

**OPTICAL AND SMALL-MOLECULE CONTROL OF PROTEIN FUNCTION  
THROUGH GENETIC CODE EXPANSION**

by

**Ji Luo**

B.S. Pharmaceutical Sciences, Tianjin University, 2008

M.S. Chemistry, Nanyang Technological University, 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

2019

UNIVERSITY OF PITTSBURGH  
THE DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Ji Luo

It was defended on

September 25, 2017

and approved by

W. Seth Childers, PhD, Assistant Professor, Department of Chemistry

Kabirul Islam, PhD, Assistant Professor, Department of Chemistry

Bennett Van Houten, PhD, Professor, Department of Pharmacology and Chemical Biology

Dissertation Advisor: Alexander Deiters, PhD, Professor, Department of Chemistry

Copyright © by Ji Luo

2019

**OPTICAL AND SMALL-MOLECULE CONTROL OF PROTEIN FUNCTION  
THROUGH GENETIC CODE EXPANSION**

Ji Luo, PhD

University of Pittsburgh, 2019

Expanding the genetic code for site-specific incorporation of various biophysical probes and labels, such as fluorescent probes, photolabile caging groups, optical probes, photoswitches, bioorthogonal chemselective groups, offers a sophisticated tool for exploring protein structure and function, dissecting the cellular processes, and developing proteins with novel properties. Site-specific installation of distinct functional groups on a protein of interest is achieved using orthogonal aminoacyl-tRNA synthetase/tRNA pairs in response to an amber stop codon (UAG) placed in the gene of interest. In this dissertation, genetic encoding of a series of diverse unnatural amino acids (UAAs), including photocaged lysines, coumarin lysines, phthalimide lysines, photocaged cysteines, photocaged tyrosines, photoisomerizable phenylalanines, and azido lysines, were achieved through the discovery of the engineered pyrrolysyl-tRNA synthetase/tRNA pair from bacteria to eukaryotic cells. Strategic placement of these functional groups renders the protein inactive until deprotection through a bioorthogonal external trigger delivers the active wild-type protein. This developed methodology enables the conditional control of various cellular processes with spatiotemporal precision in a non-invasive way, including enzymatic activation, protein folding, protein translocation, gene repair (UvrD), post-translational modification (SUMO1), DNA recombination (Cre), RNA processing (Csy4), and gene editing (CRISPR/Cas9).

## TABLE OF CONTENTS

<b>1.0</b>	<b>INTRODUCTION.....</b>	<b>1</b>
<b>1.1</b>	<b>UNNATURAL AMINO ACID INCORPORATION .....</b>	<b>1</b>
<b>1.1.1</b>	<b>Incorporation of unnatural amino acids in <i>E. coli</i> .....</b>	<b>4</b>
<b>1.1.2</b>	<b>Incorporation of unnatural amino acids in mammalian cells .....</b>	<b>4</b>
<b>1.2</b>	<b>OPTICAL CONTROL OF PROTEIN FUNCTION.....</b>	<b>6</b>
<b>1.2.1</b>	<b>Optogenetic approaches .....</b>	<b>8</b>
<b>1.2.2</b>	<b>Photocaged amino acids .....</b>	<b>16</b>
<b>1.2.3</b>	<b>Project overview .....</b>	<b>19</b>
<b>2.0</b>	<b>GENETIC ENCODING OF NITROBENZYL-CAGED LYSINES .....</b>	<b>21</b>
<b>2.1</b>	<b>INTRODUCTION TO CAGING GROUPS .....</b>	<b>21</b>
<b>2.2</b>	<b>GENETIC ENCODING OF 2-(<i>O</i>-NITROPHENYL) PROPANOL AND 3-NITRO-2-ETHYLDIBENZOFURAN LYSINE .....</b>	<b>25</b>
<b>2.2.1</b>	<b>Incorporation of 2-(<i>o</i>-nitrophenyl) propanol lysine (NPPK).....</b>	<b>25</b>
<b>2.2.2</b>	<b>Incorporation of 3-nitro-2-ethyldibenzofuran lysine (NDBFK).....</b>	<b>34</b>
<b>2.3</b>	<b>OPTICAL CONTROL OF CRE RECOMBINASE.....</b>	<b>43</b>
<b>2.3.1</b>	<b>Introduction to DNA recombination.....</b>	<b>43</b>
<b>2.3.2</b>	<b>Optical control of DNA recombination .....</b>	<b>45</b>
<b>2.4</b>	<b>CONCLUSIONS .....</b>	<b>53</b>

<b>2.5</b>	<b>EXPERIMENTAL.....</b>	<b>55</b>
<b>3.0</b>	<b>COUMARIN LYSINES AS OPTICAL PROBES .....</b>	<b>60</b>
<b>3.1</b>	<b>INTRODUCTION TO COUMARIN LYSINES.....</b>	<b>60</b>
<b>3.2</b>	<b>INCORPORATION OF COUMARIN LYSINES IN BACTERIAL AND MAMMALIAN CELLS.....</b>	<b>61</b>
<b>3.3</b>	<b>OPTICAL CONTROL OF PROTEIN FUNCTION.....</b>	<b>73</b>
	<b>    3.3.1    Optical control of firefly luciferase .....</b>	<b>73</b>
	<b>    3.3.2    Optical control of EGFP folding and formation.....</b>	<b>76</b>
	<b>    3.3.3    Optical control of SATB1 nuclear localization .....</b>	<b>82</b>
<b>3.4</b>	<b>OPTICAL CONTROL OF HELICASE UVRD FOR DNA REPAIR.....</b>	<b>84</b>
	<b>    3.4.1    Incorporation of HCK into UvrD in <i>E. coli</i> .....</b>	<b>86</b>
	<b>    3.4.2    Fluorescence-based coupled ATPase activity assays .....</b>	<b>88</b>
	<b>    3.4.3    Helicase activity assays.....</b>	<b>90</b>
	<b>        3.4.3.1    Helicase activity assay on quenched Y-shaped DNA .....</b>	<b>90</b>
	<b>        3.4.3.2    Gel-based helicase activity assays.....</b>	<b>92</b>
<b>3.5</b>	<b>OPTICAL CONTROL OF CSY4 FUNCTION.....</b>	<b>95</b>
<b>3.6</b>	<b>CONCLUSIONS .....</b>	<b>101</b>
<b>3.7</b>	<b>EXPERIMENTAL.....</b>	<b>103</b>
<b>4.0</b>	<b>GENETIC ENCODING OF FLUORESCENT AMINO ACIDS .....</b>	<b>117</b>
<b>4.1</b>	<b>INTRODUCTION TO FLUORESCENT PROBES FOR PROTEINS.....</b>	<b>117</b>
<b>4.2</b>	<b>GENETIC ENCODING OF FLUORESCENT COUMARIN LYSINES ..</b>	<b>119</b>
	<b>    4.2.1    Incorporation of amino-coumarin lysines .....</b>	<b>119</b>
	<b>    4.2.2    Fluorescent reporting of protein localization.....</b>	<b>124</b>

<b>4.3</b>	<b>GENETIC ENCODING OF PHTHALIMIDE LYSINES .....</b>	<b>126</b>
<b>4.3.1</b>	<b>Introduction to phthalimide lysines .....</b>	<b>126</b>
<b>4.3.2</b>	<b>Incorporation of phthalimide lysines .....</b>	<b>128</b>
<b>4.3.2.1</b>	<b>Incorporation of phthalimide lysines in mammalian cells .....</b>	<b>128</b>
<b>4.3.2.2</b>	<b>Incorporation of phthalimide lysines in <i>E. coli</i> .....</b>	<b>129</b>
<b>4.4</b>	<b>CONCLUSIONS .....</b>	<b>134</b>
<b>4.5</b>	<b>EXPERIMENTAL.....</b>	<b>135</b>
<b>5.0</b>	<b>GENETIC ENCODING OF CAGED CYSTEINES.....</b>	<b>140</b>
<b>5.1</b>	<b>INTRODUCTION TO CAGED CYSTEINES .....</b>	<b>140</b>
<b>5.2</b>	<b>INCORPORATION AND APPLICATION OF NITROBENZYL-CAGED CYSTEINE ANALOGS.....</b>	<b>142</b>
<b>5.2.1</b>	<b>Incorporation of nitrobenzyl-caged cysteine analogs.....</b>	<b>144</b>
<b>5.2.2</b>	<b>Application of nitrobenzyl-caged cysteine analogs.....</b>	<b>150</b>
<b>5.3</b>	<b>INCORPORATION OF COUMARIN-CAGED CYSTEINE.....</b>	<b>153</b>
<b>5.4</b>	<b>CONCLUSIONS .....</b>	<b>155</b>
<b>5.5</b>	<b>EXPERIMENTAL.....</b>	<b>156</b>
<b>6.0</b>	<b>GENETIC ENCODING OF CAGED TYROSINES .....</b>	<b>161</b>
<b>6.1</b>	<b>INTRODUCTION TO CAGED TYROSINE ANALOGS .....</b>	<b>161</b>
<b>6.2</b>	<b>INCORPORATION OF CAGED TYROSINE ANALOGS – NBY, MNPY, NPY, AND NPPY.....</b>	<b>163</b>
<b>6.3</b>	<b>OPTICAL CONTROL OF PROTEIN FUNCTION.....</b>	<b>166</b>
<b>6.3.1</b>	<b>Optical control of firefly luciferase .....</b>	<b>166</b>
<b>6.3.2</b>	<b>Optical control of EGFP maturation .....</b>	<b>172</b>

6.3.3	Optical control of TEV protease .....	173
6.4	CONCLUSIONS .....	175
6.5	EXPERIMENTAL.....	176
7.0	GENETIC ENCODING OF PHOTOSWITCHABLE AMINO ACIDS .....	182
7.1	INTRODUCTION TO AZOBENZENE PHOTOSWITCHES.....	182
7.2	INCORPORATION OF AZOPHENYLALANINE ANALOGS .....	186
7.3	REVERSIBLE OPTICAL CONTROL OF PROTEIN FUNCTION IN LIVING CELLS .....	198
7.4	CONCLUSIONS .....	208
7.5	EXPERIMENTAL.....	210
8.0	GENETIC ENCODING OF PHOSPHINE-ACTIVATABLE AMINO ACIDS 217	
8.1	INTRODUCTION .....	217
8.2	GENETIC ENCODING OF <i>ORTHO</i> -AZIDOBENZYLOXYCARBONYL LYSINE .....	219
8.2.1	Incorporation of <i>ortho</i> -azidobenzyloxycarbonyl lysine (OABK) and deprotection .....	220
8.2.2	Small molecule-triggered fluorescent protein activation .....	224
8.2.3	Small molecule-triggered enzyme activation .....	226
8.2.4	Small molecule-triggered protein translocation .....	230
8.2.5	Small molecule-triggered DNA recombination.....	233
8.2.6	Small molecule-triggered CRISPR/Cas9 gene editing .....	236
8.3	GENETIC ENCODING OF AZIDOBENZYLOXYCARBONYL LYSINE ANALOGS .....	239

8.3.1	Incorporation of azidobenzyloxycarbonyl lysine analogs in <i>E. coli</i> .....	239
8.3.2	Incorporation of the azidobenzyloxycarbonyl lysine analogs in mammalian cells .....	244
8.3.3	Triggering of protein translocation through Staudinger reduction .....	245
8.3.4	Small-molecule control of post-translational modifications .....	248
8.4	GENETIC ENCODING OF THE AZIDOMETHYLENE AND THE AZIDOETHYL LYSINE ANALOGS .....	252
8.5	GENETIC ENCODING OF THE AZIDOMETHYLENE TYROSINE....	253
8.6	CONCLUSIONS .....	256
8.7	EXPERIMENTAL.....	258
9.0	SUMMARY AND OUTLOOK.....	267
9.1	CONCLUDING REMARKS .....	267
9.2	FUTURE WORK.....	268
10.0	EXPANDED METHODS .....	271
10.1	GENERAL MOLECULAR AND CELL BIOLOGY TECHNIQUES.....	271
10.1.1	Protein expression and UAA incorporation in <i>E. coli</i> .....	271
10.1.2	Mammalian cell maintenance and passage .....	272
10.1.3	Cell transfection and UAA incorporation in mammalian cells .....	273
10.1.4	Protein isolation from mammalian cells.....	274
10.1.5	Protein analysis (western blots).....	275
10.1.6	Cell imaging .....	276
10.1.7	Cell irradiation.....	277
10.1.8	Cell fixing .....	277

<b>10.1.9</b>	<b>RNA isolation .....</b>	<b>278</b>
<b>10.1.10</b>	<b>RNA analysis by quantitative real-time PCR (qRT-PCR) .....</b>	<b>279</b>
<b>10.2</b>	<b>COMMON RECIPE.....</b>	<b>280</b>
<b>10.2.1</b>	<b>DMEM media .....</b>	<b>280</b>
<b>10.2.2</b>	<b>PBS buffer (10×).....</b>	<b>280</b>
<b>10.2.3</b>	<b>TBS buffer (10×).....</b>	<b>281</b>
<b>10.2.4</b>	<b>TBST buffer (1×) .....</b>	<b>281</b>
<b>10.2.5</b>	<b>Transfer buffer (10×) .....</b>	<b>282</b>
<b>10.2.6</b>	<b>Transfer buffer (1×) .....</b>	<b>282</b>
<b>10.2.7</b>	<b>SDS running buffer (10×) .....</b>	<b>282</b>
<b>10.2.8</b>	<b>Staining buffer (coomassie blue).....</b>	<b>283</b>
<b>10.2.9</b>	<b>Destaining buffer .....</b>	<b>283</b>
<b>10.2.10</b>	<b>10% or 12% SDS-PAGE .....</b>	<b>283</b>
<b>10.2.11</b>	<b>4% SDS-PAGE stacking gel .....</b>	<b>284</b>
<b>10.2.12</b>	<b>Lysis buffer for protein purification.....</b>	<b>284</b>
<b>10.2.13</b>	<b>Wash buffer for protein purification.....</b>	<b>284</b>
<b>10.2.14</b>	<b>Elution buffer for protein purification.....</b>	<b>285</b>
<b>APPENDIX A .....</b>		<b>286</b>
<b>APPENDIX B .....</b>		<b>291</b>
<b>APPENDIX C .....</b>		<b>294</b>
<b>BIBLIOGRAPHY .....</b>		<b>297</b>

## LIST OF TABLES

Table 1.1: Examples of cellular features controlled by genetically encoded photoactuators.....	11
Table 1.2: Comparison of the reversible photosensitive proteins used in optogenetic systems ..	13
Table 2.1: Mammalian synthetase panel.....	30
Table 3.1: Primer list.....	104
Table 3.2: ATP turnover number of UvrDs in ATPase assays.....	114
Table 4.1: Summary of coumarin lysines. ....	120
Table 4.2: Primer list.....	135
Table 6.1: Primer list.....	177
Table 7.1: Measured absorption maxima ( $\lambda_{\max}$ , orange) and photostationary states (blue) of azobenzene amino acids in DMSO/PBS (pH 7.4) at the indicated wavelengths.....	193
Table 7.2: Potential residue candidates for azobenzene incorporation in luciferase (Zscore > 2). .....	201
Table 7.3: Binding energy of D-luciferyl-AMP upon repacking at W417 for the wild-type enzyme in the AzoF mutant in the <i>trans</i> and <i>cis</i> geometry.....	204
Table 7.4: Primer list.....	211
Table 7.5: Panel of <i>MbPylRS</i> mutants screened for incorporation of azobenzene amino acids.	215
Table 8.1: <i>MbPylRS</i> panel screening for incorporation of MPABK and PABK.....	243

Table 8.2: <i>MbPylRS</i> panel screening for incorporation of AzMY in <i>E. coli</i> .....	254
Table 8.3: Primer list.....	263
Table A: List of UAA abbreviation.....	286
Table B: List of publications.....	291

## LIST OF FIGURES

Figure 1.1: Unnatural amino acid (UAA) mutagenesis in cells.....	3
Figure 1.2: The PylRS/tRNA pair is orthogonal in both bacteria and eukaryotic cells.....	5
Figure 1.3: Comparison of naturally occurring and synthetic engineered control processes.....	6
Figure 1.4: Using light to control proteins in living systems in space and time.....	8
Figure 1.5: Schematic of different methods for optical control of protein function with light .....	9
Figure 1.6: PhyB and PIF interaction triggered by 650 nm and 750 nm light.....	13
Figure 1.7: Light-induced activation of translocation and DNA recombination in mammalian cells.....	15
Figure 1.8: Examples of photocaged amino acids.....	17
Figure 1.9: Optical control of protein signaling pathways through the application of genetically encoded caged amino acids.....	19
Figure 2.1: Representative scheme of the decaging process.....	22
Figure 2.2: Photolysis mechanism of <i>o</i> -NB caging group.....	23
Figure 2.3: Photolysis mechanism of NPP caging group.....	24
Figure 2.4: Maps of the pPCKRS-mCherry-TAG-EGFP-HA ( <i>Mb</i> PylRS) and p4CMVE-U6-PylT (PyltRNA <sub>CUA</sub> ) plasmids.....	26
Figure 2.5: Reaction catalyzed by <i>P. pyralis</i> firefly luciferase.....	27

Figure 2.6: Structural analysis of optical control of <i>P. pyralis</i> firefly luciferase.	28
Figure 2.7: <i>P. pyralis</i> firefly luciferase assay for decaging of PCK <i>in vivo</i> .	29
Figure 2.8: Screening optimal synthetases for incorporation of NPPK using the synthetase panel.	
.....	31
Figure 2.9: Screening optimal synthetases for incorporation of NPPK using Bright-Glo luciferase assays.	32
Figure 2.10: Screening optimal synthetases for incorporation of NPPK using the synthetase panel.	33
Figure 2.11: Screening optimal synthetases for incorporation of NPPK using Bright-Glo luciferase assays.	33
Figure 2.12: Chemical structure of NDBFK.	34
Figure 2.13: Screening optimal synthetases for incorporation of NDBFK.	36
Figure 2.14: Screening optimal synthetases for incorporation of NDBFK using Bright-Glo luciferase assays.	37
Figure 2.15: Genetic encoding of NPPK and NDBFK.	38
Figure 2.16: Bright-Glo luciferase assay for decaging of PCK, NPPK, and NDBFK <i>in vivo</i> .	40
Figure 2.17: Light activation of EGFP-NDBFK in live cells.	41
Figure 2.18: Light activation of NDBFK-SATB1-mCherry in live cells.	42
Figure 2.19: Representative scheme of Cre DNA recombination mechanism. <sup>64</sup>	44
Figure 2.20: Photo-activation of caged Cre recombinase.	46
Figure 2.21: Western blots for incorporation of PCK into Cre recombinase.	47
Figure 2.22: Cre Stoplight plasmid (pC-SL).	48
Figure 2.23: Optical control of caged Cre recombinase in live cells.	49

Figure 2.24: Quantification of fluorescent cells .....	50
Figure 2.25: Flow cytometry analysis of light activation of non-caged and caged Cre recombinase in HEK 293T cells.....	52
Figure 2.26: Spatial control of DNA recombination in live cells.....	53
Figure 3.1: Chemical structure of BHCK, HCK, and HC <sub>2</sub> K .....	62
Figure 3.2: Crystal structures of engineered PylRS binding pocket.....	63
Figure 3.3: Genetic encoding of BHCK, HCK, and HC <sub>2</sub> K .....	64
Figure 3.4: ESI-MS analysis of sfGFP-HCK, -BHCK, and -HC <sub>2</sub> K .....	66
Figure 3.5: MS/MS analysis of sfGFP-BHCK .....	67
Figure 3.6: SDS-PAGE analysis and in-gel fluorescence imaging of sfGFP-HCK and sfGFP-HC <sub>2</sub> K.....	68
Figure 3.7: ESI-MS analysis of UV-irradiated sfGFP-HCK and sfGFP-HC <sub>2</sub> K .....	69
Figure 3.8: Genetic encoding of HCK, BHCK, and HC <sub>2</sub> K in mammalian cells.....	70
Figure 3.9: MS/MS analysis for incorporation of HCK, BHCK, and HC <sub>2</sub> K in mammalian cells.....	72
Figure 3.10: Engineering of an optochemically controlled firefly luciferase through unnatural amino acid mutagenesis .....	74
Figure 3.11: Bright-Glo luciferase assays of hydroxycoumarin lysine derivatives.....	75
Figure 3.12: Bright-Glo luciferase assays of aminocoumarin lysine.....	76
Figure 3.13: Location of K85 (yellow) and interactions with D82, C70, and S72 in EGFP. <sup>124</sup> ..	77
Figure 3.14: Fluorescence imaging of EGFP folding before and after 365 nm light-activation. .	79
Figure 3.15: Fluorescence confocal imaging of EGFP folding before and after 405 nm and 760 nm light-activation.....	81

Figure 3.16: Time-course fluorescent imaging of SATB1 nuclear translocation in live cells. ....	83
Figure 3.17: Schematic showing the role of UvrD in bacterial NER. ....	85
Figure 3.18: Site-specific incorporation of coumarin lysine into UvrD at residue K37.....	87
Figure 3.19: MS/MS analysis of incorporation of HCK into UvrD. ....	88
Figure 3.20: Fluorescence-based coupled ATPase activity assay. ....	90
Figure 3.21: Fluorescence-based UvrD helicase activity assay using a quenched Y-shaped DNA duplex.....	92
Figure 3.22: Analysis of light-activated helicase activity using a substrate resembling the post-incision intermediate of bacterial NER.....	94
Figure 3.23: Schematic of the CRISPR/Cas locus.....	96
Figure 3.24: Csy4 active site in the Csy4/substrate complex (PDB: 2XLK). ....	97
Figure 3.25: Western blots to confirm expression of wild-type and caged Csy4 in mammalian cells. ....	98
Figure 3.26: Representative vector map of the GFP reporter for Csy4 activity. ....	99
Figure 3.27: Optical control of Csy4 endonuclease in HEK 293T cells using the GFP reporter. ....	100
Figure 3.28: Optical activation of Csy4 endonuclease in HEK 293T cells using the Gluc reporter. ....	101
Figure 3.29: NADH standard curve ( $\lambda_{\text{ex}} = 384 \pm 5 \text{ nm}$ , $\lambda_{\text{em}} = 461 \pm 5 \text{ nm}$ ).....	113
Figure 3.30: Fluorescence-based coupled ATPase activity assay. ....	114
Figure 4.1: Chemical structure of ACK and AC <sub>2</sub> K. ....	120
Figure 4.2: SDS-PAGE analysis of sfGFP-Y151TAG containing ACK through incorporation in <i>E. coli</i> . ....	121

Figure 4.3: SDS-PAGE analysis of sfGFP-Y151TAG containing HC <sub>2</sub> K and AC <sub>2</sub> K through incorporation in <i>E. coli</i> .....	121
Figure 4.4: ESI-MS analysis of sfGFP-AC <sub>2</sub> K.....	122
Figure 4.5: Incorporation of ACK in mammalian cells.....	123
Figure 4.6: Nuclear localization of NLS-wt/HCK-EGFP protein in mammalian cells.....	125
Figure 4.7: Chemical structures of APIC <sub>2</sub> K, DAPIC <sub>2</sub> K, and DAPIC <sub>3</sub> K.....	127
Figure 4.8: Excitation ( $E_m = 550$ nm) and emission ( $E_x = 400$ nm) spectra of APIC <sub>2</sub> K and DAPIC <sub>2</sub> K in PBS buffer (pH 7.4).....	127
Figure 4.9: Genetic incorporation of APIC <sub>2</sub> K and DAPIC <sub>2</sub> K in mammalian cells.....	129
Figure 4.10: A 12% SDS-PAGE coomassie gel of Myo-APIC <sub>2</sub> K and Myo-DAPIC <sub>2</sub> K.....	130
Figure 4.11: ESI-MS analysis of Myo-APIC <sub>2</sub> K.....	131
Figure 4.12: A 12% SDS-PAGE Coomassie stained gel (top) and fluorescent gel (bottom) of sfGFP-APIC <sub>2</sub> K and -DAPIC <sub>3</sub> K.....	132
Figure 4.13: A 12% SDS-PAGE gel of Ub-wt and Ub-AllocK.....	133
Figure 4.14: SDS-PAGE analysis for Ub-APIC <sub>2</sub> K expression.....	133
Figure 4.15: Fluorescence analysis of Ub-APIC <sub>2</sub> K and Ub-DAPIC <sub>2</sub> K in elution buffer compared to the control of Ub-AllocK ( $E_x = 420$ nm, $E_m = 595$ nm) .....	134
Figure 5.1: Chemical structures of the reported photocaged cysteines.....	142
Figure 5.2: Structures of genetically encoded photocaged amino acids PCK, PCHC, and PCC.....	143
Figure 5.3: Genetic incorporation of photocaged lysine PCK, photocaged homocysteine PCHC, and photocaged cysteine PCC in <i>E. coli</i> .....	144
Figure 5.4: ESI-MS analysis of sfGFP-PCC and decaged sfGFP-PCC.....	146

Figure 5.5: ESI-MS analysis of sfGFP-PCHC and decaged sfGFP-PCHC.....	147
Figure 5.6: Genetic incorporation of photocaged lysine PCK, photocaged homocysteine PCHC, and photocaged cysteine PCC in mammalian cells. ....	149
Figure 5.7: Bioluminescent reaction catalyzed by <i>Renilla</i> luciferase.....	150
Figure 5.8: Optochemical control of <i>Renilla</i> luciferase.....	152
Figure 5.9: Structures of the genetically encoded hydroxycoumarin Lys (HCK) and Cys (HCC). ....	153
Figure 5.10: Genetic incorporation of photocaged cysteine HCC in <i>E. coli</i> . ....	155
Figure 6.1: Structures of the caged tyrosine analogs <b>NBY</b> , <b>NPY</b> , <b>MNPY</b> , and <b>NPPY</b> .....	163
Figure 6.2: Incorporation of new photocaged tyrosines into protein in mammalian cells. ....	165
Figure 6.3: Optical control of firefly luciferase activity.....	167
Figure 6.4: Dual luciferase assay for incorporation and decaging of NBY, MNPY, NPY, and NPPY <i>in vivo</i> .....	170
Figure 6.5: Spatial control of light-activated FLuc function. ....	171
Figure 6.6: Optical control of EGFP maturation in live cells. ....	173
Figure 6.7: Optical control of TEV protease activity in live cells.....	175
Figure 7.1: Structures of photoswitchable phenylalanine analogs.....	185
Figure 7.2: Structures of photoswitchable amino acids that were genetically encoded in <i>E. coli</i> by the Lin group. <sup>210</sup> .....	186
Figure 7.3: Photoisomerization of azobenzene analogs AzoF, F <sub>2</sub> AzoF, and F <sub>4</sub> AzoF. ....	187
Figure 7.4: HPLC profiles of the photoequilibrium of AzoF after illumination. ....	189
Figure 7.5: HPLC profiles of the photoequilibrium of F <sub>2</sub> AzoF after illumination. ....	190
Figure 7.6: HPLC profiles of the photoequilibrium of F <sub>4</sub> AzoF after illumination. ....	191

Figure 7.7: Absorption spectra of <i>trans</i> - and <i>cis</i> -photoisomers.....	192
Figure 7.8: Genetically encoding of azobenzene amino acids in pro- and eukaryotic cells.....	195
Figure 7.9: ESI-MS analysis of sfGFP-AzoF, -F <sub>2</sub> AzoF, and -F <sub>4</sub> AzoF.....	197
Figure 7.10: Crystal structures of firefly luciferase in three states.....	199
Figure 7.11: Mutant screening for reversible photoswitching of firefly luciferase containing AzoF in live mammalian cells.....	202
Figure 7.12: Reversible photoswitching of Firefly luciferase containing a W417AzoF mutation in live mammalian cells.....	205
Figure 7.13: Thermostability tests of AzoF and F <sub>4</sub> AzoF.....	207
Figure 7.14: Reversible photoswitching of FLuc-W417→F <sub>4</sub> AzoF activity in live mammalian cells.....	208
Figure 7.15: Newly constructed plasmid maps for mammalian cell experiments.....	211
Figure 8.1: Schematic of the phosphine-triggered protein activation through protecting group removal via Staudinger reduction.....	220
Figure 8.2: Genetic encoding of OABK in mammalian cells.....	221
Figure 8.3: Staudinger reduction of sfGFP-OABK.....	223
Figure 8.4: Small-molecule control of EGFP folding.....	225
Figure 8.5: Structures of the six phosphine derivatives – TCEP (1), TCEP ester (2), TPP (3), TPPTS (4), 2DPBA (5), and 2DPBM (6). .....	226
Figure 8.6: Cytotoxicity measurements of phosphine derivatives using XTT assays in HEK 293T cells.....	227
Figure 8.7: Small-molecule activation of firefly luciferase.....	229
Figure 8.8: Small-molecule activation of SatB1 translocation.....	232

Figure 8.9: Small molecule-triggered DNA recombination.....	235
Figure 8.10: Small molecule-triggered CRISPR/Cas9 gene editing.....	238
Figure 8.11: Structures of genetically encoded azido amino acids – PABK, MeOABK and MePABK.....	240
Figure 8.12: Genetic incorporation of MOABK and MPABK in <i>E. coli</i> .....	241
Figure 8.13: <i>MbPylRS</i> panel screening for incorporation of MPABK (A) and PABK (B) into sfGFP in <i>E. coli</i> .....	243
Figure 8.14: Genetically encoding of MPABK and PABK in mammalian cells.....	245
Figure 8.15: Speeding up small-molecule control of protein translocation.....	247
Figure 8.16: Small-molecule-triggered SUMOylation of SUMO1 and RanGAP1.....	251
Figure 8.17: Structures of genetically encoded azido lysine analogs – AzMK and AzEK. ....	252
Figure 8.18: Small-molecule-triggered protein translocation.....	253
Figure 8.19: Structures of genetically encoded azido tyrosine – AzMY.....	254
Figure 8.20: Genetic encoding of AzMY in <i>E. coli</i> .....	255
Figure 8.21: ESI-MS analysis of sfGFP-AzMY.....	256

## **1.0 INTRODUCTION**

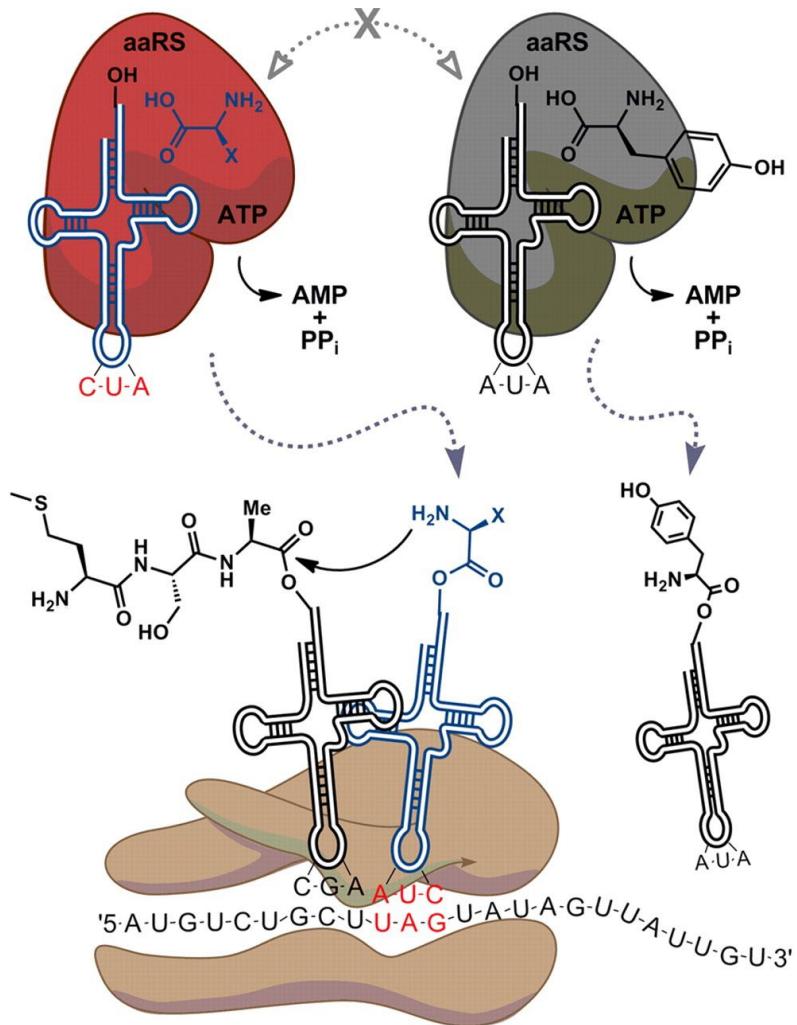
### **1.1 UNNATURAL AMINO ACID INCORPORATION**

Nature created 64 genetic codons (nucleotide triplets) to specify the canonical 20 amino acids used in protein translation in all known organisms.<sup>1</sup> However, it is clear that proteins require additional chemical complexity to carry out their functions, for example, the use of post-translational modifications and the necessity for enzyme cofactors.<sup>1-2</sup> Therefore, development of a method that enables the genetic encoding of novel amino acids will expand protein functions with new properties and provide a powerful tool for exploring protein structure and function both *in vitro* and *in vivo*.

While synthetic and semi-synthetic methods have been used to incorporate unnatural amino acids (UAs) into proteins, they have some limitations, such as small protein size, low yield, and technical complexity of large protein production.<sup>3-4</sup> It would be more beneficial to take advantage of the cellular translational machinery to introduce UAs into living cells. This was first accomplished by an *in vitro* biosynthetic method where a nonsense or frameshift suppressor tRNA was chemically aminoacylated with a specific UAA.<sup>4</sup> In the last few decades, this cell-free method was used to incorporate a large number of UAs into proteins at any site in response to a codon. However, there are limitations to this method including the small quantities of protein

produced, complexity of generating the chemical acylated aminoacyl-tRNA, and difficulty applying the system to living cells.<sup>1-2</sup>

Therefore, a strategy to incorporate UAAs directly into proteins *in vivo* is considered more advantageous for the study of protein structure and function due to its fidelity, high yield, and technical ease. Bioorthogonality is a minimal requirements to encode UAAs beyond the common 20 amino acids (Figure 1.1).<sup>1-2, 5</sup> First, a codon must specify a unique UAA at a defined site, but not encode any natural amino acid.<sup>5</sup> Second, the suppressor tRNA must only recognize the UAG amber stop codon and must be only aminoacylated with the UAA.<sup>5</sup> Moreover, the orthogonal aminoacyl-tRNA synthetase (aaRS) must efficiently aminoacylate an amber suppressor tRNA, not any endogenous tRNAs, and the suppressor tRNA must not be a substrate for any endogenous aaRS.<sup>5</sup> Also, aaRS should not use any endogenous amino acid as a substrate. An orthogonal aaRS/tRNA pair must function independently in the designed machinery and not interfere with the natural cellular aminoacylation process.<sup>5</sup> Importantly, the UAA should be non-toxic, should be stable to endogenous metabolic enzymes, and should be efficiently transported into the cytoplasm for successful incorporation into proteins.<sup>5</sup>



**Figure 1.1:** Unnatural amino acid (UAA) mutagenesis in cells.

An endogenous synthetase (right) aminoacylates a tRNA<sub>AUA</sub> with a canonical amino acid (tyrosine) while the mutant synthetase (left) aminoacylates a cognate tRNA<sub>CUA</sub> with the UAA. In the ribosome (bottom), the tRNAs recognize their corresponding codons and adds the amino acids to the growing peptide chain. Importantly, the peptide is not truncated at the amber stop codon in the presence of aminoacylated tRNA<sub>CUA</sub>. Image adapted with the permission from Young and Schultz, *J. Biol. Chem.* **2010**, 285(15):11039-11044. Copyright 2010 American Society for Biochemistry and Molecular Biology

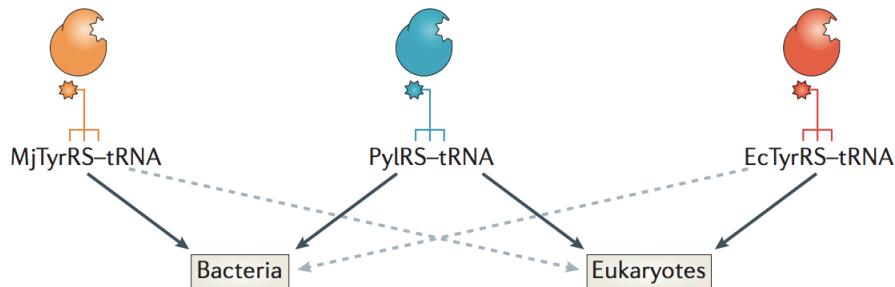
### **1.1.1 Incorporation of unnatural amino acids in *E. coli***

If bioorthogonality is not properly achieved and the suppressor tRNA is aminoacylated with an endogenous amino acid, then a dysfunctional mixture of full length protein populations will be synthesized. To overcome misaminoacetylation, the Schultz lab first evolved an engineered tyrosyl-aaRS and cognate nonsense suppressor tRNA (first orthogonal *MjTyrRS/MjtRNA<sup>Tyr</sup>*) pair from *Methanocaldococcus jannaschii* to successfully incorporate an unnatural amino acid, *O*-methyl-L-tyrosine.<sup>6</sup> The UAG (amber stop codon) was used to specify the unnatural amino acid to create an amber suppressor tRNA (*MjtRNA<sub>CUA</sub><sup>Tyr</sup>*) because it is the least-used stop codon in *E. coli*.<sup>6</sup> To select the specific orthogonal aaRS for recognition of a desired UAA, a large library of aaRS active site mutants ( $>10^8$ ) was created and subjected to a double sieve selection consisting of rounds of positive and negative selections.<sup>7</sup>

### **1.1.2 Incorporation of unnatural amino acids in mammalian cells**

Pyrrolysine (Pyl), the 22<sup>nd</sup> amino acid, was found to be encoded by UAG in *Methanosarcina barkeri* in 2002.<sup>8</sup> Pyrrolysyl-tRNA synthetase (PylRS) and pyrrolysyl-tRNA (Pylt) were identified in the *M. barkeri* genome,<sup>8</sup> and were demonstrated to be responsible for pyrrolysine incorporation. This discovery proved to be an attractive tool for expanding the genetic code for incorporating lysine analogues. Later it was demonstrated that PylRS/tRNA is a novel and orthogonal tRNA-aaRS pair, which is independent of other endogenous synthetases and tRNAs in both *E. coli* and mammalian cells. This system has the following advantages over the systems

developed previously: (a) PyltRNA<sub>CUA</sub> is a naturally perfect substrate for its cognate PylRS without being engineered; (b) with no need for mutation, the anticodon of PylT directly recognizes the UAG codon; (c) PylRS and tRNA<sup>Pyl</sup> is a perfect orthogonal pair because they have no cross-reactivity with other endogenous aaRSs or tRNAs; (d) this orthogonal pair can function in both *E. coli* and mammalian cells, which allows the mutant PylRS variants to be evolved in *E. coli* and then transferred to mammalian cells. (Figure 1.2). Therefore, in our lab, we have extensively exploited this system and further developed the *MbPylRS* to expand the genetic code and incorporate a variety of unnatural amino acids with novel functionalities and chemistries with high fidelity and efficiency.

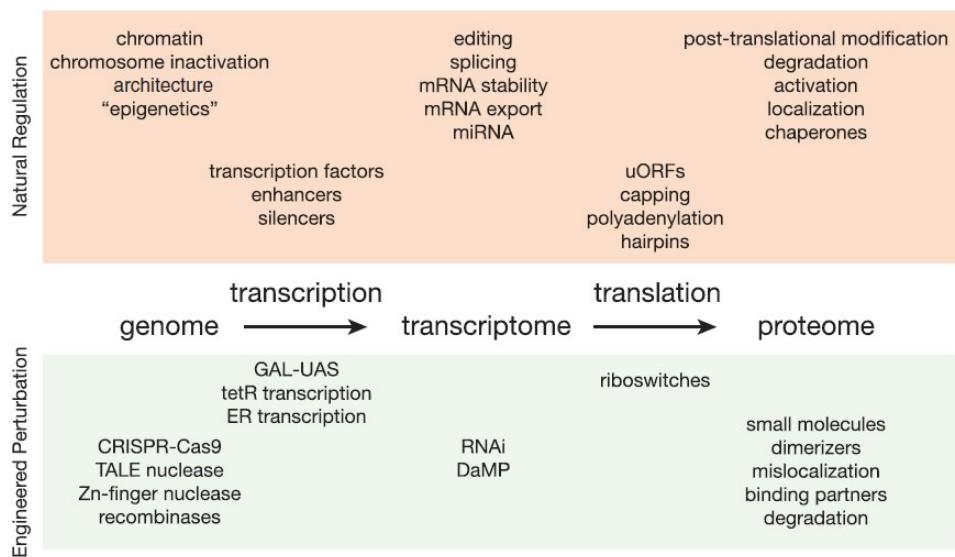


**Figure 1.2:** The PylRS/tRNA pair is orthogonal in both bacteria and eukaryotic cells.

The solid, black lines correspond to systems where the synthetase/tRNA pair has been utilized and the dotted, gray line represents non-orthogonal systems. Adapted with permission from Davis *et al.*, *Nat. Rev. Mol. Cell Biol.*, 2012, 13, 168-182. Copyright 2012 Nature Publishing Group

## 1.2 OPTICAL CONTROL OF PROTEIN FUNCTION

To dissect a complex biological process and elucidate its mechanism, it is crucial to examine it with conditional control comparable to that of the natural system (Figure 1.3). Traditional methods to regulate protein function at the genomic level, such as, knock-in, knock-out, knock-down, and small molecule approaches, provide many insights into the functions of the proteins involved, but typically suffer from long induction times, off-target effects, specificity, stability, inefficient delivery, and temporal precision.<sup>9</sup> A wide variety of perturbation strategies and technologies are available to regulate cellular systems at the genetic and transcript level, but the direct manipulation of functions at the protein level can offer significant advantages in precision, speed, and reversibility (Figure 1.3).<sup>9</sup>



**Figure 1.3:** Comparison of naturally occurring and synthetic engineered control processes.

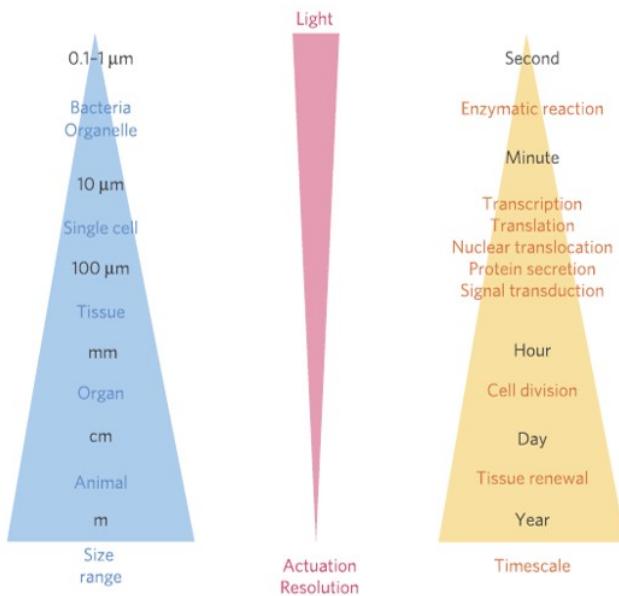
The central dogma of molecular biology governs the transcription of DNA (genome) into mRNA (transcriptome), which is then translated into proteins (proteome). Engineered experimental perturbation strategies mimic every

natural regulatory step in the flow of information from DNA into proteins. DaMP: decreased abundance of mRNA perturbation.

