine share structural similarity with photocaged lysine (PCK, Fig 3.1) for which we previously evolved a pyrrolysyl-tRNA synthetase, we first tested the incorporation of PCC and PCHC with this PylRS mutant (PCKRS, with M241F, A267S, Y271C, L274M mutations). Both unnatural amino acids could be incorporated into the Y151TAG position of sfGFP (Fig 3.1).¹¹⁶ The incorporation efficiency of PCHC (yield of 1.46 mg/L) is similar to that of PCK (yield of 1.36 mg/L), while PCC's is slightly lower (0.96 mg/L). This is not surprising, as PCC is shortened by one methylene carbon compared to PCK or PCHC. The incorporation of PCC and PCHC was further validated by ESI-MS of the purified caged-sfGFP, along with ESI-MS data for the decaged protein after light irradiation (performed by Ji Luo in our lab).

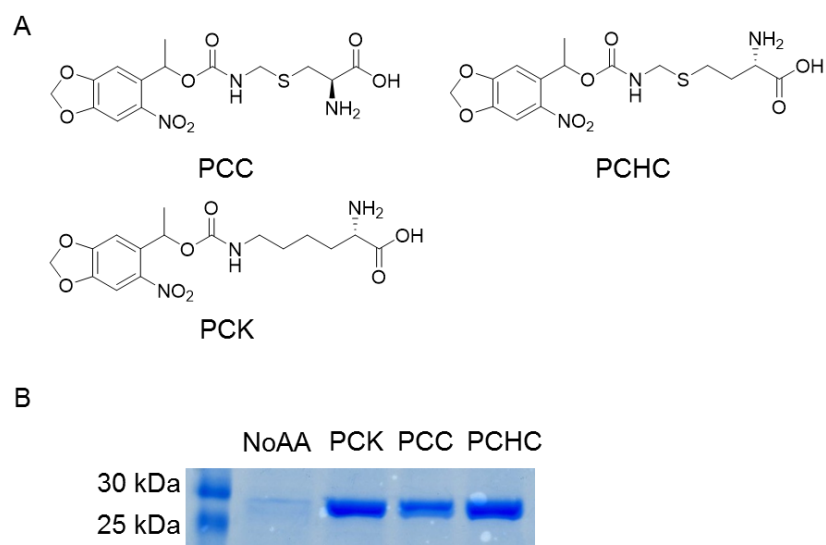


Figure 3.1. Genetic incorporation of PCC and PCHC into proteins. (A) Structure of PCC, PCHC, and PCK (as a positive control for incorporation). (B) SDS-PAGE gel demonstrating incorporation of PCC and PCHC into sfGFP at the Tyr 151 position, with yields of 0.96 mg/L and 1.46 mg/L, respectively. PCK was incorporated as a positive control, with a yield of 1.36 mg/L. A UAA concentration of 2 mM was used for expression. The PCKRS synthetase contains M241F, A267S, Y271C, L274M mutations.

After establishing successful genetic encoding of PCC and homocysteine in bacteria, we shifted our focus to targeting a critical cysteine residue on a protein in mammalian cells. We chose to engineer an intein as our first target. An intein is an internal protein domain that excises itself and catalyzes the ligation of its flanking protein sequences with a peptide bond.¹²² This posttranslational process, known as protein splicing, is a natural phenomenon found in all three kingdoms of life.^{122b} The protein sequences flanking the intein are called N-exteins and C-exteins. Inteins start with a serine or a cysteine, and the first amino acid of C-extein is an invariant serine, threonine or cysteine. Aside from full length inteins (also called *cis*-splicing inteins), split inteins (*trans*-splicing inteins) have also been described in several organisms.^{122b} In split inteins, the N- and C-terminal splicing domains are expressed as two fragments along with their respective extein sequences. They then self-associate and catalyze protein splicing. Protein splicing is a

rapid process comprising four steps: N-X acyl shift, *trans*-(thio)esterification, Asn cyclization, and X-N acyl shift (Fig 3.2).^{122b} Although how exactly an intein benefits its host evolutionarily remains elusive, inteins have been widely applied in the biotechnology field, finding applications in protein purification,^{122a} protein labeling,^{122b} and protein activity manipulation.^{122b}

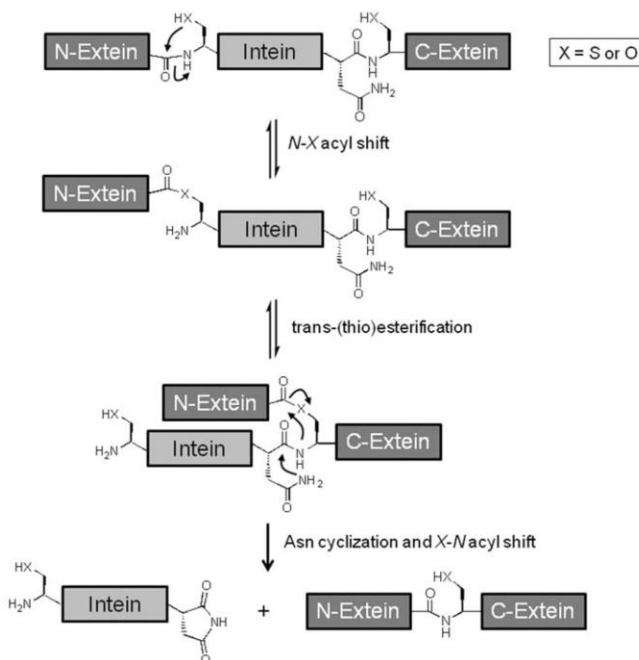


Figure 3.2. Mechanism of intein-mediated protein splicing. The process consists of four steps: N-X acyl shift, *trans*-(thio)esterification, Asn cyclization, and X-N acyl shift. Adapted from *Cell. Mol. Life Sci.* **2012**, 1-22.

The intein-based control of protein activity in living cells is of special interest to us. Small molecules and light are two major ways to control intein activity. In one example of small molecule control, rapamycin-induced FKBP and FRB dimerization was used to control the activity of a split intein.¹²³ However, such inteins have some *trans*-splicing activity even in the absence of rapamycin. To avoid this drawback, a ligand-binding domain was inserted into the *cis*-splicing intein and subjected to directed evolution until ligand binding successfully triggered the intein's

self-splicing.¹²⁴ A light-triggered intein was also reported by incorporating a bulky photocleavable group into the C-terminal half of the split intein.¹²⁵ This method, however, requires peptide synthesis with a protecting group and thus limits the length of the C-extein portion.

Inteins are excellent target proteins for PCC (Fig 3.3). Since inteins contain a conserved cysteine at its N-terminal end which directly participates in the protein splicing process, we speculate that the incorporation of a caged cysteine will block splicing activity, leaving an unspliced and inactive extein. Upon UV irradiation, the caging group is removed, restoring splicing activity. As the intein is self-spliced, an active extein is formed as its complete structure is reconstituted. Using this approach, any protein could be judiciously bisected and expressed as separate N- and a C-exteins, allowing for photochemical control of any protein of interest. Since the *cis*-splicing intein could be directly used for expression in live organisms—with no need for peptide synthesis—this method is expected to overcome issues such as low splicing efficiency and limited target protein length.

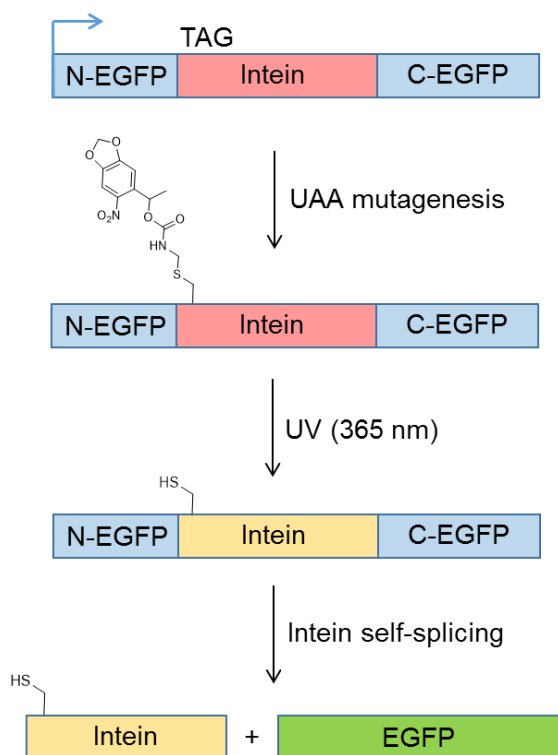


Figure 3.3. Photochemical control of protein function by incorporation of a caged cysteine into an intein. The caged cysteine blocks splicing activity, leaving an unspliced and inactive extein. UV irradiation removes the caging group, restoring splicing activity. As the intein is self-spliced, an active extein is formed. EGFP is depicted here as a model protein.

EGFP was first used as the model protein for a proof-of-principle experiment. The RecA intein was inserted in place of residues 107-109 on EGFP as previously reported.^{124a} The cDNA of EGFP(N)-Intein-EGFP(C) was then ligated into the pAG31 backbone containing the gene for PCKRS (the mammalian PyIRS mutant that incorporates PCC). The final construct, pAG31-Intein, drives the expression of EGFP(N)-Intein-EGFP(C) and PCKRS under two separate CMV promoters. pAG31-Intein was transfected into human embryonic kidney (HEK 293T) cells, and fluorescence was observed after 24 hours. However, compared to the EGFP fluorescence observed from cells transfected with the positive control plasmid pEGFP-N1, little fluorescence from spliced EGFP was observed (Fig 3.4). Thus, we began several rounds of troubleshooting.

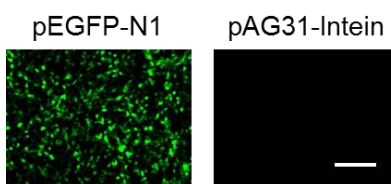


Figure 3.4. EGFP expression from the pAG31-Intein plasmid. HEK 293T cells were transfected with the pAG31-Intein, and cells were imaged at 24 h for EGFP fluorescence, with 10x objective. Scale bar indicates 200 μ m. The pEGFP-N1 plasmid was used as a control.

Since the intein we used was linked to a ligand-binding domain,^{124a} the cloning resulted in a deletion of 14 amino acids compared with the intein reference sequence. We were concerned that this deletion might affect the intein splicing activity. Therefore, other intein variants that have equivalent activities to the wild-type RecA intein, pAG31- Δ I and pAG31- Δ Ihh, were generated.¹²⁶ Δ I is a mini-intein based on the wild-type intein with the non-functional central endonuclease domain removed. Δ Ihh is based on Δ I, with the residual endonuclease loop further replaced with a seven-residue β -turn from the Hedgehog protein. Once again, however, little fluorescence from the spliced EGFP was observed using these two constructs when compared to EGFP fluorescence from the positive control plasmid pEGFP-N1.(Fig 3.5).

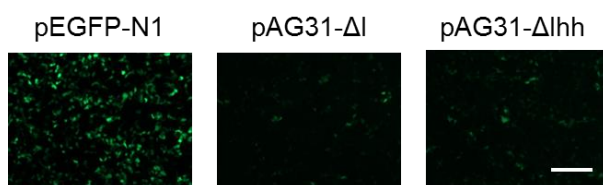


Figure 3.5. EGFP expression from the pAG31- Δ I and pAG31- Δ Ihh plasmids. HEK 293T cells were transfected with the pAG31- Δ I or pAG31- Δ Ihh plasmid (200 ng), and cells were imaged at 24 h for EGFP fluorescence, with 10x objective. Scale bar indicates 200 μ m. The pEGFP-N1 plasmid was used as a control.

Since the pAG31 construct expresses the intein construct and the PylRS gene both under the control of two separate CMV promoters, we were concerned that two CMV promoters, when presented on one plasmid, might lead to decreased gene expression. To avoid this issue, ΔI was cloned into the pEGFP-N1 plasmid. The resulting construct, p ΔI , was used for the fluorescent imaging experiments described above. This time, higher fluorescence from spliced EGFP expression was observed, although it still did not achieve the same fluorescence level as that observed using the pEGFP-N1 construct (Fig 3.6).

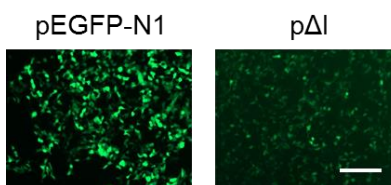


Figure 3.6. EGFP expression from the p ΔI plasmid. HEK 293T cells were transfected with the p ΔI , and cells were imaged at 24 h for EGFP fluorescence, with 10x objective. Scale bar indicates 200 μ m. The pEGFP-N1 plasmid was used as a control.

We further confirmed the identity of the intein splicing product by western blot. The p ΔI -WT plasmid (2,000 ng), encoding EGFP(N)-Intein(ΔI)-EGFP(C)-HA, was transfected into HEK 293T cells in a 6-well plate. Cell lysate was probed for the HA tag. Both unspliced (EGFP(N)-Intein-EGFP(C), 46 kDa) and spliced (EGFP, 27 kDa) products were observed, and the splicing efficiency was $\sim 70\%$ (Fig 3.7). We thus performed the incorporation and light activation experiments based on the p ΔI plasmid.

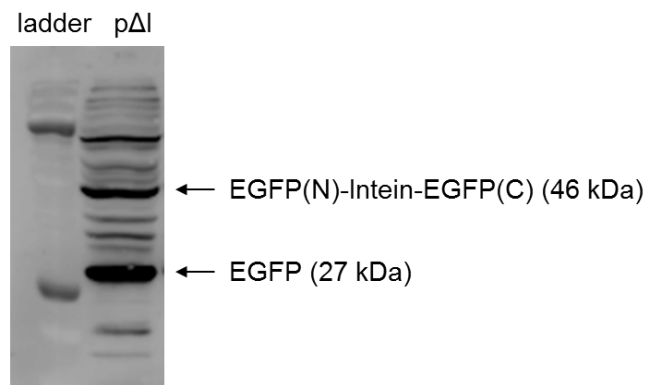


Figure 3.7. Western blot analysis of the intein splicing product. pΔI (encoding EGFP(N)-Intein-EGFP(C)-HA, 2,000 ng) was transfected to HEK 293T cells in a 6-well plate. Cells were lysed, and the lysate was probed for the HA tag. Arrows indicate unspliced (EGFP(N)-Intein-EGFP(C), 46 kDa) and spliced (EGFP, 27 kDa) products.

We then tested the incorporation of PCC into EGFP(N)-Intein-EGFP(C) (on pΔI plasmid) at position 109 by generating the C109TAG mutation. PCHC and PCK were also tested at the same time. HEK 293T cells were transfected with pΔI-C109TAG and pPCKRS-U6-PylT plasmids (1,000 ng each) in a 6-well plate. The culture medium was supplemented with the designated UAA (1 mM). After 48 h, cells were irradiated with UV light (365 nm, VWR) for 1 min, and were incubated at 37 °C for another hour. Cell lysate was then collected for western blot analysis (with anti-HA and anti-GAPDH antibodies). However, no incorporation of PCC into the intein was observed (Fig 3.8, lane 4 and 5). Although some faint incorporation of PCHC and PCK was observed (Fig 3.8, lane 6 to 9), no splicing products were present in UV-treated groups. We used pΔI-C109A as a negative control, since the C109A mutation leads to an inactive intein. As expected, no splicing product was detected (Fig 3.8, lane 2). The positive control, pΔI, worked as described above (Fig 3.8, lane 1).

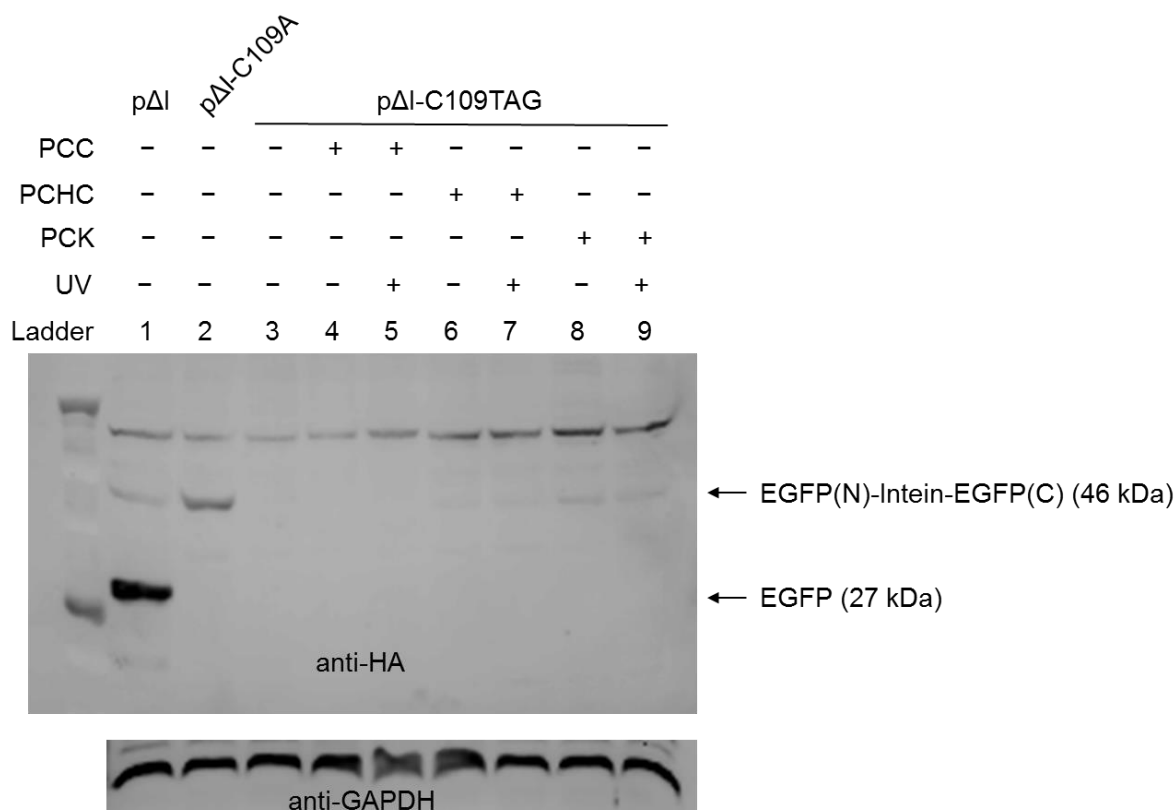


Figure 3.8. Incorporation of PCC and PCHC into an intein, followed by light-induced splicing. pΔI and pΔI-C109A were used as the positive and negative control, respectively. Cells were irradiated in 6-well plate for 1 min, then were incubated for 1 hour at 37 °C. Cell lysate was then collected for western blot analysis (with anti-HA and anti-GAPDH antibodies). Arrows indicate unspliced (EGFP(N)-Intein-EGFP(C), 46 kDa) and spliced (EGFP, 27 kDa) products.

Additionally, we performed a live cell imaging experiment. HEK 293T cells were transfected with pΔI-C109TAG and pPCKRS-U6-PyIT plasmids (200 ng each) in a 96-well plate. The culture medium was supplemented with designated UAA (1 mM). After 48 h, cells were irradiated with UV light (365 nm, VWR) for 1 min, and were incubated at 37 °C for another hour. Cells were then imaged for EGFP fluorescence (Zeiss). However, no fluorescence was observed for any UAAs (Fig 3.9).

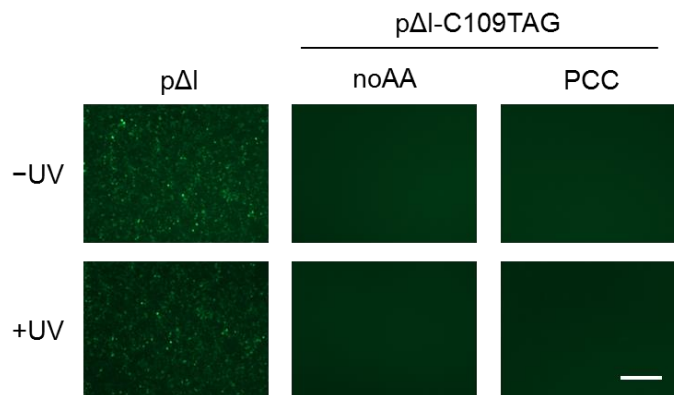


Figure 3.9. Light-activation of intein function with live cell imaging. HEK 293T cells were transfected with pΔI-C109TAG and pPCKRS-U6-PyIT plasmids (200 ng each) in a 96-well plate. The culture medium was supplemented with PCC (1 mM). After 48 h, cells were irradiated with UV light (365 nm, VWR) for 1 min, and were incubated at 37 °C for another hour. Cells were then imaged for EGFP fluorescence, with 5x objective. Scale bar indicates 100 μm. The pΔI plasmid was used as a control.

Based on the western blot results, we speculated that the reason for unsuccessful EGFP activation is the low expression level of PCC-containing protein. We later developed a one-plasmid system for better incorporation efficiency (discussed in chapter 2.2); this system was designed to boost the expression level of mutant protein. As we worked on this project, Huiwang Ai's lab reported the light activation of protein splicing with a photocaged intein.¹¹⁹ They used the *E. coli* leucyl-tRNA synthetase (*EcLeuRS*)/tRNA pair for genetic incorporation of a caged cysteine in mammalian cells, achieving higher incorporation efficiency than our current PyIRS/PyIT pair (discussed in chapter 2.4). Through incorporation of a caged cysteine into the active cysteine position of a DnaE intein from *Nostoc punctiforme* (*Npu*), they achieved light-activation of a red fluorescent protein (RFP) and a human Src tyrosine kinase.¹¹⁹

We then decided to target another protein, Tobacco Etch Virus (TEV) protease, for the incorporation of PCC. TEV protease is a cysteine protease that recognizes a specific peptide sequence, Glu-Asn-Leu-Tyr-Phe-Gln-(Gly/Ser), and cleaves between the Gln and Gly/Ser

residues. Due to its high sequence specificity and cleavage activity, TEV protease is one of the most frequently used enzymatic reagents to remove affinity tags after protein purification.¹²⁷ The peptide bond cleavage is achieved through its catalytic triad, consisting of His46, Asp81, and Cys151. The catalytic cysteine is responsible for attacking the carbonyl carbon of the scissile bond through its nucleophilic thiol.

We therefore sought to generate a caged TEV protease through the incorporation of a caged cysteine. We hypothesized that the caging group will block proteolytic activity. Upon UV irradiation, the caging group will be removed, restoring the function of TEV protease. A plasmid that expressed TEV protease, pTEVP (from Kalyn Brown in our lab), was used, and a C152TAG mutation was further introduced on pTEVP. Another two plasmids were also used: pPCKRS-U6-PyIT, which expresses both PCKRS and PyIT, and pGloSensor (Promega). The GloSensor plasmid expresses a circularly permuted firefly luciferase with a protease recognition site. Cleavage of the recognition site by TEV protease dramatically increases luciferase activity (Fig 3.10).

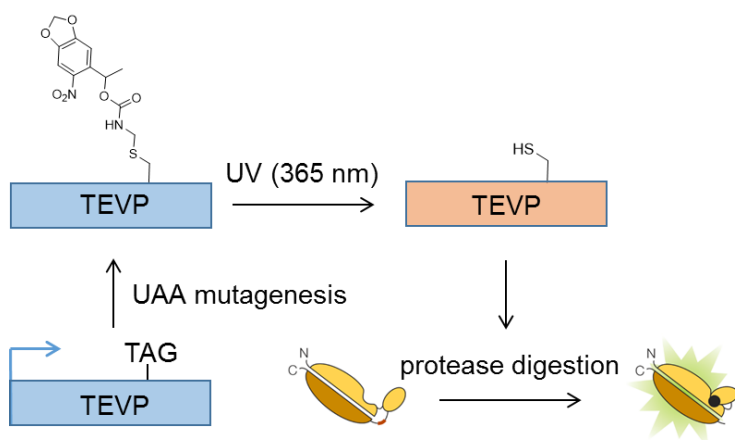


Figure 3.10. Light activation of TEV protease activity by incorporation of a caged cysteine. The caging group blocks proteolytic activity. UV irradiation removes the caging group, restoring TEV protease activity, which is detected using GloSensor.

pTEVP-C152TAG, pPCKRS-U6-PyIT and pGloSensor (200 ng each) were co-transfected into HEK 293T cells. The culture medium was supplemented with the designated UAA (2 mM). After 24 h, the designated wells were irradiated with UV light (365 nm, VWR) for 2 minutes. After another 6 hours, luciferase activity was detected using the BrightGlo reagent (Promega). The TEVP-WT, as a positive control, was active (Fig 3.11, lane 2 and 3). However, for both PCC (lane 4 and 5) and PCHC (lane 6 and 7), no increase of luciferase activity was observed after UV irradiation.

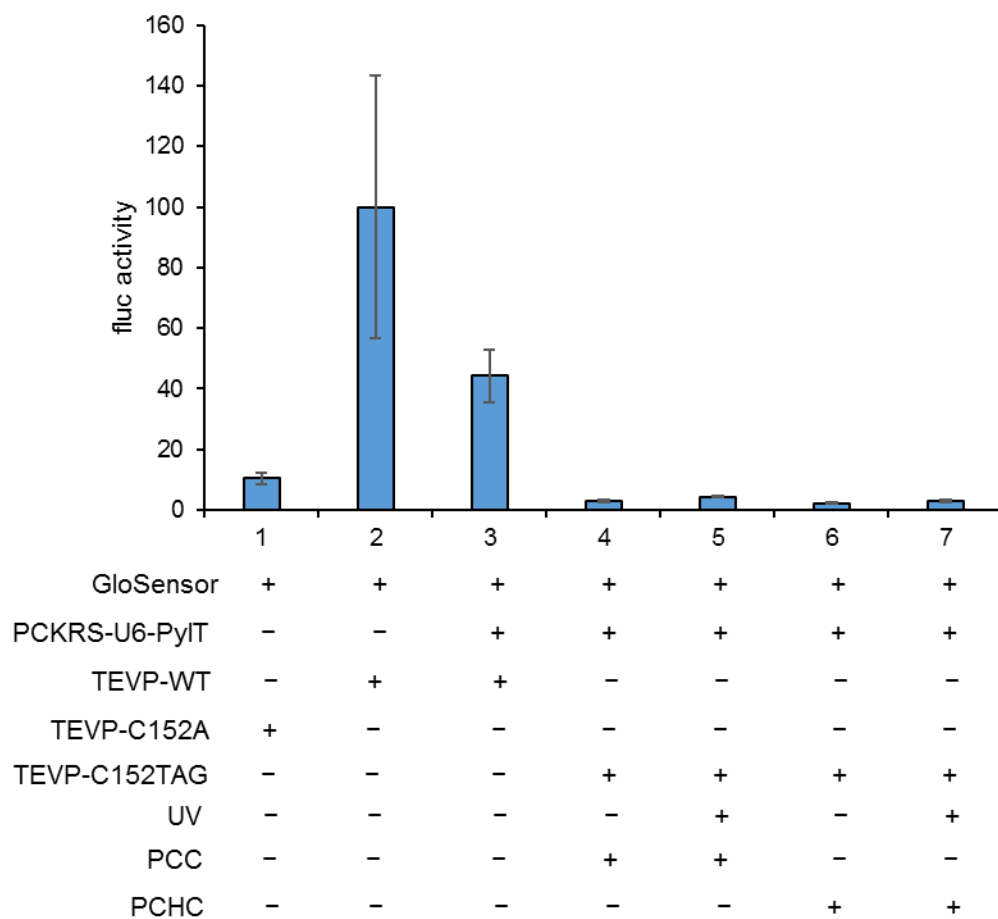


Figure 3.11. TEV protease activity assay with incorporation of PCC or PCHC, followed by light activation. TEVP-C152A is an inactive negative control (lane 1) and TEVP-WT is an active positive control (lane 2). TEVP-WT is still active when transfected with one more plasmid, pPCKRS-U6-PyIT (lane 3). For both PCC (lane 4 and 5) and PCHC (lane 6 and 7), no increase of luciferase activity was observed with

UV irradiation (365 nm, 2 min). All luciferase activities were normalized to that of TEVP-WT. Error bars, s.d.; n=4.

To confirm the incorporation of PCC or PCHC into TEV protease, we performed western blot analysis. While the expression of wild-type TEVP was detected, no expression of mutant TEVP with PCC/PCHC/PCK at position 152 was observed (Fig 3.12). The result indicated that, similar to the intein case, the low incorporation efficiency of PCC was the main reason for unsuccessful light-activation of TEV protease in mammalian cells. While we worked on this project, Jason Chin's lab reported the photoactivation of TEV protease through incorporation of a caged cysteine.²³ They evolved a PyIRS mutant for the genetic incorporation of a caged cysteine with different chemical structure from ours (discussed in chapter 2.4). However, it is surprising that the incorporation efficiency of their caged cysteine (at the concentration of 2 mM) was lower than that of our caged cysteine (tested by Subhas Samanta in our lab).

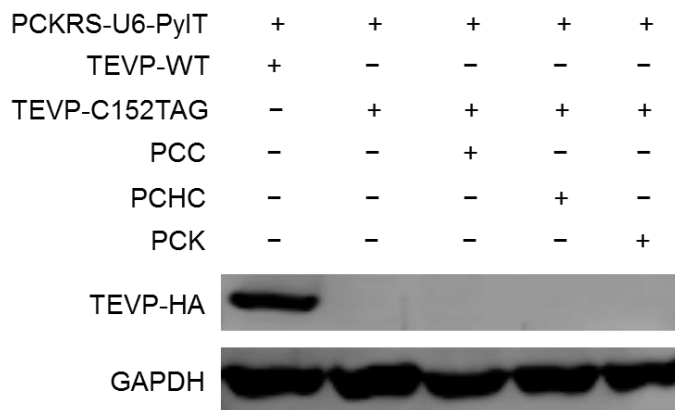


Figure 3.12. Western blot analysis of the incorporation of PCC/PCHC/PCK into TEVP in HEK293T cells. Cells were transfected with pPCKRS-U6-PyIT and pTEVP-C152TAG (1,000 ng each) in a 6-well plate. The culture medium was supplemented with designated UAA (2 mM). After 48 h, cell lysate was collected for western blot (with anti-HA and anti-GAPDH antibodies). pTEVP-WT was used as a positive control.

Since no incorporation was observed in mammalian cells, we sought to express TEV protease in bacterial cells. To this end, the pBAD-MBP-TEVP-pylT plasmid was constructed. The MBP-TEVP construct encodes MBP fused to the catalytic domain of TEV protease; the protein cleaves itself *in vivo* to yield a TEV protease catalytic domain.¹²⁸ When analyzed on SDS-PAGE, the cell lysate contained a mixture of both MBP and TEVP, while the nickel affinity purified protein (using the His₆ tag) only contained TEVP. Incorporation of PCC into TEVP-C152TAG was observed (Fig 3.13), although the yield was lower than that of the wild-type TEV protease. In the future, purified TEVP-C152TAG-PCC can be used for light activation of TEV protease *in vitro*.

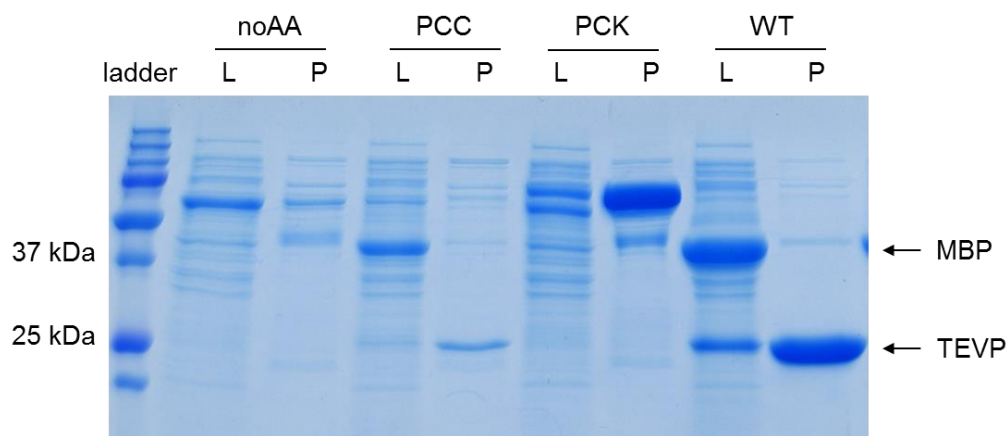


Figure 3.13. Incorporation of PCC into MBP-TEVP in *E. coli*. L indicates crude lysate. P indicates Ni-purified protein. The cell lysate contains a mixture of MBP and TEVP, while the Ni-purified protein (using the His₆ tag) only contains TEVP. Wild-type MBP-TEVP was used as a positive control.

Next, we tried light activation of another protein target, Src tyrosine kinase. It has been previously reported that the gatekeeper threonine in Src can be mutated to cysteine (T338C), which retains the function of the wild-type kinase and allows targeting with electrophilic inhibitors.¹²⁹ We hypothesized that the incorporation of a caged cysteine at the gatekeeper position could efficiently block Src activity until the removal of caging group by light. This strategy differs from other reported strategies for achieving conditional control of Src activity. In one example, Huiwang

Al's lab used a photocaged intein to control the Src activity with light.¹¹⁹ In another example, small molecule-triggered decaging was achieved by incorporating a *trans*-cyclooctene-caged lysine in Src.¹³⁰

To develop photocaged Src, we first generated a pSrc-T338TAG-EGFP-HA construct. EGFP was attached at the C terminus of Src as an incorporation reporter. HEK 293T cells were transfected with pSrc-T338TAG-EGFP-HA (300 ng) and pPCKRS-U6-PyIT (100 ng) in a 96-well plate. The culture medium was supplemented with PCC (1 mM). At 48 h, the incorporation of PCC into Src-T338TAG-EGFP-HA was detected by fluorescence microscopy (Fig 3.14).

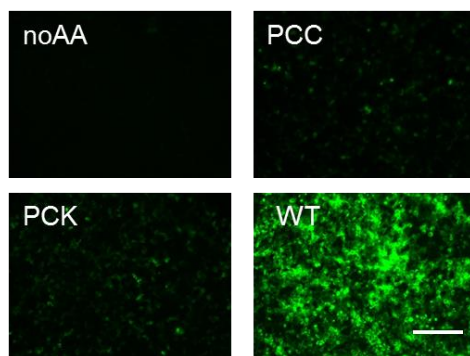


Figure 3.14. Incorporation of PCC into Src-T338TAG-EGFP-HA. HEK 293T cells were transfected with pSrc-T338TAG-EGFP-HA (300 ng) and pPCKRS-U6-PyIT (100 ng) in a 96-well plate. The culture medium was supplemented with PCC or PCK (1 mM). Scale bar indicates 200 μ m. Fluorescent images were taken at 48 h post-transfection.

We then sought to confirm the incorporation by western blot analysis. We tested both Src-T338TAG-EGFP-HA and Src-T338TAG-HA constructs, and the expressions of full-length Src-EGFP-HA and Src-HA proteins in the presence of PCC were both observed (Fig 3.15). This result indicated successful incorporation of PCC into the T338 position of Src kinase.

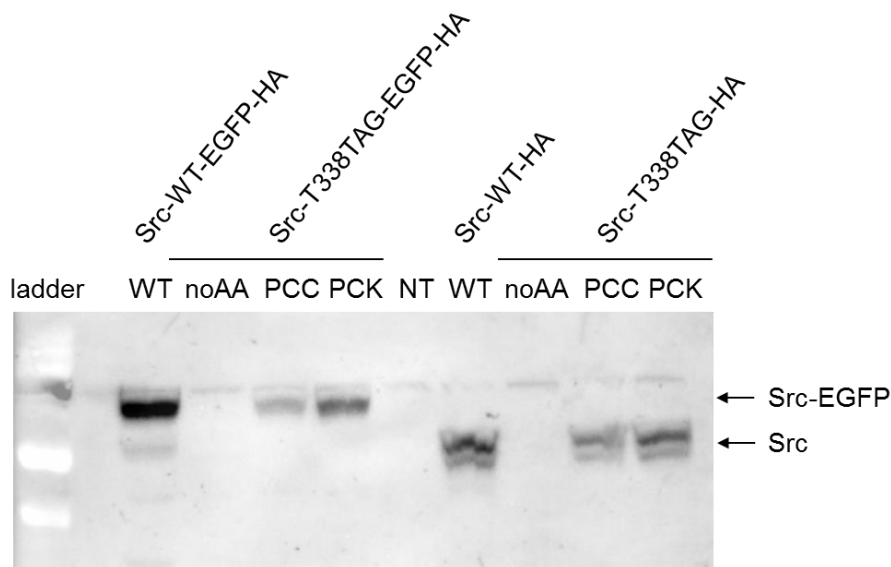


Figure 3.15. Incorporation of PCC into T338TAG position of Src or Src-EGFP. HEK293T cells were transfected with pPCKRS-U6-PyIT (1,000 ng) and pSrc-T338TAG-EGFP-HA (or pSrc-T338TAG-HA, 3,000 ng). The culture medium was supplemented with the designated UAA (1 mM). After 48 h, cell lysate was collected for western blot analysis (with anti-HA antibody). PCK was used as a positive control for incorporation. NT indicates non-treated cells. WT indicates wild-type Src.

We further performed the light-activation of photocaged Src kinase *in vitro*. The Src proteins were purified from the cell lysates (expressed in mammalian cells), using the HA-antibody-based immunoprecipitation. The Src proteins (Src-WT-EGFP-HA, Src-T338PCC-EGFP-HA) were irradiated with UV light (365 nm, VWR) for 10 min, and the *in vitro* Src kinase assays were performed on both irradiated and non-irradiated samples. While we observed the kinase activity of Src-WT-EGFP-HA, no activity of Src-T338PCC-EGFP-HA was observed with UV treatment (Fig 3.16). Since the expression of Src-T338PCC-EGFP-HA was lower than that of Src-WT-EGFP-HA (Fig 3.15), we speculated that increasing expression level is needed for successful light-activation assay. In the future, we are going to apply the *EcLeuRS*/tRNA pair (discussed in chapter 2.4), which showed higher incorporation efficiency, for the genetic incorporation of

photocaged cysteine into Src kinase. We might also test the photoactivation of the Src kinase function through incorporation of a photocaged lysine at its critical K295 position.¹³¹ The corresponding K295TAG mutation has already been made on the pSrc-EGFP plasmid.

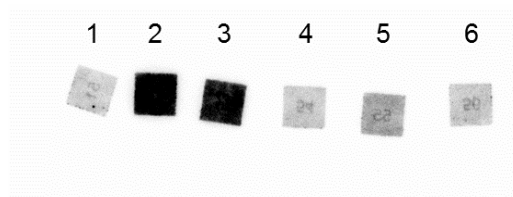


Figure 3.16. Light-activation assay for the purified Src-T338PCC-EGFP-HA protein. The cell lysates were purified by immunoprecipitation. The Src proteins were irradiated with UV light (365 nm) for 10 min, and the *in vitro* Src kinase assays were performed on both irradiated and non-irradiated samples. The numbers indicate: 1. non-treated; 2. Src-WT-EGFP-HA, -UV; 3. Src-WT-EGFP-HA, +UV; 4. Src-T338TAG-EGFP-HA, noAA; 5. Src-T338TAG-EGFP-HA, PCC, -UV; 6. Src-T338TAG-EGFP-HA, PCC, +UV.

In summary, we incorporated caged cysteine and caged homocysteine into proteins using the PylRS system. The attempts to photoactivate intein function and TEV protease function through incorporation of caged cysteines were not successful. The limitation for the successful incorporation of caged cysteine was the low efficiency, and we later found that the *E. coli* LeuRS could incorporate caged cysteine with a higher efficiency in mammalian cells (discussed in Chapter 2.4). Our future work will be to test the photoactivation of the Src kinase function, either through the incorporation of a caged cysteine at the gatekeeper position, or through the incorporation of a caged lysine at the critical lysine position. We will also confirm the light activation of RFP with DnaE intein (a gift from Dr. Huiwang Ai),¹¹⁹ and further test whether the incorporation of a homocysteine could trigger the splicing event as well.

Experimental

Protein expression

The incorporations of PCC, PCHC, or PCK into the Tyr151 position of sfGFP were performed as described in chapter 2.1. Briefly, pBAD-sfGFP-Y151TAG-pyIT and the PyIRS mutant EV2 were co-transformed (50 ng each) into *E. coli* Top10 cells. A single colony was grown overnight (37 °C, 250 rpm) in LB (3 ml) supplemented with Amp (50 µg/ml) and Tet (25 µg/ml), and the overnight culture (250 µl) was added to LB (25 ml) supplemented with the designated unnatural amino acid (2 mM), Tet (25 µg/ml), and Amp (50 µg/ml). Cells were grown at 37 °C, 250 rpm, and protein expression was induced with arabinose (0.1 %) when OD₆₀₀ reached 0.4 (measured using a Nanodrop). After overnight expression at 37 °C, cells were pelleted (5000 g, 10 min) and resuspended in phosphate lysis buffer (pH 8.0, 50 mM phosphate, 6 ml). Triton X-100 (60 µl, 10%) was added to the mixture. The lysate was incubated on ice for 1 h, sonicated (power level 5, pulse 'on' for 30 sec, pulse 'off' for 30 sec, with a total of 4 min, Fisher Scientific 550 sonic dismembrator), and then centrifuged (13,000 x g) at 4 °C for 10 min. The supernatant was transferred to a 15 ml conical tube, and Ni-NTA resin (Qiagen, 100 µl) was added. The mixture was incubated at 4 °C for 2 h with gentle shaking. The resin was then collected by centrifugation (1,000 g, 10 min) and washed with lysis buffer (300 µl); this was repeated twice, followed by two washes with wash buffer (300 µl) containing imidazole (20 mM). The protein was eluted with elution buffer (300 µl) containing imidazole (250 mM). The purified proteins were analyzed by SDS-PAGE (10%), and stained with Coomassie Brilliant Blue. The incorporations of PCC and PCK into the Cys152 position of TEV protease were performed in the same way.

Plasmid construction

To construct pAG31-Intein, we first amplified the EGFP(N)-Intein(N) fragment (Fwd: 5'-aattgctagcaccatggcaagcaaaggagaagaac-3', Rev: 5'-ctgcacgcgcggtccggccaccctgtct-3') and the intein(C)-EGFP(C) fragment (Fwd: 5'-gccggaccgcgcgtgcaggcgttcgcggat-3', Rev: 5'-

ataccaattgtaagcgtaatctggaacatcgatgggtattttagagctcatccatg-3') from the EGFP-RecA intein plasmid (gift from Dr. David Liu).^{124a} The two fragments were then pieced together with an overlap PCR (Fwd: 5'-aattgctagcaccatggcaagcaaaggagaagaac-3', Rev: 5'-ataccaattgtaagcgtaatctggaacatcgatgggtattttagagctcatccatg-3'). This step deleted the ligand-binding domain that was present in the original plasmid. The overlap PCR product was further digested with NheI and MfeI restriction enzymes (37 °C, 2 h). The pAG31 plasmid (8,000 ng, containing a PCKRS gene and a mCherry-TAG-EGFP reporter gene) was digested with NheI and MfeI enzymes, followed by the gel purification of the band corresponding to the correct length of the backbone (~6,500 bp). This step cut the mCherry-TAG-EGFP reporter gene. The pAG31-Intein plasmid was constructed by ligating the intein into the pAG31 backbone using NheI and MfeI restriction sites (T4 ligase, 16 °C, overnight). The pAG31-ΔI and pAG31-ΔIhh plasmid were constructed in a similar way. For pAG31-ΔI, the EGFP(N)-ΔI(N) fragment was amplified with Fwd: 5'-aattgctagcaccatggcaagcaaaggagaagaac-3', Rev: 5'-gctgtcgccgaagccgtcgaatcttctgggctgggccaccctgtctcccttg-3', and the ΔI(C)-EGFP(C) fragment was amplified with Fwd: 5'-gacggcttcggcgacagcgccccatccccgcccgcgtgcaggcgttcgcg-3', Rev: 5'-ataccaattgtaagcgtaatctggaacatcgatgggtattttagagctcatccatg-3'. For pAG31-ΔIhh, the EGFP(N)-ΔIhh(N) fragment was amplified with Fwd: 5'-aattgctagcaccatggcaagcaaaggagaagaac-3', Rev: 5'-gccggctctccacgtctctcacggccaccctgtctcccttg-3', and the ΔIhh(C)-EGFP(C) fragment was amplified with Fwd: 5'-gtgagagacgtggagaccggcgaaactccgctattccgtg-3', Rev: 5'-ataccaattgtaagcgtaatctggaacatcgatgggtattttagagctcatccatg-3'.

To construct pΔI, the EGFP(N)-ΔI-EGFP(C) was amplified from the pAG31-ΔI plasmid (Fwd: 5'-aattgaattcatggcaagcaaaggagaagaac-3', Rev: 5'-atacgcgccgcttaagcgtaatctggaacatcg-3'), and was digested with EcoRI and NotI enzymes (37 °C, 2 h). pEGFP-N1 (8,000 ng, Clontech) was digested with EcoRI and NotI enzymes, followed by the gel purification of the band corresponding to the correct length of the backbone (~4,500 bp). The pΔI plasmid was

constructed by ligating the EGFP(N)- Δ I-EGFP(C) into the pEGFP-N1 backbone using EcoRI and NotI restriction sites (T4 ligase, 16 °C, overnight).

Site-directed mutagenesis was used for the generation of single mutation on a target plasmid, including the C109TAG mutation on p Δ I (Fwd: 5'-acgggaactacgcatagccttgccgaggggtacc-3', Rev: 5'-ggtagcctcggaaggatgcgtagttcccg-3', TAG mutation underlined), the C109A mutation on p Δ I (Fwd: 5'-agatgacgggaactacgcaggccttgccgagg-3', Rev: 5'-cctcggaagggctgcgtagttcccgatcatct-3', alanine mutation underlined), the C152TAG mutation on pTEVP (Fwd: 5'-cagactaaggacggccaggagggcagcccactggtgagc-3', Rev: 5'-gctcaccagtgggctgcctactggccgtccttagtctg-3', TAG mutation underlined), the C152A mutation on pTEVP (Fwd: 5'-cagactaaggacggccaggccggcagcccactggtgagc-3', Rev: 5'-gctcaccagtgggctgccggctggccgtccttagtctg, alanine mutation underlined).

To construct pBAD-MBP-TEVP-pyIT, the backbone of the pBAD plasmid was generated by digesting pBAD-Myo4TAG-pyIT plasmid (8,000 ng, gift from Dr. Ashton Cropp) with NcoI and XhoI restriction enzymes, followed by gel purification of the band corresponding to the correct length of the backbone (~5,500 bp). MBP-TEVP was amplified from the pRK793 plasmid (from Addgene, #8827, Fwd: 5'-aattccatggcgatcgaagaaggtaaactggtaatctgg-3', Rev: 5'-aattctcgagttagcgacggcgacgacgattcatg-3'), and digested with NcoI and XhoI restriction enzymes (37 °C, 2 h). The pBAD-MBP-TEVP-pyIT plasmid was constructed by ligating the MBP-TEVP fragment into the pBAD backbone using NcoI and XhoI restriction sites (T4 ligase, 16 °C, overnight). The C152TAG mutation was generated on the pBAD-MBP-TEVP-pyIT plasmid by site-directed mutagenesis (Fwd: 5'-caaaccaaggatgggcaggagggcagtcattagatc-3', Rev: 5'-gatactaattggactgcctactgcccaccttggttg-3', TAG mutation underlined).

To construct the pSrc-EGFP-HA plasmid, the pEGFP-N1 plasmid (8,000 ng, Clontech) was first digested with EcoRI and BamHI restriction enzymes, followed by gel purification of the band corresponding to the correct length of the backbone (~4,500 bp). Src was amplified from the pLNCX-Src plasmid (from Addgene, #13665, Fwd: 5'-gcggaattcaccatggggagcagcaagagc-3', Rev:

5'-gcgggatccgagccggagcctaggttctctccaggctgg-3'), and digested with EcoRI and BamHI restriction enzymes (37 °C, 2 h). The pSrc-EGFP-HA plasmid was constructed by ligating the Src fragment into the pEGFP backbone using the EcoRI and BamHI restriction sites (T4 ligase, 16 °C, overnight). The T338TAG mutation was generated on the pSrc-EGFP-HA plasmid by site-directed mutagenesis (Fwd: 5'-gagcccatctacatcgctctaggagtacatgagcaagggg-3', Rev: 5'-ccccttgctcatgtactccctagacgatgtagatgggctc-3', TAG mutation underlined). To facilitate western blot analysis of Src-WT, the HA tag (YPYDVPDYA) was added at the C-terminus of the Src on pLNCX-Src plasmid by site-directed mutagenesis (Fwd: 5'-cagcctggagagaacctatacccatacgatgttccagattacgcttaagcctggcagatctgcgcgagcaaaatttaagc-3', Rev: 5'-taggttctctccaggctggtactggggctctgtcgaggtgaag-3', HA tag underlined). The K295TAG mutation was generated on the pSrc-EGFP-HA plasmid by site-directed mutagenesis (Fwd: 5'-gtggccatatagactctgaagcccgccaccatg-3', Rev: 5'-cttcagagtctatgcccactctggtggtgc-3', TAG mutation underlined).

Fluorescent imaging of intein splicing

Human embryonic kidney (HEK) 293T cells were transfected with the single intein plasmid (200 ng, pAG31-Intein, pAG31-ΔI, pAG31-ΔIhh, pΔI) at ~75% confluency. pEGFP-N1 (200 ng) was also transfected as a control for fluorescence. For genetic incorporation of PCC/PCHC/PCK, HEK 293T cells were transfected with pEGFP(N)-Intein-C109TAG-EGFP(C) and pPCKRS-U6-PyIT plasmids (200 ng each) in a 96-well plate. The culture medium was supplemented with designated UAA (1 mM). Branched polyethylene imine (bPEI, 1.5 μl, 1 mg/ml) was used as the transfection reagent. After 48 h, cells were irradiated with UV light (365 nm, VWR transilluminator) for 1 min, and were incubated in the 37 °C incubator for another 1 h. Cells were then imaged on a Zeiss Observer Z1 microscope using the with EGFP channel (filter set 38 HE; ex. BP470/40; em. BP525/50).

Fluorescent imaging of Src-EGFP

HEK 293T cells were transfected with pSrc-T338TAG-EGFP-HA (300 ng) and pPCKRS-U6-PyIT plasmids (100 ng) in a 96-well plate. The culture medium was supplemented with the designated UAA (1 mM). Cells transfected with pSrc-WT-EGFP-HA (300 ng) were used as a positive control. Branched polyethylene imine (bPEI, 1.5 μ l, 1 mg/ml) was used as the transfection reagent. After 48 h, cells were then imaged on a Zeiss Observer Z1 microscope using the EGFP channel (filter set 38 HE; ex. BP470/40; em. BP525/50).

Luciferase activity assay for TEV protease function

HEK 293T cells were transfected with pTEVP-C152TAG, pPCKRS-U6-PyIT, and pGloSensor (Promega) plasmids (200 ng each) in a 96-well plate. The culture medium was supplemented with the designated UAA (2 mM). Branched polyethylene imine (bPEI, 1.5 μ l, 1 mg/ml) was used as the transfection reagent. After 24 h, the cells were irradiated with UV light (365 nm, VWR) for 2 min, and were further incubated at 37 °C incubator. After another 6 h, half of the culture medium (100 μ l) was removed from each well, and BrightGlo luciferase assay reagent (100 μ l, Promega) was added, followed by vigorous shaking on an orbital shaker (1 min). Luminescence was then read on a Tecan M1000 PRO microplate reader.

Western blot

For detecting the intein splicing product, HEK 293T cells were transfected with p Δ I (2,000 ng, encoding EGFP(N)- Δ I-EGFP(C)-HA) in a 6-well plate. For the genetic incorporation of PCC/PCHC/PCK, HEK 293T cells were co-transfected with pEGFP(N)-intein-C109TAG-EGFP(C) (1,000 ng) and pPCKRS-U6-PyIT (1,000 ng) in the presence or absence of UAA (1 mM) in a 6-well plate. Branched polyethylene imine (bPEI, 15 μ l, 1 mg/ml) was used as the transfection reagent. After 48 h, cells were irradiated with UV light (365 nm, VWR transilluminator) for 1 min,

and were incubated at 37 °C for another hour. The cells were then washed with chilled phosphate-buffer saline (PBS, 1 ml), and lysed in mammalian protein extraction buffer (250 µl, GE Healthcare). The cell lysates were separated by 10% SDS-PAGE (run at 60 V for 15 min, and 150 V for 45 min) and were transferred to a PVDF membrane (GE Healthcare). The membrane was blocked in tris-buffer saline (TBS) with 0.1% Tween 20 and 5% milk for 1 h. The blots were probed with the primary antibody (1:1,000, anti-HA (sc-805) or anti-GAPDH (sc-25778), Santa Cruz Biotechnology) overnight at 4 °C, followed by incubation with secondary goat anti-rabbit IgG-HRP antibody (1:20,000, sc-2004, Santa Cruz Biotechnology) for 1 h at room temperature. After washing, the blots were incubated with SuperSignal West Pico working solution (mixture of the Stable Peroxide Solution and the Luminol/Enhancer Solution, 500 µl each, Thermo Scientific) for 5 min at room temperature. The luminescence signal was detected using a Bio-Rad ChemiDoc (Chemi Hi Sensitivity setting, exposure time: 10 sec).

For the genetic incorporation of PCC/PCHC/PCK in TEV protease, HEK 293T cells were transfected with pPCKRS-U6-PylT and pTEVP-C152TAG (1,000 ng each) in a 6-well plate. The culture medium was supplemented with the designated UAA (2 mM). After 48 h, the cells were washed with chilled phosphate-buffer saline (PBS, 1 ml), and lysed in mammalian protein extraction buffer (250 µl, GE Healthcare). Western blot analysis (with anti-HA and anti-GAPDH antibodies) was performed as described above. Cells transfected with pTEVP-WT (1,000 ng) was used as a positive control.

For the genetic incorporation of PCC/PCK in Src kinase, HEK 293T cells were transfected with pPCKRS-U6-PylT (1,000 ng) and pSrc-T338TAG-EGFP-HA (or pSrc-T338TAG-HA) (3,000 ng) in a 6-well plate. The culture medium was supplemented with the designated UAA (1 mM). After 48 h, the cells were washed with chilled phosphate-buffer saline (PBS, 1 ml), and lysed in mammalian protein extraction buffer (250 µl, GE Healthcare). Western blot analysis (with anti-HA

antibody) was performed as described above. Cells transfected with pSrc-WT-EGFP-HA or pSrc-WT-HA (3,000 ng) were used as the positive control.

Src kinase assay

The purified Src proteins for *in vitro* kinase assays were first obtained through immunoprecipitation. The cell lysate (200 μ l, in mammalian protein extraction buffer) from mammalian cell expression (6-well format, the same as above) was incubated with the anti-HA antibody (10 μ l, sc-805, Santa Cruz) at 4 °C for 1 h. The protein A-sepharose (40 μ l, Sigma) was then added and the mixture was further incubated at 4 °C overnight (according to the manufacturers instructions). The mixture was washed with RIPA buffer (400 μ l, Thermo) four times (centrifugation at 1,000 g for 2 min for each round), and resuspended in 80 μ l of kinase buffer (50 mM hepes, 150 mM NaCl, 1 mM DTT, 5 mM MgCl₂, 30 μ M ATP, 200 μ M sodium orthovanadate, pH 7.4). The still immobilized Src proteins (Src-WT-EGFP-HA and Src-T338PCC-EGFP-HA) were irradiated with UV light (365 nm, VWR transilluminator) for 10 min. The kinase assays were performed as previously reported.¹³² First, a master mix of reaction buffer (150 μ l of kinase buffer, 15 μ l of peptide substrate (1 mM in water, #89161-058, Enzo), 0.4 μ l of γ -³²P ATP (4 μ Ci, 3,000 Ci/mmol)) was prepared. The purified Src protein (20 μ l) and the reaction buffer (30 μ l) were mixed and reacted at room temperature for 20 min. Acetic acid (25 μ l, 50% (v/v)) was added to the mixture, which was then spotted on a P81 phosphocellulose paper (Millipore). The paper was washed with 0.43% (v/v) phosphoric acid five times, rinsed with 100% acetone, and air-dried. The paper was then exposed to a phosphor storage screen overnight and scanned using a Typhoon Phosphorimager (GE Healthcare).

3.2 Incorporation of a Photoisomerizable Amino Acid into Horseradish Peroxidase

Although the “caging” strategy has been successfully applied to control protein function,¹³³ one limitation is that a protein will remain active once the caging group is removed. In other words, the control process is irreversible. To reversibly control the function of biomolecules, another set of chemical compounds, named “photoswitches”, are being used.¹³⁴ Since photoswitches undergo reversible configurational changes, multiple rounds of active/inactive states of a biomolecule could be achieved (Fig 3.17).¹³⁵ To reversibly control biological events, the incorporation of photoswitches has been applied to nucleic acids, peptides, enzymes, receptors and ion channels.¹³⁴ Among the diverse choice of photoswitches, azobenzene has been the most widely used one.¹³⁵ Azobenzene exists in *trans* configuration in the dark, but turns to *cis* configuration when irradiated with 340 nm light. The *trans* isomer could be regenerated when the compound is irradiated with 450 nm light or through thermal re-equilibration.¹³⁵ The azobenzene photoisomerization can proceed in this manner with alternating light treatments.

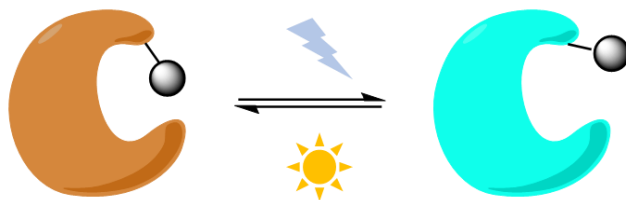


Figure 3.17. Reversible photocontrol of protein activity. Protein is inactive (orange) when the photoswitch is in one configuration, and active (cyan) when the photoswitch is in another configuration. Conversion between configurations can be achieved by different light sources.

One of the azobenzene derivatives, 4-azophenylalanine (AzoF), has been incorporated into proteins for reversible control of protein function (Fig 3.18).¹³⁶ When AzoF is placed close to the functional residues of a protein, the two isomers (*cis* and *trans*) may have different effects on

the protein function in a reversible manner. Two scenarios can occur: a switch with the default “ON” function and a switch with the default “OFF” function. In the first case, the *trans* configuration of the azobenzene enables protein function while the *cis* configuration blocks protein function, thus the protein is active until the *cis* configuration is triggered by light. In the second case, on the contrary, the *trans* configuration blocks protein function while the *cis* configuration enables protein function. Therefore, the protein is inactive until the *cis* configuration is triggered by light. Both types of switches can be useful tools in biomedical research, depending on what initial function of a protein of interest is required in each specific case.

Figure 3.18. Structure of AzoF (*cis* and *trans* state) unnatural amino acid studied in this chapter.

1),¹⁶ the synthetase pair could be transferred to the mammalian system, enabling broader applications in live cells.

Using the sfGFP-Y151TAG-pylT reporter plasmid (described in chapter 2.1), we screened four PylRS synthetases for the incorporation of AzoF in 5 ml expression scale. Although the SDS-PAGE gels for EV17 and EV20 were not successful (the incorporation of ONBY, a positive control UAA for EV20, was hardly observed on the gel), one mutant PylRS, EV16-5 (with Y271M, L274A, N311A, C313A, and Y349F mutations), was identified as a hit synthetase for AzoF (Fig 3.19). Since EV16-5 also incorporated ONBY (discussed in chapter 2.3), we speculated that the binding mode for AzoF was similar to that of ONBY.

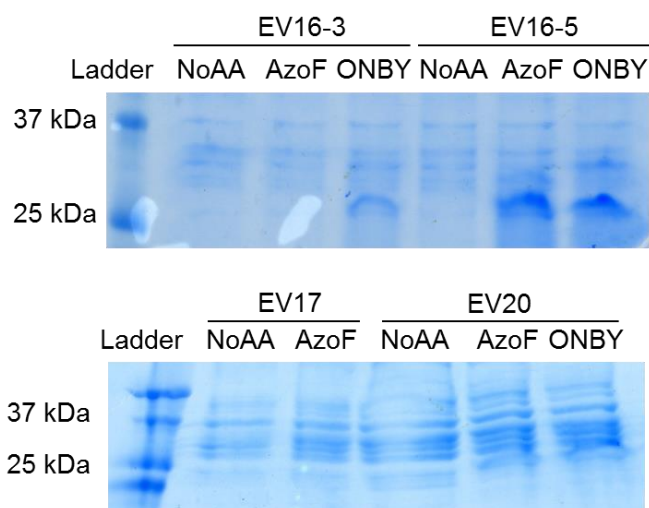


Figure 3.19. Screening of the synthetase panel reveals EV16-5 as a hit synthetase for AzoF. ONBY was used as a positive control for synthetase EV16-3, EV16-5 and EV20. Protein expression was performed in 5 ml scale. The sfGFP was extracted by three phase partitioning (TPP), and analyzed on a 10% SDS-PAGE.

We further confirmed the expression of sfGFP-Y151TAG-AzoF in *E. coli* in 25 ml expression scale, with the yield of 7 mg/L (Fig 3.20). The molecular weight of sfGFP-Y151TAG-AzoF determined by ESI-MS agrees well with its predicted molecular weight (found: 28,350.87Da,

calculated: 28,350.86Da). As expected, the mammalian counterpart of EV16-5 was confirmed to incorporate AzoF in mammalian cells (experiment performed by Ji Luo in our lab).

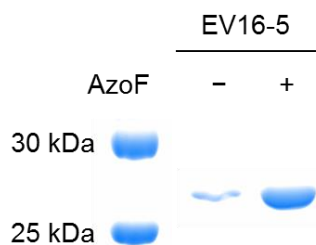


Figure 3.20. Incorporation of AzoF into sfGFP at the Tyr151 position. The PyIRS mutant, EV16-5 (with Y271M, L274A, N311A, C313A, and Y349F mutations), was used for incorporation.

We first asked whether AzoF undergoes isomerization in the context of a protein. We incorporated AzoF into sfGFP at the Tyr66 position, as what we have performed for photocaged thityrosine and photocaged azatyrosine (discussed in chapter 2.1). The sfGFP fluorescence served as a reporter for phenylalanine analogues at the Tyr66 position.⁷² Since AzoF replaced Tyr66 in the tri-peptide (Thr65-Tyr66-Gly67) chromophore (Fig 3.21), we tested whether its *trans/cis* isomerization led to reversible change of sfGFP fluorescence.

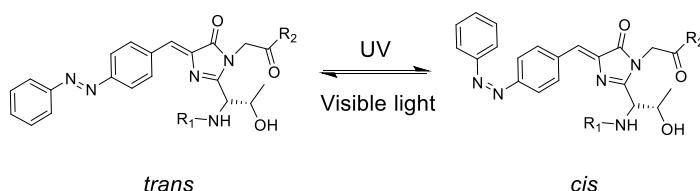


Figure 3.21. Configuration change of AzoF when incorporated into sfGFP at the Tyr66 position. The sfGFP fluorescence might be altered with different AzoF configuration.

To this end, the pBAD-sfGFP-Y66TAG-pylT plasmid was co-transformed with EV16-5 (50 ng each) into TOP10 cells, and the cells were expressed with or without AzoF (1 mM). Incorporation of AzoF (1 mM) into sfGFP at the Tyr66 position was successful (Fig 3.22). However,

some background incorporation was observed even in the absence of AzoF. This was not surprising, because EV16-5 was known to present some background incorporation (discussed in chapter 2.1 and 2.3). ONBY (0.25 mM) was used as a positive control for incorporation.

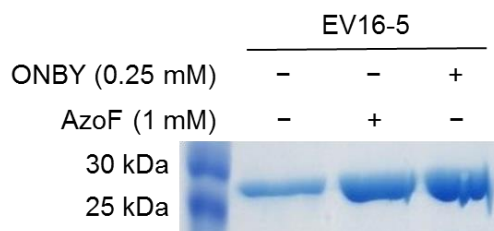


Figure 3.22. Incorporation of AzoF at the Tyr66 position of sfGFP. The PylRS mutant synthetase EV16-5 was used for incorporation. ONBY was used as a positive control for incorporation.

The sfGFP-Y66TAG-AzoF protein samples (5 μ M) were then irradiated with 365 nm UV light (VWR) for 1, 2, 5, 10 minutes, respectively. The protein sample with 10 minutes UV treatment was further treated with visible light (Just Normlicht Transparency Flat Viewer) for 10 minutes. The sfGFP fluorescence from each sample was measured (excitation wavelength: 488 nm; emission wavelength: 510 nm). The sfGFP-Y66TAG-ONBY protein samples (5 μ M) were treated in the same manner as a positive control. For sfGFP-Y66TAG-AzoF samples, the fluorescence remained the same either with UV or visible light treatment (Fig 3.23). The no UAA sample (protein purified from the cells that were expressed in the absence of AzoF) also showed same level of fluorescence, suggesting that the fluorescence was due to the background incorporation of the natural amino acid, not AzoF. The sfGFP-Y66TAG-ONBY samples presented increased fluorescence upon UV irradiation, which was in agreement with previous study.⁷² In this experiment, we only probed the fluorescence by a defined wavelength (excitation: 488 nm; emission: 510 nm). The Young lab incorporated the same AzoF using the *Mj* TyrRS system, and they probed the full spectra of the mutant GFP instead.¹³⁷ They found that when the excitation wavelength of 300 nm was used, the emission spectra showed a 50 nm blue-shift upon 365 nm

irradiation. Their results suggested that the azobenzene moiety impacted the fluorescence during isomerization.

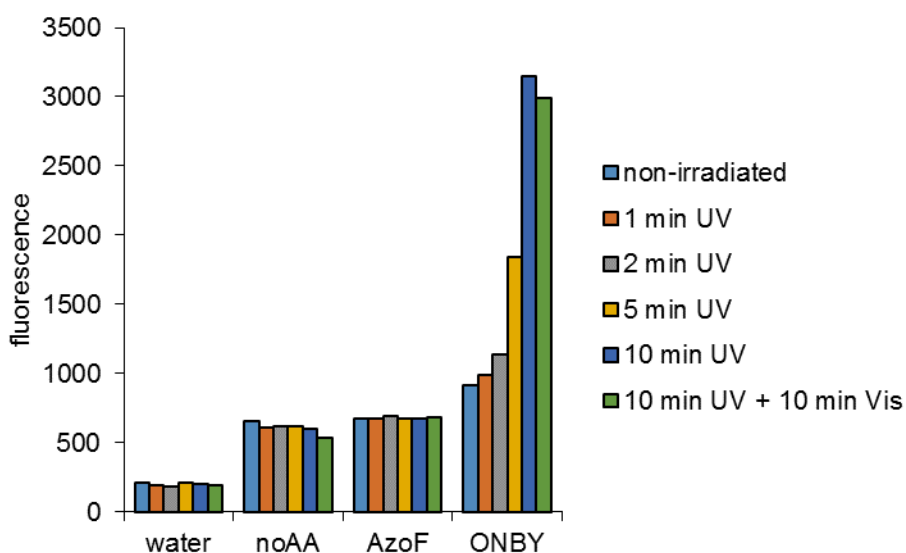


Figure 3.23. No fluorescent changes were observed for the sfGFP-Y66TAG-AzoF proteins treated with UV and visible light. The sfGFP-Y66TAG-AzoF protein samples (5 μ M) were irradiated with 365 nm UV light (VWR) for 1, 2, 5, 10 minutes, respectively. The protein sample with 10 minutes UV treatment was further treated with visible light (Just Normlicht Transparency Flat Viewer) for 10 minutes. The sfGFP fluorescence from each sample was measured (excitation wavelength: 488 nm; emission wavelength: 510 nm). The sfGFP-Y66TAG-AzoF proteins were used as the positive control. Only single measurement was performed.

We then tested another protein target, Horseradish Peroxidase (HRP). HRP is a heme protein that catalyzes the oxidation of electron donor substrates by H_2O_2 .¹³⁸ HRP activity can be assayed by several chromogenic substrates, which form colored products when oxidized. Due to its specificity and sensitivity, HRP has been extensively applied in analytical fields, such as biosensor and immunoassay design.¹³⁸ HRP has been previously used as a target protein for AzoF incorporation.^{136c} The authors incorporated AzoF using a four-base codon *in vitro* translation method. Among several incorporation positions they tested, two positions (F71 and F182) were

found to allow reversible control of HRP activity with alternating UV light and visible light treatments.^{136c} Therefore, we decided to incorporate AzoF into HRP at the same positions with UAA mutagenesis technique, and tested the reversible control of HRP activity.

To this end, a pBAD-HRP-pyIT plasmid was constructed and was co-transformed with the PyIRS synthetase EV16-5 (50 ng each) into TOP10 cells. The F71 and F182 positions were mutated to TAG respectively. Cells were expressed with or without AzoF (1 mM), and HRP proteins were extracted as inclusion bodies according to a previous protocol.¹³⁹ For the wild-type HRP, a distinct band at the expected molecular weight (34 kDa) was observed by SDS-PAGE analysis. However, no bands were observed for the HRP-F71TAG-AzoF and HRP-F182TAG-AzoF proteins (Fig 3.24).

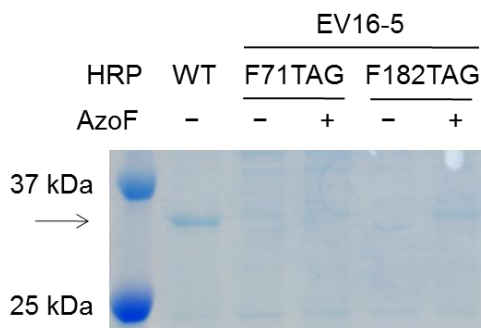


Figure 3.24. Incorporation of AzoF into HRP at the F71 and F182 positions, by PyIRS synthetase EV16-5. Wild-type HRP was expressed as a positive control. Arrow indicates expected molecular weight (34 kDa).

Despite undetectable protein expression on SDS-PAGE, we decided to directly test enzyme activity of the mutant HRP. The HRP samples were refolded and reconstituted with hemin,¹³⁹ followed by activity measurement using an ABTS-based spectrophotometric method.¹³⁹ Both reconstituted HRP-F71-AzoF and HRP-F182-AzoF proteins were active (Fig 3.25). The wild-type HRP and bovine serum albumin (BSA) were used as a positive and a negative control, respectively. It should be noted that some background activity was also observed in no UAA

samples (Fig 3.25), where proteins were purified from cells expressed in the absence of AzoF. This was most likely due to the background incorporation by EV16-5, as discussed in chapter 2.1 and 2.3.

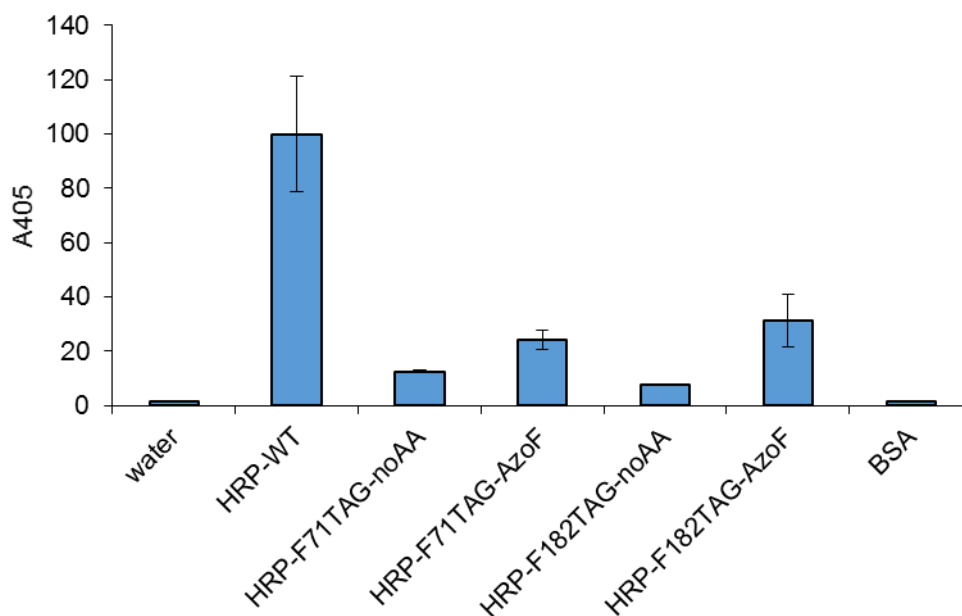


Figure 3.25. Enzymatic activities of HRP-F71TAG-AzoF and HRP-F182TAG-AzoF. Following refolding and reconstitution, the HRP samples (5 μ l) were directly used for ABTS assay. Wild-type HRP and BSA (250 ng/ μ l) were used as a positive and a negative control, respectively. The absorbance (at 405 nm) of each sample was measured and normalized based on wild-type HRP. Error bars, s.d.; n=3.

Since mutant HRP proteins were active, we further tested the reversible control with AzoF. The protein samples (HRP-WT, HRP-F71TAG-AzoF and HRP-F182TAG-AzoF, 10 μ l, directly used following refolding and reconstitution) were treated with consecutive UV (365 nm, VWR) and visible light (Just Normlicht Transparency Flat Viewer) (3 min for each exposure), and the protein activities were measured after each treatment (Fig 3.26). The activity of HRP-WT remained the same within each treatment, suggesting that light did not affect HRP activity. The activity of HRP-F182TAG-AzoF showed a reversible trend within each treatment: the activity decreased after UV

exposure, then increased after visible light exposure, and the trend continued for an additional round of light exposure (Fig 3.26). We also found that, contrary to that previous study,^{136c} the HRP activity did not decrease to zero with the UV treatment. This was most likely because the background activity of HRP (also shown in Fig 3.25) was not responsive to light. Nevertheless, our result suggested that AzoF functioned as a photoswitch at the F182 position of HRP. However, the HRP-F71TAG-AzoF protein did not present reversible activities with alternating light treatments (Fig 3.26). We speculated that lower expression and activity at the F71 position, based on result in Fig 3.25, might be the reason for unsuccessful reversible control.

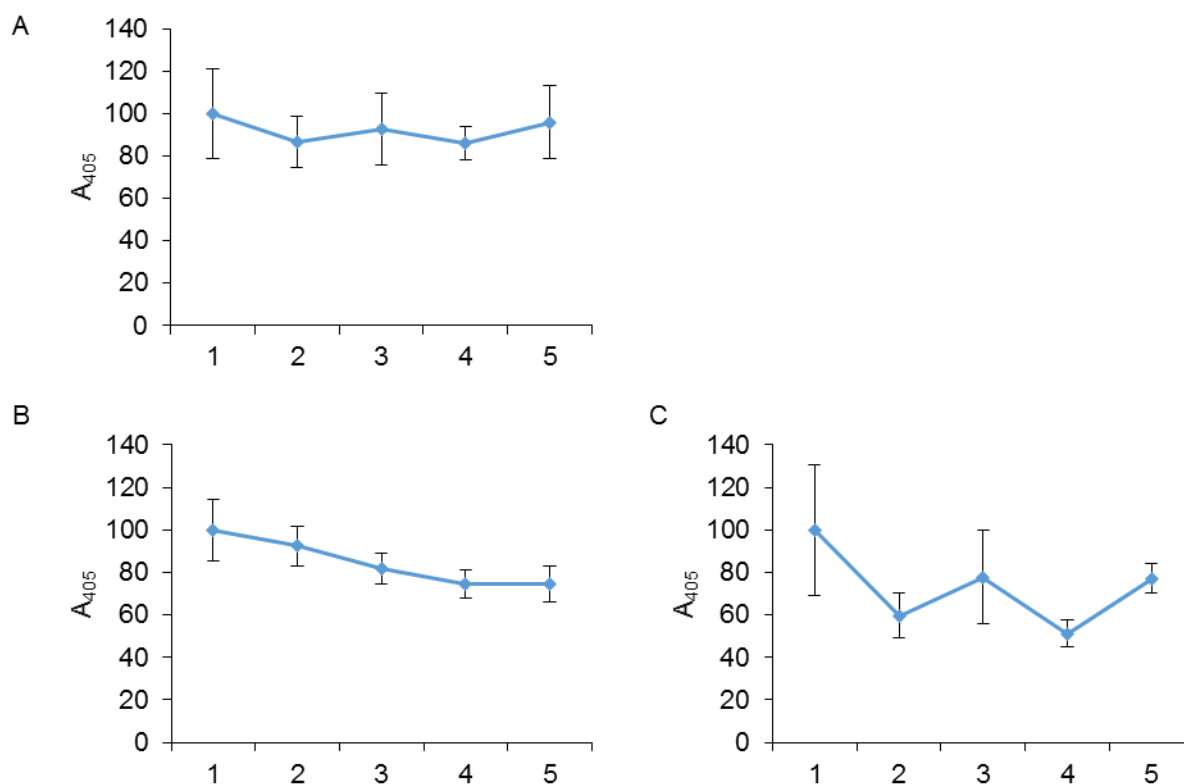


Figure 3.26. Enzymatic activity of (A) HRP-WT, (B) HRP-F71TAG-AzoF, (C) HRP-F182TAG-AzoF with alternating light treatments (365 nm UV light and visible light, 3 min each). The numbers indicate: 1. non-treated; 2. UV; 3. UV-Vis; 4. UV-Vis-UV; 5. UV-Vis-UV-Vis. The absorbance (at 405 nm) of each sample was measured and normalized based on non-treated sample. Error bars, s.d.; n=3.

Since many azobenzene-based derivatives showed different photoswitching properties (half-life, light wavelength),¹³⁵ we further sought to incorporate AzoF derivatives. We designed a fluorine-based AzoF (2f-AzoF), which might present extended half-life at room temperature.¹⁴⁰ Subhas Samanta in our lab synthesized 2f-AzoF and screened PylRS synthetase panel. The PylRS mutant, EV20 (with L270F, L274M, N311G, C313G, and Y349F mutations), was found to incorporate f-AzoF. The N311G and C313G might facilitate anchoring of the phenylalanine ring (discussed in chapter 2.3), while L270F and L274M mutations adjusted the space in the binding pocket for recognition of the second phenyl ring.

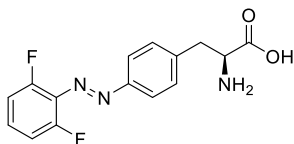


Figure 3.27. Structure of 2f-AzoF studied in this chapter.

We therefore tested the photoswitching of 2f-AzoF on HRP. The pBAD-HRP-F182TAG-pylT plasmid was co-transformed with the PylRS synthetase EV20 (50 ng each) into TOP10 cells, and the cells were expressed with or without 2f-AzoF (1 mM). The HRP proteins were extracted as inclusion bodies, and were further refolded and reconstituted with hemin. The protein samples (HRP-WT and HRP-F182TAG-2f-AzoF, 10 μ l, directly used following refolding and reconstitution) were treated with consecutive UV (365 nm, VWR) and visible light (Just Normlicht Transparency Flat Viewer) (3 min for each exposure), and protein activities were measured after each treatment. Similar to AzoF, 2f-AzoF also presented reversible HRP activities with alternating light treatments (Fig 3.28). While the initial result was promising, we later encountered difficulties expressing functional HRP again. We speculated that the HRP protein might not be folded properly. Since protein folding is a critical step for reconstituting a functional HRP, a robust folding protocol will be explored in the future for the expression of mutant proteins.

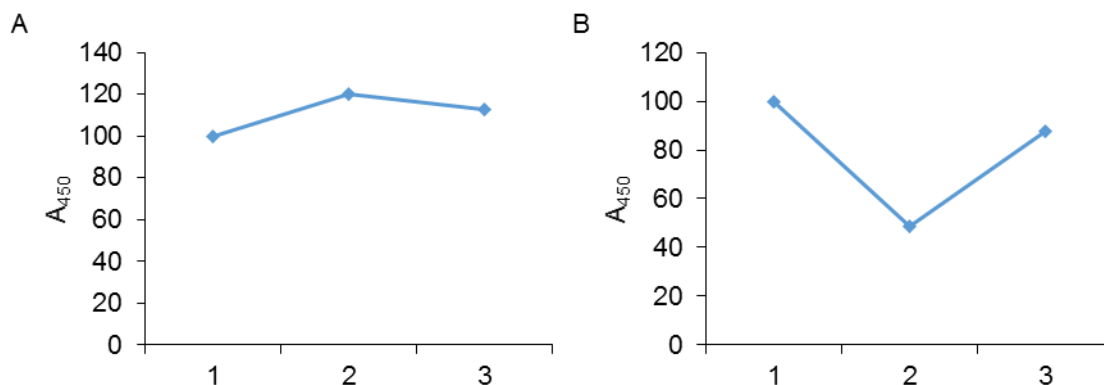


Figure 3.28. Enzymatic activity of (A) HRP-WT, (B) HRP-F182TAG-2f-AzoF with alternating light treatments (365 nm UV light and visible light, 3 min each). The numbers indicate: 1. non-treated; 2. UV; 3. UV-Vis. The absorbance (at 405 nm) of each sample was measured and normalized based on non-treated sample. Only single measurement was performed.

In summary, two photoisomerizable amino acids were genetically incorporated into proteins by the PyIRS/PyIT system. The incorporation of a photoisomerizable amino acid into HRP allowed for reversible control of HRP activity with light. As we worked on this project, the Lei Wang lab reported the genetic encoding of a photoswitchable click amino acid in bacterial and mammalian cells.³⁹ They designed an AzoF derivative bearing an additional benzyl chloride (Fig 3.29), which could be used for generation of a protein bridge by proximity-enabled reactivity towards cysteine. Incorporation of this UAA into calmodulin allowed reversible conformational change of calmodulin with light.³⁹ They later incorporated a fluorine-based photoswitchable click amino acid (Fig 3.29), which underwent isomerization in response to the visible light.¹⁴¹ In general, this reactivity-based strategy could potentially induce larger conformational change of a protein by introducing an intermolecular bond. However, it is questionable whether protein maintains its function with the newly introduced intermolecular bond. Although several photoswitchable UAAs have been genetically incorporated, they have not been applied to the reversible control of protein activity in mammalian cells. Ji Luo in our lab has confirmed the incorporation of AzoF and 2f-AzoF in

mammalian cells, and has achieved photoswitching of luciferase activity with AzoF and 2f-AzoF. In the future, we are going to target proteins that play important roles in cellular physiology, and may study cell signaling networks through photoswitching of protein activity.

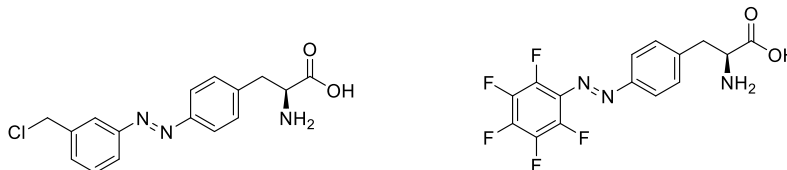


Figure 3.29. Structures of photoswitchable click amino acids reported by the Wang lab.

Experimental

Plasmid construction

To construct the pBAD-HRP-pylT plasmid, the backbone of the pBAD plasmid was generated by digesting pBAD-Myo4TAG-pylT plasmid (8,000 ng, gift from Dr. Ashton Cropp) with NcoI and NdeI restriction enzymes, followed by gel purification of the band corresponding to the correct length of the backbone (~5,500 bp). HRP was amplified from the plasmid pCMV-EGFP-ssHRP-TM (Addgene: #31156, Fwd: 5'- aattccatgggcggccagttaacc-3', Rev: 5'- aattcatatgttaatggtgatggtgatgtgccgagccagagttgctgttgaccac-3'), and digested with NcoI and NdeI restriction enzymes (37 °C, 2 h). The pBAD-HRP-pylT plasmid was constructed by ligating the HRP fragment into the pBAD backbone using NcoI and NdeI restriction sites (T4 ligase, 16 °C, overnight). Site-directed mutagenesis was used to generate F71TAG (Fwd: 5'- gtttccgcactgaaaaggatgcataggggaacgctaacag-3', Rev: 5'- ctgttagcgttcccctatgcatccttttcagtgcggaaac-3', TAG mutation underlined) and F182TAG (Fwd: 5'- ggaaagaaccagtgtaggtagatcatggataggctctaca-3', Rev: 5'- tgtagagcctatccatgatctacactactggttcttcc-3', TAG mutation underlined) mutations.

Protein expression in *E. coli*

Methods for the synthetase panel screening and the protein extraction by three-phase partitioning (TPP) were described in chapter 2.1. For the expression of sfGFP-Y151TAG-AzoF, pBAD-sfGFP-Y151TAG-pyIT and synthetase EV16-5 were co-transformed (50 ng each) into *E. coli* Top10 cells. A single colony was grown overnight (37 °C, 250 rpm) in LB (3 ml) supplemented with Kan (50 µg/ml) and Tet (25 µg/ml), and the overnight culture (250 µl) was added to LB (25 ml) supplemented with AzoF (1 mM), Tet (25 µg/ml), and Kan (50 µg/ml). Cells were grown at 37 °C, 250 rpm, and protein expression was induced with arabinose (0.1 %) when OD₆₀₀ reached 0.4 (measured by Nanodrop). After overnight expression at 37 °C, cells were pelleted (5,000 g, 10 min) and resuspended in phosphate lysis buffer (pH 8.0, 50 mM of phosphate, 6 ml). Triton X-100 (60 µl, 10%) was added to the mixture. The lysate was incubated on ice for 1 h, sonicated (power level 5, pulse 'on' for 30 sec, pulse 'off' for 30 sec, with a total of 4 min, 550 sonic dismembrator), and then pelleted (13,000 g) at 4 °C for 10 min. The supernatant was transferred to a 15 ml conical tube, and Ni-NTA resin (Qiagen, 100 µl) was added. The mixture was incubated at 4 °C for 2 h with mild shaking. The resin was then collected by centrifugation (1,000 g, 10 min) and washed with lysis buffer (300 µl), this was repeated twice, followed by two washes with wash buffer (300 µl) containing imidazole (20 mM). The protein was eluted with elution buffer (300 µl) containing imidazole (250 mM). The purified proteins were analyzed by SDS-PAGE (10%), and stained with Coomassie Brilliant Blue. For protein mass spectrometry, the purified protein (in dialysis tubing, 6,000-8,000 MWCO, Fisher) was dialyzed in 50 mM ammonia acetate solution (1 L) overnight at 4 °C. The concentration of the dialyzed sample was determined by the intensity of protein band on SDS-PAGE, using BSA as the standard (ImageJ). An aliquot of the dialyzed protein solution (10 µl, concentration adjusted to 5 µM with water) was analyzed by electrospray ionization mass spectrometry (Thermo Scientific Q-Exactive Orbitrap, connected to a Dionex Ultimate 3000 UHPLC system), via a ProSwift RP-10R, 1 mm by 5 cm column, flow rate of 200 µl/min and ACN gradient (0.1% formic acid) 26-80% for 30 min. The mass spectrometer was operated in ESI

positive-ion mode with a capillary voltage of 3.5 kV and resolution at 17,500. The Protein Deconvolution 3.0 software was used for the data analysis. The expression of sfGFP-Y66TAG-AzoF was performed in the same way, except using the pBAD-sfGFP-Y66TAG-pyIT plasmid (construction method described in chapter 2.1), instead of the pBAD-sfGFP-Y66TAG-pyIT plasmid, for transformation.

For the expression of HRP-F71TAG-AzoF, pBAD-HRP-F71TAG-pyIT and pBK-PylRS (EV16-5 mutant) were co-transformed (50 ng each) into the *E. coli* Top10 cells. A single colony was grown overnight (37 °C, 250 rpm) in LB (3 ml) supplemented with Kan (50 µg/ml) and Tet (25 µg/ml), and the overnight culture (250 µl) was added to LB (25 ml) supplemented with AzoF (1 mM), Tet (25 µg/ml), and Kan (50 µg/ml). Cells were grown at 37 °C, 250 rpm, and protein expression was induced with arabinose (0.1%) when OD₆₀₀ reached 0.4 (measured by Nanodrop). After overnight expression at 37 °C, cells were pelleted (5,000 g, 10 min), resuspended in 5 ml of buffer A (50 mM Tris-HCl, 1 mM EDTA, 10 mM DTT, pH 8), and incubated on ice for 30 min. The lysate was sonicated (power level 5, pulse 'on' for 30 sec, pulse 'off' for 30 sec, with a total of 4 min, 550 sonic dismembrator), and then pelleted (8,000 g) at 4 °C for 40 min, to allow isolation of inclusion bodies. The supernatant was discarded, and the pellet was resuspended in 5 ml of buffer B (50 mM Tris-HCl, 1 mM EDTA, 10 mM DTT, 2 M urea, pH 8). The lysate was pelleted (8,000 g) again at 4 °C for 40 min, and the supernatant was discarded. The pellet was resuspended in 5 ml of buffer C (50 mM Tris-HCl, 1 mM EDTA, 10 mM DTT, 6 M urea, pH 8). For HRP refolding, the solubilized inclusion bodies (500 µl) was added to 1 ml of refolding medium (20 mM Tris, 1.7 M urea, 4% glycerol, 2 mM CaCl₂, pH 8.5). Oxidized L-glutathione (70 mM, 10 µl) was then added, and the mixture was incubated at 4 °C overnight. Hemin (1 mM, 4 µl) was added the next day, and the mixture was buffer exchanged in 50 µl of phosphate buffer saline (PBS, pH 7.4) by ultra centrifugal filters (Amicon, 10K MWCO). The expressions for HRP-

F182TAG-AzoF and HRP-F182TAG-2f-AzoF were performed in the same way, except using the pBAD-sfGFP-F182TAG-pylT plasmid for transformation.

Fluorescence activation assay

The concentration of sfGFP-Y66TAG-AzoF was quantified based on the intensity of the bands on SDS-PAGE, using BSA as the standard (ImageJ). The sfGFP-Y66TAG-AzoF protein samples (5 μ M) were irradiated in a PCR tube with 365 nm UV light (VWR) for 1, 2, 5, 10 min, respectively. The sample with 10 min UV irradiation was further placed under visible light (Just Normlicht Transparency Flat Viewer) for 10 minutes. The sfGFP fluorescence from each sample was measured (excitation wavelength: 488 nm; emission wavelength: 510 nm) on a Tecan M1000 PRO microplate reader. The sfGFP-Y66TAG-ONBY protein samples (5 μ M) were used as the positive control.

Photoswitching and activity assay for HRP

The protein samples (HRP-WT, HRP-F71TAG-AzoF, HRP-F182TAG-AzoF, HRP-F182TAG-2f-AzoF, 10 μ l, directly used following refolding and reconstitution) were placed in PCR tubes and were treated with consecutive UV (365 nm, VWR) and visible light (Just Normlicht Transparency Flat Viewer) (3 min for each exposure), and protein activities were measured after each treatment. An ABTS-based spectrophotometric method was applied for measuring HRP activity *in vitro*.¹³⁹ The following reagents were mixed in an eppendorf tube at 25 °C: 290 μ l of ABTS (2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid), 9.1 mM), 10 μ l of 0.3% (w/w) H₂O₂, 5 μ l of reconstituted HRP protein. After 3 min, the absorbance at 405 nm ($A_{405\text{ nm}}$) was measured on a Nanodrop spectrophotometer.

3.3 Photo-deactivation of Protein Function through Degron or Incorporation of Nitrophenylalanine

Besides activation of protein function with light, we also asked if deactivation of protein function could be achieved with light, using UAA mutagenesis. We hypothesized that a photocleavable amino acid could be incorporated into a protein, such that light exposure would cleave the polypeptide backbone, thus splitting the protein into two non-functional fragments. To this end, Qingyang Liu (Deiters' lab member) synthesized a photocleavable hydroxyl acid (PHA) (Fig 3.30A). Previously, several hydroxyl acids have been genetically incorporated into proteins using the wild-type, non-engineered PylRS/PylT system,^{31, 142} indicating the efficient recognition of the hydroxyl acid moiety by the same PylRS as for its amino acid counterpart. We therefore performed screening of three PylRS synthetases for PHA in *E. coli* EV16-3 (with N311A, C313A, and Y349F mutations) was found to be the hit synthetase for PHA in the initial screening (Fig 3.30C). However, SDS-PAGE analysis of the expressed protein suggested the incorporation of natural amino acid, but not PHA (Fig 3.30D). It was most likely that the N311A and C313A double mutations facilitated recognition of phenylalanine.^{20a} In the future, screening of other synthetases in our current panel should be performed, as well as expression in minimal media.

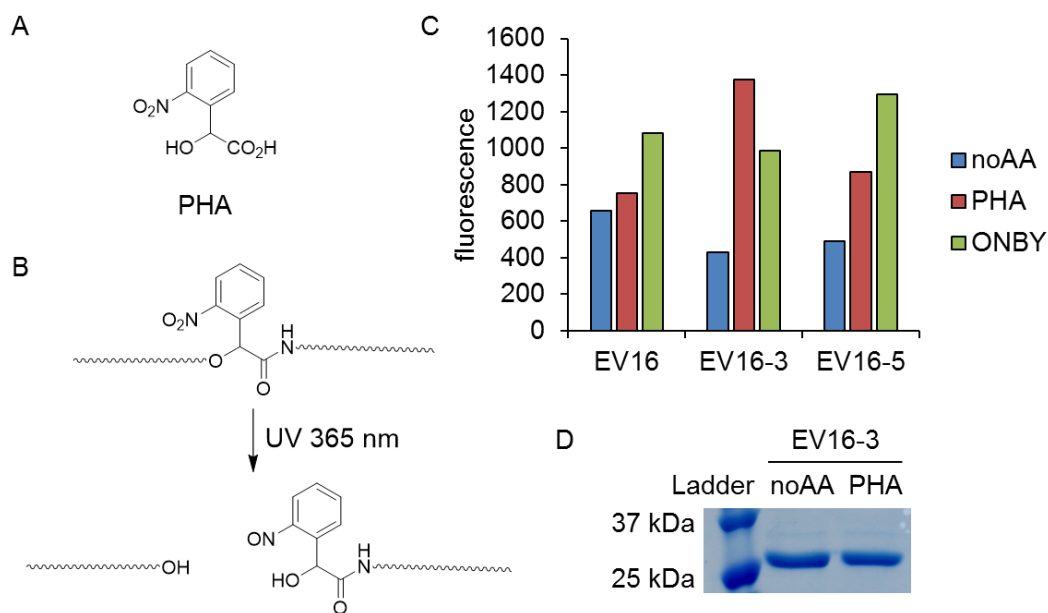


Figure 3.30. Incorporation of a photocleavable hydroxyl acid (PHA) into protein. (A) Structure of PHA. (B) Incorporation of PHA allows light-triggered peptide cleavage, and thus deactivation of protein function. (C) The screening of three PyIRS synthetases for the incorporation of PHA reveals EV16-3 as a hit. (D) SDS-PAGE analysis suggests incorporation of the natural amino acid by EV16-3.

Previously, 2-nitrophenylalanine (2-NF) has been genetically incorporated into proteins using the *M. jannaschii* tyrosyl-tRNA synthetase/tRNA^{Tyr} pair.¹⁴³ Through the *in vitro* peptide-based assay, the authors found that photolysis of 2-NF led to the cleavage of the peptide backbone, generating a C-terminus carboxylate group and a N-terminus cinnoline group.¹⁴³ The 2-NF was further site-specifically incorporated into T4 lysozyme. Upon light irradiation (365 nm) for 60 min, the T4 lysozyme was cleaved into two fragments, with the yield of 22% (quantified by optical densitometry of a Coomassie-stained gel).¹⁴³ Recently, the 2-NF was also found to be a substrate of a PyIRS mutant (EV16, with N311A and C313A mutations).^{20b} The N311A and C313A double mutations facilitate the incorporation of phenylalanine derivatives.^{20a} Moreover, the same mutant could also incorporate 3-nitrophenylalanine (3-NF).^{20c} Since the PyIRS/PyIT system could

be readily applied in mammalian cells,¹⁶ we decided to test the incorporation and photocleavage of 2-NF and 3-NF in mammalian cells.

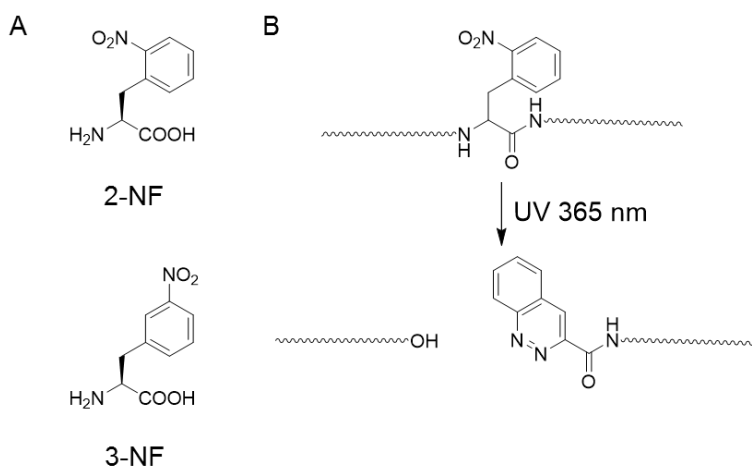


Figure 3.31. Structures of photocleavable unnatural amino acids. (A) Structure of 2-nitrophenylalanine (2-NF) and 3-nitrophenylalanine (3-NF). (B) Incorporation of 2-NF allows light-triggered peptide cleavage, and thus deactivation of protein function.

We first used the mCherry-TAG-EGFP-HA construct as a reporter, where EGFP indicates incorporation efficiency, and mCherry serves as a control for transfection efficiency. We coined the PyIRS mutant with N311A and C313A mutations as NFRS. HEK 293T cells were transfected with pNFRS-U6-PyIT (constructed by Ji Luo, simultaneously expressing NFRS and PyIT) and pmCherry-TAG-EGFP-HA plasmids (200 ng each), and were supplemented with 2-NF (2 mM) or 3-NF (2 mM). Fluorescence images were taken after 48 h. Some moderate incorporation of 2-NF and 3-NF was observed (Fig 3.32) when compared to the noAA control. Although the 2-NF (5 mM) has been genetically incorporated in mammalian cells, the reported fluorescence intensity was only 2.2-fold higher than the noAA group).^{20b}

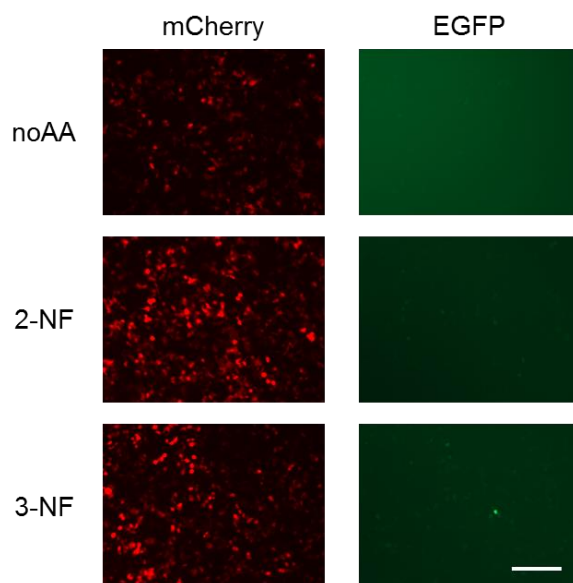


Figure 3.32. Genetic incorporation of 2-nitrophenylalanine (2-NF) and 3-nitrophenylalanine (3-NF) into mCherry-TAG-EGFP-HA. HEK293T cells were transfected with pNFRS-U6-PylT and pmCherry-TAG-EGFP-HA (200 ng each) in a 96-well plate. The culture medium was supplemented with 2-NF or 3-NF (2 mM). Images were taken at 48 h, with 10x objective. Scale bar indicates 200 μ m.

We proceeded to test the photocleavage of 2-NF and 3-NF in the mCherry-TAG-EGFP-HA fusion protein. With the incorporation of 2-NF or 3-NF at the position between mCherry and EGFP, light-triggered cleavage should generate two separate proteins (mCherry and EGFP-HA). The un-cleaved fusion protein (mCherry-TAG-EGFP-HA) and one of the cleaved protein products (EGFP-HA) could then be differentiated on an anti-HA western blot. We tested UAAs at the concentration of either 2 mM (used in the above experiments) or 5 mM (reported in the literature^{20b}). HEK293T cells were transfected with pNFRS-U6-PylT and pmCherry-TAG-EGFP-HA (1,000 ng each), and were supplemented with 2-NF or 3-NF (at concentrations of 2 mM or 5 mM). After 48 h, cell lysates were collected and irradiated in an Eppendorf tube with UV light (365 nm, VWR transilluminator) for 10 min. Non-irradiated and irradiated cell lysates were then analyzed by western blot with an anti-HA antibody. We did not observe incorporation of 2-NF into

mCherry-TAG-EGFP-HA at either concentration (2 mM or 5 mM) (Fig 3.33, lane 3 and 4). The incorporation of 3-NF was successful (Fig 3.33, lane 5 and 6). With UV irradiation, a cleaved band was observed, with cleavage efficiency of ~10% (Fig 3.33, lane 5, +UV). In the future, screening of the PyIRS panel might identify better synthetases for 2-NF and 3-NF incorporation. For the photocleavage assay, higher cleavage efficiency might be achieved with longer irradiation times (> 10 min). Lastly, after demonstrating successful protein cleavage with the mCherry-TAG-EGFP-HA construct, the strategy might be further applied to a functional enzyme in live cells.

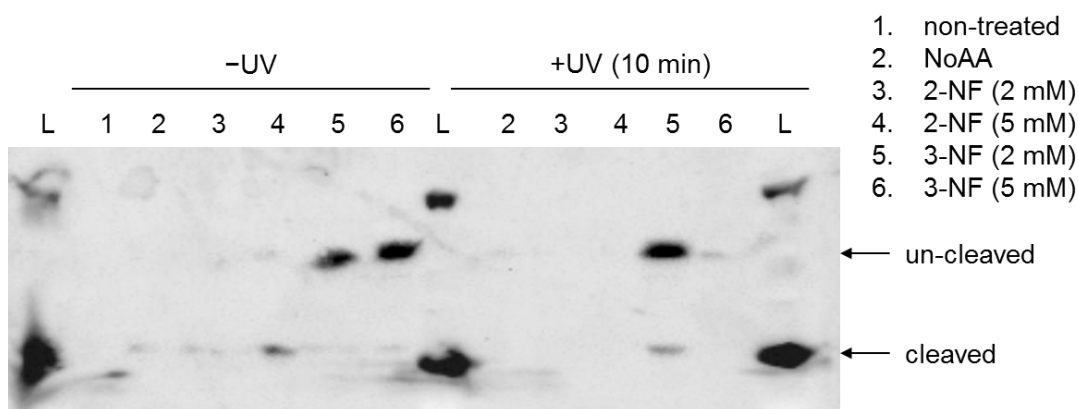


Figure 3.33. Photo-cleavage assay with the incorporation of 2-NF or 3-NF. HEK293T cells were transfected with pNFRS-U6-PylT and pmCherry-TAG-EGFP-HA (1,000 ng each) in a 6-well plate. The culture medium was supplemented with 2 mM or 5 mM of 2-nitrophenylalanine (2-NF) or 3-nitrophenylalanine (3-NF). After 48 h, the cell lysates were collected and irradiated with UV light (365 nm, VWR transilluminator) for 10 min. Non-irradiated and irradiated cell lysate were then analyzed by western blot with anti-HA antibody. Un-cleaved product (mCherry-TAG-EGFP-HA, 56 kDa) and cleaved product (EGFP-HA, 28 kDa) are indicated by arrows.

In addition to deactivation of protein function through cleavage of the peptide backbone, the ubiquitin system has also been applied for conditional protein degradation.¹⁴⁴ For example, the PROteolysis TArgeting Chimeric moleculeS (PROTACS) technology allows for small molecule-induced protein degradation by recruiting the target protein to the proteasome.¹⁴⁵

Protein degradation has also been achieved through hydrophobic tagging of a Halo Tag fusion protein.¹⁴⁶ In addition to small molecule-induced degradation, light has also been used as an external trigger for protein degradation.¹⁴⁷ In one example, a blue-light inducible degron domain (B-LID) was developed, where a small peptide degron is fused to the C-terminus of the LOV2 domain.^{147a} The protein with the B-LID fused at its C-terminus is stable, but upon exposure to blue light (2 h), the degron is exposed, leading to proteasomal degradation of the protein. In another example, the protein is fused to an auxin-inducible degron, and a photoactivable auxin is further used for spatiotemporal control of protein stability.^{147b} In the above examples, the protein of interest needs to be attached to an exogenous domain, which might affect protein function. Therefore, we asked if we could develop a light-triggered protein degradation strategy using UAA mutagenesis.

Based on our previous success with the light-activation of protein function with photocaged lysine,^{19d, 60, 148} we first sought to find a degradation domain (also called degron) where the lysine residue plays an essential role. We ruled out possible degrons that need to be fused to the C-terminus of the target protein¹⁴⁹ since truncated but functional proteins (where the TAG codon was recognized as a stop codon) would be generated but would not be responsive to optical control. For the N-terminal degron, we found an excellent model system where protein proteolysis was dependent on an N-terminal fused ubiquitin.¹⁵⁰ In this study, ubiquitin was fused to the N-terminus of GFP, with a variable amino acid at the first position of GFP. Shortly after protein synthesis in cells, the N-terminal ubiquitin moiety was rapidly cleaved from the precursor by endogenous deubiquitination enzymes (DUBs).¹⁵¹ The half-life of the cleaved product (GFP) was dependent on the first amino acid of the GFP fragment. Known as the N-end rule, this residue determines the half-life of a protein.¹⁵² For example, if it is one of these four amino acids: Arg/Lys/Phe/Leu, the protein bears a short half-life (~ 2 min). This is because the residue is recognized by an ubiquitin ligase, recruiting it to the protein, and subsequently ubiquitinating another Lys residue, resulting in rapid degradation by the proteasome.¹⁵³ Alternatively, if it is an

Ala/Met/Ser, the protein shows a long half-life (> 600 min).¹⁵⁴ Based on this knowledge, we speculated that light-triggered protein deactivation could be achieved through the incorporation of a photocaged lysine into the critical lysine position of the Ub-K-POI (protein of interest). The caging group blocks the recognition between the N-terminus residue and the ubiquitin ligase, rendering the protein stable and functional. Removal of the caging group by light exposes the lysine residue, thus leading to a short-lived protein. This strategy will allow fast degradation of the target protein without appending an external degron domain.

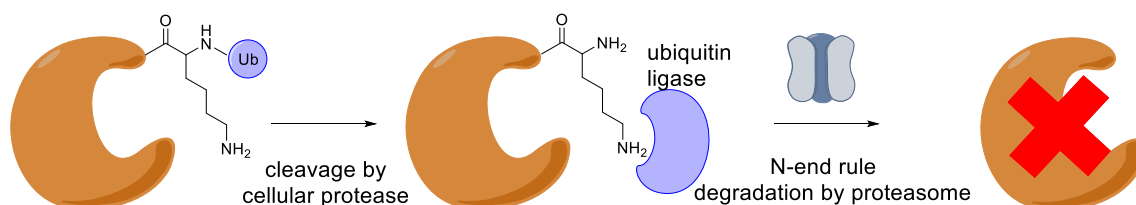


Figure 3.34. Scheme for protein degradation. After protein synthesis, ubiquitin moiety is cleaved by cellular protease, exposing N-terminus residue. If the residue is one of the four amino acids: Arg/Lys/Phe/Leu (Lys was shown here as an example), an ubiquitin ligase will be recruited, resulting in protein ubiquitination, and subsequent degradation.

To confirm the N-end rule with the lysine residue, we first generated plasmids that express EGFP-HA, Ub-K-EGFP-HA, and Ub-A-EGFP-HA. Since the N-terminus alanine leads to a long half-life protein, the Ub-A-EGFP-HA was used as a control. HEK 293T cells were co-transfected with pUb-K-EGFP-HA and pmCherry plasmids (200 ng each) and fluorescence images were taken after 48 h. Although the Ub-K-EGFP-HA and the Ub-A-EGFP-HA group showed lower fluorescence compared to the EGFP-HA group, no difference was observed within these two groups (Fig 3.35). The mCherry expression served as a control for transfection efficiency. This indicated that the Ub-K-EGFP-HA construct was not destabilized in HEK 293T cells.

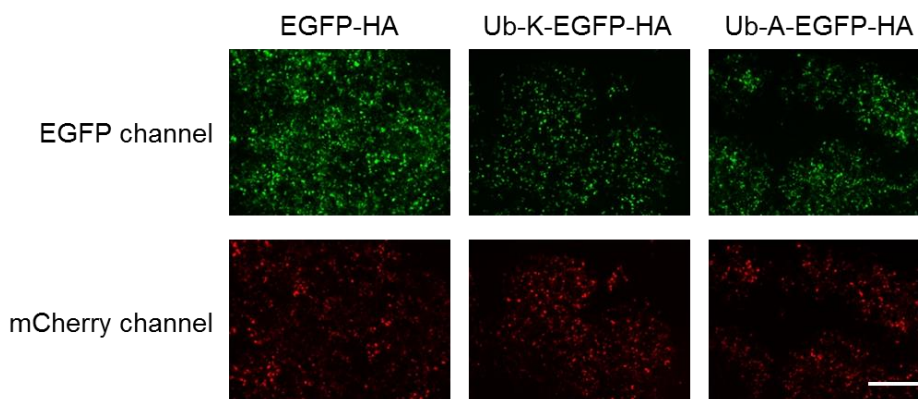


Figure 3.35. Ub-K-EGFP-HA is as stabilized as Ub-A-EGFP-HA in HEK 293T cells. Ub-A-EGFP-HA is used as the stability control. Cells were co-transfected with pUb-K-EGFP-HA and pmCherry (200 ng each). Images were taken at 48 h. Scale bar indicates 100 μ m. EGFP fluorescence indicated the stability of the Ub-K-EGFP-HA protein. mCherry fluorescence was control for transfection efficiency. Although both Ub-K-EGFP-HA and Ub-A-EGFP-HA showed lower fluorescence compared to EGFP-HA, no difference was observed between each other.

Later, a literature analysis revealed that HeLa cells were most frequently used for the ubiquitin-mediated protein degradation experiments.^{150, 155} Although specific reasons were not given, it is likely that an ubiquitin ligase complex is essential for the degradation of Ub-EGFP in HeLa cells.¹⁵⁵ We therefore repeated the same experiment in HeLa cells, except that the mCherry expression control was not used (due to the low transfection efficiency of two plasmids in HeLa cells). We observed lower fluorescence of Ub-K-EGFP-HA compared to that of Ub-A-EGFP-HA (Fig 3.36A). This demonstrated that the Ub-K-EGFP is destabilized in HeLa cells. We further confirmed that the protein level of Ub-K-EGFP-HA was lower than that of Ub-A-EGFP-HA, through the western blot analysis (Fig 3.36B).

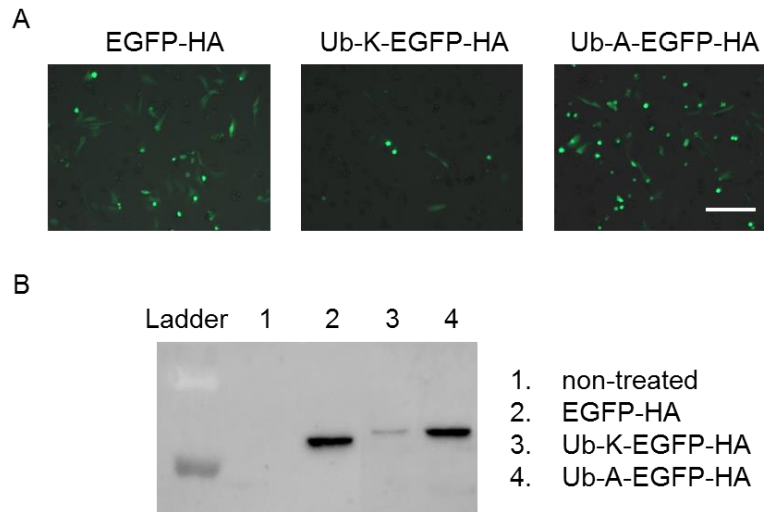


Figure 3.36. The Ub-K-EGFP construct is unstable in HeLa cells. The HeLa cells were transfected with Ub-K-EGFP (1,000 ng). Fluorescence imaging was performed after 48 h, with the 10x objective. The scale bar indicates 200 μ m. The cell lysate was then collected for western blot analysis (with anti-HA antibody). Ubiquitin was cleaved by endogenous protease, leaving a N-terminus lysine. Ub-A-EGFP was a control for stabilized EGFP. (A) Fluorescence image of Ub-K-EGFP-HA indicated that it is destabilized. (B) Western blot analysis confirmed that the protein level of Ub-K-EGFP-HA was lower than that of Ub-A-EGFP-HA.

We then generated the TAG mutation at the lysine position of Ub-K-EGFP-HA. HeLa cells were transfected with pUb-TAG-EGFP-HA and pHCKRS-U6-PyIT (200 ng each) in a 96-well plate. HCK-dependent expression of Ub-HCK-EGFP-HA was confirmed by EGFP fluorescent imaging (Fig 3.37A). At 24 h, cells were irradiated for 30 sec at 365 nm light (DAPI filter set 68, 358–365 nm), and were incubated at 37 °C for 2 h – although longer incubation times or time course experiments should have been conducted. However, no decrease in EGFP fluorescence compared to the $t = 0$ min timepoint was observed for the light-treated group (Fig 3.37B). We noticed that the incorporation efficiency is low in HeLa cells (Fig 3.37A), and cell death was observed with the presence of toxic transfection reagent (Lipofectamine 2000). In the future, we plan to further increase the incorporation efficiency with the one-plasmid system (discussed in

Chapter 2.2). Additional transfection reagents might also be screened for the efficiency and nontoxic transfection of HeLa cells and Western blot analysis of optically triggered protein degradation should be performed.

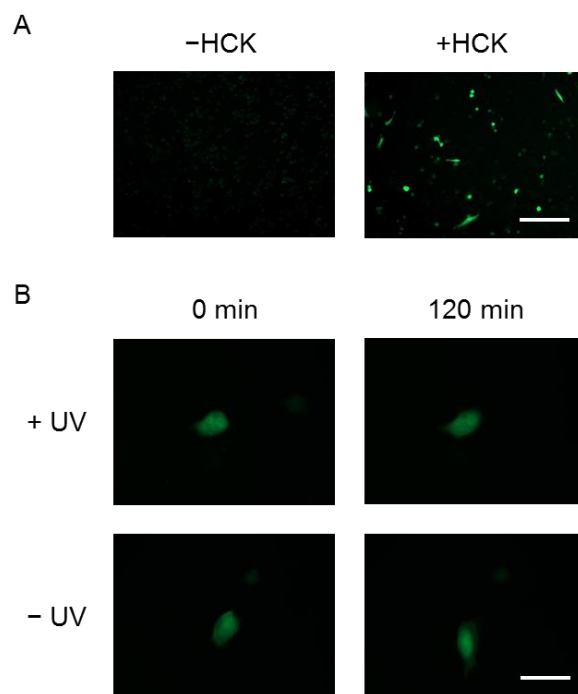


Figure 3.37. Light-deactivation of Ub-HCK-EGFP-HA. (A) HCK (1 mM) was incorporated into Ub-TAG-EGFP-HA. HeLa cells were transfected with pUb-TAG-EGFP-HA and HCKRS-U6-PyIT (200 ng each). Fluorescent images were taken at 24 h. Scale bar indicates 200 μ m. (B) At 24 h, HeLa cells were irradiated for 30 sec with 365 nm light (DAPI filter, 358–365 nm). EGFP fluorescence was acquired at 0 min and 120 min. Scale bar indicates 40 μ m.

In summary, we genetically incorporated 2-nitrophenylalanine and 3-nitrophenylalanine into proteins, and achieved light cleavage of mCherry-TAG-EGFP with 3-nitrophenylalanine. However, the cleavage efficiency was low (~10%). We further designed a ubiquitin-based strategy for protein degradation, through the incorporation of caged lysine at the first position of a target protein. This lysine position was demonstrated to be critical for protein degradation. However, we

have not yet observed the light-induced protein degradation. In the future, we will further optimize the expression system.

Experimental

Incorporation of PHA in *E. coli*

The screening of three synthetases (EV16, EV16-3, EV16-5) and large-scale expression (25 ml) of the mutant sfGFP protein were performed as described in chapter 2.1.

Incorporation of 2-NF and 3-NF in mammalian cells

HEK 293T cells were seeded at ~50,000 cells per well, and were transfected (when at ~80% confluency) with pmCherry-TAG-EGFP-HA and pNFRS-U6-PyIT (constructed by Ji Luo, simultaneously expressing NFRS and PyIT) (200 ng each) in a 96-well plate. The culture medium was supplemented with 2 mM of 2-nitrophenylalanine (2-NF) or 3-nitrophenylalanine (3-NF). Branched polyethylene imine (bPEI, 1.5 μ l, 1 mg/ml) was used as the transfection reagent. After 48 h incubation at 37 °C, cells were imaged on a Zeiss Observer Z1 microscope, with EGFP channel (filter set 38 HE; ex. BP470/40; em. BP525/50) and mCherry channel (filter set 43 HE; ex. BP575/25; em. BP605/70).

Photocleavage assay for 2-NF and 3-NF

HEK 293T cells were co-transfected with pmCherry-TAG-EGFP-HA (1,000 ng) and pNFRS-U6-PyIT (1,000 ng) in the presence or absence of UAA (2 mM or 5 mM) in a six-well plate. Branched polyethylene imine (bPEI, 15 μ l, 1 mg/ml) was used as the transfection reagent. After 48 h, the cells were washed with chilled phosphate-buffer saline (PBS, 1 ml) and lysed in mammalian protein extraction buffer (250 μ l, GE Healthcare). Cell lysate were irradiated in an eppendorf tube with UV light (365 nm, VWR transilluminator) for 10 min. Non-irradiated and irradiated cell lysates were separated by 10% SDS-PAGE (run at 60V for 15 min, and 150 V for 45 min), and were

transferred to a PVDF membrane (GE Healthcare). The membrane was blocked in tris-buffer saline (TBS) with 0.1% Tween 20 and 5% milk for 1 h. The blots were probed with the primary antibody (1:1,000, anti-HA (sc-805), Santa Cruz) overnight at 4 °C, followed by incubation with secondary goat anti-rabbit IgG-HRP antibody (1:20,000, sc-2004, Santa Cruz) for 1 h at room temperature. The blots were further incubated with the SuperSignal West Pico working solution (mixture of the Stable Peroxide Solution and the Luminol/Enhancer Solution, 500 µl each, Thermo Scientific) for 5 min at room temperature. The luminescence signal was detected by ChemiDoc (Chemi Hi Sensitivity setting, exposure time: 10 sec). The un-cleaved product (mCherry-TAG-EGFP-HA, 56 kDa) and cleaved product (EGFP-HA, 28 kDa) were identified on the blot based on the molecular weight.

Plasmid construction

To enable western blot analysis, an HA tag (YPYDVPDYA) was added at the C-terminus of EGFP on the EGFP-N1 plasmid by site-directed mutagenesis (Fwd: 5'-gcatggacgagctgtacaagtaccatacgatgttcagattacgctaaagcgccgcgactctagatcataatcagc-3', Rev: 5'-cttgtagacgctcgccatgccgagagtgatcccgcgccggtc-3', HA tag underlined).

To construct the pUb-K-EGFP-HA, the backbone was generated by digesting pEGFP-HA plasmid (8,000 ng) with EcoRI and BamHI restriction enzymes, followed by gel purification of the band corresponding to the correct length of the backbone (~ 4,500 bp). The Ub gene was amplified from the pET-Ub plasmid (gift from Dr. Ashton Cropp, Fwd: 5'-gcggaattcaccatgcagattttgtgaaaacc-3', Rev: 5'-gcgggatcctgtcgaccaagcttgccgccacgcagacgcagc-3', lysine position underlined), and digested with EcoRI and BamHI restriction enzymes (37 °C, 2 h). The pUb-K-EGFP-HA plasmid was constructed by ligating the Ub fragment into the pEGFP-HA backbone using EcoRI and BamHI restriction sites (T4 ligase, 16 °C, overnight). Site-directed mutagenesis was applied to generate Ub-A-EGFP-HA (Fwd: 5'-

gcgtggcggcccggcggccttggtcgacaggatcc-3', Rev: 5'-caaggccgcccggcgcccgccacgcagacgcagcacc-3', alanine mutation underlined) and Ub-TAG-EGFP-HA (Fwd: 5'-gcgtggcggcctagggcggccttggtcgacaggatcc-3', Rev: 5'-caaggccgcccctagcccgccacgcagacgcagcacc-3', TAG mutation underlined).

Photo-deactivation assay with degron

To confirm the N-end rule by fluorescent imaging, HeLa cells were transfected with pEGFP-HA (or pUb-K-EGFP-HA, or pUb-A-EGFP-HA, 200 ng) in a 96-well plate. Branched polyethylene imine (bPEI, 15 µl, 1 mg/ml) was used as the transfection reagent. After 24 h of incubation at 37 °C, cells were imaged on a Zeiss Observer Z1 microscope, with the EGFP channel (filter set 38 HE; ex. BP470/40; em. BP525/50).