Adapted with permission from Rakhit, R.; Navarro, R.; Wandless, T. J. *Chemistry & Biology* 2014, 21, 1238.

Copyright 2014 Elsevier Ltd.

In this context, light is an attractive external trigger applied to the control of biological processes with high spatiotemporal resolution. Light is an excellent non-invasive technique owning all desirable attributes, such as allowing control in an acute, reversible, timed, and localized fashion.<sup>9-12</sup> Compared to standard genetic or pharmacologic perturbations, light illumination can be controlled with millisecond and submicron resolutions (Figure 1.4).<sup>13</sup> Light has high spatiotemporal resolution, which makes it possible to address biological processes with not only a wide range of time (from seconds for enzymatic reaction to days for tissue renewal) but also various spatial scales (from less than micrometers for organelles to more than centimeters for animals) (Figure 1.4).<sup>13</sup> Therefore, the introduction sections presented below mainly focus on optogenetic and optochemical control of protein function.



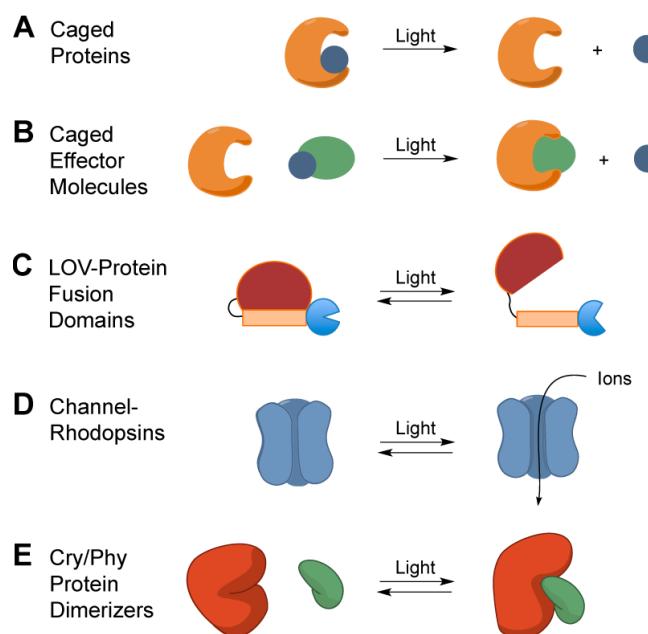
**Figure 1.4:** Using light to control proteins in living systems in space and time.

Adapted with permission from Gautier, A.; Gauron, C.; Volovitch, M.; Bensimon, D.; Jullien, L.; Vriz, S. *Nature Chemical Biology* 2014, 10, 533. Copyright 2014 Nature Publishing Group.

### 1.2.1 Optogenetic approaches

Optical control of proteins can be achieved through the application of multiple methodologies recent developments in the engineering of light-sensitive proteins and their application to the optical control of enzymatic and cellular processes will be discussed (Figure 1.5).<sup>14</sup> Two general approaches are most commonly used to control biological molecules with light. One approach to the light-regulation of protein activity is the application of optogenetic tools, which use genetically encoded, natural light-responsive proteins or protein domains to control biological processes. These optogenetic tools include genetically encodable optogenetic approaches, such as rhodopsins,<sup>15</sup> light-oxygen-voltage-sensing (LOV) domains, blue light-utilizing flavin

(BLUF),<sup>16</sup> the plant light-sensitive cryptochromes (Cry) and phytochromes (Phy), as well as the fluorescent protein Dronpa.<sup>17-19</sup> Alternatively, the other approach is the installation of exogenous synthetic light-removable protecting groups – “caging groups” (discussed in Section 2.1).<sup>13-14</sup> Caging groups can be added to a critical position or active site within the protein to render the molecule inactive until the group is removed through light exposure.<sup>20</sup>



**Figure 1.5:** Schematic of different methods for optical control of protein function with light.

(A) Caged amino acids can be genetically incorporated into proteins of interest, thereby rendering them inactive, until UV irradiation removes the caging group and restores protein activity. (B) An effector molecule can be caged, and UV irradiation can regenerate the active molecule. (C) Protein activity can be controlled by fusion of a light-oxygen-voltage (LOV) sensing domain that generates an active enzyme upon irradiation. (D) Light-sensitive ion channels, such as channel-rhodopsin, can be activated upon light exposure. (E) Protein dimerization can be induced upon illumination, such as, cryptochrome (Cry) and the Cry interacting basic helix-loop-helix (CIB), phytochrome (Phy) and the Phy interaction factor (PIF).

Adapted with permission from Baker, A. S.; Deiters, A. *ACS Chemical Biology* 2014, 9, 1398. Copyright 2014 American Chemical Society.

In optogenetic approaches, light-activation induces changes in structural perturbations, conformations, protein-protein interactions, or active site accessibility.<sup>23</sup> These optogenetic methods rely on the engineering of constructs with a naturally occurring light-sensitive protein (photoreceptor) from microbial or plant sources fused to the protein of interest in order to render it light-responsive.<sup>11</sup> The fusion proteins can be photoswitched through protein dimerization and protein conformation, enabling optical control of protein function. In most cases, under light, these naturally light-sensitive fusion proteins undergo conformational changes or changes of aggregation states due to isomerization of the associated chromophore (Figure 1.6).<sup>24</sup> Specifically, photon absorption induces conformational changes in these proteins through isomerization of a double bond or covalent adduct formation between the protein and its chromophore.<sup>13, 17-18</sup> These photogenetic approaches have been employed to optically control a variety of cellular functions, including kinase activity,<sup>25</sup> protein localization,<sup>26-27</sup> transcription,<sup>27-</sup><sup>28</sup> cytoskeleton dynamics,<sup>29</sup> protein trafficking,<sup>30</sup> and signaling (Table 1.1).<sup>13</sup>

**Table 1.1:** Examples of cellular features controlled by genetically encoded photoactuators.

Light target	Light-gated module	Biochemical output	Targeted feature
UVR8	VSVG-YFP-UVR8-UVR8	ER retention	Protein secretion
Retinal animal opsin	NinaE* 'chARGe' GPCR	IP <sub>3</sub> , DAG	Neuron activation
Retinal micr. opsin	ChR2 channel (Na <sup>+</sup> , K <sup>+</sup> ) <sup>a</sup>	Ion flux	Neuron activation
Retinal animal opsin	Rh-5HT1A GPCR	G-gated channel	Neuron activation
Retinal animal opsin	Rh-β2AR GPCR	cAMP	Signaling
Retinal animal opsin	Rh-α <sub>2</sub> AR GPCR	IP <sub>3</sub> , DAG	Signaling
Retinal micr. opsin	MChR1 channel (Na <sup>+</sup> , K <sup>+</sup> ) <sup>a</sup>	Ion flux	Neuron activation
Retinal micr. opsin	ChR1 channel (H <sup>+</sup> ) <sup>a</sup>	Ion flux	Neuron activation
Retinal micr. opsin	VChR1 channel (Na <sup>+</sup> , K <sup>+</sup> ) <sup>a</sup>	Ion flux	Neuron activation
Retinal micr. opsin	eNpHR3.0 pump (Cl <sup>-</sup> ) <sup>a</sup>	Ion flux	Neuron inhibition
Retinal micr. opsin	NpHR pump (Cl <sup>-</sup> ) <sup>a</sup>	Ion flux	Neuron inhibition
FAD CRY	CRY2-Gal4BD transcription factor	CIB1-Gal4AD binding	Transcription initiation
FAD CRY	CRY2-CreN recombinase	CIBN-CreC binding	Recombination
FAD CRY	Tale-CRY DNA binding	CIB1-VP64 binding	Transcription initiation
FAD CRY	Tale-CRY DNA binding	CIB1-Sin3IDx4 binding	Histone acetylation
FMN LOV2	DHFR(x2)-LOV2 DHRF	TH-folate synthesis	Nucleotide biosynthesis
FMN LOV2	LOV2-TAP Trp repressor	DNA binding	Gene expression
FMN LOV2	Lov-Rac1 small GTPase	GTPase	Actin dynamics
FMN FKT1	FKT1-VP16AD transcriptional activator	GI-Gal4BD binding	Transcription initiation
FMN FKT1	FKT1-Rac1 small GTPase	G1Cher-CAAX binding	Actin dynamics
FMN LOV2	LOV2-degron-targeted protein	Ubiquitination	Protein degradation
FMN LOV	mPAC adenylate cyclase <sup>a</sup>	cAMP	Signaling
FAD VVD-LOV	Gal4-vivid transcription factor	DNA binding	Gene expression
FAD BLUF	bPAC adenylate cyclase <sup>a</sup>	cAMP	Signaling
Bilin PHY	PHY-Gal4BD transcript. factor	PIF-Gal4AD binding	Transcription initiation
Bilin PHY	PHY-Cdc42 small GTPase	PIF-WASP binding	Actin dynamics
Bilin PHY	PHY-mCherry-CAAX mb. anchor	PIF -YFP binding	Protein trafficking
Bilin PHY	PHY-mCherry-CAAX mb anchor	iSH-YFP-PIF PI3K activation	Signaling

Reprinted with permission from Gautier, A.; Gauron, C.; Volovitch, M.; Bensimon, D.; Jullien, L.; Vriz, S. *Nature Chemical Biology* 2014, 10, 533. Copyright 2014 Nature Publishing Group.

Rhodopsins are the first-used optogenetic tools and found in all kingdoms of life, which consist of a chromophore and retinal bound to a seven-transmembrane protein. Upon illumination, the retinal undergoes isomerization, leading to conformational changes in the opsin backbone (Figure 1.5).<sup>31</sup>

Additionally, the LOV domains from several different organisms have been successfully employed as genetic tools.<sup>32-33</sup> LOV domains are photoreceptors sensitive to blue light (440~473 nm) and use endogenous flavin mononucleotide as a chromophore to modulate conformational changes (Table 1.2).<sup>17, 33</sup> Photoexcitation of the LOV domain induces a structural change that allows for activation. In some LOV systems, the LOV domains heterodimerize with natural or engineered binding partners to recruit signaling domains.<sup>32-33</sup> The LOV domain has also been used to control the activity of dihydrofolate reductase,<sup>34</sup> GTPases,<sup>35</sup> and tryptophan repressors.<sup>36-</sup>

<sup>37</sup> In other optogenetic systems, the domains homodimerize and bind to DNA, thereby regulating gene expression.<sup>32</sup> In a recently general application, the LOV domain fused to a peptide sequence was used to optically control peptide affinity for their binding partners by reversibly blocking binding-critical residues. The LOV domain binds a carboxy-terminal J $\alpha$  helix that unfolds through conformational changes upon light absorption by flavin mononucleotide chromophore. Light-driven unfolding of J $\alpha$  helix results in protein-peptide interaction.

Phytochromes (Phy) are natural photoreceptors that change conformation through absorptions of a photon.<sup>29</sup> Phy is activated by red light (650 nm) and inactivated by infrared red light (750 nm, Table 1.2).<sup>17</sup> Upon exposure to red light, Phy that is bound to a chromophore (phycocyanobilin, PCB) undergoes conformational change and binds to a phytochrome interaction factor (PIF) within seconds (Table 1.2, Figure 1.7).<sup>14</sup> The association is reversed

within seconds upon exposure to infrared light or is stable for hours in the dark (Figure 1.7). This technology has been used to control transcription, histone acetylation, signaling, and membrane localization (Table 1.1).<sup>13-14</sup> A major advantage of the Phy/PIF system is the ability to reversibly photoswitch between active (red) and inactive (far-red) states using light, while a significant drawback of the Phy/PIF system is the requirement to supplement the cells with exogenous PCB.<sup>17-18, 20</sup>



**Figure 1.6:** PhyB and PIF interaction triggered by 650 nm and 750 nm light.

Images adapted with permission from Levskaya, A.; Weiner, O. D.; Lim, W. A.; Voigt, C. A., Spatiotemporal control of cell signaling using a light-switchable protein interaction. 2009, *461* (7266), 997-1001. Copyright 2009 Nature Publishing Group.

**Table 1.2:** Comparison of the reversible photosensitive proteins used in optogenetic systems.

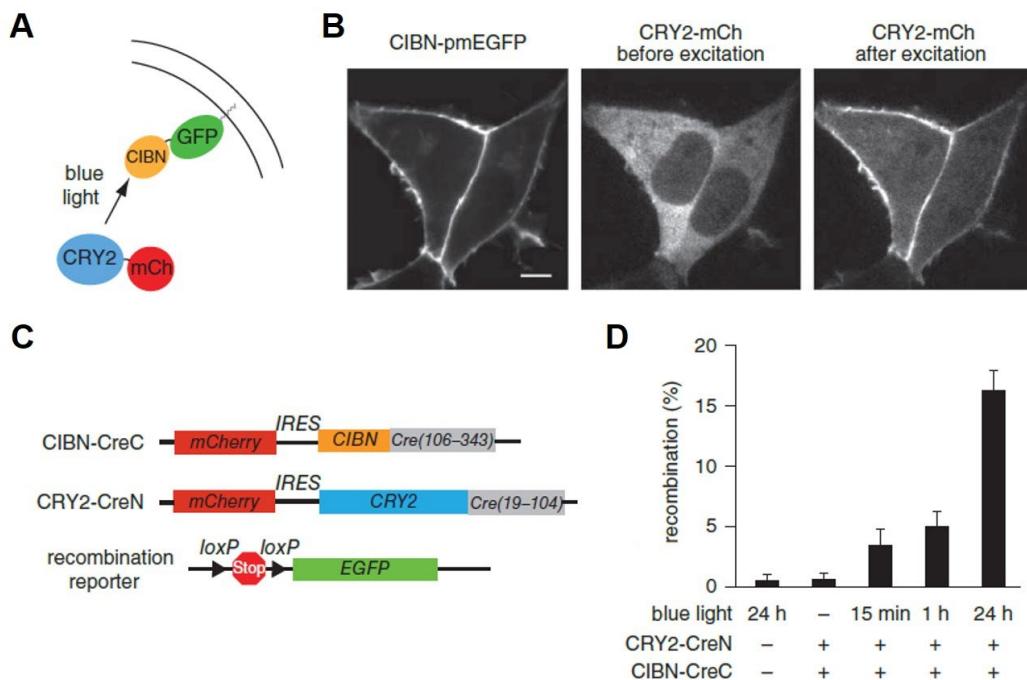
Photosensitive protein	Turn-on speed	Turn-off speed ( $t_{1/2}$ )	Chromophore requirement	Compatible imaging wavelengths (nm)	$\lambda_{on}$ (nm)	$\lambda_{off}$ (nm)	Effector affinity
PHYB	Seconds	<ul style="list-style-type: none"> <li>Seconds (illuminated at 750 nm)</li> <li>Hours (dark reversion)</li> </ul>	PCB; exogenous or synthesized <i>in situ</i>	$\leq 514$	650	750	<ul style="list-style-type: none"> <li>&lt;100 nM (post 650 nm)</li> <li>&gt;100 <math>\mu</math>M (post 750 nm)</li> </ul>
CRY2	Seconds	5 minutes	Flavin; endogenous	$\geq 561$	405–488	NA	Not determined
LOV	Seconds	Tens of seconds to minutes	Flavin; endogenous	$\geq 514$	440–473	NA	<ul style="list-style-type: none"> <li>1 <math>\mu</math>M (dark)</li> <li>100 <math>\mu</math>M (light)</li> </ul>
Dronpa	Seconds	<ul style="list-style-type: none"> <li>Tens of seconds (illuminated at 390 nm)</li> <li>Tens of minutes (dark reversion)</li> </ul>	None	$\geq 600$	390	490	<ul style="list-style-type: none"> <li>10 <math>\mu</math>M (post 490 nm)</li> <li>&gt;100 <math>\mu</math>M (post 390 nm)</li> </ul>

Adapted with permission from Tischer, D.; Weiner, O. D., *Nat. Rev. Mol. Cell Biol.*, 2014, *15*, 551. Copyright 2012 Nature Publishing Group.

As an alternative approach to the optical control of protein activity, cryptochromes (Cry) is a protein from *A. thaliana*, which is sensitive to blue light (405~488 nm), and undergoes conformational change through absorptions of a photon, allowing for the binding of a cryptochrome-interact basic-helix-loop-helix protein (CIB), and the Cry-CIB dimerization relies on an endogenous chromophore cofactor, flavin adenine dinucleotide (FAD), and thus requires no exogenous ligands (Table 1.2).<sup>14, 38</sup> In the Cry-CIB system, optical control of membrane localization has been achieved (Figure 1.8).<sup>27</sup> In this study, the full-length Cry2 was fused to a monomeric Cherry fluorescent protein (CRY2-mCherry) and its binding partner, a truncated version of CIB (CIBN), was fused to a membrane-localized EGFP (CIBN-pmEGFP, Figure 1.8A).<sup>27</sup> Before illumination with blue light, red fluorescence was shown throughout the cell (Figure 1.8B).<sup>27</sup> After a pulse of blue light, the red fluorescence was translocated to the membrane co-localized with pmEGFP fluorescence through a light-induced dimerization of Cry and CIB (Figure 1.8B).

In addition, optical control of DNA recombination through light-induced protein dimerization was developed using a split Cre recombinase system, fused to Cry2 and CIBN domains. (Figure 1.7C).<sup>27</sup> Cre recombinase targets palindromic *loxP* sites, cleaves the dsDNA and either insert or excise sequences through a double recombination event.<sup>39</sup> A plasmid containing a transcriptional stop sequence flanked by *loxP* sites preceding EGFP was used as a reporter for Cre recombinase activity (Figure 1.7C). This reporter expresses EGFP when the Cre recombinase is activated through dimerization. The Cre expression and reporter constructs were transfected into cells, and irradiated with pulses of blue light, ranging up to 24 h (Figure 1.7D).<sup>27</sup>

Before exposure to blue light, a very low level of recombination was observed in cells. Upon irradiation with blue light (450 nm), cells transfected with both Cry2-CreN and CIBN-CreC showed increased EGFP expression and the most significant increase (158-fold) in recombinase activity was observed with 24 h irradiation (Figure 1.7D).<sup>27</sup> This Cry-CIB system is entirely genetically encoded and can be activated using common light sources, so it could be utilized for optical control of a wide range of biological phenomena.



**Figure 1.7:** Light-induced activation of translocation and DNA recombination in mammalian cells.

**A)** Schematic showing fusion proteins (CRY-mCherry and CIBN-pmEGFP) for plasma membrane translocation through Cry and CIB dimerization induced by blue light. **B)** Fluorescence images of CIBN-pmEGFP and CRY2-mCh coexpressed in HEK 293T cells before and after a pulse of blue light (488 nm, 25  $\mu$ W). Scale bar, 5  $\mu$ m. **C)** Schematic showing the two split Cre recombinase constructs (CIBN-CreC and CRY2-CreN) and the Cre EGFP reporter construct. **D)** Cre reporter recombination was measured 48 h after cellular transfection with the Cre reporter and CRY2/CIBN constructs. Cells were exposed to blue light pulses (450 nm) for the indicated durations or kept in the dark.

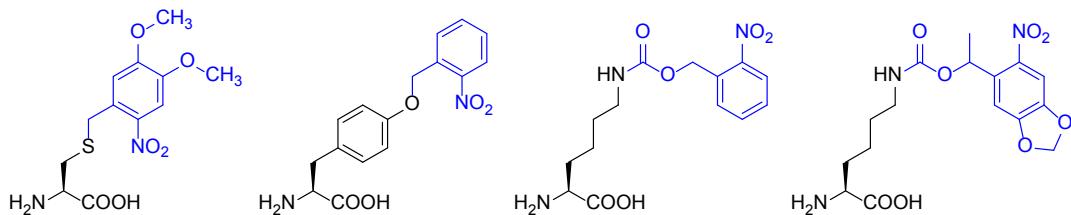
Adapted with permission from Kennedy, M.; Hughes, R.; Peteya, L.; Schwartz, J.; Ehlers, M.; Tucker, C., Rapid blue-light-mediated induction of protein interactions in living cells. *Nature Methods*, 2010, 7 (12), 973. Copyright 2010 Nature Publishing Group.

### 1.2.2 Photocaged amino acids

The optogenetic methods discussed above have been used successfully to control a wide range of biological processes, but they have some limitations.<sup>20</sup> First, these relatively large protein domains can cause significant structural perturbations and might thus influence protein expression, localization, or function. Furthermore, these methods require fusions of the light-responsive proteins and the protein of interest, which is difficult to implement and may not be generalizable to all proteins. The engineering of light-responsive fusion constructs with the protein of interest requires significant rounds of trial-and-error and intensive protein engineering.<sup>11, 17</sup> Therefore, the use of smaller site-specific caging modifications has been developed through incorporation of caged amino acids into proteins.

Several light-responsive unnatural amino acids for the photochemical control of protein function have been synthesized and successfully incorporated in response to an amber codon, such as nirobenzyl and nitropiperonyl derivatives of Cys,<sup>41</sup> Tyr,<sup>42</sup> and Lys (Figure 1.8).<sup>43</sup> The caging of proteins in cells through UAA incorporation has several advantages over other optogenetic approaches. First, the caging group is very small (~200 Da) and brings minimal or no perturbation to the overall protein structure. Second, the site-specific installation of the caged amino acid can be determined based on protein structure or protein activity. Moreover, light illumination removes the photolabile caging group restoring the native amino acid residue, so the

light-activated proteins are produced in live cells. Irradiation with non-toxic UV light (>360 nm) has been shown to cause no or minimal damage to biological systems.<sup>20</sup> The advantages of optogenetic approaches over UAA mutagenesis method with caging groups include: 1) there is no additional engineering of the protein biosynthetic machinery to expand the genetic code with an exogenous, caged amino acid; 2) it is reversible to optically modulate protein function using natural light-responsive proteins or protein domains.



**Figure 1.8:** Examples of photocaged amino acids.

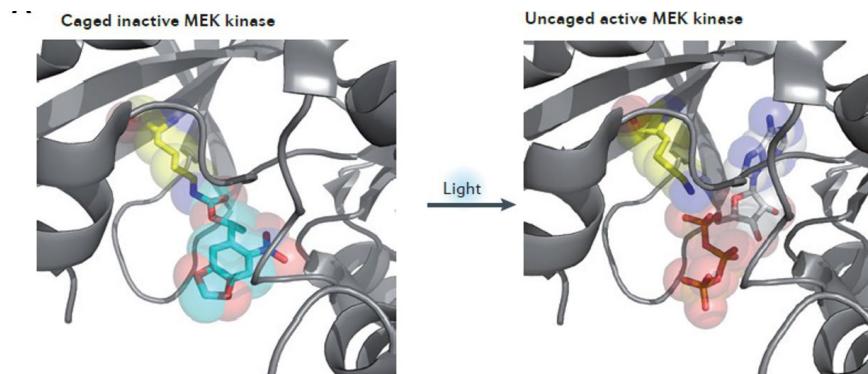
From left to right: Cys caged with a 4,5-dimethoxy-2-nitrobenzyl group; Tyr caged with a 2-nitrobenzyl group; Lys caged with a 2-nitrobenzyloxycarbonyl group; and Lys caged with a 6-nitropiperonyloxycarbonyl group. Photocaging groups are indicated in blue.

This developed optogenetic methodology enables the conditional control of various cellular processes with spatiotemporal precision in a non-invasive way, including enzymatic activation,<sup>44</sup> protein folding,<sup>44</sup> protein translocation,<sup>45</sup> gene repair (UvrD),<sup>46</sup> post-translational modification (SUMO1), DNA recombination (Cre),<sup>47</sup> RNA processing (Csy4), and gene editing (CRISPR/Cas9),<sup>45</sup> which will be discussed in the following chapters in detail. Additionally, the site-specific incorporation of caged unnatural amino acids has been used to induce light-mediated control of signal transduction,<sup>48-49</sup> protein localization,<sup>50</sup> intein,<sup>51</sup> ion channels,<sup>52</sup> DNA and RNA polymerase,<sup>53-54</sup> zinc finger nuclease,<sup>55</sup> transcription factors,<sup>56</sup> gene silencing,<sup>57</sup> gene editing,<sup>58</sup> and other fundamental biological processes. These proteins are caged

using the UAA mutagenesis in a manner that deactivates function, typically through inhibition of active site residues or steric blocking of conformational changes required for protein activity. Light irradiation then activates the caged protein to initiate a biological process of interest, which is monitored to demonstrate optical regulation of gene expression or protein function.

For example, this method is generally applicable to creating photoactivated kinases for other signaling pathways to enable the study of kinetics and elementary steps in signaling cascades and provide precise quantitative insights into kinase subnetworks in space and time. As an example for photochemical control of signal transduction, a photocaged lysine derivative (Lys caged with a 6-nitropiperonyloxycarbonyl group shown in Figure 1.8) was incorporated into MAP kinase kinase (MEK1) to achieve optical control over kinase signaling networks within the Raf/MEK/ERK pathway in mammalian cells. The photocaged lysine (**PCK**) was placed at a highly conserved Lys residue (K97) in the ATP binding pocket of the MEK1 catalytic domain, blocking the MEK1 function until light-induced decaging generates the native lysine (Figure 1.9).<sup>48</sup> In this study, isolating a subnetwork in MAP kinase cascade via genetically encoding a photocaged lysine in MEK1 active site was achieved through the UAA mutagenesis strategy for creating site-specific caged kinases. Furthermore, a tyrosine derivative caged with a 2-nirobenzyl group (shown in Figure 1.8) was applied to optochemical control of gene expression through DNA and RNA polymerase. The caged tyrosine was incorporated at a site crucial for DNA binding in order to obtain optical control of oligonucleotide polymerization. This example illustrates the ability to optically control DNA and RNA editing and expression through site-specific incorporation of genetically encoded photocaged amino acids into oligonucleotide-processing enzymes. Additionally, a photocaged cysteine (Cys caged with a 4,5-dimethoxy-2-nitrobenzyl group shown in Figure 1.8) was incorporated into the Tobacco Etch Virus (TEV)

protease in live HEK293T cells and optochemical control of protease activity was monitored with a FRET assay.



**Figure 1.9:** Optical control of protein signaling pathways through the application of genetically encoded caged amino acids.

Caging a highly conserved lysine (K97) in the MEK1 active site inactivates the enzyme by sterically blocking ATP binding. Decaging with light rapidly removes the caging group and activates the kinase. (PDB: 1S9J)

Adapted with permission from Gautier, A.; Deiters, A.; Chin, J. W. *J. Am. Chem. Soc.* **2011**, 133(7): 2124. Copyright 2011 American Chemical Society.

### 1.2.3 Project overview

The research presented here attempts to expand on UAA mutagenesis and photochemical approaches via the development of methodologies and technologies interfacing chemistry and biology. In this dissertation, genetic encoding of a series of diverse unnatural amino acids (UAAAs) have been achieved including photocaged lysines (Chapter 2), coumarin lysines (Chapter 3), phthalimide lysines (Chapter 4), photocaged cysteines (Chapter 5), photocaged tyrosines (Chapter 6), photoisomerizable phenylalanines (Chapter 7), and azido lysines (Chapter 8).

Overall, the research presented here has the following objectives (Chapter 2-8):

- 1) Genetic encoding of distinct amino acids with new functionality and chemistry;
- 2) Application of optically and chemically activatable amino acids in the spatial and temporal control of protein function;
- 3) Engineering of a light-switchable protein regulation system based on UAAs.

These aims will ultimately afford new means to investigate and modulate protein function with unprecedented spatiotemporal precision.

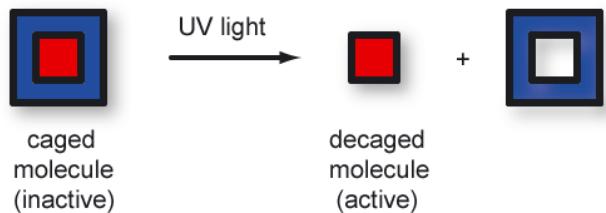
## **2.0 GENETIC ENCODING OF NITROBENZYL-CAGED LYSINES**

This chapter will describe genetic encoding and application of the known and new nitrobenzyl-caged lysine analogs – **PCK**, **NPPK**, and **NDBFK** in mammalian cells. A tightly regulated, light-activated Cre recombinase was developed through site-specific incorporation of the genetically-encoded photocaged lysine **PCK** in human cells. Excellent optical off to on switching of DNA recombination was obtained and spatial control of Cre recombinase was accomplished through patterned illumination.

### **2.1 INTRODUCTION TO CAGING GROUPS**

Photocaged biomolecules have been developed to study cellular or multicellular processes in a high spatio-temporal manner. Light is an unique external modulator for the study of various biological processes due to the following advantages over other regulatory tools: (a) Light acts as an ideal non-invasive technique for use in biology and as an orthogonal tool in most of live cell circumstances;<sup>21</sup> (b) Light causes minimal perturbation of cellular processes; (c) Light can be regulated in a high spatial and temporal fashion; (d) Light can easily be tuned in wavelength, amplitude, and intensity.<sup>12</sup> Photochemical control of biomolecules can be achieved through the installation of light-removable protecting groups, so-called “caging groups”. The addition of a caging group at a specific position renders the biomolecule inactive. Upon a UV light trigger, a

photochemical decaging reaction is induced to activate the caged biomolecule and the activity is restored (Figure 2.1).

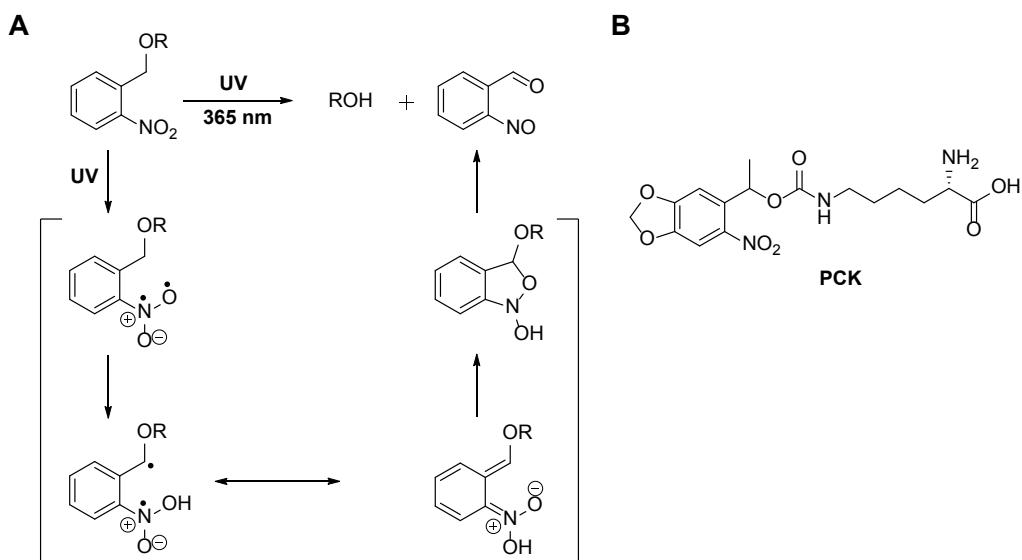


**Figure 2.1:** Representative scheme of the decaging process.

In a caging approach, the photolabile groups installed on the biomolecule of interest must meet the following requirements: (a) addition of the caging group must render the biomolecule nonfunctional, (b) upon photolysis, the biomolecule must be released quickly and efficiently in a functional form, (c) the wavelength necessary for decaging must not be harmful to the biological system, and (d) the byproduct of the decaging reaction must be biological inert.<sup>10,47</sup> For instance, our group, together with the Chin lab, developed the photocaged lysine (**PCK**) that is efficiently decaged at 365 nm, which results in an active lysine residue and a benign byproduct.<sup>48</sup> In our lab, with the development of unnatural amino acid mutagenesis, this strategy has been successfully applied to engineer several important light-activatable enzymes, including  $\beta$ -lactamase,<sup>49</sup> zinc finger nuclease,<sup>50</sup> DNA and RNA polymerase,<sup>51-52</sup> Cre recombinase,<sup>53</sup> and protein kinase MEK1.<sup>46</sup>

A variety of single-photon excitation and two-photon excitation caging groups have been reported in the literature.<sup>54-55</sup> One of the most common single-photon caging groups is the *ortho*-nitrobenzyl group. The *o*-nitrobenzyl (*o*-NB) type caged substrate absorbs light in the UV range (250–365 nm).<sup>56</sup> Photolysis of the caged substrate occurs via single-photon excitation. The

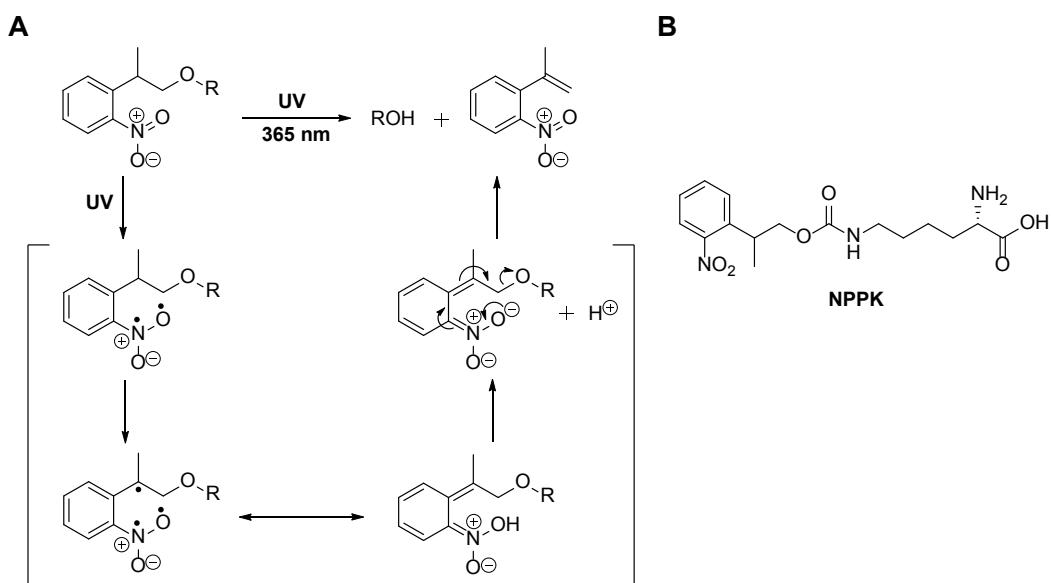
decaging process is based on a Norrish type II mechanism resulting in an active substrate and a nitroso-benzocarbonyl byproduct (Figure 2.2). The rate of photolysis depends on the nature of the caged substrate, pH, and dielectric constant of the medium.<sup>57</sup> One limitation is the formation of an *o*-nitroso carbonyl compound produced by the irradiation of the *o*-nitrobenzyl caging group, which is potentially toxic to cells. In the Deiters lab, an improved type of *o*-NB type caged group, *ortho*-nitropiperonyl, was applied to the caging of unnatural amino acids, including **PCK** (Figure 2.2B).<sup>58</sup> Through the same mechanism, however, photolysis at 365 nm of **PCK** generates a ketone byproduct, which is less toxic than a nitroso-benzaldehyde.<sup>39</sup> With the development of unnatural amino acid mutagenesis, this strategy has been successfully applied to engineer several important light-activable enzymes, including  $\beta$ -lactamase,<sup>59</sup> zinc finger nuclease,<sup>60</sup> DNA and RNA polymerase,<sup>61-62</sup> Cre recombinase,<sup>63</sup> and the protein kinase MEK1.<sup>64</sup>



**Figure 2.2:** Photolysis mechanism of *o*-NB caging group.

A) Photolysis mechanism of an *o*-NB type substrate. B) Chemical structure of the *o*-NB caged lysine.

In addition to the classical *o*-nitrobenzyl caging groups, a 2-(*o*-nitrophenyl) propanol (NPP) caging group has been developed.<sup>65-66</sup> Photolysis of the NPP caging group proceeds through a  $\beta$ -elimination pathway resulting in an active substrate and a less toxic *o*-nitro styrene byproduct. This has been applied to the generating of caged amino acids, including **NPPK** (Figure 2.3). The mechanistic pathway of decaging is shown in Figure 2.3A. An additional benefit of applying an NPP caging group is, that its photolysis rate is 1.8 fold faster than that of the *o*-NB caging group.<sup>67</sup>



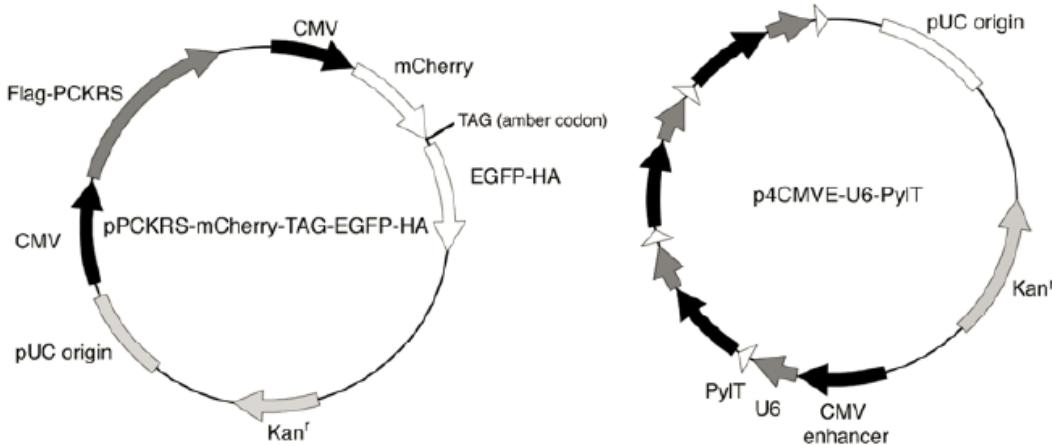
**Figure 2.3:** Photolysis mechanism of NPP caging group.

A) Photolysis mechanism of an NPP type substrate. B) Chemical structure of the NPP type lysine.

## **2.2 GENETIC ENCODING OF 2-(*O*-NITROPHENYL) PROPANOL AND 3-NITRO-2-ETHYLDIBENZOFURAN LYSINE**

### **2.2.1 Incorporation of 2-(*o*-nitrophenyl) propanol lysine (NPPK)**

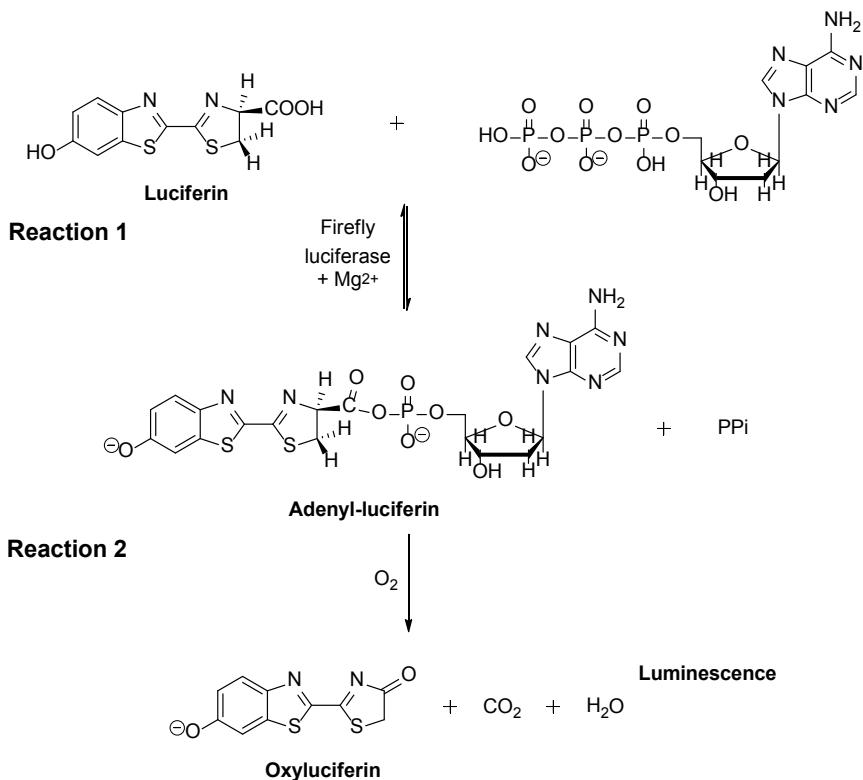
In order to screen for a functional synthetase capable of incorporating a desired unnatural amino acid in mammalian cells, an established mCherry-EGFP assay was used.<sup>48</sup> The PylT gene encoding the pyrrolysyl tRNA<sub>CUA</sub> requires an external promoter for transcription in eukaryotes. Thus, the PylT plasmid, named p4CMVE-U6-PylT, was constructed and allows for the expression of the pyrrolysyl tRNA<sub>CUA</sub> in mammalian cells (Figure 2.4 right).<sup>48</sup> The expression of the tRNA<sub>CUA</sub> is driven by a U6 promoter with an upstream CMV enhancer (CMVE). For copies of CMVE-U6-PylT were required for efficient transcription of PylT.<sup>48</sup> The mCherry-EGFP reporter (mCherry-TAG-EGFP-ha gene) contains an N-terminal mCherry gene, a linker containing a TAG amber stop codon, a C-terminal enhanced green fluorescent protein EGFP gene, and the HA-tag coding sequence.<sup>48</sup> The designed pMbPylRS-mCherry-TAG-EGFP-HA plasmid is for the expression of MbPylRS (with an N-terminal Flag-tag) with mCherry-TAG-EGFP (with a C-terminal HA-tag), both under the control of a CMV promoter (Figure 2.4 left). In cells expressing the MbPylRS/tRNA<sub>CUA</sub> pair, mCherry fluorescence can be detected in the presence or absence of UAA, but EGFP fluorescence can only be observed upon the incorporation of an UAA.



**Figure 2.4:** Maps of the pPCKRS-mCherry-TAG-EGFP-HA (*MbPylRS*) and p4CMVE-U6-PylT (*PyltRNA<sub>CUA</sub>*) plasmids.

Adapted with permission from Gautier *et al*, *J Am Chem Soc* 2010, 132(12): 4086. Copyright 2010 American Chemical Society.

To develop a sophisticated tool for offering precise tracking of protein expression and decaging within intact cells, firefly luciferase was considered to be an ideal enzyme. Luciferase enables quantitative measurements of decaging of differently caged lysine analogues *in vivo* because the bioluminescence technique shows low background, high sensitivity, and quantitative capability. A critical lysine residue (K206) on firefly luciferase was identified. Firefly luciferase (FLuc) is a 62 kDa protein from *Photinus pyralis* that catalyzes the production of light. In the presence of Mg-ATP and molecular oxygen, the enzyme oxidizes its substrate, luciferin, emitting yellow-green light.<sup>68</sup> Firefly luciferase catalyzes the conversion of luciferin to the highly luminescent oxyluciferin through a two-step process (Figure 2.5).<sup>68</sup> In the first step, the carboxylate group on luciferin is adenylated with Mg-ATP by FLuc (Figure 2.5).<sup>68</sup> In the second step, the luciferyl adenylate intermediate is oxidized, which releases a photon.<sup>69</sup>

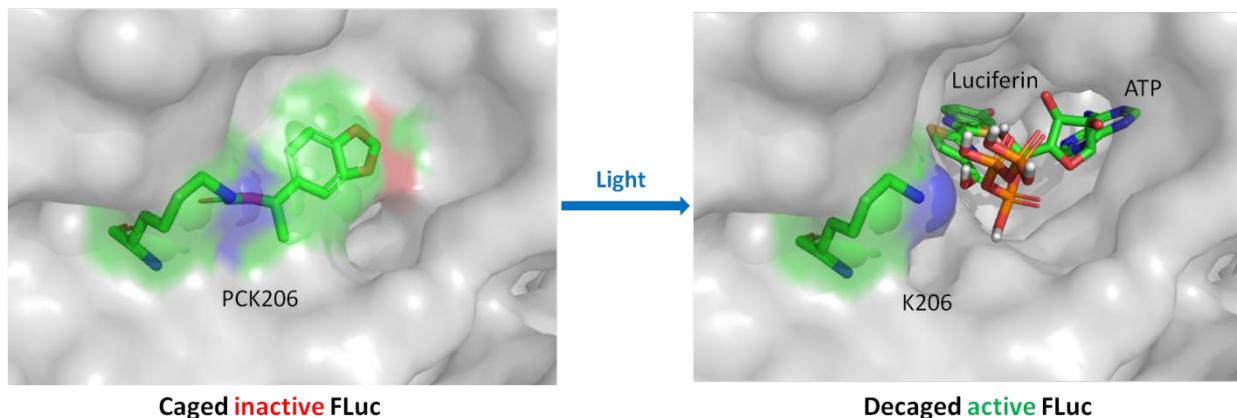


**Figure 2.5:** Reaction catalyzed by *P. pyralis* firefly luciferase.

ATP is hydrolyzed and light is emitted when firefly luciferase catalyzes the oxidation of D-luciferin.<sup>68</sup>

Based on the firefly luciferase crystal structure, K206 is located in the loop connecting the two  $\beta$ -strands of 6 and 7 in the  $\beta$ -sheet subdomain, and is a residue at the edge of the binding pocket of FLuc (Figure 2.6).<sup>68</sup> It can stabilize ATP in the active site when ATP enters towards the cleft (binding pocket) between the two subdomains of FLuc,<sup>68</sup> and it has been shown to be a key residue for the effective orientation of ATP in the interaction with luciferin.<sup>68</sup> The  $\epsilon$ -amino group ( $-\text{NH}_3^+$ ) on K206 provides a favorable hydrogen-bond interaction to stabilize a phosphate residue of ATP to promote the adenylation reaction with firefly luciferin. Thus, the conserved K206 site is essential for catalytic activity as shown by dramatically decreasing the enzymatic activity of the K206R mutant.<sup>68</sup> Therefore, we hypothesized that a photocaging group attached to K206 could potentially prevent access of the substrates (ATP and luciferin, Figure 2.6) towards

the binding pocket of FLuc. Photolysis of the caged lysine analogue by UV irradiation can remove the caging group to produce a native lysine and restore the catalytic activity of the firefly luciferase (Figure 2.6). Therefore, we assumed that PCK206 could be essential for the regulation of the catalytic activity of firefly luciferase by light. The stimulated structure shows a binding interaction of K206 with the phosphate adjacent to the adenosine at the entrance of the binding pocket (Figure 2.6).

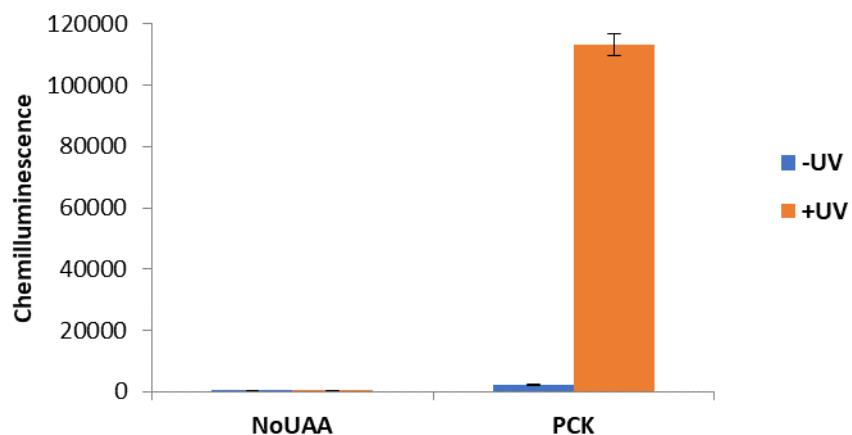


**Figure 2.6:** Structural analysis of optical control of *P. pyralis* firefly luciferase.

A) Photocaged K206 (PCK206) is blocking the entrance of the binding pocket from access by luciferin and ATP. B) The “enzyme-substrate” complex of FLuc interacting with ATP upon decaging, which can enter the cleft. (PDB: 2D1S)

Site-directed mutagenesis of the corresponding lysine to the amber codon (TAG) was performed on the firefly luciferase plasmid (pGL3) using Quikchange mutagenesis (Agilent). HEK293T cells were co-transfected with the pMbCKRS/PyltRNA<sub>CUA</sub> expression plasmids and pGL3-K206TAG in the absence or presence of **PCK** (1 mM). After a 24 hr incubation, the cells were either irradiated for 5 min (365 nm, 25 W) or kept in the dark, followed by a Bright-Glo luciferase assay. The incorporation of **PCK** into FLuc caused complete inhibition of luciferase

activity before UV irradiation, comparable to the negative control (no UAA, Figure 2.7). This demonstrates that installation of a caging group at K206 inhibited enzymatic catalysis. After UV irradiation, **PCK** was decaged to produce a free lysine, resulting in the activation of firefly luciferase. Therefore, the activity of firefly luciferase could be tightly regulated with spatial and temporal resolution by incorporation of a photocaged lysine residue into the key active site of the luciferase protein. The developed system was then applied to the testing of different caging groups.



**Figure 2.7:** *P. pyralis* firefly luciferase assay for decaging of **PCK** *in vivo*.

The FLuc activity of FLuc-**PCK** increased 55-fold after UV exposure. Virtually no activity was observed in the absence of **PCK**, regardless of UV exposure. Error bars represent standard deviation from three independent experiments.

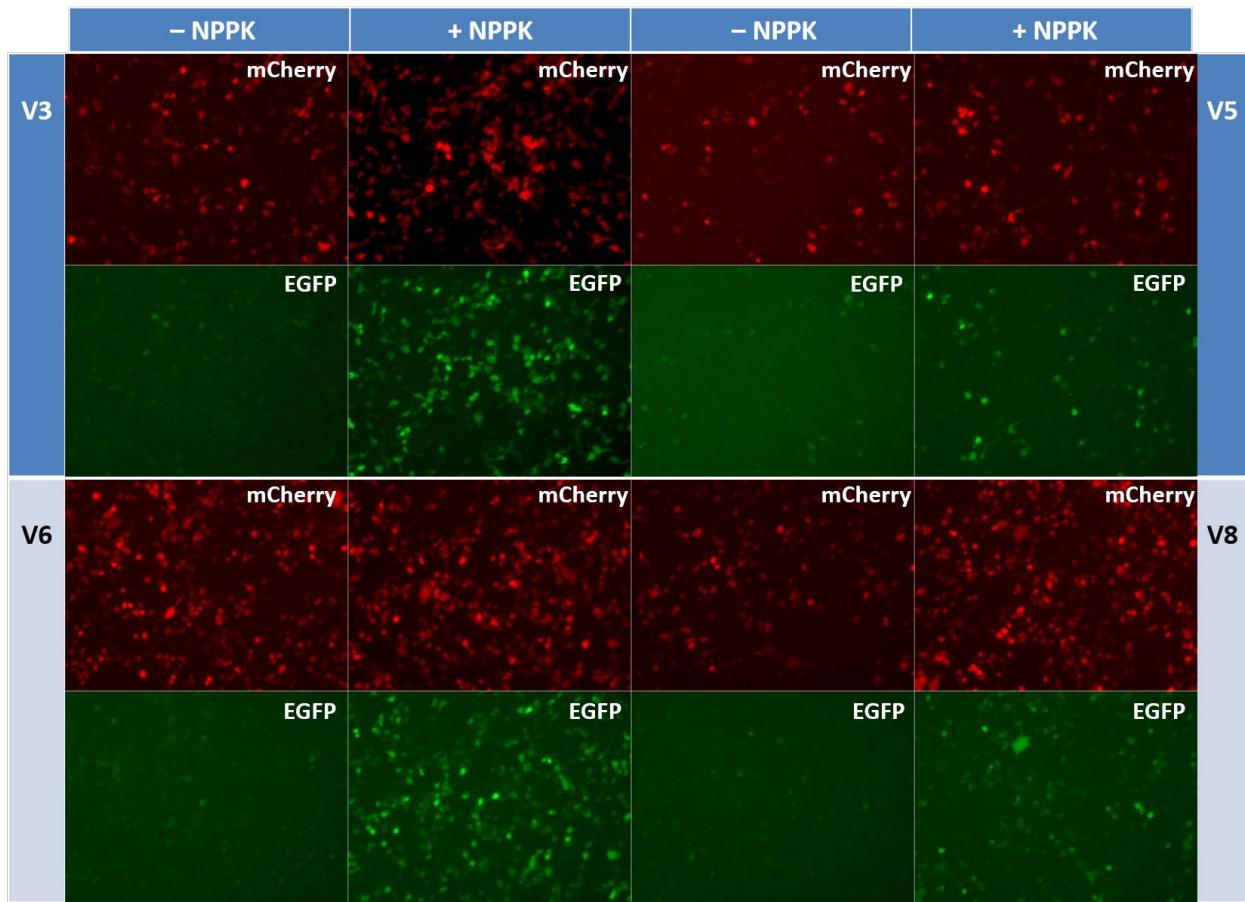
In order to test the decaging efficiency of the caged lysine analogs in living cells, the developed cagedFLuc strategy above was applied to the nitrobenzyl-caged lysine analogs below.

The luciferase assay could also be used as an alternative tool for screening the synthetases to demonstrate the fidelity of incorporation of caged lysine analogues.

NPP-caged lysine (**NPPK**, Figure 2.3) was synthesized by Dr. Rajendra Uprety in the Deiters Lab. To obtain an optimal synthetase for **NPPK**, the mCherry\_TAG\_EGFP assay was performed in mammalian cells (Figure 2.8). Through a synthetase screening (Table 2.1), Based on the fluorescent imaging results, synthetase V3 and V6 looked promising (Figure 2.8).

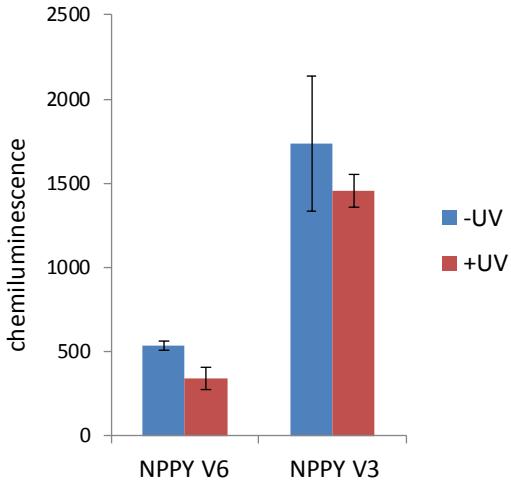
**Table 2.1:** Mammalian synthetase panel.

Code	Mutations							
	S 149	D 203	M 241	A 267	Y 271	L 274	C 313	M 315
V1								
V2			F	S	C	M		
V3						V	V	Q
V4	F					V		
V5						V	V	
V6						V		Q
V7	F					V	V	
V8							V	
V9						V		
V10					A			
V11		N				V	V	
V12					A	M		



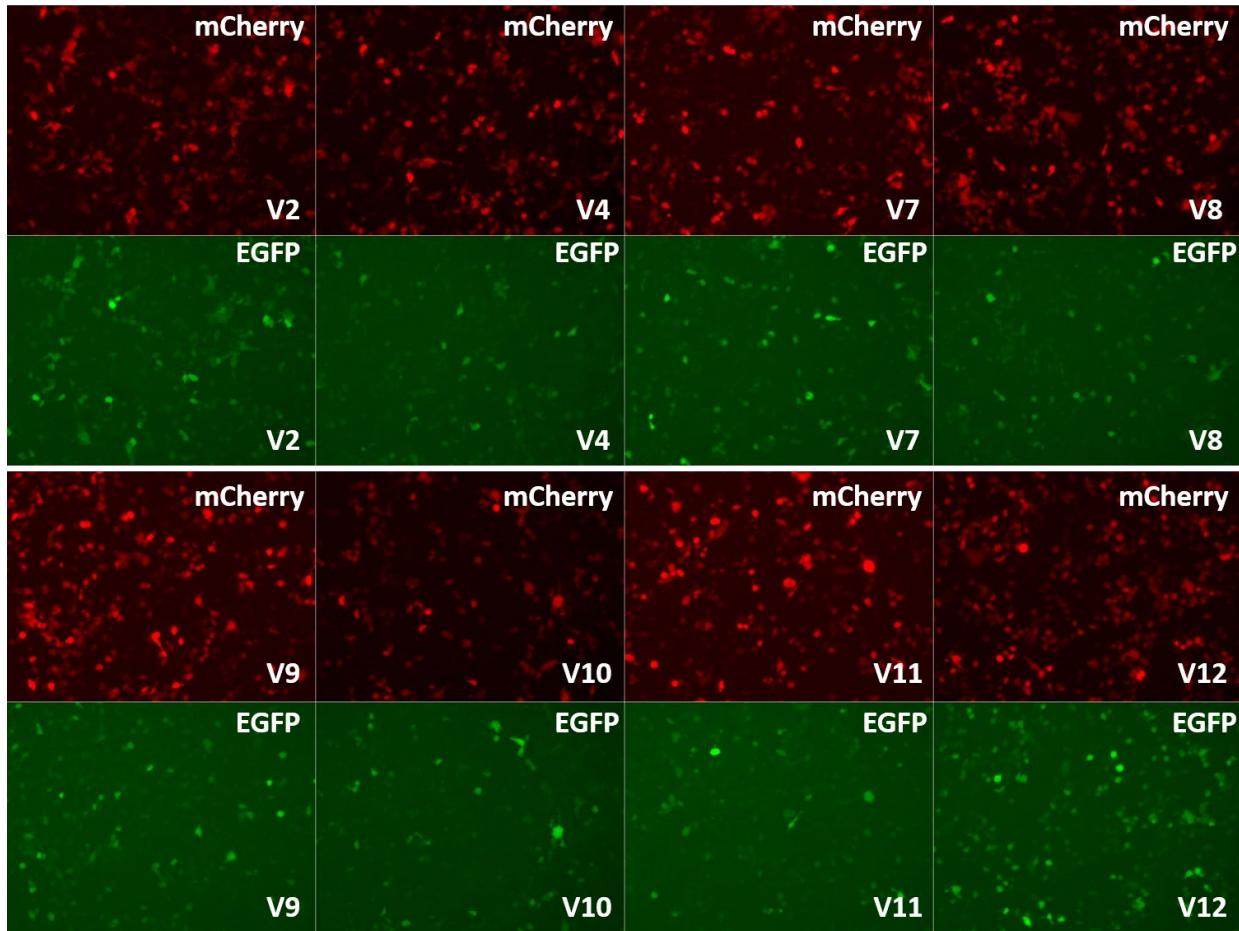
**Figure 2.8:** Screening optimal synthetases for incorporation of **NPPK** using the synthetase panel.

Bright-Glo luciferase assays were used to further confirm the screening results using mCherry-EGFP assays (Figure 2.9). However, the light activation of FLuc-**NPPK** was not observed using these two synthetases, indicating the potential synthetases were false positive ones for incorporation of **NPPK**.

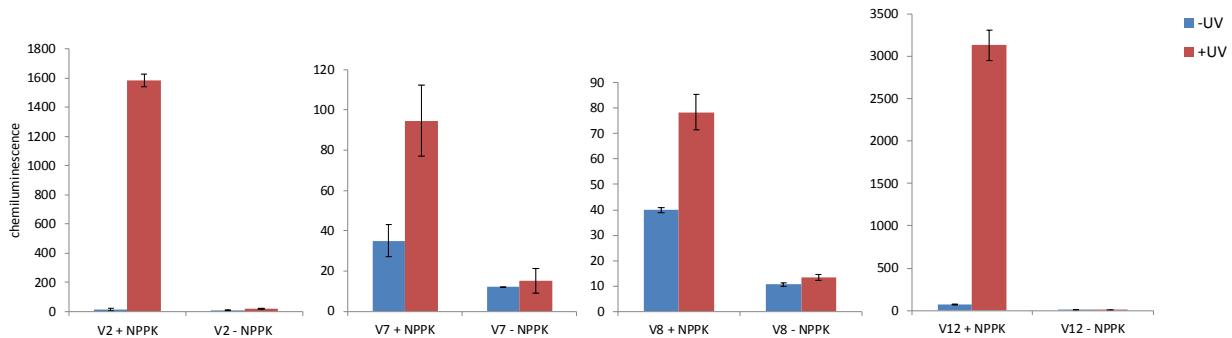


**Figure 2.9:** Screening optimal synthetases for incorporation of **NPPK** using Bright-Glo luciferase assays.

The screening was further carried out using other synthetases in the synthetase panel (Table 2.1, Figure 2.10). Based on the fluorescent imaging results, synthetase V2, V7, V8, and V12 might be hit synthetases for incorporation of **NPPK** or incorporate an endogenous amino acid. Subsequently, the Bright-Glo luciferase assays were conducted using synthetase V2, V7, V8, and V12 to confirm the results from the fluorescent assays. Gratifyingly, it showed that the light activation of FLuc-**NPPK** was observed using synthetase V2 and V12, but not synthetase V7 and V8 (Figure 2.11). Additionally, synthetase V12 has higher incorporation efficiency than V2.



**Figure 2.10:** Screening optimal synthetases for incorporation of **NPPK** using the synthetase panel.



**Figure 2.11:** Screening optimal synthetases for incorporation of **NPPK** using Bright-Glo luciferase assays.

Therefore, the pMbPylRS synthetase variant containing two mutations at Y271A and L274M was the optimal synthetase for incorporation of **NPPK** into protein (Figure 2.15). Bright-

Glo luciferase assays were further conducted to demonstrate the incorporation and activation of **NPPK** in mammalian cells (Figure 2.16). No further experiments, such as Western blots, expression in *E. coli*, and applications were conducted due to a lack of **NPPK**.

## 2.2.2 Incorporation of 3-nitro-2-ethyldibenzofuran lysine (NDBFK)

To enable decaging using IR light for decaging via two-photon excitation, *o*-nitrodibenzofuran lysine (3-nitro-2-ethyldibenzofuran, **NDBFK**, Figure 2.12) was synthesized by Dr. Qingyang Liu in the Deiters lab. Two properties, quantum yield ( $\phi$ ) and extinction coefficient ( $\varepsilon$ ), define the photochemical efficiency of all caging groups.<sup>70</sup> For the NDBF caging group, the quantum yield ( $\phi$ ) of its photolysis is 0.7 and the extinction coefficient ( $\varepsilon$ ) at the  $\lambda_{\text{max}}$  of 330 nm is 18,400 M<sup>-1</sup> cm<sup>-1</sup>,<sup>71-72</sup> thus photolysis of a NDBF derivative is more efficient than the other commonly-used nitrobenzyl caging groups.<sup>73</sup>



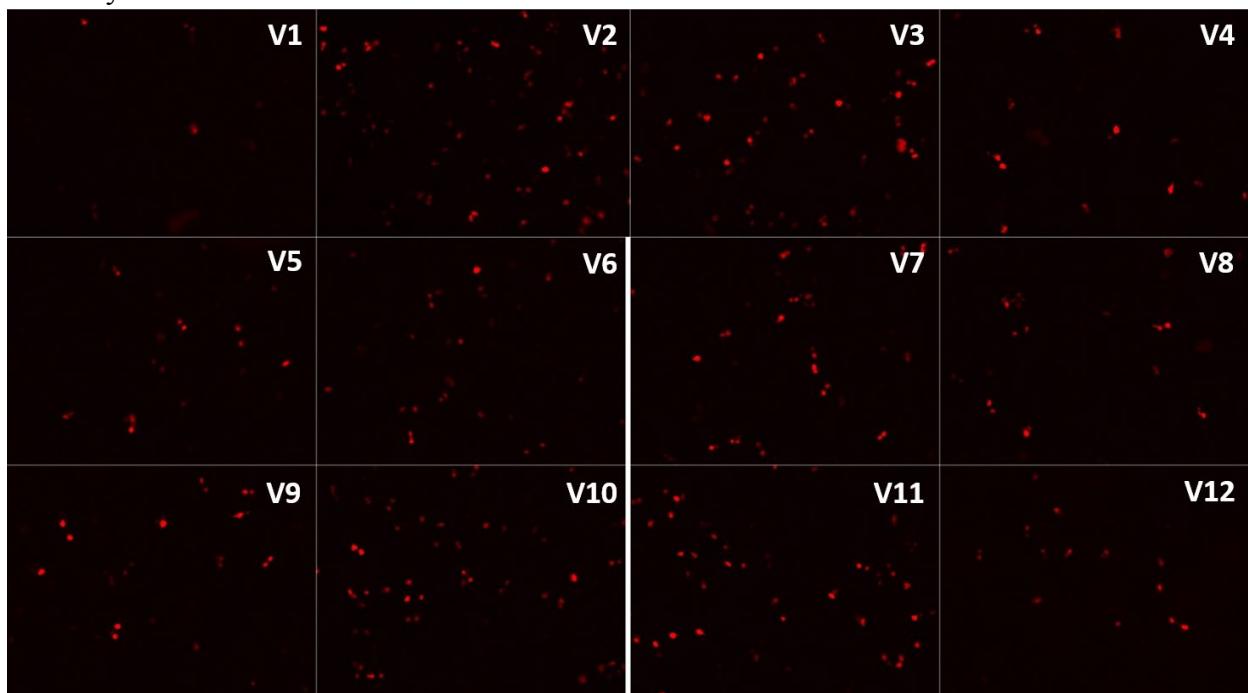
Figure 2.12: Chemical structure of **NDBFK**.

Previously, in collaboration with the Chin lab, a *MbPylRS* library was generated by randomizing 6 amino acid positions (S149, D203, Y271, L274, C313, and M315) in the binding site of the pyrrolysine ring.<sup>74</sup> Positive and negative selections were performed in *E. coli* in the presence and absence of **NDBFK**. From this selection, three potential NDBFK pyrrolysyl-tRNA

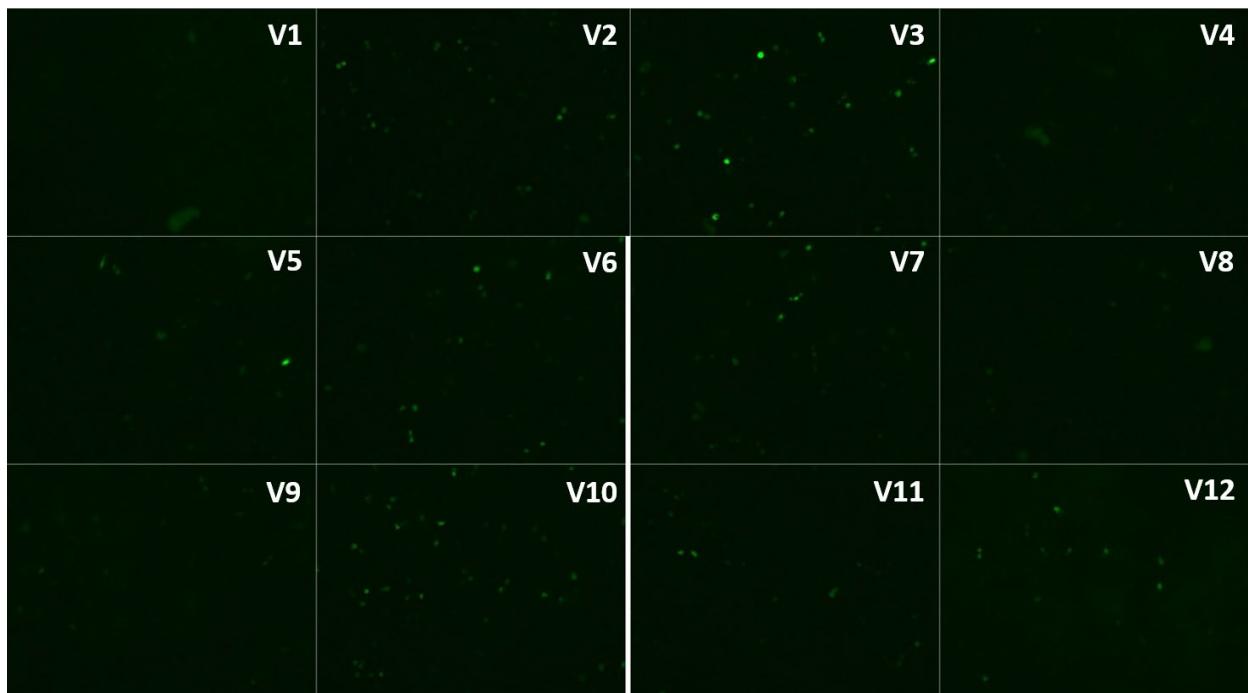
synthetases (NDBFKRS-1, -2, and -3) were obtained, including S149F\_L274V, S149F\_L274V\_C313V, and D203N\_Y271C\_L274V\_C313V\_M315Y.

However, for reasons unknown, **NDBFK** could not be incorporated into protein in mammalian cells by these evolved *MbPylRS* synthetases. Following this conventional evolution, we designed a small library of PylRS mutants with four active-site residues mutated to select amino acids (Y271A/M/V, L274A/M/V, C313A/V, and M315A/V/Q). The panel was screened using the mCherry\_TAG\_EGFP reporter and active synthetases contained the following mutations: Y271A, L274M/V, C313V, and M315Q (Table 2.1). Based on the screening results (Figure 2.13), the synthetases V3, V6, V10, and V12 were able to incorporate **NDBFK**.

mCherrychannel:

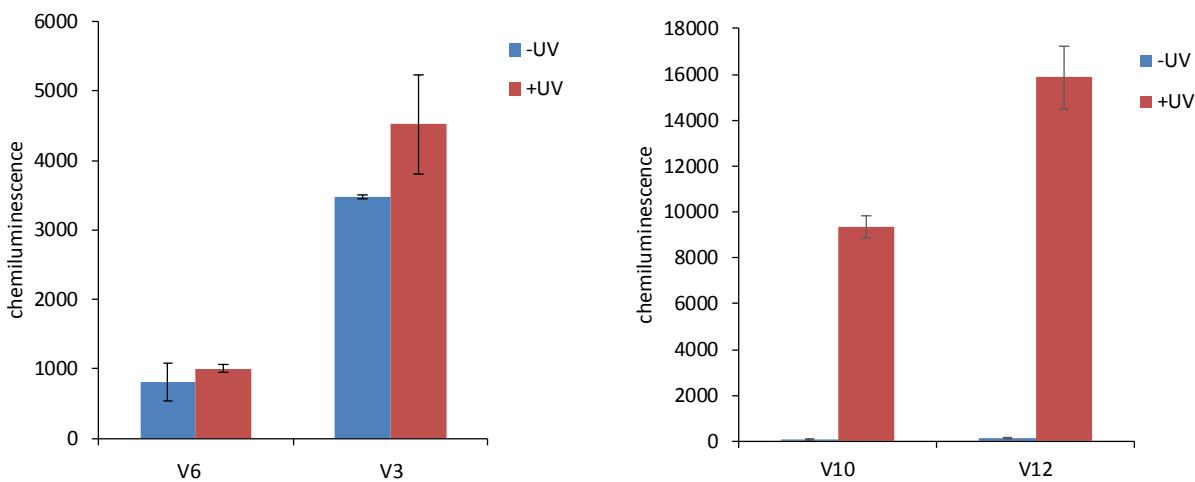


EGFPchannel:



**Figure 2.13:** Screening optimal synthetases for incorporation of **NDBFK**.

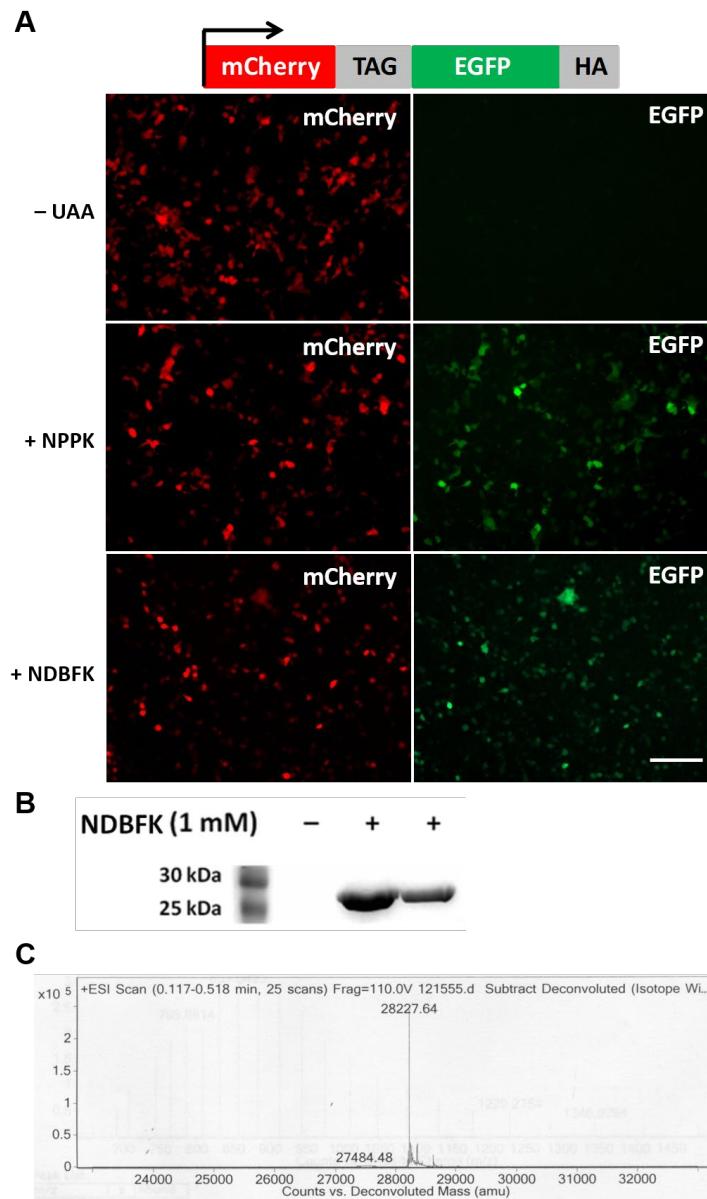
Similarly, to determine the fidelity of the incorporation and the efficiency of photolysis of **NDBFK** using the hit synthetases, Bright-Glo luciferase assays were further carried out to confirm the fluorescent screening results (Figure 2.14). The light activation of FLuc-**NDBFK** was observed using synthetase V10 and V12, but not synthetase V3 and V6, indicating V10 and V12 are the potential synthetases for incorporation of **NDBFK**. Based on the chemiluminescence signals, the luciferase assays also demonstrate that synthetase V12 is the most efficient one for incorporation of **NDBFK**. Therefore, the most active synthetase, Y271A\_L274M, termed NDBFKRS, was used in subsequent incorporation experiments.



**Figure 2.14:** Screening optimal synthetases for incorporation of **NDBFK** using Bright-Glo luciferase assays.

To confirm the site-specific incorporation of **NDBFK** using the NDBFKRS/PyltRNA<sub>CUA</sub> pair in *E. coli*, we expressed and purified sfGFP containing **NDBFK** in place of Y151 in good yield (3.5 mg/L) from Top10 competent cells (Figure 2.15). This yield is comparable to the yield obtained for other well-incorporated UAAs.<sup>75</sup> Electrospray ionization mass spectrometry (ESI-MS) of sfGFP-Y151-**NDBFK** showed a molecular weight of 28227.94 Da (expected mass

28482.27 Da), which is consistent with the molecular weight of sfGFP upon decaging. This indicates that **NDBFK** was decaged during *E. coli* expression or mass spec analysis.<sup>76</sup>

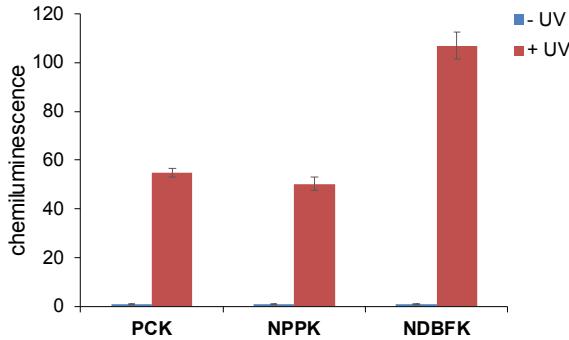


**Figure 2.15:** Genetic encoding of **NPPK** and **NDBFK**.

A) Incorporation of **NPPK** and **NDBFK** in mammalian cells. Fluorescence micrographs of HEK 293T cells expressing pMbPylRS-mCherry-TAG-EGFP-HA and PyltRNA<sub>CUA</sub> in the absence and presence of **NPPK** (1 mM) and **NDBFK** (0.25 mM). Scale bar indicates 100  $\mu$ m. B) Coommassie stained gel of Ni-NTA purified sfGFP-

Y151TAG expression in the absence and presence of **NDBFK** (1 mM). C) ESI-MS analysis confirmed that **NDBFK** was converted to lysine in sfGFP after protein expression in *E. coli*. (observed mass: 28227.64 Da, expected mass: 28482.27 Da.)

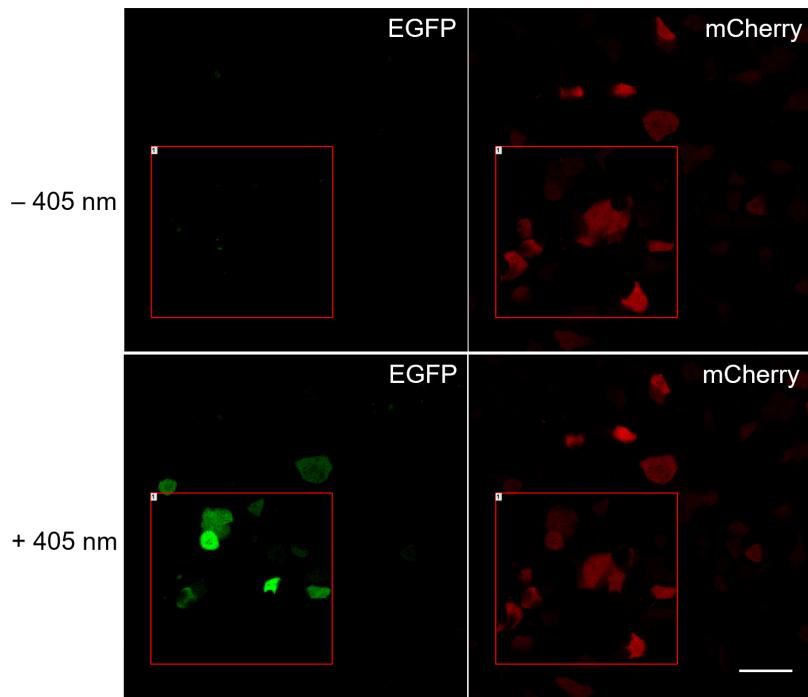
In order to compare the efficiency of photolysis of **PCK**, **NPPK**, and **NDBFK** *in vivo*, HEK 293T cells were co-transfected with pGL3\_K206TAG and the corresponding pMbPylRS/PyltRNA<sub>CUA</sub> plasmids in the absence and presence of **PCK** (1 mM), **NPPK** (1 mM), and **NDBFK** (0.25 mM). After 24 hr incubation, the cells were either irradiated for 4 min (365 nm, 25 W) or kept in the dark, followed by a Bright Glo luciferase assay. In the absence of UV light, low firefly luciferase activity was detected in cells treated with **PCK**, **NPPK**, and **NDBFK**, which indicates that the caged luciferase was inactive. However, upon irradiation, >100-fold FLuc increased activity was observed using **NDBFK** (Figure 2.16). As expected, in the absence of the UAA, only minimal background FLuc activity was detected, as no functional FLuc was expressed (Figure 2.16). Compared to **PCK**, decaging of **NPPK** and **NDBFK** with UV light resulted in a 50-fold and 107-fold increase in FLuc activity, respectively, whereas a 55-fold increase in FLuc activity was observed with **PCK** (Figure 2.16). This result is in agreement with literature reports that the NDBF caging group is more efficient at photolysis than other *o*-NB caged lysines.<sup>77</sup>



**Figure 2.16:** Bright-Glo luciferase assay for decaging of **PCK**, **NPPK**, and **NDBFK** *in vivo*.

After UV exposure, the FLuc activities in the presence of **PCK**, **NPPK**, and **NDBFK** increased by 55-fold, 50-fold, and 107-fold, respectively, relative to –UV control. Error bars represent standard deviations from three independent experiments.

In addition to controlling enzymatic function with light, optical triggering of EGFP folding (see section 3.3.2) was performed as a proof-of-principle. HEK293T cells were co-transfected with pEGPF-K85TAG-mCherry and pMbNDBFKRS-4PyIT in the presence of **NDBFK** (0.25 mM). After 24 h incubation, only mCherry fluorescence was observed in the presence of **NDBFK**. Since the mCherry serves as a transfection and expression control, it indicated that **NDBFK** was incorporated in mammalian cells. Upon 405 nm light irradiation (DMD stimulation for 30 s), time-lapse imaging of cellular fluorescence was performed for activation of EGFP for 90 min (Figure 2.17). A steady increase in EGFP fluorescence suggested protein folding and fluorophore maturation upon illumination, further verifying the stability and optical deprotection of **NDBFK** in HEK293T cells. Thus, genetically encoded **NDBFK** in conjunction with light activation allows for the conditional regulation of intracellular protein maturation in mammalian cells.