To confirm the N-end rule by Western blot, HeLa cells were transfected with pEGFP-HA (or pUb-K-EGFP-HA, or pUb-A-EGFP-HA, 1,000 ng) in a 6-well plate. Lipofectamine 2000 (Invitrogen, 10 µl per well) was used as the transfection reagent. After 48 h, cells were washed with chilled phosphate-buffer saline (PBS, 1 ml), and lysed in mammalian protein extraction buffer (250 µl, GE Healthcare). The cell lysates were separated by 10% SDS-PAGE (run at 60 V for 15 min, and 150 V for 45 min) and were transferred to a PVDF membrane (GE Healthcare). The membrane was blocked in tris-buffer saline (TBS) with 0.1% Tween 20 and 5% milk for 1 h. The blots were probed with the primary antibody (1:1,000, anti-HA (sc-805) or anti-GAPDH (sc-25778), Santa Cruz Biotechnology) overnight at 4 °C, followed by incubation with secondary goat anti-rabbit IgG-HRP antibody (1:20,000, sc-2004, Santa Cruz Biotechnology) for 1 h at room temperature. After washing, the blots were incubated with SuperSignal West Pico working solution (mixture of the Stable Peroxide Solution and the Luminol/Enhancer Solution, 500 µl each, Thermo Scientific) for 5 min at room temperature. The luminescence signal was detected using a Bio-Rad ChemiDoc (Chemi Hi Sensitivity setting, exposure time: 10 sec).

To incorporate HCK into the Ub-TAG-EGFP-HA construct, HeLa cells were transfected with pUb-TAG-EGFP-HA and pHCKRS-U6-PyIT (200 ng each) in a 96-well plate. Lipofectamine 2000 (Invitrogen, 1 μ l per well) was used as the transfection reagent. The culture medium was supplemented with HCK (1 mM). At 24 h, cells were irradiated for 30 sec with 365 nm light (DAPI filter set 68, 358–365 nm), and were further incubated at 37 °C for 2 h. EGFP fluorescence (filter set 38 HE; ex. BP470/40; em. BP525/50) was acquired at 0 and 120 min.

3.4 Small molecule Activation of Protein Function using Proc and Alloc Groups

Small molecules are important external triggers for the modulation of protein function.¹⁵⁶ Small-molecule-induced protein activation can be achieved through the genetic incorporation of an UAA that contains a small-molecule-removable group. For example, our lab reported the activation of protein function by a phosphine-mediated Staudinger reduction, through the genetic incorporation of an *ortho*-azidobenzyloxycarbonyl amino acid.⁶¹ A similar strategy was also pursued for aliphatic azido lysines (discussed in chapter 2.1). Furthermore, transition metals can also be used as external triggers.¹⁵⁷ A palladium-mediated deprotection reaction was applied to activate a protein in living cells, through the genetic incorporation of a proc (propargyl carbamate) lysine.^{157b} Through a screening of several palladium catalysts, [PdCl(allyl)]₂ was found as the most efficient catalyst for proc group deprotection (82% cleavage on proc lysine, 37 °C, 8 h; 90% cleavage on GFP-N149-proc lysine, 25 °C, 1 h). The same catalyst showed much less efficiency for alloc group deprotection (22% cleavage on alloc lysine, 37 °C, 8 h). Pd(OAc)₂ was a much less efficient catalyst for proc deprotection (4% cleavage on GFP-N149-proc lysine, 25 °C, 1 h). The activation kinetics for both phosphine (~200 min to reach completion)⁶¹ and palladium (180-240 min to reach completion)^{157b} were slow in live cell experiments. Significantly faster chemical protein activation was achieved through an inverse-electron-demand Diels-Alder reaction. Through the genetic incorporation of a cyclooctene-containing TCO lysine, protein function was inhibited until tetrazine-induced decaging lead to rescue of 80% of protein activity within 15 min.¹⁵⁸ Recently, the Koide lab at Pitt discovered a combination of catalysts (tri(2-furyl)phosphine (75 μM) and PdCl₂ (100 μM)) that allows fast deprotection of the proc group (unpublished data). Therefore, we were interested in applying this strategy to the small-molecule triggered protein activation. Alloc and proc lysine have previously been incorporated into proteins by the wild-type PylRS, and consequently allowed us to directly test the protein activation without synthetase screening.^{18b}

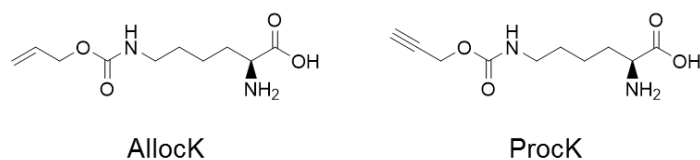


Figure 3.38. Structure of alloc lysine and proc lysine.

We first proposed that a small-molecule activatable firefly luciferase could be generated. Previously, incorporation of a photocaged lysine at a critical lysine residue (K206) blocked the firefly luciferase activity, which was later restored upon UV irradiation.⁶⁰ We therefore hypothesized that the incorporation of an alloc or proc lysine at the same position will have a similar effect. The activation of the firefly luciferase was tested by removing the protecting group using the new catalysts. To this end, HEK 293T cells were transfected with pFluc-K206TAG and pWTRS-U6-PyIT plasmids (200 ng each). The culture medium was supplemented with either alloc lysine (1 mM) or proc lysine (1 mM). After 24 h, the culture medium was changed to the medium containing Pd(OAc)₂ (50 μM), or [PdCl(allyl)]₂ (50 μM), or tri(2-furyl)phosphine (75 μM)/PdCl₂ (100 μM), and the cells were further incubated at 37 °C for 3 h. The luciferase assay was then performed on live cells. Unexpectedly, we observed the firefly luciferase activity either with or without the catalyst (Fig 3.39). The result indicated that alloc or proc handle was not able to block the luciferase activity. Interestingly, in the study where the proc lysine was incorporated for Pd-triggered activation, the firefly luciferase was not applied as a model protein,^{157b} even though the same group reported a photoactivable luciferase before.¹⁵⁹ We speculated that the alloc or proc handle was too small to block the active site of luciferase.

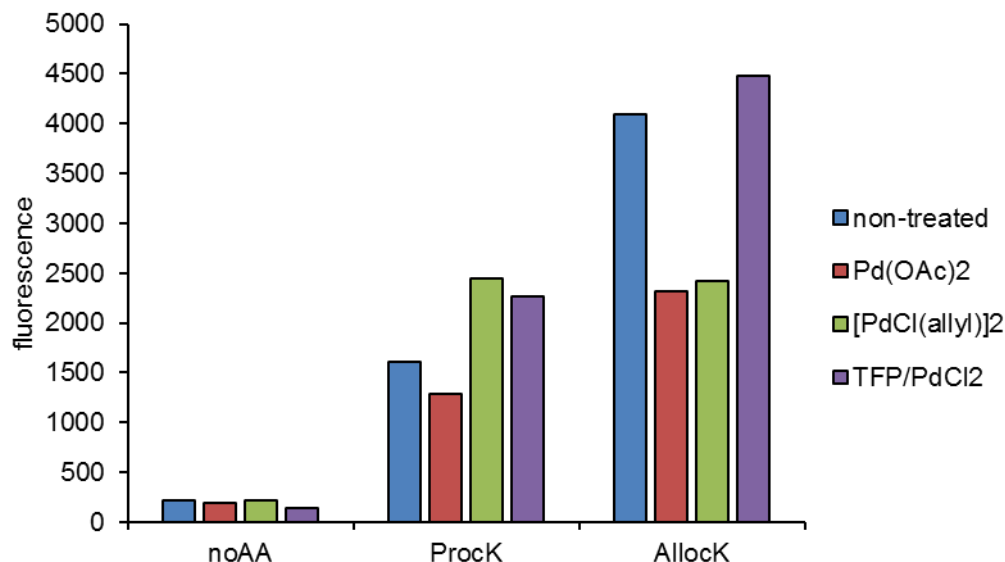


Figure 3.39. Small molecule activation of Fluc-K206TAG with different catalysts. HEK 293T cells were transfected with pFluc-K206TAG and pWTRS-U6-PyIT plasmids. The culture medium was supplemented with either alloc lysine (1 mM) or proc lysine (1 mM). After 24 h, the culture medium was changed to the medium containing Pd(OAc)₂ (50 μM), or [PdCl(allyl)]₂ (50 μM), or tri(2-furyl)phosphine (75 μM)/PdCl₂ (100 μM), and the cells were further incubated at 37 °C for 3 h. The luciferase assay was then performed on live cells.

We then tested another protein target, Cre recombinase. Incorporation of a photocaged lysine into Cre recombinase at a critical K201 position allowed for light-activation of recombinase activity.¹⁶⁰ We ask if this system could be adapted to small-molecule deprotection. We employed the Stoplight plasmid as a reporter, which encodes a DsRed and a transcription termination region, both flanked by loxP sites and located upstream of an EGFP gene.¹⁶¹ Cre recombinase leads to the activation of EGFP expression through excision of the DsRed-terminator cassette. We first confirmed that the wild-type Cre resulted in the expression of EGFP, but not DsRed in mammalian cells (Fig 3.40A). The incorporation of a photocaged lysine (HCK, 1 mM) into the K201 position of Cre allowed for light-activation of the Cre recombinase, although the EGFP expression was not as high as that of wild-type Cre (Fig 3.40B).

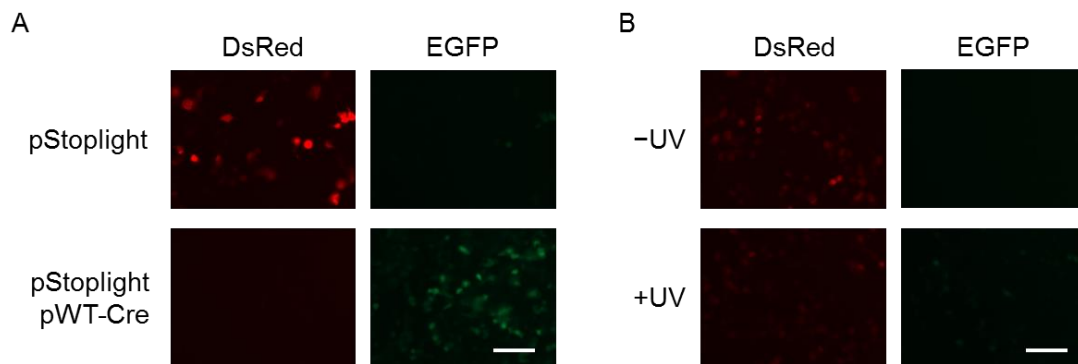


Figure 3.40. Cre recombinase assay and light-activation through HCK incorporation. (A) Wild-type Cre is active in mammalian cells. HEK 293T cells were transfected with the indicated plasmid(s) (200 ng each). Images were taken after 24 h. The expression of WT-Cre led to activation of EGFP expression through excision of the DsRed-terminator cassette. (B) Light-activation of the photocaged Cre. HEK 293T cells were transfected with pStoplight, pCre-K201TAG, pHCKRS-U6-PyIT plasmids (100 ng each). The culture medium was supplemented with HCK (1 mM). After 24 h, the cells were irradiated with UV light (365 nm, 2 min), and were further incubated at 37 °C for 24 h. Images were taken at 48 h. Scale bar indicates 100 μ m.

After establishing the Cre recombinase assay, we incorporated alloc lysine or proc lysine into the K201 position of Cre recombinase; however, protein expression, while expected, was not confirmed by Western blot. Catalysts (tri(2-furyl)phosphine (75 μ M)/PdCl₂ (100 μ M) for alloc lysine, [PdCl(allyl)]₂ (50 μ M) for proc lysine) were added at 24 h and the cells were incubated for another 24 h before imaging. However, despite repeated attempts, we did not observe an increase in EGFP fluorescence with the addition of catalysts (Fig 3.41).

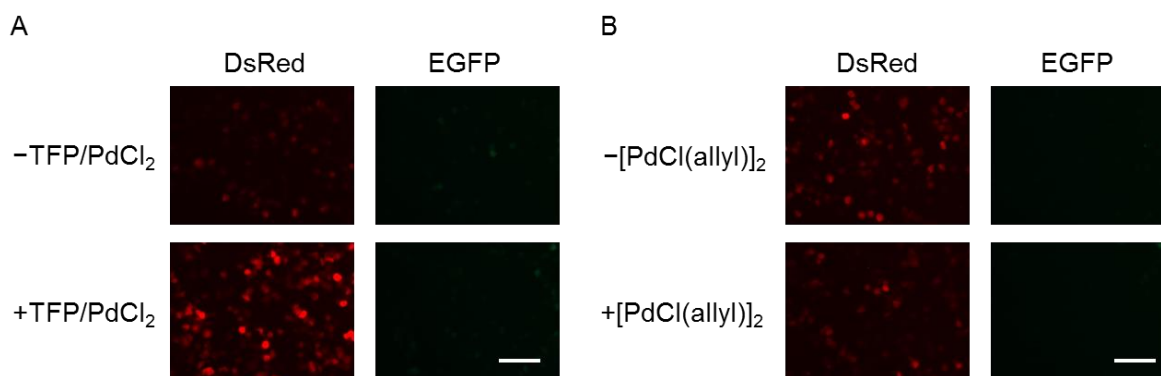


Figure 3.41. Activation of the Cre recombinase through Pd catalyst. HEK 293T cells were transfected with pStoplight, pCre-K201TAG, and pWTRS-U6-PylT plasmids (100 ng each). The culture medium was supplemented with alloc lysine (1 mM, in A) or prop lysine (1 mM, in B). After 24 h, the culture medium was replaced with the medium containing tri(2-furyl)phosphine (75 μ M)/PdCl₂ (100 μ M) (in A) or [PdCl(allyl)]₂ (50 μ M) (in B). The cells were further incubated at 37 °C for 24 h. Images were taken at 48 h. Scale bar indicates 100 μ m.

Lastly, we tested small-molecule triggered nucleus translocation signaling. Previously, a photocaged lysine was incorporated into a critical lysine position (K29) of the nuclear localization signal (NLS) of the transcription factor SATB1.¹⁶² The caging group blocked nuclear import of the protein until its removal by light. EGFP-SATB1-K29TAG-mCherry was used as a reporter, where the EGFP was localized both in cytoplasm and nucleus, and the mCherry was translocated from cytoplasm to nucleus with a functional SATB1 NLS.¹⁶² We first confirmed the incorporation of alloc lysine into EGFP-SATB1-K29TAG-mCherry (Fig 3.42A). With no treatment, the mCherry fluorescence located in the cytoplasm, indicating an inactive SATB1 NLS function blocked by the alloc group (Fig 3.42B). However, with the addition of tri(2-furyl)phosphine (75 μ M)/PdCl₂ (100 μ M), no nuclear translocation was observed up to 12 h (Fig 3.42B). We speculated that the catalyst complex was not able to deprotect the alloc group under these conditions.

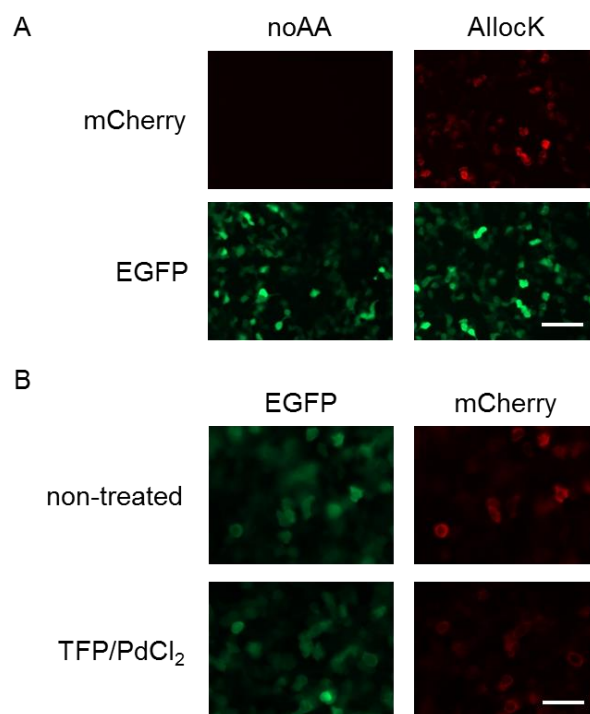


Figure 3.42. Small molecule activation of protein translocation. HEK 293T cells were transfected with pEGFP-SATB1-K29TAG-mCherry and pWTRS-U6-PylT (200 ng each). The culture medium was supplemented with alloc lysine (1 mM). After 24 h, the culture medium were replaced with the medium containing tri(2-furyl)phosphine (75 μ M)/PdCl₂ (100 μ M). The cells were further incubated at 37 °C for up to 12 h. Images were taken at 36 h. (A) The mCherry fluorescence confirmed the incorporation of the alloc lysine into the EGFP-SATB1-K29TAG-mCherry. EGFP was used as a control for transfection efficiency. Scale bar indicates 100 μ m. (B) No nucleus translocation of mCherry was observed upon TFP/PdCl₂ treatment. The scale bar indicates 50 μ m.

In summary, we tested small-molecule activation of protein function through the genetic incorporation of alloc or proc lysine. We attempted to achieve alloc or proc deprotection with a combination of catalysts (phosphine and Pd). However, we found that the protecting group was too small to block firefly luciferase activity, and that the current catalyst conditions did not activate activity of Cre recombinase or SATB1 translocation. Future work will focus on screening catalyst combinations and concentrations in the context of mammalian cells using the small molecule

fluorescence sensor.¹⁶³ Optimized condition will further be applied for controlling Cre recombinase activity or nuclear localization by small molecules.

Experimental

Small-molecule activation of luciferase activity

HEK 293T cells were transfected with the pFluc-K206TAG and pWTRS-U6-PyIT plasmids (200 ng each) in a 96-well plate. The culture medium was supplemented with either alloc lysine (1 mM) or proc lysine (1 mM). Branched polyethylene imine (bPEI, 1.5 μ l, 1 mg/ml) was used as the transfection reagent. After 24 h, the culture medium was changed to the medium containing Pd(OAc)₂ (50 μ M), or [PdCl(allyl)]₂ (50 μ M), or tri(2-furyl)phosphine (75 μ M)/ PdCl₂ (100 μ M), and the cells were further incubated at 37 °C for another 3 h. The luciferase assay was then performed on live cells. Half of the culture medium (100 μ l) was removed from each well, and BrightGlo luciferase assay reagent (100 μ l, Promega) was added, followed by vigorous shaking on an orbital shaker (1 min). Luminescence was then read on a Tecan M1000 PRO microplate reader.

Small-molecule activation of Cre recombinase activity

To confirm the light activation of Cre recombinase activity in mammalian cells, HEK 293T cells were transfected with pStoplight, pCre-K201TAG, and pHCKRS-U6-PyIT plasmids (100 ng each, provided by Ji Luo) as previously described.¹⁶⁰ The culture medium was supplemented with HCK (1 mM). Branched polyethylene imine (bPEI, 1.5 μ l, 1 mg/ml) was used as the transfection reagent. After 24 h, the cells were irradiated with the UV light (365 nm, VWR) for 2 min, and were further incubated at 37 °C for 24 h. Cells were imaged on a Zeiss Observer Z1 microscope, with EGFP channel (filter set 38 HE; ex. BP470/40; em. BP525/50) and mCherry channel (filter set 43 HE; ex. BP575/25; em. BP605/70).

To activate Cre recombinase with small molecules, HEK 293T cells were transfected with pStoplight, pCre-K201TAG, and pWTRS-U6-PyIT plasmids (100 ng each). Branched

polyethylene imine (bPEI, 1.5 μ l, 1 mg/ml) was used as the transfection reagent. The culture medium was supplemented with alloc lysine (1 mM) or proc lysine (1 mM). After 24 h, the culture medium was replaced with medium containing tri(2-furyl)phosphine (75 μ M)/PdCl₂ (100 μ M) or [PdCl(allyl)]₂ (50 μ M). The cells were further incubated at 37 °C for 24 h. Cells were imaged on a Zeiss Observer Z1 microscope, with EGFP channel (filter set 38 HE; ex. BP470/40; em. BP525/50) and mCherry channel (filter set 43 HE; ex. BP575/25; em. BP605/70).

Small-molecule triggered nucleus translocation

HEK 293T cells were transfected with pEGFP-SATB1-K29TAG-mCherry¹⁶² and pWTRS-U6-PylT plasmids (200 ng each). Branched polyethylene imine (bPEI, 1.5 μ l, 1 mg/ml) was used as the transfection reagent. The culture medium was supplemented with alloc lysine (1 mM). After 24 h, the culture medium was replaced with the medium containing tri(2-furyl)phosphine (75 μ M)/PdCl₂ (100 μ M). The cells were further incubated at 37 °C for 12 h. At 36 h, cells were imaged on a Zeiss Observer Z1 microscope, with EGFP channel (filter set 38 HE; ex. BP470/40; em. BP525/50) and mCherry channel (filter set 43 HE; ex. BP575/25; em. BP605/70).

4.0 Installation of an Aminoxy Functional Group for Protein Labeling

Site-specific protein conjugation reactions are essential for biophysical and functional studies of proteins.¹⁶⁴ Unnatural amino acid (UAA) mutagenesis provides a versatile approach to achieve this goal, by site-specifically installing a biorthogonal functional group that enables protein conjugation chemistries with high selectivity.¹⁶⁵ More than thirty UAAs have been genetically incorporated in *E. coli* and mammalian cells for protein labeling purpose,^{165a} including electrophiles (ketones¹⁶⁶ and aldehydes¹⁶⁷), azides,^{18a, 18b, 168} alkynes,^{18b, 35, 168-169} alkenes,^{36, 170} and tetrazenes.^{18d, 35} For example, Cu(I)-catalyzed azide-alkyne cycloaddition has been widely used for *in vitro* labeling of biomolecules (Fig 4.1A).^{165a} Recently, inverse electron-demand Diels-Alder cycloaddition reactions were used for fast protein labeling in live cells (Fig 4.1B).^{18d} Electrophiles were also genetically incorporated for oxime ligation with hydroxyamines or hydrazines (Fig 4.1C). However, the genetic encoding of a nucleophile had been lacking (Fig 4.1D). Here, we show the site-specific incorporation of (S)-2-amino-5-(aminoxy)pentanoic acid (AOK) as a nucleophilic tag into proteins in *E. coli*. The installed aminoxy handle enables rapid and mild oxime ligations that have previously been utilized in single molecule fluorescence resonance energy transfer (smFRET) studies,¹⁷¹ site-specific protein spin-labeling,¹⁷² fluorescent labeling of membrane proteins,¹⁷³ and production of antibody-drugs.¹⁷⁴ The installation of a nucleophile onto proteins through unnatural amino acid mutagenesis adds a new chemical reactivity to the expanding toolbox of site-specific protein modifications.¹⁷⁵

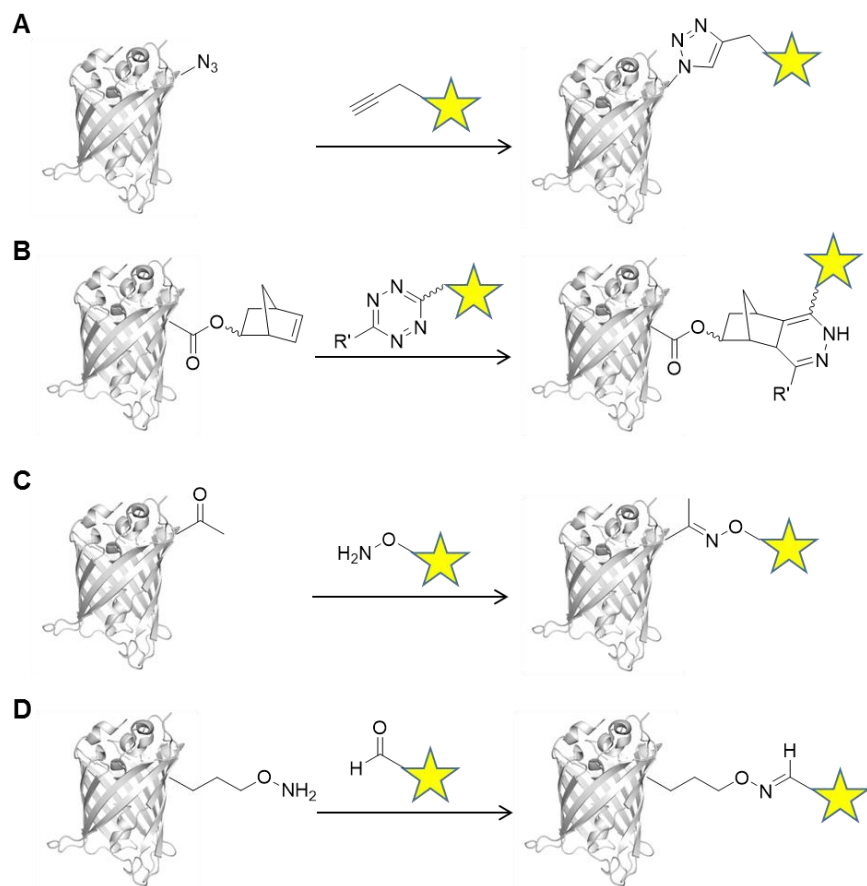


Figure 4.1. Site-specific protein conjugation using unnatural amino acids. (A) Cu(I)-catalyzed cycloaddition. (B) Inverse electron-demand Diels-Alder cycloaddition. (C) Oxime ligation with a genetically encoded ketone. (D) Structure used in this study. Oxime ligation, with a genetically encoded aminoxy group.

Previously an aminoxy functional group was generated by reacting bis-(oxyamine) with thioester group on C-terminus of a protein.¹⁷⁶ Although very efficient, this method is limited to modify proteins only at the C-terminus. It is therefore not applicable when labeling needs to be achieved at other positions of a protein. Recently an aminoxy side-chain was introduced on a miniprotein through solid-phase peptide synthesis.¹⁷⁷ Oxime ligation was utilized to form protein dimers, which showed promising therapeutic use to suppress tumor progression. However, this synthetic strategy is limited to miniproteins with less than fifty amino acids. We hypothesized that

the incorporation of an UAA with an aminooxy functional group would circumvent above limitations. First, since the incorporation is residue-specific, labeling could be achieved at any positions of a protein. Second, protein of any length could be readily expressed in *E. coli*, greatly simplifying the process and increasing the yield. Besides, several nucleophilic catalysts can dramatically accelerate the oxime ligation,¹⁷⁸ with the rate constant of $10^2 \text{ M}^{-1} \text{ S}^{-1}$. Therefore, the incorporation of an aminooxy functional group will provide a simple and robust way to label proteins through oxime ligation (Fig 4.1D). As we worked on this project, the Virdee group reported the genetic incorporation of an UAA with an aminooxy group,¹⁷⁹ with the same chemical structure as we proposed (see below). They generated several nonhydrolyzable ubiquitin conjugates (diubiquitin, polymeric ubiquitin chain, ubiquitinated SUMO) by oxime ligation. They further confirmed that these ubiquitin conjugates function as inhibitors of deubiquitinating enzymes, thus providing a new tool to study deubiquitinating enzymes.¹⁷⁹

We selected (S)-2-amino-5-(aminooxy)pentanoic acid (AOK) as a target, since genetic code expansion of many lysine analogues has been successfully achieved with an engineered pyrrolysyl-tRNA/tRNA synthetase (PylRS) system.¹⁶ However, due to the structural similarity between AOK and lysine, we did not expect to discover a PylRS that could recognize AOK, but discriminate lysine. Instead, we first disguised AOK with a protecting group, thereby enabling differentiation from lysine, followed by incorporation and protecting group removal, liberating a free aminooxy group on proteins. Here we utilized *tert*-butyl-oxycarbonyl (Boc) group and photo-caging group as the protecting group respectively, and designed two unnatural amino acids, BAOK and PAOK, as the disguised form of AOK. This strategy has been previously applied to install near-natural amino acids on proteins. For example, incorporation of fluorotyrosine with a photo-caging group,⁶⁹ and *N*^ε-methyl-L-lysine with a *tert*-butyl-oxycarbonyl (Boc) protection group,¹⁸⁰ followed by deprotection step, generates fluorotyrosine and *N*^ε-methyl-L-lysine on proteins respectively.

Since *N*^ε-tert-butyl-oxycarbonyl-L-lysine (BocK) is a known substrate for both wild-type PylRS and for a PylRS with a single mutation, Y349F,^{18a} we first explored whether BAOK is recognized by the same synthetases. The incorporation of BAOK is assessed by the expression of superfolder green fluorescent protein (sfGFP) containing an amber stop codon TAG at Tyr 151 position. Two plasmids, pBAD-sfGFPY151TAG-PylIT and pBK-PylRS, were co-transformed into Top10 cells for sfGFP expression. As expected, both the wild-type PylRS and the Y349F mutant could incorporate BocK in excellent yields of 23 mg/L and 20 mg/L, respectively. Interestingly, only PylRS-Y349F was able to incorporate BAOK, with a yield of 9 mg/L (Fig 4.2B). When using the wild-type PylRS, no sfGFP expression was observed in the presence or absence of BAOK, suggesting that BAOK is not a substrate for this PylRS. The Y349F mutation is proposed to increase the aminoacylation activity of PylRS,^{18a} and it has been found in several PylRS mutants that accept UAAs of diverse structures.^{19b, 21, 169d, 181} Electrospray Ionization Mass Spectrometry (ESI-MS) confirmed the expression of sfGFP-BAOK (expected mass: 28329.8 Da, observed mass: 28329.9 Da). The Boc group on sfGFP-BAOK was then deprotected through treatment with 50% trifluoroacetic acid (TFA) at 37 °C for 3 h as previously described.¹⁸² The ESI-MS of the deprotection product showed that the previous peak for sfGFP-BAOK disappeared and instead a single peak with a mass 22 Da higher than sfGFP containing AOK at position 151 appeared (expected mass: 28229.7 Da, observed mass: 28252.0 Da). As we performed our experiment, another group reported the incorporation of BAOK by a slightly different PylRS mutant (with mutation Y349W).¹⁷⁹ They showed that the incorporation efficiency of BAOK (by Y349W mutant) was similar to that of BocK (by wild-type PylRS), and confirmed the deprotection product (by 60% TFA, 23 °C, 2 h) with ESI-MS. As our result suggested that incorporation efficiency of BAOK (by Y349F mutant) was less than that of BocK (by both wild-type PylRS and Y349F mutant), it will be worthwhile to use Y349W PylRS mutant for future experiments.

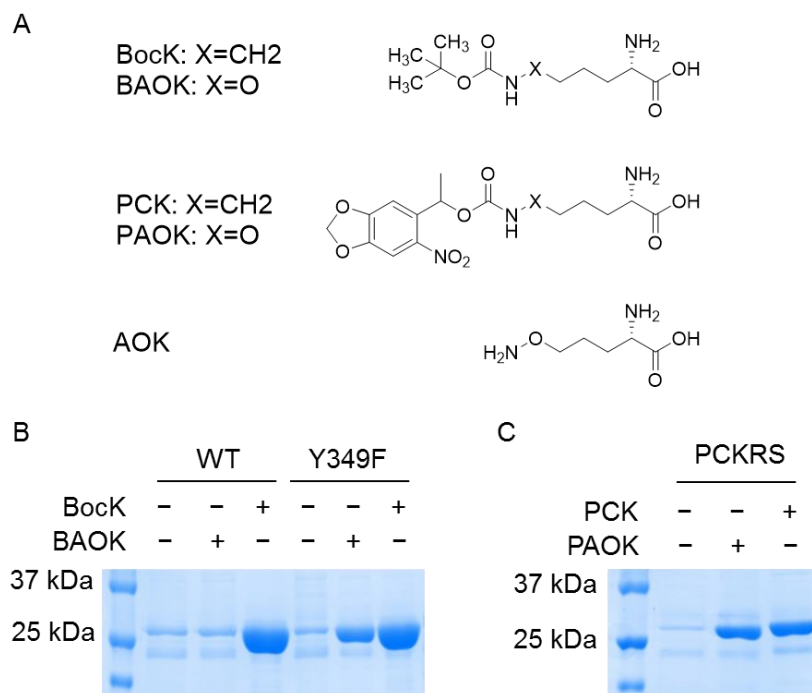


Figure 4.2. Genetic incorporation of Boc-aminooxy and photocaged-aminoxy amino acid. (A) Structure of the molecules studied in this research. (B) Genetic incorporation of BAOK by Y349F mutant PylRS. (C) Genetic incorporation of PAOK by PCKRS.

We then asked if the acid-catalyzed deprotection step could be avoided by applying a light-removable protecting group on AOK. We previously incorporated the photocaged lysine PCK into proteins in bacterial and mammalian cells by applying a *M. barkeri* PylRS harboring the mutations M241F, A267S, Y271C, and L274M – termed PCKRS.^{19d} The same synthetase also accepts a caged cysteine and a caged homocysteine as substrates,¹¹⁶ thereby showing broader substrate specificities. We therefore tested whether PAOK is also a substrate for PCKRS. We used sfGFP-Y151TAG as a reporter, as described above, and gratifyingly, PAOK showed similar incorporation efficiency as PCK, with a yield of 3 mg/L (Fig 4.2C). The ESI-MS spectrum revealed a peak corresponding to sfGFP-PAOK (expected mass: 28466.9 Da, observed mass: 28466.8 Da). Yet another equivalent peak is also present in the spectrum (observed: 28255.7 Da), corresponding to partial decaging product of sfGFP-PAOK. To verify this hypothesis, we fully

removed the photocaging group through UV exposure (365 nm, 10 min), and subsequent ESI-MS analysis revealed a single peak that is 24 Da higher than sfGFP-AOK (expected mass: 28229.7 Da, observed mass: 28253.4 Da), a molecular weight shift that was also observed in the spectrum of the deprotected sfGFP-BAOK. This result suggested that the photocaging group was completely removed upon UV irradiation. Although the exact cause of this molecular weight shift was not fully understood, we speculated that aminooxy might react with trace amount of acetaldehyde during MS analysis. Similar molecular weight shift has been observed before, as aminooxy shows high reactivity with small amount of carbonyl-containing compounds.¹⁸³

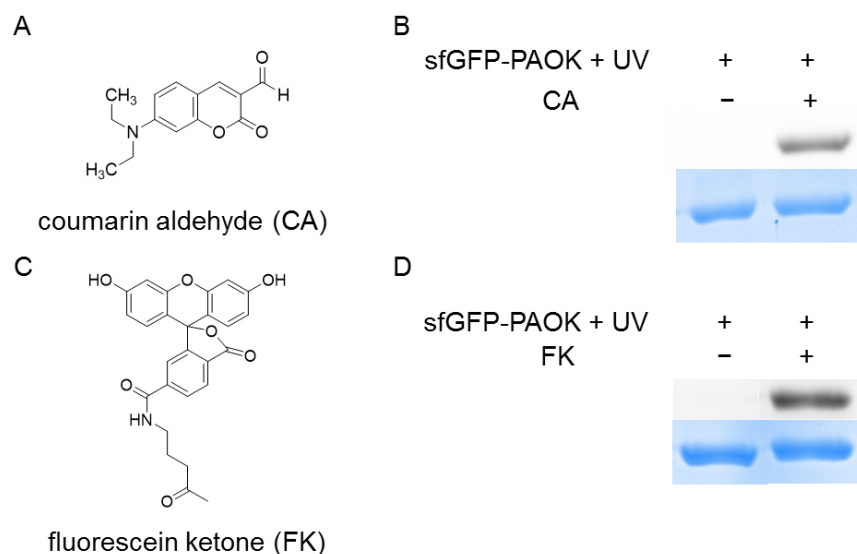


Figure 4.3. Site-specific protein labeling with genetically encoded aminooxy group. (A) Structure of coumarin aldehyde (CA). (B) Labeling of sfGFP-aminooxy with coumarin aldehyde. (C) Structure of fluorescein ketone (FK). (D) Labeling of sfGFP-aminooxy with fluorescein ketone. Fluorescence was analyzed in-gel by ChemiDoc (Alexa 488 setting for 6, Fluorescein setting for 7, for both settings, ex: 470 nm, em: 530 nm). The same gel was further stained with Coomassie Brilliant Blue and imaged by ChemiDoc (Coomassie Blue setting).

We then tested whether site-specific protein labeling could be achieved through genetically encoded aminooxy group. To this end, decaged sfGFP-PAOK (5 μ M), the coumarin-

aldehyde (abbreviated as CA, 1 mM, Sigma), and aniline (100 mM) were incubated in NaOAc buffer (pH 5.5). Aniline was used as a nucleophilic catalyst for the oxime ligation.^{178a} In-gel fluorescence showed that sfGFP with aminooxy group could be labeled with CA (Fig 4.3B). The fluorescence reached its peak within 10 min at room temperature (Fig 4.4). The reaction rate is comparable to that of the previously reported aldehyde/aminooxy reaction.¹⁶⁷ Virdee group did not report conjugation kinetics, but a different condition was used for diubiquitin generation (pH 6, 37 °C for 1 h).¹⁷⁹ As expected, protein with lysine at position (sfGFP-Y151K) did not show any labeling under the same conditions, suggesting that CA was selectively reacting with aminooxy group. The ESI-MS spectrum revealed a single peak that matched the oxime ligation product of sfGFP-AOK with CA (expected mass: 28457.0 Da, observed mass: 28457.4 Da).



Figure 4.4. Time-course study of sfGFP-aminooxy labeling with coumarin aldehyde. Fluorescence was analyzed in-gel by ChemiDoc (Alexa 488 setting, ex: 470 nm, em: 530 nm). The same gel was further stained with Coomassie Brilliant Blue and imaged by ChemiDoc (Coomassie Blue setting).

To further explore the reactivity of the aminooxy group, we tested protein labeling with the fluorescein-ketone (abbreviated as FK, synthesized by Subhas Samanta). In-gel fluorescence showed that FK could be successfully labeled on sfGFP (Fig 4.3D). The ESI-MS spectrum revealed a single peak that matched the oxime ligation product of sfGFP-AOK with FK (expected mass: 28671.16 Da, observed mass: 28670.23 Da). Protein with lysine at position (sfGFP-Y151K) did not show any labeling under the same conditions, suggesting that FK was selectively reacting

with aminoxy group, but not lysine (Fig 4.5). A time-course study revealed that the fluorescence intensity was highest after 24 h (Fig 4.5). The slower reaction rate of the ketone FK compared to the aldehyde CA is not surprising.¹⁶⁷



Figure 4.5. Time-course study of sfGFP-aminoxy labeling with fluorescein ketone. Fluorescence was analyzed in-gel by ChemiDoc (Fluorescein setting, ex: 470 nm, em: 530 nm). The same gel was further stained with Coomassie Brilliant Blue and imaged by ChemiDoc (Coomassie Blue setting).

To demonstrate that this labeling strategy is applicable to other proteins, we also incorporated AOK into the fourth position of myoglobin. Facile labeling of decaged myo-PAOK with CA was observed (Fig 4.6A), and was further confirmed by the ESI-MS analysis (expected mass: 18625.4 Da, observed mass: 18625.4 Da), demonstrating the robustness of the developed bioconjugation methodology. Similarly, oxime ligation with FK was detected as well (Fig 4.6B).

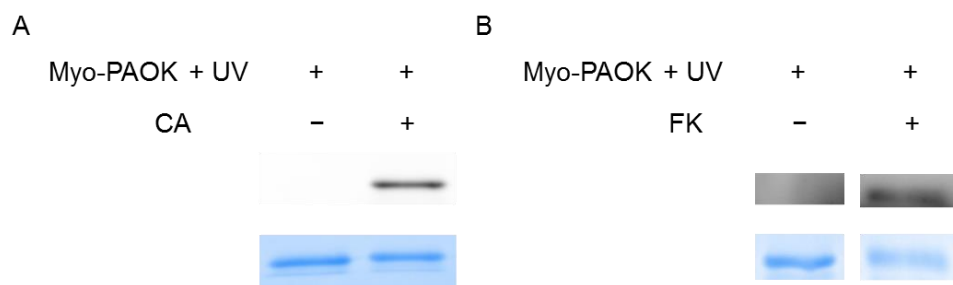


Figure 4.6. Labeling of myoglobin with (A) coumarin aldehyde and (B) fluorescein ketone. Fluorescence was analyzed in-gel by ChemiDoc (Alexa 488 setting for 6, Fluorescein setting for 7, for both settings, ex: 470 nm, em: 530 nm). The same gel was further stained with Coomassie Brilliant Blue and imaged by ChemiDoc (Coomassie Blue setting).

We then sought to conduct protein conjugation with large molecules. PEGylation is a frequent protein modification to improve efficacy and stability of therapeutic proteins.¹⁸⁴ Reaction of decaged sfGFP-PAOK with PEG-5k (Sigma, 37 °C, 2 days) resulted in a band that corresponds to PEGylated sfGFP on SDS-PAGE (Fig 4.7B). However, the reaction showed only 21% PEGylation based on image quantification (ImageJ). The reaction yield is lower than PEGylation through an acrylamide unnatural amino acid (~ 50%).³⁶ We then sought to generate a diubiquitin with oxime linkage through the reaction of decaged Ub-PAOK and Ub-aldehyde (Enzo Life Sciences). The Ub-aldehyde has the C-terminus glycine carboxyl modified into an aldehyde. We observed 24% of the diubiquitin product (Fig 4.7C). As we performed this experiment, Virdee group generated the diubiquitin-oxime conjugate using the same strategy, and tested that it is resistant to hydrolysis of deubiquitylating enzymes (DUBs).¹⁷⁹

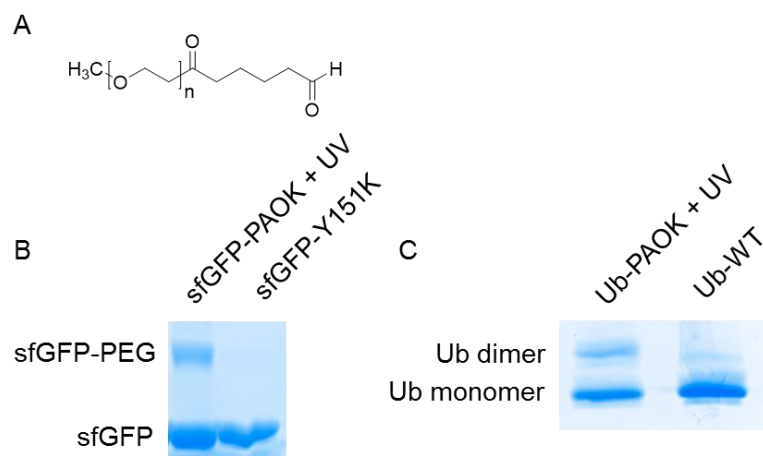


Figure 4.7. sfGFP PEGylation and ubiquitin conjugation through oxime ligation. (A) Structure of PEG-5k used for conjugation. (B) sfGFP-aminooxy reacted with PEG-aldehyde, generating PEGylated sfGFP. (C) Ub-aminooxy reacted with Ub-aldehyde, generating diubiquitin with oxime linkage. The gel was stained with Coomassie Brilliant Blue and imaged by ChemiDoc (Coomassie Blue setting).

We also explored if PAOK could be genetically incorporated in mammalian cells. Treatment of HEK293T cells with 1 or 1.5 mM of PAOK did not show apparent toxicity. We used mCherry-TAG-EGFP-HA as a reporter, where the TAG amber codon resides between mCherry and EGFP. Cells will only show EGFP fluorescence when the unnatural amino acid is successfully incorporated, and mCherry serves as a control for transfection efficiency. We observed EGFP fluorescence only in the presence of PAOK, indicating successful incorporation of PAOK in mammalian cells (Fig 4.8B). The western blotting analysis further confirmed PAOK-dependent expression of mCherry-TAG-EGFP-HA protein (Fig 4.8D).

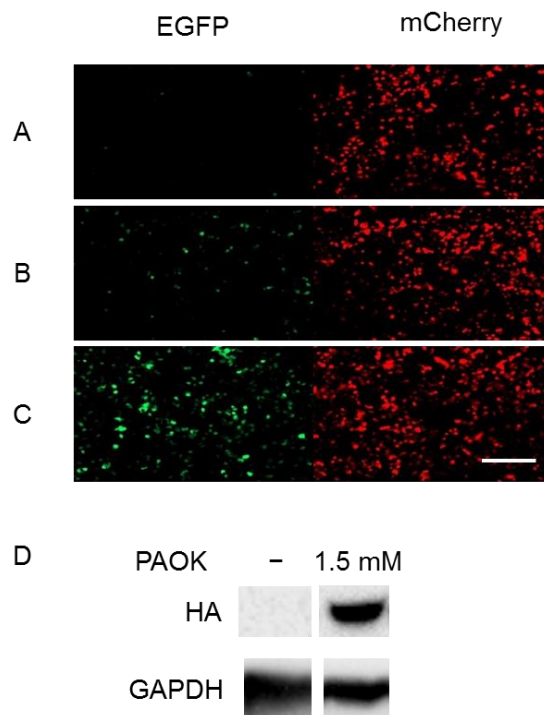


Figure 4.8. Genetic incorporation of PAOK into mCherry-TAG-EGFP in mammalian cells. Cells were transfected with pmCherry-TAG-EGFP-HA and pPCKRS-4CMV-U6-PylT plasmids, and were further treated with (A) no UAA, (B) 1 mM of PAOK, (C) 0.25 mM of PCK. Images were taken at 24 h, with 10x objective. Scale bar indicates 200 μ m. (D) Incorporation of PAOK into EGFP-TAG-mCherry-HA, analyzed by western blotting. GAPDH was used as loading control.

We then tried to achieve cell labeling using oxime ligation. We first incorporated PAOK into mCherry-actin at lysine 144 position, and observed PAOK-dependent expression of mCherry-actin (Fig 4.9A). The mCherry expression showed expected actin filament in both HEK293T cells (Fig 4.9B) and HeLa cells (Fig 4.9C). At 24 h, cells were fixed with 3.7% formaldehyde and permeabilized with 0.5% Triton X100. Fixed cells were further irradiated with UV light, and reacted with CA (coumarin-aldehyde). After reaction, cells were washed with PBS twice and imaged. However, while cells expressed mCherry-actin in the presence of PAOK, they were not labeled with CA after oxime ligation (Fig 4.9B and Fig 4.9C). We speculate that the aminooxy group might

react with free carbohydrates in cells,¹⁸⁵ thus making the labeling with an aldehyde dye inefficient. Besides, the EGFP channel (ex. BP470/40; em. BP525/50) is not ideal for imaging of CA. Other fluorescent dyes that show higher brightness under these imaging conditions might be pursued.

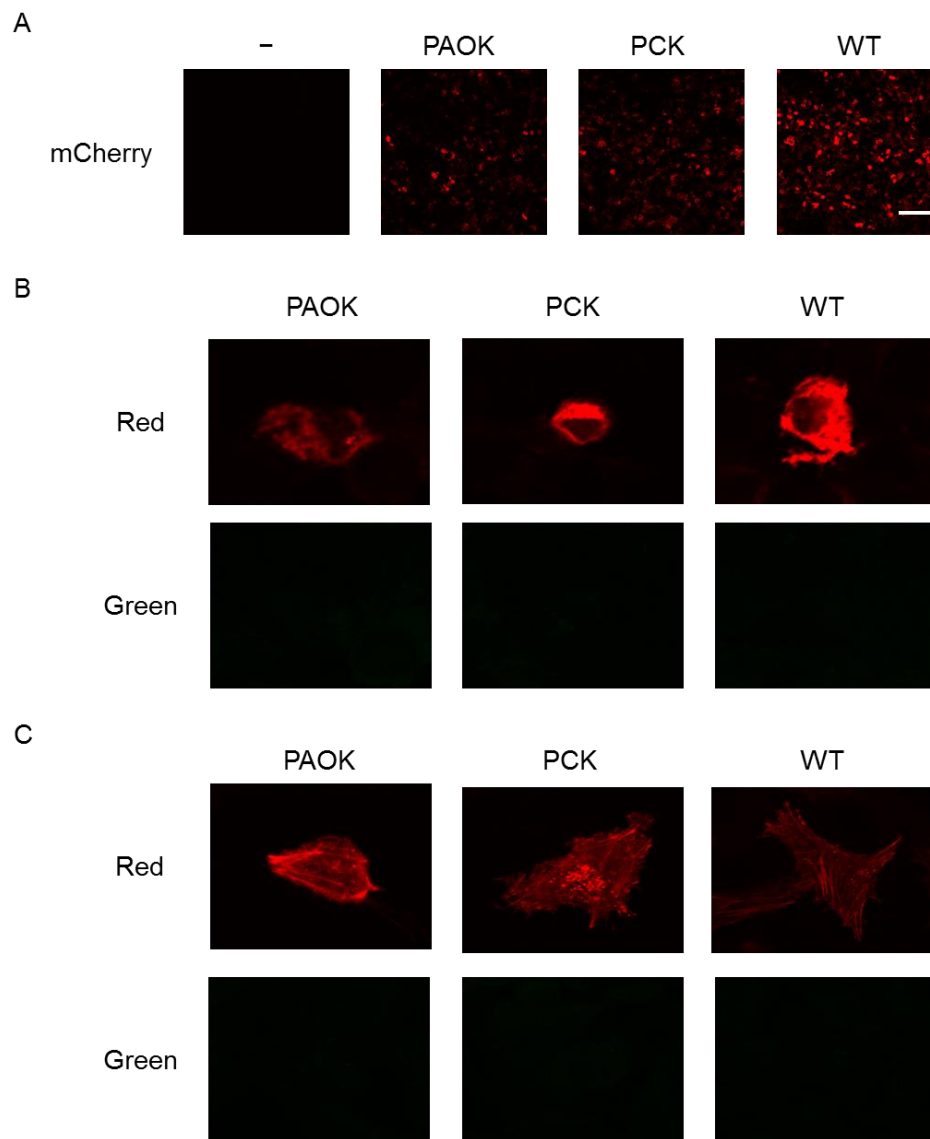


Figure 4.9. Cellular labeling with genetically encoded aminoxy group. (A) Genetic incorporation of PAOK into mCherry-actin-K144TAG in HEK293T cells. Cells were transfected with pmCherry-actin-K144TAG and pPCKRS-4CMV-U6-PyIT plasmids, and were treated with PAOK (1 mM). Photocaged lysine (PCK) was used as a control. Images were taken at 24 h with an 20x objective. Scale bar indicates 100 μ m. (B) Cell labeling with CA in HEK293T cells. Fixed cells were reacted with CA

(coumarin aldehyde) for 2 h at room temperature. Images were taken with a Nikon confocal microscope (60x objective). (C) Cell labeling with CA in HeLa cells. Fixed cells were reacted with CA (coumarin aldehyde) for 2 h at room temperature. Images were taken with Nikon confocal microscope (60x objective).

In summary, we report the genetic encoding of an aminooxy nucleophile. Site-specific protein labeling was achieved through oxime ligation between an incorporated aminooxy side-chain and aldehyde or ketone dyes. Recently it was found that ketone-selective reagents are not fully orthogonal in mammalian cells, probably due to the existence of carbonyl derivatives on protein side-chains following post-translational oxidation processes.¹⁸⁶ Thus, an aminooxy-reactive reagent might show less background labeling. One potential concern is the non-specific reaction between the aminooxy group and free carbohydrates in cells.¹⁸⁵ And we have not found other reports using aminooxy group in cells and animals. However, the installation of an aminooxy group onto a protein still holds promise in the selective and rapid conjugations of proteins *in vitro* and *in vivo*. For example, it might allow spatio-temporal control of bioconjugation in mammalian cells.^{165a} The efficient protein labeling strategy is also expected to present new approaches for protein engineering, including synthesis of site-specific antibody-drug conjugate¹⁸⁷ and self-assembled antibody multimers.¹⁸⁸ It can also uniquely generate nonhydrolyzable ubiquitin conjugate in test tubes, as has been shown before.¹⁷⁹ In the future, effort is needed to further increase the incorporation efficiency of photocaged aminooxy UAA (currently the yield is only 3 mg/L). This could be done by identifying a better mutant PylRS that incorporates PAOK. Other aminooxy UAA with a different photocaging group, for example, a coumarin caging group,⁶⁰ could also be explored. For cellular labeling experiment, other fluorescent dyes that show higher brightness under microscope might also be tested.

Experimental

Plasmid construction

The backbone of pBAD plasmid was generated by digesting pBAD-Myo4TAG-PylT plasmid (8000 ng, gift from Dr. Ashton Cropp lab) with NcoI and NdeI restriction enzymes, followed by gel extraction of the band corresponding to the correct length of the backbone (~5500 bp). The pBAD-Ub-K48TAG-PylT plasmid was constructed by cloning Ub-K48TAG fragment into a pBAD backbone using NcoI and NdeI restriction sites (Fwd: 5'-gacccatggcgattttgtgaaaaccctgacc-3', Rev: 5'-tagcatatgttaatggtgatggtgatgatgGCCGCTGCCgccgccacgcagacgcagcac-3'). A Gly-Ser-Gly linker was added between Ubiquitin and Histidine tag to allow adequate flexibility of ubiquitin (capitalized letter). The pmCherry-actin plasmid was a gift from the Roy lab (University of Pittsburgh).¹⁸⁹ The original TAG stop codon on the pmCherry-actin plasmid was mutated to TGA by site-directed mutagenesis (Fwd: 5'-cgtccaccgcaaagtcttctgaggatccaccggatctagataac-3', Rev: 5'-gttatctagatccggtggatcctcagaagcatttgcggtggacg-3'). The TAG mutation was introduced at lysine 144 position by site-directed mutagenesis (Fwd: 5'-ggccccgtaatgcagaagtagaccatgggctgggaggcc-3', Rev: 5'-ggcctcccagcccatggtctacttctgcattacggggcc-3').

Protein expression in *E. coli*

pBAD-sfGFP-Y151TAG-pylT (or pBAD-Myo4TAG-pylT or pBAD-Ub-K48TAG-pylT) and pBK-PCKRS (*M. barkeri* PylRS with mutations M241F, A267S, Y271C, and L274M, generated by Dr. Jason Chin)^{19d} were co-transformed (50 ng for each) into *E. coli* Top10 cells. A single colony was grown overnight (37 °C, 250 rpm) in LB supplemented with Amp (50 µg/ml) and Tet (25 µg/ml), and the overnight culture (250 µl) was added to LB (25 ml) supplemented with the designated unnatural amino acid, Tet (25 µg/ml), and Amp (50 µg/ml). Cells were grown at 37 °C, 250 rpm, and the protein expression was induced with arabinose (0.1 %) when OD₆₀₀ reached 0.4 (measured by Nanodrop). After overnight expression at 37 °C, cells were pelleted (5000 g, 10

min) and resuspended in phosphate lysis buffer (pH 8.0, 50 mM of phosphate salt, 6 ml). Triton X-100 (60 μ l, 10%) was added to the mixture. The lysate was incubated on ice for 1 h, sonicated (power level 5, pulse 'on' for 30 sec, pulse 'off' for 30 sec, with a total of 4 min, 550 sonic dismembrator), and then centrifuged (13,000 g) at 4 °C for 10 min. The supernatant was transferred to a 15 mL conical tube, and Ni-NTA resin (Qiagen, 100 μ l) was added. The mixture was incubated at 4 °C for 2 h with mild shaking. The resin was then collected by centrifugation (1,000 g, 10 min) and washed with lysis buffer (300 μ l), this was repeated twice, followed by two washes with wash buffer (300 μ l) containing imidazole (20 mM). The protein was eluted with elution buffer (300 μ l) containing imidazole (250 mM). The purified proteins were analyzed by SDS-PAGE (10%), and stained with Coomassie Brilliant Blue. For protein mass spectrometry, the purified protein was dialyzed in 50 mM ammonia acetate solution overnight at 4 °C. The dialyzed sample (10 μ l, 5 μ M) was analyzed by electrospray ionization mass spectrometry (Thermo Scientific Q-Exactive Oribtrap).

Protein labeling

sfGFP-PAOK (or Myo-PAOK) (5 μ M, 100 μ l) was irradiated in a PCR tube with a UV transilluminator (10 min, 365 nm, VWR), and was mixed with CA or FK (final concentration of 1 mM) in the presence of aniline (final concentration of 100 mM). The reaction was performed in NaOAc buffer (pH 5.5) at room temperature. Samples (20 μ l) were taken out at different reaction time points (1/6 h, 2 h, 6 h, 24 h), and were analyzed by SDS-PAGE (10%). SDS-PAGE was run at 60 V for 15 min, and 150 V for 45 min. Fluorescence was analyzed in-gel by ChemiDoc (Alexa 488 setting for CA, Fluorescein setting for FK, both settings utilize the same excitation (470 nm) and emission (530 nm) filters). The same gel was further stained with Coomassie Brilliant Blue and imaged by ChemiDoc (Coomassie Blue setting).

Protein PEGylation

sfGFP-PAOK (5 μ M, 40 μ l) was irradiated with a UV transilluminator (10 min, 365 nm, VWR), and was incubated with PEG-5k-aldehyde (final concentration of 10 mM, Sigma #41964) in the presence of aniline (final concentration of 100 mM). The reaction was performed at 37 °C for 2 days, and the product was analyzed by SDS-PAGE (10%). SDS-PAGE was run at 60 V for 15 min, and 150 V for 45 min. Protein was stained with Coomassie Brilliant Blue, and imaged with ChemiDoc (Coomassie Blue setting). PEGylation yield was calculated based on the intensity of protein bands. The protein bands were selected on ImageJ, and the intensity was analyzed by “Plot Lane” function.

Ubiquitin dimerization

Ub-PAOK (5 μ M, 100 μ l) was irradiated with a UV transilluminator (10 min, 365 nm, VWR), and was incubated with Ub-aldehyde (final concentration of 5 μ M, Enzo Life Sciences) in the presence of aniline (final concentration of 100 mM). The reaction was performed at 37 °C for 2 days, and was analyzed by SDS-PAGE (10%). SDS-PAGE was run at 60 V for 15 min, and 150 V for 45 min. Protein was stained with Coomassie Brilliant Blue, and imaged with ChemiDoc (Coomassie Blue setting). The dimer formation was calculated based on intensity of protein band. The protein bands were selected on ImageJ, and the intensity was analyzed by “Plot Lane” function.

Genetic incorporation of PAOK in mammalian cell

Human embryonic kidney (HEK) 293T cells (ATCC, #CRL-11268) were grown in DMEM (Dulbecco's Modified Eagle Medium) supplemented with 10% FBS, 1% Pen-Strep, and glutamine (2 mM) in 96-well plates in a humidified atmosphere with 5% CO₂ at 37 °C. HEK 293T cells were transiently transfected with pmCherry-TAG-EGFP-HA (200 ng) and pPCKRS-4CMV-U6-PylT (200 ng) at ~75% confluency in the presence of PAOK (1 mM) or PCK (0.25 mM). Branched

polyethylene imine (bPEI, 1.5 μ l, 1 mg/ml) was used as transfection reagent for each well. After an overnight incubation at 37 °C, cells were imaged on a Zeiss Observer Z1 microscope, with EGFP channel (filter set 38 HE; ex. BP470/40; em. BP525/50) and mCherry channel (filter set 43 HE; ex. BP575/25; em. BP605/70).

Western blotting

HEK 293T cells were co-transfected with pmCherry-TAG-EGFP-HA (2000 ng) and pPCKRS-4CMV-U6-PyIT (2000 ng) in the presence or absence of PAOK (1.5 mM) in a six-well plate. Branched polyethylene imine (bPEI, 15 μ l, 1 mg/ml) was used as transfection reagent for each well. After 48 h of incubation, the cells were washed with chilled phosphate-buffer saline (PBS, 1 ml), and lysed in mammalian protein extraction buffer (250 μ l, GE Healthcare). The cell lysates were separated by 10% SDS-PAGE (run with 60V for 15 min, and 150 V for 45 min) and were transferred to a PVDF membrane (GE Healthcare). The membrane was blocked in tris-buffer saline (TBS) with 0.1% Tween 20 and 5% milk for 1 h. The blots were probed with the primary antibody (1:1,000, anti-HA (sc-805) or anti-GAPDH (sc-25778), Santa Cruz) overnight at 4 °C, followed by incubation with secondary goat anti-rabbit IgG-HRP antibody (1:20,000, sc-2004, Santa Cruz) for 1 h at room temperature. The blots were further incubated with the SuperSignal West Pico working solution (mixture of the Stable Peroxide Solution and the Luminol/Enhancer Solution, 500 μ l each, Thermo Scientific) for 5 min at room temperature. The luminescence signal was detected by ChemiDoc (Chemi Hi Sensitivity setting, manual exposure time: 10 sec).