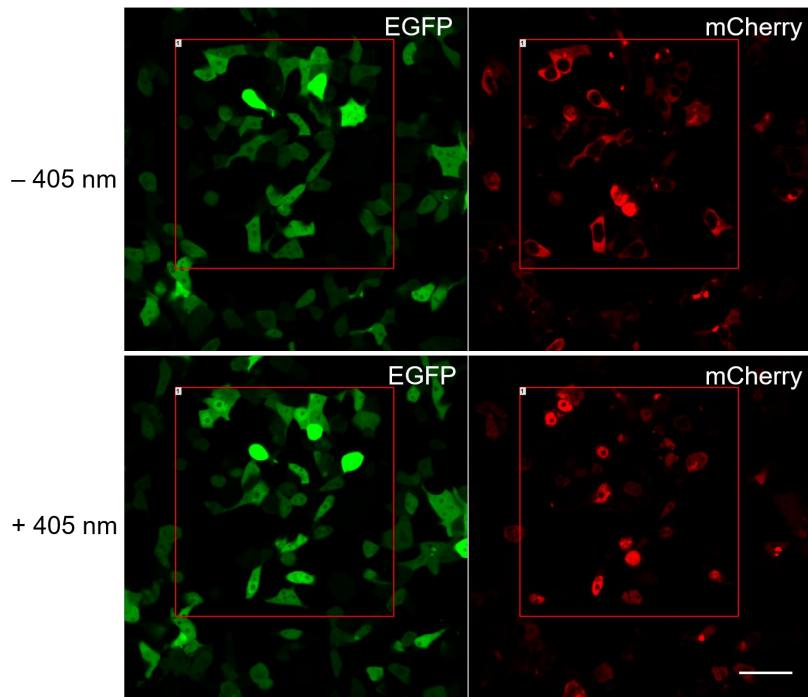


**Figure 2.17:** Light activation of EGFP-**NDBFK** in live cells.

Fluorescence imaging of HEK 293T cells expressing EGFP-K85TAG-mCherry, 90 min after irradiation by a 405 nm DMD system (30 s, in the red square) in the presence of **NDBFK** (0.25 mM). The fluorescent cells were imaged using Nikon A1R confocal microscope with 20× objective (2-fold zoom, fluorescein (Ex = 488 nm) and Texas Red (Ex = 560 nm) channels). Scale bar indicates 50  $\mu$ m.

In addition, the activation assay of SatB1 nuclear translocation (see 3.3.3) was also carried out to further showcase the applicability of **NDBFK** and efficient optical protein activation in mammalian cells. The presence of the NDBF group at position K29 completely inhibited translocation of an mCherry reporter as an NLS fusion. mCherry fluorescence localized in the cytoplasm was observed in HEK 293T cells expressing EGFP-SATB1\_K29→**NDBFK**-mCherry. After 24 hr of incubation, the cells were treated with 405 nm light (DMD stimulation for 30 s) and nuclear translocation of mCherry was triggered through decaging and formation of the native SATB1-NLS (Figure 2.18). Importantly, complete mCherry protein translocation from

the cytoplasm to the nucleus was observed, demonstrating complete **NDBFK-SATB1** decaging in live cells through the optical triggering of a caged SATB1-NLS.



**Figure 2.18:** Light activation of **NDBFK-SATB1-mCherry** in live cells.

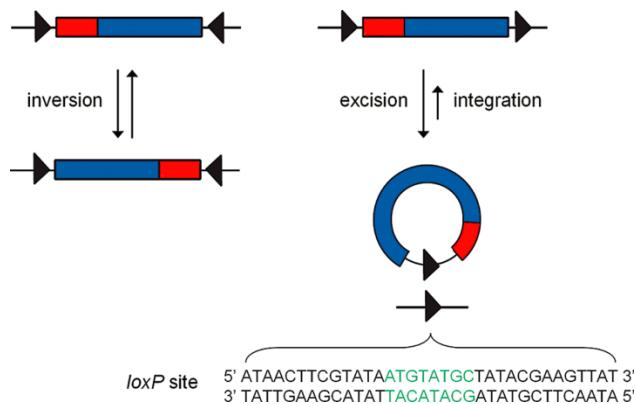
Fluorescence micrographs of HEK293T cells showing no nuclear localization of **NDBFK-SATB1-mCherry** and complete activation of translocation from the cytoplasm to the nucleus after 405 nm light activation. EGFP expression was used as a transfection and imaging control. The nuclear translocation occurred only in irradiated cells within the red square. The cells were imaged using a Nikon A1 confocal microscope with a Plan Apochromat 40 $\times$ oil objective (fluorescein ( $E_x = 488$  nm) and Texas Red ( $E_x = 560$  nm) channels). Scale bar indicates 50  $\mu$ m.

## 2.3 OPTICAL CONTROL OF CRE RECOMBINASE

### 2.3.1 Introduction to DNA recombination

This material was reprinted, in part, with permission from [Luo, J.; Arbely, E.; Zhang, J.; Chou, C.; Uprety, R.; Chin, J. W.; Deiters, A. \*Chem Commun\* 2016, 135\(36\), 13433-13439.](#)

DNA site-specific recombination is described as the DNA strand exchange between defined target sequences on each of two DNA segments.<sup>78</sup> Cre-loxP recombination is a special type of site-specific DNA recombination initially developed by Dr. Brian Sauer for activation of gene expression in mammalian cell lines.<sup>79</sup> Cre recombinase, a 38 kD protein from Bacteriophage P1<sup>80</sup> that recognizes *loxP* sites on dsDNA, is the most widely applied recombinase (Figure 2.19).<sup>81</sup> *loxP* sites consist of two 13-base pair repeats flanking an 8 base pair asymmetric sequence. Cre recombinase targets the *loxP* sites cleaves the dsDNA and either insert or excise sequences through a double recombination event that proceeds through a holiday junction.<sup>39</sup>



**Figure 2.19:** Representative scheme of Cre DNA recombination mechanism.<sup>64</sup>

Adapted with permission from Edwards *et al*, ACS Chem. Biol., 2009, 4 (6), 441–445. Copyright 2009 American Chemical Society.

DNA recombination is an important biological process that is commonly used for the activation and deactivation of gene expression in single cells and multicellular organisms and thus has been extensively engineered into a versatile genome manipulation tool.<sup>82</sup> The Cre-loxP recombination system has proven to be a robust and reliable site-specific recombination tool due to its efficient function in multiple organisms including *E. coli*,<sup>39</sup> yeast,<sup>63</sup> mice,<sup>57, 83-84</sup> zebrafish,<sup>85</sup> and drosophila.<sup>86</sup> Moreover, the Cre/loxP system is active on any type of DNA, such as linear, supercoiled, or circular,<sup>87</sup> and has been extensively applied in genetic engineering, enabling efficient conditional gene knock-out and knock-in for functional genetics studies.<sup>88</sup>

However, a limitation of the Cre/loxP recombination system is the lack of tight spatial and temporal control over Cre recombinase activity. Initial attempts to temporally regulate Cre expression used inducible promoters<sup>89-91</sup> and fusion proteins with ligand binding domains, such as the rapamycin inducible FKBP-FRB dimerization pair or the estrogen receptor (ER).<sup>21, 92</sup> However, triggering Cre function with small molecules limits the ability to achieve spatial control. In order to address this limitation, three different light-activation strategies were developed: 1) The two fragments of a split Cre recombinase were fused to cryptochrome 2

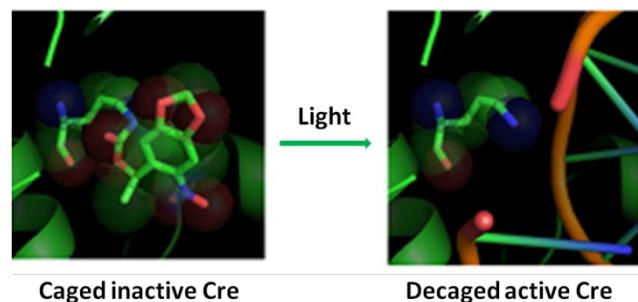
(CRY2) and *Arabidopsis* CIB1(cryptochrome-interacting basic-helix-loop-helix), and exposure to blue light (450 nm) induced dimerization of CIB1 and CRY2 and Cre activation.<sup>12</sup> However, no spatial control was reported and the frequency of recombination events was limited. 2) The Cre-ER system was expanded by introducing a photocaged tamoxifen or its analogs for photochemical control of DNA recombination.<sup>41, 93-95</sup> However, limited recombination activity was observed after UV exposure and the diffusible small molecule ligand could induce off-target effects. 3) The catalytically active Tyr324 in Cre recombinase was replaced with a photocaged *ortho*-nitrobenzyl tyrosine (**ONBY**) using unnatural amino acid mutagenesis in *E. coli*.<sup>96</sup> The caged Cre enzyme was expressed in *E. coli*, isolated and purified, transfected into mammalian cells.<sup>97</sup> This enabled spatio-temporal activation of DNA recombination in tissue culture, but the required enzyme isolation, purification, and protein transfection limit its application.

### 2.3.2 Optical control of DNA recombination

In order to address all limitations of the aforementioned approaches by developing a genetically encoded, tightly regulated, light-activated Cre recombinase through utilization of the *MbPylRS/tRNA<sub>CUA</sub>* pair,<sup>98</sup> which have been previously used to optically control transcription, nuclear localization, and kinase function.<sup>98-99</sup> This system allows for the site-specific incorporation of photocaged amino acids into proteins in mammalian cells, enabling the genetic encoding of light-activated processes. The methylated 6-nitropiperonylmethyl group in **PCK**, are more suitable to live cell applications than the *ortho*-nitrobenzyl group in **ONBY**. Substitutions on the aromatic chromophore in **PCK** result in a bathochromic shift of the effective irradiation wavelength (>350 nm) and consequently lower phototoxicity. In addition, methyl substitution at the benzylic position improves the biocompatibility of the protecting group since it transforms

the photolysis product from an aldehyde to a less toxic ketone. In addition, **PCK** shows improved solubility compared to **ONBY** in cell culture media.

In order to generate a light-activated Cre recombinase in a eukaryotic system, the photocaged lysine (**PCK**) was incorporated at position K201 within the active site of Cre recombinase. This lysine was selected as a target for caging group introduction, since it is highly conserved in almost all of the 80 Int family recombinases<sup>99</sup> and since it is essential for catalytic activity as shown by the enzymatic inactivity of the K201A mutant.<sup>98, 100</sup> K201 is located in the loop connecting β-2 and β-3 and is directed into the minor groove near the cleavage site of the DNA, undergoing interactions with the N3 of the +1 guanine base, 5'-O and 4'-O of the -1 sugar, and a tightly bound water molecule within hydrogen bonding distance (Figure 2.20).<sup>101</sup>

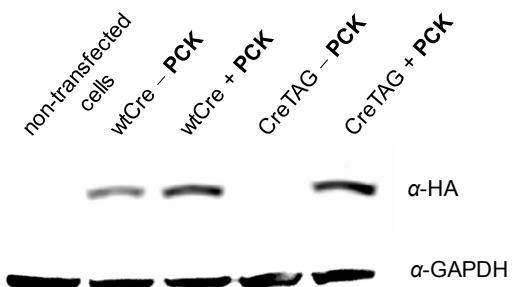


**Figure 2.20:** Photo-activation of caged Cre recombinase.

Before light trigger, caging of K201 with an *o*-nitropiperonyl group in the Cre active site inhibits the enzyme activity by blocking DNA cleavage. After light irradiation, the caging group is removed and Cre recombinase is activated. PDB: 1CRX.

First, a nuclear targeted and affinity tagged Cre recombinase was engineered through amplification of the Cre gene from pET21-Cre<sup>102</sup> and cloning into the pCS2-HA-NLS backbone, creating pCS2-HA-NLS-Cre.<sup>103</sup> In order to improve the expression level of the Cre gene, the

HA-NLS-Cre construct aforementioned was inserted into the CMV promoter-driven coding region of the pCKRS vector, which contains the tRNA synthetase evolved for **PCK**, generating pCKRS-HA-NLS-Cre. An amber stop codon, TAG, was then introduced at position K201, creating pCKRS-HA-NLS-CreK201TAG. The pCKRS-HA-NLS-Cre or pCKRS-HA-NLS-CreK201TAG plasmid was co-transfected with the p4CMVE-U6-PylT plasmid into HEK 293T cells in the absence or presence of **PCK** (1 mM) and a Western blot was performed (Figure 2.21). A band corresponding to a 39 kDa HA-tagged Cre recombinase was detected for wild-type expression (pNLS-CreWT) in the absence and presence of **PCK** (lanes 2 and 3). While in the case of caged Cre recombinase expression (pNLS-CreK201TAG), the same band was only observed in the presence but not the absence of **PCK** (lanes 4 and 5). This demonstrates that **PCK** was successfully incorporated into the mutant Cre protein with high fidelity.

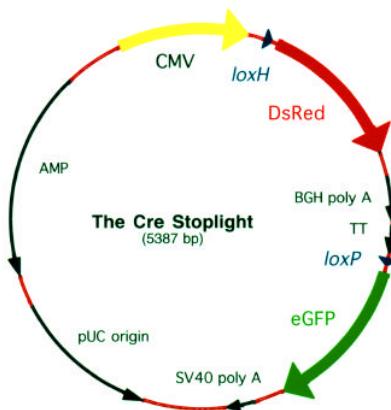


**Figure 2.21:** Western blots for incorporation of **PCK** into Cre recombinase.

Anti-HA Western blot showing amino acid dependent expression of caged Cre recombinase in HEK293T cells in the absence or presence of **PCK** (1.0 mM). Expression levels of CreK201PCK are similar to wild-type Cre recombinase expression.

Cre Stoplight (pC-SL) encodes the DsRed gene followed by a transcription terminator and the EGFP reporter (Figure 2.22).<sup>104</sup> If cells are transiently transfected with pC-SL, in the

absence of Cre, only DsRed is expressed; however, in the presence of Cre, DsRed is excised through DNA recombination, removing the DsRed gene and the stop codon, resulting in the expression of EGFP. Therefore, DsRed expression serves as a transfection control without Cre expression and EGFP expression is a reporter for Cre activity.

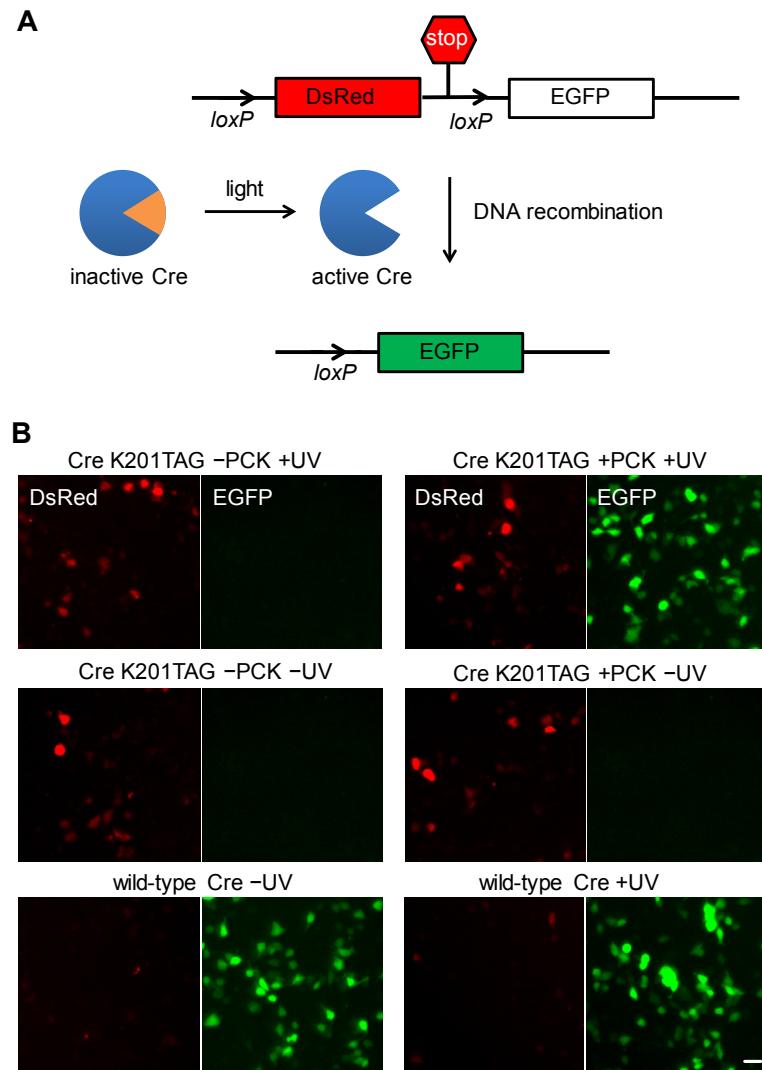


**Figure 2.22:** Cre Stoplight plasmid (pC-SL).

Image adapted with permission from *BioTechniques* 2001, 31 (5), 1036–1041.

To demonstrate photoactivation of K201-caged Cre recombinase in live cells, HEK293T cells were co-transfected with the above described expression system and the pC-SL reporter plasmid. As expected, the expression of wild-type Cre recombinase led to intracellular recombination and observation of EGFP expression (Figure 2.22). As a negative control, the expression of caged Cre in the absence of PCK showed exclusive DsRed expression, indicating that no functional Cre recombinase was generated. This is further supported by Western blot analysis (Figure 2.23). Importantly, cells transfected with pCKRS-HA-NLS-CreTAG in the presence of PCK but in the absence of UV irradiation exclusively showed DsRed expression, verifying the complete catalytic inactivity of the caged Cre recombinase containing the K201→PCK mutation. The protein was activated through a brief UV exposure (4 min, 365

nm), which induced decaging and subsequent Cre-catalyzed DNA recombination, activating EGFP expression. Residual DsRed fluorescence was still observed, due to reporter expression during the 24 hr incubation before light activation.

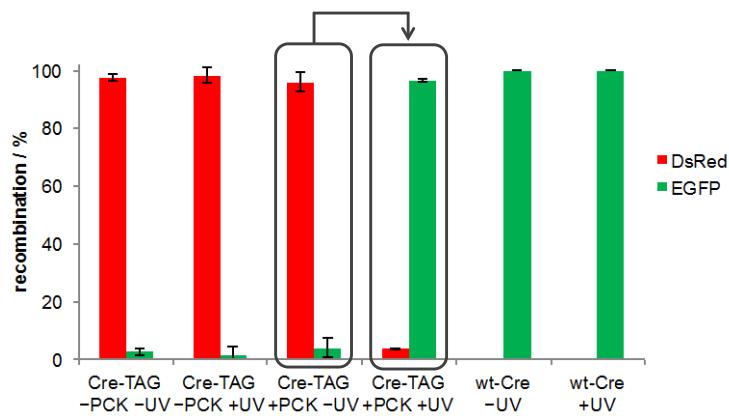


**Figure 2.23:** Optical control of caged Cre recombinase in live cells.

A) The Cre-Stoplight reporter encodes for DsRed and transcription termination signal, flanked by *loxP* sites and followed by a GFP gene. Light activation of caged Cre recombinase results in Cre-mediated recombination, and turns on the expression of GFP. B) Fluorescence microscopy images of HEK293 cells expressing the Cre-Stoplight

reporter and the caged Cre recombinase, CreK201PCK. The caged enzymes are completely inactive until UV exposure (365 nm) triggers enzymatic activity and DNA recombination. Scale bar represents 100  $\mu$ m.

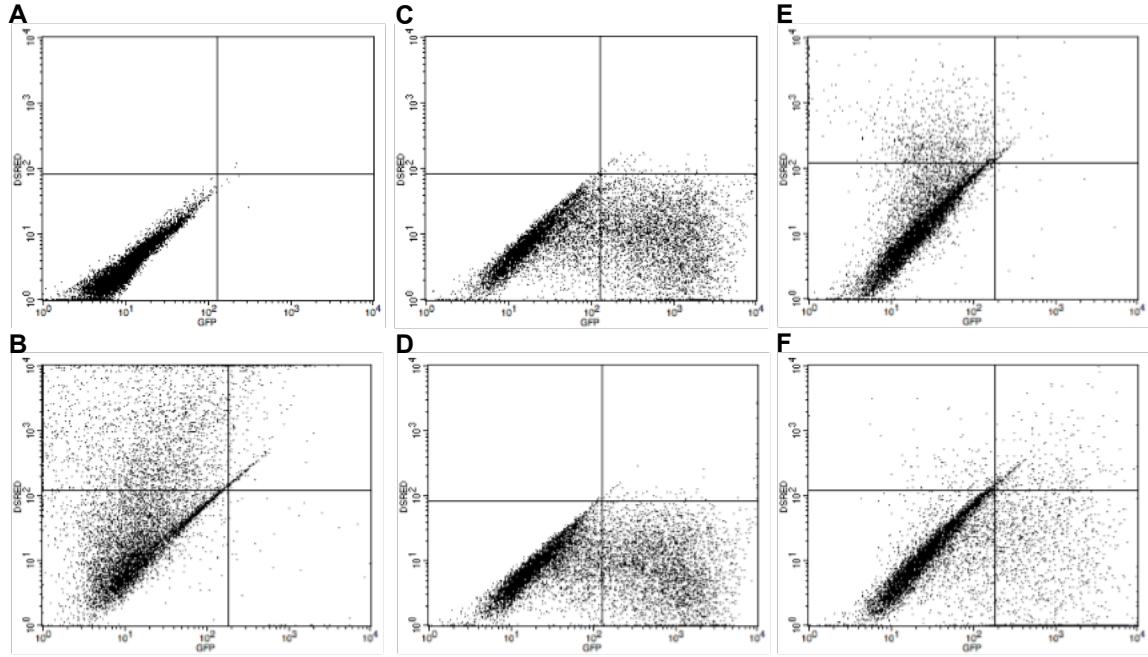
In order to quantify the observed recombination activity of caged Cre recombinase in the absence or presence of UV exposure, the number of DsRed- and GFP-positive cells were counted in six randomly selected areas for each well (Figure 2.24). The results showed no DNA recombination activity of the caged enzyme in the absence of UV exposure and an optical activation that almost reaches the activity of the expressed wild-type enzyme. To further verify the micrograph results, a flow cytometry analysis was conducted (Figure 2.24). The results confirmed the imaging analysis and showed a high level of optically triggered DNA recombination activity (80% of wild-type levels) was observed.



**Figure 2.24:** Quantification of fluorescent cells.

Fluorescent cells from the light-activation of CreK201PCK (Figure 2.23) were counted in both the DsRed and EGFP channel over six randomly selected fields of view per well and counts were normalized to the total number of fluorescent cells (Nikon HCA software). Error bars represent standard deviations from three independent experiments.

In addition, fluorescence-activated cell sorting (FACS) provides a quantitative analysis of the fluorescent signals generated by the Cre Stoplight reporter. FACS was used to further confirm the observed recombination activity of **Cre-PCK**. The cells in each sample were analyzed for DsRed and EGFP expression (DsRed-GFP plot) on a FACS Calibur instrument. As expected, in negative controls, no fluorescence was shown in the non-transfected cells (Figure 2.25A), and exhibited a high level of red fluorescence was detected in the cells transfected with pC-SL only (Figure 2.25B). The cells co-transfected with pC-SL and pCreWT showed the consistent level of GFP expression in the absence and presence of UV treatment (Figure 2.25C and D), indicating that UV irradiation did not have effect on DNA recombination. In the caged Cre sample, only DsRed expression was detected in the absence of UV trigger (Figure 2.25E), however, Upon UV irradiation, only EGFP expression was observed (Figure 2.25F). It was notable that the proportion of DsRed cells before UV irradiation (Figure 2.25E) was less than the proportion of EGFP expressing cells after UV irradiation (Figure 2.25F), since DsRed was expressed during the first 24 hr incubation before light activation (Figure 2.25F). These results are similar to the caged Cre data obtained above (Figure 2.23 and Figure 2.24). Analysis was performed on a FACSCalibur (Becton-Dickinson) instrument, using a 488 nm excitation laser with a 530 nm band pass filter (GFP) and a 633 nm excitation argon laser and 661 nm band pass filter (DsRed)

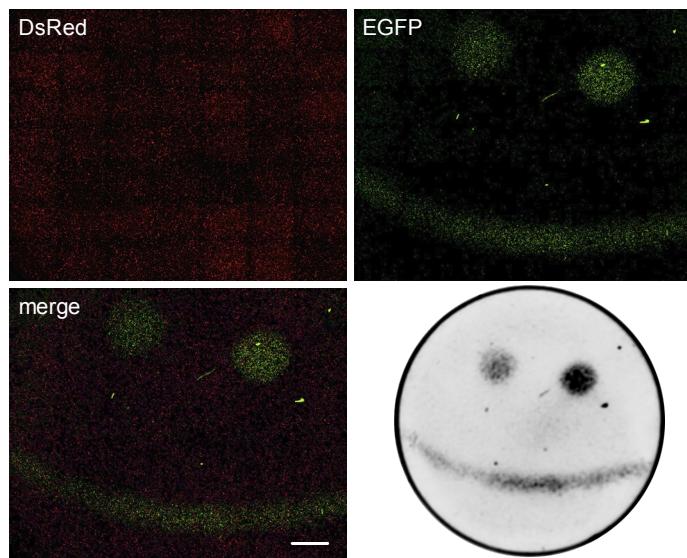


**Figure 2.25:** Flow cytometry analysis of light activation of non-caged and caged Cre recombinase in HEK 293T cells.

The X-axis presents the intensity of EGFP fluorescence and the Y-axis present the intensity of DsRed fluorescence. A) (-) control: non-transfected cells; B) pC-SL only; C) pCreWT + pC-SL – UV; D) pCreWT + pC-SL + UV; E) pCreKTAG + pC-SL + **PCK** – UV; F) pCreKTAG + pC-SL + **PCK** + UV. Proportion of red (upper left quadrant, E) to green fluorescent cells (lower right quadrant, F) is shifted after Cre recombinase is activated *in vivo*.

In order to demonstrate spatial control over DNA recombination in mammalian cells, HEK293T cells expressing both the caged Cre recombinase and the pC-SL reporter were exposed to patterned irradiation. After light exposure, the cells were incubated for 24 h and fluorescence in live cells was imaged (Figure 2.24). DsRed fluorescence was detected in all transfected cells but EGFP fluorescence was only detected in the areas exposed to localized irradiation showing tight spatial control over the DNA recombination event. In addition, fine-tuning of the DNA recombination in response to different exposure times was demonstrated by irradiating certain areas for different durations (in the same experiment and the same plate).

Longer UV treatment (5 min) led to a greater Cre activation and increased EGFP expression (see left versus right eye in Figure 2.26).



**Figure 2.26:** Spatial control of DNA recombination in live cells.

Spatial control of DNA recombination through patterning of a smiley face onto a layer of HEK293T cells in a 35 mm imaging  $\mu$ -dish (observation area 21 mm, ibidi) expressing caged Cre recombinase and the Cre-Stoplight reporter. The left eye and mouth area were exposed to 365 nm light for 3.5 min, while the right eye was exposed for 5 min, which led to increased recombination activity. DsRed and EGFP channels were imaged through tiling (7 $\times$ 7, 4 $\times$  objective, Nikon A1R microscope). Scale bar represents 2 mm. For improved contrast, fluorescence was also imaged using a Bio-Rad ChemiDoc system.

## 2.4 CONCLUSIONS

In conclusion, new single-photon (**NPPK**) and two-photon (**NDBFK**) nitrobenzyl-caged lysine analogues have been incorporated into proteins in mammalian cells using engineered *MbPylRS*

synthetases. As a proof-of-principle, optical control of firefly luciferase was achieved in live cells by caging a key lysine residue with **NPPK** and **NDBFK**, and excellent OFF to ON light-switching ratios were observed. Furthermore, **NDBFK** was applied to photocontrol of EGFP maturation and SatB1 nuclear translocation with a high spatiotemporal resolution. Also, decaging of **NDBFK** was readily achieved with blue light of 405 nm, while a previously encoded *ortho*-nitrobenzyl-caged lysine requires UV activation. Taken together, these results demonstrate that **NPPK** and **NDBFK** are new and valuable types of optical probes that can potentially be used for the investigation and regulation of protein function and cellular processes in live cells.

In addition, a genetically encoded, light-activated Cre recombinase has been engineered through site-specific incorporation of **PCK** in mammalian cells. The activity of the enzyme can be stringently regulated both spatially and temporally through the use of a light-removable caging group installed directly on the essential lysine residue K201. The overall efficiency of light-activation of recombination was improved over other photoresponsive Cre systems and enabled the spatial control of DNA recombination.<sup>105-106</sup> By applying an engineered pyrrolysyl tRNA synthetase/tRNA system for the genetic encoding of a photocaged lysine, the developed Cre recombinase system can be easily adapted to other eukaryotic cells and multicellular model organisms.<sup>105</sup> The highly conserved lysine site can also be found in large numbers of other Int family proteins, such as FLP recombinase, XerD resolvase,  $\lambda$  phage integrase, and transposase, further expanding this approach in future applications.<sup>107</sup>

## 2.5 EXPERIMENTAL

### **Screening optimal synthetases for incorporation of NPPK and NDBFK using mCherry-EGFP assay in mammalian cells.**

See Section 9.1.3 for specific information on cell transfection and epifluorescent imaging in 96-well plates. HEK 293T cells were seeded at ~50,000 cells per well and grown in Dulbecco's Modified Eagle's Medium (200 µL, DMEM, Gibco) supplemented with FBS (Sigma, 10%), Pen-Strep (Corning Cellgro, 1%) and L-glutamine (Alfa Aesar, 2 mM) in 96-well plates (Greiner) under a humidified atmosphere containing CO<sub>2</sub> (5%) at 37 °C. Cells were transiently transfected with pM<sub>b</sub>PylRS(V1-V12)-mCherry-TAG-EGFP-HA and p4CMVE-U6-PylT (25 ng of each plasmid) at ~80% confluence using linear polyethylenimine (LPEI; 1.5 µL, 0.323 mg/mL) in DMEM (200 µL) without or with NPPK and NDBFK (0.25, 0.5, 1.0 mM). After a 24 h incubation at 37 °C, the media were replaced with PBS and the cells were imaged with a Zeiss Axio Observer Z1 Microscope (10× objective) using EGFP (38HE: Ex 470/40; Em 525/50) and mCherry (43HE: Ex 550/25; Em 605/70) filter cubes.

### **Incorporation of NDBFK into sfGFP in *E. coli*.**

See Section 9.1.1 for specific information on UAA incorporation in *E. coli* and protein purification. The plasmid pBAD-sfGFP-Y151TAG-pylT<sup>108</sup> was co-transformed with pBK-BHCKRS into chemically competent *E. coli* Top10 cells. A single colony was grown in LB media overnight and 300 µL of the overnight culture was added to 20 mL of LB media

supplemented with 1 mM of **NDBFK** and 25 µg/mL of tetracycline and 50 µg/mL of kanamycin. Cells were grown at 37 °C, 250 rpm, and protein expression was induced with 0.1% arabinose when OD<sub>600</sub> reached ~0.5. After overnight expression at 37 °C, cells were harvested and washed with PBS. Cell pellets were re-suspended in 5 mL of phosphate lysis buffer (50 mM, pH 8.0), Triton X-100 (60 µL, 10%) and protease inhibitor (5 µL, Sigma), gently mixed, and incubated on ice for 1 h. Cell mixtures were sonicated (Fisher Scientific 550 Sonic Dismembrator) with six short bursts of 30 s on ice followed by 30 s intervals for cooling, and then the cell lysates were centrifuged at 4 °C and 13,000 g for 10 min. The supernatant was transferred to a 15 mL conical tube and 100 µL Ni-NTA resin (Qiagen) was added. The mixture was incubated at 4 °C for 2 h under mild shaking. The resin was then collected by centrifugation (1,000 g, 10 min) at 4 °C, washed three times with 300 µL of lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, pH 8.0), followed by two washes with 300 µL of wash buffer (lysis buffer containing 20 mM imidazole). The protein was eluted twice with 300 µL of elution buffer (lysis buffer containing 250 mM imidazole). The purified proteins were analyzed by 12% SDS-PAGE (60 V for 20 min, and then 150 V for 1 h), and stained with Coomassie Blue. The protein mass was obtained by electrospray ionization mass spectrometry.

### **Light activation of EGFP folding and SatB1 translocation.**

See Section 9.1.2 for specific information on cell transfections in 8 well chamber slide. The 8-well chamber slide (Lab-Tek) was first coated using poly-D-lysine (0.1 mg/mL, Sigma), and then HEK293T cells were plated at ~50,000 cells per well (400 µL). At ~75% confluency, cells were co-transfected with pEGFP-K85TAG-mCherry or pEGFP-K29TAG-SatB1-mCherry and pBHCKRS-4PyIT (200 ng of each plasmid) using linear PEI (3 µL, 0.323 mg/mL). After 20 hrs

of incubation at 37 °C and 5% CO<sub>2</sub> in DMEM with 10% FBS in the presence of **NDBFK** (0.25 mM), the cells were washed once with phenol-red-free DMEM (200 µl), followed by an hour of incubation to remove any non-incorporated **NDBFK**. Before light activation, the cells were focused using the Texas Redchannel, and imaged with a Nikon A1 confocal microscope (20×objective, 2× zoom, fluorescein (Ex = 488 nm) and Texas Red (Ex = 560 nm) channels). Subsequently, the cells were exposed to by 405 nm light (DMD) for 30 s, and then EGFP and mCherry fluorescence were imaged every five minutes for two hours (scan resolution 512 × 512, 2× scan zoom, dwell time 1.9 ms).

### **Western blots.**

See Section 9.1.4 and 9.1.5 for specific information on protein analysis by Western blots. HEK 293T cells were passaged into 6-well plates, and grown to 80% confluence. The cells were then transfected with the pCKRS\_NLS\_HA\_Cre or pCKRS\_NLS\_HA\_CreKTAG and p4CMVE-U6-PyLT (2 µg each plasmid per well; plasmids were constructed by Dr. Chungjung Chou and Jie Zhang in the Deiters lab) using linear polyethylenimine (PEI) transfection reagent in antibiotic free DMEM medium (Invitrogen) with 10% FBS. The transfection mixture was incubated at 37 °C overnight. The cells were gently washed with ice-cold PBS and lysed with NE-PER nuclear protein extraction kit (Thermo Scientific) following the manufacturer's protocol. The protein extraction was analyzed on a 10% SDS-PAGE and transferred to a PVDF membrane (80 V, 1.5 h). The membrane was incubated in blocking buffer (5% nonfat milk/TBST (0.2% Tween 20, 20 mM Tris-HCl, 150 mM NaCl, pH 7.4) for 30 min and washed three times with TBST buffer. The membrane was then incubated with the primary antibody, α-HA-probe (Y-11) rabbit polyclonal IgG (sc-805, Santa Cruz Biotech), overnight at 4 °C, rinsed and incubated with a

fluorescent secondary antibody, goat- $\alpha$ -rabbit IgG-Cy3 antibody (GE Healthcare), for 1 h at room temperature. Cy3 imaging was performed with a Typhoon FLA7000 scanner(GE Healthcare). The other blots as a loading control were probed and incubated with the primary GAPDH antibody (G-9) mouse monoclonal IgG<sub>1</sub> (sc-365062, Santa Cruz Biotech) overnight at 4 °C, followed by a goat anti-mouse IgG-HRP secondary antibody (sc-2031, Santa Cruz Biotech) for 1 h at room temperature. The blots were washed by TBST for three times, and then incubated in working solution (Thermo Scientific SuperSignal West Pico Chemiluminescent substrate) for 5 min. The blots were imaged by ChemiDoc (BioRad).

### **Photoactivation of CreK201PCK.**

See Section 9.1.3 for specific information on cell transfection and epifluorescent imaging in 96-well plates. HEK 293T cells were plated in a 96-well cell culture (Greiner) microplate and grown to 75% confluence in Dubelcco's modified Eagle's medium (DMEM) supplemented with 10% FBS and 1% penicillin/streptomycin at 37 °C in 5% CO<sub>2</sub> atmosphere. The cells were co-transfected with the pCKRS-Cre or pCKRS-CreK201TAG, p4CMVE-U6-PylT, and the pC-SL reporter plasmid (100 ng of each) using linear PEI (2  $\mu$ L per well) in the presence or absence of **PCK** (1 mM) overnight at 37 °C. The media was replaced with fresh DMEM without **PCK**, and cells were incubated for 2 h at 37 °C in 5% CO<sub>2</sub> atmosphere. Subsequently, the cells were exposed to 365 nm UV light for 4 min using a 365 nm UV lamp (high performance UV transilluminator, UVP, 25 W) or kept in the dark. Expression level of the Cre activation reporter was measured after 24 h incubation.

### **Spatial control of Cre activation.**

See Section 9.1.3 and 9.1.7 for specific information on cell transfection and irradiation. In order to demonstrate spatial control of light-activated Cre-**PCK** via locally controlled light irradiation in living cells, HEK293T cells expressing both the caged Cre recombinase and the pC-SL reporter in a 35 mm imaging  $\mu$ -dish (observation area 21 mm, ibidi) were exposed to UV irradiation in a “smiley face” pattern, followed by microscopy and ChemiDoc (Bio-Rad) imaging. Bright EGFP fluorescence was only observed for cells containing activated Cre recombinase in the patterned region. Non-irradiated cells were dark. All cells showed DsRed expression, as a control. In response to different UV exposure times, different EGFP fluorescence intensity was observed in right eye (5 min’s irradiation) and the left eye and mouth (3.5 min’s exposure), demonstrating tuning of Cre decaging and activation.

### **3.0 COUMARIN LYSINES AS OPTICAL PROBES**

This chapter will describe a new and valuable class of optical probes – coumarin lysine analogs and its biological applications in genetically encoded photocontrol systems. Importantly, the use of coumarin lysines for the investigation and regulation of protein structure, dynamics, function, and localization was demonstrated in live cells, allowing for optically controlled *in vivo* studies in a precise spatial and temporal fashion.

#### **3.1 INTRODUCTION TO COUMARIN LYSINES**

This material was reprinted, in part, with permission from [Luo, J.; Uprety, R.; Naro, Y.; Chou, C.; Nguyen, D. P.; Chin, J. W.; Deiters, A. J Am Chem Soc 2014, 136\(44\), 15551-15558.](#)

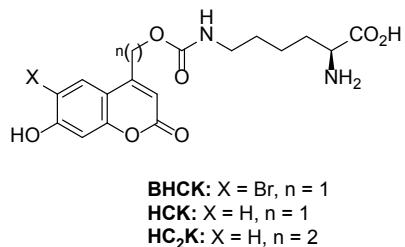
Good photochemical properties, chemical stability, and ease of synthesis make coumarins an important class of fluorescent probes for biological studies.<sup>109-111</sup> In addition to being versatile fluorophores, coumarin chromophores can be used as light-removable protecting groups, so called “caging groups”, that are photolysed through one- and two-photon irradiation.<sup>112</sup> Caged molecules have been extensively applied in the optical control of cellular processes.<sup>12, 14, 94, 113-114</sup>

In particular, the 6-bromo-7-hydroxycoumarinmethyl caging group undergoes fast two-photon photolysis at 740 nm and has been used to optically control neurotransmitters, secondary messengers, and oligonucleotides.<sup>115-116,117</sup> Two-photon irradiation enables optical activation of

biological processes with enhanced tissue penetration of up to one millimeter. Moreover, two-photon caging groups can be released with greater precision in 3-dimensional space than simple one-photon caging groups.<sup>112, 118</sup> Therefore, in our lab, a series of new fluorescent amino acids, unique hydroxyl- and amino-coumarin lysines, have been developed and genetically encoded in both *E. coli* and mammalian cells. In addition, the fluorescent subcellular localization of proteins and optochemical regulation of protein functions in mammalian cells is being studied.

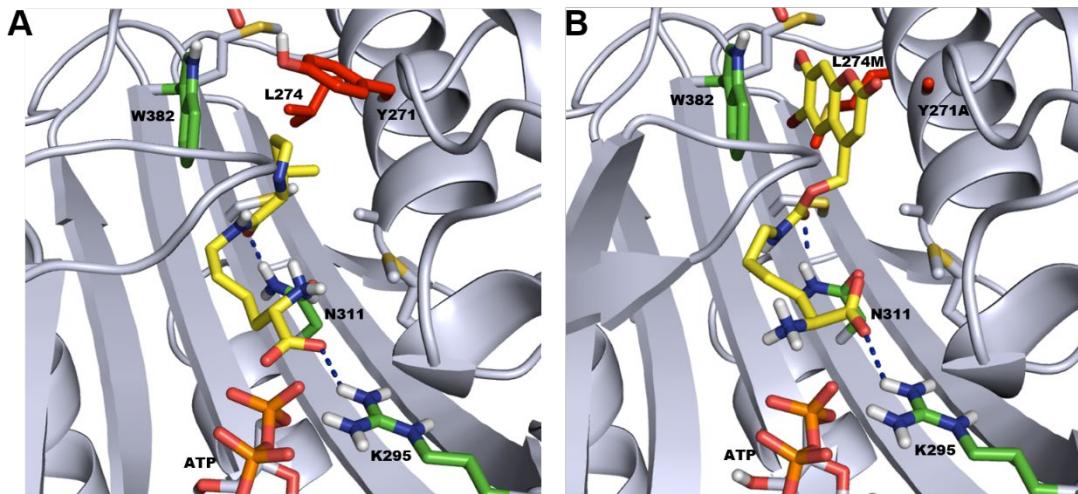
### 3.2 INCORPORATION OF COUMARIN LYSINES IN BACTERIAL AND MAMMALIAN CELLS

All three investigated coumarin lysines contain identical benzopyrone cores as fluorescent probes (Figure 3.1). However, subtle substitutions result in a set of coumarin derivatives with unique photochemical properties. Introduction of a bromine at the 6-position enables decaging not only with UV (single photon) light (in case of **HCK**), but also near IR (two-photon) excitation (in case of **BHCK**).<sup>116</sup> In contrast, extension of the coumarin-carbamate linker by a single carbon atom results in the coumarin lysine **HC<sub>2</sub>K**, which does not undergo photolysis and thus represents a stable coumarin amino acid probe. Thus, coumarin lysines **BHCK** and **HCK** can be used as both fluorescent and light-activated probes for optochemical control of protein function using UV or near-IR light, while the coumarin lysine **HC<sub>2</sub>K** may serve as a stable fluorescent probe that does not decage under UV excitation. All three compounds were synthesized by Dr. Rajendra Uprety in the Deiters lab.