Cellular labeling with the genetically encoded aminooxy group

Human embryonic kidney (HEK) 293T cells or HeLa cells were co-transfected with pmCherry-actin-K144TAG (200 ng) and pPCKRS-4CMV-U6-PyIT (200 ng) at ~75% confluency in the presence of PAOK (1 mM) or PCK (0.25 mM). Branched polyethylene imine (bPEI, 1.5 μ l, 1 mg/ml)

was used as transfection reagent for each well. After an overnight incubation at 37 °C, cells were washed with chilled PBS, fixed with 3.75% formaldehyde (200 µl), and permeabilized with 0.5% Triton X100 (100 µl). Cells were then irradiated in PBS (100 µl) with UV light (365 nm, VWR UV transilluminator) for 2 min, and were reacted with CA (coumarin-aldehyde, final concentration of 1 µM) in the presence of aniline (final concentration of 100 mM) for 2 h at room temperature. After reaction, cells were washed with PBS (200 µl) twice, and were imaged with a confocal microscope (60x objective, Nikon A1).

5.0 Expanding the Genetic Code in Zebrafish Embryos

5.1 Incorporation of Unnatural Amino Acids into Proteins in Zebrafish

This material was reprinted, in part, with permission from [Liu, J.; Hemphill, J.; Samanta, S.; Tsang, M.; Deiters, A. *J. Am. Chem. Soc.* 2017, 139 \(27\).](#)

Although site-specific incorporation of unnatural amino acids has been performed in metazoans,^{24-26, 190} previous experiments were limited to reporter genes and no expression of functional proteins that affect animal physiology has been reported. The zebrafish is a commonly employed model organism for vertebrate development,¹⁹¹ disease modeling¹⁹² and drug discovery.¹⁹³ The *ex vivo* development and transparency of the embryo make it an excellent system for the application of non-invasive optical tools, including light-activated antisense agents,¹⁹⁴ thereby providing insight into gene regulatory processes and networks with spatial and temporal resolution. Moreover, microinjection of mRNA into the 1-cell stage embryo is a standard and rapid approach for delivery of exogenous genes that can be readily adapted to encode for any gene product and provides homogenous protein expression in zebrafish.¹⁹⁵ Taken together, these distinct advantages over other model organisms make the zebrafish an ideal system for a wide range of biological studies.¹⁹⁶

Light-regulation of protein activity in zebrafish has been reported using natural photoreceptor domains,¹⁹⁷ however, the genetic encoding of photocaged amino acids will further expand the optogenetic toolbox and will enable a rational design of light-activated proteins based on their function or structure. For example, photocaged lysine analogs have been applied to the optical control of protein localization,^{19d} kinase function,¹⁴⁸ and CRISPR/Cas9 gene editing¹⁹⁸ in

human cells. By genetically encoding a photocaged lysine using the PylRS system in zebrafish embryos, we demonstrate the consequences of optical control of MEK activation at different stages in development. Temporal control of kinase function led to the identification of a critical time window for activity of the MEK/ERK pathway in order to establish dorsal/ventral polarity in the early embryo.

We first tested incorporation of unnatural amino acids using a *Renilla* luciferase (Rluc) reporter assay. Wild-type Rluc was active in zebrafish embryos (Fig 5.1), and was used as a positive control.

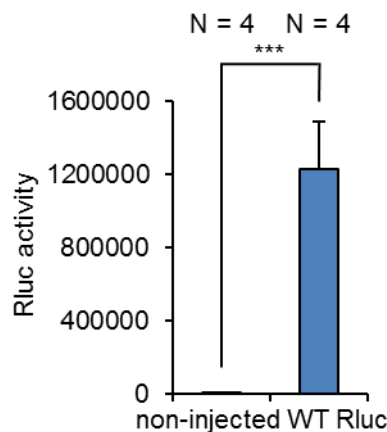


Figure 5.1. Activity of wild-type Rluc in zebrafish embryos. Embryos were injected with wild-type Rluc mRNA (100 pg). Embryos were collected at 48 hpf, and luciferase assays were performed. N indicates the number of pooled samples (4 embryos each). Statistical significance is indicated by *** $p < 0.001$ (unpaired t-tests).

We predicted that incorporation of an unnatural amino acid at a leucine residue (L95) located at the surface of Rluc would not interfere with Rluc function, thereby generating a highly specific reporter for amber codon suppression (Fig 5.2B). Thus, L95 was mutated to a TAG codon to probe read-through during translation. After injection of all the necessary genetic components

to the zebrafish embryos, the fish lysate was collected at 48 hpf for luciferase assay. The Rluc activity indicated the incorporation efficiency of the UAA (Fig 5.2A).

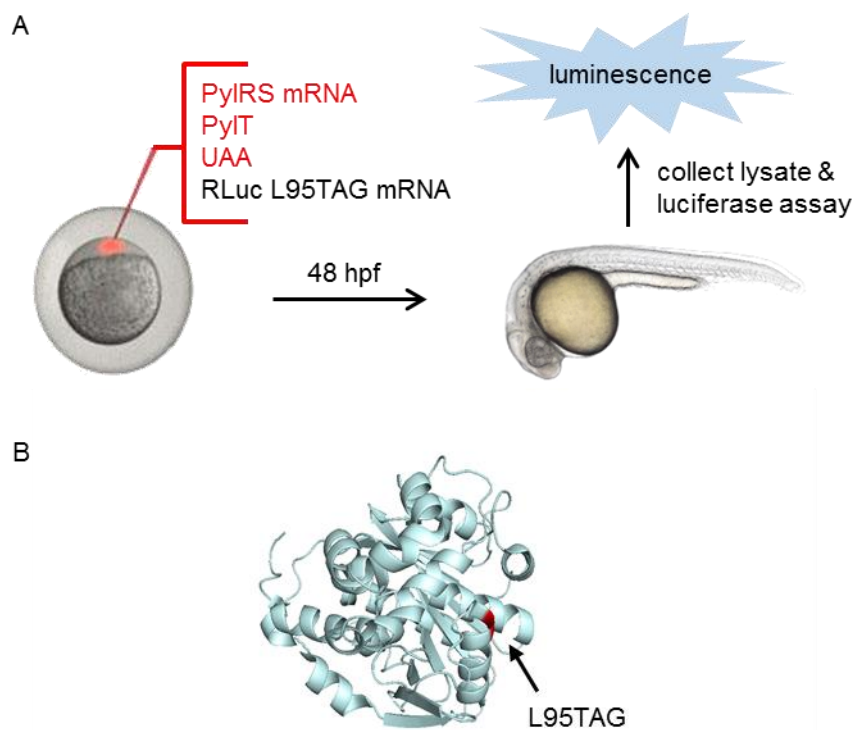


Figure 5.2. Injection, gene expression, and Rluc reporter assay. (A) Zebrafish embryos were injected with PyIRS mRNA, PyIT, RLuc-L95TAG mRNA, PyIT, and UAA at 1-2 cell stage. At 48 hpf, fish lysate was collected for luciferase assay. (B) Renilla luciferase as a reporter for unnatural amino acid incorporation. The unnatural amino acids were incorporated at a permissive leucine residue (L95) located at the protein surface (shown in red). PDB: 2PSD.

To this end, the *M. barkeri* wild-type PyIRS (WTRS), RLuc-L95TAG mRNAs and PyIT were synthesized through *in vitro* transcription. WTRS mRNA and PyIT were injected together with the unnatural amino acid (UAA) into zebrafish embryos. After 48 h, zebrafish lysate was collected for luciferase assays and a 217- and 161-fold increase of Rluc activity was observed in the presence of the UAAs **1** and **2**, respectively (Fig 5.3). Negligible Rluc activity in the absence of the UAA

demonstrated the excellent fidelity of the PylRS system in zebrafish embryos, as none of the common 20 amino acids were recognized as substrates.

Inspired by the success of using wild-type PylRS for incorporating **1** and **2**, which could be applied in both protein labeling and protein activation experiments,^{18b, 157b} we examined if mutant PylRS enzymes can be employed in embryos to incorporate more structurally complex amino acids. To this end, we synthesized mRNA of HCKRS⁶⁰ and OABKRS,⁶¹ which have been shown to incorporate **3** and **4**, respectively. We performed injections as described above, and observed a 70- and 34-fold increase of Rluc activity in the presence of **3** and **4**, respectively (Fig 5.3).

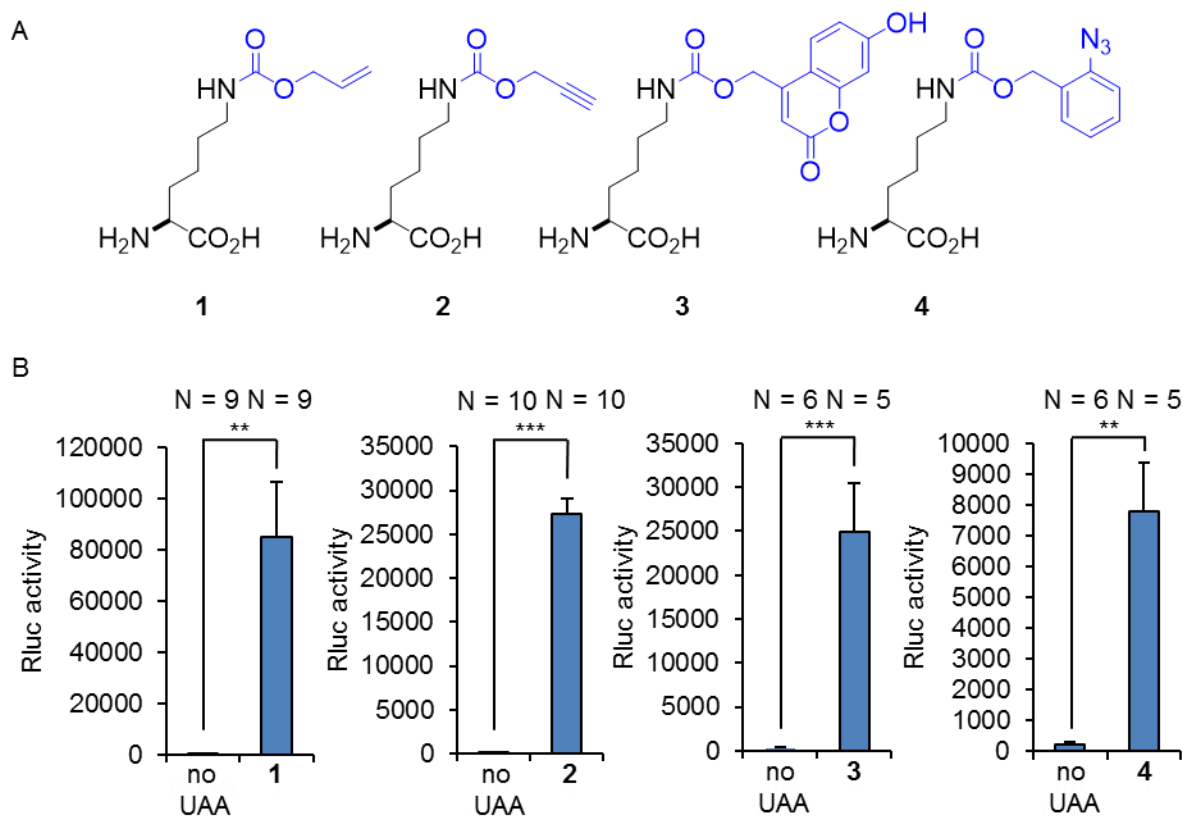


Figure 5.3. Genetic encoding of the unnatural amino acids 1-4 in zebrafish embryos. (A) Structures of the unnatural lysine derivatives modified with an alkene, an alkyne, a coumarin caging group, and an azido benzyl group. (B) Incorporation of 1-4 into Rluc-L95TAG quantified by Rluc activity.

We also tested the toxicity of these UAAs in zebrafish. No toxicity was observed for any of the four UAAs (Fig 5.4). Taken together, we demonstrated successful genetic encoding of four different UAAs in zebrafish embryos.

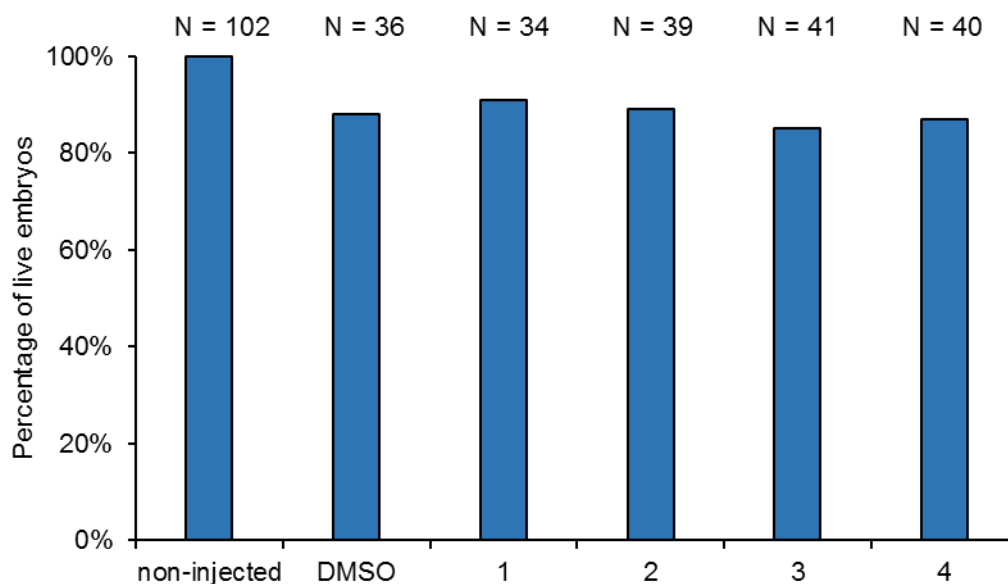


Figure 5.4. Toxicity test of the unnatural amino acids **1-4**. Embryos were injected with 1% DMSO or unnatural amino acid, and the numbers of live and dead embryos were counted at 24 hpf. No developmental defects were observed in live embryos. N represents the number of embryos observed.

We further explored the effect of PylT on incorporation efficiency. We found that chemically synthesized PylT showed similar efficacy compared to *in vitro* transcribed PylT (Fig 5.5). However, when we tested *in vitro* transcribed PylT without a CCA tail, significantly lower efficacy was noted (Fig 5.5). The CCA tail is a conserved sequence at the 3' end of tRNA, which is acylated with the amino acid. Although this sequence can be added by CCA-adding enzymes in cells,¹⁹⁹ our results suggested that the direct addition of a CCA tail on the PylT improved incorporation efficiency, and thus was used in all subsequent experiments.

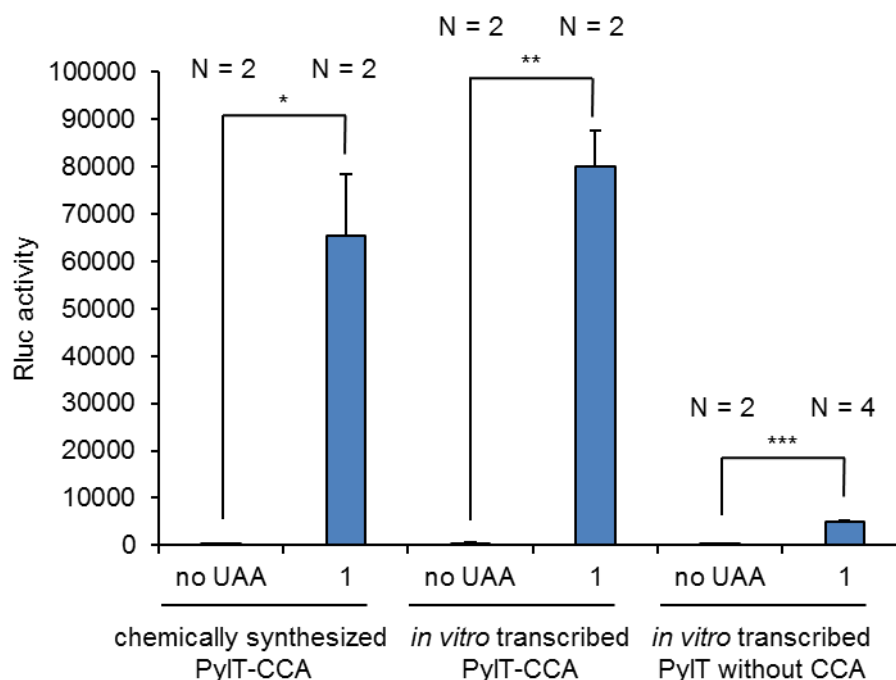


Figure 5.5. Comparison of incorporation efficiency with different PyITs. The unnatural amino acid **1** was incorporated into Rluc-L95TAG. N indicates the number of pooled samples (4 embryos each). Statistical significance is indicated by * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$ (unpaired t-tests).

We then tested the effect of PyIT concentration on incorporation efficiency. Different PyIT amount (2 ng, 1 ng, 500 pg) was injected per embryo, and the incorporation was assayed by luciferase activity. We found that 2 ng was critical for high incorporation efficiency (Fig 5.6). This is in agreement with previous findings of the PyIT amount possibly being a limiting factor for genetic encoding of UAAs.²⁰⁰

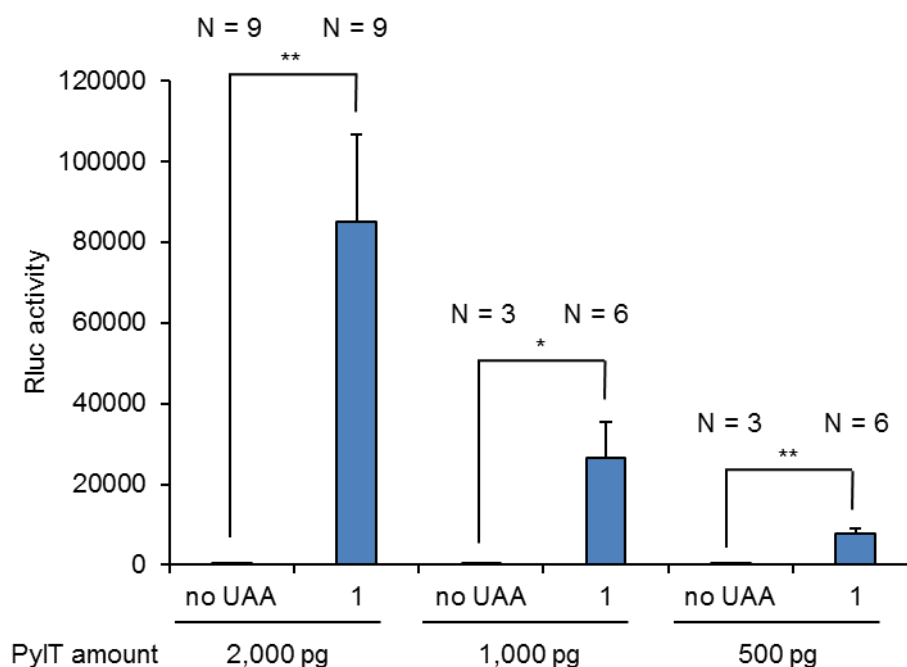


Figure 5.6. Comparison of incorporation efficiency with different amounts of PylT-CCA (*in vitro* transcribed). The unnatural amino acid **1** was incorporated into Rluc-L95TAG. N indicates the number of pooled samples (4 embryos each). Statistical significance is indicated by * $p < 0.05$ and ** $p < 0.01$ (unpaired t-tests).

We then asked if UAAs other than lysine derivatives could also be incorporated. We chose to test two photocaged tyrosines first (ONBY and NPY, discussed in Chapter 2.1), because they have been successfully incorporated in *E. coli* and mammalian cells with high yield.^{22, 66} Therefore, we synthesized mRNA of ONBYRS (corresponding to EV20 in *E. coli*, with L270F, L274M, N311G, C313G, Y349F mutations),⁶⁶ and performed injections as described above. We observed a 23-fold increase of Rluc activity in the presence of ONBY (Fig 5.7B). However, no significant Rluc activity was detected in the presence of NPY. The result is surprising and the experiment might worth repeating, because the incorporation efficiency of ONBY and NPY were similar in *E. coli* (data in Chapter 2.1). Moreover, I previously identified a better synthetase (ONBYRS1, with Y271M, L274T, N311A, C313A, Y349W mutations) for ONBY incorporation using the library

selection (data in chapter 2.3). This PyIRS mutant should be tested in the context of zebrafish as well.

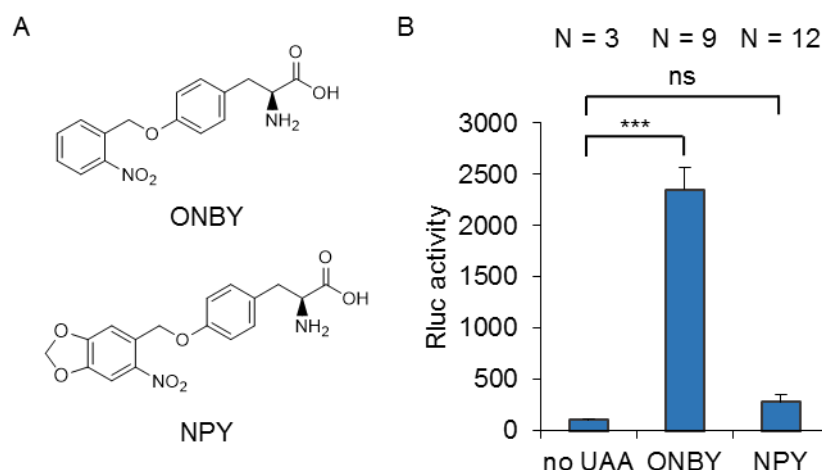


Figure 5.7. Genetic incorporation of caged tyrosine (ONBY, NPY) in zebrafish embryos. (A) Structures of ONBY and NPY. (B) Incorporation of ONBY and NPY into Rluc-L95TAG, quantified by Rluc activity. N indicates the number of pooled samples (4 embryos each). Statistical significance is indicated by ns: not significant, and *** $p < 0.001$ (unpaired t-tests).

The photocaged lysine **3** has previously been applied to control protein function in mammalian cells using 365 nm, 405 nm, and 760 nm (two-photon) irradiation.⁶⁰ With successful genetic encoding of **3** in zebrafish, we tested if protein function can be manipulated with light in developing embryos. As an initial proof-of-concept, we utilized firefly luciferase (Fluc) with a TAG amber codon at position lysine 206, because installation of **3** blocks Fluc activity until light exposure.⁶⁰ In order to create an internal control for incorporation efficiency, we fused Rluc to the C-terminus of Fluc-K206TAG (Fig 5.8A). To this end, Fluc-K206TAG-Rluc mRNA was injected, together with HCKRS mRNA, PyIT, and **3**, into zebrafish zygotes. After 48 h, embryos were either briefly irradiated at 365 nm or kept in the dark. Embryo lysate was subsequently collected and a luciferase assay was performed for both Fluc and Rluc. Excellent optical OFF to ON switching of Fluc function was observed, with negligible background activity before irradiation (Fig 5.8B).

Normalization of Fluc activity to Rluc activity, as a TAG codon suppression control, revealed a 26-fold increase of Fluc activity upon light-triggering. This result shows that light-activation of protein function can be achieved in live zebrafish embryos with an expanded genetic code.

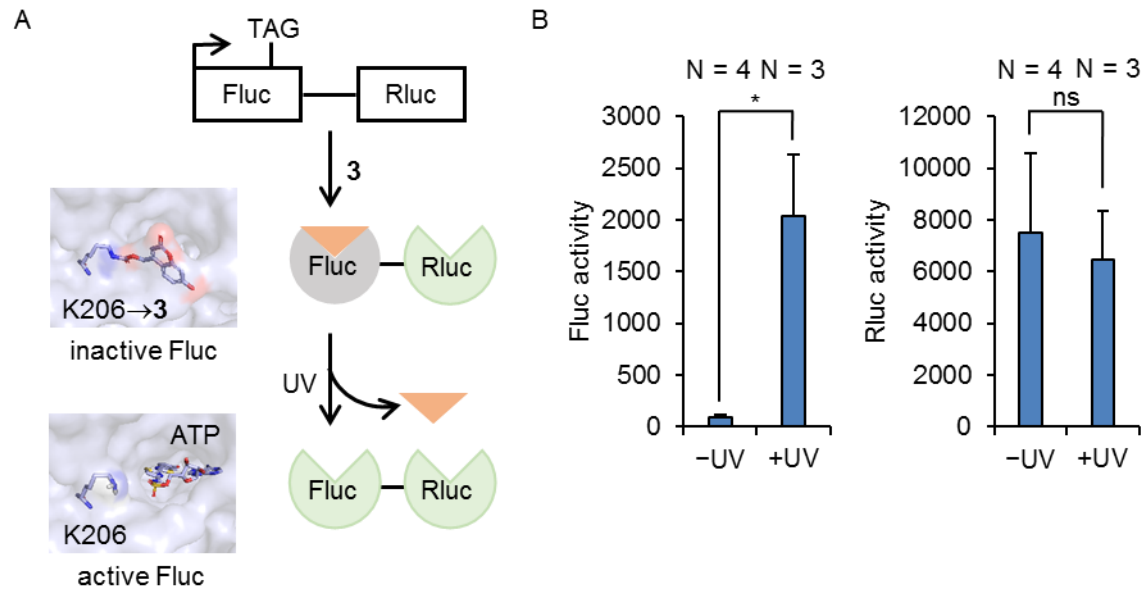


Figure 5.8. Optical control of luciferase activity (A) Construct used for generating a photocaged firefly luciferase followed by light activation. *Renilla* luciferase was used as an internal control for incorporation efficiency. Fluc containing **3** at position K206 is inactive, as ATP is blocked from the active site (PDB 2D1S). Decaging restores Fluc activity. (B) Fluc activity was observed after UV exposure, while Rluc activity was not affected and was used as an internal control. N indicates the number of pooled samples (4 embryos each). Statistical significance is indicated by ns (not significant), * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$ (unpaired t-tests)

We then sought to apply genetic code expansion in zebrafish to an enzyme with endogenous function in order to demonstrate its utility in altering embryonic development. Incorporation of photocaged amino acids into proteins enables precise dissection of signaling pathways with light, and this approach has been applied to study the dynamics of MEK/ERK

signaling in mammalian cells.¹⁴⁸ While the photocaged lysine **3** has not been applied on MEK1 before, we first confirmed the incorporation of **3** into MEK1 in *E. coli* (Fig 5.9).

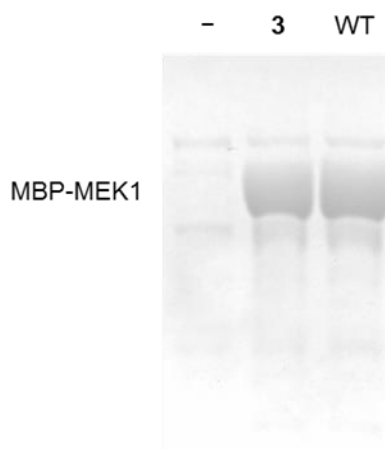


Figure 5.9. Coomassie stained gel of Ni-purified MBP-MEK1-K97TAG-**3** expressed in *E. coli*. No protein expression was detected in the absence of **3**. MBP-MEK1-WT was used as a positive control.

We further confirmed the subsequent decaging through UV exposure, by MS/MS analysis of recombinantly expressed MEK1 protein (Fig 5.10). This result demonstrated that light-activation of the MEK1 could be achieved through the incorporation of **3**.

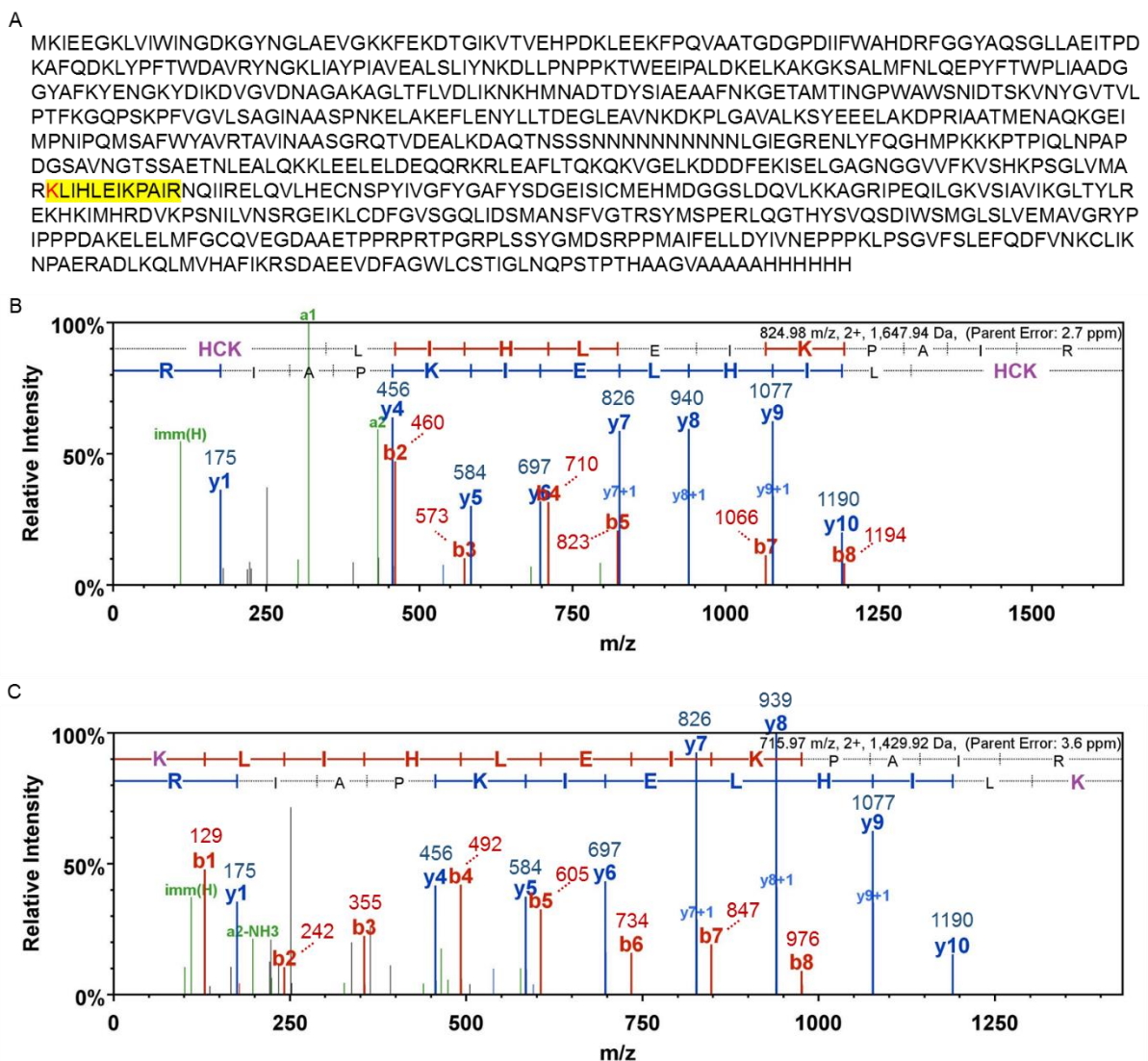


Figure 5.10. MS/MS analysis confirms incorporation of **3** into MEK1 and subsequent decaging with UV light. (a) Full-length protein sequence of MBP-MEK1. Target fragment with **3** is highlighted in yellow, and the K97 position is labeled in red. (b) The spectrum before UV irradiation indicates **3** at the 97 position. (c) The spectrum after UV irradiation indicates lysine at the 97 position. The MBP-MEK1-K97TAG-**3** protein was recombinantly expressed in *E. coli*. MS/MS analyses were conducted by MS Bioworks (Ann Arbor, MI).

While optical activation on the second to minute timescale in mammalian cells provided further insight into adaptive behavior of the MEK/ERK network in single cells, in the context of zebrafish biology, kinase signaling pathways are important regulators throughout embryogenesis.²⁰¹ The MEK/ERK pathway is a well-known downstream target of Fibroblast Growth Factor (FGF) signaling and plays an important role in mesendoderm induction and dorsoventral patterning of the zebrafish embryo. FGF signaling induces expression of chordin and noggin, secreted inhibitors of ventralizing bone morphogenetic proteins resulting in dorsalization (Fig 5.11A).²⁰² An inhibitor-based chemical approach has previously been used for perturbation of the MEK/ERK pathway during zebrafish development;²⁰³ however, pharmacological inhibitors only allow for the deactivation of kinase function – not activation, and their specificity is often limited. We reasoned that optical activation of the MEK/ERK pathway in zebrafish provides an innovative tool to study its role, as site-specific incorporation of the caged amino acid **3** conveys complete kinase specificity. Substituting the critical lysine 97 with **3**, the caging group blocks the ability of the enzyme to correctly position ATP in the MEK1 active site (Fig 5.11B). We then generated mRNA of constitutively active MEK1 (caMEK1, containing S218D and S222D mutations), and confirmed that injection of caMEK1 led to dorsalized embryos at 10 hpf (Fig 5.11D,E), as previously reported.²⁰³ We further generated caMEK1-K97TAG mRNA and injected it into zebrafish embryos, together with HCKRS mRNA, PylT, and **3**. When these embryos were left in the dark, they developed normally, indicating that caged MEK1 was inactive (Fig 5.11D). To activate caged MEK1 at different developmental stages, we irradiated embryos for 30 seconds at 2 h, 5 h or 8 h post injection. Light-activation of MEK1 can efficiently increase ERK phosphorylation at all three time points (Fig 5.11C). Embryos irradiated at 2 h and 5 h showed an elongated phenotype at 10 hpf (Fig 5.11D,E). However, the majority of embryos irradiated at 8 h appeared normal at 10 hpf (Fig 5.11E), indicating that active MEK was not able to efficiently trigger an elongated phenotype after 8 hpf.

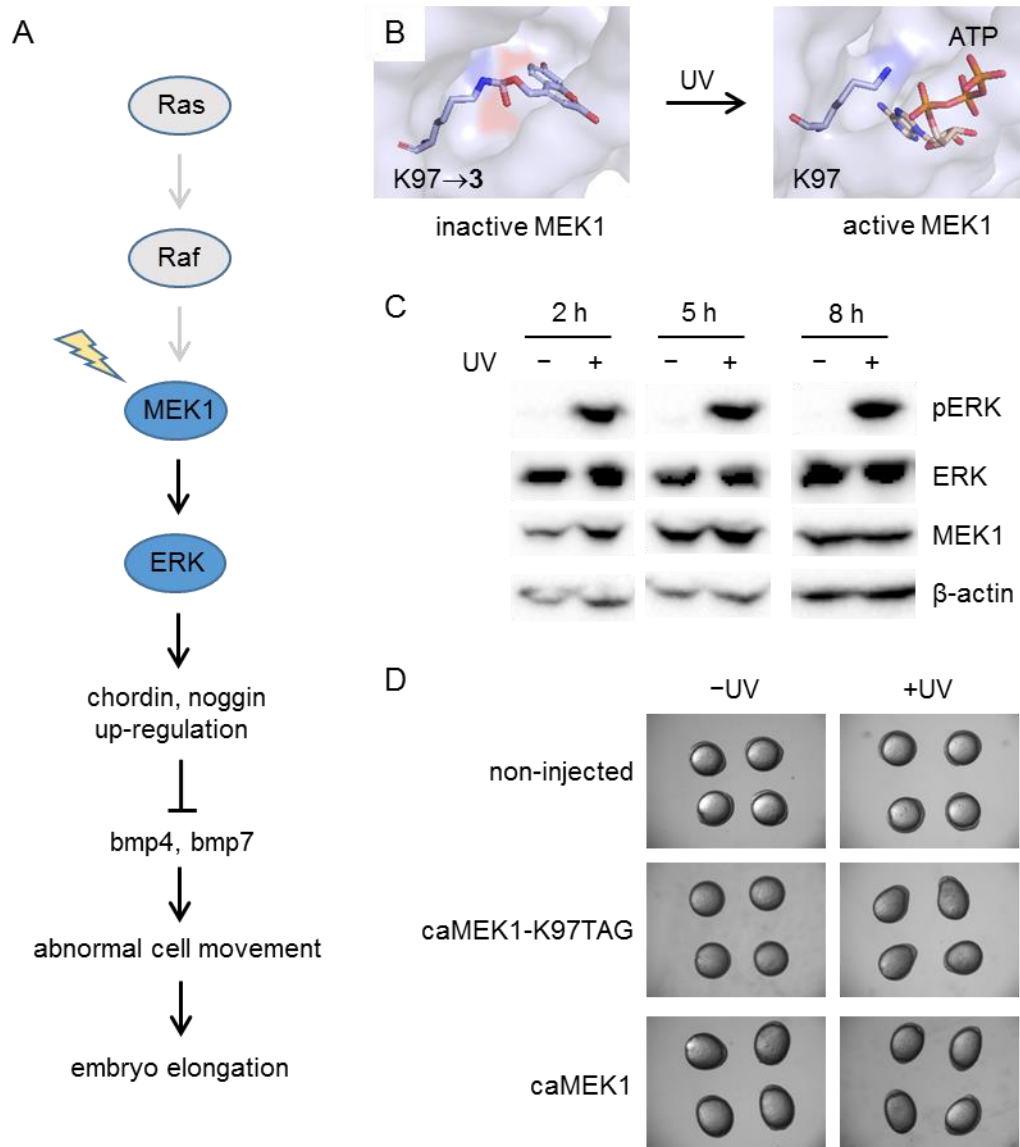


Figure 5.11. Light-activation of MEK1 leads to elongated zebrafish embryos. (A) Activation of caged MEK1 induces an elongated phenotype through the secreted bmp inhibitors chordin and noggin. (B) MEK1 containing **3** at position K97 is inactive, as ATP is blocked from the active site (PDB 1S9J). Removal of the caging group through light exposure restores MEK1 activity. (C) Time-course analysis of ERK phosphorylation by activated MEK1. (D) Micrographs of embryos imaged at 10 hpf; irradiation was performed at 5 h post injection. Embryos expressing caged MEK1 only displayed an elongation phenotype when activated through light exposure. (E) Temporal activation of MEK1 reveals a critical

time window for activity of the MEK/ERK pathway in the early embryo. N indicates the number of phenotypically scored embryos.

We then tested if optical activation of the MEK/ERK pathway resulted in a change at the gene expression level. We probed expression of the *brachyury homolog a (ta)* gene, a well-known downstream target of the FGF/MEK/ERK pathway.²⁰⁴ At shield stage (6 hpf), embryos that were exposed to UV light showed broader expression of *ta* in the margin when compared to embryos that were kept in the dark (Fig 5.12A). In some instances, *ta* expression was detected at the animal pole of the embryos, a pattern that is similar to embryos injected with constitutively active MEK1. At bud stage (~10 hpf), the expression of the *ta* was also wider along the notochord in light-activated embryos compared to embryos that were not irradiated (Fig 5.12B).

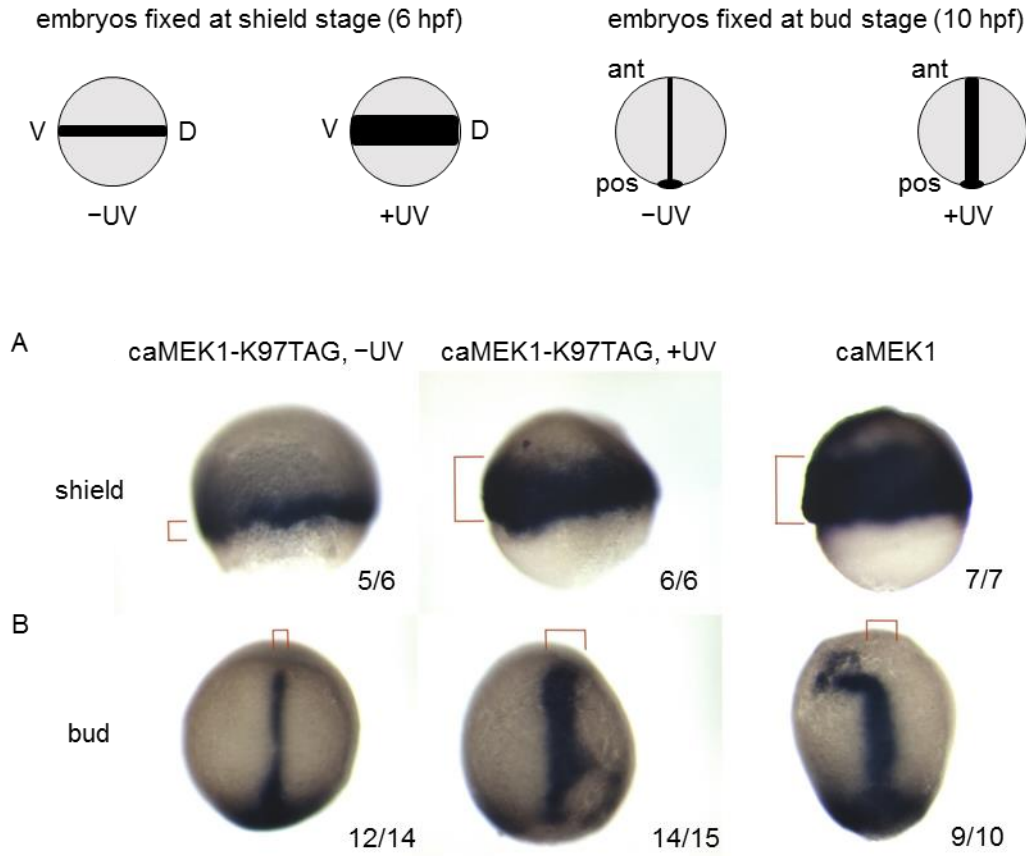


Figure 5.12. Upregulation of downstream targets in response to light-activated MEK1. (A) Expression of the *brachyury homolog a (ta)* gene at shield stage. Embryos that express caged MEK1 show broader expression after exposure to UV light, compared to embryos that were kept in the dark (see red brackets). (B) The same experiments conducted at bud stage showed wider expression along the notochord after light-activation. The number of embryos with the displayed expression and the total number of embryos is indicated. Red brackets mark the *ta* expression area. Lateral views (A) and dorsal views with the anterior at the top (B).

We also probed expression of the *chordin (chd)* gene, a marker for dorsalized embryos that is known to be induced following activation of the FGF/Ras/MAPK pathway.²⁰⁵ As expected, embryos that were exposed to UV light showed expanded expression of *chd* at shield stage, compared to embryos that were kept in the dark (Fig 5.13).

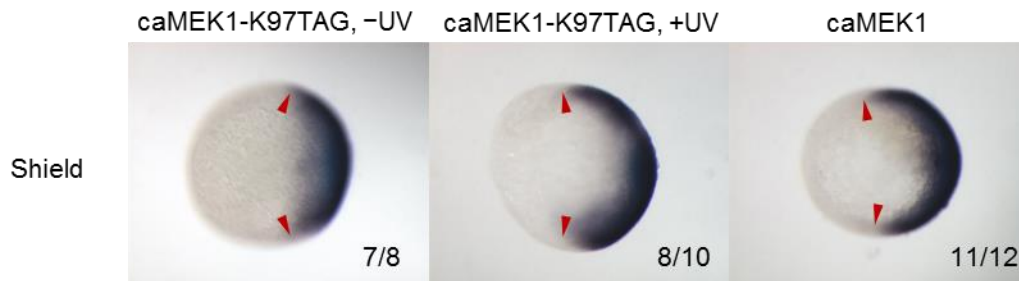


Figure 5.13. Upregulation of *chordin* (*chd*) expression in response to light-activation of MEK1, imaged at shield stage. Embryos that express caged MEK1 show expanded expression after exposure to UV light, compared to embryos kept in the dark (see red arrows). Animal views with the dorsal side to the right. Number of embryos that displayed the shown expression pattern and total number of embryos are shown.

Taken together, the observed *ta* and *chd* expression patterns in response to optical MEK1 activation and the time-resolved phenotypic studies demonstrate that the MEK/ERK pathway influences dorsal/ventral patterning in zebrafish development before 8 hpf, thereby providing support for early intervention with pharmacological MEK inhibitors for related congenital defects in humans, such as cardio-facio-cutaneous syndrome.^{203, 206}

In conclusion, we incorporated four unnatural amino acids into proteins in zebrafish embryos through genetic code expansion using injection methods that are applicable to many zebrafish studies. We demonstrated light-activation of enzymatic function, specifically luciferase activity, through site-specific incorporation of a photocaged unnatural amino acid in live embryos. We then applied this methodology to the temporal activation of the MEK/ERK pathway in zebrafish, and identified a time window for MEK activity that can influence dorsoventral patterning. Besides controlling protein function with light, other potential applications of unnatural amino acids in live zebrafish embryos include small molecule triggered protein activation, site-specific labeling of proteins with fluorescent and biophysical probes, and probing protein interactions through covalent bond formation with electrophilic or photocrosslinking groups. The zebrafish is a well-

established model organism for human development and disease, and we anticipate that the ability to genetically encode a 21st amino acid will become a powerful tool to manipulate and study protein function in animals.

Experimental

Materials

The zebrafish experiments were performed according to a protocol approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Pittsburgh. The AB* strain was maintained under standard conditions at the University of Pittsburgh School of Medicine in accordance with Institutional and Federal guidelines. Unnatural amino acid **1** was purchased from Chem-Impex International. Unnatural amino acids **2** to **4**, ONBY, and NPY were synthesized as described previously.^{18b, 22, 60-61, 66}

Statistical methods

Error bars represent s.e.m. for the N numbers shown in figures. Sample sizes were chosen because preliminary experiments suggested that it would be sufficient to indicate incorporation of unnatural amino acid. No statistical methods were used to predetermine sample size. Unpaired t-tests were used to calculate statistical significance.

Plasmid construction

The *Renilla* luciferase (Rluc) gene was amplified from pGL4.7 (Promega) using primers P1 and P2 (for sequence information see Table 5.1), and was cloned into BamHI and XbaI sites of the pCS2+ plasmid through an in-fusion cloning kit (Clontech). The L95TAG mutation was introduced into pCS2-Rluc by QuikChange site-directed mutagenesis (Agilent) using P3 and P4. PyIRS (WTRS, HCKRS, OABKRS, and ONBYRS) was amplified using P5 and P6, and was cloned into the pCS2+ plasmid as described above. The firefly luciferase (Fluc) gene was amplified from

pGL3 (Promega) using P7 and P8, and was cloned into the pCS2+ plasmid. The K206TAG mutation was introduced into pCS2-Fluc using P9 and P10. pCS2-Fluc-K206TAG-Rluc was assembled from two overlapping fragments, pCS2-Fluc-K206TAG (amplified with P11 and P12) and Rluc (amplified with P13 and P14), using the Gibson assembly method. Maltose binding protein (MBP) tagged MEK1 gene (MBP-MEK1) was amplified from Addgene plasmid # 68300,²⁰⁷ using primers P15 and P16, and was cloned into NcoI and AseI sites of the pBAD-pyIT plasmid. The K97TAG mutation was introduced using P17 and P18. Constitutively active caMEK1 (containing S218D and S222D mutations) and MEK1-K97TAG were amplified using P19 and P20, and were cloned into the pCS2+ plasmid as described above. All plasmids were confirmed by Sanger sequencing. For map of newly constructed plasmid, see Appendix B.

reached 0.4. After overnight expression at 27 °C, cells were pelleted and resuspended in phosphate lysis buffer (pH 8.0, 50 mM, 6 ml). Triton X-100 (60 µl, 10%) was added to the mixture. The lysate was incubated on ice for 1 h, sonicated, and then centrifuged (13,000 g) at 4 °C for 10 min. The supernatant was transferred to a 15 ml conical tube and Ni-NTA resin (Qiagen, 100 µl) was added. The mixture was incubated at 4 °C for 2 h with mild shaking. The resin was then collected by centrifugation (1,000 g, 10 min), and washed with lysis buffer (800 µl) and wash buffer (800 µl) containing imidazole (20 mM). The protein was eluted with elution buffer (200 µl) containing imidazole (250 mM). Expression of MBP-MEK1-K97TAG-3 was confirmed by SDS-PAGE (8%). For protein decaging, purified proteins (20 µl) were irradiated in a PCR tube for 5 min with a 365 nm UV transilluminator (8 mW/cm²).

Protein MS/MS

The non-irradiated and irradiated protein samples were analyzed by SDS-PAGE (8%), and stained with Coomassie Brilliant Blue. Regions corresponding to the expected molecular weight of MBP-MEK1 were excised and sent for LC-MS/MS analysis (MS Bioworks). The in-gel trypsin digests were analyzed by nano LC-MS/MS with a Waters NanoAcquity HPLC system interfaced to a ThermoFisher Q Exactive. Peptides were loaded on a trapping column and eluted over a 75 µm analytical column at 350 nL/min. The mass spectrometer was operated in data-dependent mode, with MS and MS/MS performed in the Orbitrap at 70,000 FWHM resolution and 17,500 FWHM resolution, respectively. The fifteen most abundant ions were selected for MS/MS. The peptide mass tolerance was set as 10 ppm, and the fragment mass tolerance was set as 0.02 Da. Data were filtered using a minimum protein value of 90%, a minimum peptide value of 50% and requiring at least two unique peptides per protein.

mRNA synthesis

The corresponding pCS2+ plasmid (10 µg) was linearized through NotI digestion (2 µl of NotI in 40 µl reaction, 37 °C, overnight). The linearized product (40 µl) was mixed with water (10 µl) and phenol:chloroform:isoamyl alcohol (PCIA, 50 µl), and centrifuged at maximum speed for 5 min. The top layer (~50 µl) was collected, and further mixed with NaOAc (5 µl, 3 M, pH 5.2) and 100% ethanol (125 µl). The mixture was centrifuged at maximum speed for 5 min. The pellet was washed with 70% ethanol (600 µl), and dissolved in 10 µl of water. The concentration of linear DNA was determined by Nanodrop. Linear DNA (1 µg) was used to generate mRNA in a 20 µl reaction (37 °C, 4 h) with the mMESSAGE mMACHINE SP6 Transcription Kit (Ambion). DNase (1 µl, provided by the kit) was then added to remove the linear DNA template (37 °C, 30 min). The reaction (20 µl) was mixed with water (30 µl) and PCIA (50 µl), and was centrifuged at maximum speed for 5 min. The top layer (~50 µl) was purified through a G-50 sephadex spin column (Roche, #11274015001) according to the product manual. The RNA solution (50 µl) was mixed with water (50 µl), NaOAc (10 µl, 3 M, pH 5.2), and 100% ethanol (300 µl). The mixture was placed at -20 °C freezer overnight, and was centrifuged at maximum speed for 15 min. The RNA pellet was washed with 70% ethanol, and dissolved in water (15 µl). The quality was verified by 1% agarose gel (run at 80 V for 45 min).

PylT synthesis

The PylT sequence contained a U25C mutation in order to increase incorporation efficiency.²⁰⁸ The PylT DNA with a truncated T7 promoter was amplified by PCR from a PylT oligonucleotide (IDT gBlock), using P21 and P22. The PCR was performed with 10 µM PylT template (4 µl), 10 mM dNTP (4 µl), 100 µM forward and reverse primers (2 µl each), 10x Taq PCR buffer (20 µl), Taq polymerase (2 µl), water (166 µl). The PCR product (100 µl) was mixed with phenol:chloroform:isoamyl alcohol (PCIA, 100 µl), and centrifuged at maximum speed for 5 min. The top layer (~100 µl) was collected, and further mixed with NaOAc (10 µl, 3 M, pH 5.2) and

100% ethanol (250 μ l). The mixture was centrifuged at maximum speed for 5 min. The pellet was washed with 70% ethanol (600 μ l), and dissolved in 10 μ l of water. The concentration of PylT DNA was obtained on a nanodrop. The pylT DNA (1.5 μ g) was used as a template to generate PylT RNA in a 20 μ l reaction (37 °C, 4 h), using the MEGAscript T7 Transcription Kit (Ambion). DNase (1 μ l, provided by the kit) was then added to remove the linear DNA template (37 °C, 30 min). The reaction (20 μ l) was mixed with water (30 μ l) and PCIA (50 μ l), and was centrifuged at maximum speed for 5 min. The top layer (~50 μ l) was mixed with water (50 μ l), NaOAc (10 μ l, 3 M, pH 5.2), and 100% ethanol (300 μ l). The mixture was placed at -20 °C freezer overnight, and was centrifuged at maximum speed for 15 min. The RNA pellet was washed with 70% ethanol (600 μ l), and dissolved in 10 μ l of water, and the quality was verified by 1.5% agarose gel. The PylT RNA without CCA was generated in a similar manner. Chemically synthesized PylT RNA (HPLC purified) was purchased from Integrated DNA Technologies (IDT). It was used directly without further purification. The PylT sequence is shown in Table 5.1.

Microinjection of embryos

For Rluc and Fluc experiment, the injection mixture (3 μ l) was prepared as 50 ng/ μ l of reporter mRNA, 100 ng/ μ l of PylRS mRNA, and 1,000 ng/ μ l of PylT. For the unnatural amino acid incorporation, 0.15 μ l of unnatural amino acid (100 mM stock) was added to 1.5 μ l of injection mixture. Embryos from natural mating were obtained and microinjected at the 1 to 2 cell stage with 2 nL of the injection mixture using a World Precision Instruments Pneumatic PicoPump injector. For Fluc activation experiments, at 48 hpf the embryos were irradiated for 2 minutes with a 365 nm UV transilluminator (8 mW/cm²). For constitutively active MEK1 experiments, 2 nL of 10 ng/ μ l caMEK1 mRNA was injected per embryo. For caged MEK1 experiments, the injection mixture was prepared as 50 ng/ μ l of MEK1-K97TAG mRNA, 100 ng/ μ l of HCKRS mRNA, 1,000 ng/ μ l of PylT, and 2 nL of mixture were injected per embryo. The embryos were irradiated for 30 seconds at 2 h, 5 h or 8 h post injection. Embryos were dechorinated and imaged at 10 hpf using

a stereomicroscope (Leica MZ 16 FA). Embryos irradiated after 5 h post injection were used for whole-mount RNA in situ hybridization. For all injections, zebrafish embryos with early mortality (< 6 hpf) were excluded from further analysis. Zebrafish embryos were randomly assigned to experimental groups during irradiation. Blinding was not used for scoring dorsalized embryos.

Embryo toxicity assay

For toxicity tests of unnatural amino acids, embryos were injected with 2 nL of injection solution (0.15 µl of 100 mM unnatural amino acid in DMSO or just DMSO (vehicle control), 0.75 µl of phenol red (0.5% in Dulbecco's phosphate-buffered saline), and 0.6 µl of water). Live (phenotypically normal) and dead embryos were counted at 24 hpf.

Luciferase assay

Luciferase assays for zebrafish embryos were performed as described previously.²⁰⁹ Briefly, 4 embryos were collected at 48 hpf, washed twice with 1 ml of phosphate-buffered saline (PBS), and were incubated with 50 µl of 1x passive lysis buffer (Promega) for 30 minutes. After incubation, the embryos were manually homogenized with pipette tips, and the extract was centrifuged for 3 minutes at 13,000 rpm to remove cellular debris. For Rluc assays, 10 µl of lysate was mixed with 10 µl of Renilla Luciferase Assay Reagent (Promega), and luminescence was read immediately on a microplate reader (Tecan M1000 PRO), with an integration time of 1,000 ms. For Fluc activation assays, a dual luciferase reporter assay system (Promega) was used. Fluc activity was measured by mixing 10 µl of lysate with 50 µl of LAR II reagent. Rluc activity was measured by adding 50 µL of Stop & Glo reagent. Luminescence was read on a Tecan M1000 PRO microplate reader.

Western blotting

For detection of caged MEK1 expression, zebrafish embryos were collected at 6 h post injection and washed twice with Ringer's solution. For time-course analysis of MEK1 expression and ERK phosphorylation, zebrafish embryos were collected 30 min after UV irradiation at specified time points. The samples were incubated with 80 μ l of 2% SDS buffer for 20 minutes, homogenized manually with pipette tips, heated at 95 °C for 5 minutes, and centrifuged for 1 minute at 13,000 rpm. The protein extract was analyzed by Western blot. Briefly, after gel electrophoresis and transfer to a nitrocellulose membrane (GE Healthcare), the membrane was blocked in Tris-buffered saline with 0.1% Tween 20 and 5% milk powder for one hour. The blots were probed with an anti-phospho-ERK antibody (1:1,000 dilution, Cell Signaling, #9101S), an anti-ERK antibody (1:1,000 dilution, Sigma, #M-7431), an anti-HA antibody (1:1,000 dilution, Cell Signaling, #3724), or an anti- β -actin antibody (1:1,000 dilution, Santa Cruz, #81178) overnight at 4 °C, followed by incubation with secondary antibodies (goat anti-rabbit IgG-HRP for phospho-ERK blot and HA blot, goat anti-mouse for ERK blot and β -actin blot, 1:20,000 dilution) conjugated to horseradish peroxidase (HRP) for 1 hour at room temperature. The blots were further incubated with the SuperSignal West Pico working solution (mixture of the Stable Peroxide Solution and the Luminol/Enhancer Solution, 500 μ l each, Thermo Scientific) for 5 min at room temperature. The luminescence signal was detected by ChemiDoc (setting: Chemi Hi Sensitivity, exposure time: 10 sec).

Whole-mount RNA in situ hybridization

Embryos collected at shield stage (6 hpf) or bud stage (10 hpf) were fixed overnight in 4% paraformaldehyde/PBS at 4 °C. Whole-mount *in situ* hybridization was performed with standard procedures, using described probes for *ta* and *chd*.²¹⁰ Specifically, the fixed embryos were dehydrated in 100% methanol at room temperature for 10 min, followed by incubation with fresh 100% methanol at -20 °C for 30 min. The embryos were then treated with acetone at -20 °C for

8 min, and rehydrated with the following reagents in sequence: 100% methanol for 5 min, 50% methanol (in PBTw, PBS with 0.1% Tween-20) for 5 min, 30% methanol (in PBTw) for 5 min, PBTw for 5 min. The embryos were incubated with hybridization buffer (HYB, 50% formamide, 5 x SSC, 5 mM EDTA, 0.1% Tween-20, 0.1% CHAPS) at 65 °C for 1 h, and then incubated with DIG labeled antisense RNA (2 µl RNA probe in 500 µl of HYB buffer) at 65 °C overnight. The embryos were then washed with the following reagents in sequence: 1) 50% formamide, 2 x SSC, 0.3% CHAPS, 65 °C for 30 min; 2) 2 x SSC, 0.3% CHAPS, 65 °C for 15 min; 3) 0.2 x SSC, 0.3% CHAPS, 65 °C for 20 min; 4) MAB solution (0.1 M maleic acid, 150 mM NaCl), room temperature for 15 min. The embryos were then transferred to the blocking solution (2% blocking reagent (Sigma, #11096176001), 5% lamb serum (Sigma, #S4877), filled with MAB to 50 ml), and were incubated at room temperature for 1 h. The embryos were incubated with anti-digoxigenin-AP (1:2,000 dilution, Roche, #11093274910) for 4 hours at room temperature, and were stained in 500 µl of BM Purple alkaline phosphatase substrate (Roche #11442074001) for 1 hour at room temperature. The embryos were transferred into 1x PBS and imaged using a stereomicroscope (Leica MZ 16 FA). All stained zebrafish embryos were scored without blinding.

5.2 Optical Control of Alk5 Function in Mammalian Cells and Zebrafish

The Transforming Growth Factor beta (TGF- β) pathway regulates a wide range of biological events, including cell differentiation, growth, and apoptosis.²¹¹ The dimeric TGF- β molecule binds to a heterotetrameric complex, consisting of two TGF- β Type I receptors and two TGF- β Type II receptors. Upon ligand binding, the TGF- β Type II receptors are first autophosphorylated, and they further phosphorylate the TGF- β Type I receptors. The TGF- β Type I receptors then activate pathway-restricted Smad2 and Smad3 (Smads that are the substrates of specific receptor), which forms a complex with the common-mediator Smad4 (Smad that is involved in all TGF- β pathways). The Smad complex then functions as a transcription factor and regulates gene expressions.^{211a} The TGF- β pathway plays a critical role in establishing left-right asymmetry in early embryonic development.²¹² However, little is known about the importance of the spatiotemporal characteristic of the signal during the formation of asymmetry.²¹³ To address the timing and positional restrictions on laterality in the zebrafish embryos, we are interested in activating the TGF- β pathway with precise spatial and temporal control. We hypothesized that this could be achieved through the generation of a photocaged, constitutively active Alk5 (containing a T206D mutation)²¹⁴ in zebrafish embryos. Alk5 is a TGF- β Type I receptor, which harbors a crucial lysine residue (K324) in the active site that supports ATP anchoring.²¹⁵ The T206D mutation is proposed to mimic the phosphorylation site and renders the kinase constitutively active.²¹⁶ The caging group will block the Alk5 function, until its removal by light exposure. Light activation in defined regions in early zebrafish embryos will allow precise control of the TGF- β pathway.