**Figure 3.1:** Chemical structure of **BHCK**, **HCK**, and **HC<sub>2</sub>K**.

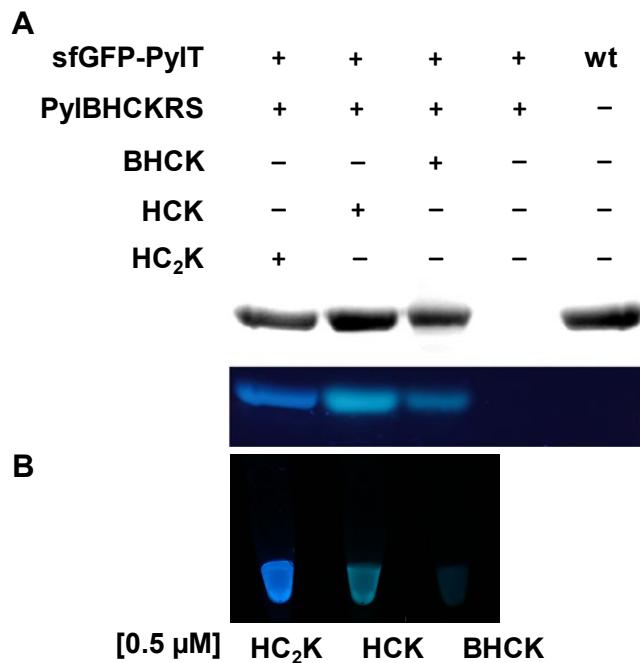
Here the PylRS/tRNA<sub>CUA</sub> pair was applied to the expansion of the genetic code of bacterial and mammalian cells with coumarin amino acids. Through screening of a synthetase panel of lysine analogues in mammalian cells and *E. coli*, the most active synthetase identified contains the mutations Y271A and L274M (Figure 3.2B) for incorporation of **BHCK** with respect to wild-type PylRS (Figure 3.2A).<sup>8, 119-120</sup> Compared to the pyrrolysine structure, the coumarin caging group is more bulky than the pyrroline ring. The five residues (A267, Y271, L274, N311, and C313) are crucial for the binding pocket (Figure 3.2A). The two mutations, Y271A and L274M, create a larger pocket for the coumarin ring (Figure 3.2B). This synthetase, named BHCKRS, enabled the site-specific incorporation of not only **BHCK**, but also **HCK** and **HC<sub>2</sub>K**, in response to the TAG amber codon into sfGFP-Y151TAG-6×His in *E. coli* (Figure 3.3).



**Figure 3.2:** Crystal structures of engineered PylRS binding pocket.

A) Crystal structure of wild type PylRS binding pocket with pyrrolysine (Pyl-AMP) substrate. B) 3D structure of BHCKRS binding pocket with **HCK** (HCK-AMP). PDB: 2Q7H.

SDS-PAGE analysis also revealed coumarin fluorescence of the expressed proteins containing **BHCK**, **HCK**, or **HC<sub>2</sub>K**. No fluorescence was observed for wt-sfGFP since its excitation wavelength does not match that of coumarin lysines and because of the denaturing conditions of the gel (Figure 3.3). The dependence of protein expression on the addition of coumarin lysines demonstrates that the identified evolved BHCKRS has a high specificity for coumarin lysines, as it does not incorporate any of the common 20 amino acids, and also allowed for selective bacterial expression of ubiquitin and myoglobin genes containing an amber codon at a defined site.

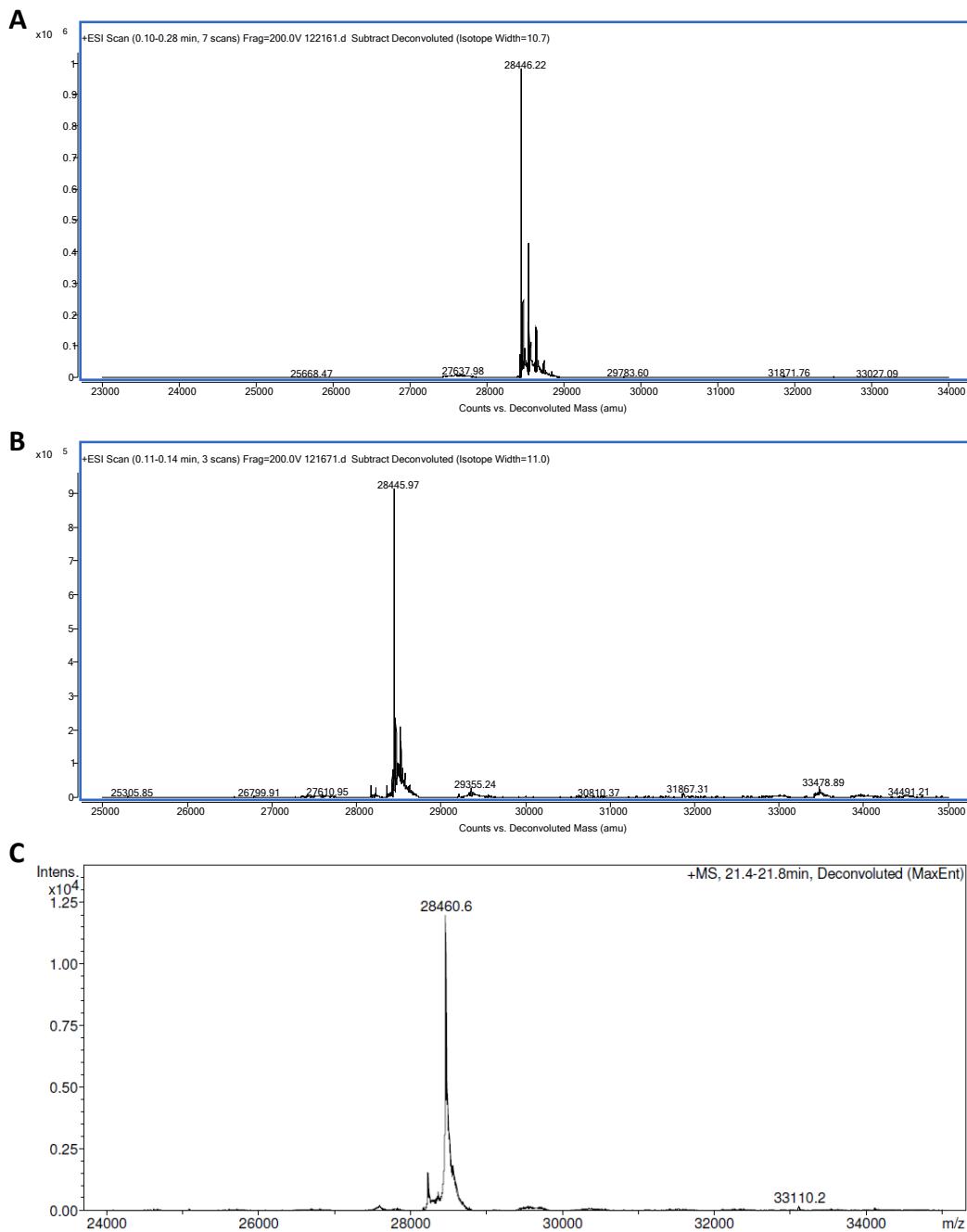


**Figure 3.3:** Genetic encoding of **BHCK**, **HCK**, and **HC<sub>2</sub>K**.

A) SDS-PAGE analysis of sfGFP-Y151TAG containing **BHCK**, **HCK**, and **HC<sub>2</sub>K** through incorporation in *E. coli*. The gel was stained with Coomassie blue (top) and fluorescence (bottom) was imaged via excitation at 365 nm, showing only amino acid fluorescence but not sfGFP fluorescence. (Protein yields: 1.6 mg/L, 8.0 mg/L, and 2.5 mg/L, respectively) B) Fluorescence of small molecule fluorophores **BHCK**, **HCK**, and **HC<sub>2</sub>K** (0.5  $\mu$ M).

Electrospray ionization mass spectrometry showed that recombinantly expressed sfGFP-**HCK** and **HC<sub>2</sub>K** had a mass of 28446.60 Da and 28460.60 Da, which was in agreement with the expected mass of 28446.03 Da and 28460.04 Da, respectively (Figure 3.4A, Figure 3.4C). ESI-MS analysis conclusively demonstrated that **HCK** and **HC<sub>2</sub>K** was incorporated into sfGFP. The observed mass of sfGFP-**BHCK** is 28445.97 Da (Figure 3.4B), while its expected mass containing sfGFP was 28524.91 Da. Interestingly, the difference between expected and observed masses of sfGFP-**BHCK** was 78.94 Da, equal to the mass of Br, 78.92 Da. This indicated that the bromide group was removed during *E. coli* expression, which agrees well with the discovery

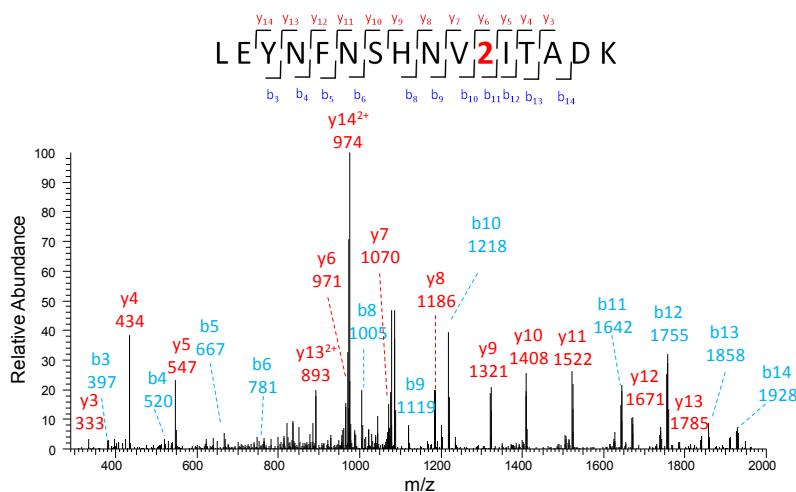
of *E. coli* reductive dehalogenation.<sup>121</sup> Overall, these results demonstrated that all three coumarin lysines could be incorporated into proteins in *E. coli* in good yield (8.0 mg/L, 1.6 mg/L, and 2.5 mg/L, respectively, for sfGFP) and with high specificity.



**Figure 3.4:** ESI-MS analysis of sfGFP-**HCK**, -**BHCK**, and -**HC<sub>2</sub>K**.

A) sfGFP-**HCK**: observed MS: 28446.22 Da, expected MS: 28446.03 Da; B) sfGFP-**BHCK**: observed MS: 28445.97 Da, expected MS: 28524.91 Da; C) sfGFP-**HC<sub>2</sub>K**: observed MS: 28460.60 Da, expected MS: 28460.04 Da.

In order to further investigate the incorporation of **BHCK** in *E. coli*, liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis of tryptic peptides was performed on sfGFP-**BHCK** (Figure 3.5). This result suggested that **BHCK** was site-specifically incorporated into sfGFP. In contrast, the mass of sfGFP-HCK (without bromine) was observed through ESI-MS (Figure 3.4B). Taken together, both analyses indicated a partial loss of bromine during *E. coli* expression, potentially due to bacterial reductive dehalogenation.<sup>121</sup>

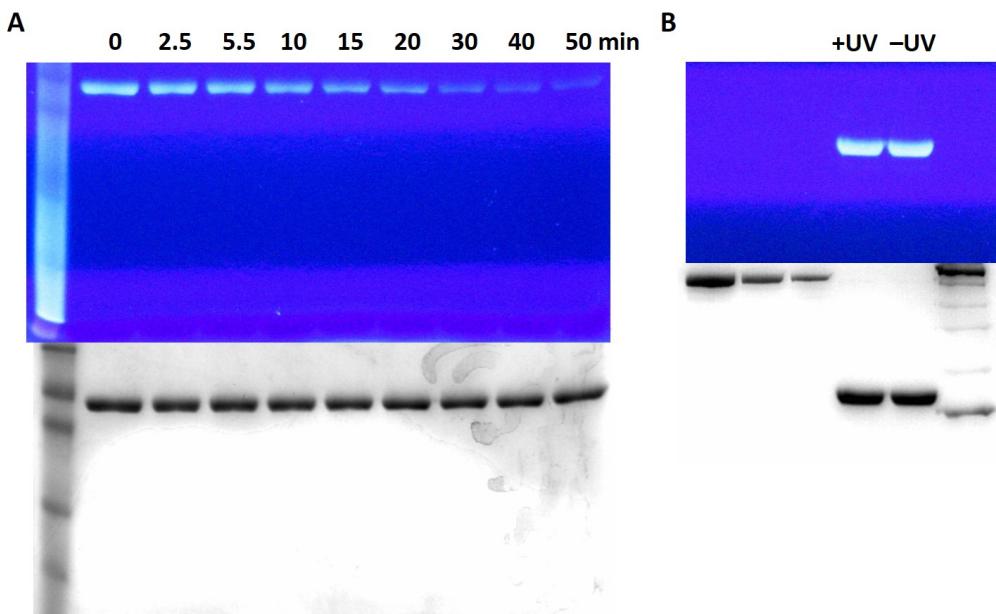


**Figure 3.5:** MS/MS analysis of sfGFP-**BHCK**.

sfGFP-6×His incorporating **BHCK** (**2**) expressed in Top10 competent cells was purified by Ni-NTA column for MS/MS analysis. The MS/MS fragmentation spectrum of a tryptic peptide from the purified protein confirms the incorporation of **BHCK** at the specific Y151 site.

The intrinsic fluorescence of **BHCK** and **HCK** once incorporated into a protein could potentially be used as an indicator of protein decaging through UV irradiation, as shown in Figure 3.6A. This enables the determination of protein expression, protein localization, and protein decaging using a single optochemical probe in a single experiment. In contrast, insertion

of an extra methylene unit fully abrogates photocleavage and thus establishes **HC<sub>2</sub>K** as a UV-stable amino acid for the site-specific fluorescent labeling of proteins (Figure 3.6B).

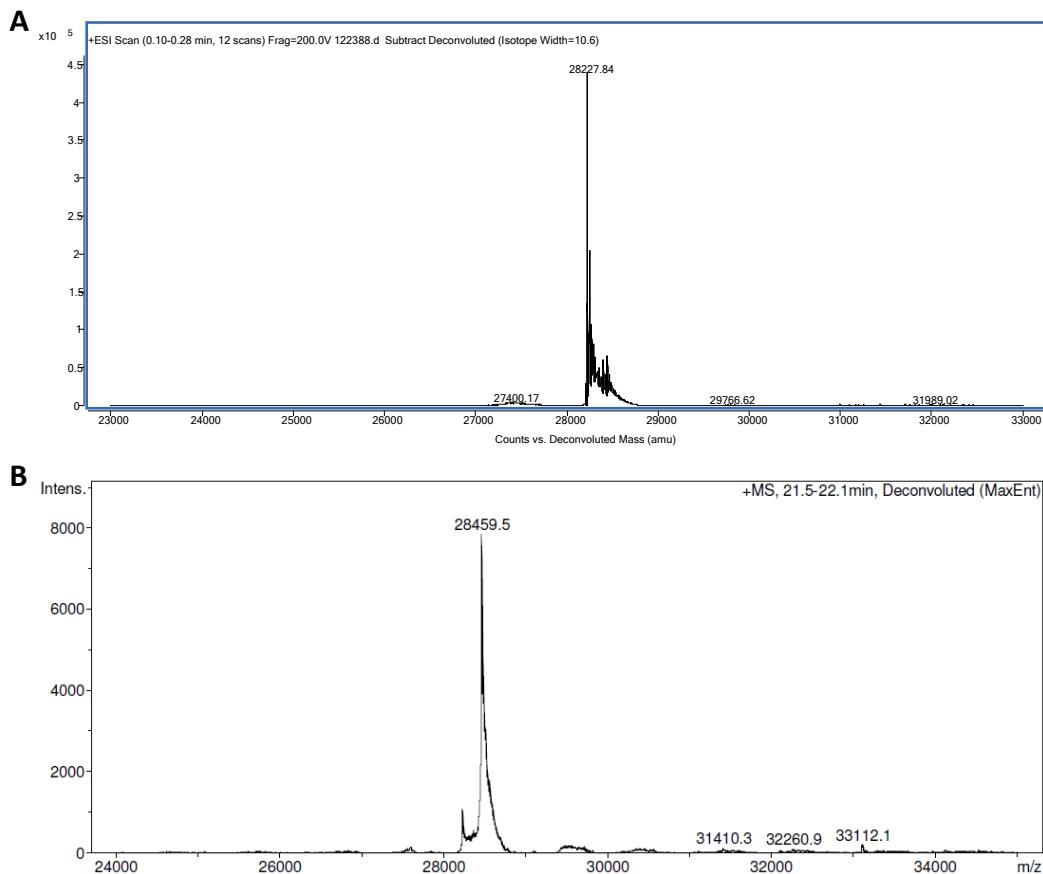


**Figure 3.6:** SDS-PAGE analysis and in-gel fluorescence imaging of sfGFP-**HCK** and sfGFP-**HC<sub>2</sub>K**.

A) SDS-PAGE analysis and photodecaging of sfGFP-**HCK**. Fluorescent imaging of the incorporated **HCK** through 365 nm UV excitation shows coumarin incorporation and decaging over time leading to diminishing fluorescence (top). Coomassie blue staining of the same gel indicates the same amount of protein in each lane (bottom). B) SDS-PAGE analysis of sfGFP-**HC<sub>2</sub>K** before and after UV exposure at 365 nm (20 min) indicates the stability of the incorporated amino acid toward photolysis. Fluorescence imaging at 365 nm excitation (top) and staining with Coomassie blue (bottom).

Furthermore, ESI-MS analyses were performed on the proteins irradiated by UV light. It showed that the decaged sfGFP-**HCK** had a mass of 28227.84 Da, in agreement with the expected mass of 28228.06 Da (Figure 3.7A), while the observed mass of the irradiated sfGFP-**HC<sub>2</sub>K** was 28459.50 Da (Figure 3.7B), corresponding to the mass of sfGFP-**HC<sub>2</sub>K**, 28460.60 Da

(Figure 3.4C). This further confirmed that **HCK** was a photocaging group, while **HC<sub>2</sub>K** acted as a UV-stable chromophore.

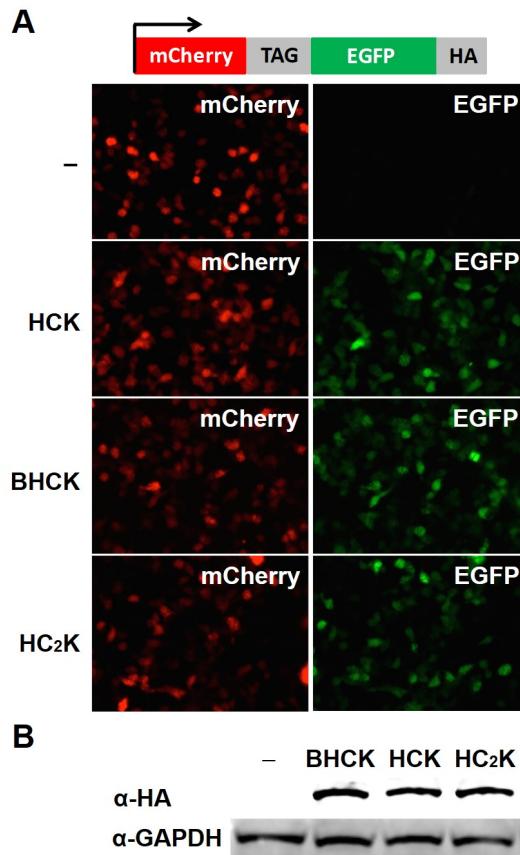


**Figure 3.7:** ESI-MS analysis of UV-irradiated sfGFP-**HCK** and sfGFP-**HC<sub>2</sub>K**.

A) decaged sfGFP-**HCK**: observed MS: 28227.84 Da, expected MS: 28228.06 Da; and B) irradiated sfGFP-**HC<sub>2</sub>K**: observed MS: 28459.50 Da, expected MS: 28460.04 Da.

To demonstrate the incorporation of three coumarin lysines into proteins in mammalian cells, pBHCKRS-mCherry-TAG-EGFP-HA and p4CMVE-U6-PylT were co-transfected into human embryonic kidney (HEK) 293T cells. Cells were incubated for 24 hr in the absence of any unnatural amino acid or in the presence of **BHCK**, **HCK**, and **HC<sub>2</sub>K** (0.25 mM). Fluorescence

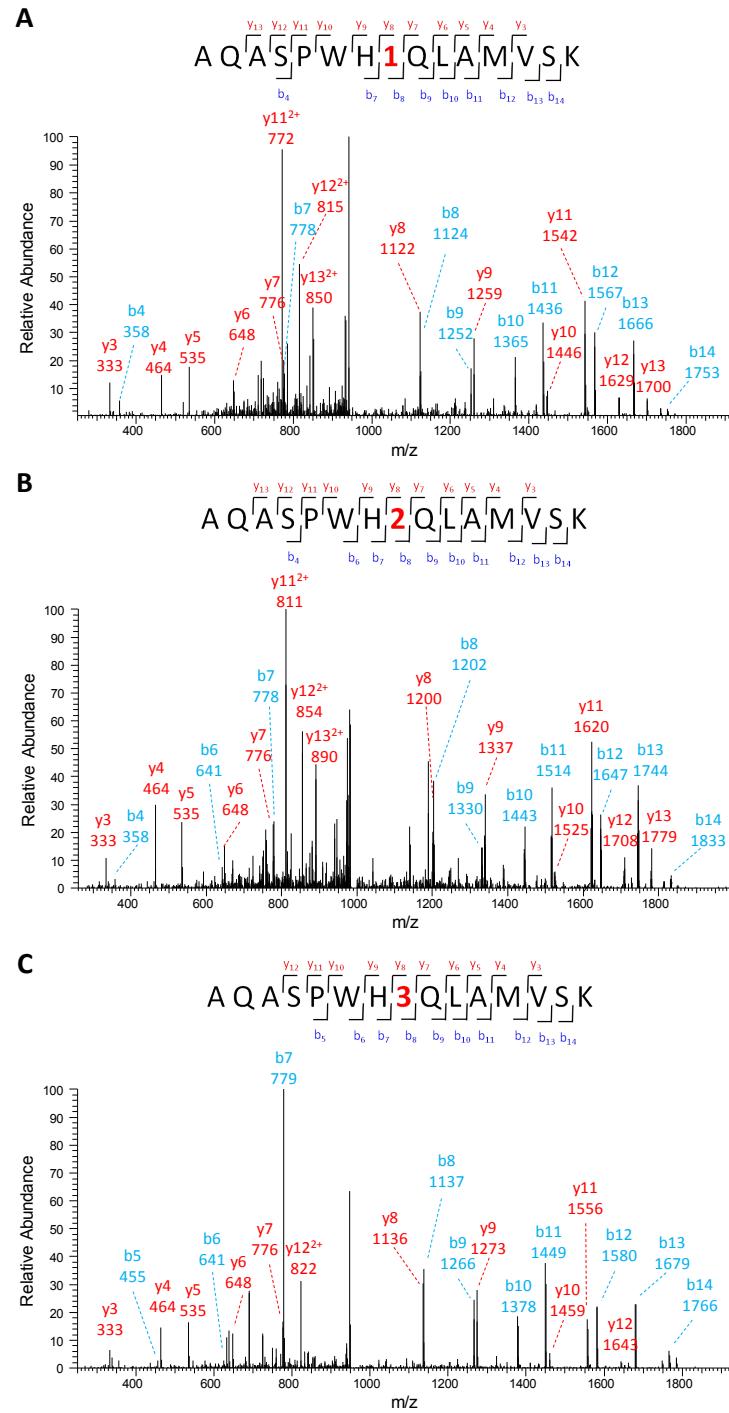
imaging revealed EGFP expression only in the presence of **BHCK**, **HCK**, and **HC<sub>2</sub>K**, indicating specific incorporation of the coumarin lysines without incorporation of any endogenous amino acid (Figure 3.8A). This was further confirmed by Western blot analysis of mammalian cell lysates from the same experiment with an anti-HA antibody (Figure 3.8B).



**Figure 3.8:** Genetic encoding of **HCK**, **BHCK**, and **HC<sub>2</sub>K** in mammalian cells.

A) Fluorescence micrographs of HEK293T cells co-expressing the BHCKRS/tRNA<sub>CUA</sub> pair and mCherry\_TAG\_EGFP\_HA in the presence or absence of **HCK**, **BHCK**, and **HC<sub>2</sub>K** (0.25 mM). After 24 h, cells were imaged using Zeiss epi-fluorescence microscopy via 10× objective with both EGFP (488 nm excitation) and mCherry (561 nm excitation) filter cubes. The mCherry channel is shown to the left and the EGFP channel is shown on the right. Scale bar indicates 50 μm. B) Western blot analysis of cell lysates using an anti-HA antibody and anti-GAPDH antibody as a loading control.

In addition, full-length mCherry-EGFP protein was immunoprecipitated from HEK 293T cells using an immobilized antibody against the HA-tag and mass spectrometry sequencing confirmed that coumarin lysines are site-specifically incorporated into protein in mammalian cells (Figure 3.9). Importantly, the presence of bromine was verified for protein containing **BHCK**, confirming the genetic encoding of the BHC-caged lysine. It demonstrated that a partial loss of bromine during *E. coli* expression is due to bacterial reductive dehalogenation, but it does not happen to mammalian cells.<sup>121</sup> These results demonstrate that **BHCK**, **HCK**, and **HC<sub>2</sub>K** can be selectively incorporated into proteins in mammalian cells in response to the amber codon (TAG).



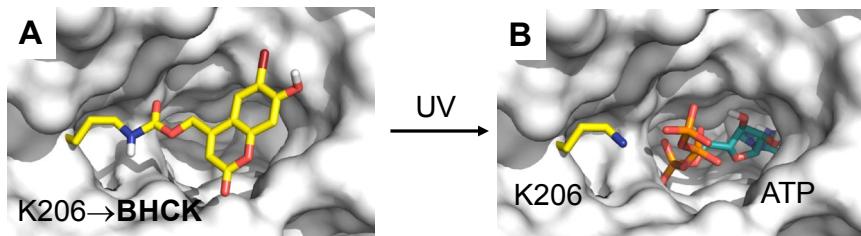
**Figure 3.9:** MS/MS analysis for incorporation of **HCK**, **BHCK**, and **HC<sub>2</sub>K** in mammalian cells.

A–C) mCherry-EGFP-HA containing **HCK** (1), **BHCK** (2), and **HC<sub>2</sub>K** (3) was expressed in HEK 293T cells and purified by anti-HA immunoprecipitation for MS/MS analysis. The MS/MS fragmentation spectrum of a tryptic peptide from the purified protein confirms the incorporation of **BHCK**, **HCK**, and **HC<sub>2</sub>K** at the expected site.

### **3.3 OPTICAL CONTROL OF PROTEIN FUNCTION**

#### **3.3.1 Optical control of firefly luciferase**

In order to apply the coumarin lysines in the optical control of protein function in live cells, firefly luciferase (Fluc) was selected as an initial target since bioluminescence measurements afford low background, high sensitivity, and easy quantification. Based on the Fluc crystal structure, a critical lysine residue, K206, was identified, which is positioned at the edge of the substrate-binding pocket (Figure 3.10B). It has been proposed that this residue stabilizes and orients ATP in the active site.<sup>68-69</sup> The  $\epsilon$ -amino group on K206 provides a hydrogen-bond interaction with the  $\gamma$ -phosphate of ATP and promotes the adenylation reaction with luciferin, thus being essential for catalytic activity as shown by the dramatic decrease in enzymatic activity displayed by the K206R mutant.<sup>68</sup> Therefore, we hypothesized that a sterically demanding coumarin caging group placed on K206 would prevent the interaction with ATP and limit the overall access of the substrates to the active site (Figure 3.10A). Photolysis of the coumarin lysine would remove the caging group and produce a native lysine residue, restoring the catalytic activity of the enzyme (Figure 3.10B). A genetically encoded photocaged lysine at K206 would enable the enhanced regulation of the catalytic activity of firefly luciferase via light activation.

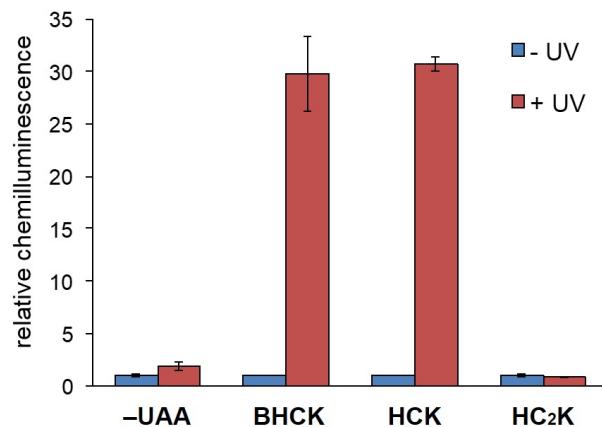


**Figure 3.10:** Engineering of an optochemically controlled firefly luciferase through unnatural amino acid mutagenesis.

A) Caging groups at position K206 are blocking access to the binding pocket by luciferin and ATP and are disrupting a required hydrogen bonding network. B) After decaging, wild-type Fluc is generated and the substrates can now enter the active site. PDB: 2D1S.

To determine the decaging of **BHCK**, **HCK**, and **HC<sub>2</sub>K** *in vivo*, the photocaged firefly luciferase activity quantification was applied to three coumarin lysines (Figure 3.11). Site-directed mutagenesis of the corresponding K206 residue to the amber codon (TAG) enabled incorporation of **BHCK**, **HCK**, and **HC<sub>2</sub>K** into firefly luciferase in mammalian cells. HEK 293T cells were co-transfected with the mutated firefly luciferase plasmid (pGL3-K206TAG) and pMbPylBHCKRS/PyltRNA<sub>CUA</sub> pair in the presence or absence of **BHCK**, **HCK**, and **HC<sub>2</sub>K** (0.25 mM). After a 24 hr incubation, the cells were either irradiated for 4 min (365 nm, 25W) or kept in the dark, followed by a Bright-Glo luciferase assay. The incorporation of **BHCK**, **HCK**, and **HC<sub>2</sub>K** in FLuc caused complete inhibition of luciferase activity before UV irradiation, comparable to the negative control (No UAA, Figure 3.11). This demonstrates that **BHCK**, **HCK**, and **HC<sub>2</sub>K** were capable of inhibiting catalysis of firefly luciferase. After UV irradiation, **BHCK** and **HCK** were decaged to produce native lysine, resulting in the activation of firefly luciferase. UV irradiation of FLuc-**BHCK**, -**HCK**, and -**HC<sub>2</sub>K** led to a 34-fold, 31-fold, and 0.8-fold increase in activity. These results demonstrate that FLuc-**BHCK** and -**HCK** can efficiently be activated *in vivo* and **HC<sub>2</sub>K** cannot be decaged as designed. The decaging efficiency (365 nm,

4 min) is similar between **BHCK** and **HCK**. From the chemiluminescence intensity, **HCK** shows higher activation than **BHCK**, indicating that the incorporation of **HCK** is higher than that of **BHCK**, which is consistent with the results of protein expression described previously. These results further determined the fidelity of incorporation of **BHCK**, **HCK**, and **HC<sub>2</sub>K** by the *MbPylBHCKRS/PyltRNA<sub>CUA</sub>* pair. Therefore, the activity of firefly luciferase can be tightly regulated in spatial and temporal resolution by incorporation of a coumarin lysine residue into the associated active site of the constitutively active luciferase protein.

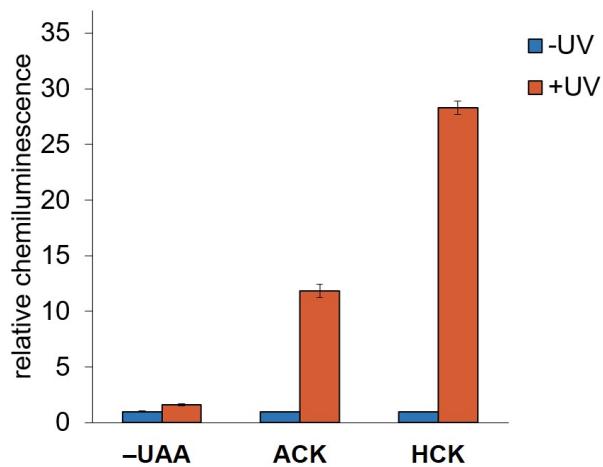


**Figure 3.11:** Bright-Glo luciferase assays of hydroxycoumarin lysine derivatives.

Bright-Glo luciferase assay of cells that were either kept in the dark or irradiated (365 nm, 4 min). Chemiluminescence was normalized to the -UAA/-UV control. No enzymatic activity was observed for the caged proteins and significant increases in luminescence were observed after photolysis of luciferase containing **BHCK** or **HCK**, while the K206→**HC<sub>2</sub>K** mutant was permanently deactivated, as expected. Error bars represent standard deviations from three independent experiments.

To investigate decaging of **ACK** *in vivo*, the same caged firefly luciferase experiment was carried out. HEK 293T cells were co-transfected with the mutated firefly luciferase plasmid (pGL3-K206TAG) and the *MbBHCKRS/PyltRNA<sub>CUA</sub>* pair (pBHCKRS-4PylT) in the absence

and presence of **HCK** (as a positive control) and **ACK** (0.25 mM). After UV irradiation, **ACK** was decaged to produce native lysine, resulting in the activation of firefly luciferase by 12-fold, which was comparable to the positive control, activation of FLuc-HCK by 28-fold (Figure 3.12). However, differences in incorporation efficiency between **ACK** and **HCK** were not taken into account.



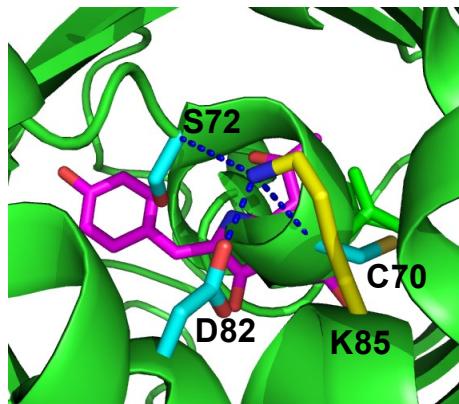
**Figure 3.12:** Bright-Glo luciferase assays of aminocoumarin lysine.

Bright-Glo luciferase assay of cells that were either kept in the dark or irradiated (365 nm, 4 min). Chemiluminescence units were normalized to the NoUAA/-UV control. No enzymatic\ activity was observed for the caged proteins and significant increases in luminescence were observed after photolysis of luciferase containing **ACK** or **HCK**. Error bars represent standard deviations from three independent experiments.

### 3.3.2 Optical control of EGFP folding and formation

In order to observe the optical triggering of protein function via decaging of **HCK** and **BHCK** in real time, enhanced green fluorescent protein (EGFP) was selected as a second target protein for

caging. EGFP consists of an 11-stranded  $\beta$ -barrel and a central  $\alpha$ -helix with the Thr65-Tyr66-Gly67 chromophore.<sup>122</sup> The chromophore plays a crucial role in EGFP fluorescence and stability.<sup>123</sup> Correctly folded EGFP is a prerequisite for mature chromophore formation, with a number of lysine residues being essential to its successful folding.<sup>124</sup> Most notable is that only one lysine (K85) out of twenty is buried within the protein.<sup>124</sup> K85 forms a salt bridge with D82 and H-bonding interactions with the backbone of C70 and S72,<sup>124</sup> all of which are in close proximity to the chromophore (Figure 3.13). It has been shown that C70, S72, and D82 are key residues for control of chromophore formation and oxidation.<sup>125-126</sup> It was hypothesized that introduction of coumarin-caged lysines at K85 would affect D82, C70, and S72 - interrupting the  $\alpha$ -helix bending, and thus indirectly inhibiting chromophore maturation. To this end, UV activation would yield native EGFP that rapidly undergoes maturation (Figure 3.13).

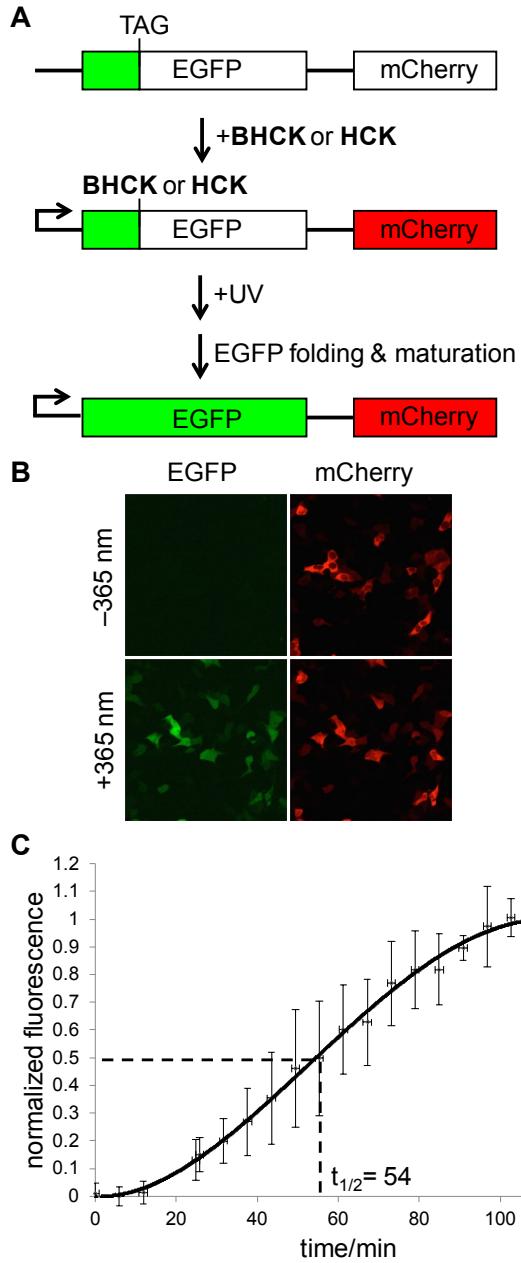


**Figure 3.13:** Location of K85 (yellow) and interactions with D82, C70, and S72 in EGFP.<sup>124</sup>

The chromophore is shown in magenta (model based on PDB 4EUL).

An EGFP mutant with an amber codon at position K85 (pEGFP-K85TAG) was generated as a fusion construct with mCherry, to provide a second reporter for successful plasmid transfection and incorporation of **HCK** and **BHCK** (Figure 3.14A). HEK 293T cells were co-

transfected with pEGFP-K85TAG-mCherry and the BHCKRS/PyltRNA pair in the presence of **HCK** and **BHCK** (0.25 mM). After 24 hrs, the cells were washed and incubated in fresh media for 1 hr. mCherry expressing cells were observed by fluorescence imaging to confirm that EGFP-**HCK/BHCK**-mCherry is generated in the presence of **HCK** or **BHCK**. Cells were irradiated for 30 s at 365 nm and fluorescence was imaged by time-lapse microscopy. After photolysis of EGFP-**HCK**, green fluorescence started to appear around 10 min, and over time the fluorescence intensity gradually increased reaching a plateau at 120 min, while the mCherry fluorescence intensity did not change (Figure 3.14B). A half-life of 54 min was observed matching reports of EGFP chromophore maturation as the rate-limiting step.<sup>127</sup> The C-terminal-fused mCherry, as a transfection and incorporation reporter, had no effect on the EGFP folding. Previous measurements of EGFP folding and maturation have been exclusively performed in test tubes.<sup>128</sup> No cellular studies have been conducted, as a precise starting point for kinetic analysis could not be provided (Figure 3.14C).



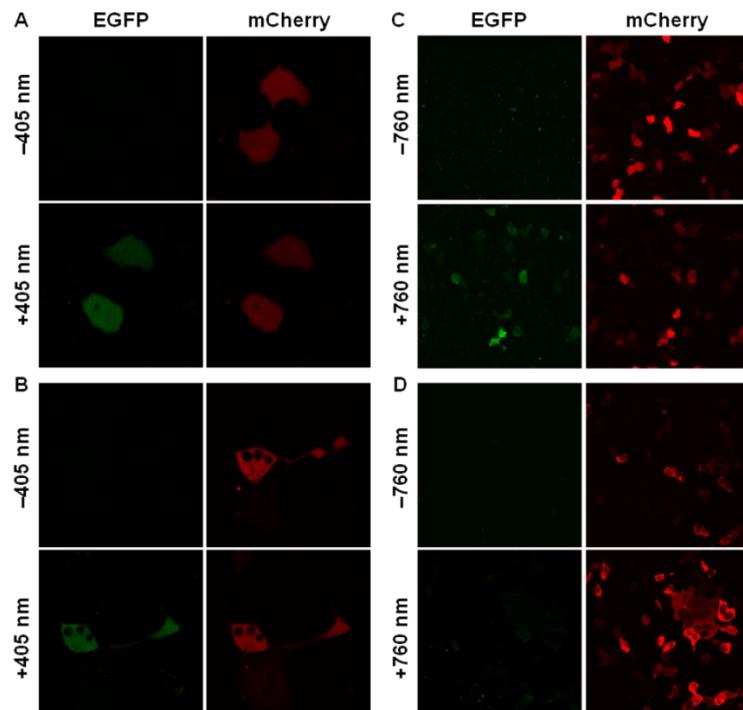
**Figure 3.14:** Fluorescence imaging of EGFP folding before and after 365 nm light-activation.

A) Schematic of the pEGFP-K85TAG-mCherry construct and its application in light activation studies. B) Fluorescence imaging of HEK 293T cells expressing EGFP-K85TAG-mCherry, 90 min after irradiation at 365 nm (30 s, DAPI filter, 358-365 nm) in the presence of **BHCK**. (Nikon A1R confocal microscope, 20× objective, 2-fold zoom). C) Normalized EGFP fluorescence as a function of time after 365 nm light-activation (error bars represent standard deviations from the measurement of three independent cells,  $t_{1/2} = 54$  min).