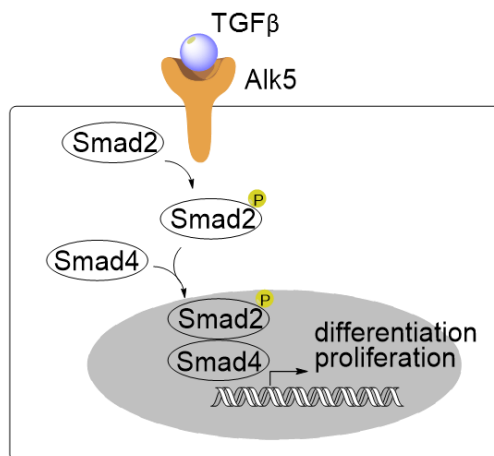


Figure 5.14. Scheme of TGF- β mediated Alk5 signaling pathway. Activated Alk5 phosphorylates Smad2, which further forms the complex with Smad4. The Smad complex functions as transcription factor in the nucleus, regulating gene expressions.

We first sought to confirm the light activation of Alk5 in mammalian cells. We reasoned that the incorporation of a photocaged lysine (HCK) at K234 position might block the activity of Alk5 kinase; identical to the approach we have taken to the optical control of MEK1 (see Chapter 5.1). The removal of the caging group by light then restores the protein activity. We probed the phosphorylation level of Smad2, the downstream target of Alk5, as a readout for the activity.

We first tested if overexpression of the Alk5 kinase increased phosphorylation level of Smad2. While the constitutively active Alk5 expressed well, no phosphorylated Smad2 was observed (Fig 5.15A). We speculated that this was due to the low level of endogenous Smad2 in cells, and increasing the level of Smad2 might facilitate detection. To this end, HEK 293T cells were co-transfected with pAlk5-Myc (a gift from Dr. Beth Roman at Pitt) and pFlag-Smad2 (2,000 ng each). The cell lysate was collected at 48 h for Western blot analysis. This time, we observed elevated level of phosphorylated Smad2 in the presence of constitutively active Alk5 (Fig 5.15B). One caveat in this experiment is that overexpression of Smad2 alone could result in increased phosphorylated Smad2 levels. However, a previous report showed that overexpression of Smad2

by itself did not increase Smad2 phosphorylation in PC3 cells.²¹⁷ Our findings are that the simultaneous overexpression of the constitutively active Alk5 kinase and the Smad2 is required for the detection of phosphorylated Smad2 in HEK293T cells. Thus endogenous Smad2 levels were not expressed at high enough levels for Western Blot detection.

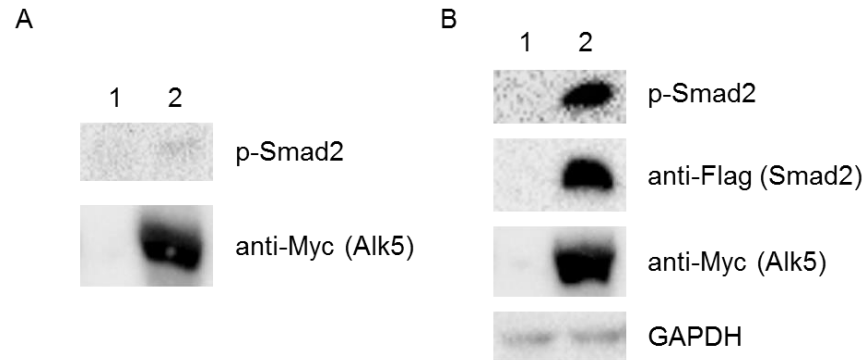


Figure 5.15. Simultaneous overexpression of Alk5 kinase and Smad2 enables detection of phosphorylated Smad2 in HEK 293T cells. (A) Overexpression of Alk5 kinase alone did not increase the phosphorylation level of Smad2. HEK 293T cells were transfected with pAlk5-Myc plasmid (2,000 ng), and the cell lysate was collected at 48 h for western blot analysis (with anti-Myc and anti-pSmad2 antibodies). 1, non-treated cells. 2, pAlk5-Myc transfected cells. (B) Simultaneous overexpression of Alk5 kinase and the Smad2 increases the phosphorylation level of Smad2. HEK 293T cells were transfected with pAlk5-Myc and pFlag-Smad2 plasmids (2,000 ng each), and the cell lysate was collected at 48 h for western blot analysis (with anti-Myc, anti-Flag, anti-pSmad2 and anti-GAPDH antibodies). 1, non-treated cells. 2, pAlk5-Myc and pFlag-Smad2 co-transfected cells.

A TAG mutation was inserted at the K234 position of the Alk5 kinase. HEK 293T cells were co-transfected with pAlk5-K234TAG-Myc, pFlag-Smad2 and pHCKRS-U6-PyIT plasmids (1,500 ng each). The culture medium was supplemented with HCK (1 mM). At 42 h, the culture medium was changed to the serum-free DMEM. At 46 h, the cells were irradiated with UV light (365 nm, VWR) for 2 min and were incubated at 37 °C. After another 2 h, the cell lysate was collected for western blot analysis. We first confirmed the HCK-dependent expression of the full-

length Alk5 protein (Fig 5.16A). An elevated phosphorylation level of Smad2 was observed in response to the UV irradiation (Fig 5.16B). The result suggested that optical activation of the Alk5 function could be achieved through the incorporation of a photocaged lysine at the K234 position. However, we missed an important control experiment where cells in the absence of HCK were also irradiated. This control experiment will be performed in the future, to suggest that UV light itself does not trigger Smad2 phosphorylation.

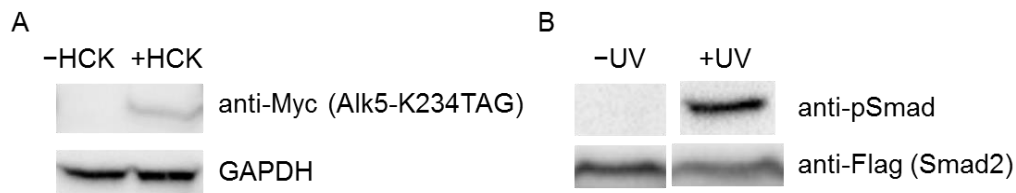


Figure 5.16. Optical activation of Alk5 activity with UV light. HEK 293T cells were co-transfected with pAlk5-K234TAG-Myc, pFlag-Smad2, pHCKRS-U6-PyIT plasmids (1,500 ng each). The culture medium was supplemented with HCK (1 mM). At 42 h, the culture medium was changed to the serum-free DMEM (Dulbecco's Modified Eagle Medium, Gibco). At 46 h, the cells were irradiated with UV light (365 nm, VWR) for 2 min and were incubated at 37 °C. After another 2 h, the cell lysate was collected for western blot analysis (with anti-Myc, anti-Flag, anti-pSmad2 and anti-GAPDH antibodies). For panel A, the cell lysate without UV irradiation was used. For panel B, the cell lysate with or without UV irradiation was used. (A) Incorporation of HCK into the K234 position of Alk5. GAPDH was used as a loading control. (B) UV irradiation activated Alk5 and increased phosphorylation level of Smad2.

We then pursued other molecular readouts for Alk5 activation. Previously, a fluorescent reporter, EGFP-Smad2, was used for probing the activation of the TGF- β pathway.²¹⁸ Upon TGF- β induction, the EGFP-Smad2 will be phosphorylated and transported to the nucleus. We therefore constructed the same reporter (see experimental section for details). To test the translocation assay, HEK 293T cells were transfected with pEGFP-Smad2 (200 ng). After 20 h, the culture medium was changed to the serum-free DMEM. After another 4 h, TGF- β (2 ng/ml) was added to the culture medium, and the time-lapse imaging (every 10 min for a total of 80 min)

was performed for EGFP fluorescence (Zeiss Observer.Z1 Microscope). EGFP (expressed from the pEGFP-C2 plasmid) did not respond to TGF- β , and was used as a negative control. However, we did not observe the nuclear translocation upon TGF- β induction (Fig 5.17). To ensure that TGF- β (2 ng/ml) could activate the TGF- β pathway as expected, the phosphorylation of Smad2 will need to be confirmed by western blot in the future. The expression of Smad2 might be tuned so that the EGFP is not dramatically overexpressed in the cytoplasm, potentially obscuring nuclear translocation events. Previously, the Smad2 translocation was observed in HaCaT cells that stably expressed the EGFP-Smad2.²¹⁸ We might also test this cell line in the future, as HeCaT cells are frequently used in the study of the TGF- β pathway.²¹⁹

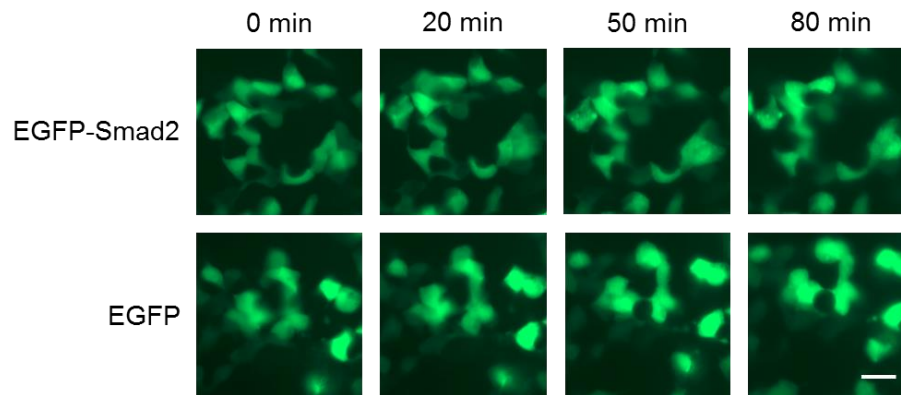


Figure 5.17. Overexpression of Alk5 did not lead to EGFP-Smad2 nuclear translocation in HEK 293T cells. The pEGFP-Smad2 plasmid (200 ng) was transfected into HEK 293T cells. After 20 h, the culture medium was changed to the serum-free DMEM. After another 4 h, TGF- β (2 ng/ml) was added to the culture medium, and the time-lapse imaging (every 10 min for a total of 80 min) was performed for EGFP fluorescence (Zeiss Observer.Z1 Microscope, 20x objective). Scale bar indicates 200 μ m. EGFP did not respond to TGF- β , and was used as a negative control.

The TGF- β pathway undergoes highly complex and context-dependent crosstalk with other signaling pathways, resulting in synergistic or antagonistic effects and eventually desirable biological outcomes.²²⁰ For example, TGF- β can activate non-smad pathways, including the ERK

MAPK pathway.^{211c} We used an ERK-KTR-Clover reporter to monitor the MEK/ERK signaling pathway, where the ERK-KTR-Clover will be transported out of the nucleus upon the pathway activation.²²¹ While we observed the expected translocation between 20 and 30 min upon EGF (100 ng/ml) induction,¹⁴⁸ no translocation was observed upon TGF- β induction (Fig 5.18). Since the kinetics of Erk phosphorylation induced by TGF- β varies with cell types and culture conditions,²²² and the expression of TGF- β receptors differs in cells, the cell lines that present rapid Erk activation (5-10 min) upon TGF- β stimulation might be further pursued.²²³

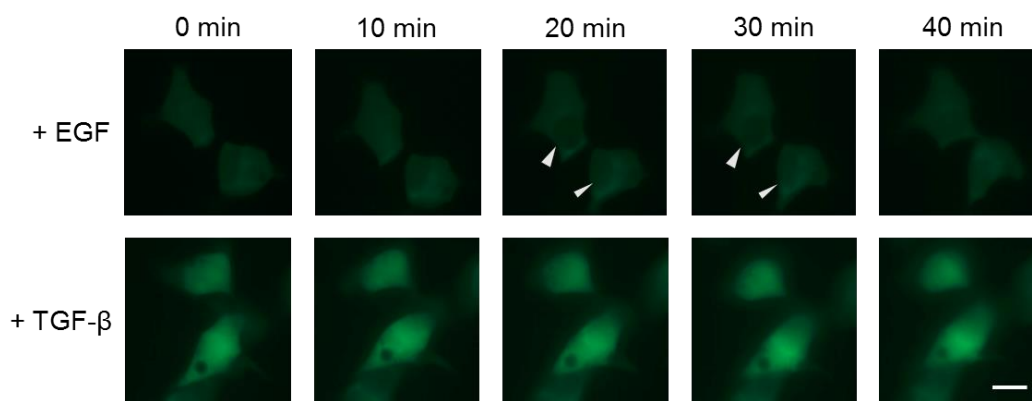


Figure 5.18. TGF- β did not activate MEK/ERK signaling pathway. HEK 293T cells were transfected with pKTR-Clover plasmid (200 ng). After 20 h, the culture medium was changed to the serum-free DMEM. After another 4 h, EGF or TGF- β was added to the culture medium, and the time-lapse imaging (every 10 min for the total of 40 min) was performed for EGFP fluorescence (Zeiss Observer.Z1 Microscope). The scale bar indicates 40 μ m. While EGF successfully activated MEK/ERK signaling pathway, TGF- β did not.

We then shifted our system to the zebrafish. It has been known that activation of the TGF- β pathway in early zebrafish embryos leads to an elongated phenotype.²²⁴ We first prepared the constitutively active Alk5 (caAlk5, with T206D mutation) mRNA from the pCS2-caAlk5 plasmid (gift from Dr. Michael Tsang), and injected it into zebrafish embryos at the 1-2 cell stage. When

the embryos were imaged at 12 hpf, they showed a distinct elongated phenotype as previously reported.²²⁴ We therefore used this phenotypic assay for probing light-activation of Alk5 function.

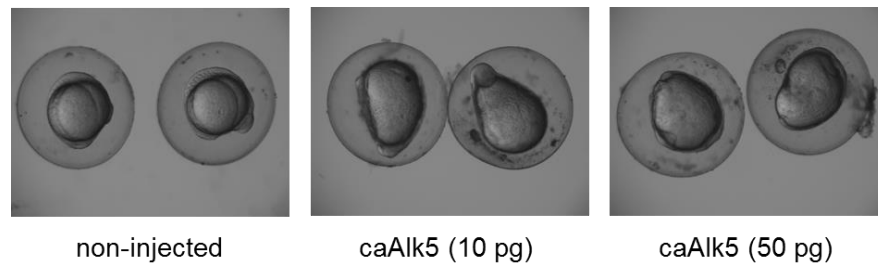


Figure 5.19. Expression of caAlk5 in zebrafish embryos activates the TGF- β pathway, leading to elongated embryos at 12 hpf. Zebrafish embryos were injected with caAlk5 (10 pg or 50 pg per embryo) at the 1-2 cell stage. Zebrafish were imaged at 12 hpf.

To this end, we generated a TAG mutation at the position K234 of caAlk5. Alk5-K234TAG mRNA, HCKRS mRNA, PylT, and HCK were injected into zebrafish embryos at the 1-2 cell stage. The RNA quality was confirmed by agarose gel (Fig 5.20).

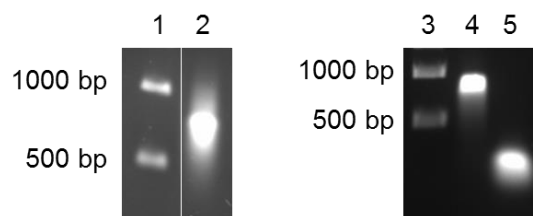


Figure 5.20. Agarose gel images confirming RNA quality. The lanes indicate: 1. 1 kb ladder, 2. Alk5-K234TAG mRNA, 3. 1 kb ladder, 4. HCKRS mRNA, 5. PylT.

Short UV irradiation (1 min or 2 min, 365 nm) was performed on live embryos at the indicated time points (2 h, 3.5 h or 6 h post injection). Normal and elongated embryos were counted at 12 hpf. While we observed that some embryos in the UV-irradiated group were elongated, there were some elongated embryos in the non-irradiated group as well (Fig 5.21). UV

light had no effect on the phenotype of non-injected embryos. To further study whether light treatment led to the Alk5 activation, we decided to detect the change in gene expression level.

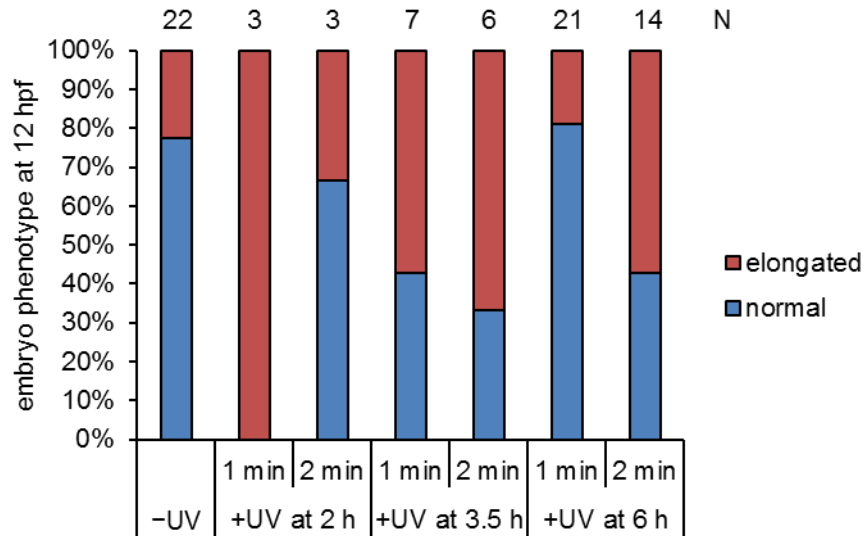


Figure 5.21. Phenotypic assay for probing light-activation of Alk5 function in zebrafish. Alk5-K234TAG mRNA, HCKRS mRNA, PylT, and HCK were injected into zebrafish embryos at the 1-2 cell stage. Short UV irradiation (1 min or 2 min, 365 nm) was performed on live embryos at the indicated time points (2 h, 3.5 h or 6 h post injection). Normal and elongated embryos were counted at 12 hpf. N indicates the total number of embryos.

Using the standard in-situ hybridization technique,²¹⁰ we probed expression of the *goosecoid* (*gsc*) gene and the *brachyury homolog a* (*ta*) gene, two well-known downstream targets of the TGF- β pathway.²²⁴⁻²²⁵ Zebrafish embryos were injected with caAlk5 mRNA at the 1-2 cell stage. The embryos were fixed with 4% paraformaldehyde at 7 hpf, and the gene expression patterns were detected by in-situ hybridization. Basal expression of *gsc* gene was detected in non-injected embryos, and an ectopic lateral expression for *gsc* was observed in embryos injected with caAlk5 mRNA, as previously reported.²²⁶ The *ta* gene was expressed throughout the margin at the shield stage in non-injected embryos,²²⁷ and caAlk5 mRNA injection induced an expanded expression towards the animal pole (Fig 5.22), which is in agreement with a previous report.²²⁶

These results demonstrate that activation of TGF- β pathway by Alk5 induces elevated expression of *gsc* and *ta* genes in zebrafish embryos.

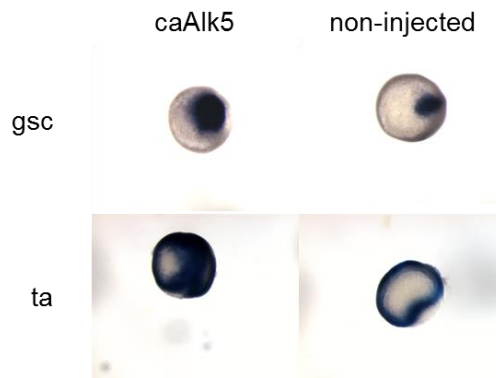


Figure 5.22. Expression of caAlk5 in zebrafish embryos activates downstream targets, *gooseoid* (*gsc*) and *brachyury homolog a* (*ta*). Zebrafish embryos were injected with caAlk5 mRNA at the 1-2 cell stage. The embryos were fixed with 4% paraformaldehyde at 7 hpf. Gene expression patterns were detected by in-situ hybridization. Dorsal views for *gsc*, and animal pole views for *ta*.

We then studied the change of gene expression patterns with light activation of the photocaged Alk5. Zebrafish embryos were injected with Alk5-K234TAG mRNA, HCKRS mRNA, PyIT, and HCK at the 1-2 cell stage. The embryos were irradiated with UV light (365 nm) for 1 min at 3.5 h post injection, and were further incubated at 28.5 °C. The irradiation time point was chosen as an initial single-point to study the gene expression upon optical Alk5 activation at 3.5 h post injection. The embryos were fixed with 4% paraformaldehyde at 7 hpf, and the gene expression patterns (*gsc* and *ta*) were probed by in-situ hybridization. In the –HCK group, embryos presented basal expression of *gsc* and *ta* genes, either with or without UV irradiation (Fig 5.23). This result demonstrates that UV exposure (1 min, 365 nm) does not activate the TGF- β pathway. In embryos expressing photocaged Alk5, we observed over-expression of both *gsc* and *ta* genes, either with or without UV treatment (Fig 5.23). Their expression patterns were similar to those of embryos injected with caAlk5 mRNA, indicating that the photocaged Alk5 was

active even without the UV irradiation. We speculated that this might also be the reason for elongated embryos in the non-irradiated group (Fig 5.21).

We then asked if tuning the expression level of the photocaged Alk5 could avoid light-independent activation of the TGF- β pathway. One possibility is that the caged Alk5 dimerizes with the functional endogenous Alk5, thus activating the TGF- β pathway in the absence of light. To this end, we performed the injection and the in-situ hybridization as described above, except that we decreased the Alk5-K234TAG mRNA amount from 100 pg per embryo to 40 or 10 pg per embryo. However, over-expression of the *gsc* gene was still observed in both non-irradiated and irradiated embryos (Fig 5.23C). The test of even lower mRNA amount (for example, 1 pg) might be further pursued.

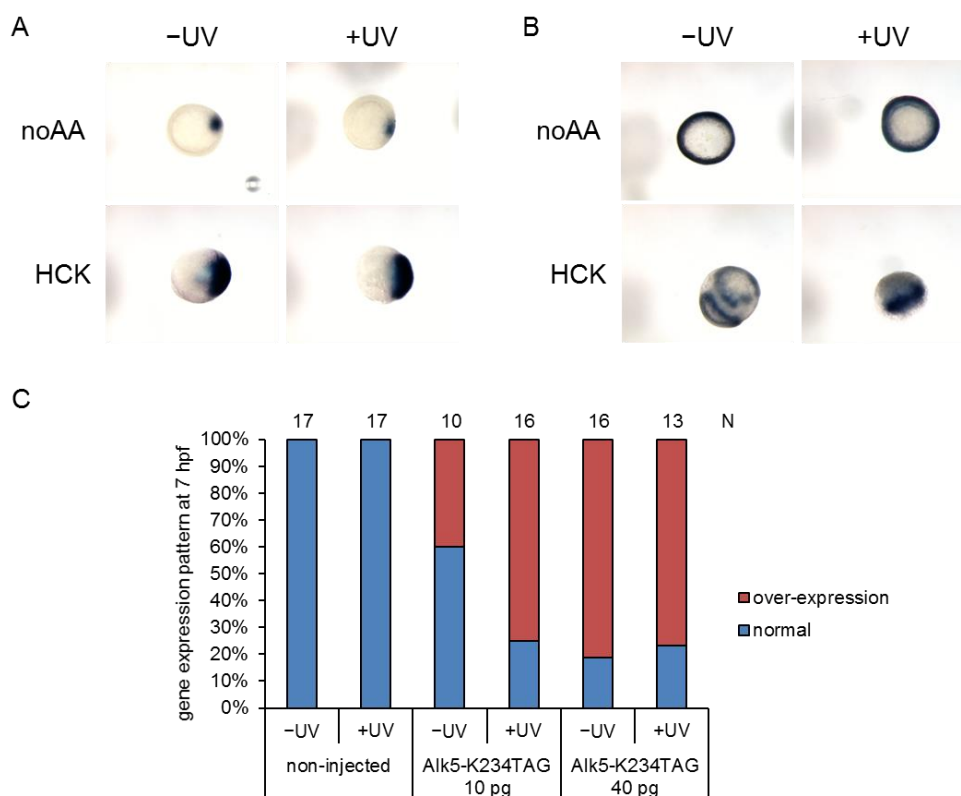


Figure 5.23. Photocaged Alk5 activates gene expression even without the UV irradiation. Zebrafish embryos were injected with Alk5-K234TAG mRNA (100 pg in A and B, 10 pg or 40 pg in C), HCKRS mRNA, PylT, and HCK at the 1-2 cell stage. The embryos were irradiated with UV light (365 nm) for 1 min at 3.5 h post injection, and were further incubated at 27 °C. The embryos were fixed with 4% paraformaldehyde at 7 hpf. Gene expression patterns were probed by in-situ hybridization. (A) Expression of *goosecoid* (*gsc*) gene. Animal views with dorsal to the right. (B) Expression of *brachyury homolog a* (*ta*) gene. Animal views with dorsal to the right. (C) The expression pattern of the *goosecoid* (*gsc*) gene. N indicates the total number of embryos. The percentage of normal and over-expression was inferred by the number of zebrafish with corresponding expression patterns shown in A and B.

The light-independent activity of the photocaged Alk5 was unexpected, because experiments in mammalian cells indicated that phosphorylation of Smad2, the downstream target of Alk5, was only induced with light exposure. To rule out the possibility that the photocaged Alk5

was unintentionally degraded by ambient light during the experiment, we further incorporated Alloc Lys into the position K234 of Alk5, which should block the Alk5 function regardless of light. However, over-expression of the *gsc* gene was observed in embryos expressing the Alk5-K234TAG-Alloc Lys (Fig 5.24). To determine whether inactive Alk5 could activate the TGF- β pathway, we further tested the activity of kinase-deficient Alk5 (with K234R mutation). Similar to the experiment in mammalian cells,²²⁸ Alk5-K234R was inactive in zebrafish embryos (Fig 5.24). The result rule out the possibility that inactive Alk5 could still activate the TGF- β pathway.

In summary, we potentially achieved optical activation of Alk5 function in mammalian cells, through the genetic incorporation of a photocaged lysine, thus allowing for the control of TGF- β pathway with light. The experiments need to be repeated though and appropriate controls need to be included. We planned to test the nuclear translocation of Smad2 upon light activation, but the EGFP-based translocation assay was not successful and will need further characterization. We further shifted this system to zebrafish. However, we observed unexpected activation of the TGF- β pathway with the incorporation of HCK or Alloc-Lys at position K234 of Alk5. Surprisingly, the light-independent activation was not observed in mammalian cells. In the future, additional control experiment (e.g., effect of UV light alone on TGF- β activation) and more characterization (activity, decaging efficiency) of the photocaged Alk5 enzyme in mammalian cells will need to be performed.

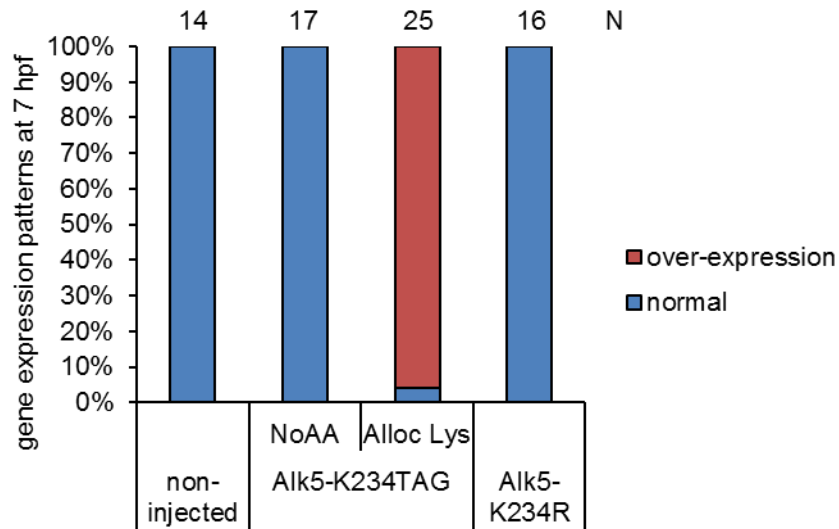


Figure 5.24. Alk5 was active with the incorporation of Alloc Lys at the position K234, while Alk5-K234R mutant was inactive. For Alloc Lys incorporation, zebrafish embryos were injected with Alk5-K234TAG mRNA, WTRS mRNA, PylT, and Alloc Lys at the 1-2 cell stage. For Alk5-K234R mutant, zebrafish embryos were injected with Alk5-K234R mRNA at the 1-2 cell stage. The embryos were fixed with 4% paraformaldehyde at 7 hpf. Expression pattern of *goosecoid* (*gsc*) gene was probed by in-situ hybridization. N indicates the total number of embryos. Percentages of normal and over-expression patterns were calculated based on the counts.

Experimental

Plasmid construction

The K234R mutation was introduced on pCS2-caAlk5 (gift from Dr. Michael Tsang) using site-directed mutagenesis (Fwd: 5'-ggagaggaggtggcgggtgcggatcttctcctccagagagg-3', Rev: 5'-cctctctggaggagaagatccgcaccgccacctcctctcc-3', arginine mutation underlined).

The pCS2-Alk5-Myc plasmid (constitutively active, with T206D mutation) was from Dr. Michael Tsang. Since TAG was used as the stop codon in the original plasmid, it was first mutated to TGA by site-directed mutagenesis (Fwd: 5'-cctctcgagcctctagaactatgatgagtcgtattacgtagatcc-3', Rev: 5'-ggatctacgtaatacgaactcatcatagtcttagaggctcgagagg-3', TGA mutation underlined). The TAG

mutation was then introduced at the K234 position by site-directed mutagenesis (Fwd: 5'-ggagaggaggtggcgggtgtagatcttctcctccagagagg-3', Rev: 5'-cctctctggaggagaagatctacaccgccacctctctcc-3', TAG mutation underlined).

The pFlag-Smad2 plasmid was from Addgene (#14042). To construct the pEGFP-Smad2 plasmid, the pEGFP-C2 plasmid (8,000 ng, Clontech) was first digested with XhoI and PstI restriction enzymes, followed by gel purification of the band corresponding to the correct length of the backbone (~4,500 bp). Smad2 was amplified from the pFlag-Smad2 plasmid (Fwd: 5'-atcgctcgagcttcgtccatcttgccatt-3', Rev: 5'-acgtctgcagatgacatgcttgagcaacgcac-3'), and digested with XhoI and PstI restriction enzymes (37 °C, 2 h). The pEGFP-Smad2 plasmid was constructed by ligating the Smad2 fragment into the pEGFP-C2 backbone using the XhoI and PstI restriction sites (T4 ligase, 16 °C, overnight).

Western blot

For the wild type Alk5 activity assay, HEK 293T cells were transfected with pAlk5-Myc and pFlag-Smad2 (2,000 ng each), and the cell lysate was collected at 48 h for western blot analysis. For the incorporation of HCK into the K234 position of Alk5, HEK 293T cells were co-transfected with pAlk5-K234TAG-Myc, pFlag-Smad2 and pHCKRS-U6-PyIT plasmids (1,500 ng each). The culture medium was supplemented with HCK (1 mM). Branched polyethylene imine (bPEI, 15 µl, 1 mg/ml) was used as the transfection reagent. At 42 h, the culture medium was changed to the serum-free DMEM. At 46 h, the cells were irradiated with UV light (365 nm, VWR) for 2 min, and were further incubated at 37 °C incubator. After another 2 h, the cell lysate was collected for western blot analysis. The cell lysates (in mammalian protein extraction buffer (250 µl, GE Healthcare)) were separated by 10% SDS-PAGE (run with 60V for 15 min, and 150 V for 45 min) and were transferred to a PVDF membrane (GE Healthcare). The membrane was blocked in tris-buffer saline (TBS) with 0.1% Tween 20 and 5% milk for 1 h. The blots were probed with the primary antibody (1:1,000, anti-GAPDH (sc-25778, Santa Cruz), anti-Myc (sc-40, Santa Cruz),

anti-Flag (F1804-200UG, Sigma), anti-pSmad (3108P, Cell Signaling)) overnight at 4 °C, followed by incubation with secondary goat anti-rabbit IgG-HRP antibody (1:20,000, sc-2004, Santa Cruz) for 1 h at room temperature. The blots were further incubated with the SuperSignal West Pico working solution (mixture of the Stable Peroxide Solution and the Luminol/Enhancer Solution, 500 µl each, Thermo Scientific) for 5 min at room temperature. The luminescence signal was detected by ChemiDoc (Chemi Hi Sensitivity setting, exposure time: 10 sec).

Translocation assay in live cells

For EGFP translocation assay, HEK 293T cells were transfected with pEGFP-Smad2 (200 ng). Branched polyethylene imine (bPEI, 1.5 µl, 1 mg/ml) was used as the transfection reagent. After 20 h, the culture medium was changed to the serum-free DMEM. After another 4 h, TGF-β (2 ng/ml, #NC0779906, Fisher) was added to the culture medium, and the time-lapse imaging (every 10 min for a total of 80 min) was acquired on a Zeiss Observer Z1 microscope, with EGFP channel (filter set 38 HE; ex. BP470/40; em. BP525/50).

For KTR-Clover translocation assay, HEK 293T cells were transfected with ERK-KTR-Clover reporter (gift from Dr. Markus Covert, 200 ng).²²¹ Branched polyethylene imine (bPEI, 1.5 µl, 1 mg/ml) was used as the transfection reagent. After incubation at 37 °C for 20 h, the culture medium was changed to the serum-free DMEM. After another 4 h of incubation, either TGF-β (2 ng/ml) or EGF (100 ng/ml, #AF-100-15, PeproTech) was added to the culture medium, and the time-lapse imaging (every 10 min for a total of 80 min) was acquired on a Zeiss Observer Z1 microscope, with EGFP channel (filter set 38 HE; ex. BP470/40; em. BP525/50).

mRNA synthesis

The caAlk5, Alk5-K234TAG, Alk5-K234R mRNA were generated as described in chapter 5.1, using the pCS2-caAlk5, pCS2-Alk5-K234TAG, pCS2-Alk5-K234R plasmids.

Microinjection of embryos

For caAlk5 experiments, 2 nL of 5 ng/μl (or 25 ng/μl) caAlk5 mRNA was injected per embryo.²²⁹ For caged Alk5 experiments, the injection mixture (3 μl) was prepared as 50 ng/μl (20 ng/μl or 5 ng/μl for dose-response study) of Alk5-K234TAG mRNA, 100 ng/μl of HCKRS mRNA, and 1,000 ng/μl of PylT. For the UAA incorporation, 0.15 μl of UAA (HCK or Alloc Lys, 100 mM stock in 100% DMSO) was added to 1.5 μl of injection mixture. Embryos from natural mating were obtained and microinjected at the 1 to 2 cell stage with 2 nL of the injection mixture using a World Precision Instruments Pneumatic PicoPump injector. For light activation experiments, the embryos were irradiated for either 1 min or 2 min with a 365 nm UV transilluminator (8 mW/cm²) at 2 h, 3.5 h, or 6 h post injection. Embryos were dechorionated and imaged at 12 hpf using a stereomicroscope (Leica MZ 16 FA). Normal and elongated embryos were also counted.

Whole-mount RNA in situ hybridization

Embryos collected at 7 hpf were fixed overnight in 4% paraformaldehyde/PBS at 4 °C. Whole-mount in situ hybridization was performed with standard procedures (detailed experimental in chapter 5.1), using described probes for *gsc* and *ta*.²¹⁰ The embryos were incubated with anti-digoxigenin-AP (1:2,000 dilution, Roche #11093274910) for 4 hours at room temperature, and were stained in 500 μl of BM Purple alkaline phosphatase substrate (Roche #11442074001) for 1 hour at room temperature. The embryos were transferred into 1x PBS and imaged using a stereomicroscope (Leica MZ 16 FA).

5.3 Optical Control of Cre Recombinase Function in Zebrafish

Besides temporal control of protein function in zebrafish embryos, as demonstrated by activation of photocaged MEK1, light also allows for spatial control in living animals. To explore the ability to control protein function in specific regions (or cells), we sought to generate a photocaged Cre recombinase in zebrafish. Previously, we have demonstrated that the incorporation of a photocaged lysine into Cre recombinase at the critical K201 position allows for light-activation of DNA recombination in mammalian cells.¹⁶⁰ Moreover, the *ubi:loxP-GFP-loxP-mCherry* lineage tracer transgenic zebrafish, which provide strong reporter activity upon Cre exposure, have been established and widely used.²³⁰ The Cre-dependent DNA recombination switches the EGFP fluorescence to the mCherry fluorescence, by removing the EGFP gene from the genome, thereby functioning as an excellent reporter for spatial activity of Cre recombinase. Previously, both temporal²³¹ and spatial²³² control of Cre recombinase activity have been reported in zebrafish embryos, based on ligand-inducible CreER^{T2} fusion protein. A photoactivable Cre recombination system has been applied in live mice.²³³ However, no light-mediated spatiotemporal control of Cre recombinase activity has been achieved in zebrafish so far. Our method to expand the toolbox for conditional site-specific recombination in zebrafish will further facilitate the study of physiological processes in zebrafish.

To this end, we introduced the K201TAG mutation into the pCS2-Cre-WT plasmid, and prepared Cre-K201TAG mRNA. The RNA quality was confirmed by the agarose gel (Fig 5.25).

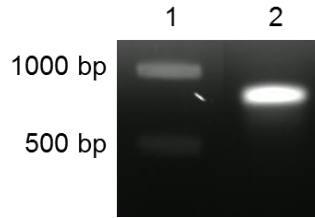


Figure 5.25. Agarose gel images confirming RNA quality. The lanes indicate: 1. 1 kb ladder, 2. Cre-K201TAG mRNA.

Transgenic zebrafish embryos (*ubi:loxP-EGFP-loxP-mCherry*) were injected with Cre-K201TAG mRNA, HCKRS mRNA, PylT, and HCK at the 1-2 cell stage. At 6 hpf (selected based on previous experience with caged MEK1), embryos were irradiated with UV light (365 nm) for 30 sec, and were further incubated at 27 °C. At 24 hpf, embryos were imaged with EGFP, mCherry, and bright field channels. For non-injected fish, EGFP fluorescence, but not mCherry fluorescence, was detected, suggesting the correct functionality of the transgenic fish line (Fig 5.26). Injection of the wild type Cre mRNA switched the fluorescence from EGFP to mCherry, indicating that the Cre recombinase was active in zebrafish (Fig 5.26). The zebrafish did not develop normally, probably due to the high injection dose of the wild type Cre mRNA (200 pg per fish). Lower injection dose (50 pg per fish) will be tested in the future to obtain phenotypically normal zebrafish. Zebrafish embryos that were irradiated with UV light showed mCherry expression, while non-irradiated zebrafish did not (Fig 5.26). This result suggested that global activation of the Cre recombinase activity could be achieved with UV light on zebrafish. In the future, we plan to perform local irradiation, and test the spatial control of Cre recombinase activity on zebrafish.

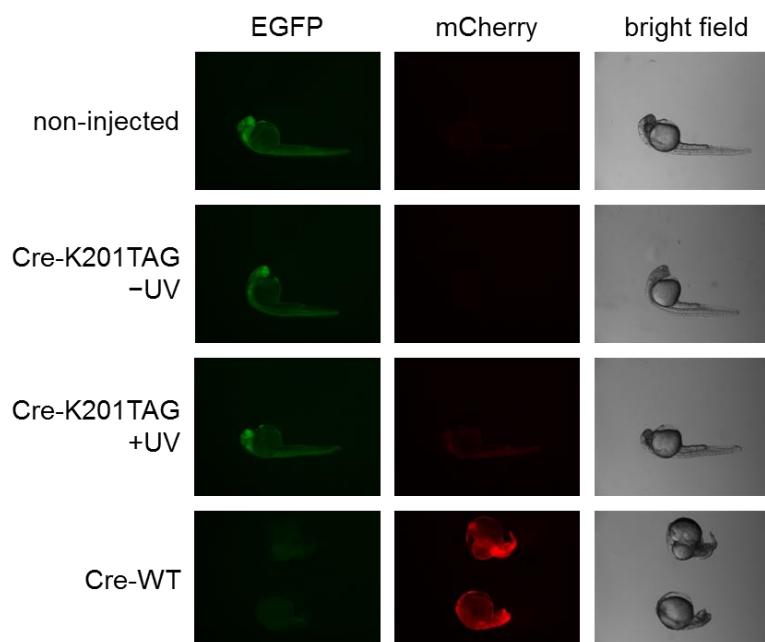


Figure 5.26. Light-activation of Cre recombinase activity in zebrafish. The Cre-K201TAG mRNA, HCKRS mRNA, PyIT, and HCK were injected into the transgenic zebrafish embryos (*ubi:loxP-EGFP-loxP-mCherry*) at the 1-2 cell stage. At 6 hpf, embryos were irradiated with UV light (365 nm) for 30 sec. At 24 hpf, embryos were imaged with EGFP, mCherry, and bright field channels. The wild type Cre (Cre-WT) was used as a positive control.

Experimental

Plasmid construction

The pCS2-Cre-WT plasmid was generated by Jie Zhang in our lab. The K201TAG mutation was introduced on pCS2-Cre-WT using site-directed mutagenesis (Fwd: 5'-gttaatccatattggcagaacgtagacgctggtagcaccgcagg-3', Rev: 5'-cctgcggtgctaaccagcgctctacgttctgccaatatggattaac-3', TAG mutation underlined).

mRNA synthesis

The Cre-WT, Cre-K201TAG mRNA were generated as described in Chapter 5.1, using the pCS2-Cre-WT, and pCS2-Cre-K201TAG plasmids.

Microinjection of embryos

For Cre recombinase experiments we used the transgenic zebrafish lines (ubi:loxP-EGFP-loxP-mCherry, from Dr. Donghun Shin).²³⁰ To express wild type Cre recombinase, 2 nL of 100 ng/μl Cre-WT mRNA was injected per embryo. To express photocaged Cre, the injection mixture (3 μl) was prepared as 50 ng/μl of Cre-K201TAG mRNA, 100 ng/μl of HCKRS mRNA, and 1,000 ng/μl of PyIT. For the HCK incorporation, 0.15 μl of HCK (100 mM stock in 100% DMSO) was added to 1.5 μl of injection mixture. Embryos from natural mating were obtained and microinjected at the 1 to 2 cell stage with 2 nL of the injection mixture using a World Precision Instruments Pneumatic PicoPump injector. For light activation experiments, at 6 hpf the embryos were irradiated for 30 sec (based on previous MEK1 results) with a 365 nm UV transilluminator (8 mW/cm²). At 24 hpf, the embryos were imaged with a Leica M205 FA microscope with EGFP, mCherry, and bright field channels, with a 0.63x objective.

5.4 Small Molecule Control of Protein Function in Zebrafish

Small molecules have been widely applied as modulators of protein function in zebrafish.^{193, 234} With successful generation of light-activatable proteins (MEK1, Cre, and luciferase) through UAA mutagenesis in zebrafish embryos, we asked if small molecule-induced protein activation could be achieved. The ability to control protein function with small molecules in zebrafish will greatly expand our toolbox to study zebrafish biology. Previously, several approaches for small molecule-induced protein activation have been developed in mammalian cells, using UAA mutagenesis (discussed in detail in Chapter 3.5).^{61, 157b} For example, our lab reported activation of protein function by a phosphine-mediated Staudinger reduction, through the genetic incorporation of an *ortho*-azidobenzoyloxycarbonyl amino acid (OABK).⁶¹ With successful incorporation of OABK into proteins in zebrafish (discussed in Chapter 5.1), we decided to test phosphine-triggered protein activation in zebrafish.

We first tried Fluc activation assay using the Fluc-K206TAG-Rluc construct, as described in Chapter 5.1 for the activation of photocaged lysine (HCK). Zebrafish embryos were injected with Fluc-K206TAG-Rluc mRNA, OABKRS mRNA, PyIT, and OABK at the 1-2 cell stage. At 24 hpf, zebrafish embryos were transferred to fish E3 medium (5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl₂, 0.33 mM MgSO₄) containing 2DPBM (final concentration of 500 μ M, from 100 mM stock solution dissolved in 100% DMSO). The zebrafish embryos were incubated for another 4 h at 27 °C, before the fish lysates were collected for Fluc and Rluc activity assays. The Rluc activities were detected, indicating successful incorporation of OABK. However, no Fluc activity was observed in response to treatment with 2DPBM. We speculated that the penetration of 2DPBM into the zebrafish might be inefficient.

In summary, we incorporated OABK into the K206TAG position of Fluc, but the activation of Fluc function was unsuccessful, presumably due to inefficient penetration of the phosphine. In

the future, we will first test the Fluc activation in zebrafish lysate, to avoid the possible penetration issue of 2DPBM. We will further test the Fluc activation *in vivo* by direct injection of the phosphine into the zebrafish yolk sack (as demonstrated by Luis Angel Vázquez in our lab for an unrelated project).

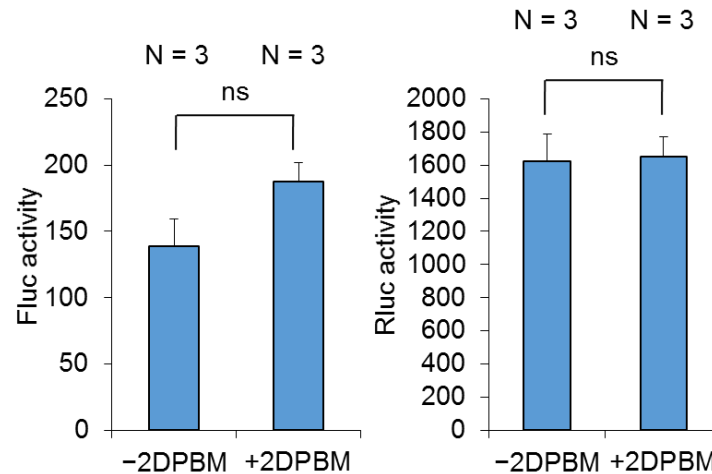


Figure 5.27. Fluc activation assay with the incorporation of OABK into Fluc-K206TAG-Rluc in zebrafish embryos. Zebrafish embryos were injected with Fluc-K206TAG-Rluc mRNA, OABKRS mRNA, PylT, and OABK at the 1-2 cell stage. At 24 hpf, zebrafish embryos were treated with fish E3 water containing 2DPBM (500 μ M) for 4 h. The fish lysates were collected for Fluc and Rluc activity assays. N indicates the number of pooled samples (4 embryos each). Statistical significance is indicated by ns: not significant (unpaired t-tests)

Experimental

Microinjection of embryos

The injection mixture (3 μ l) was prepared as 50 ng/ μ l of Fluc-K206TAG-Rluc mRNA, 100 ng/ μ l of OABKRS mRNA, and 1,000 ng/ μ l of PylT. For the OABK incorporation, 0.15 μ l of OABK (100 mM stock in 100% DMSO) was added to 1.5 μ l of injection mixture. Embryos from natural mating were obtained and microinjected at the 1 to 2 cell stage with 2 nL of the injection mixture using a World Precision Instruments Pneumatic PicoPump injector. At 24 hpf, zebrafish embryos were removed

to fish E3 medium (5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl₂, 0.33 mM MgSO₄) containing 2DPBM (500 µM, dissolved in 100% DMSO). Zebrafish embryos were incubated for another 4 h at 27 °C. At 28 hpf, zebrafish lysates were collected, and Fluc and Rluc activities were measured as described in Chapter 5.1.

5.5 Optical Control of CRISPR/Cas9 Gene Editing in Zebrafish

Many bacteria utilize clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated (Cas) systems to silence foreign nucleic acids.²³⁵ Bacterial type II CRISPR systems have been adapted to create a single guide RNA (gRNA),²³⁶ and have further been applied for direct site-specific DNA cleavage in cultured cells²³⁷ and animals.²³⁸ Optical regulation of Cas9 activity allows spatial and temporal control of gene editing. For example, photoactivable Cas9 was generated by splitting Cas9 into two fragments and attaching each of them to a photoinducible dimerization domain.²³⁹ Our lab also created a light-activatable Cas9 through genetic incorporation of a photocaged lysine at a critical lysine residue on Cas9.²⁴⁰ The caging group blocks the CRISPR/Cas9 function, which was later restored upon removal of the caging group by UV irradiation. By screening several lysine residues that were placed in close proximity to the gRNA nucleic acid binding sites, the K866 position was found to allow most efficient activation of Cas9 function upon UV irradiation.²⁴⁰ This method was further applied for optical deactivation of exogenous and endogenous gene function.²⁴⁰

While optical control of gene editing in mammalian cells holds promise to facilitate the understanding of gene interactions and networks, expansion of this technology to animal models will create broader impact on the study of complex biological questions and human-related diseases. We therefore asked if optical control of gene editing could be adapted to the zebrafish system. While CRISPR/Cas9 has become a powerful tool for the induction of genetic modifications in zebrafish embryos,^{238b} no optical control of this process has been reported to date. We first sought to confirm the genome modification of zebrafish embryos with the Cas9 mRNA injection. We utilized an enzyme digestion-based method (developed by Manush Sayd Mohammed in Dr. Michael Tsang lab) to detect Cas9 function. Briefly, the gRNA was designed to target a NotI restriction site within the endogenous centrosomal protein 290 (cep290) gene in

zebrafish.²⁴¹ If Cas9 efficiently modifies the NotI site, the PCR-amplified cep290 DNA will be resistant to NotI digestion. Otherwise, the cep290 DNA will be digested into two fragments by NotI.

To this end, Cas9 mRNA was generated through *in vitro* transcription. Cep290 gRNA was generated by Manush Sayd Mohammed. Zebrafish embryos were injected with the wild-type Cas9 mRNA (200 pg per embryo), cep290 gRNA (200 pg or 500 pg per embryo) at the 1-cell stage. Zebrafish embryos were collected (three fish) at 48 hpf. The Cep290 region was amplified from genomic DNA, followed by digestion with NotI restriction enzyme (3 h, 37 °C). The digestion products were analyzed on a 2% agarose gel. In non-injected zebrafish embryos, the cep290 region was successfully amplified (281 bp). Two fragments (174 bp and 107 bp) were generated upon NotI digestion, suggesting that the NotI recognition site was unmodified and was digested by NotI (Fig 5.28). In zebrafish embryos injected with wild-type Cas9 mRNA and cep290 gRNA, a distinct undigested product was present with NotI treatment (Fig 5.28). This result demonstrated successful genome modification with the Cas9 mRNA injection in zebrafish embryos.

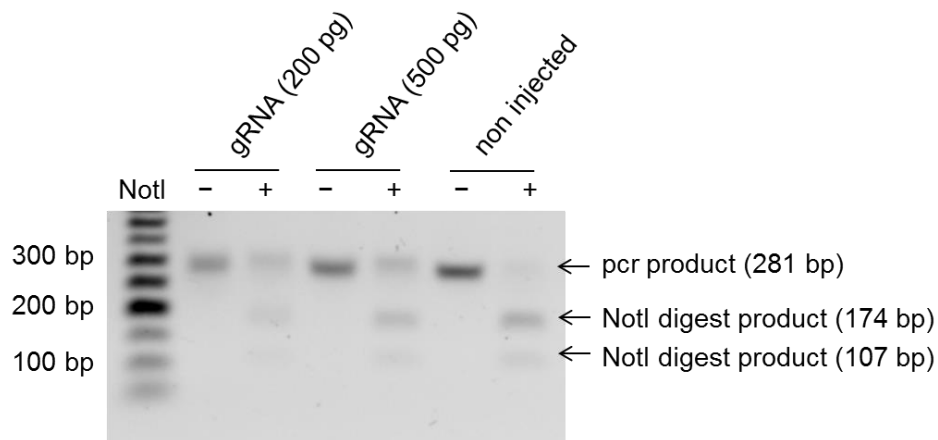


Figure 5.28. Genome modification with the Cas9 mRNA injection in zebrafish embryos. Zebrafish embryos were injected with wild-type Cas9 mRNA (200 pg per embryo), cep290 gRNA (200 pg or 500 pg per embryo) at the 1-cell stage. Zebrafish embryos were collected (three fish) at 48 hpf. The Cep290 region was amplified from genomic DNA, followed by digestion with NotI restriction enzyme. The digestion products were analyzed on a 2% agarose gel. CRISPR/Cas9 modified the NotI recognition site, thus resulting in undigested product with NotI treatment.

Using the UAA mutagenesis in zebrafish (discussed in Chapter 5.1), we incorporated a photocaged lysine (HCK) into the K866 position of Cas9. Zebrafish embryos were injected with Cas9-K866TAG mRNA, HCKRS mRNA, PylT, cep290 gRNA and HCK at the 1-cell stage. At 6 hpf, zebrafish embryos were irradiated with UV light (365 nm, 30 sec), and were further incubated at 27 °C. At 48 hpf, zebrafish embryos were collected (three fish) for Cas9 activity assay described above. However, for UV-irradiated zebrafish samples, we did not observe undigested product with NotI treatment (Fig 5.29). The result suggested that no functional CRISPR/Cas9 was generated upon UV irradiation, presumably because of low expression levels of photocaged Cas9. However, a Western blot analysis of caged Cas9 expression was not conducted. We therefore further pursued the direct injection of Cas9/gRNA ribonucleoprotein complexes (Cas9 RNPs), followed by light activation.

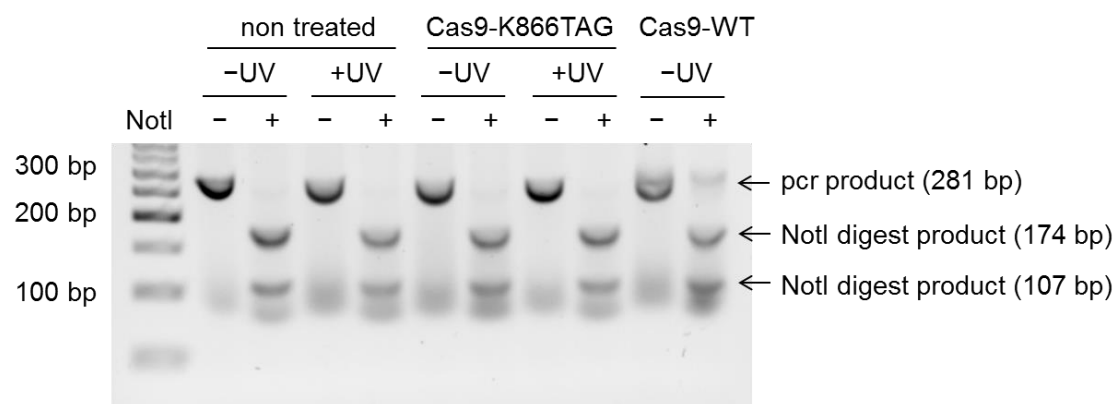


Figure 5.29. Light activation of Cas9 function through the genetic incorporation of a photocaged lysine (HCK). Zebrafish embryos were injected with Cas9-K866TAG mRNA, HCKRS mRNA, PylIT, cep290 gRNA and HCK at the 1-cell stage. At 6 hpf, zebrafish embryos were irradiated with UV light (365 nm, 30 sec), and were further incubated at 27 °C. At 48 hpf, zebrafish embryos were collected (three fish) for activity assay as described above. Zebrafish embryos injected with wild-type Cas9 mRNA and cep290 gRNA were used as the positive control.

Cas9 RNPs, consisting of purified Cas9 protein in complex with a gRNA, are capable of gene editing with similar efficiency as compared to plasmid-based expression of Cas9/gRNA.²⁴² Direct injection of Cas9 RNPs has shown efficient genome modifications in zebrafish embryos, and in some cases, even higher modification efficiency compared to the Cas9 mRNA injection.²⁴³ We tested the efficacy of two Cas9 proteins: NLS-Cas9, expressed recombinantly from *E. coli* (experiment performed by Wenyuan Zhou in our lab), and NLS-Cas9-NLS, purchased from NEB (# M0646T). Both Cas9 proteins were fused with a nuclear localization signal (NLS) at the N-terminus, but the protein from NEB also contained another NLS at the C-terminus. The Cas9 protein (250 ng/μl) was mixed with the cep290 gRNA (100 ng/μl) *in vitro* for 10 min. Zebrafish embryos were injected with the Cas9/gRNA ribonucleoproteins (RNPs) at the 1-cell stage. At 48 hpf, zebrafish embryos were collected (three fish per tube) for activity assay as described above. Gratifyingly, we observed almost 100% undigested PCR product with NotI treatment (Fig 5.30).

This result indicated efficient genome modification with the injection of Cas9 RNPs in zebrafish embryos. The efficiency was higher than that of Cas9 mRNA injection, which is in agreement with a previous report.²⁴⁴

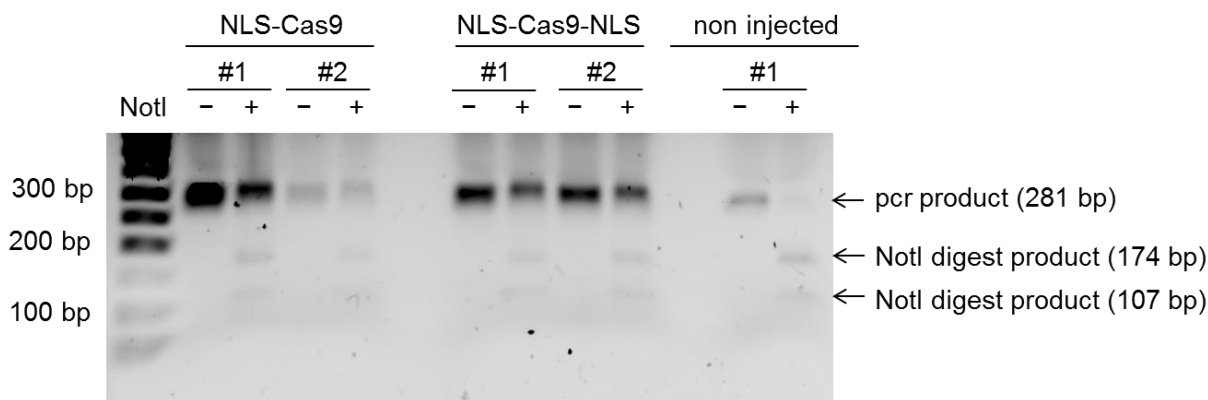


Figure 5.30. Genome modification of zebrafish embryos with Cas9 RNPs injection. NLS-Cas9 was recombinantly expressed from *E. coli*. NLS-Cas9-NLS was purchased from NEB. The Cas9 protein (250 ng/μl) was mixed with the cep290 gRNA (100 ng/μl) *in vitro* for 10 min. Zebrafish embryos were injected with Cas9/gRNA Ribonucleoproteins (RNPs) at the 1-cell stage. At 48 hpf, zebrafish embryos were collected (three fish per tube) for activity assay as described above. Two samples were analyzed for each NLS-Cas9 and NLS-Cas9-NLS group.

We then prepared recombinantly expressed, caged NLS-Cas9-K866TAG-HCK protein from *E. coli* (experiment performed by Wenyuan Zhou in our lab). The concentration of the protein sample was determined by the intensity of protein band on SDS-PAGE, using BSA as the standard (ImageJ). The protein (250 ng/μl) was incubated with the cep290 gRNA (100 ng/μl) *in vitro* for 10 min. The NLS-Cas9-K866TAG-HCK/gRNA ribonucleoproteins were irradiated in an Eppendorf tube with UV light (365 nm, 8 mW/cm²) for 1 min. Zebrafish injections and activity assays were performed as described above. Cas9 activities were detected in all three UV-irradiated samples (Fig 5.31). Among them, two samples (#1 and #2) showed almost 100% undigested PCR product, while one sample (#3) showed ~20% undigested PCR product. While

the variations between different samples would need to be minimized in the future, the efficiency of photoactivated Cas9 RNPS showed promises to be equal to that of the wild-type Cas9 RNPs. For non-irradiated samples, one out of two presented Cas9 activity (Fig 5.31), which might due to the undesired photo-decaging during microinjection or incubation. This *in vitro* irradiation test showed promises for light-activation of Cas9 activity in zebrafish.

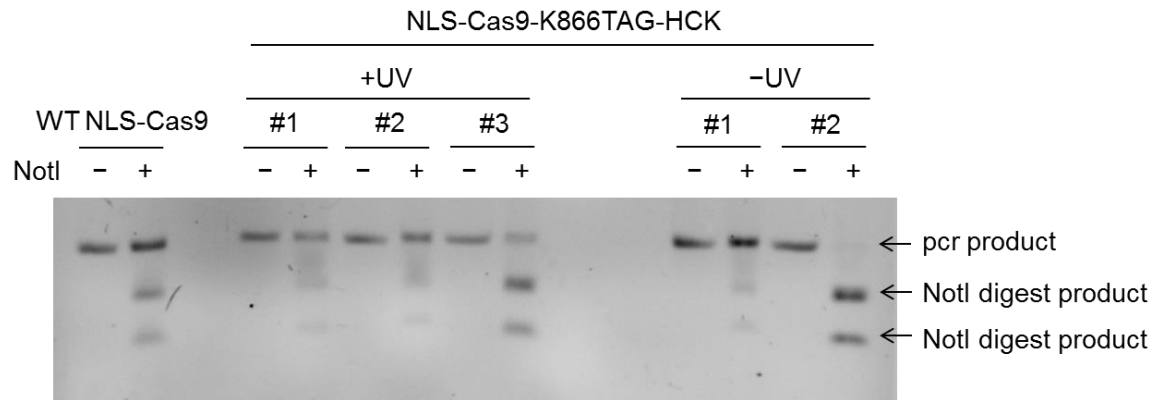


Figure 5.31. Optical control of genome modification with photocaged Cas9 RNPs injection. WT NLS-Cas9 and NLS-Cas9-K866TAG-HCK were recombinantly expressed from *E. coli*. The Cas9 protein (250 ng/μl) was mixed with the cep290 gRNA (100 ng/μl) *in vitro* for 10 min. The NLS-Cas9-K866TAG-HCK/gRNA ribonucleoproteins were irradiated in an eppendorf tube with UV light (365 nm, 8 mW/cm²) for 1 min. Zebrafish embryos were injected with Cas9/gRNA ribonucleoproteins (RNPs) at the 1-cell stage. At 48 hpf, zebrafish embryos were collected (three fish per tube) for activity assays as described above. Three samples were analyzed for UV-irradiated group, and two samples were analyzed for non-irradiated group. WT (wild-type) NLS-Cas9 was used as a positive control.

In summary, we confirmed successful genome modification on zebrafish through the injection of either Cas9 mRNA or Cas9 protein – in conjunction with the corresponding gRNA. The light activation of Cas9 function through *in vivo* incorporation of photocaged lysine was not successful, probably due to low incorporation efficiency (not confirmed by Western blot). However, the light activation of Cas9/gRNA RNPs showed some promising results. In the future, we are

going to inject the NLS-Cas9-K866TAG-HCK/gRNA ribonucleoproteins into zebrafish embryos, and perform UV irradiation on live zebrafish embryos.

Experimental

Plasmid construction

The K866TAG mutation was introduced on pCS2-Cas9-WT by site-directed mutagenesis (Fwd: 5'-caaaaacagaggatagtccagataatgtccctagtgaggaagtgg-3', Rev: 5'-cattatctgactatcctctgttttgcggacctggcagc-3', TAG mutations underlined).

mRNA synthesis

The pCS2-Cas9-WT plasmid was a gift from Dr. Michael Tsang. The Cas9-WT mRNA was generated as described in Chapter 5.1. Cep290 gRNA was generated through *in vitro* transcription (provided by Manush Sayd Mohammed; Tsang lab).

Microinjection of embryos

To confirm the Cas9-mediated genome modification in zebrafish, 2 nL of injection mixture containing wild-type Cas9 mRNA (100 ng/μl) and cep290 gRNA (100 ng/μl or 250 ng/μl) was injected per embryo. To express photocaged Cas9, the injection mixture (3 μl) was prepared as 100 ng/μl of Cas9-K866TAG mRNA, 100 ng/μl of HCKRS mRNA, 1,000 ng/μl of PylT, and cep290 gRNA (100 ng/μl). For the UAA incorporation, 0.15 μl of HCK (100 mM stock in 100% DMSO) was added to 1.5 μl of injection mixture. Embryos from natural mating were obtained and microinjected at the 1 to 2 cell stage using a World Precision Instruments Pneumatic PicoPump injector. For light activation experiments, at 6 hpf the embryos were irradiated for 30 sec with a 365 nm UV transilluminator (8 mW/cm²). At 48 hpf, zebrafish embryos were collected (3 fish per PCR tube) for NotI digestion assays.

For injection of Cas9 RNPs, the Cas9 protein (250 ng/μl) was mixed with the cep290 gRNA (100 ng/μl) *in vitro* for 10 min at room temperature.²⁴⁴ Zebrafish embryos were injected with the Cas9/gRNA ribonucleoproteins (RNPs) at 1-cell stage. At 48 hpf, zebrafish embryos were collected (3 fish per PCR tube) for NotI digestion assays.

NotI digestion assay for Cas9 genome editing

The zebrafish embryos in each PCR tube were digested in 20 μl of NaOH (60 mM) at 95 °C for 20 min. The fish lysate was neutralized to pH 8 with ~2.2 μl of Tris-HCl (1 M), and was centrifuged at max speed for 5 min. The supernatant (2 μl) was used as the template for amplifying the cep290 region through PCR (20 μl reaction, 0.3 μl of 10 mM dNTPs, 0.4 μl of 5 μM forward and reverse primers, 0.2 μl of Taq polymerase, in Taq polymerase buffer, Fwd: 5'-actcttttgctgtgacggta-3', Rev: 5'-tgcctcaaacgtgtcagcttaccatt-3', primers provided by Manush Sayd Mohammed). The PCR program was set as: 95 °C for 4 min, 30 cycles of (95 °C for 30 sec, 53 °C for 30 sec, 72 °C for 30 sec), 72 °C for 5 min. Half of the PCR product (10 μl) was digested with NotI restriction enzyme (NEB, # R3189S, 0.2 μl) at 37 °C for 3 h (in CutSmart buffer provided by NEB). Both non-treated and NotI-treated samples were analyzed on a 2% agarose gel (run at 80 V for 45 min).

5.6 Generation of Stable Fishline for Genetic Code Expansion

With the successful incorporation of UAAs into proteins through transient mRNA injection into zebrafish embryos, our next goal was to generate transgenic fish lines that stably express PylRS and its cognate PylT. The transgenic fish lines will allow sustained expression of necessary components (PylRS and PylT) for UAA mutagenesis, thus possibly increasing the incorporation efficiency. Moreover, the transgenic fish lines could be conveniently shared with the scientific community, allowing for broader applications in zebrafish studies.

We first constructed a pCS2-WTRS-2A-CFP plasmid. A cyan fluorescent protein (CFP) was fused to the WTRS (wild-type PylRS) through a 2A peptide, to indicate WTRS expression in zebrafish. The 2A peptide is a short, self-cleaving peptide that produces equal amount of proteins from the same mRNA transcript.²⁴⁵ The 2A peptide was used to ensure that the WTRS function is not affected by the CFP. We confirmed the CFP expression in mammalian cells (Fig 5.32A), and in zebrafish embryos (Fig 5.32B). The successful incorporation of Alloc Lys, based on a Rluc-L95TAG reporter, further indicated functional WTRS expression in zebrafish embryos (Fig 5.32C). We then cloned the WTRS-2A-CFP fragment into the pISceI plasmid, under an *Ubiquitin* promoter. The pISceI plasmid contains an I-SceI endonuclease sequence, which increases the efficiency of transgene integration.²⁴⁶ The *Ubiquitin* promoter drives the expression of WTRS throughout the whole animal.

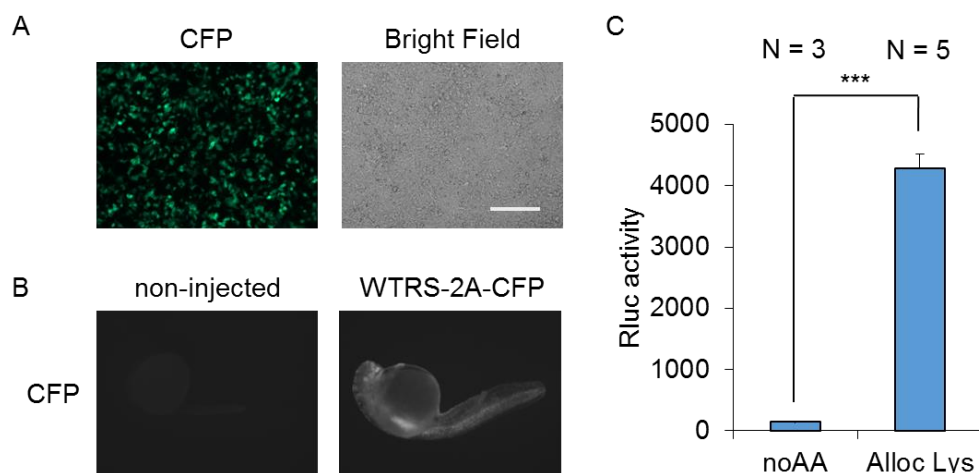


Figure 5.32. Functionality tests for the pCS2-WTRS-2A-CFP construct. (A) Expression of CFP in mammalian cells. HEK 293T cells were transfected with pCS2-WTRS-2A-CFP (200 ng). After 24 h, cells were imaged with CFP and bright field channels. Scale bar indicates 200 μ m. (B) Expression of CFP in zebrafish embryos. The WTRS-2A-CFP mRNA, RLuc-L95TAG mRNA, PylT, and Alloc Lys were injected into zebrafish embryos at the 1-2 cell stage. At 24 hpf, zebrafish embryos were imaged with a Leica M205 FA microscope with CFP and bright field channels. (C) Incorporation of Alloc Lys in zebrafish embryos. The fish lysates from the above experiment were collected at 24 hpf, and the RLuc activities were measured. N indicates the number of pooled samples (4 embryos each). Statistical significance is indicated by *** $p < 0.001$ (unpaired t-tests)

The pISceI-WTRS-2A-CFP plasmid (50 ng/ μ l) was injected into zebrafish embryos at the 1-cell stage. The embryos were raised to adulthood, and F0 adults (4 pairs) were crossed. However, no CFP expression was observed from the offsprings (~100 embryos) at 24 hpf, indicating that the WTRS-2A-CFP cassette was not incorporated into the F0 fish genome. Further effort to screen more F0 adults will be needed. We also constructed the pISceI-HCKRS-2A-CFP in the same way. The pISceI-HCKRS-2A-CFP plasmid (50 ng/ μ l) were injected to zebrafish embryos at the 1-cell stage. The zebrafish were raised to adult, and the offsprings from F0 adults will be screened for CFP expression. It should be noted that the polyA sequence was not present

in either plasmids. Current effort to construct the plasmid containing polyA signal is under way, and the new plasmid will be used for generating stable fishline.

Future work also includes generating a transgenic vector for PylT expression, as well as inserting the PylT expression cassette into our current pISceI-Ubb-PylRS-2A-CFP vector. In a recent study of transgenic fish line generation for UAA mutagenesis, four copies of U6-promoter driven tRNAs were designed in the transgene construct.²⁴⁷ U6 promoter has also been used for stable expression of small RNAs in zebrafish.²⁴⁸ Therefore, we plan to place four copies of PylT (driven by four separate U6 promoters) into our current vector. The final construct will be injected into the zebrafish embryos at 1-cell stage. The embryos will be raised to adulthood, and the offsprings from F0 adults will be screened for CFP expression at 24 hpf. CFP-positive embryos, which indicates successful insertion of the transgene vector, will be raised to adults. To validate the functionality of the engineered PylRS/PylT transgenic fish line, embryos from the transgenic fish line will be injected with Rluc-L95TAG mRNA and corresponding UAA at the 1-2 cell stage. At 48 hpf the fish lysate will be collected for the luciferase assay (discussed in Chapter 5.1). The Rluc activity will be indicative of the PylRS/PylT functionality.

Experimental

Plasmid construction

To generate the pCS2-WTRS-2A-CFP plasmid, the pCS2-Cyto-CFP plasmid (8,000 ng, gift from Dr. Michael Tsang)²⁴⁹ was digested with NcoI and AgeI restriction enzymes, followed by gel purification of the band corresponding to the correct length of the backbone (~5,000 bp). The WTRS-2A fragment was amplified from the pCS2-WTRS plasmid (construction described in Chapter 5.1, Fwd: 5'-tgaccatggactacaaggacgacgacg-3', Rev: 5'-gaaccggtaggaccgggggttttctccacgtctcctgcttgcttaacagagagaagttcgtggctccggatcccaggttggtgctgatg-3', 2A peptide sequence underlined). The 2A peptide sequence, encoding GSGATNFSLLKQAGDVEENPGP, was designed on the reverse primer. The PCR product was

digested with NcoI and AgeI restriction enzymes (37 °C, 2 h). The pCS2-WTRS-2A-CFP plasmid was constructed by ligating the WTRS-2A fragment into the pCS2-CFP backbone using NcoI and AgeI restriction sites (T4 ligase, 16 °C, overnight). The pCS2-HCKRS-2A-CFP plasmid was also constructed in the same way.

To generate the pISceI-WTRS-2A-CFP plasmid, the pISceI plasmid (8,000 ng, gift from Dr. Michael Tsang) was digested with SpeI and NotI restriction enzymes, followed by gel purification of the band corresponding to the correct length of the backbone (~6,500 bp). The WTRS-2A-CFP fragment was amplified from the pCS2-WTRS-2A-CFP plasmid (Fwd: 5'-gtccactagtagtgactacaaggacgacgacg-3', Rev: 5'-tactgcgccgcttactgtacagctcgtccatgc-3'). The PCR product was digested with SpeI and NotI restriction enzymes (37 °C, 2 h). The pISceI-WTRS-2A-CFP plasmid was constructed by ligating the WTRS-2A-CFP fragment into the pISceI backbone using SpeI and NotI restriction sites (T4 ligase, 16 °C, overnight). The pISceI-HCKRS-2A-CFP plasmid was also constructed in the same way. It should be noted that the polyA sequence was not present in either plasmids. Current effort to construct the plasmid containing polyA signal is under way.

Expression of WTRS-2A-CFP in mammalian cells

HEK 293T cells were transfected with pCS2-WTRS-2A-CFP (200 ng) in a 96-well plate. Branched polyethylene imine (bPEI, 1.5 µl, 1 mg/ml) was used as the transfection reagent. At 24 h, cells were imaged on a Zeiss Observer Z1 microscope with CFP channel.

Microinjection of embryos

The WTRS-2A-CFP mRNA was generated as described in Chapter 5.1, using the pCS2-WTRS-2A-CFP plasmid. The injection mixture (3 µl) was prepared as 50 ng/µl of Rluc-L95TAG mRNA, 100 ng/µl of WTRS-2A-CFP mRNA, and 1,000 ng/µl of PyIT. For the UAA incorporation, 0.15 µl of Alloc Lys (100 mM stock in 100% DMSO) was added to 1.5 µl of injection mixture. Embryos

from natural mating were obtained and microinjected at the 1- to 2-cell stage with 2 nL of the injection mixture using a World Precision Instruments Pneumatic PicoPump injector. At 24 hpf, zebrafish embryos were imaged with a Leica M205 FA microscope with CFP and bright field channels. The fish lysates were further collected, and the Rluc activity assays were performed as described in Chapter 5.1.

Generation of transgenic fish line

To generate the transgenic fish line expressing wild-type PylRS, the pISceI-WTRS-2A-CFP plasmid (50 ng/ μ l) was injected into zebrafish embryos at the 1-cell stage. The embryos were raised to adulthood (takes ~3 months), and F0 adults (4 pairs, one male and one female for each pair) were crossed in separate tanks. At 24 hpf, zebrafish embryos were imaged with a Leica M205 FA microscope with CFP and bright field channels.

Appendix A. Optical control of other biomolecules in zebrafish

A1. Optical control of circular morpholino function in zebrafish

Morpholino oligomers (MOs) are the most frequently used gene-silencing reagents in the study of zebrafish embryos.²⁵⁰ MOs, usually around 25 base pairs, bind to mRNA and silence protein translation. In a typical experimental setting, MO that targets gene of interest is synthesized and injected into zebrafish embryos at the 1-2 cell stage. Observation of any phenotypic defects would indicate critical role of the target gene for embryonic development. Therefore, MOs provide a versatile method to study gene function in zebrafish embryos.

One disadvantage of the MO injection, however, is the lack of spatiotemporal control on gene regulation. Once MO is injected into an embryo, it will distribute throughout the whole embryo, thus inhibiting translation of the protein of interest globally. However, gene expression patterns are spatially and temporally defined during embryonic development.²⁵¹ A tool to selectively control MO activities within a defined time and space will allow us to study the mechanisms of complicated gene networks, which cannot be easily achieved with traditional MO injections.

Several strategies have been applied to achieve the spatiotemporal control of MO activity.²⁵² Recently, circular morpholinos were generated with a photocleavable oligonucleotide linker.^{194d, 253} Due to the steric constraint, the circular morpholino does not bind to mRNA and thus does not inhibit protein translation. The linker is cleaved with the light exposure, generating a linear, functional morpholino. MOs can also be cyclized with different caging groups that are responsive to wavelength-selective light, thus allowing for selective silencing of multiples target

genes with sequential light exposure.^{253b} This strategy has been applied to selectively control gene expression of flatting head (*flh*) and spadetail (*spt*) in live zebrafish embryos.^{253b}

Previously, morpholinos have also been applied to inhibit miRNA function in zebrafish.²⁵⁴ miRNA is a 22 nucleotide non-coding RNA that plays important roles in gene regulation.²⁵⁵ To enable spatiotemporal control of miRNA function in zebrafish, We asked if a photocleavable morpholino that targets miRNA could be generated. We first chose to target miR-125b, because injection of morpholino targeting miR-125b (5'-UCCCUGAGACCCUAACUUGUGA-3') has led to dramatic phenotypic defects (severe cell death in the brain, dorsalized embryo) at 24 hpf.²⁵⁶ To this end, different amount of scrambled morpholino (targeting a random sequence) or miR-125b morpholino were injected into zebrafish embryos at the 1-2 cell stage. In agreement with the previous study, phenotypic defects (dorsalized embryo, cell death in the brain, as demonstrated by opaque tissues next to the zebrafish eyes) were observed in miR-125b morpholino injected embryos, but not scrambled morpholino-injected embryos (Fig A1).

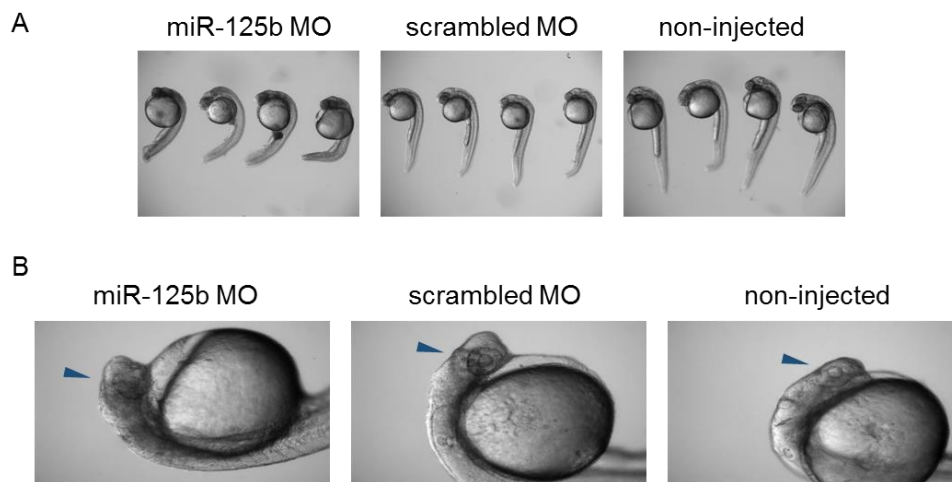


Figure A1. Phenotypic defects of the embryos injected with miR-125b morpholino (10 ng). (A) Dorsalized embryos. (B) Severe cell death in the brain. Arrows indicate positions of the cell death. Images were taken at 24 hpf.

We further counted the normal and defective embryos at 24 hpf. miR-125b morpholino (10 ng or 15 ng) led to phenotypic defects on 90% of the embryos (Fig A2). A lower amount of miR-125b morpholino (6 ng) showed milder effect (53%). Scrambled morpholino showed no effect on phenotype, suggesting that the phenotypic defects were caused by targeting miR-125b. Taken together, these data suggest that miR-125b morpholino injection (10 ng or 15 ng) leads to distinct loss-of-function embryo morphology. Luis Angel Vazquez in the Deiters lab further pursued the synthesis of circular miR-125b morpholino. However, the synthesis failed after several attempts.

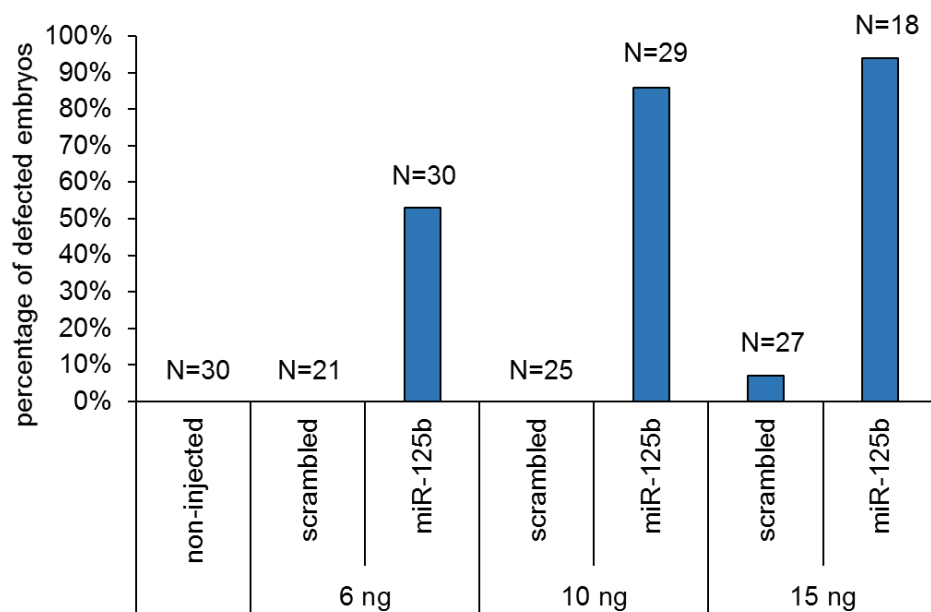


Figure A2. Injection of miR-125b morpholino led to phenotypic defects. Different amount (6 ng, 10 ng, 15 ng) of scrambled morpholino (targeting random sequences) or miR-125b morpholino were injected into zebrafish embryos at the 1-2 cell stage. Normal and defected embryos were counted at 24 hpf. N represents number of embryos observed.

We therefore looked to target another microRNA, miR-30, which is involved in the regulation of Hedgehog signaling pathway.²⁵⁷ Injection of miR-30 morpholino leads to elevated expression of Ptc1, an Hedgehog ligand receptor that is activated by Hedgehog signaling.²⁵⁷ We are currently confirming the expression of Ptc1 by in-situ hybridization.

A2. Optical control of peptide nucleic acid (PNA) in zebrafish

Peptide nucleic acids (PNAs) are synthetic homologs of nucleic acids, in which the phosphate–sugar polynucleotide backbone is replaced by a flexible pseudo-peptide polymer.²⁵⁸ PNA oligomers show great specificity in binding to complementary DNAs or RNAs. PNAs are also resistance to DNAses and proteinases, making them an ideal molecular tool for a range of biological research,²⁵⁹ as well as gene therapies.²⁶⁰ For example, PNAs were used for targeted gene knockdown in zebrafish embryos.²⁶¹ Light-activatable PNAs have also been developed for conditional downregulation of gene expression in zebrafish.^{261b}

We are interested in developing methods for optochemical gene activation in live animals. For example, a caged promoter strategy was applied for light-activation of gene expression in zebrafish.²⁶² However, this method requires incorporation of caged oligonucleotides into double-stranded circular DNA. The injection of plasmid into the zebrafish also suffers from the mosaic expression.²⁶³ Alternatively, an antisense photo-morpholinos could be designed to optically activate gene function.²⁶⁴ The antisense photo-morpholino contains a photocleavable group in the middle. It first binds the target RNA and inhibits its function. Upon UV light irradiation, it is cleaved into two fragments, liberating RNA function. Recently, γ -modified peptide nucleic acids (γ PNAs) were used to control protein translation *in vitro*.²⁶⁵ The γ PNA blocks protein translation through targeting one of the two mRNA sites, the 5'-terminal end of the mRNA, or the Kozak initiation sequence.²⁶⁶ We therefore hypothesized that a photoactivable γ PNA could be used to control protein translation, a similar strategy that was applied to antisense reagents.²⁶⁷

To this end, we first optimized the Kozak sequence of our Rluc expression construct (from GGATCCATGG to GCAAACATGG, start codon underlined), to ensure maximum translation efficiency in zebrafish.²⁶⁸ We then designed γ PNA sequences that target the 5'-terminal end, or the Kozak initiation sequence. The γ PNA (100 nM or 100 μ M) was incubated with Rluc mRNA

(100 ng/μl, or 3 μM) at 37 °C for 1 h, condition used in a previous report.²⁶⁵ The mixture was then injected into zebrafish embryos at the 1-2 cell stage. At 24 hpf, fish lysates were collected for Rluc activity assays. However, we did not observe any decrease of luciferase activity with the γPNA treatment (Fig A3). Since 100 nM (concentration of γPNA) is much lower than 3 μM (concentration of Rluc mRNA), and thus efficient inhibition is not expected, this condition should not be tested experimentally. We did not observe efficient inhibition when γPNA was in excess (100 μM) though.

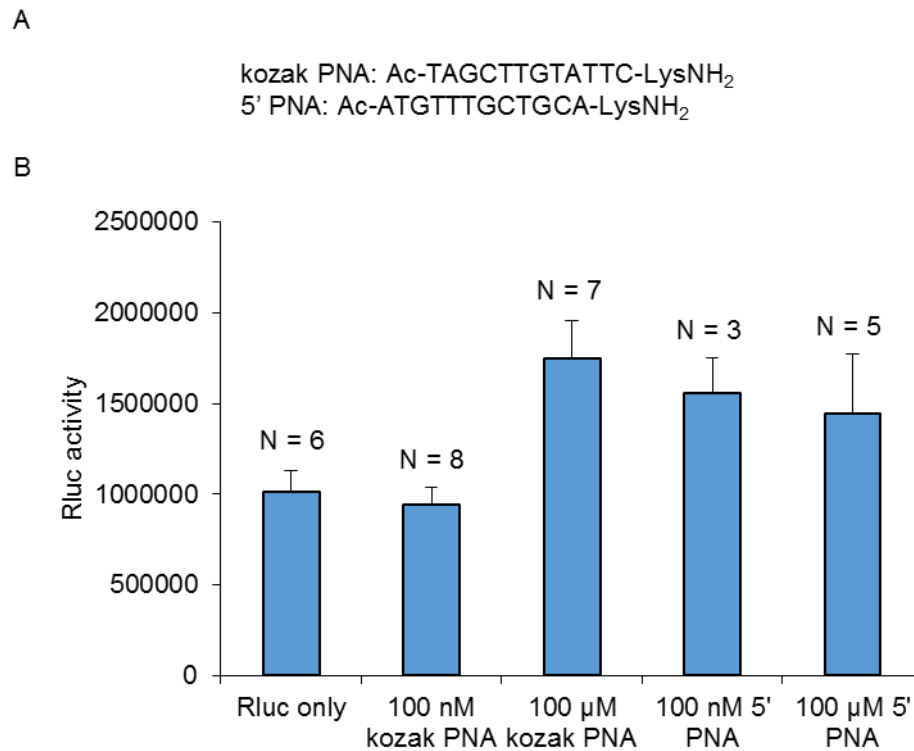


Figure A3. γPNAs targeting either kozak sequence or 5'-terminal end did not inhibit Rluc translation. (A) Sequences of γPNAs used in this study. (B) The Rluc mRNA (100 ng/μl, or 3 μM) was mixed with the γPNA targeting kozak sequence (kozak PNA) or the γPNA targeting 5'-terminal end (5' PNA). The mixture was incubated at 37 °C for 1 h, and was injected into zebrafish embryos at the 1-2 cell stage. At 24 hpf, fish lysates were collected for Rluc activity assays. N indicates the number of pooled samples (4 embryos each). Standard error of mean (SEM) was indicated.

We further tested *in vitro* translation of Rluc. But the γ PNA treatment still did not inhibit Rluc translation (Fig A4, experiment performed by rotation student, Sara Whitlock). In some cases (kozak PNA at 100 nM or 10 μ M), a surprising increase of translation level was even observed. Altogether, the tested γ PNAs could not inhibit Rluc mRNA translation.

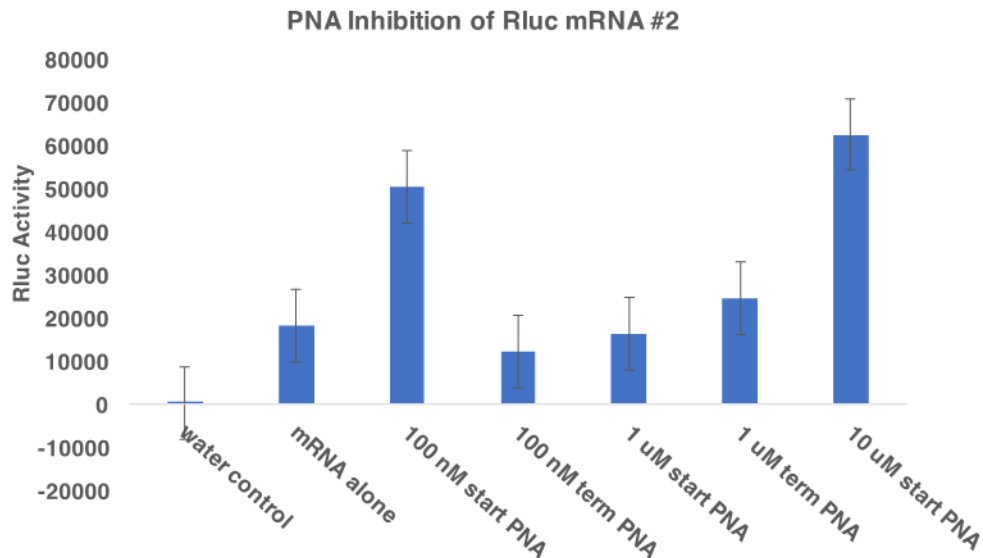


Figure A4. γ PNA that targets the 5'-terminal end (term PNA) or kozak sequence (start PNA) of Rluc did not inhibit Rluc translation. The sequences of γ PNAs are shown in the previous figure. The Rluc mRNA (3 μ M) was mixed with the γ PNA targeting the 5'-terminal end or kozak sequence of Rluc. The mixture was incubated at 37 °C for 1 h, and the *in vitro* inhibition assays were performed. The error bars indicated standard deviation, with N = 3 (technical replicates).

We then decided to test the inhibition of Fluc translation, which has been successfully targeted with a γ PNA in test tube experiments.²⁶⁵ A 10-mer or 12-mer γ PNA targeting the 5'-terminal end of Fluc (gifts from Dr. Bruce Armitage) was used. The 10-mer or 12-mer γ PNA (7.5 μ M or 15 μ M) was incubated with the Fluc mRNA (3 μ M) at 37 °C for 1 h. The mixture was injected into zebrafish embryos at the 1-2 cell stage. At 24 hpf, fish lysates were collected for Fluc activity

assays (Fig A5). Gratifyingly, we observed a 75% reduction of Fluc activity with the treatment of γ PNA (10-mer or 12-mer, 15 μ M). In the future, we plan to further test the inhibition efficiency with longer γ PNAs (14-mer and 16-mer), and the best γ PNA will be used for light-activation of Fluc translation assay.

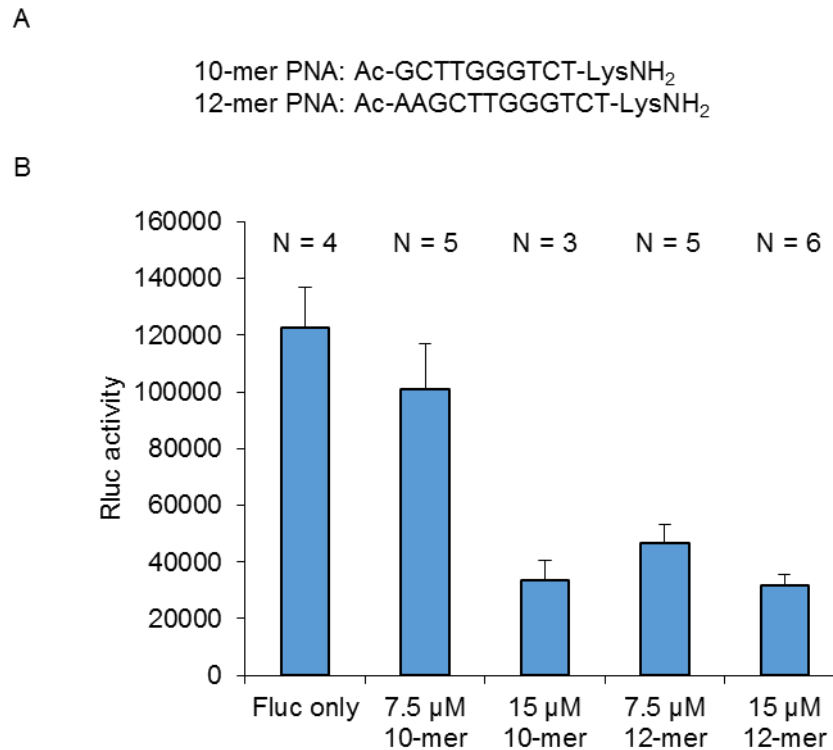


Figure A5. γ PNA that targets the 5'-terminal end of Fluc inhibits Fluc translation. (A) Sequences of γ PNAs used in this study. (B) The Fluc mRNA (3 μ M) was mixed with the γ PNA targeting the 5'-terminal end (12-mer or 15-mer). The mixture was incubated at 37 $^{\circ}$ C for 1 h, and was then injected into zebrafish embryos at the 1-2 cell stage. At 24 hpf, fish lysates were collected for Fluc activity assays. N indicates the number of pooled samples (4 embryos each). Standard error of mean (SEM) was indicated.

In summary, we confirmed the phenotypical effect of miR-125b inhibition in zebrafish. Furthermore, we also sought to block protein translation through targeting mRNA sequence with γ -modified peptide nucleic acids (γ PNAs). We observed inhibition of firefly luciferase expression

by γ PNA (10-mer or 12-mer), targeting 5'-terminal end of the mRNA, with the most efficiency at the concentration of 15 μ M. In the future, since the circular miR-125b morpholino cannot be generated, we will confirm the phenotypic effect of another microRNA target, miR-30, and generate a circular miR-30 morpholino for optical control of microRNA function in zebrafish embryos. For γ PNA, we will test the inhibition efficiency of γ PNAs with different length, and further generate photocaged γ PNAs for optical control of mRNA function in zebrafish.

Experimental

Plasmid construction

The kozak sequence of pCS2-Rluc plasmid (construction described in chapter 5.1) was first optimized to ensure maximum translation efficiency in zebrafish, using the site-directed mutagenesis (Fwd: 5'-cttttgagcaaacatggctccaaggtgtacgaccccgag-3', Rev: 5'-ggaagccatgtttgctgcaaaaagaacaagtagctgtattc-3', mutations on kozak sequence were underlined). γ PNAs targeting the Rluc (Ac-TAGCTTGTATTC-LysNH₂, targeting kozak sequence 5'-GAAUACAAGCUA-3'; Ac-ATGTTTGCTGCA-LysNH₂, targeting 5'-terminal sequence 5'-UGCAGCAAACAU-3') were purchased from PNA Innovations, and 100 μ M stock solutions were made with water. The pT7-Fluc plasmid and γ PNAs targeting the Fluc (12-mer: Ac-AAGCTTGGGTCT-LysNH₂, 10-mer: Ac-GCTTGGGTCT-LysNH₂, both targeting 5'-AGACCCAAGCUUUCA-3') were gifts from Dr. Bruce Armitage.²⁶⁵

mRNA synthesis

The generation of Rluc mRNA was performed following the protocol described in Chapter 5.1, except using the kozak sequence-optimized pCS2-Rluc plasmid (construction method described in the above paragraph) as the template.

The pT7-Fluc plasmid (7.5 μ g, gift from Dr. Bruce Armitage) was linearized through ApaL1 digestion (1 μ l of ApaL1 in 30 μ l reaction, in CutSmart buffer provided by NEB, 37 °C, overnight).

The linearized product (30 μ l) was mixed with water (20 μ l) and phenol:chloroform:isoamyl alcohol (PCIA, 50 μ l), and centrifuged at maximum speed for 5 min. The top layer (~50 μ l) was collected, and further mixed with NaOAc (5 μ l, 3 M, pH 5.2, pH adjusted with acetic acid) and 100% ethanol (125 μ l). The mixture was centrifuged at maximum speed for 5 min. The pellet was washed with 70% ethanol (600 μ l), and dissolved in 10 μ l of water. The concentration of the linear DNA was determined by Nanodrop. Linear DNA (1 μ g) was used to generate capped mRNA in a 20 μ l reaction (37 $^{\circ}$ C, 4 h) with the T7 Megascript Transcription Kit (Ambion), together with the cap analog (3 μ l of 40 mM stock) (m⁷G(5')ppp(5')G) (Fisher). DNase (1 μ l, provided by the kit) was then added to remove the linear DNA template (37 $^{\circ}$ C, 30 min). The reaction (20 μ l) was mixed with water (30 μ l) and PCIA (50 μ l), and was centrifuged at maximum speed for 5 min. The top layer (~50 μ l) was purified through a G-50 sephadex spin column (Roche, #11274015001, column buffered with 10 mM Tris-HCl, pH 7.5; 1 mM EDTA; 100 mM NaCl) according to the product manual. The RNA solution (50 μ l) was mixed with water (50 μ l), NaOAc (10 μ l, 3 M, pH 5.2), and 100% ethanol (300 μ l). The mixture was placed at -20 $^{\circ}$ C freezer overnight, and was centrifuged at maximum speed for 15 min. The RNA pellet was washed with 70% ethanol, and dissolved in water (10 μ l). The quality was verified by running 0.5 μ l of the RNA (in 10 μ l of water) on a 1% agarose gel (run at 80 V for 45 min). The RNA concentration following this protocol is normally around 1 μ g/ μ l.

Microinjection of embryos

The miR-125b morpholino (3 μ g/ μ l, 5 μ g/ μ l, or 7.5 μ g/ μ l, in water) was injected into zebrafish embryos (2 nL) at the 1-2 cell stage. Scrambled morpholino of the same concentration was used as the negative control. At 24 hpf, the embryos were imaged using a stereomicroscope (Leica MZ 16 FA). Normal and defective embryos were counted.

For *in vivo* Rluc inhibition assay, the injection mixture was prepared as γ PNA (100 nM or 100 μ M) and Rluc mRNA (3 μ M) in water. For *in vivo* Fluc inhibition assay, the injection mixture

was prepared as γ PNA (7.5 μ M or 15 μ M) and Fluc mRNA (3 μ M) in water. In both experiments, the mixture was incubated at 37 °C for 1 h to facilitate duplex formation. Embryos from natural mating were obtained and microinjected at the 1 to 2 cell stage with 2 nL of the injection mixture using a World Precision Instruments Pneumatic PicoPump injector. At 24 hpf, fish lysates (4 embryos per pooled sample, homogenized in 50 μ l of 1x passive lysis buffer (Promega)) were collected and the luciferase assays were performed as described in Chapter 5.1.

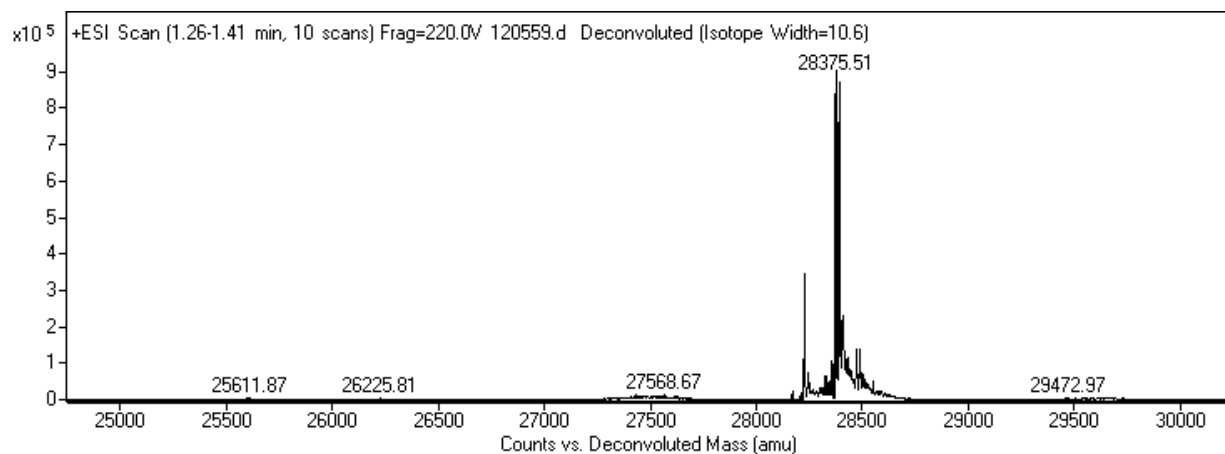
in vitro inhibition assay

The Rluc mRNA (3 μ M) was mixed with γ PNA (100 nM, or 1 μ M, or 10 μ M) in water, and the mixture was incubated at 37 °C for 1 h to facilitate duplex formation. The Rluc mRNA (either with or without incubation with γ PNA) was translated with the Promega Rabbit Reticulocyte Lysate kit (#L4960) following the kit recommended conditions at 37 °C for 2 h. Then 3 μ l of the translation product was mixed with 30 μ l of Renilla Luciferase Assay Reagent (Promega), and luminescence was read immediately on a microplate reader (Tecan M1000 PRO), with an integration time of 1000 ms.

Appendix B. ESI-MS spectrum, plasmid map, general protocol, buffer composition

ESI-MS spectrum

A



B

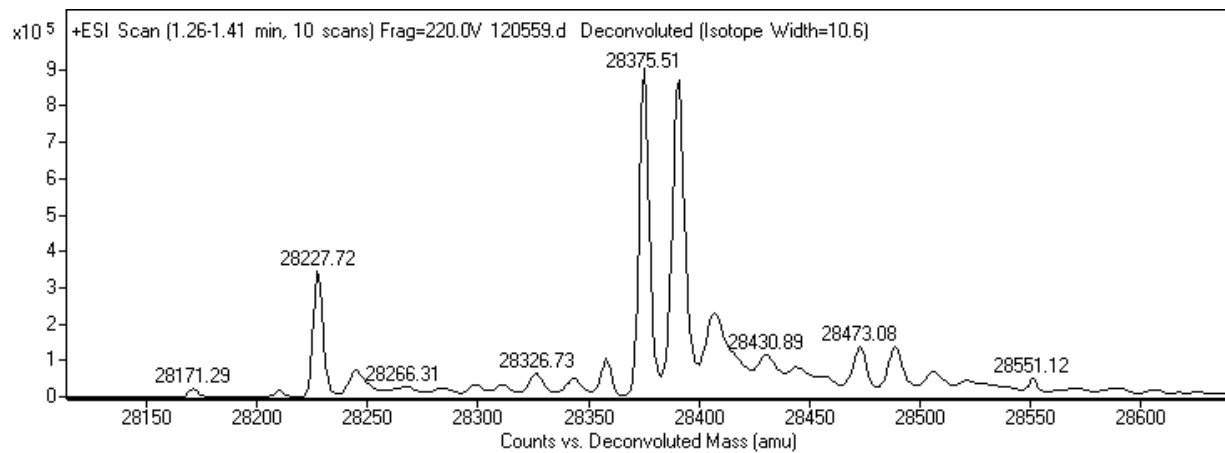
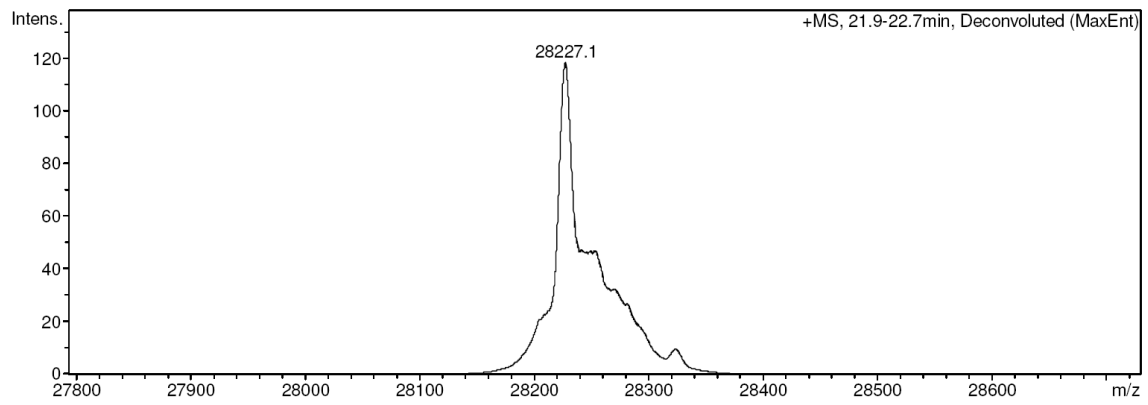


Figure A6. ESI-MS analysis for sfGFP-Y151TAG-DK. (A) spectrum displayed with range from 25000 to 30000 Da. (B) spectrum displayed with range from 28150 to 28600 Da.

A



B

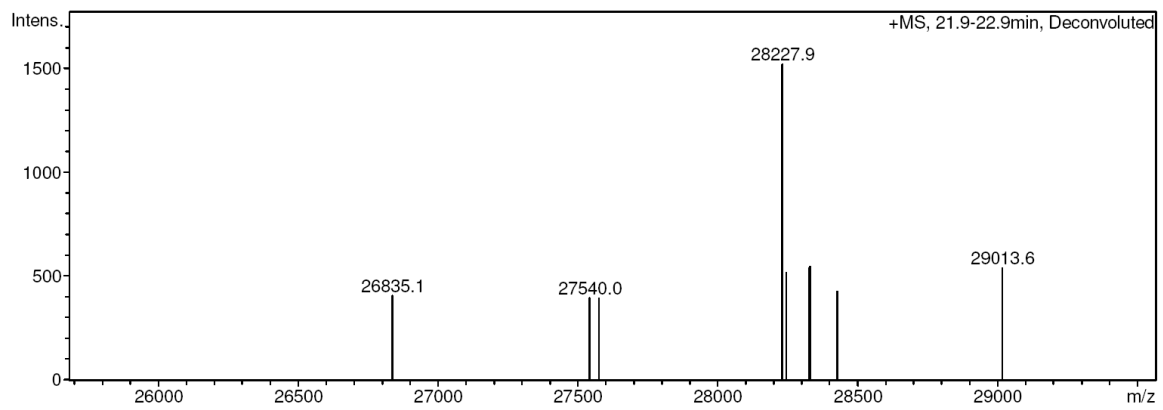
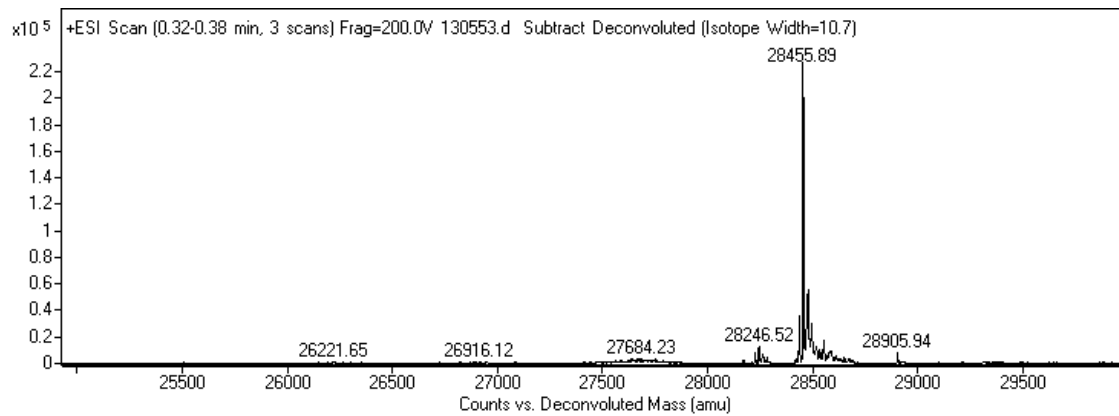
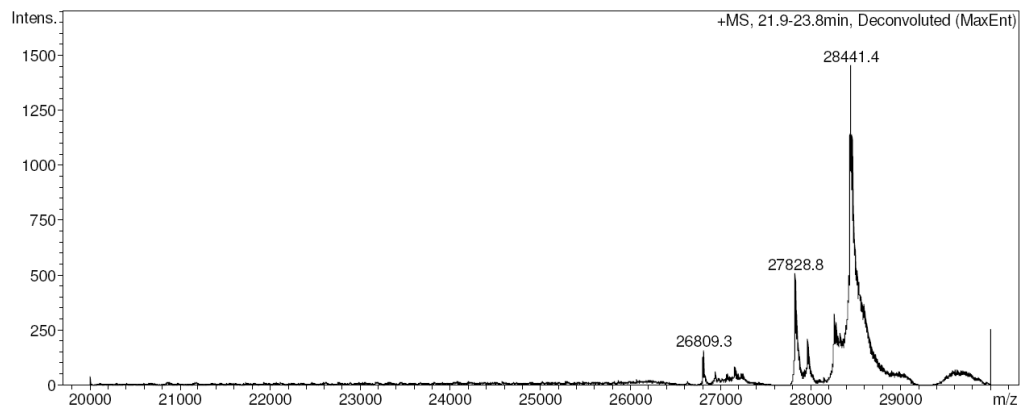


Figure A7. ESI-MS analysis for (A) sfGFP-Y151TAG-AzEK and (B) sfGFP-Y151TAG-AzMK.

A



B



C

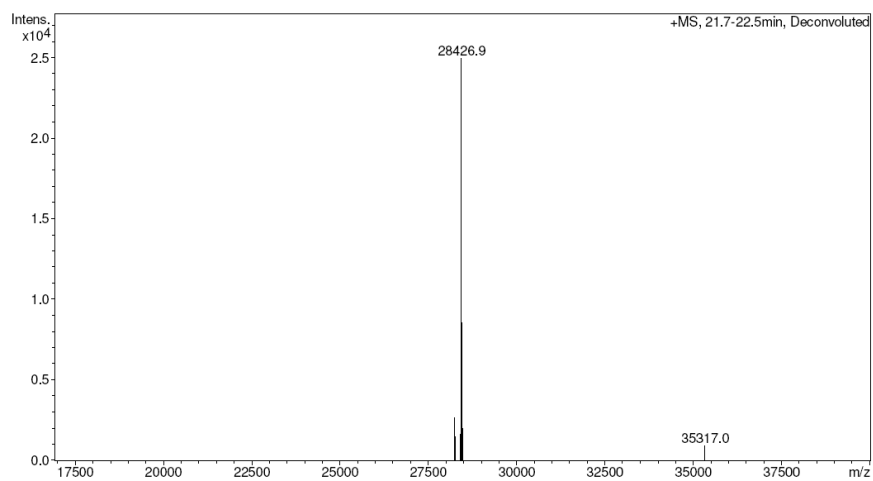
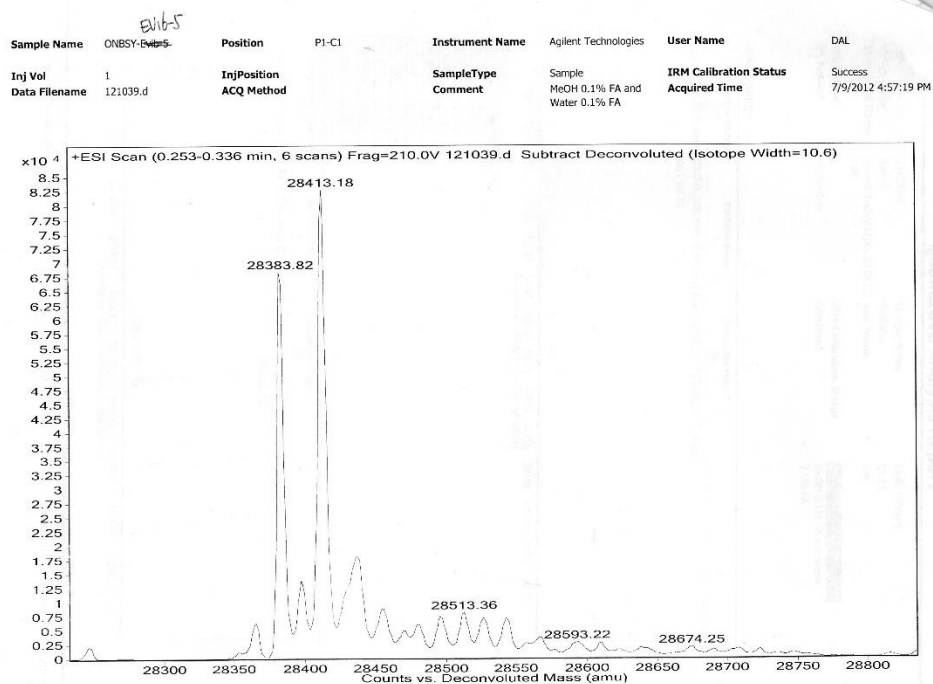


Figure A8. ESI-MS analysis for (A) sfGFP-Y151TAG-MNPY, (B) sfGFP-Y151TAG-NPY, (C) sfGFP-Y151TAG-NPEY.

A



B

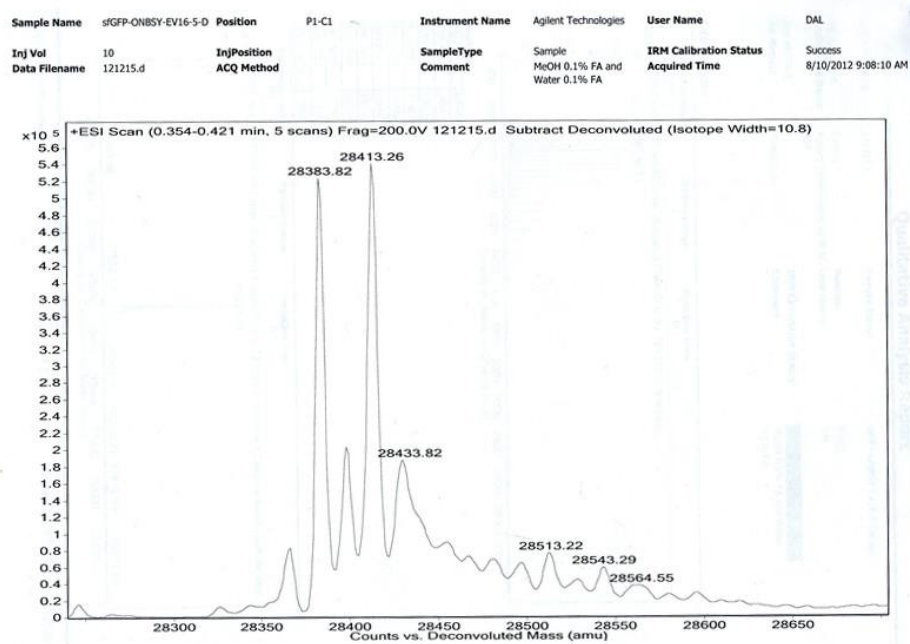
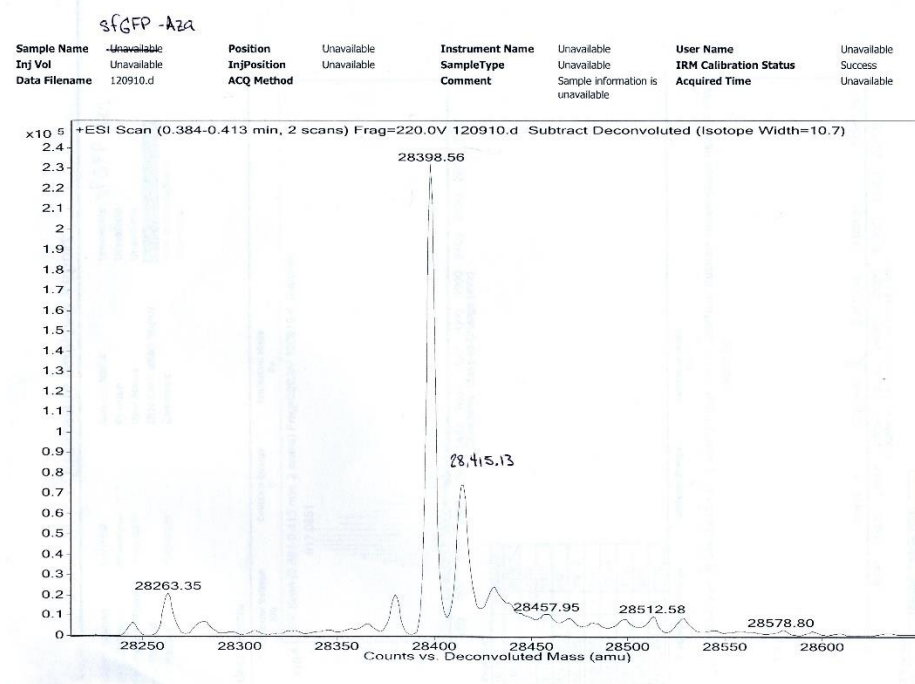


Figure A9. ESI-MS analysis of sfGFP-Y151TAG-ONBSY (A) before and (B) after UV irradiation.

A



B

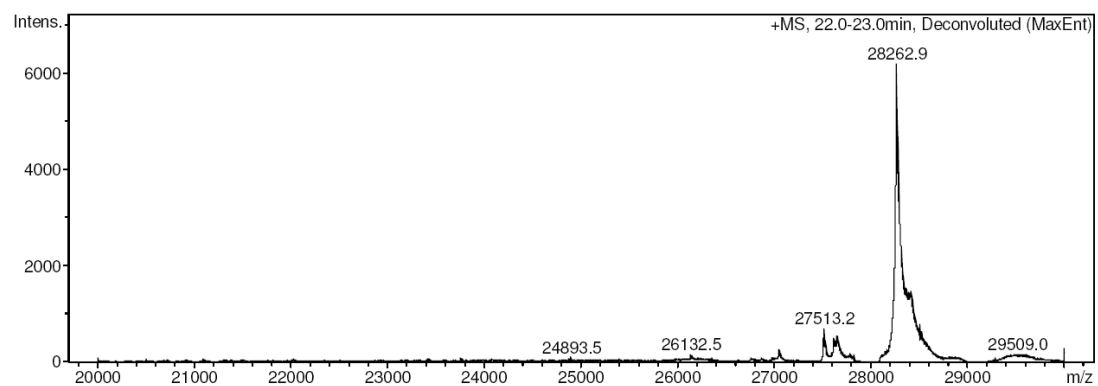


Figure A10. ESI-MS analysis of sfGFP-Y151TAG-ONBAY (A) before and (B) after UV irradiation.

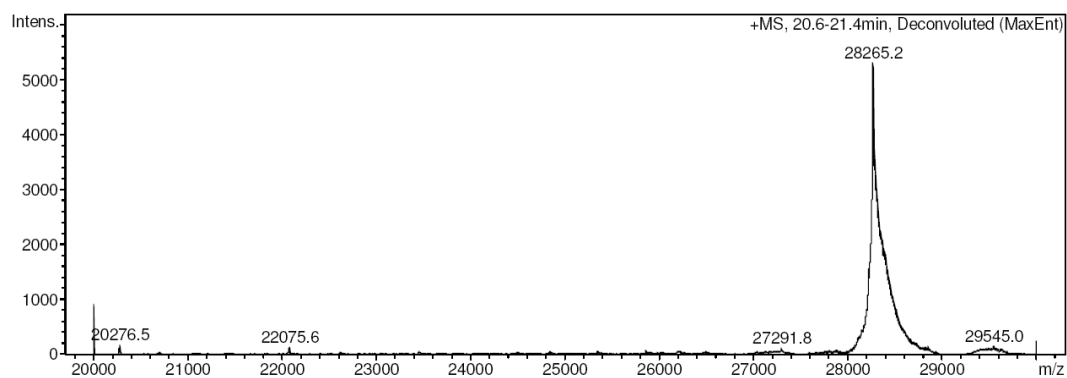
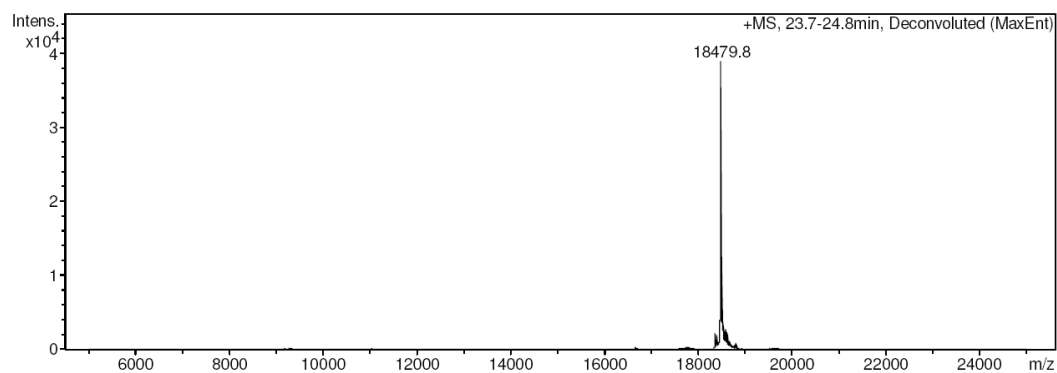


Figure A11. ESI-MS analysis of sfGFP-Y66TAG-ONBAY after UV irradiation.