Given that the coumarin lysines have relatively broad absorbance spectra (Figure 3.11) and that the hydroxyl group of **HCK** contributes to a shoulder,<sup>129</sup> a longer wavelength (such as 405 nm in the range of visible light) might efficiently decage **BHCK** and **HCK**. Thus, similar experiments as described previously were performed in COS-7 cells. Photodecaging by 405 nm laser on a Zeiss confocal microscope enabled rapid protein decaging. As expected, **BHCK** and **HCK** blocked the EGFP fluorescence before irradiation, and upon irradiation, EGFP fluorescence gradually increased. The EGFP fluorescence intensity reached a maximum and remained stable after 2 hrs upon photolysis of **BHCK** and **HCK** (Figure 3.15). An exact time course was performed for kinetics analysis of EGFP folding. Since attempts to decage a previously incorporated nitrobenzyloxycarbonyl lysine<sup>46, 48, 130</sup> and nitrobenzyl tyrosine<sup>50, 53, 131</sup> at 405 nm were not successful on comparable timescales and at comparable illumination power, the caged lysines **HCK** and **BHCK** may enable multi-wavelength activation of proteins caged with the two different optical probes.

Taking advantage of the two-photon decaging feature of **BHCK**,<sup>116</sup> photocontrol of EGFP folding by two-photon activation of EGFP-**BHCK** was performed. HEK 293T cells were co-transfected with pEGFP-K85TAG-mCherry and pBHCKRS-4PyIT in the absence or presence of **HCK** and **BHCK** (0.25 mM). After a 24 h incubation, the cells were washed and incubated in fresh media for 1 h, and irradiated with a multi-photon laser (760 nm, 130 mW, 2 μm/s dwell time, 30 cycles, Olympus Fluoview FV1000 MPE Multiphoton laser scanning microscope FV10-ASW, MaiTai DSBB-OL IR pulsed laser). Images were acquired before and after two-photon irradiation using both EGFP (488 nm) and mCherry (561 nm) excitation. Gratifyingly, an EGFP fluorescent signal was observed after photolysis of **BHCK** at 760 nm (Figure 3.15C). The cells

expressing EGFP containing **HCK**, as a control, were also exposed to two-photon excitation (760 nm) and imaged in the same fashion (Figure 3.15D); no EGFP activation was observed. In addition to the increased three-dimensional resolution that is provided through two-photon excitation, effectively shifting the activation wavelength to the near-IR will enable multi-wavelength activation in conjunction with other optically triggered biological processes, while also preventing any overlap with established fluorescent reporter proteins.

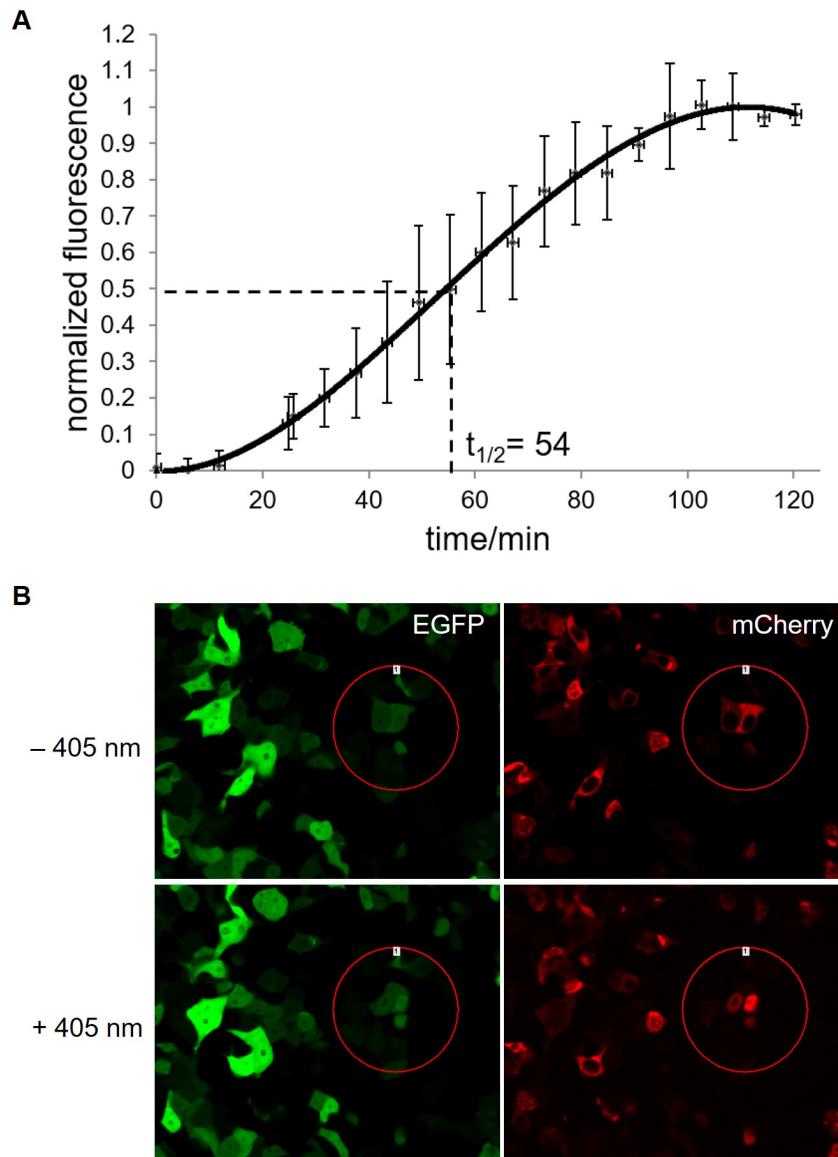


**Figure 3.15:** Fluorescence confocal imaging of EGFP folding before and after 405 nm and 760 nm light-activation. Fluorescence confocal imaging of COS-7 cells expressing EGFP-KTAG-mCherry, before and after irradiation at 405 nm (30 mW diode laser, 20% laser power, 12.6  $\mu$ s dwell time, 8 cycles) in the presence of **BHCK** (A) or **HCK** (B) (Zeiss confocal LSM710 microscope, 40 $\times$  oil objective). Similar light-activation experiments before and after irradiation of HEK 293T cells at 760 nm (130 mW, 2  $\mu$ m/s dwell time, 30 cycles, Olympus Fluoview FV1000 MPE, MaiTai DSBB-OL IR pulsed laser) in the presence of **BHCK** (C) or **HCK** (D). (Olympus microscope, 40 $\times$  oil objective)

Herein, the new coumarin lysine analogues, **BHCK**, **HCK**, and **HC<sub>2</sub>K**, were site-specific and genetically incorporated into proteins in both *E. coli* and human cells. **HCK** was utilized to fluorescent nuclear localize the protein in mammalian cells. The two-photon **BHCK** and single-photon **HCK** coumarin lysines was applied to optochemical control of EGFP formation upon irradiation of different wavelength of light or laser (365 nm, 405 nm, and 740 nm) *in vivo*.

### 3.3.3 Optical control of SATB1 nuclear localization

In addition to the light triggering of enzymatic function (FLuc) and protein maturation (EGFP), we wanted to demonstrate the optical control of a protein that is the substrate for another protein. Thus, we investigated light-activated nuclear translocation by genetically incorporating **HCK** into the nuclear localization signal (NLS) of the transcription factor SATB1. The presence of the coumarin group at position K29 completely inhibited translocation of an mCherry reporter as an NLS fusion, as observed in HEK 293T cells expressing SATB1K29**HCK**-mCherry. After 24 hr of incubation, the cells were treated with UV light and nuclear translocation of mCherry was triggered through light-induced formation of the native SATB1-NLS (Figure 3.16B). Translocation was completed within 90 min, as shown by time-lapse imaging, and fluorescence quantification over time revealed a  $t_{1/2}$  of 54 min (Figure 3.16A). Importantly, complete protein translocation from the cytoplasm to the nucleus was observed, demonstrating complete **HCK**-SATB1 decaging in live cells through the optical triggering of a caged SATB1-NLS.



**Figure 3.16:** Time-course fluorescent imaging of SATB1 nuclear translocation in live cells.

Site-specific incorporation of **HCK** at K29 in the SATB1 nuclear localization sequence leads to complete exclusion of mCherry from the nucleus, until the wild-type NLS is generated through light treatment. A) A time-course of the normalized mCherry fluorescence in the nucleus indicates a combined  $t_{1/2}$  of decaging and protein translocation of 54 min. Error bars represent standard deviations from three cells. B) Micrographs of HEK293T cells showing no nuclear localization of **HCK-SATB1-mCherry** and complete activation of translocation from the cytoplasm to the nucleus after light activation. EGFP expression was used as a transfection and imaging control. The nuclear translocation occurred only in irradiated cells within the red circle. The cells were imaged using a Nikon A1 confocal microscope with a Plan Apochromat 40 $\times$  oil objective.

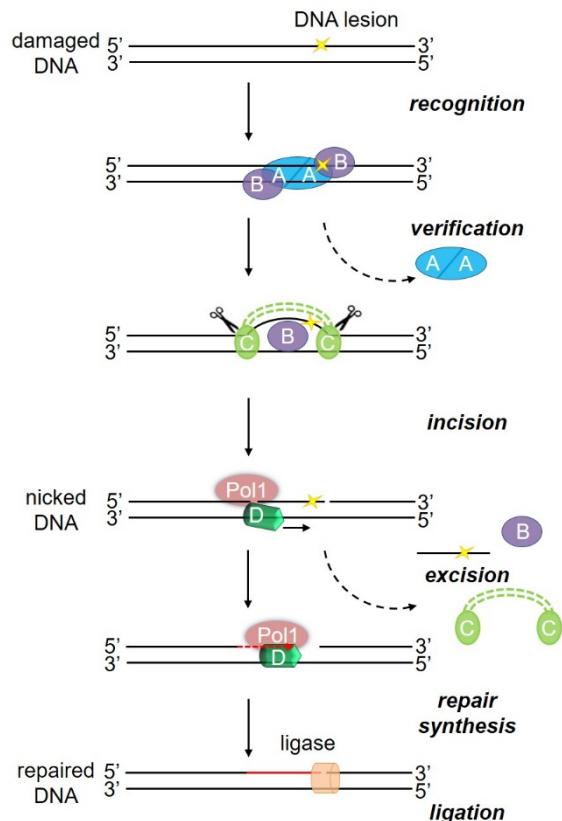
### **3.4 OPTICAL CONTROL OF HELICASE UVRD FOR DNA REPAIR**

This material was reprinted, in part, with permission from [Luo, J.; Kong, M.; Li, L.; Samanta, S.; Van Houten, B.; Deiters, A. \*Chembiochem\* 2017, 18\(5\), 466-469.](#)

UvrD is a DNA-dependent ATPase and helicase that belongs to the helicase SF1 superfamily and catalyzes the unwinding of duplex DNA in a 3' to 5' direction.<sup>132</sup> UvrD plays essential roles in both methyl-directed mismatch repair and nucleotide excision repair (NER) in bacteria<sup>133-135</sup> and corresponding functions show a high degree of conservation in yeast, and human cells. NER is a generalized DNA repair mechanism that is capable of removing a wide variety of DNA lesions resulting from physical agents, such as sunlight or chemicals, such as cisplatin or polycyclic hydrocarbons. In bacteria, UvrA and UvrB recognize and bind the DNA lesion. Once UvrB is loaded at the site of damage, it helps to recruit UvrC to perform dual incisions on both sides of the lesion, leaving a UvrBC-DNA post-incision complex. UvrD and DNA polymerase I work in concert to turnover the UvrBC-DNA complex and UvrD unwinds the doubly nicked duplex DNA and displaces the lesion-containing strand (Figure 3.17).<sup>136</sup> DNA polymerase I is then able to undergo repair synthesis to generate a repair patch that is ligated during the final step.

Helicase defects have been linked to several human diseases including cancer and genetic disorders.<sup>137</sup> However, determining the molecular basis of helicase function defects that cause the pathophysiology associated with human diseases has been challenging, due to the involvement of a complex network of interconnected roles of DNA helicases and their protein

interactions.<sup>137</sup> For example, XPD and XPB are key helicases that play important roles in human NER and are part of the seven-member TFIIH core complex.<sup>138</sup> Since many disease-causing mutations in either XPD or XPB affect the stability of TFIIH,<sup>138</sup> having the ability to directly initiate helicase activity of either protein within the TFIIH complex would be most desirable. Therefore, as proof of principle, we aimed to develop a strategy for conditional control of helicase function. In order to provide a general approach for the optical activation of helicase function, we photocaged a key amino acid residue that is conserved throughout the helicase protein family. The precise temporal control provided through photocaging<sup>20, 94, 139-140</sup> may allow assessment of heretofore intractable problems, such as how helicase-dependent DNA repair pathways are coordinated.



**Figure 3.17:** Schematic showing the role of UvrD in bacterial NER.

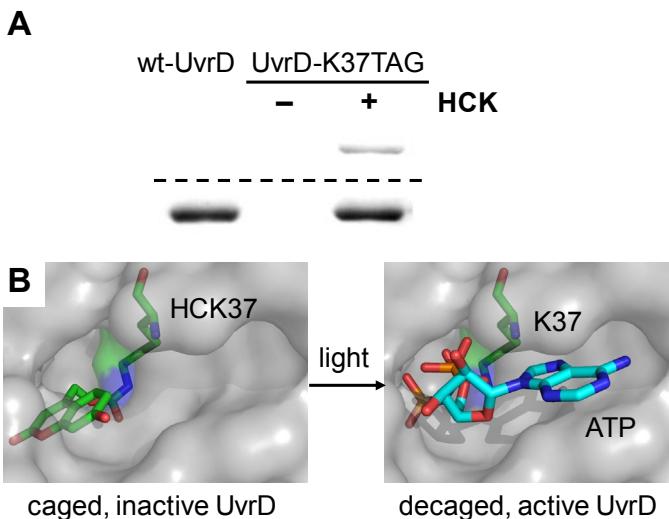
During damage detection, UvrA and UvrB proteins bind at the site of damaged DNA. Once UvrB is loaded at the site of damage during the verification step, UvrA dissociates and UvrC is recruited and produces two incisions on the damaged strand both 5' and 3' to the damaged nucleotide. The dual action of UvrD and PolI are necessary to remove the damaged oligonucleotide and carry out repair synthesis using the complementary strand as a template. DNA ligase seals the newly made repair patch.

Here, we are utilizing the *Methanoscarcina barkeri* pyrrolysyl tRNA synthetase/tRNA<sub>CUA</sub> pair,<sup>104</sup> which we have previously applied to the optical control of transcription, nuclear translocation, protein folding, DNA recombination, genome editing, and kinase function.<sup>48, 107, 141-144</sup> The BHCKRS/tRNA<sub>CUA</sub> pair previously identified in our lab directs the incorporation of hydroxycoumarin lysine (**HCK**, Figure 3.1) in response to an amber stop codon, TAG.<sup>107</sup> The genetically encoded coumarin lysine has been successfully applied as both a fluorescent cellular probe and a light-activated caging group for the optical activation of protein function. Therefore, site-specific incorporation of **HCK** into UvrD will enable the genetic encoding of light-activated DNA helicase function for the first time.

### 3.4.1 Incorporation of HCK into UvrD in *E. coli*

We selected the lysine at position 37 as a target for hydroxycoumarin caging group introduction (Figure 3.18B), since it is highly conserved in the helicase superfamily and is essential for ATP binding.<sup>145</sup> K37 is localized in the nucleotide binding motif I and is one of the four basic residues that are responsible for coordination of the triphosphate moiety, in particular the  $\gamma$ -phosphate.<sup>145</sup> Herein, we reasoned that photocaging K37 in UvrD (K37→**HCK**) prevents ATP-dependent DNA unwinding and thus creates an inactive protein. UvrD-K37**HCK** could be temporally

controlled through activation by light-induced decaging (Figure 3.18B), which in turn generates wild-type UvrD.

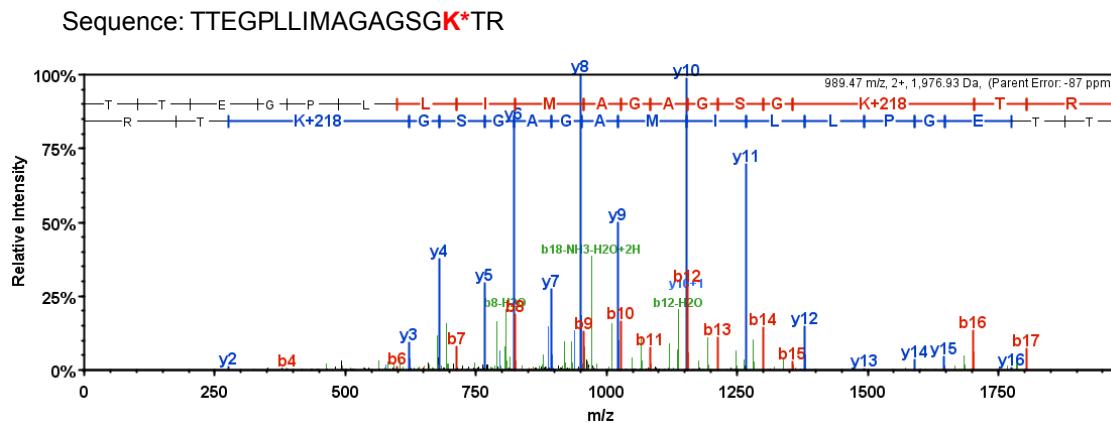


**Figure 3.18:** Site-specific incorporation of coumarin lysine into UvrD at residue K37.

A) SDS-PAGE analysis of wild-type UvrD and photocaged UvrD (UvrD-K37HCK) expressed in *E. coli*. The gel was stained with Coomassie blue (bottom), and the in-gel fluorescence was imaged via excitation at 365 nm (top). B) **HCK** modeled into the active site of UvrD (PDB 4JA8) obstructs the ATP binding site until photochemically removed through irradiation, delivering the wild-type enzyme.

A 6×HIS-tagged UvrD was generated through amplification of the UvrD gene from pETM11-UvrD and cloning into the pBAD-PylT backbone, creating pBAD-UvrD-PylT. A TAG codon and an alanine codon were introduced at position 37, and the mutated UvrD-K37TAG (pETM11-UvrD-K37TAG) and UvrD-K37A (pETM11-UvrD-K37A) were cloned into the pBAD-PylT expression vector as well, creating the pBAD-UvrD-K37TAG-PylT and pBAD-UvrD-K37A-PylT plasmids, respectively. Expression of full-length UvrD from *E. coli* bearing

UvrD-K37TAG and BHCKRS/tRNA<sub>CUA</sub> pair was dependent in the presence of **HCK**. The purified protein was analyzed by SDS-PAGE (Figure 3.18A) and confirmed by MS/MS sequencing (Figure 3.19).



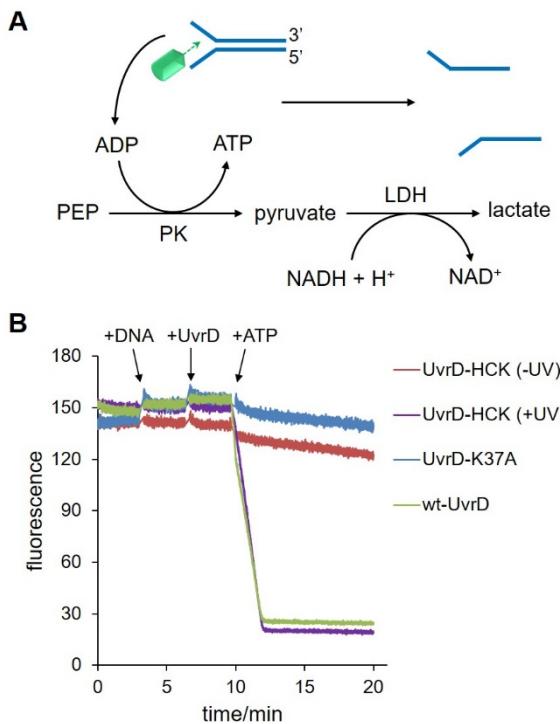
**Figure 3.19:** MS/MS analysis of incorporation of **HCK** into UvrD.

UvrD-6×His incorporating **HCK** expressed in Top10 competent cells was purified by Ni-NTA column for the MS/MS analysis. The MS/MS fragmentation spectrum of a tryptic peptide from the purified protein confirms the incorporation of **HCK** at the specific K37 site.

### 3.4.2 Fluorescence-based coupled ATPase activity assays

In order to evaluate the optical control of UvrD function through introduction of the K37**HCK** mutation, we applied several fluorescence-based assays due to their high sensitivity and ease of quantification. UvrD is a DNA-dependent ATPase,<sup>146</sup> and its ATPase activity was characterized and quantified using an assay design that is based on the reaction of phosphoenolpyruvate (PEP) to pyruvate by pyruvate kinase (PK), which is then coupled to the conversion of pyruvate to lactate by lactate dehydrogenase (LDH).<sup>147</sup> In the first reaction, ATP is hydrolyzed to ADP by

the ATPase domain of UvrD, and then converted back to ATP by PK, which in turn leads to the conversion of PEP to pyruvate. In the second reaction, the pyruvate is reduced to lactate by LDH, which in turn oxidizes NADH to NAD<sup>+</sup> (Figure 3.20A). The ATPase activities of wild-type, photocaged, decaged, and inactive UvrD were determined by time-course measurement of the fluorescence signal generated by NADH, as it is directly linked to the formation of ADP by UvrD (Figure 3.20B). The rapid drop of NADH concentration within minutes of light-activation of the caged UvrD indicates a sudden, dramatic increase in ATP consumption due to decaging of the ATPase domain. Similar results were observed after addition of wild-type UvrD (Figure 3.20B), suggesting complete activation of ATPase function of photocaged UvrD upon illumination. ATPase activities of wild-type UvrD and light-activated UvrD were calculated at 21.0 s<sup>-1</sup> and 36.9 s<sup>-1</sup>, respectively (Table 3.2). In addition, fluorescence intensity in the case of the photocaged UvrD (**UvrD-HCK**) decreased at a very slow rate, similar with that of the inactive UvrD-K37A mutant as a negative control, indicating that the introduction of the caging group into the ATPase active site fully inhibited its function. Because ATP hydrolysis is a shared feature of all cellular ATPases and other ATP-dependent enzymes, this may represent a general approach to the optical control of ATP-dependent enzymes.



**Figure 3.20:** Fluorescence-based coupled ATPase activity assay.

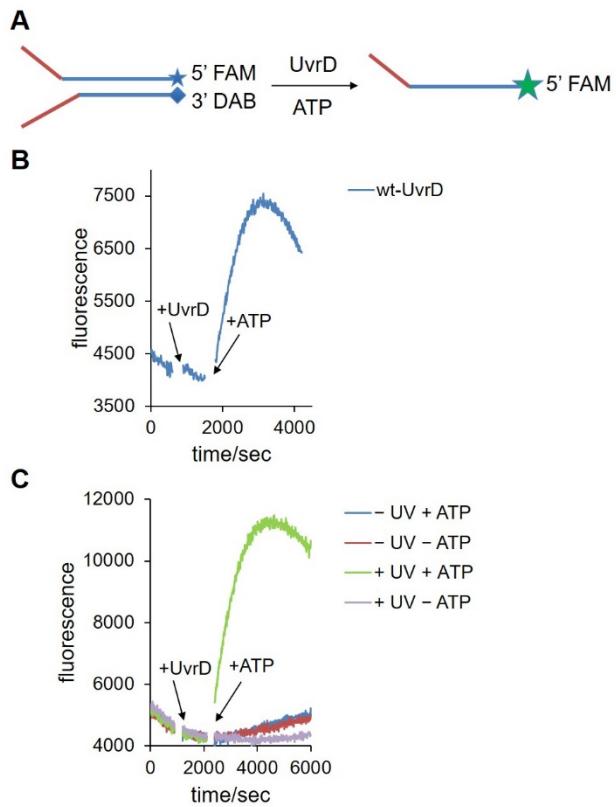
A) Mechanism of the coupled ATPase activity assay. B) Time-course fluorescence measurement of NADH consumption driven by UvrD ATPase. This trace is representative of three independent experiments. Fluorescence-based coupled ATPase activity assay.

### 3.4.3 Helicase activity assays

#### 3.4.3.1 Helicase activity assay on quenched Y-shaped DNA

Since the ATPase activity of UvrD-**HCK** could be restored through UV exposure, actual helicase activities were characterized by using a quenched Y-shaped double-stranded DNA substrate, containing a 5' FAM (fluorescein) labeled top strand and a 3' DAB (dimethylaminophenylazophenyl) labeled bottom strand. FAM fluorescence is quenched by DAB until the strands are separated by UvrD helicase activity (Figure 3.21A). The DNA

substrate and wild-type enzyme were combined and, after initial stabilization of the fluorescence readout, ATP was added to initiate the reaction. A rapid increase in fluorescence was observed, indicating that UvrD binds the top strand at the 3' end followed by ATP-driven unwinding of the DNA duplex. Generation of single-stranded nucleic acids leads to activation of FAM fluorescence. After reaching maximum fluorescence intensity, fluorescence dropped back down due to consumption of ATP and annealing of the fluorescent top strand back to the quencher bottom strand. (Figure 3.21B). Following validation of the assay for the wild-type helicase, the same concentration of photocaged UvrD was irradiated (365 nm) and the same helicase assay was performed. Only in the presence of light exposure and ATP, was an increase in total fluorescence intensity detected, revealing light-activated UvrD-**HCK** activity comparable to wild-type UvrD (Figure 3.21C). As expected, three negative control experiments showed that only minimal fluorescence was observed in case of the photocaged, non-irradiated UvrD system in the presence or absence of ATP and in the light-activated UvrD system without ATP (Figure 3.21C). Therefore, optical control of actual UvrD helicase was achieved as determined by Y-shaped double-stranded DNA substrate assay. In order to further investigate light-activation of UvrD in a context more relevant to the NER process, an unwinding assay using nicked plasmid DNA as a substrate was performed.



**Figure 3.21:** Fluorescence-based UvrD helicase activity assay using a quenched Y-shaped DNA duplex.

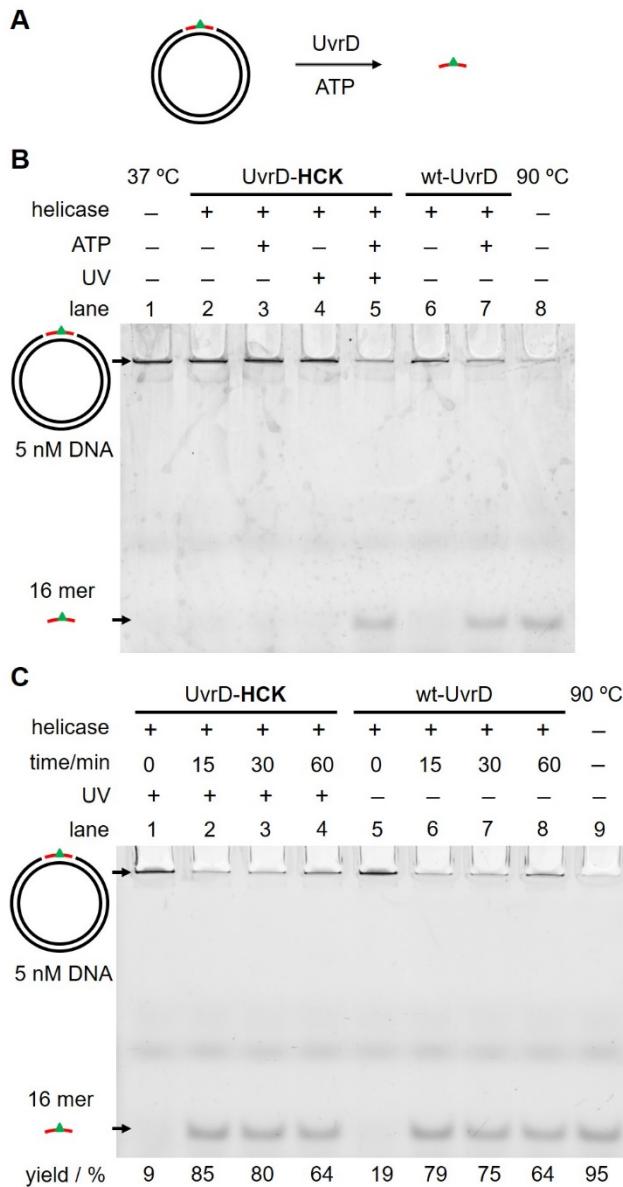
A) Assay design based on fluorophore activation as a result of DNA duplex dehybridization. B) Time-course measurement of the helicase activity of wild-type UvrD, C) Time-course measurement of the helicase activities of photocaged UvrD-HCK, with or without light activation, in the absence or presence of ATP.

### 3.4.3.2 Gel-based helicase activity assays

Gel-based helicase assays were performed to not only complement and support the validity of the fluorescence-based helicase activity measurements, but also to simulate the real UvrD helicase function in the general NER pathway in bacteria. These assays were carried out with a nicked plasmid substrate that contains a fluorescein-labeled 16-mer DNA strand between the two nicks, and displacement of the fluorescent strand through unwinding by the helicase was analyzed by gel electrophoresis (Figure 3.22A). In bacteria, UvrA and UvrB recognize and bind the DNA lesion, and recruit UvrC to excise the fragment, cleave the impaired strand on both sides of the

lesion, resulting in a nicked strand; the resulting oligonucleotide is unwound and displaced by UvrD. Here, the nicked substrate designed can mimic the intermediate state of NER processes.

The incorporation of **HCK** into UvrD caused complete inhibition of unwinding activity before UV irradiation, comparable to the absence of activity in the negative controls (nicked substrate only and UvrD-WT without ATP; inactive UvrD-K37A, Figure 3.22). As expected, UV irradiation of UvrD-**HCK** led to activation and unwinding activity upon the addition of ATP, which is consistent with wild-type UvrD, while light-activated UvrD did not show any helicase activity in the absence of ATP (Figure 3.22B). Without UvrD (lane 1) or with UvrD lacking ATP (lane 6), the 16-mer unwound ssDNA was not observed. In addition, in order to evaluate the rates of unwinding between wild-type and light-activated UvrDs, the time-course helicase assays were performed (Figure 3.22C). It was observed that the *light-activated UvrD* significantly stimulated *wild-type* UvrD unwinding activity in a time-dependent manner, based on the measured unwinding yields over time (Figure 3.22C). Therefore, this gel-based helicase assay is robust and reliable to complement and crosscheck with the fluorescence-based helicase assay. It demonstrates that UvrD helicase function could be optically controlled for repair of nicked DNA in the NER pathway.



**Figure 3.22:** Analysis of light-activated helicase activity using a substrate resembling the post-incision intermediate of bacterial NER.

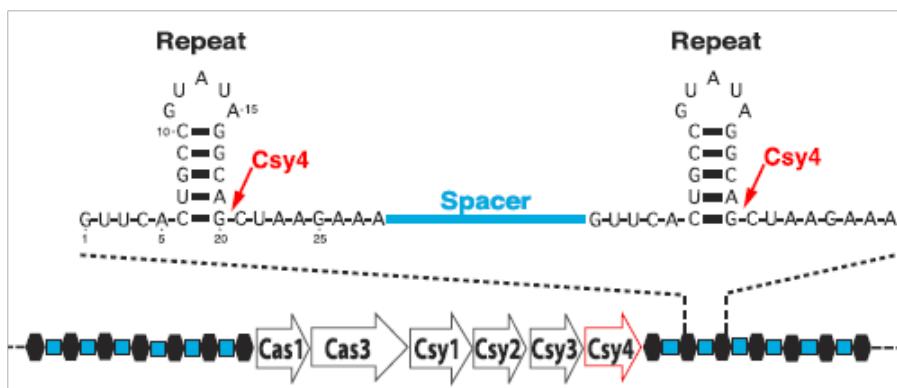
A) Design of the nicked plasmid unwinding assay: released 16-mer ssDNA is shown in red; the Fl-dT fluorophore is indicated by the green triangle. B) Unwinding reactions of wild-type and photocaged UvrD on the nicked plasmid DNA substrate. Unwound Fl-dT-containing 16-mer ssDNA was separated from the plasmid substrate by a 12% native PAGE. Top band: plasmid DNA containing the 16-mer ssDNA insert. Bottom band: unwound 16-mer ssDNA. C) Time-course of the nicked plasmid unwinding assays of wild-type and light-activated UvrD analyzed by native PAGE.

Herein, we have engineered a genetically encoded, light-activated *E. coli* helicase UvrD. The activity of the enzyme can be stringently regulated through the use of a light-removable caging group installed directly on the essential residue K37 in the active site. This was achieved by adding an engineered pyrrolysyl tRNA synthetase/tRNA pair for the genetic encoding of **HCK** to the protein expression host, *E. coli*. The use of lysine protected with a hydroxycoumarin caging group (compared to the more common *ortho*-nitrobenzyl caging group) provides improved expression levels of the caged UvrD protein. In addition, we have developed three different fluorescence-based assays using different template substrates for the characterization of the enzymatic activities of light-activated helicase UvrD. The overall activity of light-activation of UvrD was comparable to that of wild-type UvrD, and thus enabled stringent optical control of helicase activity. Moreover, the targeted, highly conserved lysine site in the ATP binding domain can also be found in other helicase II superfamily proteins, such as PcrA, Rep, and Srs2, further expanding this approach in future applications. Therefore, this is a general approach to the optical control of all ATP-dependent helicases (and ATP-dependent enzymes in general).

### 3.5 OPTICAL CONTROL OF CSY4 FUNCTION

Many prokaryotes contain clustered regularly interspaced short palindromic repeats (CRISPRs) to resist viral infection.<sup>148</sup> Any invasive genetic elements (comprised of DNA) integrated into CRISPR loci are transcribed into long pre-crRNAs (CRISPR transcripts).<sup>149</sup> Through post-transcription, pre-crRNAs are cleaved into short crRNAs by CRISPR-associated (Cas) proteins

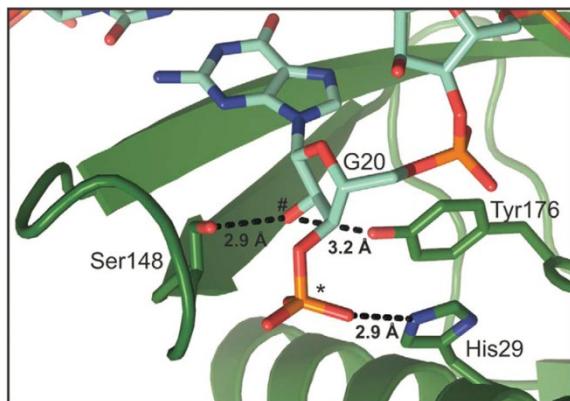
against the propagation of foreign plasmids.<sup>149</sup> Therefore, CRISPR loci and their co-existing Cas proteins constitute a CRISPR/Cas adaptive immune system (Figure 3.23).<sup>150</sup> In this “immune system”, it was discovered in 2010 that the Csy4 endoribonuclease is responsible for CRISPR transcript processing.<sup>150</sup> CRISPR transcript cleavage by Csy4 is a rapid, metal ion-independent reaction.<sup>150</sup> The catalytic dyad, His29 and Ser148, contains two strictly conserved residues participating in cleavage activity (Figure 3.24).<sup>151</sup> Based on the crystal structure of Csy4 (PDB: 2XLK) and its cognate substrate RNA, Lys106 is positioned at the pocket of RNA binding,<sup>150</sup> its side chain winds through the RNA hairpin binding site, and stacks on top of guanine (G).<sup>151</sup> Therefore, we hypothesized that Lys106 is closely associated with RNA binding and the catalytic dyad, and Csy4 could be photochemical-controlled by incorporation of **HCK** upon light irradiation. Significantly, RNA-binding proteins regulate cellular pathways at the level of transcription, posttranscriptional modification, and translation.<sup>152-153</sup> This strategy can offer us a sophisticated tool for studying RNA function with specific binding proteins with tight spatial-temporal resolution.



**Figure 3.23:** Schematic of the CRISPR/Cas locus.

Csy4 specifically cleaves its cognate pre-crRNA substrate. The six Cas genes are flanked by two CRISPR loci, each consisting of a series of 28-nucleotide repeats separated by 32-nucleotide distinct spacer sequences (blue). Red

arrows denote the cleavage site. Adapted from Haurwitz *et al*, *Science* 2010, 329(5997):1355–1358. Reprinted with permission from AAAS.

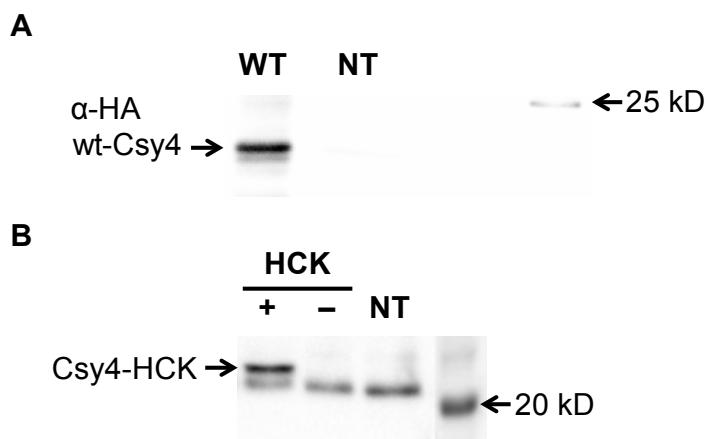


**Figure 3.24:** Csy4 active site in the Csy4/substrate complex (PDB: 2XLK).

Adapted from Haurwitz *et al*, *EMBO J.* 2012, 31(12): 2824–2832. Reprinted with permission from John Wiley and Sons.

In order to verify whether the wild type Csy4 and Csy4-HCK could be expressed in mammalian cells, Western blots were carried out in HEK 293T cells. An HA-tag was inserted into the N terminus of the Csy4 coding sequence, generating pTR-wt-Csy4-HA and pTriSys-Csy4-KTAG-HA. First, pTR-wt-Csy4-HA was transfected into HEK 293T cells and a Western blot was carried out (Figure 3.25A). For wild-type expression, a band corresponding to a 21.4 kDa HA-tagged Csy4 was detected (Figure 3.25A). Subsequently, pTriSys-Csy4-K106TAG-HA was co-transfected with the pBHCKRS-4CMVE-U6-PyLT plasmid into HEK 293T cells in the absence and presence of **HCK** (0.25 mM). In the case of caged Csy4 expression, the corresponding band at the molecular weight of the wild-type protein was only observed in the presence but not the absence of **HCK** (Figure 3.25B). Cell lysates from the non-transfected HEK 293T cells were analyzed as another negative control, indicating the bottom band visible is the

antibody light chain background due to non-specific binding. This demonstrates that **HCK** was successfully incorporated into the mutant Csy4 protein in mammalian cells.

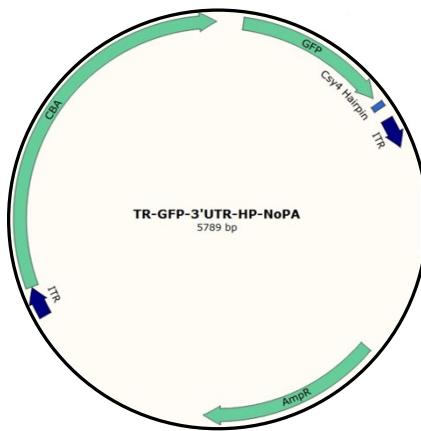


**Figure 3.25:** Western blots to confirm expression of wild-type and caged Csy4 in mammalian cells.

A) wt-Csy4; B) Incorporation of **HCK** into Csy4.

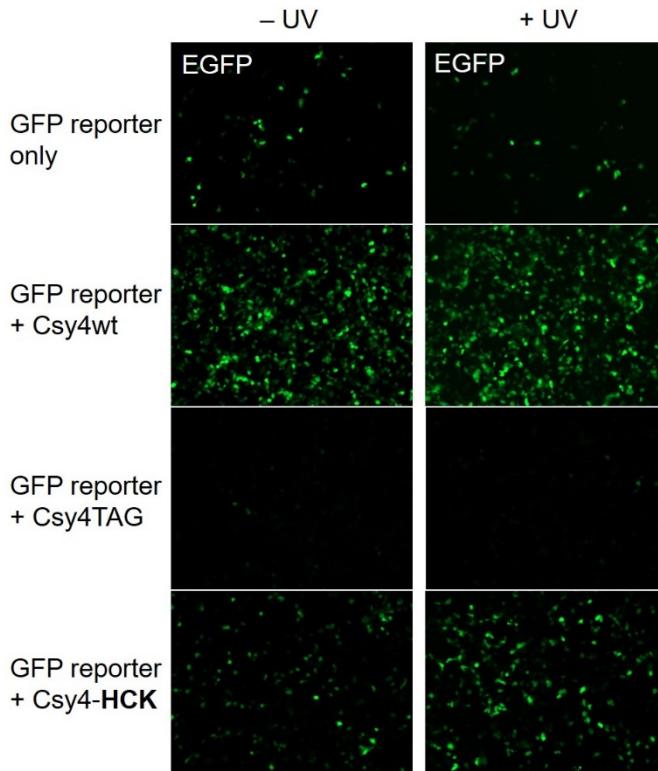
To demonstrate optical activation of caged Csy4 in live cells, a GFP-based fluorescence Csy4 activity reporter constructed by the Aravind lab (UNC) was used (Figure 3.26). Based on the Csy4-mediated RNA recognition mechanism (shown in Figure 3.23),<sup>150</sup> the CRISPR repeat stem-loop sequence (hairpin), which is the cleavage recognition sequence for Csy4, was inserted downstream of GFP. Additionally, the poly(A) site downstream of the GFP coding sequence was removed. GFP expression is low without polyadenylation, while expression of GFP is switched on and normally regulated once Csy4 is activated. If Csy4 is deactivated by caging a key lysine residue (as in Csy4-**HCK**), polyadenylation does not occur, and thus mature, polyadenylated mRNA is not able to be produced for translation, resulting in low GFP expression and low fluorescence intensity. If Csy4-**HCK** is irradiated with UV light, activated

Csy4 can cleave the pre-crRNAs, so GFP gene expression is turned on, and thus GFP fluorescence is observed.



**Figure 3.26:** Representative vector map of the GFP reporter for Csy4 activity.

The activity of wild type Csy4 was tested using the EGFP reporter (Figure 2.26). As expected, the expression of wild-type Csy4 led to intracellular transcript cleavage and turning on GFP expression (Figure 3.27). Compared to the negative control transfected with GFP reporter only, a low level of GFP expression was observed. Subsequently, K106 of Csy4 was mutated to a TAG amber codon and then genetically encoded with **HCK** to obtain a caged, inactive Csy4. As a negative control, the expression of caged Csy4 in the absence of **HCK** showed very low level of GFP expression, indicating that no functional Csy4 was generated. Notably, cells transfected with pTriSys-Csy4-KTAG-HA in the presence of **HCK** but in the absence of UV irradiation also showed a low level of GFP expression, suggesting the complete inactivity of the caged Csy4-**HCK**. The caged Csy4 protein was activated by UV irradiation, which induced decaging and switching on GFP expression (Figure 3.27).

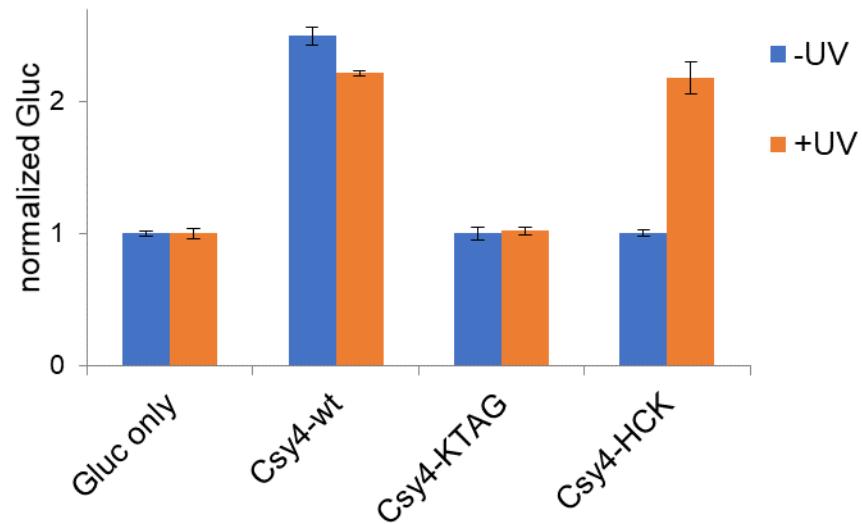


**Figure 3.27:** Optical control of Csy4 endonuclease in HEK 293T cells using the GFP reporter.

Fluorescent cell micrographs of detection of Csy4 activity in the absence and presence of UV irradiation. 1) Negative control: transfection of EGFP reporter only; 2) Co-transfection of wt-Csy4 and EGFP reporter induced EGFP switch-on and fluorescence shown up; 3) Negative control: triple transfection of Csy4-K106TAG, pBHKRS-4PyIT, and EGFP reporter in the absence of **HCK**; 4) Triple transfection of Csy4-K106TAG, pBHKRS-4PyIT, and EGFP reporter in the presence of **HCK**, UV irradiation led to Csy4 activated, and thus EGFP switch-on and fluorescence shown up.

Similarly, to apply **HCK** to optical control of Csy4 in live cells, a Gluc-based fluorescence Csy4 activity reporter was also constructed by the Aravind lab. As expected, the expression of wild-type Csy4 led to intracellular transcript cleavage and detection of Gluc chemiluminescence, which was normalized to the negative control (transfected with Gluc reporter only) as 1 (Figure 3.28). In the case of caged Csy4 protein, weak Gluc signals were detected in the absence and presence of **HCK** before UV treatment, indicating that Csy4 was not

functional. Importantly, high level of Gluc chemiluminescence was detected only in the presence of **HCK** through UV irradiation (Figure 3.28).



**Figure 3.28:** Optical activation of Csy4 endonuclease in HEK 293T cells using the Gluc reporter.

Gluc assays for detection of Csy4 activity in the absence and presence of UV irradiation – 1) Negative control: transfection of Gluc reporter only; 2) Co-transfection of wt-Csy4 and Gluc reporter induced Gluc switch-on and detection of Gluc; 3) Negative control: triple transfection of Csy4-K106TAG, pBHCKRS-4PyIT, and Gluc reporter without **HCK**; 4) Triple transfection of Csy4-K106TAG, pBHCKRS-4PyIT, and Gluc reporter with **HCK**, UV irradiation led to Csy4 activated, and thus Gluc switch-on and fluorescence shown up. Error bars represent standard deviations from three independent experiments.

### 3.6 CONCLUSIONS

In summary, the site-specific genetic incorporation of three new coumarin lysine analogues into proteins was achieved in bacterial and mammalian cells using an engineered BHCKRS synthetase system. The genetically encoded coumarin lysines were successfully applied as

fluorescent cellular probes for protein localization and the small size of these coumarin lysines is expected to minimally perturb protein structure and function, unless they are placed at critical sites. In addition to their small size, the spectral properties of new coumarin lysines do not interfere with common fluorescent proteins (e.g., EGFP). While the amino acid **HC<sub>2</sub>K** showed stability under irradiation conditions, the coumarins **BHCK** and **HCK** were readily decaged, generating wild-type lysine residues. As a proof-of-principle, photo-regulation of firefly luciferase was achieved in live cells by caging a key lysine residue and excellent OFF to ON light-switching ratios were observed for **BHCK** and **HCK**. As expected, the stable fluorescent amino acid **HC<sub>2</sub>K** did not undergo photolysis. Furthermore, two-photon and single-photon optochemical control of EGFP maturation was demonstrated, enabling the use of different, potentially orthogonal, excitation wavelengths (365 nm, 405 nm, and 760 nm) for the sequential activation of protein function in live cells. While the caged lysine **BHCK** could be activated using two-photon irradiation at 760 nm, the lysine **HCK** was stable under these conditions. However, decaging of **HCK** was readily achieved with blue light of 405 nm, while a previously encoded *ortho*-nitrobenzyl-caged lysine requires UV activation.<sup>46, 48, 130</sup> These results demonstrate that coumarin lysines are a new and valuable class of optical probes that can potentially be used for the investigation and regulation of protein structure, dynamics, function, localization, and DNA repair in live cells. The small size of coumarin, the application as both a light-activated caging group and a fluorescent probe, and the broad range of excitation wavelengths are advantageous over other genetically encoded photocontrol systems and provide a unique and multi-functional tool for cellular biology.

### 3.7 EXPERIMENTAL

#### **Plasmid cloning.**

- (1) Construction of pBKM<sub>b</sub>BHCKRS (bacterial expression system for BHCKRS): The pBKM<sub>b</sub>BHCKRS plasmid was obtained by making two mutations to the wild-type *MbPylRS*<sup>48</sup>: Y271A and L274M, using the primer pair P1/P2 (Table 3.1), with the QuikChange site-directed mutagenesis kit (Agilent).
- (2) Construction of p*MbBHCKRS-mCherry-TAG-EGFP*: The p*MbBHCKRS-mCherry-TAG-EGFP* plasmid was obtained by making Y271A and L274M mutations to the wild-type p*MbPylRS-mCherry-TAG-EGFP*<sup>48</sup> plasmid using the primer pair MP1/MP2, with the QuikChange site-directed mutagenesis method.
- (3) Construction of pBHCKRS-4PyIT: The plasmid was obtained by ligating the p4CMVE-U6-PyIT fragment from p*MbPylT*<sup>48</sup> between the restriction sites *NheI* and *MfeI* sites of p*MbBHCKRS-mCherry-TAG-EGFP*.
- (4) Construction of pETM11-UvrD-K37A and pETM11-UvrD-K37TAG: Plasmids were obtained by converting the AAG (Lys) codon of pETM11-UvrD-wt into an TAG and GCG (Ala) codons using primers QC1/QC2 and QC3/QC4, respectively, with a QuikChange site-directed mutagenesis kit (Agilent).
- (5) Construction of pBAD-UvrD-wt-PyltRNA, pBAD-UvrD-K37TAG-PyltRNA, and pBAD-UvrD-K37A-PyltRNA: A 6×HIS-tagged UvrD was generated through amplification of the UvrD gene from pETM11-UvrD (a gift from Dr. Caroline Kisker, University of Wurzburg)

using primers N1/N2 and cloning into the pBAD-PyltRNA backbone,<sup>107</sup> creating pBAD-6HIS-UvrD-PyltRNA. The mutated UvrD-K37TAG (pETM11-UvrD-K37TAG) and UvrD-K37A (pETM11-UvrD-K37A) genes were cloned into the pBAD-PyltRNA expression vector with the *Nco*I and *Nde*I sites to creating pBAD-6HIS-UvrD-K37TAG-PyltRNA and pBAD-6HIS-UvrD-K37A-PyltRNA plasmids, respectively.

**Table 3.1:** Primer list.

Primer	Sequences
P1	5' GATGCTGGCCCCGACCCTGGCGAACTATATGCGTAAACTGGATCGTATT 3'
P2	5' GAATACGATCCAGTTACGCATATAGTTGCCAGGGTCGGGCCAGC 3'
MP1	5' CTATGCTGGCCCCCACCTGGCCAACTACATGCGGAACTGGACAGAAC 3'
MP2	5' GATTCTGTCCAGTTCCGCATGTAGTTGCCAGGGTGGGCCAGC 3'
QC1	5' CGGGAGCGGGCAGCGGATAGACGCGCGTGTGACGCAC 3'
QC2	5' GTGCGTCAGCACGCGCGTCTATCCGCTGCCGCTCCG 3'
QC3	5' GAGCGGGCAGCGGAGCAACGCGCGTGTGAC 3'
QC4	5' GTCAGCACGCGCGTTGCTCCGCTGCCGCTC 3'
N1	5' GAATTAACCATGGTGCATCATCATCATCATATGAATTTTATCGGAAAAGC 3'
N2	5' CGAATTCCCATAATGTCATTATACTTCTCAATCGCG 3'

### Expression and purification of proteins in *E. coli*.

The plasmid, pBAD-sfGFP-Y151TAG-pylT<sup>108</sup> was co-transformed with pBK-BHCKRS<sup>48</sup> into *E. coli* Top10 cells. A single colony was grown in LB media overnight and 500 µL of the overnight culture was added to 25 mL of LB media, supplemented with 1 mM of the designated unnatural amino acid and 25 µg/mL of tetracycline and 50 µg/mL of kanamycin. Cells were grown in the presence of coumarin lysine derivatives at 37 °C, 250 rpm, and protein expression 104

was induced with 0.1% arabinose when the OD<sub>600</sub> reached ~0.6. After overnight expression at 37 °C, cells were harvested and washed by PBS. The cell pellets were re-suspended in 6 mL of phosphate lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, pH 8.0, pH 8.0) and Triton X-100 (60 µL, 10%), gently mixed, and incubated for an hour at 4 °C. The cell mixtures were sonicated (Fisher Scientific 550 Sonic Dismembrator) with six short bursts of 30 s on ice followed by 30 s intervals for cooling, and then the cell lysates were centrifuged at 4 °C, 13,000 g, for 10 minutes. The supernatant was transferred to a 15 mL conical tube and 100 µL of Ni-NTA resin (Qiagen) was added. The mixture was incubated at 4 °C for 2 hours under mild shaking. The resin was then collected by centrifugation (1,000 g, 10 min), washed twice with 400 µL of lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, pH 8.0), followed by two washes with 400 µL of wash buffer (lysis buffer containing 20 mM imidazole). The protein was eluted with 400 µL of elution buffer (lysis buffer containing 250 mM imidazole). The purified proteins were analyzed by 10% SDS-PAGE and stained with Coomassie Blue.

### **Protein analysis by ESI-MS.**

Two different instruments were used: (A) Protein samples were analyzed using capillary LC ESI-TOF MS. The protein samples were loaded onto a PRLP-S column (Thermo Fisher 5 µm, 1000 Å, 300 µm i.d. × 100 mm) on an LC system (Ultimate 3000, Dionex, Sunnyvale, CA). The LC system was directly coupled to an electrospray ionization time-of-flight mass spectrometer (microTOF, BrukerDaltonics, Billerica, MA). Chromatographic separation was performed at a constant flow rate of 3.5 µL/min using a binary solvent system (solvent A: 2.5% acetonitrile and 0.1% formic acid; solvent B: 80% acetonitrile and 0.1% formic acid) and a linear gradient program (0–5 min, 5% B; 5–10 min, 5–30% B; 10–30 min, 30–75% B; 30–35 min, 75–100% B;

35–45 min, 100–5% B; 45–60 min, 5% B). Mass spectra were acquired in positive ion mode over the mass range m/z 50 to 3000. ESI spectra were deconvoluted with the MaxEnt algorithm (Data Analysis 3.3, Bruker Daltonics, Billerica, MA), obtaining molecular ion masses with a mass accuracy of 1-2 Da. (B) High-resolution exact mass measurement were conducted on an Agilent Technologies (Santa Clara, California) 6210 LC-TOF mass spectrometer. Samples were analyzed via a 1  $\mu$ L flow injection at 300  $\mu$ L/min in a water:methanol mixture (25:75 v/v) with 0.1% formic acid. The mass spectrometer was operated in positive ion mode with a capillary voltage of 4 kV, nebulizer pressure of 35 psi, and a drying gas flow rate of 12 L/min at 350°C. The fragmentor and skimmer voltages were 200 V and 60 V, respectively. Reference ions of purine at *m/z* 121.0509 and HP-0921 at *m/z* 922.0098 were simultaneously introduced via a second orthogonal sprayer, and used for internal calibration.

### **Coumarin lysine incorporation in human cells.**

Human embryonic kidney (HEK) 293T cells were seeded at ~50,000 cells per well and grown in DMEM (Dulbecco's Modified Eagle Medium, Gibco) supplemented with 10% FBS (Gibco), 1% Pen-Strep (Corning Cellgro) and 2 mM L-glutamin (Alfa Aesar) in 96-well plates (Costar) in a humidified atmosphere with 5% CO<sub>2</sub> at 37 °C. HEK 293T cells were transiently transfected with the pM**b**HCKRS-mCherry-TAG-EGFP-HA and p4CMVE-U6-PylT<sup>48</sup> at ~75% confluency in the presence or absence of coumarrin lysines (0.25 mM) in 96-well plates. Double transfections were performed with equal amounts of both plasmids (100 ng of each plasmid) using linear polyethylenimine (LPEI; 1.5  $\mu$ L, 0.323 mg/mL). After an overnight incubation at 37 °C, the cells were washed with PBS (200  $\mu$ L) at room temperature and imaged with a Zeiss Axio Observer

Z1 Microscope (10 $\times$  objective) using EGFP (38HE: Ex 470/40; Em 525/50) and mCherry (43HE: Ex 550/25; Em 605/70) filter cubes.

### **Western blots.**

Western blot was performed. HEK 293T cells (~500,000 cells per well) were co-transfected with pMbBHCKRS-mCherry-TAG-EGFP-HA and p4CMVE-U6-PyIT (1.5  $\mu$ g of each plasmid) using LPEI (15  $\mu$ L, 0.323 mg/mL) in the presence or absence of **BHCK**, **HCK**, and **HC<sub>2</sub>K** (0.25 mM) in 6-well plates. After 24 hrs of incubation, the cells were washed with chilled PBS (2 mL), lysed in mammalian protein extraction buffer (200  $\mu$ L, GE Healthcare) with complete protease inhibitor cocktail (Sigma) on ice, and the cell lysates were cleared at 13,200 rpm centrifugation (4 °C, 20 min). The protein lysate (16  $\mu$ L) was boiled with loading buffer and then analyzed by 10% SDS-PAGE. After gel electrophoresis and transfer to a PVDF membrane (GE Healthcare), the membrane was blocked in TBS with 0.1% Tween 20 (TBST) and 5% milk for 1 hr. The blots were probed and incubated with the primary antibody –  $\alpha$ -HA-probe (Y-11) rabbit polyclonal IgG (sc-805, Santa Cruz Biotech, 1:1000 dilution, 5  $\mu$ L in 5 mL TBST) overnight at 4 °C, followed by a fluorescent secondary antibody – goat- $\alpha$ -rabbit IgG Cy3 (GE Healthcare, 1:2000 dilution, 2.5  $\mu$ L in 5 mL TBST) for 1 hr at room temperature. The binding and washing steps were performed in TBS with 0.1% Tween 20. The Western blots were imaged using the Cy3 settings on a Typhoon Imager (GE Healthcare).

### **Protein sequencing by LC-MS/MS.**

HEK 293T cells were transfected with pBHCKRS-mCherry-TAG-EGFP-HA and p4CMVE-U6-

PyLT in a 10 cm petri dish and incubated with DMEM containing coumarin lysines **HCK**, **BHCK**, or **HC<sub>2</sub>K** (0.25 mM) for 24 h. Cells were lysed with extraction buffer (GE Healthcare) and the mCherry-**HCK/BHCK/HC<sub>2</sub>K**-EGFP-HA protein was immunoprecipitated using the Pierce HA Tag IP/Co-IP kit (Pierce) according to manufacturer's protocol. The proteins were separated on SDS-PAGE gels and stained with silver stain. Regions corresponding to the expected molecular weight of mCherry-EGFP-HA were excised, washed with HPLC water and destained with 50% acetonitrile/25 mM ammonium bicarbonate until no visible staining. Gel pieces were dehydrated with 100% acetonitrile, reduced with 10 mM dithiothreitol at 56 °C for 1 hr, followed by alkylation with 55 mM iodoacetamide at room temperature for 45 min in the dark. Gel pieces were then again dehydrated with 100% acetonitrile to remove excess alkylating and reducing agents, and rehydrated with 20 ng/μl trypsin/25 mM ammonium bicarbonate and digested overnight at 37 °C. The resultant tryptic peptides were extracted with 70% acetonitrile/5% formic acid, speed-vac dry and re-constituted in 18 μl 0.1% formic acid. Tryptic digests were analyzed by reverse-phased LC-MS/MS using a nanoflow LC (Waters nanoACQUITY UPLC system, Waters Corp., Milford, MA) coupled online to LTQ/Orbitrap Velos hybrid mass spectrometer (Thermo-Fisher, San Jose, CA). Separations were performed using a C18 column (PicoChip™ column packed with 10.5 cm ReproSil C18 3 μm 120Å chromatography media with a 75 μm ID column and a 15 μm tip, New Objective, Inc., Woburn, MA). Mobile phase A was 0.1% formic acid in water and mobile phase B was 0.1% formic acid in acetonitrile. Samples were injected onto a trap column (nanoACQUITY UPLC trap column, Waters Corp., Milford, MA) and washed with 1% mobile phase B at a flow rate of 5 μl/min for 3 min. Peptides were eluted from the column using a 90 min gradient running at 300 nl/min (5% B

for 3 min, 5-36% B in 62 min, 36-95% B in 2 min, 95% B for 8 min, 95%-5% B in 1 min, 5% B for 16 min). The LTQ/Orbitrap instrument was operated in a data-dependent MS/MS mode in which each high resolution broad-band full MS spectra ( $R = 60,000$  at mass to charge ( $m/z$ ) 400, precursor ion selection range of  $m/z$  300 to 2000) was followed by 13 MS/MS scans in the linear ion trap where the 13 most abundant peptide molecular ions dynamically determined from the MS scan were selected for tandem MS using a relative collision-induced dissociation (CID) energy of 35%. Dynamic exclusion was enabled to minimize redundant selection of peptides previously selected for CID. MS/MS spectra were searched with the MASCOT search engine (version 2.4.0, Matrix Science Ltd) against a UniProt jellyfish proteome database (June 2014 release) from the European Bioinformatics Institute (<http://www.ebi.ac.uk/integr8>) combined with endogenous mCherry-EGFP fasta sequences. The following modifications were used: static modification of cysteine (carboxyamidomethylation, +57.0214 Da) and variable modification of methionine (oxidation, +15.9949 Da) for all searches, variable modifications of lysine for mCherry-EGFP-HA (**HCK**, +218.17 Da; **BHCK**, +295.93 Da; **HC<sub>2</sub>K**, +231.03 Da). The mass tolerance was set at 20 ppm for the precursor ions and 0.8 Da for the fragment ions. Peptide identifications were filtered using PeptideProphet™ and ProteinProphet® algorithms with a protein threshold cutoff of 99% and peptide threshold cutoff of 95% implemented in Scaffold™ (Proteome Software, Portland, Oregon, USA).

### **Expression of caged firefly luciferase and light activation.**

HEK 293T cells were cultured in DMEM (Dulbecco's Modified Eagle Medium, Gibco) supplemented with 10% FBS (Gibco), 1% Pen-Strep (Gibco) and 2 mM L-glutamin (Alfa Aesar) in 96-well plates (BD Falcon) in a humidified atmosphere with 5% CO<sub>2</sub> at 37 °C. At 80-90%

confluence, cells seeded on plates were transfected and medium was changed to fresh DMEM supplemented without or with **HCK**, **BHCK**, **HC<sub>2</sub>K**, or **ACK** (0.25 mM). The plasmid p*Mb*BHCKRS-4PyIT was constructed containing both pCMV-*Mb*BHCKRS and p4CMVE-U6-PyIT. A TAG amber stop codon was introduced at the K206 site using primers GL1/GL2 and a QuikChange mutagenesis kit (Agilent Technologies). A pGL3-control plasmid containing the gene encoding *P. pyralis* firefly luciferase with the TAG amber mutation at residue K206 (pGL3-K206TAG) was co-transfected into cells with the plasmid pBHCKRS-4PyIT using linear PEI according to the manufacturer's protocol (Millipore). After double transfection and 24 h incubation, the medium was changed to DMEM without phenol red, and the cells were irradiated with UV light (365 nm) for 4 min using a 365-nm UV lamp (high performance UV transilluminator, UVP, 25 W) or kept in the dark. Cells were lysed by addition of 100 µL of substrate solution (Promega) in a 96-well plate (BD Falcon) and luminescence was measured on a Synergy 4 multi-mode microplate reader with an integration time of 2 s and a sensitivity of 150 or on a Tecan M1000 microplate reader with an integration time of 1 s.

#### **Visualization of nuclear localization through coumarin lysine incorporation.**

CHO K1 cells were plated into a poly-lysine coated 4-well chamber slide (Lab-Tek) and, after growing to 75% confluence, were transfected with 1 µg of pNLS-KTAG-EGFP and pBHCKRS-4PyIT each. After 16 hr incubation at 37 °C / 5% CO<sub>2</sub> in DMEM with 10% FBS in the presence of **HCK** (0.25 mM), cells were washed with DMEM without phenol red and then incubated for 2 hrs. The cells were washed with PBS, fixed with 4% formaldehyde, and stained with Rhodamine-phalloidin (Life Technologies). The chamber slide was dried in the dark overnight and cells were imaged on a Zeiss 710 confocal microscope (40× water objective).

### **One-photon light-activation of EGFP.**

HEK 293T cells were plated into a poly-D-lysine coated 8-well chamber slide (Lab-Tek). After growing to 70% confluence, cells were transfected with pEGFP-K85TAG-mCherry and pBHCKRS-4PyLT (200 ng each). After a 20 hr incubation at 37 °C / 5% CO<sub>2</sub> in DMEM with 10% FBS in the presence of **HCK** (0.25 mM), cells were washed with DMEM without phenol red, and then incubated for 1 hr. Before light activation, cells were focused using the TXRED channel, and imaged with a Nikon A1Rsi confocal microscope (20× objective, 2-fold zoom, EGFP (ex. 488 nm) and mCherry (ex. 560 nm) channels). Subsequently, cells were illuminated for 15 s at 365 nm light (DAPI filter, 358-365 nm), and then EGFP and mCherry fluorescence was acquired by real-time imaging (every 1 min for the first 15 min, every 5 min for the following 150 min, scan resolution 512×512, scan zoom 2×, dwell time 1.9 ms). The mean EGFP fluorescence intensities were quantified using Nikon Elements software.

### **Two-photon light-activation of EGFP.**

HEK 293T cells were plated into a poly-lysine coated μ-dish (ibidi) and after growing to 50% confluence were transfected with 1 µg each of pEGFP-KTAG-mCherry and pBHCKRS-4PyLT. After a 20 hr incubation at 37 °C / 5% CO<sub>2</sub> in DMEM with 10% FBS in the presence of **HCK** or **BHCK** (0.25 mM, 0.5% DMSO), cells were washed with DMEM without phenol red and then incubated for 1 hr. Cells were imaged with an Olympus Fluoview confocal microscope before two-photon irradiation (40× oil objective, EGFP (ex. 488 nm) and mCherry (ex. 560 nm) channels), imaging positions for mCherry-expressing cells were recorded, and the cell μ-dish was transferred to an Olympus multiphoton microscope for irradiation (Olympus Fluoview

FV1000 MPE). Cells were localized at the previously recorded positions, focused using the mCherry channel, and then irradiated using a 760 nm laser (130 mW, 5% of laser power, 30 cycles of scanning, 2  $\mu$ m/s dwell time, MaiTai DSBB-OL IR pulsed laser). After irradiation, the cell  $\mu$ -dish was transferred back to the original microscope for imaging.

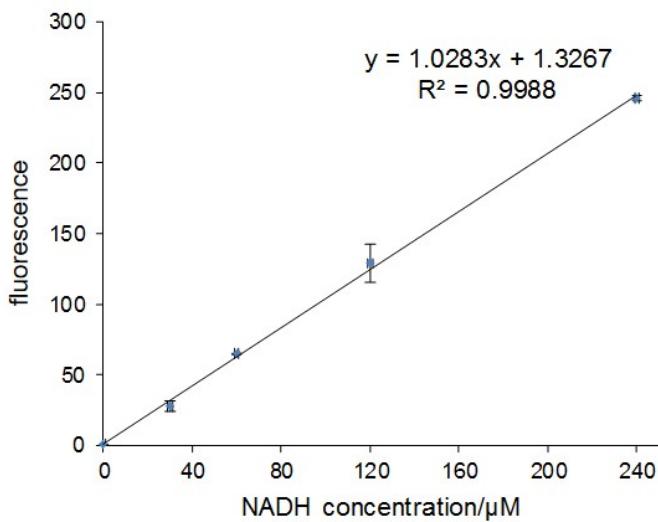
### **Light-activation of SATB1 translocation.**

HEK293T cells were plated at  $\sim$ 60,000 cells per well (400  $\mu$ L) into a poly-D-lysine-coated 8-well chamber slide (Lab-Tek). At  $\sim$ 75% confluence, cells were co-transfected with pEGFP-K29TAG-SatB1-mCherry and pMbBHCKRS-4PyLT (100 ng of each plasmid) using linear PEI (2  $\mu$ L, 0.323 mg/ml). After 20 hrs of incubation at 37 °C and 5% CO<sub>2</sub> in DMEM with 10% FBS in the presence of HCK (0.25 mM), the media was changed with fresh phenol-red-free DMEM (400  $\mu$ L). Before light irradiation, the cells were focused using the Texas Red channel, and imaged with a Nikon A1 confocal microscope ( $\times$ 40 oil objective, fluorescein isothiocyanate (ex = 488 nm) and Texas Red (ex = 560 nm) channels). After 30 s' irradiation using 405 nm LED light, EGFP and mCherry fluorescence were subsequently imaged every five minutes for 2 hrs (scan resolution 512  $\times$  512, dwell time 1.9 ms). The mean fluorescence intensities in the nucleus, in the case of the light activation of SatB1-mCherry were quantified using NIS Elements software to calculate  $F_{nt}/F_{nmax}$  (where  $F_{nt}$  is the fluorescence intensity in the nucleus over time) followed by normalization.

### **Fluorescence-based coupled ATPase activity assays.**

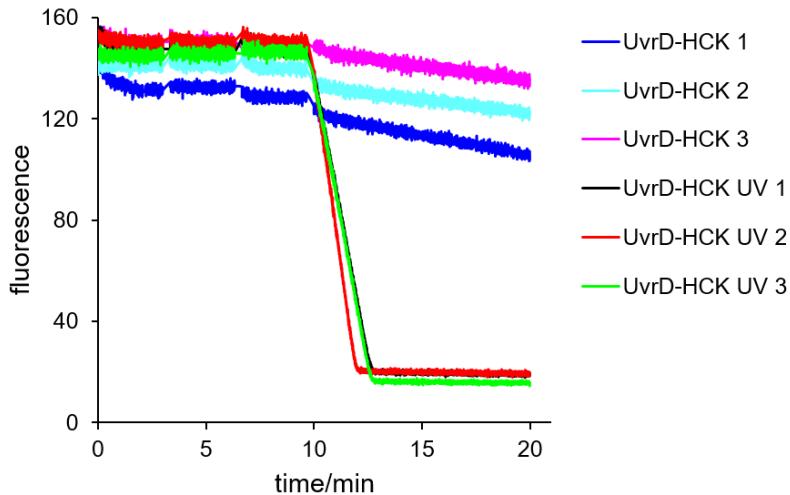
This ATPase assay design is based on the reaction of the conversion of phosphoenolpyruvate (PEP) to pyruvate by pyruvate kinase (PK), which is then coupled to the conversion of pyruvate

to lactate by lactate dehydrogenase (LDH).<sup>147</sup> In the first reaction, one ATP molecule is hydrolyzed by the ATPase to ADP, and then converted back to ATP by PK. In the second reaction, one NADH molecule is oxidized into a NAD<sup>+</sup> by LDH. A master mix containing 20 U/ml PK, 20 U/ml LDH, 2 mM PEP, 100 μM NADH in 1× reaction buffer (20 mM HEPES pH 7.5, 50 mM KCl, 3 mM MgCl<sub>2</sub>, 5 mM DTT) was prepared in a 400 μl quartz cuvette (Agilent). The concentration of NADH is monitored throughout the experiment by a Cary Eclipse Fluorescence Spectrophotometer (Agilent) via excitation at 384 ± 5 nm and emission at 461 ± 5 nm with a PMT voltage of 700 V at 37 °C. Data collection started with only the master mix in the absence of DNA, protein, or ATP. Fluorescence intensity was allowed to stabilize before 100 nM of ssDNA (39/20t, Table 3.2), 25 nM of protein, and 1 mM of ATP (final concentrations) were added sequentially. Typically, additions in each step were timed five minutes apart to allow the signal to stabilize again. ATPase assays were performed for wild-type UvrD, photocaged UvrD, decaged UvrD, and inactive UvrD-K37A.



**Figure 3.29:** NADH standard curve ( $\lambda_{\text{ex}} = 384 \pm 5 \text{ nm}$ ,  $\lambda_{\text{em}} = 461 \pm 5 \text{ nm}$ ).

$\Delta 1.028$  unit of fluorescence intensity equals to  $\Delta 1 \mu\text{M}$  NADH (ATP). Error bars represent standard deviations from three independent experiments.



**Figure 3.30:** Fluorescence-based coupled ATPase activity assay.

Time-course fluorescence measurement of NADH consumption driven by UvrD-**HCK** in the absence and presence of UV treatment (three independent experiments).

Based on the standard curve and fitting curve established in Figure 3.29 and Figure 3.30, the ATPase activities were calculated as shown in Table 3.2.

**Table 3.2:** ATP turnover number of UvrDs in ATPase assays.

Protein	ATP turnover number* ( $\text{s}^{-1}$ , mean $\pm$ s.d.)
wt-UvrD	$46.0 \pm 20.9$
UvrD-K37A	$2.6 \pm 1.3$
UvrD- <b>HCK</b> (+UV)	$41.2 \pm 14.6$
UvrD- <b>HCK</b> (-UV)	$2.0 \pm 0.5$

\*Error bars represent standard deviations (s.d.) from three independent experiments.

### **Fluorescence-based helicase activity assay.**

Quenched Y-shaped double-stranded DNA substrate (5 nM, final concentration) was added to the solution as a convenient reporter of the helicase UvrD activity. Such Y-shaped dsDNA substrate has a 5' FAM (fluorescein) labeled top strand, whose fluorescence is quenched by a 3' DAB (dimethylaminophenylazophenyl) label on the bottom strand until these two strands are separated by helicase activity of the protein (Figure 3.22A). Measurements started with monitoring total fluorescence intensity of 400 µL of reaction buffer in a cuvette, and continued throughout the rest of the experiment with excitation at 485±5 nm and emission at 520±5 nm wavelengths and PMT voltages at 750 V. After readings got stabilized, proteins (50 nM, final concentration) was added, and ATP (3 mM) was sequentially added to the mix to initiate the reaction.

### **Gel-based helicase activity assay.**

The gel-based assays were carried out with UvrD and nicked plasmid substrate that is not quenched in different conditions. Nicked double-stranded DNA with fluorescein labeled on the free 5' of the top strand (that is not quenched) is designed and used in the gel-based helicase assays. Reactions contain 5 nM substrate, 3 mM ATP, and same amount of protein incubated for different time periods. The two controls were used in the gel-based experiments, one is the DNA substrate only as a negative control, and the other one is the substrate after heating at 90 °C for 3 min as a positive control, both were stopped immediately after prepared. In the ATP-dependent experiment, all incubations of wild-type and photocaged UvrD were carried out at 37 °C for 30 min. All controls and reactions of 20 µL volumes were stopped by the addition of 10 µL of Stop Buffer containing 60 mM EDTA, loading dye, and 50× excess non-labeled chilled top strand to

prevent re-annealing of separated single strands. 10  $\mu$ L of each tube was loaded into lanes of a pre-cast 8% TBE gel, ran at 200 V for 40 min, imaged by a GE Typhoon 9400 scanner, and quantified by ImageJ. In addition, in the time course helicase assay experiments, reactions of wild-type UvrD and light-activated UvrD at the same concentration were both stopped at different time points (0, 15, 30, and 60 min) incubated at 37 °C.

## **4.0 GENETIC ENCODING OF FLUORESCENT AMINO ACIDS**

This chapter will describe the biological applications of two new optical probes, coumarin and phthalimide lysine analogs, through genetic code expansion. Specifically, protein localization was explored through coumarin lysine incorporation in live cells.

### **4.1 INTRODUCTION TO FLUORESCENT PROBES FOR PROTEINS**

The ability to selectively introduce fluorescent probes into proteins has greatly facilitated the study of protein structure, localization, dynamics, interaction and function both *in vitro* and *in vivo*. One widely used method involves the fusion of a fluorescent protein to the protein of interest;<sup>1-3</sup> however, this strategy has some limitations. First, the significant size of the fluorescent proteins (FPs, ~238 amino acids, >25 kDa) can interfere with the localization, structure, or activity of the proteins to which they are fused.<sup>2</sup> Second, this strategy is generally limited to placement of the fused fluorescent protein at either the C- or N-terminus. Furthermore, FPs have typical barrel structures which are highly packed to isolate the fluorescent chromophore inside from the cellular environment, therefore, this type of fluorescent probes are relatively insensitive to the local environment in living cells, such as hydrophobicity, pH, and ion concentrations.<sup>3</sup> Lastly, the brightness or photostability of FPs is not as good as some small

molecule fluorophores. Therefore, this strategy is not ideal for single-molecule fluorescence studies *in vivo*.

In addition to the use of fluorescent proteins, a variety of chemical and enzymatic strategies have also been developed to selectively modify proteins with various synthetic fluorophores, by employing the attachment of uniquely reactive unnatural amino acid (UAA) residues or short recognized peptide tags in the target proteins. For example, bioorthogonal conjugation is applied to uniquely reactive unnatural side chains. Enzymatic examples include the use of fluorescent probes that enzymatically label a genetically encoded target peptide tag or motif, such as sortase,<sup>154</sup> biotin ligase,<sup>155</sup> lipoic acid ligase,<sup>156</sup> farnesyltransferase,<sup>157</sup> SNAP-tag<sup>158-159</sup> and CLIP-tag,<sup>160</sup> FlAsH and ReAsH,<sup>161-162</sup> etc. These methods are limited by labeling selectivity and efficiency, and compatibility of the reaction of the synthetic fluorescent dye within living cells.

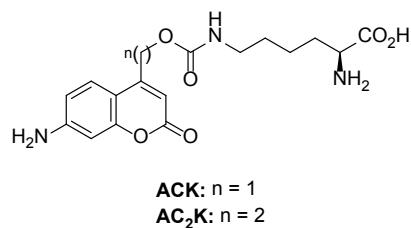
To circumvent the problems above, an approach to translationally incorporate a small UAA with a fluorescent side chain directly into the target protein has been developed to allow for a small fluorescent probe labeling at any site along the target protein sequence. To this end, an unnatural amino acid mutagenesis strategy has been developed to site-specifically incorporate amino acids containing 5-hydroxycoumarin, dansyl, and 6-acetylnaphthalen side chains using orthogonal tRNA/aminoacyl-tRNA synthetase (aaRS) pairs in response to the amber nonsense codon in *E. coli*, *S. cerevisiae*, and mammalian cells, respectively.<sup>163-165</sup>

## 4.2 GENETIC ENCODING OF FLUORESCENT COUMARIN LYSINES

### 4.2.1 Incorporation of amino-coumarin lysines

In order to improve the fluorescence characteristics of coumarin amino acids, aminocoumarin lysine (ACK) and aminocoumarinethyl lysine (AC<sub>2</sub>K) were synthesized by Dr. Subhas Samanta in our lab (Figure 4.1). In Schmidt's research, the absorption maximum ( $\lambda_{\text{max}}$ ) of 7-amino-substituted coumarin, reported as a caging group, showed a bathochromic shift (>20 nm) compared with 7-hydroxy and 7-methoxy coumarin derivatives.<sup>166</sup> Photolysis of this caging group was applied to caged neurotransmitters on auditory neurons in brainstem slices of mice. It is notable that 7-amino-substituted coumarin caging group is chemically stable and shows no intrinsic activity under physiological dark conditions, allowing for long-term biological studies.<sup>166</sup> The photochemical properties of coumarin lysine analogs were measured by Dr. Subhas Samanta (Table 4.1). The aminocoumarin lysine chromophore shows better quantum yield of fluorescence compared with hydroxycoumarin lysines (Table 4.1). In the same design, extension of the coumarin-carbamate linker by a single carbon atom results in the coumarin lysine **AC<sub>2</sub>K**, which does not undergo photolysis and thus represents a stable coumarin amino acid probe. **ACK** can be used as both fluorescent and light-activated probes for optochemical control of protein function using UV light, while **AC<sub>2</sub>K** serves as a stable fluorescent probe that does not decay under light excitation. One limitation of the 7-hydroxycoumarin fluorophore is that fluorescence is pH dependent. The phenolic hydroxyl group has a pKa of ~7.5,<sup>167</sup> and the

fluorophore is only emissive in its anionic form, thus it cannot be imaged in the acidic environments. In contrast to 7-hydroxycoumarin, 7-aminocoumarin, as an alternative coumarin structure, is not only uncharged at physiological pH, but also highly fluorescent at a wide range of pH values. Since it is sterically similar to 7-hydroxycoumarin, the same synthetase for incorporation of **HCK** could be used for genetically encoding of **ACK**.



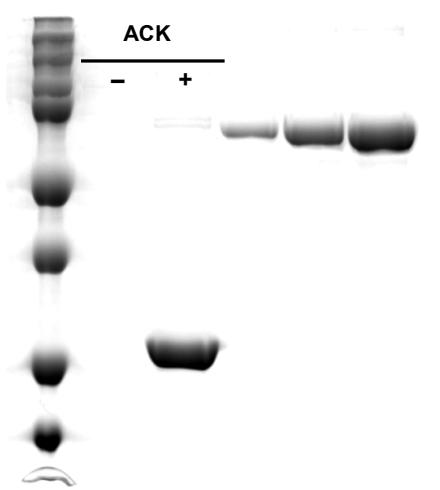
**Figure 4.1:** Chemical structure of **ACK** and **AC<sub>2</sub>K**.

**Table 4.1:** Summary of coumarin lysines.

UAA	$\lambda_{\text{ex}}$	absorption coefficient	$\lambda_{\text{em}}$	quantum yield of fluorescence
<b>HCK</b>	335 nm	6034 M <sup>-1</sup> cm <sup>-1</sup> (phosphate buffer, pH 7.5)	478 nm	$\Phi_{365\text{nm}} = 0.174$ (phosphate buffer, pH 7.5)
	370 nm	4331 M <sup>-1</sup> cm <sup>-1</sup> (phosphate buffer, pH 7.5)		
<b>HC<sub>2</sub>K</b>	335 nm	7517 M <sup>-1</sup> cm <sup>-1</sup> (phosphate buffer, pH 7.5)	450 nm	$\Phi_{365\text{nm}} = 0.899$ (phosphate buffer, pH 7.5)
	370 nm	6138 M <sup>-1</sup> cm <sup>-1</sup> (phosphate buffer, pH 7.5)		
<b>ACK</b>	365 nm	13793 M <sup>-1</sup> cm <sup>-1</sup> (DMSO)	450 nm	$\Phi_{365\text{nm}} = 0.900$ (DMSO)
	350 nm	10055 M <sup>-1</sup> cm <sup>-1</sup> (phosphate buffer, pH 7.5)	474 nm	$\Phi_{365\text{nm}} = 0.353$ (phosphate buffer, pH 7.5)
<b>AC<sub>2</sub>K</b>	365 nm	13103 M <sup>-1</sup> cm <sup>-1</sup> (DMSO)	430 nm	$\Phi_{365\text{nm}} = 0.910$ (DMSO)
	350 nm	9741 M <sup>-1</sup> cm <sup>-1</sup> (phosphate buffer, pH 7.5)	446 nm	$\Phi_{365\text{nm}} = 0.920$ (phosphate buffer, pH 7.5)

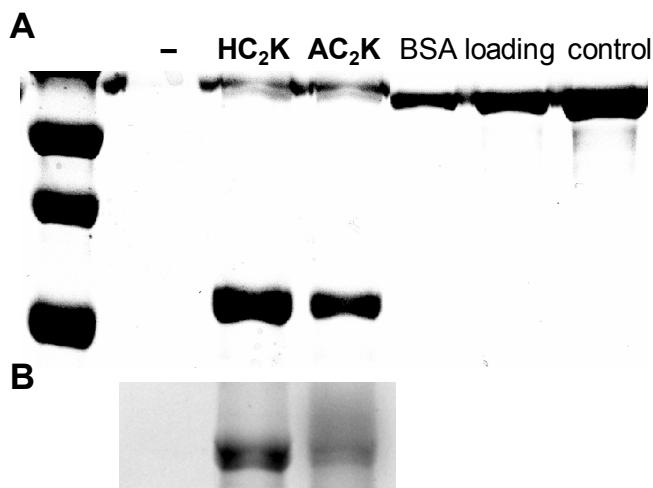
Given the very similar structures between hydroxy- and amino-coumarin lysines and previous observations of the high promiscuity of PylRS, the site-specific incorporation of amino-coumarin lysine analogues could be accomplished using the same BHCKRS. SDS-PAGE

analysis demonstrated that **ACK** and **AC<sub>2</sub>K** were incorporated in *E. coli* (Figure 4.2 and Figure 4.3) and revealed coumarin fluorescence of the expressed proteins containing the amino-coumarin lysines (Figure 4.3). The dependence of protein expression on the presence of **ACK** and **AC<sub>2</sub>K** demonstrates that the engineered BHCKRS synthetase has a high specificity for coumarin lysines, and does not significantly incorporate any of the common 20 amino acids.