A



B

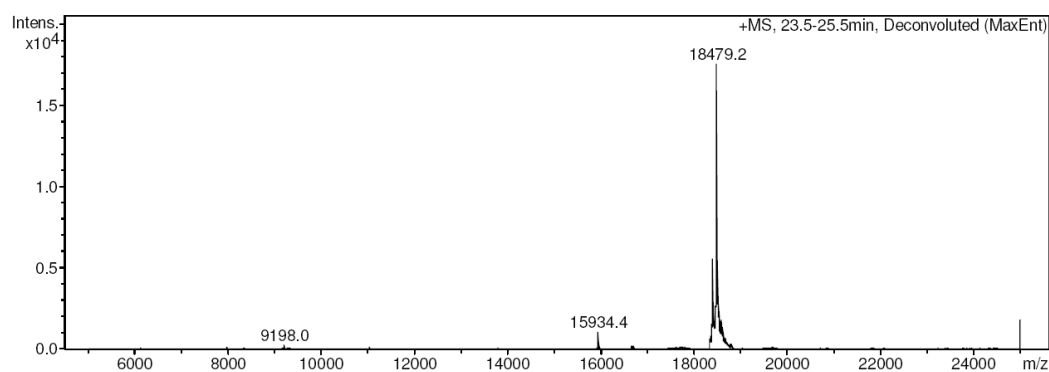
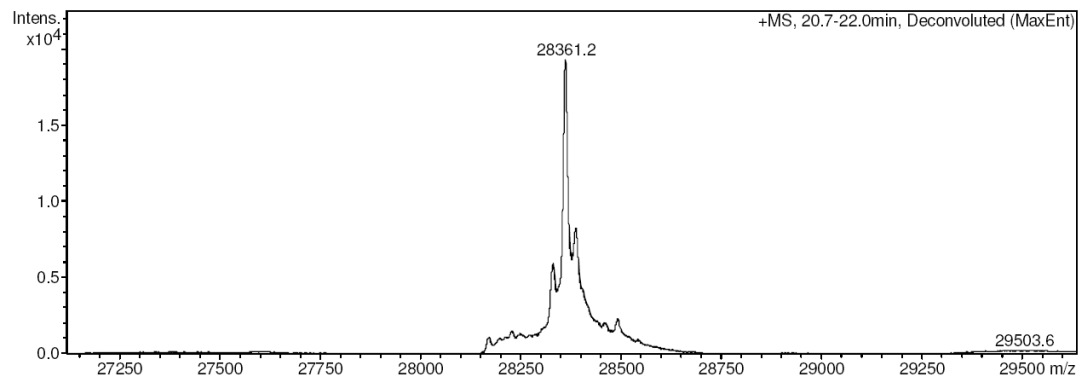


Figure A12. ESI-MS analysis for genetic incorporation of alkene lysine (A) **7** and (B) **9** into Myo4TAG.

A



B

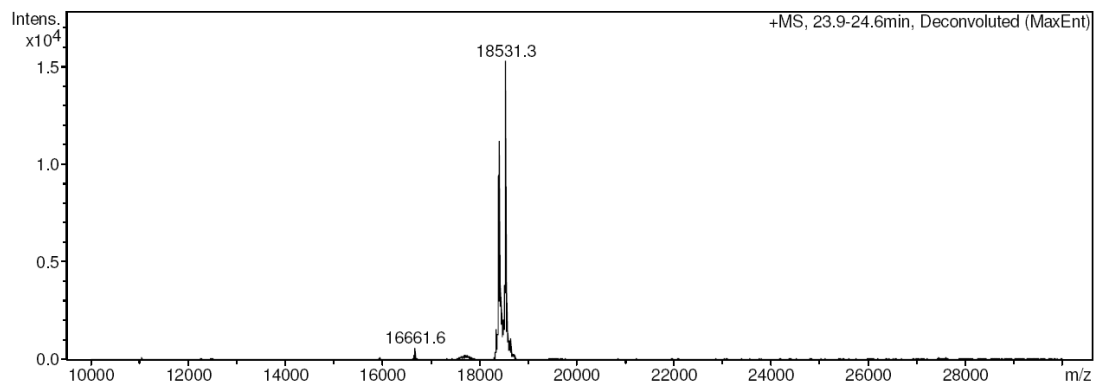


Figure A13. ESI-MS analysis of (A) sfGFP-Y151TAG-BPK and (B) Myo-4TAG-BPK.

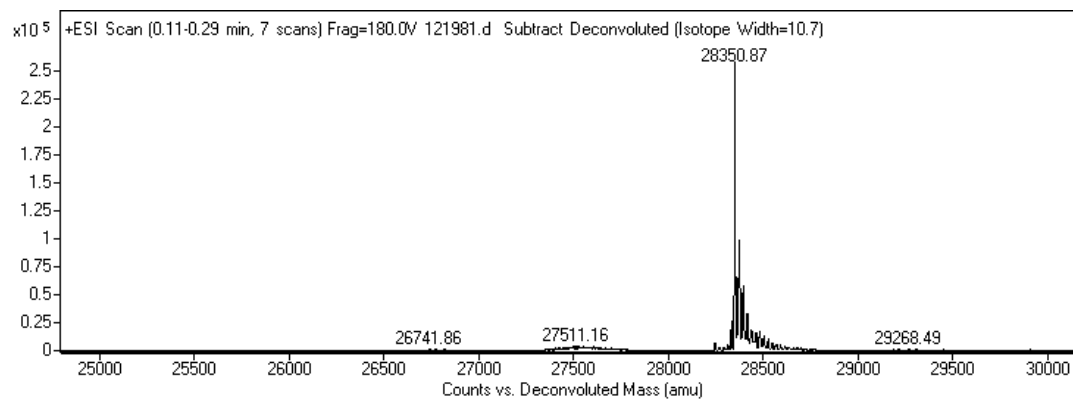
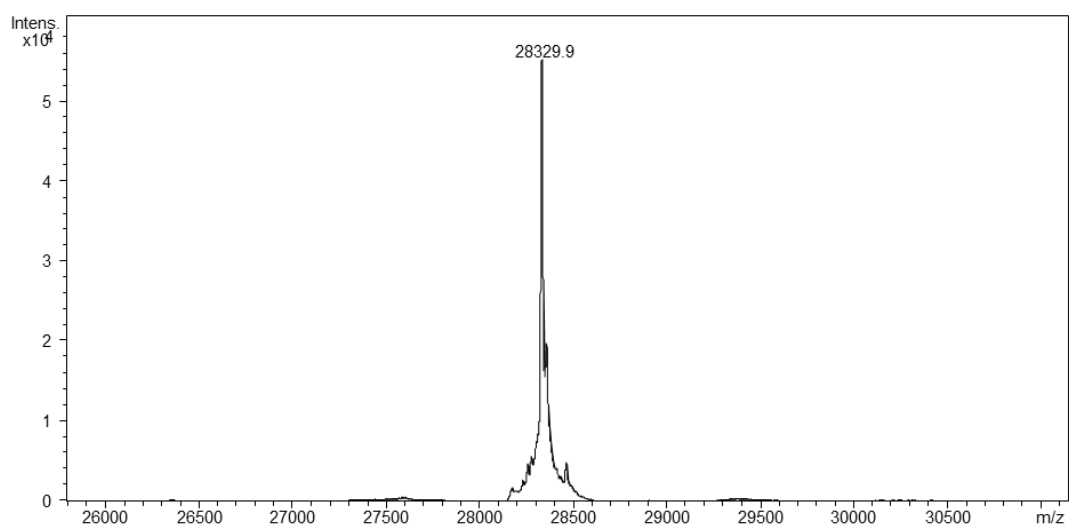


Figure A14. ESI-MS analysis of sfGFP-Y151TAG-AzoF.

A



B

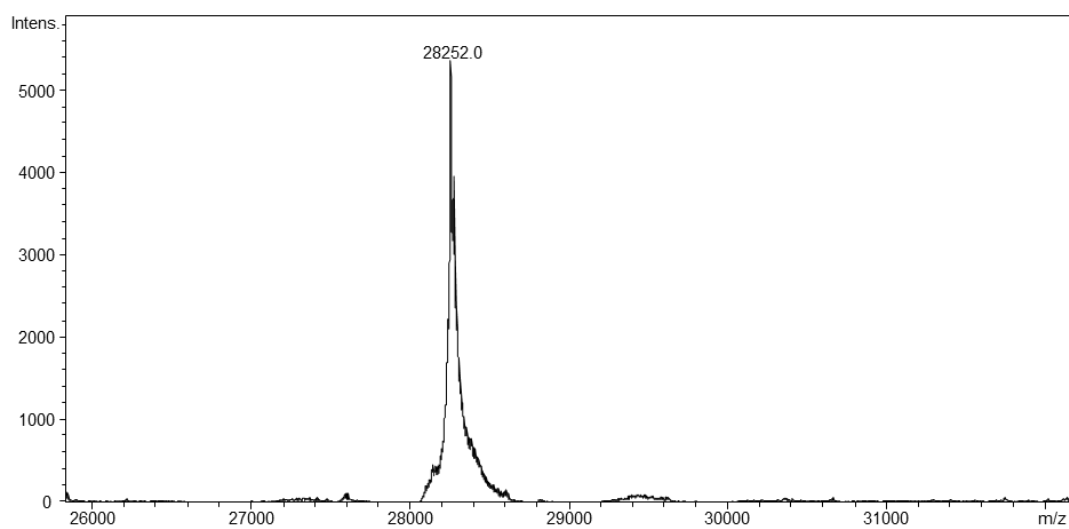
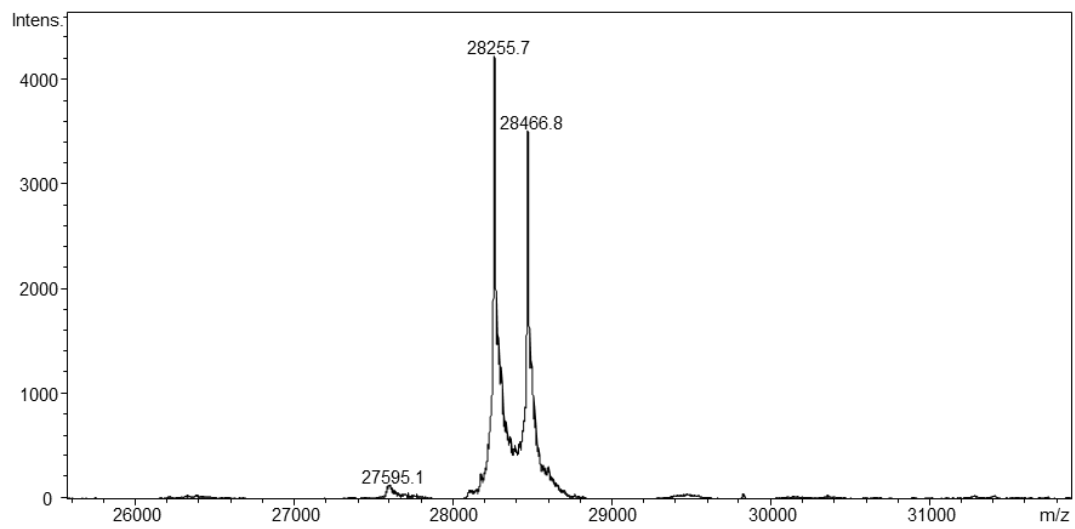


Figure A15. ESI-MS analysis of sfGFP-Y151TAG-BAOK (A) before and (B) after TFA deprotection.

A



B

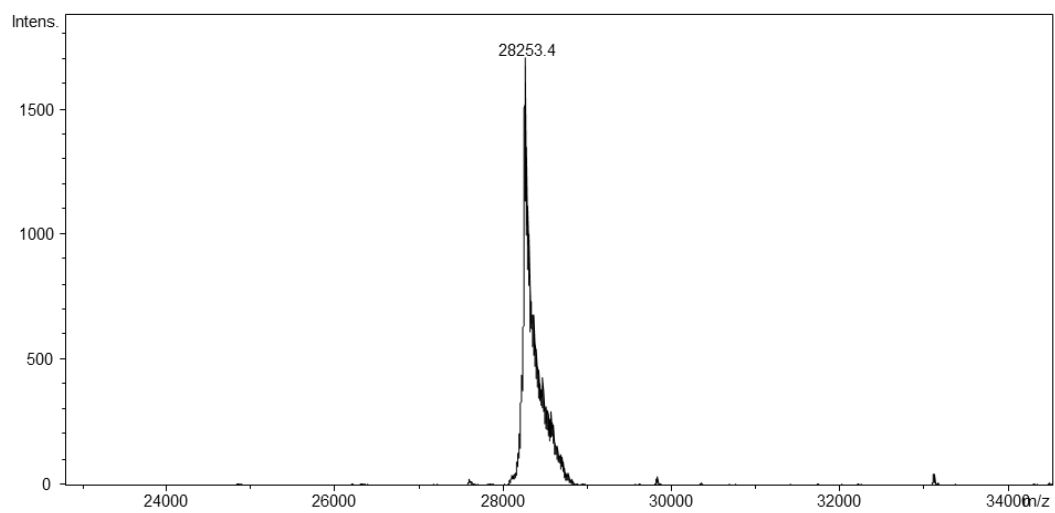


Figure A16. ESI-MS analysis of sfGFP-Y151TAG-PAOK (A) before and (B) after UV irradiation.

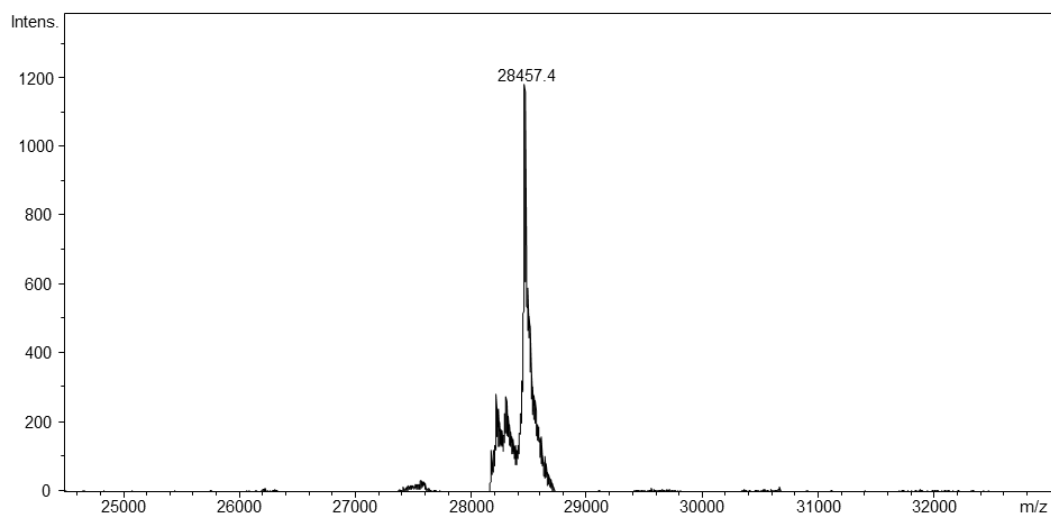


Figure A17. ESI-MS analysis of oxime ligation product of degraded sfGFP-PAOK with coumarin aldehyde (CA).

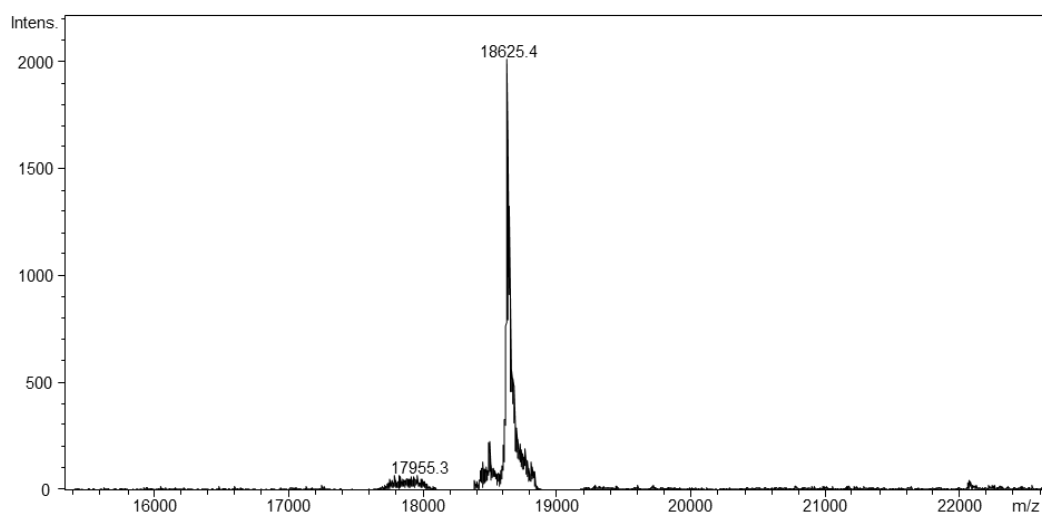


Figure A18. ESI-MS analysis of ligation product of degraded Myo-4TAG-PAOK with coumarin aldehyde (CA).

Plasmid maps

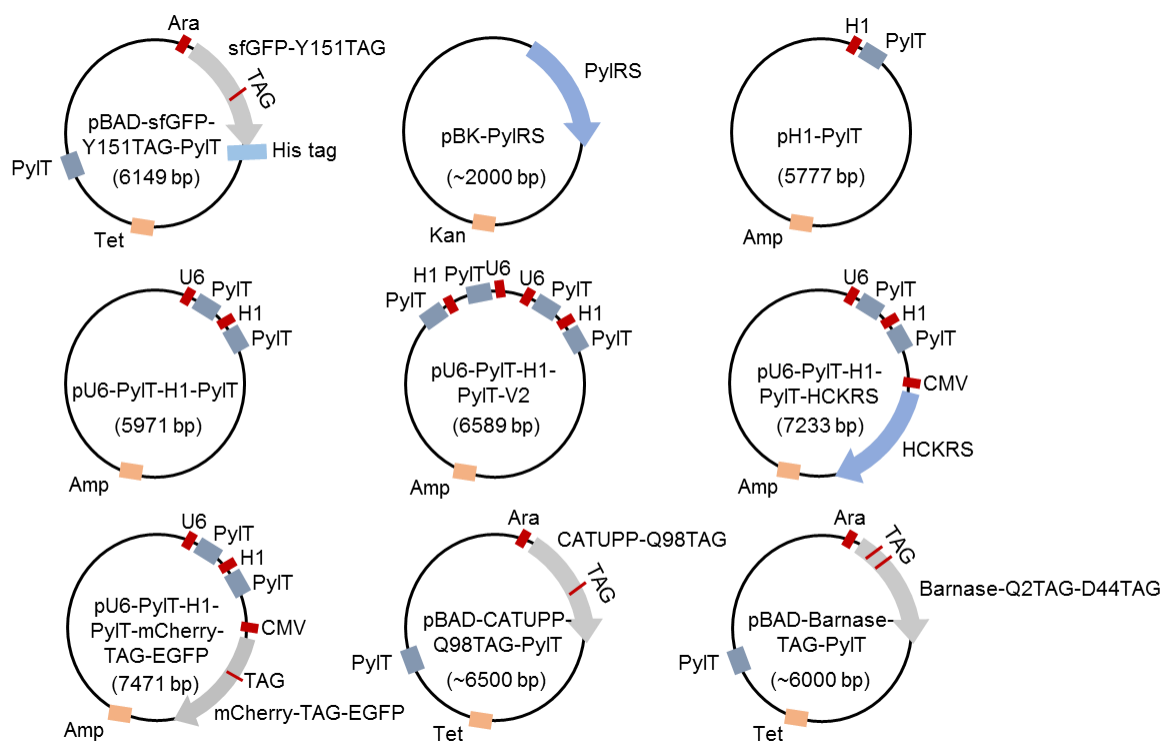


Figure A19. Newly constructed plasmids in Chapter 2.

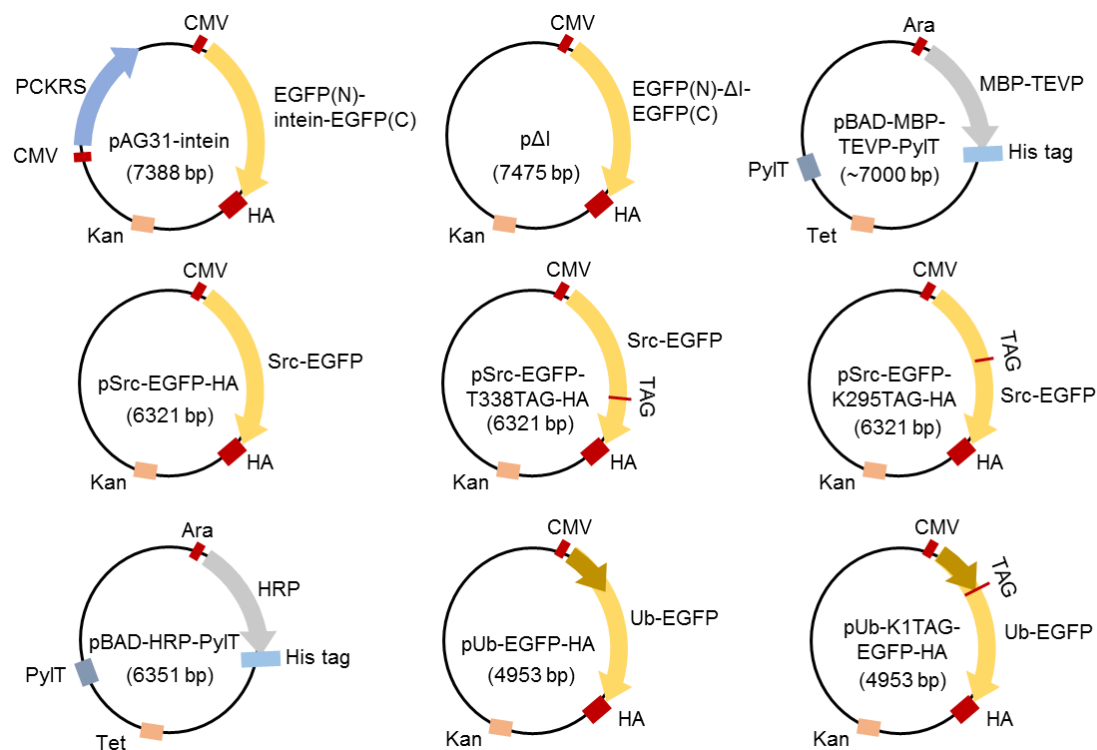


Figure A20. Newly constructed plasmids in Chapter 3.

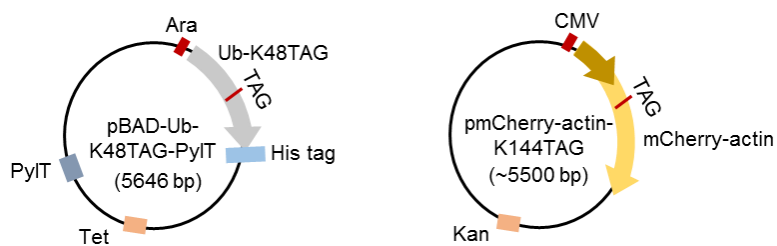


Figure A21. Newly constructed plasmids in Chapter 4.

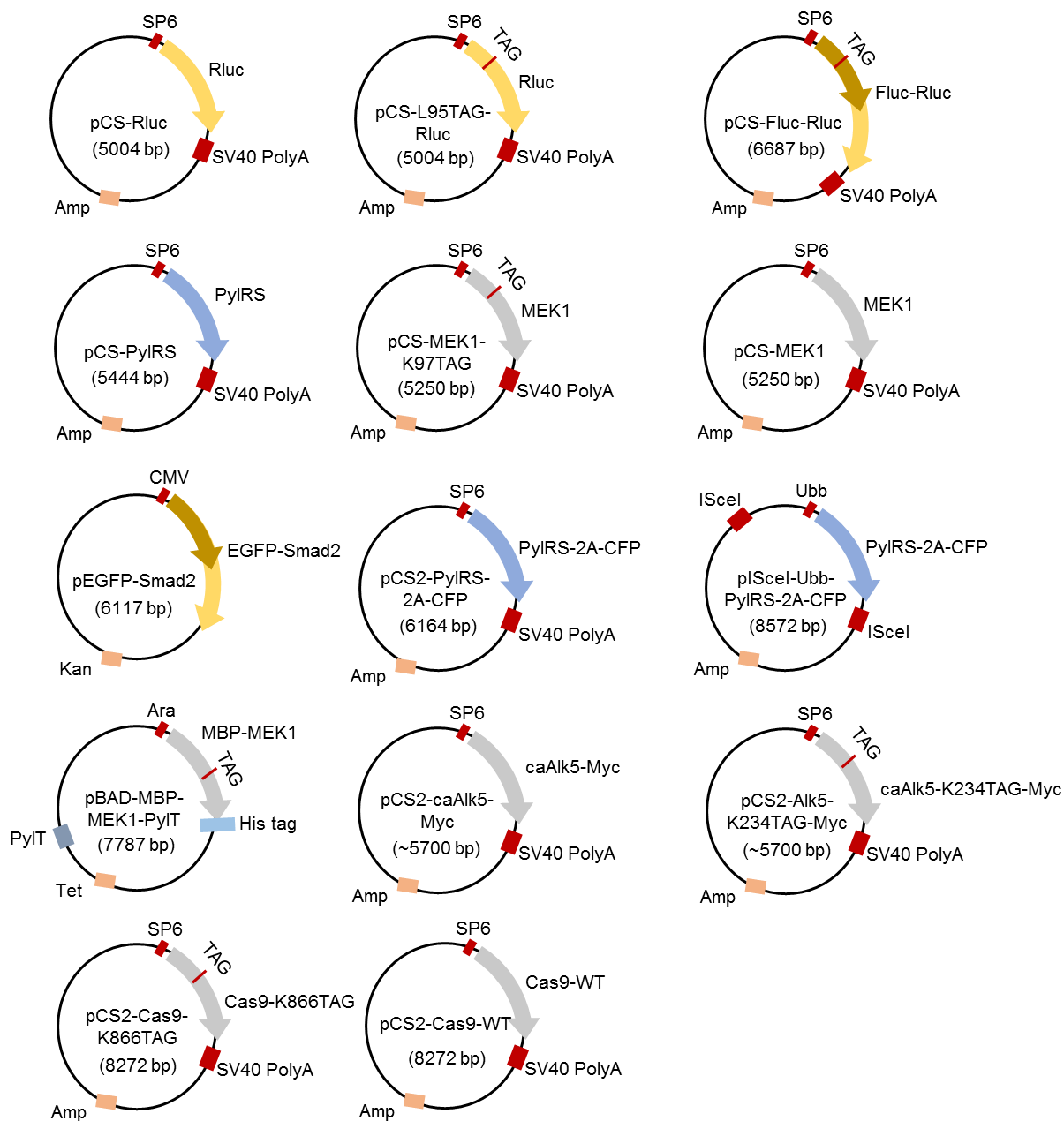


Figure A22. Newly constructed plasmids in Chapter 5.

General protocol

Restriction enzyme-based cloning

1. Digest the parent plasmid with restriction enzymes, using the protocol below. Incubate at 37 °C for 2 h, heat-inactivate at 75 °C for 30 min.

	Volume (μl)
DNA	80
Enzyme A	2
Enzyme B	2
10xCutSmart Buffer	20
H ₂ O	96
Total	200

2. Add 2 μl of Antarctic Phosphatase (NEB), 20 μl of 10x AP buffer to the digestion solution. Incubate at 37 °C for 1 h, run on 0.8% agarose gel (80 V for 45 min), extract the band corresponding to the backbone using the gel extraction kit (Omega). Elute with 30 μl of water.

3. Perform PCR for the target fragment (25 μl reaction). Purify the PCR product using the PCR cycle pure kit (Omega). Elute with 30 μl of water. Concentration should be around 30 ng/μl. Digest the elution product with 0.3 μl of enzyme A and 0.3 μl of enzyme B in a 30 μl reaction.

4. Calculate the volume of backbone (B) and insert (I): $c(B) \cdot \text{volume}(B) / \text{length}(B) \cdot 3 = c(I) \cdot \text{volume}(I) / \text{length}(I)$. The total backbone DNA is 50 ng.

5. Set up ligation reaction with 10x T4 ligase buffer (1 μl), T4 ligase buffer (0.2 μl). Fill up with water to 10 μl. Incubate at 16 °C overnight. Transform ligation reaction (3 μl) to the chemical competent cells (50 μl) the next day.

PCR reaction

1. Set up the PCR reaction according to the protocol below.

	Volume (µl)	Final Concentration
Plasmid (adjust to 5 ng/µl)	0.5	
Forward Primer (10 µM)	0.625	0.25 µM
Reverse Primer (10 µM)	0.625	0.25 µM
10 mM dNTP	0.5	20 µM
5xBuffer	5	
dd H ₂ O	17.5	
Polymerase (Fusion)	0.25	
Total	25	

2. Run PCR reaction according to the program below. Adjust the annealing temperature if the PCR amplification is not successful.

		30 cycles				
Temp (°C)	98	98	58	72	72	12
Time	30 s	10 s	30 s	30 s/kb	10 min	hold

Site directed mutagenesis

Note: The protocol is based on Agilent Quikchange Site-Directed Mutagenesis Kit:
<http://www.agilent.com/cs/library/usermanuals/Public/210513.pdf>

Primer design:

For the forward primer, at the 5' end and 3' end of the desired mutations, 18 bp of the original sequences adjacent to the target mutation site are designed. Both ends are further extended until a C/G is met. The reverse primer is the reverse complimentary sequence of the forward primer.

Experimental:

1. Make stock solution of primers, to the final concentration of 1 µg/µl.
2. Make working solution of primers, with the final concentration of 0.1 µg/µl.
3. Set up the PCR reaction according to the protocol below.

	Volume (µl)
H ₂ O	11.6
5x Phusion Buffer	4
Fwd Primer	0.5
Rev Primer	0.5
100% DMSO	0.6
5 ng/µl DNA	2
10 mM dNTP	0.4
Phusion Polymerase	0.4
Total	20

4. Run PCR reaction according to the program below.

		18 cycles				
Temp (°C)	95	95	58	68	68	12
Time	2 min	20 s	10 s	30 s/kb	5 min	hold

5. Add 0.4 µl of DpnI (NEB) enzyme to the PCR reaction. Incubate at 37 °C for 10 min. Transform 3 µl of the final product to chemical competent cells (50 µl).

Gibson assembly

Note: The protocol is from Miller lab: <http://miller-lab.net/MillerLab/protocols/molecular-biology-and-cloning/gibson-assembly/>

Primer design:

The fragments for Gibson assembly are first defined (and could be designed on vector NTI *in silico*). The Gibson assembly primers are typically 60 bp in length (30 bp of vector and 30 bp of insert). If the primers are not completely complimentary to each other, the overlap sequence needs to be at least 40 bp.

Experimental:

1. Make 5x ISO reaction buffer according to the protocol below.

	Volume	Final Concentration
1M Tris-HCl, pH 7.5	100 µl	500 mM
1M MgCl ₂	10 µl	50 mM
10 mM dNTP	20 µl	1 mM
1M DTT	10 µl	50 mM
PEG-8000	50 mg	25%
50 mM NAD	20 µl	5 mM
H ₂ O	adjusted	
Total	200 µl	

2. Make enzyme reaction master mix according to the protocol below. Aliquot in PCR tubes (15 µl each), and store at –20 °C fridge.

	Volume
5xISO reaction buffer	40 µl
10 U/µl T5 exo	0.1 µl
2 U/µl Phusion polymerase	2.5 µl
40 U/µl Taq ligase	20 µl
H ₂ O	87.4 µl
Total	150 µl

3. Prepare DNA mixture, with 50 ng of backbone, and 100 ng of each DNA fragments. Fill up with water to 5 μ l.

4. Add DNA mixture (5 μ l) to enzyme reaction master mix (15 μ l). Incubate at 50 °C for 1 h. Transform the product (3 μ l) to chemical competent cells (50 μ l).

Making chemical competent cells

Note: The protocol below (including transformation) was written and used by Hank Chou (former postdoc at Deiters group). A standard protocol could be found on NEB: <https://www.neb.com/protocols/2012/06/21/making-your-own-chemically-competent-cells>.

1. Inoculate a single colony into LB medium (3 ml) with proper antibiotic. Grow overnight at 37 °C shaker.

2. The next morning, add 1 ml of overnight culture to 50 ml of LB medium (with no antibiotics). Grow at 37 °C shaker, until OD₆₀₀ reaches 0.4~0.5.

IMPORTANT: frequently check OD after 2h. OD could not be higher than 0.5.

3. Chill the flask on the ice for 20 min.

4. Centrifuge the culture at 3,200 rpm in 50 ml tube, 10 min at 4 °C.

5. Prepare 10 ml of TSS (Transformation and Storage Solution): 8.5 ml of water, 0.212 g of LB broth, 1 g of PEG (MW: 5,600 - 8,000), 0.5 ml of 100% DMSO, 0.1 g of MgCl₂.

6. Resuspend the cells in N ml of ice-cold TSS. $N = 10 * OD_{600}$

7. Aliquot (50 μ l each) into prechilled 1.7ml tubes. Flash freezing in isopropanol/dry ice bath. Store in -80 °C freezer.

Testing the competency of chemical competent cells

1. Add DNA (< 5 μ l) to 50 μ l of chemical competent cells. Incubate on ice for 30 min. Heat shock at 42 °C for 30 s. Add 500 μ l of SOC medium. Shake at 37 °C for 1 h. Plate 100 μ l culture on LB agar. Incubate at 37 °C overnight.

2. Calculate the competency based on colony count the next day:

Competency = colony count * 5 / amount of DNA (μ g)

Making electrocompetent cells

Note: The protocol below (including transformation) is based on the NEB protocol, with some modifications: <https://www.neb.com/protocols/2012/06/21/making-your-own-electrocompetent-cells>.

1. Inoculate a single colony into LB medium (3 ml) with proper antibiotic. Grow overnight at 37 °C shaker.

2. The next morning, add 2 ml of overnight culture to 200 ml of LB medium (with no antibiotics). Grow at 37 °C shaker, until OD₆₀₀ reaches 0.4~0.5.

IMPORTANT: frequently check OD after 2h. OD could not be higher than 0.5.

3. Chill the flask on the ice for 20 min.

IMPORTANT: Prepare ice cold H₂O, cold 10% glycerol, chilled pipette tips and 50 ml tubes beforehand. All subsequent procedures should be performed on ice.

4. Centrifuge the culture at 4,000 rpm in 50 ml tube, 15 min at 4 °C.

5. Decant the supernatant, and resuspend the pellet in 50 ml of ice cold H₂O.

6. Centrifuge again using the same condition as above.

7. Decant the supernatant, and resuspend the pellet in 50 ml of ice cold H₂O.

8. Centrifuge again using the same condition as above.

9. Decant the supernatant, and resuspend the pellet in 40 ml of 10% cold glycerol.

10. Centrifuge again using the same condition as above.

11. Decant the supernatant, and add 500 µl of 10% cold glycerol. Resuspend.

IMPORTANT: slowly decant supernatant, as pellets lose adherence in 10% glycerol. Carefully aspirate the remaining supernatant with pipette.

12. Aliquot (50 µl each) into prechilled 1.7ml tubes. Flash freezing in isopropanol/dry ice bath. Store in -80 °C freezer.

Testing the competency of electrocompetent cells

1. Add 1 µl of DNA (1 ng/µl) to 50 µl of electrocompetent cells.

2. Set the electroporator to 1800 V.

3. Transfer electrocompetent cells to a prechilled 1 mm cuvette (VWR), tap on countertop twice, wipe moisture from the cuvette, and insert the cuvette into the electroporator.

4. Select the pre-programmed electroporation protocol for 1 mm cuvette.

5. Immediately add 1 ml of SOC medium to the cells, and transfer all to a sterile culture tube. Shake for 1.5 h at 37 °C.

6. Make 10x, 100x, 1000x dilution of the culture. Plate 100 µl of the diluted culture on LB plates containing the appropriate antibiotics. Grow overnight at 37 °C.

7. Calculate the competency using the following equation.

$$\text{Competency (CFU/}\mu\text{g)} = \text{colonies} / 0.001 \mu\text{g} * \text{total volume (1050 }\mu\text{l)} / \text{volume plated}$$

$$\text{For 100x dilution: volume plated} = 1 \mu\text{l}$$

For 10x dilution: volume plated = 10 μ l

Transfection in mammalian cells

Human embryonic kidney (HEK) 293T cells (ATCC, #CRL-11268) were grown in 200 μ l of DMEM (Dulbecco's Modified Eagle Medium) supplemented with 10% FBS, 1% Pen-Strep, and glutamine (2 mM) in 96-well plates in a humidified atmosphere with 5% CO₂ at 37 °C. At ~75% confluency, the culture medium were replaced with 180 μ l of fresh DMEM. UAA (at the final concentration of 1 mM concentration unless noted) was supplemented in the culture medium if needed. Branched polyethylene imine (BPEI, 1.5 μ l, 1 mg/ml, transfection reagent) was mixed with the designated plasmids (normally 200 ng for each) in 20 μ l of Opti-MEM. The mixture was incubated at room temperature for 10 min, and was then added to the culture medium. The cells were incubated in a humidified atmosphere with 5% CO₂ at 37 °C, and they were analyzed after 24 h or 48 h.

Western blot

Human embryonic kidney (HEK) 293T cells (ATCC, #CRL-11268) were grown in 2 ml of DMEM (Dulbecco's Modified Eagle Medium) supplemented with 10% FBS, 1% Pen-Strep, and glutamine (2 mM) in 6-well plates in a humidified atmosphere with 5% CO₂ at 37 °C. At ~75% confluency, HEK 293T cells were co-transfected with the designated plasmids (normally 2000 ng each). Branched polyethylene imine (BPEI, 15 μ l, 1 mg/ml) was used as transfection reagent for each well, using the transfection protocol described above. After 48 h of incubation, the cells were washed with chilled phosphate-buffer saline (PBS, 1 ml), and lysed in mammalian protein extraction buffer (250 μ l, GE Healthcare). The cell lysates were separated by 10% SDS-PAGE (run with 60V for 15 min, and 150 V for 45 min) and were transferred to a PVDF membrane (GE Healthcare). The membrane was blocked in tris-buffer saline (TBS) with 0.1% Tween 20 and 5% milk for 1 h. The blots were probed with the primary antibody (normally diluted as 1:1,000 in TBST) overnight at 4 °C, followed by incubation with secondary goat anti-rabbit IgG-HRP antibody (normally diluted as 1:20,000 in TBST) for 1 h at room temperature. The blots were further incubated with the SuperSignal West Pico working solution (mixture of the Stable Peroxide Solution and the Luminol/Enhancer Solution, 500 μ l each, Thermo Scientific) for 5 min at room temperature. The luminescence signal was detected by ChemiDoc (Chemi Hi Sensitivity setting, manual exposure time: 10 sec).

Buffer composition

SDS-PAGE:

Separating/Running gel:

Solution	Volume
40% acrylamide	1.25 ml
1.5 M Tris (pH 8.8)	1.25 ml
10% APS	50 μ l
10% SDS	50 μ l
TEMED	3 μ l
H ₂ O	2.4 ml
Total	5 ml

Stacking gel:

Solution	Volume
40% acrylamide	250 μ l
1 M Tris (pH 6.8)	250 μ l
10% APS	20 μ l
10% SDS	20 μ l
TEMED	2 μ l
H ₂ O	1.46 ml
Total	2 ml

10x SDS running buffer

- dissolve 144 g of glycine and 30.2 g of Tris Base in 900 ml of DI H₂O.
- add 10 g of SDS and mix.
- fill up to 1 L with DI H₂O.

SDS-PAGE staining buffer:

- add 100 ml of glacial acetic acid to 500 ml of DI H₂O.
- add 400 ml of methanol and mix.
- add 1g of Coomassie R250 dye (Briliant Blue) and mix.

SDS-PAGE destaining buffer:

- add 100 ml of glacial acetic acid to 700 ml of DI H₂O.
- add 200 ml of methanol.

4x SDS loading buffer (10 ml)

8 ml of 50% glycerol
2.4 ml of 1M Tris/HCl, pH 6.8
0.8 g of SDS
4 mg of bromophenol blue
0.5 ml of beta-mercaptoethanol

Phosphate-based buffer for protein purification

Elution buffer

NaH ₂ PO ₄	0.3 g
NaCl	0.877 g
Imidazole	0.85 g
10 M NaOH	107 µl

Lysis buffer

3 M NaCl	4.8 ml
1M Na ₂ HPO ₄	2.237 ml
1M NaH ₂ PO ₄	0.163 ml
elution buffer	2 ml
10 M NaOH	13 µl

Wash buffer

10 ml lysis buffer + 2.5 ml elution buffer

DMEM (Dulbecco's Modified Eagle's medium)

5.89 g DMEM
1.63 g Sodium bicarbonate
0.145 g L-glutamine
440 ml H₂O
Adjust pH to 7.4 (~500 µl of 1N HCl)
Add 50 ml of 10x FBS (Fetal Bovine Serum), and 5 ml of 100x P/S (Penicillin/Streptomycin) solution

Western blot

Transfer buffer:

2.25 g	Tris base
--------	-----------

10.5 g Glycine
1 g SDS
200 ml Methanol
Fill up to 1 L with DI H₂O

10x TBS (Tris-Buffer Saline)

24.2 g Tris
84 g NaCl
Adjust pH to 7.6 with 1N HCl
Fill up to 1 L with DI H₂O

1x TBST (Tris-Buffer Saline with 0.1% Tween 20)

To 500 ml of 1x TBS, add 0.5 ml of Tween 20

Blocking buffer

To 10 ml of 1x TBST, add 0.5 g of nonfat dry milk powder

Bibliography

1. Wang, L.; Schultz, P. G., Expanding the genetic code. *Angew Chem Int Edit* **2005**, *44* (1), 34-66.
2. Young, T. S.; Schultz, P. G., Beyond the canonical 20 amino acids: expanding the genetic lexicon. *The Journal of biological chemistry* **2010**, *285* (15), 11039-44.
3. Merrifield, R. B., Solid-Phase Peptide Synthesis. *Adv Enzymol Ramb* **1969**, *32*, 221.
4. Dawson, P. E.; Kent, S. B. H., Synthesis of native proteins by chemical ligation. *Annual review of biochemistry* **2000**, *69*, 923-960.
5. Muir, T. W., Semisynthesis of proteins by expressed protein ligation. *Annual review of biochemistry* **2003**, *72*, 249-289.
6. Noren, C. J.; Anthonycahill, S. J.; Griffith, M. C.; Schultz, P. G., A General-Method for Site-Specific Incorporation of Unnatural Amino-Acids into Proteins. *Science* **1989**, *244* (4901), 182-188.
7. Link, A. J.; Mock, M. L.; Tirrell, D. A., Non-canonical amino acids in protein engineering. *Curr Opin Biotech* **2003**, *14* (6), 603-609.
8. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., Expanding the genetic code of Escherichia coli. *Science* **2001**, *292* (5516), 498-500.
9. Sharp, P. M.; Cowe, E.; Higgins, D. G.; Shields, D. C.; Wolfe, K. H.; Wright, F., Codon Usage Patterns in Escherichia-Coli, Bacillus-Subtilis, Saccharomyces-Cerevisiae, Schizosaccharomyces-Pombe, Drosophila-Melanogaster and Homo-Sapiens - a Review of the Considerable within-Species Diversity. *Nucleic Acids Res* **1988**, *16* (17), 8207-8211.
10. Xie, J. M.; Schultz, P. G., An expanding genetic code. *Methods* **2005**, *36* (3), 227-238.
11. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z. W.; Schultz, P. G., An expanded eukaryotic genetic code. *Science* **2003**, *301* (5635), 964-967.
12. Zhang, Z. W.; Alfonta, L.; Tian, F.; Bursulaya, B.; Uryu, S.; King, D. S.; Schultz, P. G., Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *P Natl Acad Sci USA* **2004**, *101* (24), 8882-8887.
13. (a) Srinivasan, G.; James, C. M.; Krzycki, J. A., Pyrrolysine encoded by UAG in Archaea: Charging of a UAG-decoding specialized tRNA. *Science* **2002**, *296* (5572), 1459-1462; (b) Hao, B.; Gong, W. M.; Ferguson, T. K.; James, C. M.; Krzycki, J. A.; Chan, M. K., A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **2002**, *296* (5572), 1462-1466.
14. Gaston, M. A.; Zhang, L. W.; Green-Church, K. B.; Krzycki, J. A., The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. *Nature* **2011**, *471* (7340), 647-U131.
15. Blight, S. K.; Larue, R. C.; Mahapatra, A.; Longstaff, D. G.; Chang, E.; Zhao, G.; Kang, P. T.; Church-Church, K. B.; Chan, M. K.; Krzycki, J. A., Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature* **2004**, *431* (7006), 333-335.
16. Wan, W.; Tharp, J. M.; Liu, W. R., Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochimica et biophysica acta* **2014**, *1844* (6), 1059-70.
17. Kavran, J. M.; Gundliapalli, S.; O'Donoghue, P.; Englert, M.; Soell, D.; Steitz, T. A., Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *P Natl Acad Sci USA* **2007**, *104* (27), 11268-11273.
18. (a) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S., Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-

azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chemistry & biology* **2008**, 15 (11), 1187-97; (b) Nguyen, D. P.; Lusic, H.; Neumann, H.; Kapadnis, P. B.; Deiters, A.; Chin, J. W., Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/tRNA(CUA) pair and click chemistry. *Journal of the American Chemical Society* **2009**, 131 (25), 8720-1; (c) Kuboe, S.; Yoda, M.; Ogata, A.; Kitade, Y.; Tomari, Y.; Ueno, Y., Diazirine-containing RNA photocrosslinking probes for the study of siRNA-protein interactions. *Chemical communications* **2010**, 46 (39), 7367-9; (d) Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W., Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nature chemistry* **2012**, 4 (4), 298-304.

19. (a) Hao, Z.; Song, Y.; Lin, S.; Yang, M.; Liang, Y.; Wang, J.; Chen, P. R., A readily synthesized cyclic pyrrolysine analogue for site-specific protein "click" labeling. *Chemical communications* **2011**, 47 (15), 4502-4; (b) Chen, P. R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Geierstanger, B. H.; Schultz, P. G., A facile system for encoding unnatural amino acids in mammalian cells. *Angewandte Chemie* **2009**, 48 (22), 4052-5; (c) Virdee, S.; Kapadnis, P. B.; Elliott, T.; Lang, K.; Madrzak, J.; Nguyen, D. P.; Riechmann, L.; Chin, J. W., Traceless and site-specific ubiquitination of recombinant proteins. *Journal of the American Chemical Society* **2011**, 133 (28), 10708-11; (d) Gautier, A.; Nguyen, D. P.; Lusic, H.; An, W.; Deiters, A.; Chin, J. W., Genetically encoded photocontrol of protein localization in mammalian cells. *Journal of the American Chemical Society* **2010**, 132 (12), 4086-8; (e) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W., Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nature chemical biology* **2008**, 4 (4), 232-4.

20. (a) Wang, Y. S.; Fang, X.; Wallace, A. L.; Wu, B.; Liu, W. R., A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum. *Journal of the American Chemical Society* **2012**, 134 (6), 2950-3; (b) Tharp, J. M.; Wang, Y. S.; Lee, Y. J.; Yang, Y.; Liu, W. R., Genetic incorporation of seven ortho-substituted phenylalanine derivatives. *ACS chemical biology* **2014**, 9 (4), 884-90; (c) Wang, Y. S.; Fang, X.; Chen, H. Y.; Wu, B.; Wang, Z. U.; Hilty, C.; Liu, W. R., Genetic incorporation of twelve meta-substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant. *ACS chemical biology* **2013**, 8 (2), 405-15.

21. Xiao, H.; Peters, F. B.; Yang, P. Y.; Reed, S.; Chittuluru, J. R.; Schultz, P. G., Genetic incorporation of histidine derivatives using an engineered pyrrolysyl-tRNA synthetase. *ACS chemical biology* **2014**, 9 (5), 1092-6.

22. Arbely, E.; Torres-Kolbus, J.; Deiters, A.; Chin, J. W., Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *Journal of the American Chemical Society* **2012**, 134 (29), 11912-5.

23. Nguyen, D. P.; Mahesh, M.; Elsasser, S. J.; Hancock, S. M.; Uttamapinant, C.; Chin, J. W., Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *Journal of the American Chemical Society* **2014**, 136 (6), 2240-3.

24. Greiss, S.; Chin, J. W., Expanding the genetic code of an animal. *Journal of the American Chemical Society* **2011**, 133 (36), 14196-9.

25. Bianco, A.; Townsley, F. M.; Greiss, S.; Lang, K.; Chin, J. W., Expanding the genetic code of *Drosophila melanogaster*. *Nature chemical biology* **2012**, 8 (9), 748-50.

26. Ernst, R. J.; Krogager, T. P.; Maywood, E. S.; Zanchi, R.; Beranek, V.; Elliott, T. S.; Barry, N. P.; Hastings, M. H.; Chin, J. W., Genetic code expansion in the mouse brain. *Nature chemical biology* **2016**, 12 (10), 776-8.

27. Young, D. D.; Young, T. S.; Jahnz, M.; Ahmad, I.; Spraggon, G.; Schultz, P. G., An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **2011**, 50 (11), 1894-900.

28. Pedelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S., Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology* **2006**, 24 (1), 79-88.

29. Kim, J.; Seo, M. H.; Lee, S.; Cho, K.; Yang, A.; Woo, K.; Kim, H. S.; Park, H. S., Simple and efficient strategy for site-specific dual labeling of proteins for single-molecule fluorescence resonance energy transfer analysis. *Analytical chemistry* **2013**, 85 (3), 1468-74.
30. Wang, Y. S.; Russell, W. K.; Wang, Z.; Wan, W.; Dodd, L. E.; Pai, P. J.; Russell, D. H.; Liu, W. R., The de novo engineering of pyrrolysyl-tRNA synthetase for genetic incorporation of L-phenylalanine and its derivatives. *Molecular bioSystems* **2011**, 7 (3), 714-7.
31. Kobayashi, T.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S., Recognition of non-alpha-amino substrates by pyrrolysyl-tRNA synthetase. *Journal of molecular biology* **2009**, 385 (5), 1352-60.
32. Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S., Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochemical and biophysical research communications* **2008**, 371 (4), 818-22.
33. Zhang, M.; Lin, S.; Song, X.; Liu, J.; Fu, Y.; Ge, X.; Fu, X.; Chang, Z.; Chen, P. R., A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. *Nature chemical biology* **2011**, 7 (10), 671-7.
34. Nguyen, D. P.; Elliott, T.; Holt, M.; Muir, T. W.; Chin, J. W., Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. *Journal of the American Chemical Society* **2011**, 133 (30), 11418-21.
35. Lang, K.; Davis, L.; Wallace, S.; Mahesh, M.; Cox, D. J.; Blackman, M. L.; Fox, J. M.; Chin, J. W., Genetic Encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic Diels-Alder reactions. *Journal of the American Chemical Society* **2012**, 134 (25), 10317-20.
36. Lee, Y. J.; Wu, B.; Raymond, J. E.; Zeng, Y.; Fang, X.; Wooley, K. L.; Liu, W. R., A genetically encoded acrylamide functionality. *ACS chemical biology* **2013**, 8 (8), 1664-70.
37. Schmidt, M. J.; Borbas, J.; Drescher, M.; Summerer, D., A genetically encoded spin label for electron paramagnetic resonance distance measurements. *Journal of the American Chemical Society* **2014**, 136 (4), 1238-41.
38. Lacey, V. K.; Louie, G. V.; Noel, J. P.; Wang, L., Expanding the library and substrate diversity of the pyrrolysyl-tRNA synthetase to incorporate unnatural amino acids containing conjugated rings. *ChemBiochem* **2013**, 14 (16), 2100-5.
39. Hoppmann, C.; Lacey, V. K.; Louie, G. V.; Wei, J.; Noel, J. P.; Wang, L., Genetically encoding photoswitchable click amino acids in Escherichia coli and mammalian cells. *Angewandte Chemie* **2014**, 53 (15), 3932-6.
40. Jain, S.; Singh, R.; Gupta, M. N., Purification of recombinant green fluorescent protein by three-phase partitioning. *Journal of chromatography. A* **2004**, 1035 (1), 83-6.
41. Aubin-Tam, M. E.; Hamad-Schifferli, K., Structure and function of nanoparticle-protein conjugates. *Biomedical materials* **2008**, 3 (3), 034001.
42. Giljohann, D. A.; Seferos, D. S.; Daniel, W. L.; Massich, M. D.; Patel, P. C.; Mirkin, C. A., Gold nanoparticles for biology and medicine. *Angewandte Chemie* **2010**, 49 (19), 3280-94.
43. Kim, Y. P.; Daniel, W. L.; Xia, Z.; Xie, H.; Mirkin, C. A.; Rao, J., Bioluminescent nanosensors for protease detection based upon gold nanoparticle-luciferase conjugates. *Chemical communications* **2010**, 46 (1), 76-8.
44. (a) Lee, J. M.; Park, H. K.; Jung, Y.; Kim, J. K.; Jung, S. O.; Chung, B. H., Direct immobilization of protein g variants with various numbers of cysteine residues on a gold surface. *Analytical chemistry* **2007**, 79 (7), 2680-7; (b) Aubin-Tam, M. E.; Hamad-Schifferli, K., Gold nanoparticle-cytochrome C complexes: the effect of nanoparticle ligand charge on protein structure. *Langmuir : the ACS journal of surfaces and colloids* **2005**, 21 (26), 12080-4.
45. (a) Frasconi, M.; Mazzei, F.; Ferri, T., Protein immobilization at gold-thiol surfaces and potential for biosensing. *Analytical and bioanalytical chemistry* **2010**, 398 (4), 1545-64; (b) Abad, J. M.; Mertens, S. F.; Pita, M.; Fernandez, V. M.; Schiffrin, D. J., Functionalization of thioctic acid-

capped gold nanoparticles for specific immobilization of histidine-tagged proteins. *Journal of the American Chemical Society* **2005**, 127 (15), 5689-94.

46. Ulman, A., Formation and Structure of Self-Assembled Monolayers. *Chemical reviews* **1996**, 96 (4), 1533-1554.

47. (a) Prisco, U.; Leung, C.; Xirouchaki, C.; Jones, C. H.; Heath, J. K.; Palmer, R. E., Residue-specific immobilization of protein molecules by size-selected clusters. *Journal of the Royal Society, Interface* **2005**, 2 (3), 169-75; (b) Patel, J. D.; O'Carra, R.; Jones, J.; Woodward, J. G.; Mumper, R. J., Preparation and characterization of nickel nanoparticles for binding to his-tag proteins and antigens. *Pharmaceutical research* **2007**, 24 (2), 343-52.

48. Valera, E.; Ramon-Azcon, J.; Sanchez, F. J.; Marco, M. P.; Rodriguez, A., Conductimetric immunosensor for atrazine detection based on antibodies labelled with gold nanoparticles. *Sensor Actuat B-Chem* **2008**, 134 (1), 95-103.

49. Thobhani, S.; Attree, S.; Boyd, R.; Kumarswami, N.; Noble, J.; Szymanski, M.; Porter, R. A., Bioconjugation and characterisation of gold colloid-labelled proteins. *J Immunol Methods* **2010**, 356 (1-2), 60-69.

50. Jans, H.; Liu, X.; Austin, L.; Maes, G.; Huo, Q., Dynamic Light Scattering as a Powerful Tool for Gold Nanoparticle Bioconjugation and Biomolecular Binding Studies. *Analytical chemistry* **2009**, 81 (22), 9425-9432.

51. Boulos, S. P.; Davis, T. A.; Yang, J. A.; Lohse, S. E.; Alkilany, A. M.; Holland, L. A.; Murphy, C. J., Nanoparticle-Protein Interactions: A Thermodynamic and Kinetic Study of the Adsorption of Bovine Serum Albumin to Gold Nanoparticle Surfaces. *Langmuir : the ACS journal of surfaces and colloids* **2013**, 29 (48), 14984-14996.

52. Zheng, M.; Huang, X., Nanoparticles comprising a mixed monolayer for specific bindings with biomolecules. *Journal of the American Chemical Society* **2004**, 126 (38), 12047-54.

53. Dzimitrowicz, A.; Jamroz, P.; Greda, K.; Nowak, P.; Nyk, M.; Pohl, P., The influence of stabilizers on the production of gold nanoparticles by direct current atmospheric pressure glow microdischarge generated in contact with liquid flowing cathode. *Journal of nanoparticle research : an interdisciplinary forum for nanoscale science and technology* **2015**, 17 (4), 185.

54. Sedighi, A.; Krull, U. J., Rapid Immobilization of Oligonucleotides at High Density on Semiconductor Quantum Dots and Gold Nanoparticles. *Langmuir : the ACS journal of surfaces and colloids* **2016**, 32 (50), 13500-13509.

55. De, M.; Rana, S.; Akpinar, H.; Miranda, O. R.; Arvizo, R. R.; Bunz, U. H.; Rotello, V. M., Sensing of proteins in human serum using conjugates of nanoparticles and green fluorescent protein. *Nature chemistry* **2009**, 1 (6), 461-5.

56. Bunz, U. H. F.; Rotello, V. M., Gold Nanoparticle-Fluorophore Complexes: Sensitive and Discerning "Noses" for Biosystems Sensing. *Angew Chem Int Edit* **2010**, 49 (19), 3268-3279.

57. Saiyed, Z. M.; Sharma, S.; Godawat, R.; Telang, S. D.; Ramchand, C. N., Activity and stability of alkaline phosphatase (ALP) immobilized onto magnetic nanoparticles (Fe₃O₄). *J Biotechnol* **2007**, 131 (3), 240-244.

58. Shay, K. P.; Moreau, R. F.; Smith, E. J.; Smith, A. R.; Hagen, T. M., Alpha-lipoic acid as a dietary supplement: molecular mechanisms and therapeutic potential. *Biochimica et biophysica acta* **2009**, 1790 (10), 1149-60.

59. Huang, C.; Yin, Q.; Zhu, W.; Yang, Y.; Wang, X.; Qian, X.; Xu, Y., Highly selective fluorescent probe for vicinal-dithiol-containing proteins and in situ imaging in living cells. *Angewandte Chemie* **2011**, 50 (33), 7551-6.

60. Luo, J.; Upreti, R.; Naro, Y.; Chou, C.; Nguyen, D. P.; Chin, J. W.; Deiters, A., Genetically encoded optochemical probes for simultaneous fluorescence reporting and light activation of protein function with two-photon excitation. *Journal of the American Chemical Society* **2014**, 136 (44), 15551-8.

61. Luo, J.; Liu, Q.; Morihira, K.; Deiters, A., Small-molecule control of protein function through Staudinger reduction. *Nature chemistry* **2016**, 8 (11), 1027-1034.

62. Amyes, T. L.; Jencks, W. P., Lifetimes of Oxocarbenium Ions in Aqueous-Solution from Common Ion Inhibition of the Solvolysis of Alpha-Azido Ethers by Added Azide Ion. *Journal of the American Chemical Society* **1989**, *111* (20), 7888-7900.
63. Polcarpo, C.; Ambrogelly, A.; Berube, A.; Winbush, S. M.; McCloskey, J. A.; Crain, P. F.; Wood, J. L.; Soll, D., An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc Natl Acad Sci U S A* **2004**, *101* (34), 12450-4.
64. Mayer, G.; Heckel, A., Biologically active molecules with a "light switch". *Angewandte Chemie* **2006**, *45* (30), 4900-21.
65. Woll, D.; Laimgruber, S.; Galetskaya, M.; Smirnova, J.; Pfeleiderer, W.; Heinz, B.; Gilch, P.; Steiner, U. E., On the mechanism of intramolecular sensitization of photocleavage of the 2-(2-nitrophenyl)propoxycarbonyl (NPPOC) protecting group. *Journal of the American Chemical Society* **2007**, *129* (40), 12148-58.
66. Luo, J.; Torres-Kolbus, J.; Liu, J.; Deiters, A., Genetic Encoding of Photocaged Tyrosines with Improved Light-Activation Properties for the Optical Control of Protease Function. *Chembiochem* **2017**.
67. Stephanopoulos, N.; Francis, M. B., Choosing an effective protein bioconjugation strategy. *Nature chemical biology* **2011**, *7* (12), 876-84.
68. Forbes, C. R.; Zondlo, N. J., Synthesis of thiophenylalanine-containing peptides via Cu(I)-mediated cross-coupling. *Organic letters* **2012**, *14* (2), 464-7.
69. Wilkins, B. J.; Marionni, S.; Young, D. D.; Liu, J.; Wang, Y.; Di Salvo, M. L.; Deiters, A.; Cropp, T. A., Site-specific incorporation of fluorotyrosines into proteins in Escherichia coli by photochemical disguise. *Biochemistry* **2010**, *49* (8), 1557-9.
70. Davidenko, T. I.; Bondarenko, G. I., Reduction of nitro-substituted compounds by native and immobilized Escherichia coli cells. *Appl Biochem Micro* **2000**, *36* (1), 63-68.
71. (a) Hensarling, R. M.; Hoff, E. A.; LeBlanc, A. P.; Guo, W.; Rahane, S. B.; Patton, D. L., Photocaged pendent thiol polymer brush surfaces for postpolymerization modifications via thiol-click chemistry. *J Polym Sci Pol Chem* **2013**, *51* (5), 1079-1090; (b) Delaittre, G.; Pauloehrl, T.; Bastmeyer, M.; Barner-Kowollik, C., Acrylamide-Based Copolymers Bearing Photoreleasable Thiols for Subsequent Thiol-Ene Functionalization. *Macromolecules* **2012**, *45* (4), 1792-1802.
72. Wang, F.; Niu, W.; Guo, J.; Schultz, P. G., Unnatural amino acid mutagenesis of fluorescent proteins. *Angewandte Chemie* **2012**, *51* (40), 10132-5.
73. Monden, Y.; Hamano Takaku, F.; Shindo Okada, N.; Nishimura, S., Azatyrosine. Mechanism of action for conversion of transformed phenotype to normal. *Annals of the New York Academy of Sciences* **1999**, *886*, 109-21.
74. Barone, V.; Adamo, C., Proton-Transfer in Excited Electronic States - Environmental-Effects on the Tautomerization of 2-Pyridone. *J Photoch Photobio A* **1994**, *80* (1-3), 211-219.
75. Burke, T. R.; Yao, Z. J.; Ye, B.; Miyoshi, K.; Otaka, A.; Wu, L.; Zhang, Z. Y., Phospho-azatyrosine, a less effective protein-tyrosine phosphatase substrate than phosphotyrosine. *Bioorganic & medicinal chemistry letters* **2001**, *11* (10), 1265-1268.
76. Hamano-Takaku, F.; Iwama, T.; Saito-Yano, S.; Takaku, K.; Monden, Y.; Kitabatake, M.; Soll, D.; Nishimura, S., A mutant Escherichia coli tyrosyl-tRNA synthetase utilizes the unnatural amino acid azatyrosine more efficiently than tyrosine. *The Journal of biological chemistry* **2000**, *275* (51), 40324-8.
77. Wu, Y.; Fried, S. D.; Boxer, S. G., Dissecting Proton Delocalization in an Enzyme's Hydrogen Bond Network with Unnatural Amino Acids. *Biochemistry* **2015**, *54* (48), 7110-9.
78. Sigala, P. A.; Fafarman, A. T.; Schwans, J. P.; Fried, S. D.; Fenn, T. D.; Caaveiro, J. M.; Pybus, B.; Ringe, D.; Petsko, G. A.; Boxer, S. G.; Herschlag, D., Quantitative dissection of hydrogen bond-mediated proton transfer in the ketosteroid isomerase active site. *Proc Natl Acad Sci U S A* **2013**, *110* (28), E2552-61.
79. Hoyle, C. E.; Bowman, C. N., Thiol-ene click chemistry. *Angewandte Chemie* **2010**, *49* (9), 1540-73.

80. Hoyle, C. E.; Lowe, A. B.; Bowman, C. N., Thiol-click chemistry: a multifaceted toolbox for small molecule and polymer synthesis. *Chem Soc Rev* **2010**, 39 (4), 1355-1387.
81. Dondoni, A.; Marra, A., Recent applications of thiol-ene coupling as a click process for glycoconjugation. *Chem Soc Rev* **2012**, 41 (2), 573-586.
82. Torres-Kolbus, J.; Chou, C. J.; Liu, J. H.; Deiters, A., Synthesis of Non-linear Protein Dimers through a Genetically Encoded Thiol-ene Reaction. *PloS one* **2014**, 9 (9).
83. Li, Y. M.; Yang, M. Y.; Huang, Y. C.; Song, X. D.; Liu, L.; Chen, P. R., Genetically encoded alkenyl-pyrrolysine analogues for thiol-ene reaction mediated site-specific protein labeling. *Chem Sci* **2012**, 3 (9), 2766-2770.
84. (a) Behbehani, G. R.; Saboury, A. A.; Taleshi, E., A direct calorimetric determination of denaturation enthalpy for lysozyme in sodium dodecyl sulfate. *Colloids and surfaces. B, Biointerfaces* **2008**, 61 (2), 224-8; (b) Michaux, C.; Pomroy, N. C.; Prive, G. G., Refolding SDS-denatured proteins by the addition of amphipathic cosolvents. *Journal of molecular biology* **2008**, 375 (5), 1477-88.
85. (a) Durham, O. Z.; Krishnan, S.; Shipp, D. A., Polymer Microspheres Prepared by Water-Borne Thiol-Ene Suspension Photopolymerization. *Acs Macro Lett* **2012**, 1 (9), 1134-1137; (b) Jiang, W.; Wang, W.; Pan, B.; Zhang, Q.; Zhang, W.; Lv, L., Facile fabrication of magnetic chitosan beads of fast kinetics and high capacity for copper removal. *ACS applied materials & interfaces* **2014**, 6 (5), 3421-6.
86. Lu, Y.; Yeung, N.; Sieracki, N.; Marshall, N. M., Design of functional metalloproteins. *Nature* **2009**, 460 (7257), 855-62.
87. DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A., De novo design and structural characterization of proteins and metalloproteins. *Annual review of biochemistry* **1999**, 68, 779-819.
88. Xie, J.; Liu, W.; Schultz, P. G., A genetically encoded bidentate, metal-binding amino acid. *Angewandte Chemie* **2007**, 46 (48), 9239-42.
89. Kang, M. C.; Light, K.; Ai, H. W.; Shen, W. J.; Kim, C. H.; Chen, P. R.; Lee, H. S.; Solomon, E. I.; Schultz, P. G., Evolution of Iron(II)-Finger Peptides by Using a Bipyridyl Amino Acid. *Chembiochem* **2014**, 15 (6), 822-825.
90. Mills, J. H.; Khare, S. D.; Bolduc, J. M.; Forouhar, F.; Mulligan, V. K.; Lew, S.; Seetharaman, J.; Tong, L.; Stoddard, B. L.; Baker, D., Computational design of an unnatural amino acid dependent metalloprotein with atomic level accuracy. *Journal of the American Chemical Society* **2013**, 135 (36), 13393-9.
91. Yanagisawa, T.; Hino, N.; Iraha, F.; Mukai, T.; Sakamoto, K.; Yokoyama, S., Wide-range protein photo-crosslinking achieved by a genetically encoded N(epsilon)-(benzyloxycarbonyl)lysine derivative with a diazirinyl moiety. *Molecular bioSystems* **2012**, 8 (4), 1131-5.
92. Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G., An enhanced system for unnatural amino acid mutagenesis in E. coli. *Journal of molecular biology* **2010**, 395 (2), 361-74.
93. Pott, M.; Schmidt, M. J.; Summerer, D., Evolved sequence contexts for highly efficient amber suppression with noncanonical amino acids. *ACS chemical biology* **2014**, 9 (12), 2815-22.
94. Chatterjee, A.; Xiao, H.; Bollong, M.; Ai, H. W.; Schultz, P. G., Efficient viral delivery system for unnatural amino acid mutagenesis in mammalian cells. *Proc Natl Acad Sci U S A* **2013**, 110 (29), 11803-8.
95. Xiao, H.; Chatterjee, A.; Choi, S. H.; Bajjuri, K. M.; Sinha, S. C.; Schultz, P. G., Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. *Angewandte Chemie* **2013**, 52 (52), 14080-3.
96. Paddison, P. J.; Caudy, A. A.; Bernstein, E.; Hannon, G. J.; Conklin, D. S., Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes & development* **2002**, 16 (8), 948-58.

97. Schmied, W. H.; Elsasser, S. J.; Uttamapinant, C.; Chin, J. W., Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. *Journal of the American Chemical Society* **2014**, *136* (44), 15577-83.
98. Cohen, S.; Arbely, E., Single-Plasmid-Based System for Efficient Noncanonical Amino Acid Mutagenesis in Cultured Mammalian Cells. *Chembiochem* **2016**, *17* (11), 1008-11.
99. Xie, J.; Schultz, P. G., An expanding genetic code. *Methods* **2005**, *36* (3), 227-38.
100. Kuhn, S. M.; Rubini, M.; Fuhrmann, M.; Theobald, I.; Skerra, A., Engineering of an orthogonal aminoacyl-tRNA synthetase for efficient incorporation of the non-natural amino acid O-methyl-L-tyrosine using fluorescence-based bacterial cell sorting. *Journal of molecular biology* **2010**, *404* (1), 70-87.
101. Melancon, C. E., 3rd; Schultz, P. G., One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases. *Bioorganic & medicinal chemistry letters* **2009**, *19* (14), 3845-7.
102. Rackham, O.; Chin, J. W., A network of orthogonal ribosome x mRNA pairs. *Nature chemical biology* **2005**, *1* (3), 159-66.
103. Tiraby, M.; Cazaux, C.; Baron, M.; Drocourt, D.; Reynes, J. P.; Tiraby, G., Concomitant expression of E. coli cytosine deaminase and uracil phosphoribosyltransferase improves the cytotoxicity of 5-fluorocytosine. *FEMS microbiology letters* **1998**, *167* (1), 41-9.
104. Kavran, J. M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Soll, D.; Steitz, T. A., Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc Natl Acad Sci U S A* **2007**, *104* (27), 11268-73.
105. Cropp, T. A.; Anderson, J. C.; Chin, J. W., Reprogramming the amino-acid substrate specificity of orthogonal aminoacyl-tRNA synthetases to expand the genetic code of eukaryotic cells. *Nature protocols* **2007**, *2* (10), 2590-600.
106. Wan, W.; Huang, Y.; Wang, Z. Y.; Russell, W. K.; Pai, P. J.; Russell, D. H.; Liu, W. R., A Facile System for Genetic Incorporation of Two Different Noncanonical Amino Acids into One Protein in Escherichia coli. *Angew Chem Int Edit* **2010**, *49* (18), 3211-3214.
107. (a) Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S., Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochemical and biophysical research communications* **2008**, *371* (4), 818-822; (b) Yanagisawa, T.; Hino, N.; Iraha, F.; Mukai, T.; Sakamoto, K.; Yokoyama, S., Wide-range protein photo-crosslinking achieved by a genetically encoded N-epsilon-(benzyloxycarbonyl)lysine derivative with a diazirinyl moiety. *Molecular bioSystems* **2012**, *8* (4), 1131-1135.
108. Pace, N. J.; Weerapana, E., Diverse functional roles of reactive cysteines. *ACS chemical biology* **2013**, *8* (2), 283-96.
109. Lindahl, M.; Mata-Cabana, A.; Kieselbach, T., The disulfide proteome and other reactive cysteine proteomes: analysis and functional significance. *Antioxidants & redox signaling* **2011**, *14* (12), 2581-642.
110. Spasser, L.; Brik, A., Chemistry and biology of the ubiquitin signal. *Angewandte Chemie* **2012**, *51* (28), 6840-62.
111. Kemp, M.; Go, Y. M.; Jones, D. P., Nonequilibrium thermodynamics of thiol/disulfide redox systems: a perspective on redox systems biology. *Free radical biology & medicine* **2008**, *44* (6), 921-37.
112. Shao, L. E.; Tanaka, T.; Gribi, R.; Yu, J., Thioredoxin-related regulation of NO/NOS activities. *Annals of the New York Academy of Sciences* **2002**, *962*, 140-50.
113. Greaves, J.; Chamberlain, L. H., Palmitoylation-dependent protein sorting. *The Journal of cell biology* **2007**, *176* (3), 249-54.
114. Giles, N. M.; Giles, G. I.; Jacob, C., Multiple roles of cysteine in biocatalysis. *Biochemical and biophysical research communications* **2003**, *300* (1), 1-4.

115. Fass, D., Disulfide bonding in protein biophysics. *Annual review of biophysics* **2012**, *41*, 63-79.
116. Uprety, R.; Luo, J.; Liu, J.; Naro, Y.; Samanta, S.; Deiters, A., Genetic encoding of caged cysteine and caged homocysteine in bacterial and mammalian cells. *Chembiochem* **2014**, *15* (12), 1793-9.
117. Kang, J. Y.; Kawaguchi, D.; Coin, I.; Xiang, Z.; O'Leary, D. D.; Slesinger, P. A.; Wang, L., In vivo expression of a light-activatable potassium channel using unnatural amino acids. *Neuron* **2013**, *80* (2), 358-70.
118. Lemke, E. A.; Summerer, D.; Geierstanger, B. H.; Brittain, S. M.; Schultz, P. G., Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nature chemical biology* **2007**, *3* (12), 769-72.
119. Ren, W.; Ji, A.; Ai, H. W., Light activation of protein splicing with a photocaged fast intein. *Journal of the American Chemical Society* **2015**, *137* (6), 2155-8.
120. Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G., A genetically encoded photocaged amino acid. *Journal of the American Chemical Society* **2004**, *126* (44), 14306-7.
121. (a) Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G., A genetically encoded fluorescent amino acid. *Proc Natl Acad Sci U S A* **2006**, *103* (26), 9785-9; (b) Lee, H. S.; Guo, J. T.; Lemke, E. A.; Dimla, R. D.; Schultz, P. G., Genetic Incorporation of a Small, Environmentally Sensitive, Fluorescent Probe into Proteins in *Saccharomyces cerevisiae*. *Journal of the American Chemical Society* **2009**, *131* (36), 12921-+.
122. (a) Elleuche, S.; Poggeler, S., Inteins, valuable genetic elements in molecular biology and biotechnology. *Applied microbiology and biotechnology* **2010**, *87* (2), 479-89; (b) Volkmann, G.; Mootz, H. D., Recent progress in intein research: from mechanism to directed evolution and applications. *Cellular and molecular life sciences : CMLS* **2013**, *70* (7), 1185-206.
123. Mootz, H. D.; Muir, T. W., Protein splicing triggered by a small molecule. *Journal of the American Chemical Society* **2002**, *124* (31), 9044-5.
124. (a) Buskirk, A. R.; Ong, Y. C.; Gartner, Z. J.; Liu, D. R., Directed evolution of ligand dependence: small-molecule-activated protein splicing. *Proc Natl Acad Sci U S A* **2004**, *101* (29), 10505-10; (b) Peck, S. H.; Chen, I.; Liu, D. R., Directed evolution of a small-molecule-triggered intein with improved splicing properties in mammalian cells. *Chemistry & biology* **2011**, *18* (5), 619-30.
125. Vila-Perello, M.; Hori, Y.; Ribo, M.; Muir, T. W., Activation of protein splicing by protease- or light-triggered O to N acyl migration. *Angewandte Chemie* **2008**, *47* (40), 7764-7.
126. Van Roey, P.; Pereira, B.; Li, Z.; Hiraga, K.; Belfort, M.; Derbyshire, V., Crystallographic and mutational studies of *Mycobacterium tuberculosis* recA mini-inteins suggest a pivotal role for a highly conserved aspartate residue. *Journal of molecular biology* **2007**, *367* (1), 162-73.
127. Waugh, D. S., An overview of enzymatic reagents for the removal of affinity tags. *Protein expression and purification* **2011**, *80* (2), 283-93.
128. Kapust, R. B.; Tozser, J.; Fox, J. D.; Anderson, D. E.; Cherry, S.; Copeland, T. D.; Waugh, D. S., Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein engineering* **2001**, *14* (12), 993-1000.
129. Garske, A. L.; Peters, U.; Cortesi, A. T.; Perez, J. L.; Shokat, K. M., Chemical genetic strategy for targeting protein kinases based on covalent complementarity. *Proc Natl Acad Sci U S A* **2011**, *108* (37), 15046-52.
130. Zhang, G.; Li, J.; Xie, R.; Fan, X.; Liu, Y.; Zheng, S.; Ge, Y.; Chen, P. R., Bioorthogonal Chemical Activation of Kinases in Living Systems. *ACS central science* **2016**, *2* (5), 325-31.
131. Kamps, M. P.; Sefton, B. M., Neither arginine nor histidine can carry out the function of lysine-295 in the ATP-binding site of p60src. *Molecular and cellular biology* **1986**, *6* (3), 751-7.
132. Bjorge, J. D.; Pang, A.; Fujita, D. J., Identification of protein-tyrosine phosphatase 1B as the major tyrosine phosphatase activity capable of dephosphorylating and activating c-Src in

several human breast cancer cell lines. *Journal of Biological Chemistry* **2000**, 275 (52), 41439-41446.

133. (a) Lee, H. M.; Larson, D. R.; Lawrence, D. S., Illuminating the chemistry of life: design, synthesis, and applications of "caged" and related photoresponsive compounds. *ACS chemical biology* **2009**, 4 (6), 409-27; (b) Riggsbee, C. W.; Deiters, A., Recent advances in the photochemical control of protein function. *Trends in biotechnology* **2010**, 28 (9), 468-75.

134. Szymanski, W.; Beierle, J. M.; Kistemaker, H. A.; Velema, W. A.; Feringa, B. L., Reversible photocontrol of biological systems by the incorporation of molecular photoswitches. *Chemical reviews* **2013**, 113 (8), 6114-78.

135. Beharry, A. A.; Woolley, G. A., Azobenzene photoswitches for biomolecules. *Chem Soc Rev* **2011**, 40 (8), 4422-37.

136. (a) Liu, D.; Karanicolas, J.; Yu, C.; Zhang, Z. H.; Woolley, G. A., Site-specific incorporation of photoisomerizable azobenzene groups into ribonuclease S. *Bioorganic & medicinal chemistry letters* **1997**, 7 (20), 2677-2680; (b) Bose, M.; Groff, D.; Xie, J. M.; Brustad, E.; Schultz, P. G., The incorporation of a photoisomerizable amino acid into proteins in E. coli. *Journal of the American Chemical Society* **2006**, 128 (2), 388-389; (c) Muranaka, N.; Hohsaka, T.; Sisido, M., Photoswitching of peroxidase activity by position-specific incorporation of a photoisomerizable non-natural amino acid into horseradish peroxidase. *FEBS Lett* **2002**, 510 (1-2), 10-12.

137. Padilla, M. S.; Young, D. D., Photosensitive GFP mutants containing an azobenzene unnatural amino acid. *Bioorganic & medicinal chemistry letters* **2015**, 25 (3), 470-3.

138. Azevedo, A. M.; Martins, V. C.; Prazeres, D. M.; Vojinovic, V.; Cabral, J. M.; Fonseca, L. P., Horseradish peroxidase: a valuable tool in biotechnology. *Biotechnology annual review* **2003**, 9, 199-247.

139. Grigorenko, V.; Andreeva, I.; Borchers, T.; Spener, F.; Egorov, A., A genetically engineered fusion protein with horseradish peroxidase as a marker enzyme for use in competitive immunoassays. *Analytical chemistry* **2001**, 73 (6), 1134-9.

140. Bleger, D.; Schwarz, J.; Brouwer, A. M.; Hecht, S., o-Fluoroazobenzenes as Readily Synthesized Photoswitches Offering Nearly Quantitative Two-Way Isomerization with Visible Light. *Journal of the American Chemical Society* **2012**, 134 (51), 20597-20600.

141. Hoppmann, C.; Maslennikov, I.; Choe, S.; Wang, L., In Situ Formation of an Azo Bridge on Proteins Controllable by Visible Light. *Journal of the American Chemical Society* **2015**, 137 (35), 11218-21.

142. Li, Y. M.; Yang, M. Y.; Huang, Y. C.; Li, Y. T.; Chen, P. R.; Liu, L., Ligation of expressed protein alpha-hydrazides via genetic incorporation of an alpha-hydroxy acid. *ACS chemical biology* **2012**, 7 (6), 1015-22.

143. Peters, F. B.; Brock, A.; Wang, J.; Schultz, P. G., Photocleavage of the polypeptide backbone by 2-nitrophenylalanine. *Chemistry & biology* **2009**, 16 (2), 148-52.

144. Raina, K.; Crews, C. M., Chemical inducers of targeted protein degradation. *The Journal of biological chemistry* **2010**, 285 (15), 11057-60.

145. Schneekloth, J. S., Jr.; Fonseca, F. N.; Koldobskiy, M.; Mandal, A.; Deshaies, R.; Sakamoto, K.; Crews, C. M., Chemical genetic control of protein levels: selective in vivo targeted degradation. *Journal of the American Chemical Society* **2004**, 126 (12), 3748-54.

146. Neklesa, T. K.; Tae, H. S.; Schneekloth, A. R.; Stulberg, M. J.; Corson, T. W.; Sundberg, T. B.; Raina, K.; Holley, S. A.; Crews, C. M., Small-molecule hydrophobic tagging-induced degradation of HaloTag fusion proteins. *Nature chemical biology* **2011**, 7 (8), 538-43.

147. (a) Bongers, K. M.; Rakhit, R.; Payumo, A. Y.; Chen, J. K.; Wandless, T. J., General method for regulating protein stability with light. *ACS chemical biology* **2014**, 9 (1), 111-5; (b) Delacour, Q.; Li, C.; Plamont, M. A.; Billon-Denis, E.; Aujard, I.; Le Saux, T.; Jullien, L.; Gautier, A., Light-Activated Proteolysis for the Spatiotemporal Control of Proteins. *ACS chemical biology* **2015**, 10 (7), 1643-7.

148. Gautier, A.; Deiters, A.; Chin, J. W., Light-activated kinases enable temporal dissection of signaling networks in living cells. *Journal of the American Chemical Society* **2011**, *133* (7), 2124-7.
149. (a) Bongers, K. M.; Chen, L. C.; Liu, C. W.; Wandless, T. J., Small-molecule displacement of a cryptic degron causes conditional protein degradation. *Nature chemical biology* **2011**, *7* (8), 531-7; (b) Touitou, R.; Richardson, J.; Bose, S.; Nakanishi, M.; Rivett, J.; Allday, M. J., A degradation signal located in the C-terminus of p21WAF1/CIP1 is a binding site for the C8 α -subunit of the 20S proteasome. *The EMBO journal* **2001**, *20* (10), 2367-75.
150. Dantuma, N. P.; Lindsten, K.; Glas, R.; Jellne, M.; Masucci, M. G., Short-lived green fluorescent proteins for quantifying ubiquitin/proteasome-dependent proteolysis in living cells. *Nature biotechnology* **2000**, *18* (5), 538-43.
151. Chung, C. H.; Baek, S. H., Deubiquitinating enzymes: their diversity and emerging roles. *Biochemical and biophysical research communications* **1999**, *266* (3), 633-40.
152. Tasaki, T.; Sriram, S. M.; Park, K. S.; Kwon, Y. T., The N-end rule pathway. *Annual review of biochemistry* **2012**, *81*, 261-89.
153. Menendez-Benito, V.; Heessen, S.; Dantuma, N. P., Monitoring of ubiquitin-dependent proteolysis with green fluorescent protein substrates. *Methods in enzymology* **2005**, *399*, 490-511.
154. Varshavsky, A., The N-end rule: functions, mysteries, uses. *Proc Natl Acad Sci U S A* **1996**, *93* (22), 12142-9.
155. Kirisako, T.; Kamei, K.; Murata, S.; Kato, M.; Fukumoto, H.; Kanie, M.; Sano, S.; Tokunaga, F.; Tanaka, K.; Iwai, K., A ubiquitin ligase complex assembles linear polyubiquitin chains. *The EMBO journal* **2006**, *25* (20), 4877-87.
156. (a) Buskirk, A. R.; Liu, D. R., Creating small-molecule-dependent switches to modulate biological functions. *Chemistry & biology* **2005**, *12* (2), 151-61; (b) Zorn, J. A.; Wells, J. A., Turning enzymes ON with small molecules. *Nature chemical biology* **2010**, *6* (3), 179-188.
157. (a) Yang, M.; Li, J.; Chen, P. R., Transition metal-mediated bioorthogonal protein chemistry in living cells. *Chem Soc Rev* **2014**, *43* (18), 6511-26; (b) Li, J.; Yu, J.; Zhao, J.; Wang, J.; Zheng, S.; Lin, S.; Chen, L.; Yang, M.; Jia, S.; Zhang, X.; Chen, P. R., Palladium-triggered deprotection chemistry for protein activation in living cells. *Nature chemistry* **2014**, *6* (4), 352-61.
158. Li, J.; Jia, S.; Chen, P. R., Diels-Alder reaction-triggered bioorthogonal protein decaging in living cells. *Nature chemical biology* **2014**, *10* (12), 1003-5.
159. Zhao, J.; Lin, S.; Huang, Y.; Zhao, J.; Chen, P. R., Mechanism-based design of a photoactivatable firefly luciferase. *Journal of the American Chemical Society* **2013**, *135* (20), 7410-3.
160. Luo, J.; Arbely, E.; Zhang, J.; Chou, C.; Uprety, R.; Chin, J. W.; Deiters, A., Genetically encoded optical activation of DNA recombination in human cells. *Chemical communications* **2016**, *52* (55), 8529-8532.
161. Yang, Y. S.; Hughes, T. E., Cre stoplight: a red/green fluorescent reporter of Cre recombinase expression in living cells. *BioTechniques* **2001**, *31* (5), 1036, 1038, 1040-1.
162. Engelke, H.; Chou, C.; Uprety, R.; Jess, P.; Deiters, A., Control of Protein Function through Optochemical Translocation. *Acs Synth Biol* **2014**, *3* (10), 731-736.
163. Song, F.; Garner, A. L.; Koide, K., A highly sensitive fluorescent sensor for palladium based on the allylic oxidative insertion mechanism. *Journal of the American Chemical Society* **2007**, *129* (41), 12354-5.
164. Chen, I.; Ting, A. Y., Site-specific labeling of proteins with small molecules in live cells. *Current opinion in biotechnology* **2005**, *16* (1), 35-40.
165. (a) Lang, K.; Chin, J. W., Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chemical reviews* **2014**, *114* (9), 4764-806; (b) Kim, C. H.; Axup, J. Y.; Schultz, P. G., Protein conjugation with genetically encoded unnatural amino acids. *Current opinion in chemical biology* **2013**, *17* (3), 412-9.

166. (a) Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G., Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100* (1), 56-61; (b) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G., An expanded eukaryotic genetic code. *Science* **2003**, *301* (5635), 964-7; (c) Zeng, H.; Xie, J.; Schultz, P. G., Genetic introduction of a diketone-containing amino acid into proteins. *Bioorganic & medicinal chemistry letters* **2006**, *16* (20), 5356-9; (d) Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Wang, Z.; Russell, D. H.; Liu, W., Genetic incorporation of an aliphatic keto-containing amino acid into proteins for their site-specific modifications. *Bioorganic & medicinal chemistry letters* **2010**, *20* (3), 878-80.
167. Tuley, A.; Lee, Y. J.; Wu, B.; Wang, Z. U.; Liu, W. R., A genetically encoded aldehyde for rapid protein labelling. *Chemical communications* **2014**, *50* (56), 7424-6.
168. Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G., Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*. *Journal of the American Chemical Society* **2003**, *125* (39), 11782-3.
169. (a) Deiters, A.; Schultz, P. G., In vivo incorporation of an alkyne into proteins in *Escherichia coli*. *Bioorganic & medicinal chemistry letters* **2005**, *15* (5), 1521-4; (b) Fekner, T.; Li, X.; Lee, M. M.; Chan, M. K., A pyrrolysine analogue for protein click chemistry. *Angewandte Chemie* **2009**, *48* (9), 1633-5; (c) Li, J.; Lin, S.; Wang, J.; Jia, S.; Yang, M.; Hao, Z.; Zhang, X.; Chen, P. R., Ligand-free palladium-mediated site-specific protein labeling inside gram-negative bacterial pathogens. *Journal of the American Chemical Society* **2013**, *135* (19), 7330-8; (d) Plass, T.; Milles, S.; Koehler, C.; Schultz, C.; Lemke, E. A., Genetically encoded copper-free click chemistry. *Angewandte Chemie* **2011**, *50* (17), 3878-81.
170. Yu, Z.; Pan, Y.; Wang, Z.; Wang, J.; Lin, Q., Genetically encoded cyclopropene directs rapid, photoclick-chemistry-mediated protein labeling in mammalian cells. *Angewandte Chemie* **2012**, *51* (42), 10600-4.
171. Brustad, E. M.; Lemke, E. A.; Schultz, P. G.; Deniz, A. A., A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer. *Journal of the American Chemical Society* **2008**, *130* (52), 17664-5.
172. Fleissner, M. R.; Brustad, E. M.; Kalai, T.; Altenbach, C.; Cascio, D.; Peters, F. B.; Hideg, K.; Peucker, S.; Schultz, P. G.; Hubbell, W. L., Site-directed spin labeling of a genetically encoded unnatural amino acid. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106* (51), 21637-42.
173. Ye, S.; Kohrer, C.; Huber, T.; Kazmi, M.; Sachdev, P.; Yan, E. C.; Bhagat, A.; RajBhandary, U. L.; Sakmar, T. P., Site-specific incorporation of keto amino acids into functional G protein-coupled receptors using unnatural amino acid mutagenesis. *The Journal of biological chemistry* **2008**, *283* (3), 1525-33.
174. Kularatne, S. A.; Deshmukh, V.; Ma, J.; Tardif, V.; Lim, R. K.; Pugh, H. M.; Sun, Y.; Manibusan, A.; Sellers, A. J.; Barnett, R. S.; Srinagesh, S.; Forsyth, J. S.; Hassenpflug, W.; Tian, F.; Javahishvili, T.; Felding-Habermann, B.; Lawson, B. R.; Kazane, S. A.; Schultz, P. G., A CXCR4-targeted site-specific antibody-drug conjugate. *Angewandte Chemie* **2014**, *53* (44), 11863-7.
175. Liu, C. C.; Schultz, P. G., Adding new chemistries to the genetic code. *Annual review of biochemistry* **2010**, *79*, 413-44.
176. Yi, L.; Sun, H.; Wu, Y. W.; Triola, G.; Waldmann, H.; Goody, R. S., A highly efficient strategy for modification of proteins at the C terminus. *Angewandte Chemie* **2010**, *49* (49), 9417-21.
177. Kim, J. W.; Cochran, F. V.; Cochran, J. R., A chemically cross-linked knottin dimer binds integrins with picomolar affinity and inhibits tumor cell migration and proliferation. *Journal of the American Chemical Society* **2015**, *137* (1), 6-9.
178. (a) Dirksen, A.; Hackeng, T. M.; Dawson, P. E., Nucleophilic catalysis of oxime ligation. *Angewandte Chemie* **2006**, *45* (45), 7581-4; (b) Wendeler, M.; Grinberg, L.; Wang, X.; Dawson,

- P. E.; Baca, M., Enhanced catalysis of oxime-based bioconjugations by substituted anilines. *Bioconjugate chemistry* **2014**, 25 (1), 93-101.
179. Stanley, M.; Virdee, S., Genetically Directed Production of Recombinant, Isosteric and Nonhydrolysable Ubiquitin Conjugates. *ChemBiochem* **2016**, 17 (15), 1472-1480.
180. Nguyen, D. P.; Garcia Alai, M. M.; Kapadnis, P. B.; Neumann, H.; Chin, J. W., Genetically encoding N(epsilon)-methyl-L-lysine in recombinant histones. *Journal of the American Chemical Society* **2009**, 131 (40), 14194-5.
181. Kim, C. H.; Kang, M.; Kim, H. J.; Chatterjee, A.; Schultz, P. G., Site-specific incorporation of epsilon-N-crotonyllysine into histones. *Angewandte Chemie* **2012**, 51 (29), 7246-9.
182. Yanagisawa, T.; Takahashi, M.; Mukai, T.; Sato, S.; Wakamori, M.; Shirouzu, M.; Sakamoto, K.; Umehara, T.; Yokoyama, S., Multiple site-specific installations of Nepsilon-monomethyl-L-lysine into histone proteins by cell-based and cell-free protein synthesis. *ChemBiochem : a European journal of chemical biology* **2014**, 15 (12), 1830-8.
183. Bure, C.; Lelievre, D.; Delmas, A., Identification of by-products from an orthogonal peptide ligation by oxime bonds using mass spectrometry and tandem mass spectrometry. *Rapid communications in mass spectrometry : RCM* **2000**, 14 (23), 2158-64.
184. Roberts, M. J.; Bentley, M. D.; Harris, J. M., Chemistry for peptide and protein PEGylation. *Advanced drug delivery reviews* **2002**, 54 (4), 459-76.
185. (a) Lohse, A.; Martins, R.; Jorgensen, M. R.; Hindsgaul, O., Solid-phase oligosaccharide tagging (SPOT): Validation on glycolipid-derived structures. *Angewandte Chemie* **2006**, 45 (25), 4167-72; (b) Pilobello, K. T.; Mahal, L. K., Deciphering the glycode: the complexity and analytical challenge of glycomics. *Current opinion in chemical biology* **2007**, 11 (3), 300-5.
186. Tian, H.; Naganathan, S.; Kazmi, M. A.; Schwartz, T. W.; Sakmar, T. P.; Huber, T., Bioorthogonal fluorescent labeling of functional G-protein-coupled receptors. *ChemBiochem : a European journal of chemical biology* **2014**, 15 (12), 1820-9.
187. Axup, J. Y.; Bajjuri, K. M.; Ritland, M.; Hutchins, B. M.; Kim, C. H.; Kazane, S. A.; Halder, R.; Forsyth, J. S.; Santidrian, A. F.; Stafin, K.; Lu, Y. C.; Tran, H.; Seller, A. J.; Biroce, S. L.; Szydlak, A.; Pinkstaff, J. K.; Tian, F.; Sinha, S. C.; Felding-Habermann, B.; Smider, V. V.; Schultz, P. G., Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. *P Natl Acad Sci USA* **2012**, 109 (40), 16101-16106.
188. Kazane, S. A.; Axup, J. Y.; Kim, C. H.; Ciobanu, M.; Wold, E. D.; Barluenga, S.; Hutchins, B. A.; Schultz, P. G.; Winssinger, N.; Smider, V. V., Self-Assembled Antibody Multimers through Peptide Nucleic Acid Conjugation. *Journal of the American Chemical Society* **2013**, 135 (1), 340-346.
189. Gau, D.; Veon, W.; Zeng, X. M.; Yates, N.; Shroff, S. G.; Koes, D. R.; Roy, P., Threonine 89 Is an Important Residue of Profilin-1 That Is Phosphorylatable by Protein Kinase A. *PloS one* **2016**, 11 (5).
190. (a) Chen, Y.; Ma, J.; Lu, W.; Tian, M.; Thauvin, M.; Yuan, C.; Volovitch, M.; Wang, Q.; Holst, J.; Liu, M.; Vriza, S.; Ye, S.; Wang, L.; Li, D., Heritable expansion of the genetic code in mouse and zebrafish. *Cell research* **2016**; (b) Parrish, A. R.; She, X.; Xiang, Z.; Coin, I.; Shen, Z.; Briggs, S. P.; Dillin, A.; Wang, L., Expanding the genetic code of *Caenorhabditis elegans* using bacterial aminoacyl-tRNA synthetase/tRNA pairs. *ACS chemical biology* **2012**, 7 (7), 1292-302.
191. Grunwald, D. J.; Eisen, J. S., Headwaters of the zebrafish -- emergence of a new model vertebrate. *Nature reviews. Genetics* **2002**, 3 (9), 717-24.
192. Lieschke, G. J.; Currie, P. D., Animal models of human disease: zebrafish swim into view. *Nature reviews. Genetics* **2007**, 8 (5), 353-67.
193. Zon, L. I.; Peterson, R. T., In vivo drug discovery in the zebrafish. *Nature reviews. Drug discovery* **2005**, 4 (1), 35-44.
194. (a) Hemphill, J.; Liu, Q.; Uprety, R.; Samanta, S.; Tsang, M.; Juliano, R. L.; Deiters, A., Conditional control of alternative splicing through light-triggered splice-switching oligonucleotides. *Journal of the American Chemical Society* **2015**, 137 (10), 3656-62; (b) Shestopalov, I. A.; Sinha,

- S.; Chen, J. K., Light-controlled gene silencing in zebrafish embryos. *Nature chemical biology* **2007**, 3 (10), 650-1; (c) Tang, X.; Maegawa, S.; Weinberg, E. S.; Dmochowski, I. J., Regulating gene expression in zebrafish embryos using light-activated, negatively charged peptide nucleic acids. *Journal of the American Chemical Society* **2007**, 129 (36), 11000-1; (d) Wang, Y.; Wu, L.; Wang, P.; Lv, C.; Yang, Z. J.; Tang, X. J., Manipulation of gene expression in zebrafish using caged circular morpholino oligomers. *Nucleic Acids Res* **2012**, 40 (21), 11155-11162; (e) Yamazoe, S.; Liu, Q.; McQuade, L. E.; Deiters, A.; Chen, J. K., Sequential gene silencing using wavelength-selective caged morpholino oligonucleotides. *Angewandte Chemie* **2014**, 53 (38), 10114-8.
195. Cheung, C. Y.; Webb, S. E.; Meng, A.; Miller, A. L., Transient expression of apoaquorin in zebrafish embryos: extending the ability to image calcium transients during later stages of development. *The International journal of developmental biology* **2006**, 50 (6), 561-9.
196. (a) Amatruda, J. F.; Shepard, J. L.; Stern, H. M.; Zon, L. I., Zebrafish as a cancer model system. *Cancer cell* **2002**, 1 (3), 229-31; (b) Hill, A. J.; Teraoka, H.; Heideman, W.; Peterson, R. E., Zebrafish as a model vertebrate for investigating chemical toxicity. *Toxicological sciences : an official journal of the Society of Toxicology* **2005**, 86 (1), 6-19; (c) Norton, W.; Bally-Cuif, L., Adult zebrafish as a model organism for behavioural genetics. *BMC neuroscience* **2010**, 11, 90.
197. (a) Portugues, R.; Severi, K. E.; Wyart, C.; Ahrens, M. B., Optogenetics in a transparent animal: circuit function in the larval zebrafish. *Current opinion in neurobiology* **2013**, 23 (1), 119-26; (b) Reade, A.; Motta-Mena, L. B.; Gardner, K. H.; Stainier, D. Y.; Weiner, O. D.; Woo, S., TAE: a zebrafish-optimized optogenetic gene expression system with fine spatial and temporal control. *Development* **2017**, 144 (2), 345-355; (c) Yoo, S. K.; Deng, Q.; Cavnar, P. J.; Wu, Y. I.; Hahn, K. M.; Huttenlocher, A., Differential regulation of protrusion and polarity by PI3K during neutrophil motility in live zebrafish. *Developmental cell* **2010**, 18 (2), 226-36; (d) Arrenberg, A. B.; Stainier, D. Y.; Baier, H.; Huisken, J., Optogenetic control of cardiac function. *Science* **2010**, 330 (6006), 971-4.
198. Hemphill, J.; Borchardt, E. K.; Brown, K.; Asokan, A.; Deiters, A., Optical Control of CRISPR/Cas9 Gene Editing. *Journal of the American Chemical Society* **2015**, 137 (17), 5642-5.
199. Xiong, Y.; Steitz, T. A., A story with a good ending: tRNA 3'-end maturation by CCA-adding enzymes. *Current opinion in structural biology* **2006**, 16 (1), 12-7.
200. Schmied, W. H.; Elsasser, S. J.; Uttamapinant, C.; Chin, J. W., Efficient Multisite Unnatural Amino Acid Incorporation in Mammalian Cells via Optimized Pyrrolysyl tRNA Synthetase/tRNA Expression and Engineered eRF1. *Journal of the American Chemical Society* **2014**, 136 (44), 15577-15583.
201. Krens, S. F.; He, S.; Spaink, H. P.; Snaar-Jagalska, B. E., Characterization and expression patterns of the MAPK family in zebrafish. *Gene expression patterns : GEP* **2006**, 6 (8), 1019-26.
202. Furthauer, M.; Van Celst, J.; Thisse, C.; Thisse, B., Fgf signalling controls the dorsoventral patterning of the zebrafish embryo. *Development* **2004**, 131 (12), 2853-64.
203. Anastasaki, C.; Estep, A. L.; Marais, R.; Rauen, K. A.; Patton, E. E., Kinase-activating and kinase-impaired cardio-facio-cutaneous syndrome alleles have activity during zebrafish development and are sensitive to small molecule inhibitors. *Human molecular genetics* **2009**, 18 (14), 2543-54.
204. Mizoguchi, T.; Izawa, T.; Kuroiwa, A.; Kikuchi, Y., Fgf signaling negatively regulates Nodal-dependent endoderm induction in zebrafish. *Developmental biology* **2006**, 300 (2), 612-22.
205. Xiong, C.; Liu, X.; Meng, A., The Kinase Activity-deficient Isoform of the Protein Araf Antagonizes Ras/Mitogen-activated Protein Kinase (Ras/MAPK) Signaling in the Zebrafish Embryo. *The Journal of biological chemistry* **2015**, 290 (42), 25512-21.
206. Anastasaki, C.; Rauen, K. A.; Patton, E. E., Continual low-level MEK inhibition ameliorates cardio-facio-cutaneous phenotypes in zebrafish. *Disease models & mechanisms* **2012**, 5 (4), 546-52.

207. Pirman, N. L.; Barber, K. W.; Aerni, H. R.; Ma, N. J.; Haimovich, A. D.; Rogulina, S.; Isaacs, F. J.; Rinehart, J., A flexible codon in genomically recoded *Escherichia coli* permits programmable protein phosphorylation. *Nature communications* **2015**, 6, 8130.
208. Chatterjee, A.; Sun, S. B.; Furman, J. L.; Xiao, H.; Schultz, P. G., A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli*. *Biochemistry* **2013**, 52 (10), 1828-37.
209. Alcaraz-Perez, F.; Mulero, V.; Cayuela, M. L., Application of the dual-luciferase reporter assay to the analysis of promoter activity in Zebrafish embryos. *BMC biotechnology* **2008**, 8, 81.
210. Thisse, C.; Thisse, B., High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nature protocols* **2008**, 3 (1), 59-69.
211. (a) Heldin, C. H.; Miyazono, K.; ten Dijke, P., TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature* **1997**, 390 (6659), 465-71; (b) Gordon, K. J.; Blobel, G. C., Role of transforming growth factor-beta superfamily signaling pathways in human disease. *Biochimica et biophysica acta* **2008**, 1782 (4), 197-228; (c) Zhang, Y. E., Non-Smad pathways in TGF-beta signaling. *Cell Res* **2009**, 19 (1), 128-39.
212. Muller, P.; Rogers, K. W.; Jordan, B. M.; Lee, J. S.; Robson, D.; Ramanathan, S.; Schier, A. F., Differential Diffusivity of Nodal and Lefty Underlies a Reaction-Diffusion Patterning System. *Science* **2012**, 336 (6082), 721-724.
213. Speder, P.; Petzoldt, A.; Suzanne, M.; Noselli, S., Strategies to establish left/right asymmetry in vertebrates and invertebrates. *Curr Opin Genet Dev* **2007**, 17 (4), 351-358.
214. Wieser, R.; Wrana, J. L.; Massague, J., Gs Domain Mutations That Constitutively Activate T-Beta-R-I, the Downstream Signaling Component in the Tgf-Beta Receptor Complex. *Embo Journal* **1995**, 14 (10), 2199-2208.
215. (a) Attisano, L.; Wrana, J. L.; Montalvo, E.; Massague, J., Activation of signalling by the activin receptor complex. *Molecular and cellular biology* **1996**, 16 (3), 1066-1073; (b) Goldberg, F. W.; Ward, R. A.; Powell, S. J.; Debreczeni, J. E.; Norman, R. A.; Roberts, N. J.; Dishington, A. P.; Gingell, H. J.; Wickson, K. F.; Roberts, A. L., Rapid Generation of a High Quality Lead for Transforming Growth Factor-beta (TGF-beta) Type I Receptor (ALK5). *J Med Chem* **2009**, 52 (23), 7901-7905.
216. Wieser, R.; Wrana, J. L.; Massague, J., GS domain mutations that constitutively activate T beta R-I, the downstream signaling component in the TGF-beta receptor complex. *The EMBO journal* **1995**, 14 (10), 2199-208.
217. Yu, J. S.; Ramasamy, T. S.; Murphy, N.; Holt, M. K.; Czapiewski, R.; Wei, S. K.; Cui, W., PI3K/mTORC2 regulates TGF-beta/Activin signalling by modulating Smad2/3 activity via linker phosphorylation. *Nature communications* **2015**, 6, 7212.
218. Nicolas, F. J.; De Bosscher, K.; Schmierer, B.; Hill, C. S., Analysis of Smad nucleocytoplasmic shuttling in living cells. *Journal of cell science* **2004**, 117 (Pt 18), 4113-25.
219. (a) Tojo, M.; Hamashima, Y.; Hanyu, A.; Kajimoto, T.; Saitoh, M.; Miyazono, K.; Node, M.; Imamura, T., The ALK-5 inhibitor A-83-01 inhibits Smad signaling and epithelial-to-mesenchymal transition by transforming growth factor-beta. *Cancer science* **2005**, 96 (11), 791-800; (b) Daly, A. C.; Randall, R. A.; Hill, C. S., Transforming growth factor beta-induced Smad1/5 phosphorylation in epithelial cells is mediated by novel receptor complexes and is essential for anchorage-independent growth. *Molecular and cellular biology* **2008**, 28 (22), 6889-902; (c) Inman, G. J.; Nicolas, F. J.; Hill, C. S., Nucleocytoplasmic shuttling of Smads 2, 3, and 4 permits sensing of TGF-beta receptor activity. *Molecular cell* **2002**, 10 (2), 283-94.
220. Guo, X.; Wang, X. F., Signaling cross-talk between TGF-beta/BMP and other pathways. *Cell Res* **2009**, 19 (1), 71-88.
221. Regot, S.; Hughey, J. J.; Bajar, B. T.; Carrasco, S.; Covert, M. W., High-sensitivity measurements of multiple kinase activities in live single cells. *Cell* **2014**, 157 (7), 1724-34.

222. Simeone, D. M.; Zhang, L.; Graziano, K.; Nicke, B.; Pham, T.; Schaefer, C.; Logsdon, C. D., Smad4 mediates activation of mitogen-activated protein kinases by TGF-beta in pancreatic acinar cells. *American journal of physiology. Cell physiology* **2001**, *281* (1), C311-9.
223. Olsson, N.; Piek, E.; Sundstrom, M.; ten Dijke, P.; Nilsson, G., Transforming growth factor-beta-mediated mast cell migration depends on mitogen-activated protein kinase activity. *Cellular signalling* **2001**, *13* (7), 483-90.
224. Toyama, R.; O'Connell, M. L.; Wright, C. V.; Kuehn, M. R.; Dawid, I. B., Nodal induces ectopic goosecoid and lim1 expression and axis duplication in zebrafish. *Development* **1995**, *121* (2), 383-91.
225. Schultemerker, S.; Hammerschmidt, M.; Beuchle, D.; Cho, K. W.; Derobertis, E. M.; Nussleinvohard, C., Expression of Zebrafish Goosecoid and No Tail Gene-Products in Wild-Type and Mutant No Tail Embryos. *Development* **1994**, *120* (4), 843-852.
226. Chang, L. L.; Kessler, D. S., Foxd3 is an essential Nodal-dependent regulator of zebrafish dorsal mesoderm development. *Dev Biol* **2010**, *342* (1), 39-50.
227. Schultemerker, S.; Ho, R. K.; Herrmann, B. G.; Nussleinvohard, C., The Protein Product of the Zebrafish Homolog of the Mouse T-Gene Is Expressed in Nuclei of the Germ Ring and the Notochord of the Early Embryo. *Development* **1992**, *116* (4), 1021-&.
228. Schniewind, B.; Groth, S.; Muerkoster, S. S.; Sipos, B.; Schaefer, H.; Kalthoff, H.; Fandrich, F.; Ungefroren, H., Dissecting the role of TGF-beta type I receptor/ALK5 in pancreatic ductal adenocarcinoma: Smad activation is crucial for both the tumor suppressive and prometastatic function. *Oncogene* **2007**, *26* (33), 4850-4862.
229. Laux, D. W.; Febbo, J. A.; Roman, B. L., Dynamic analysis of BMP-responsive smad activity in live zebrafish embryos. *Developmental dynamics : an official publication of the American Association of Anatomists* **2011**, *240* (3), 682-94.
230. Mosimann, C.; Kaufman, C. K.; Li, P.; Pugach, E. K.; Tamplin, O. J.; Zon, L. I., Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. *Development* **2011**, *138* (1), 169-77.
231. Hans, S.; Kaslin, J.; Freudenreich, D.; Brand, M., Temporally-controlled site-specific recombination in zebrafish. *PloS one* **2009**, *4* (2), e4640.
232. Sinha, D. K.; Neveu, P.; Gagey, N.; Aujard, I.; Le Saux, T.; Rampon, C.; Gauron, C.; Kawakami, K.; Leucht, C.; Bally-Cuif, L.; Volovitch, M.; Bensimon, D.; Jullien, L.; Vriz, S., Photoactivation of the CreER(T2) Recombinase for Conditional Site-Specific Recombination with High Spatiotemporal Resolution. *Zebrafish* **2010**, *7* (2), 199-204.
233. Kawano, F.; Okazaki, R.; Yazawa, M.; Sato, M., A photoactivatable Cre-loxP recombination system for optogenetic genome engineering. *Nature chemical biology* **2016**, *12* (12), 1059-1064.
234. Burns, C. G.; Milan, D. J.; Grande, E. J.; Rottbauer, W.; MacRae, C. A.; Fishman, M. C., High-throughput assay for small molecules that modulate zebrafish embryonic heart rate. *Nature chemical biology* **2005**, *1* (5), 263-4.
235. Wiedenheft, B.; Sternberg, S. H.; Doudna, J. A., RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **2012**, *482* (7385), 331-8.
236. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012**, *337* (6096), 816-21.
237. Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M., RNA-guided human genome engineering via Cas9. *Science* **2013**, *339* (6121), 823-6.
238. (a) Wang, H.; Yang, H.; Shivalila, C. S.; Dawlaty, M. M.; Cheng, A. W.; Zhang, F.; Jaenisch, R., One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **2013**, *153* (4), 910-8; (b) Hwang, W. Y.; Fu, Y. F.; Reyon, D.; Maeder, M. L.; Tsai, S. Q.; Sander, J. D.; Peterson, R. T.; Yeh, J. R. J.; Joung, J. K., Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* **2013**, *31* (3), 227-229.

239. Nihongaki, Y.; Kawano, F.; Nakajima, T.; Sato, M., Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nature biotechnology* **2015**, 33 (7), 755-760.
240. Hemphill, J.; Borchardt, E. K.; Brown, K.; Asokan, A.; Deiters, A., Optical Control of CRISPR/Cas9 Gene Editing. *Journal of the American Chemical Society* **2015**, 137 (17), 5642-5645.
241. Gorden, N. T.; Arts, H. H.; Parisi, M. A.; Coene, K. L. M.; Letteboer, S. J. F.; van Beersum, S. E. C.; Mans, D. A.; Hikida, A.; Eckert, M.; Knutzen, D.; Alswaid, A. F.; Ozyurek, H.; Dibooglu, S.; Otto, E. A.; Liu, Y. F.; Davis, E. E.; Hutter, C. M.; Bammler, T. K.; Farin, F. M.; Dorschner, M.; Topcu, M.; Zackai, E. H.; Rosenthal, P.; Owens, K. N.; Katsanis, N.; Vincent, J. B.; Hildebrandt, F.; Rubel, E. W.; Raible, D. W.; Knoers, N. V. A. M.; Chance, P. F.; Roepman, R.; Moens, C. B.; Glass, I. A.; Doherty, D., CC2D2A Is Mutated in Joubert Syndrome and Interacts with the Ciliopathy-Associated Basal Body Protein CEP290. *Am J Hum Genet* **2008**, 83 (5), 559-571.
242. Kim, S.; Kim, D.; Cho, S. W.; Kim, J.; Kim, J. S., Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome research* **2014**, 24 (6), 1012-9.
243. Gagon, J. A.; Valen, E.; Thyme, S. B.; Huang, P.; Ahkmetova, L., Efficient Mutagenesis by Cas9 Protein-Mediated Oligonucleotide Insertion and Large-Scale Assessment of Single-Guide RNAs (vol 9, e98186, 2014). *PloS one* **2014**, 9 (8).
244. Gagnon, J. A.; Valen, E.; Thyme, S. B.; Huang, P.; Ahkmetova, L.; Pauli, A.; Montague, T. G.; Zimmerman, S.; Richter, C.; Schier, A. F., Efficient Mutagenesis by Cas9 Protein-Mediated Oligonucleotide Insertion and Large-Scale Assessment of Single-Guide RNAs. *PloS one* **2014**, 9 (5).
245. Kim, J. H.; Lee, S. R.; Li, L. H.; Park, H. J.; Park, J. H.; Lee, K. Y.; Kim, M. K.; Shin, B. A.; Choi, S. Y., High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PloS one* **2011**, 6 (4), e18556.
246. Molina, G. A.; Watkins, S. C.; Tsang, M., Generation of FGF reporter transgenic zebrafish and their utility in chemical screens. *BMC developmental biology* **2007**, 7, 62.
247. Chen, Y.; Ma, J.; Lu, W.; Tian, M.; Thauvin, M.; Yuan, C.; Volovitch, M.; Wang, Q.; Holst, J.; Liu, M.; Vriz, S.; Ye, S.; Wang, L.; Li, D., Heritable expansion of the genetic code in mouse and zebrafish. *Cell Res* **2017**, 27 (2), 294-297.
248. Ablain, J.; Durand, E. M.; Yang, S.; Zhou, Y.; Zon, L. I., A CRISPR/Cas9 vector system for tissue-specific gene disruption in zebrafish. *Developmental cell* **2015**, 32 (6), 756-64.
249. Yanagawa, M.; Yamashita, T.; Shichida, Y., Comparative fluorescence resonance energy transfer analysis of metabotropic glutamate receptors: implications about the dimeric arrangement and rearrangement upon ligand bindings. *The Journal of biological chemistry* **2011**, 286 (26), 22971-81.
250. Nasevicius, A.; Ekker, S. C., Effective targeted gene 'knockdown' in zebrafish. *Nature genetics* **2000**, 26 (2), 216-220.
251. Thisse, B.; Heyer, V.; Lux, A.; Alunni, V.; Degrave, A.; Seiliez, I.; Kirchner, J.; Parkhill, J. P.; Thisse, C., Spatial and temporal expression of the zebrafish genome by large-scale in situ hybridization screening. *Method Cell Biol* **2004**, 77, 505-519.
252. (a) Shestopalov, I. A.; Sinha, S.; Chen, J. K., Light-controlled gene silencing in zebrafish embryos. *Nature chemical biology* **2007**, 3 (10), 650-651; (b) Deiters, A.; Garner, R. A.; Lusic, H.; Govan, J. M.; Dush, M.; Nascone-Yoder, N. M.; Yoder, J. A., Photocaged Morpholino Oligomers for the Light-Regulation of Gene Function in Zebrafish and Xenopus Embryos. *Journal of the American Chemical Society* **2010**, 132 (44), 15644-15650; (c) Tallafuss, A.; Gibson, D.; Morcos, P.; Li, Y. F.; Seredick, S.; Eisen, J.; Washbourne, P., Turning gene function ON and OFF using sense and antisense photo-morpholinos in zebrafish. *Development* **2012**, 139 (9), 1691-1699.
253. (a) Yamazoe, S.; Shestopalov, I. A.; Provost, E.; Leach, S. D.; Chen, J. K., Cyclic Caged Morpholinos: Conformationally Gated Probes of Embryonic Gene Function. *Angew Chem Int Edit* **2012**, 51 (28), 6908-6911; (b) Yamazoe, S.; Liu, Q. Y.; McQuade, L. E.; Deiters, A.; Chen, J. K.,

Sequential Gene Silencing Using Wavelength-Selective Caged Morpholino Oligonucleotides. *Angew Chem Int Edit* **2014**, 53 (38), 10114-10118.

254. (a) Kloosterman, W. P.; Lagendijk, A. K.; Ketting, R. F.; Moulton, J. D.; Plasterk, R. H., Targeted inhibition of miRNA maturation with morpholinos reveals a role for miR-375 in pancreatic islet development. *PLoS biology* **2007**, 5 (8), e203; (b) Choi, W. Y.; Giraldez, A. J.; Schier, A. F., Target protectors reveal dampening and balancing of Nodal agonist and antagonist by miR-430. *Science* **2007**, 318 (5848), 271-4.

255. Griffiths-Jones, S.; Grocock, R. J.; van Dongen, S.; Bateman, A.; Enright, A. J., miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **2006**, 34 (Database issue), D140-4.

256. Le, M. T.; Teh, C.; Shyh-Chang, N.; Xie, H.; Zhou, B.; Korzh, V.; Lodish, H. F.; Lim, B., MicroRNA-125b is a novel negative regulator of p53. *Genes & development* **2009**, 23 (7), 862-76.

257. Ketley, A.; Warren, A.; Holmes, E.; Gering, M.; Aboobaker, A. A.; Brook, J. D., The miR-30 microRNA family targets smoothened to regulate hedgehog signalling in zebrafish early muscle development. *PLoS one* **2013**, 8 (6), e65170.

258. Paulasova, P.; Pellestor, F., The peptide nucleic acids (PNAs): a new generation of probes for genetic and cytogenetic analyses. *Annales de genetique* **2004**, 47 (4), 349-58.

259. Ray, A.; Norden, B., Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **2000**, 14 (9), 1041-60.

260. Dean, D. A., Peptide nucleic acids: versatile tools for gene therapy strategies. *Advanced drug delivery reviews* **2000**, 44 (2-3), 81-95.

261. (a) Urtishak, K. A.; Choob, M.; Tian, X.; Sternheim, N.; Talbot, W. S.; Wickstrom, E.; Farber, S. A., Targeted gene knockdown in zebrafish using negatively charged peptide nucleic acid mimics. *Developmental dynamics : an official publication of the American Association of Anatomists* **2003**, 228 (3), 405-13; (b) Wickstrom, E.; Urtishak, K. A.; Choob, M.; Tian, X.; Sternheim, N.; Cross, L. M.; Rubinstein, A.; Farber, S. A., Downregulation of gene expression with negatively charged peptide nucleic acids (PNAs) in zebrafish embryos. *Methods in cell biology* **2004**, 77, 137-58.

262. Hemphill, J.; Govan, J.; Uprety, R.; Tsang, M.; Deiters, A., Site-specific promoter caging enables optochemical gene activation in cells and animals. *Journal of the American Chemical Society* **2014**, 136 (19), 7152-8.

263. Westerfield, M.; Wegner, J.; Jegalian, B. G.; DeRobertis, E. M.; Puschel, A. W., Specific activation of mammalian Hox promoters in mosaic transgenic zebrafish. *Genes & development* **1992**, 6 (4), 591-8.

264. Tallafuss, A.; Gibson, D.; Morcos, P.; Li, Y.; Seredick, S.; Eisen, J.; Washbourne, P., Turning gene function ON and OFF using sense and antisense photo-morpholinos in zebrafish. *Development* **2012**, 139 (9), 1691-9.

265. Canady, T. D.; Telmer, C. A.; Oyaghire, S. N.; Armitage, B. A.; Bruchez, M. P., In Vitro Reversible Translation Control Using gammaPNA Probes. *Journal of the American Chemical Society* **2015**, 137 (32), 10268-75.

266. Doyle, D. F.; Braasch, D. A.; Simmons, C. G.; Janowski, B. A.; Corey, D. R., Inhibition of gene expression inside cells by peptide nucleic acids: effect of mRNA target sequence, mismatched bases, and PNA length. *Biochemistry* **2001**, 40 (1), 53-64.

267. Young, D. D.; Lively, M. O.; Deiters, A., Activation and deactivation of DNAzyme and antisense function with light for the photochemical regulation of gene expression in mammalian cells. *Journal of the American Chemical Society* **2010**, 132 (17), 6183-93.

268. Grzegorski, S. J.; Chiari, E. F.; Robbins, A.; Kish, P. E.; Kahana, A., Natural variability of Kozak sequences correlates with function in a zebrafish model. *PLoS one* **2014**, 9 (9), e108475.