**Figure 4.2:** SDS-PAGE analysis of sfGFP-Y151TAG containing **ACK** through incorporation in *E. coli*.

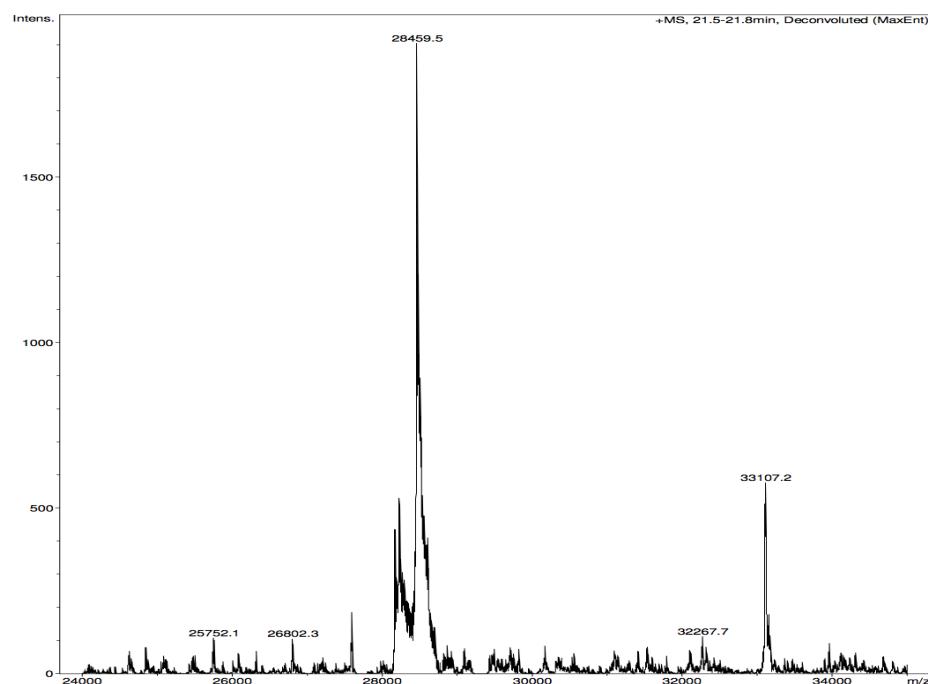
The gel was stained with Coomassie blue. The expression yield of sfGFP-**ACK** is 2.4 mg/L.



**Figure 4.3:** SDS-PAGE analysis of sfGFP-Y151TAG containing **HC<sub>2</sub>K** and **AC<sub>2</sub>K** through incorporation in *E. coli*.

A) Coomassie stained gel. B) The coumarin fluorescence was imaged using ChemiDoc via excitation at 365 nm.

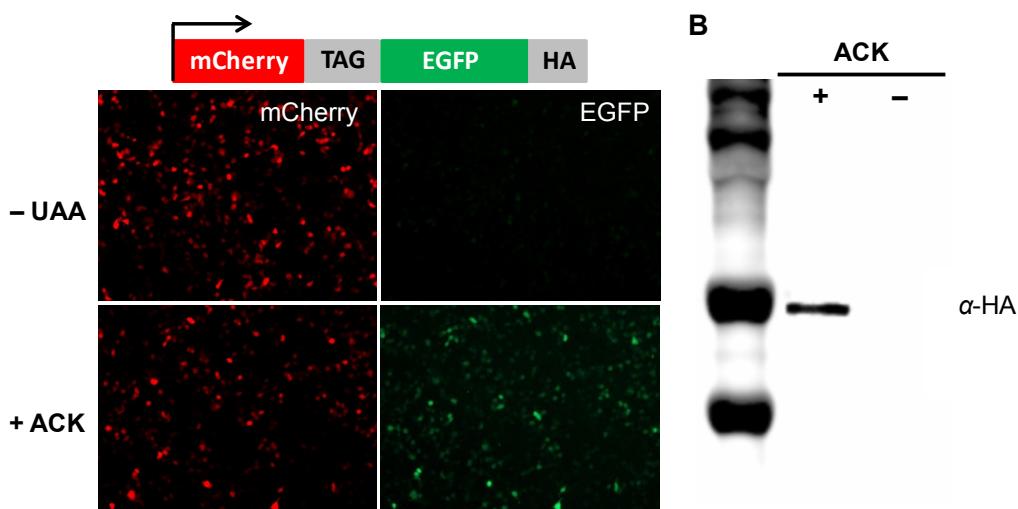
Electrospray ionization mass spectrometry (ESI-MS, Figure 4.4) showed that recombinantly expressed sfGFP-**AC<sub>2</sub>K** has a mass of 28459.5 Da, in agreement with the expected mass of 28460.03 Da. The incorporation yields of **AC<sub>2</sub>K** (0.8 mg/L) was significantly reduced in contrast with that of **HCK**. As for incorporation of the three hydroxycoumarin lysines, the Y271A and L274M mutations on BHCKRS greatly enlarge the binding pocket to accommodate the bulky bicyclic caging group, and orient it in a favorable  $\pi$ -stacking interaction with W382. This orientation also benefits from a favorable H-bond interaction between the coumarin hydroxyl group and D373. However, 7-amino substitute in **AC<sub>2</sub>K**, instead of 7-hydroxy group in hydroxycoumarin lysines, might have effect on the H-bonding, resulting in the decrease of incorporation yields.



**Figure 4.4:** ESI-MS analysis of sfGFP-**AC<sub>2</sub>K**.

Observed MS: 28459.5 Da, expected MS: 28460.03 Da.

To demonstrate the incorporation of aminocoumarin lysines into proteins in mammalian cells, pBHCKRS-mCherry-TAG-EGFP-HA and p4CMVE-U6-PylT were co-transfected into HEK 293T cells in the absence (as a negative control) or in the presence of **ACK** (0.25 mM) for 24 hr. Fluorescence imaging revealed EGFP expression only in the presence of **ACK**, indicating specific incorporation of the coumarin lysines without incorporation of any endogenous amino acid (Figure 4.5A). This was further confirmed by Western blot analysis of mammalian cell lysates from the same experiment with an anti-HA antibody (Figure 4.5B). These results demonstrate **ACK** can be selectively incorporated into proteins in mammalian cells in response to the amber codon (TAG).



**Figure 4.5:** Incorporation of **ACK** in mammalian cells.

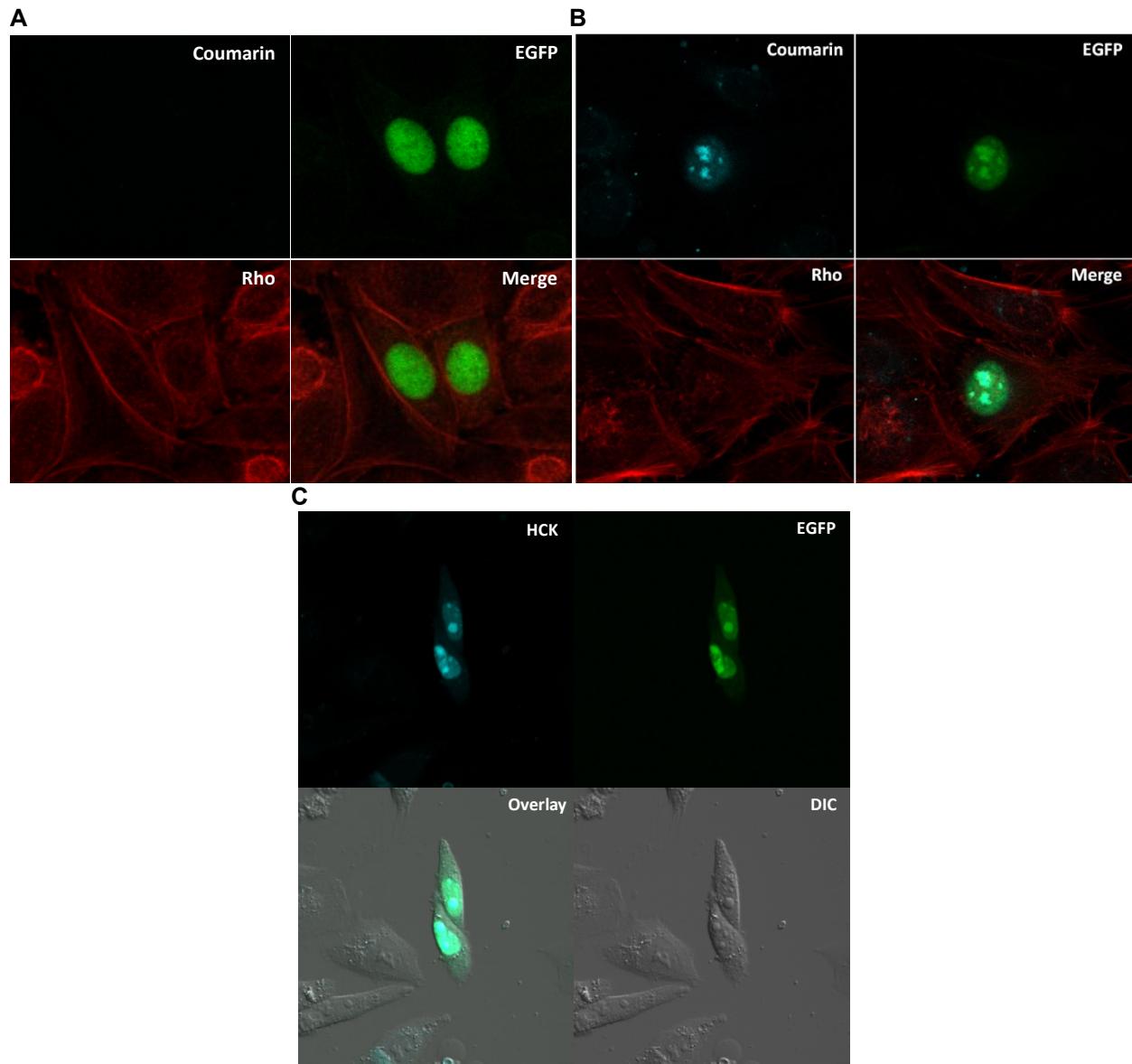
A) Fluorescence micrographs of HEK293T cells co-expressing the BHCKRS/tRNA<sub>CUA</sub> pair and mCherry\_TAG\_EGFP\_HA in the presence or absence of **ACK**. B) Western blot analysis of cell lysates using an anti-HA antibody.

#### 4.2.2 Fluorescent reporting of protein localization

This material was reprinted, in part, with permission from [Luo, J.; Uprety, R.; Naro, Y.; Chou, C.; Nguyen, D. P.; Chin, J. W.; Deiters, A. \*J Am Chem Soc\* 2014, 136\(44\), 15551-15558.](#)

In order to demonstrate the ability of the genetically encoded coumarin lysines to act as reporters for protein localization in live cells, they were first utilized as a protein nuclear localization marker. A plasmid was constructed to express EGFP-HA with an N-terminal wild-type NLS (PKKKRKV, NLS-linker-EGFP-HA).<sup>168-169</sup> The pNLS-linker-EGFP plasmid was designed to reliably localize EGFP to the nucleus, a phenotype that can be easily verified using confocal fluorescent microscope (Figure 4.6A, wild type).

EGFP was used as not only an incorporation reporter but also a control signal for coumarin comparison. A TAG amber codon was introduced in the linker between the NLS and EGFP, which allows for site-specific unnatural amino acid incorporation without affecting EGFP formation or nuclear translocation. To optimize the conditions for protein localization, HeLa, and CHO K1 cells were co-transfected with the reporter plasmid (pNLS-KTAG-EGFP) and PyLBHCKRS-4PyLT in the presence of **HCK** (0.25 mM). After a 16 hr incubation, the cells were washed with PBS, and analyzed for coumarin fluorescence (405 nm excitation, 450-480 nm emission) and EGFP fluorescence (488 nm excitation, 490-520 nm emission) by confocal microscopy, which was set up based on the absorbance and emission spectra of coumarin lysines (Figure 4.6B and Figure 4.6C). The observation of complete co-localization of both fluorophores in the nucleus (merged micrographs) demonstrates the ability to use **HCK** as a reporter of protein localization (Figure 4.6).



**Figure 4.6:** Nuclear localization of NLS-wt/HCK-EGFP protein in mammalian cells.

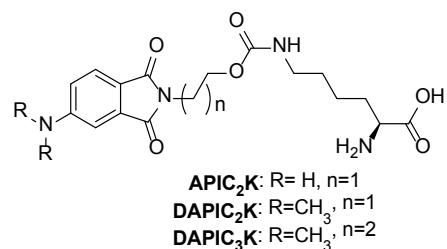
A) Nuclear localization of NLS-wt-EGFP protein in fixed CHO K1 cells with Rhodamin-phalloidine staining. Representative confocal micrographs showing the EGFP nuclear localization. B) Fluorescent nuclear localization by confocal microscopy images of HeLa cells co-transfected with pNLS-TAG-EGFP-HA and pBHCKRS-4PyIT in 0.25 mM **HCK** with Rhodamin-phalloidine staining. C) Nuclear localization of NLS-**HCK**-EGFP protein in live CHO K1 cells. Images were visualized using coumarin (Ex 405 nm) or EGFP (Ex 488 nm) or TXRED (Ex 561 nm)

channels and 40× air objective by Zeiss laser scanning confocal microscope. Representative confocal micrographs showing the coumarin fluorescent nuclear localization of NLS-**HCK**-EGFP fusions.

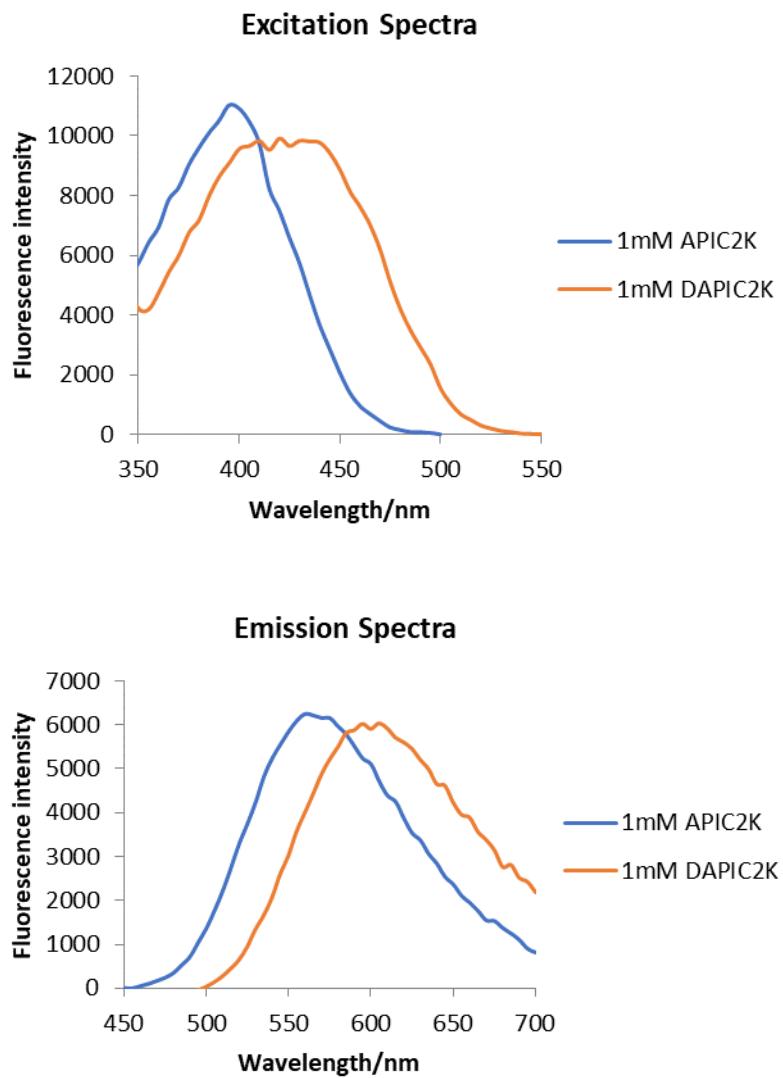
### 4.3 GENETIC ENCODING OF PHTHALIMIDE LYSINES

#### 4.3.1 Introduction to phthalimide lysines

The site-specific incorporation of a fluorescently labeled unnatural amino acid not only enables to visualize the precise location of a protein in live cells, but also to probe protein folding and function. Using unnatural amino acid mutagenesis, the fluorescent labeling of receptors, antibodies, and enzymes at specific positions has been achieved.<sup>170</sup> The aminophthalimide fluorophore was selected, in order to take advantage of its small size as well as its emission at longer wavelengths (500–600 nm). Since phthalimide is a solvatochromic fluorophore and its fluorescence intensity is dependent on the polarity of the environment, it can potentially be used as a cellular sensor.<sup>171</sup> Therefore, three phthalimide lysine analogues, including 4-aminophthalimide lysine (**APIC<sub>2</sub>K**), 4-(*N,N*-dimethylamino)phthalimide lysine (**DAPIC<sub>2</sub>K**), and 4-(*N,N*-dimethylamino)phthalimide lysine (**DAPIC<sub>3</sub>K**), were synthesized by Dr. Rajendra Uprety in the Deiters lab (Figure 4.7). The difference between **DAPIC<sub>2</sub>K** and **DAPIC<sub>3</sub>K** is just the length of the linker (extra CH<sub>2</sub>) between lysine and the fluorophore.



**Figure 4.7:** Chemical structures of **APIC<sub>2</sub>K**, **DAPIC<sub>2</sub>K**, and **DAPIC<sub>3</sub>K**.

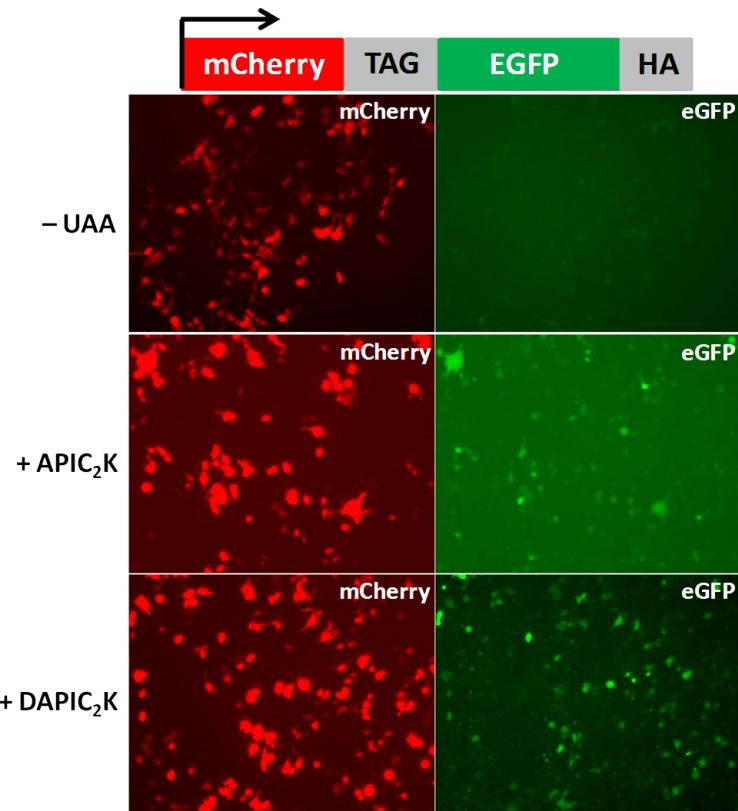


**Figure 4.8:** Excitation ( $E_m = 550$  nm) and emission ( $E_x = 400$  nm) spectra of **APIC<sub>2</sub>K** and **DAPIC<sub>2</sub>K** in PBS buffer (pH 7.4).

## 4.3.2 Incorporation of phthalimide lysines

### 4.3.2.1 Incorporation of phthalimide lysines in mammalian cells

After screening the synthetase panel, an optimal *MbPylRS* for incorporation of **APIC<sub>2</sub>K** was found. It contains three mutations, L274V, C313V, and M315Q, and was termed *MbAPIKRS*. Based on the structural similarity of phthalimide lysines, *MbAPIKRS* is sufficient for the incorporation of both **APIC<sub>2</sub>K** and **DAPIC<sub>2</sub>K**. In order to determine the incorporation of the phthalimide lysine fluorophores in mammalian cells, HEK 293T cells were co-transfected with p*MbAPIKRS*-mCherry-TAG-EGFP and p4CMVE-U6-PylT in the absence and presence of **APIC<sub>2</sub>K** and **DAPIC<sub>2</sub>K** (0.5 mM). Without the UAAs, only mCherry was expressed, while both mCherry and EGFP fluorescence was observed upon the addition of **APIC<sub>2</sub>K** and **DAPIC<sub>2</sub>K** (Figure 4.9). Therefore, the results demonstrate that the two phthalimide lysines were successfully incorporated in eukaryotes.

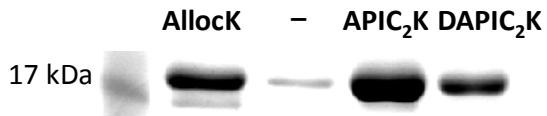


**Figure 4.9:** Genetic incorporation of **APIC<sub>2</sub>K** and **DAPIC<sub>2</sub>K** in mammalian cells.

Fluorescent micrographs of HEK 293T cells expressing pPyRS3-mCherry-TAG-EGFP and PytRNA<sub>CUA</sub> in the absence and presence of **APIC<sub>2</sub>K** and **DAPIC<sub>2</sub>K** (0.5 mM).

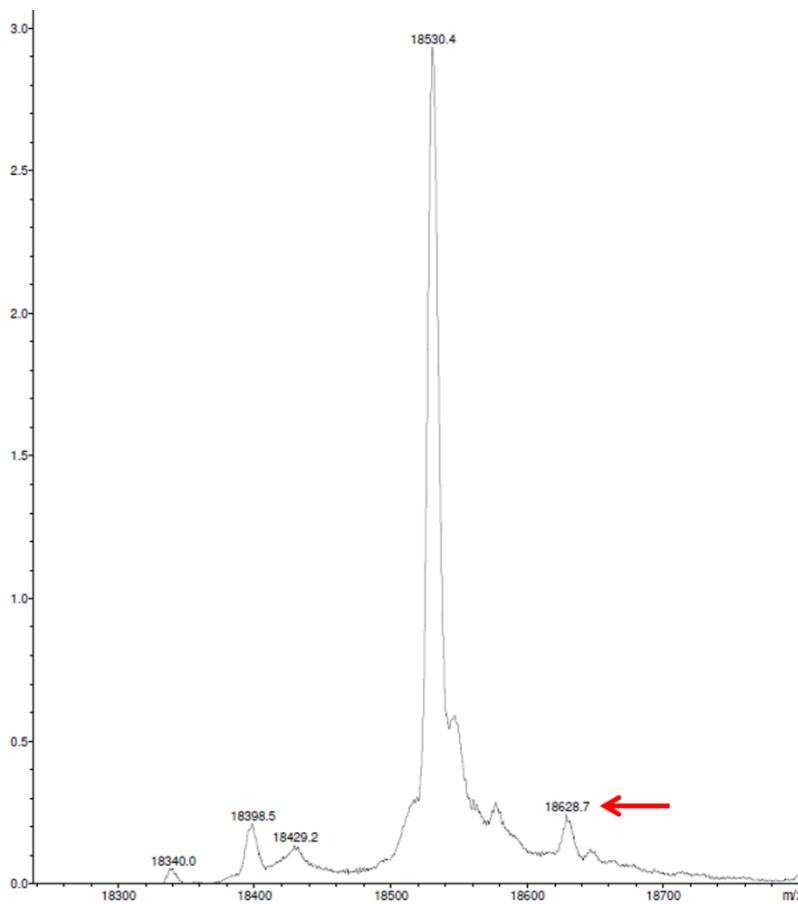
#### 4.3.2.2 Incorporation of phthalimide lysines in *E. coli*

In order to demonstrate the incorporation of **APIC<sub>2</sub>K** and **DAPIC<sub>2</sub>K** in bacteria, Myo-**APIC<sub>2</sub>K** and -**DAPIC<sub>2</sub>K** were expressed in *E. coli*. The purified Myo-**APIC<sub>2</sub>K** and -**DAPIC<sub>2</sub>K** were analyzed by 12% SDS-PAGE (Figure 4.10).



**Figure 4.10:** A 12% SDS-PAGE coomassie gel of Myo-**APIC<sub>2</sub>K** and Myo-**DAPIC<sub>2</sub>K**.

In order to confirm incorporation of phthalimide lysines, ESI-MS analyses were performed. The ESI-MS results showed a predominant peak of 18530.4 Da, which is 98 Da lower than the theoretical mass (18628 Da). This discrepancy in mass can be rationalized in two ways: 1) the amino acid is incorporated as **APIC<sub>2</sub>K** and undergoes degradation after incorporation due to the aqueous environment, or 2) the amino acid is degraded in *E. coli* and the degradation product is then incorporated. A small peak (Figure 4.11, peak indicated in red) has the expected mass for Myo-**APIC<sub>2</sub>K**; however, the signal intensity is much lower than the degraded peak. While **APIC<sub>2</sub>K** appears to be incorporated to some extent, the efficiency is very low (Figure 4.11). Interestingly, in the chemistry side, when **APIC<sub>2</sub>K** is maintained in DMSO, the expected mass is observed; however, when dissolved in water (analogous to expression conditions), a degraded byproduct is observed over time. Therefore, the additional studies need to be conducted in order to improve the stability of **APIC<sub>2</sub>K** and **DAPIC<sub>2</sub>K** such that they can be incorporated intact in cells.

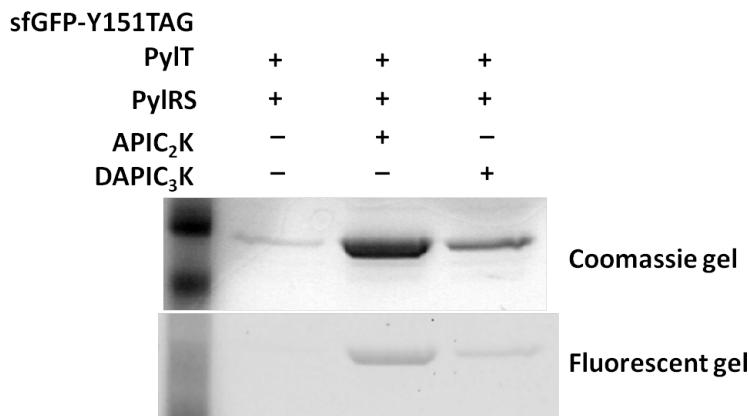


**Figure 4.11:** ESI-MS analysis of Myo-APIC<sub>2</sub>K.

Observed mass: 18530.4 Da, expected mass: 18628 Da.

To further investigate the in-gel fluorescence of phthalimide lysines, sfGFP-APIC<sub>2</sub>K and -DAPI<sub>C3</sub>K were expressed in *E. coli*. After Ni-NTA purification, sfGFP-APIC<sub>2</sub>K and -DAPI<sub>C3</sub>K were analyzed by 12% SDS-PAGE (Figure 4.12 top). Upon visualization, in-gel fluorescence was not observed under UV light (365 nm, see the spectra in Figure 4.8). Subsequently, the gel was imaged using a GE Typhoon scanner at 473 nm and in-gel fluorescence of sfGFP-APIC<sub>2</sub>K and -DAPI<sub>C3</sub>K was detected (Figure 4.12 bottom). The fluorescent bands detected were weak because the available filter sets do not contain the optimal wavelength for excitation. The detected fluorescence in the protein derived from the APIC<sub>2</sub>K

and **DAPIC<sub>3</sub>K** fluorescence, since the sfGFP was denatured on the SDS-PAGE and should no longer produce fluorescence. In centrifuge tubes of sfGFP-**APIC<sub>2</sub>K** and -**DAPIC<sub>3</sub>K**, the fluorescence of **APIC<sub>2</sub>K** and **DAPIC<sub>3</sub>K** could not be differentiated from the intrinsic green fluorescence of sfGFP. In addition, the protein expression yield of sfGFP-**APIC<sub>2</sub>K** was 6.0 µg/mL, which is comparable to the Myo-**APIC<sub>2</sub>K** expression yield of 11.6 µg/mL (Figure 4.10). Therefore, myoglobin would be a better model protein for incorporation of the phthalimide fluorophores compared with sfGFP.

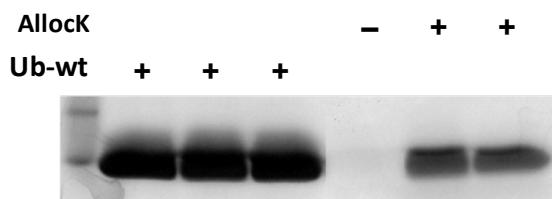


**Figure 4.12:** A 12% SDS-PAGE Coomassie stained gel (top) and fluorescent gel (bottom) of sfGFP-**APIC<sub>2</sub>K** and -**DAPIC<sub>3</sub>K**.

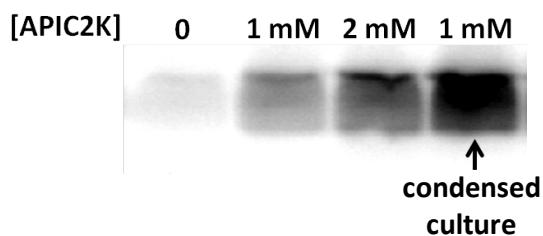
The fluorescent gel (bottom) was imaged at 473 nm (CsY2 filter) using a GE Typhoon scanner.

Ubiquitin (Ub) is a small non-fluorescent, non-colored protein and should not contribute to the background of a fluorescent assay, which is a limitation to using sfGFP and myoglobin. Therefore, Ub was used for incorporation of **APIC<sub>2</sub>K** and **DAPIC<sub>2</sub>K**. In order to improve the protein expression, different conditions were tested, including varying the UAA concentration and using a condensed cell culture method. wt-Ub, Ub-**AllocK**, and Ub-**APIC<sub>2</sub>K** were analyzed by 12% SDS-PAGE (Figure 4.13). Based on the purification conditions used, adding **APIC<sub>2</sub>K** (1

mM) using the condensed cell culture is the optimal expression condition (Figure 4.14). However, in-gel fluorescence was not detected under UV light. The fluorescence of Ub-**APIC<sub>2</sub>K** and -**DAPIC<sub>2</sub>K** was then analyzed by a fluorescence assay using a microplate reader. Based on the excitation and emission spectra of **APIC<sub>2</sub>K** and **DAPIC<sub>2</sub>K** (Figure 4.8), the fluorescence assay was performed using an excitation of 420 nm and an emission of 595 nm (Figure 4.15). Compared to the Ub-**AllocK** control, the fluorescence intensity of Ub-**APIC<sub>2</sub>K** and **DAPIC<sub>2</sub>K** is 7.3- and 6.9-fold higher than that of the control, respectively (Figure 4.15). The fluorescence increase is not very prominent, indicating that the fluorescence detected from **APIC<sub>2</sub>K**- and **DAPIC<sub>2</sub>K**-labeled ubiquitin was weak.

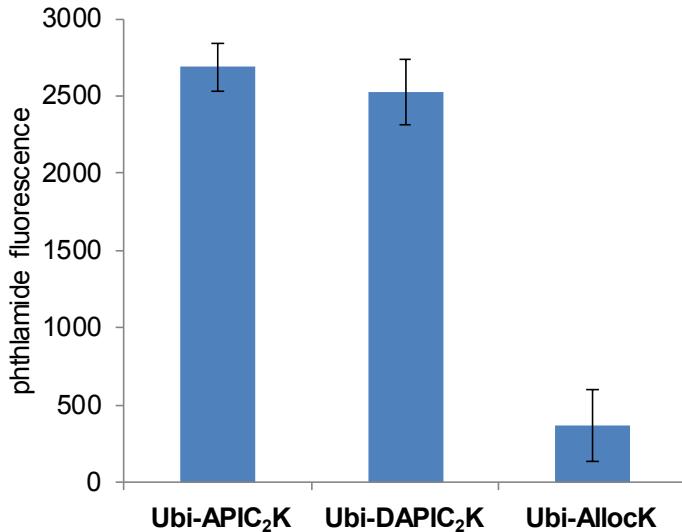


**Figure 4.13:** A 12% SDS-PAGE gel of Ub-wt and Ub-AllocK.



**Figure 4.14:** SDS-PAGE analysis for Ub-**APIC<sub>2</sub>K** expression.

Optimization of Ub-**APIC<sub>2</sub>K** expression in different conditions.



**Figure 4.15:** Fluorescence analysis of Ub-APIC<sub>2</sub>K and Ub-DAPIC<sub>2</sub>K in elution buffer compared to the control of Ub-AllocK ( $E_x = 420$  nm,  $E_m = 595$  nm).

#### 4.4 CONCLUSIONS

In summary, the site-specific genetic incorporation of new coumarin lysine analogs (**ACK** and **AC<sub>2</sub>K**) in *E. coli* and mammalian cells. The incorporation of phthalimide lysine analogs (**APIC<sub>2</sub>K**, **DAPIC<sub>2</sub>K**, and **DAPIC<sub>3</sub>K**) into proteins is not convincingly demonstrated in *E. coli* due to the stability of UAAs in aqueous environments. Coumarin lysine **HCK** has been applied to detect nuclear protein localization, demonstrating that the fluorescence of coumarin lysine **HCK** is bright enough to be useful as a cellular probe. The small size of coumarin lysines is expected to minimally perturb protein structure, function, and localization. **ACK** and **AC<sub>2</sub>K** could be further developed as more stable and brighter fluorescent probes for protein labeling in living cells. **ACK** as a stable caging group could be well suited for long-term biological studies in live

model organisms or animals. The ability to incorporate all five coumarin lysines with the same PylRS/tRNA<sub>CUA</sub> pair further facilitates their biological applications.

## 4.5 EXPERIMENTAL

### Plasmid cloning.

(1) Construction of pNLS-TAG-EGFP-HA: The pTAG-EGFP-HA fragment was amplified from pmCherry-TAG-EGFP-HA using the primers G1/G2, and digested with *Hind*III and *Bg*II. Digested product was then ligated with the pEGFP-N1 (Clontech) backbone, generating the pTAG-EGFP-HA plamid (Table 4.2). The pNLS PCR fragment was obtained by using primers N1/N2 to assemble and then ligated into the *Hind*III and *Xba*I sites of pTAG-EGFP-HA to generate the pNLS-TAG-EGFP-HA plasmid (Table 4.2).

(2) Construction of pNLS-WT-EGFP-HA: Plasmids were obtained by converting the TAG codon of pNLS-TAG-EGFP-HA into an AAG (Lys) codon using primers QC1/QC2 and a QuikChange site-directed mutagenesis kit (Agilent, Table 4.2).

**Table 4.2:** Primer list.

Primer No.	Sequences
G1	gtcctataacaagcttcgatggcttaccatggcactag
G2	gatcttagagtccgcgcggatccttatcattaagcg
N1	gactcagatcaattactcgagatggccaccgtcctgaagcgccccgtgccacaaaaaggctggccaggcca aa
N2	tcgaagcttgagctcgctgagtcggacttgtacagtgaagcgtccagcttctttggccctggccagccaaaa gg
QC1	gaaggcttaccatggcacaagcaattagccatggtag
QC2	ctcaccatggctaattgttgtccatggtaaggcttc
GL1	gatctactggctgccttaggggtcgctgcctcatag
GL2	ctataggcagagcgacaccctaaggcagaccagtagatc

### **Expression and purification of proteins in *E. coli*.**

The plasmid, pBAD-sfGFP-Y151TAG-pylT was co-transformed with pBK-BHCKRS<sup>48</sup> into *E. coli* Top10 cells. A single colony was grown in LB media overnight and 500 µL of the overnight culture was added to 25 mL of LB media, supplemented with 1 mM of the designated unnatural amino acid and 25 µg/mL of tetracycline and 50 µg/mL of kanamycin. Cells were grown in the presence of coumarin lysine derivatives at 37 °C, 250 rpm, and protein expression was induced with 0.1% arabinose when the OD<sub>600</sub> reached ~0.6. After overnight expression at 37 °C, cells were harvested and washed by PBS. The cell pellets were re-suspended in 6 mL of phosphate lysis buffer (50 mM, pH 8.0) and Triton X-100 (60 µL, 10%), gently mixed, and incubated for an hour at 4 °C. The cell mixtures were sonicated, and then the cell lysates were centrifuged at 4 °C, 13,000 g, for 10 minutes. The supernatant was transferred to a 15 mL conical tube and 100 µL of Ni-NTA resin (Qiagen) was added. The mixture was incubated at 4 °C for 2 hours under mild shaking. The resin was then collected by centrifugation (1,000 g, 10 min), washed twice with 400 µL of lysis buffer, followed by two washes with 400 µL of wash buffer containing 20 mM imidazole. The protein was eluted with 400 µL of elution buffer containing 250 mM imidazole. The purified proteins were analyzed by 10% SDS-PAGE and stained with Coomassie Blue.

### **Coumarin lysine incorporation in human cells.**

HEK 293T cells were grown in DMEM (Dulbecco's Modified Eagle Medium, Gibco) supplemented with 10% FBS (Gibco), 1% Pen-Strep (Corning Cellgro) and 2 mM L-glutamin (Alfa Aesar) in 96-well plates (Costar) in a humidified atmosphere with 5% CO<sub>2</sub> at 37 °C. HEK 293T cells were transiently transfected with the pMbBHCKRS-mCherry-TAG-EGFP-HA and

p4CMVE-U6-PylT<sup>48</sup> at ~75% confluence in the presence or absence of coumarrin lysines (0.25 mM) in 96-well plates. Double transfections were performed with equal amounts of both plasmids. After an overnight incubation at 37 °C, the cells were washed by PBS and imaged with a Zeiss Axio Observer.Z1 Microscope (10× objective).

### **Visualization of nuclear localization through coumarin lysine incorporation.**

CHO K1 cells were plated into a poly-lysine coated 4-well chamber slide (Lab-Tek) and, after growing to 75% confluence, were transfected with 1 µg of pNLS-KTAG-EGFP and pBHCKRS-4PylT each. After 16 hr incubation at 37 °C / 5% CO<sub>2</sub> in DMEM with 10% FBS in the presence of **HCK** (0.25 mM), cells were washed with DMEM without phenol red and then incubated for 2 hrs. The cells were washed with PBS, fixed with 4% formaldehyde, and stained with Rhodamine-phalloidin (Life Technologies). The chamber slide was dried in the dark overnight and cells were imaged on a Zeiss 710 confocal microscope (40× water objective).

### **Incorporation of APIC<sub>2</sub>K and DAPIC<sub>2</sub>K into myoglobin.**

In order to demonstrate the incorporation of **APIC<sub>2</sub>K** and **DAPIC<sub>2</sub>K** in bacteria, Myo-**APIC<sub>2</sub>K** and -**DAPIC<sub>2</sub>K** were expressed in *E. coli*. The plasmids, pBAD-Myo-S4TAG-PylT and pBK\_PylRS-L274V\_C313V\_M315Q, were co-transformed into *E. coli* Top10 competent cells and plated on LB agar containing tetracycline (25 µg/mL) and kanamycin (50 µg/mL). A colony was picked and grown in LB media (3 mL) overnight in a shaker at 37 °C. The cells from the 3 mL culture were then inoculated into 25 mL of LB media containing tetracycline and kanamycin for protein expression and were grown until they reached an OD<sub>600</sub> ~0.6. The cell culture was induced with arabinose (0.2%) and was incubated overnight. The cells were collected by

centrifuge (5,000 rpm, 10 min) and then stored at -80 °C until purification was performed. After Ni-NTA purification, Myo-**APIC<sub>2</sub>K** and -**DAPIC<sub>2</sub>K** were analyzed by 12% SDS-PAGE (Figure 4.10). However, in-gel fluorescence was not detected upon exposure to UV light (365 nm). The purified protein solutions were stored in centrifuge tubes and were excited under UV light. The fluorescence in the tubes of Myo-**APIC<sub>2</sub>K** and Myo-**DAPIC<sub>2</sub>K** was compared to the control protein solution of Myo-Alloc (Figure 4.10). After comparing Myo-**APIC<sub>2</sub>K** and -**DAPIC<sub>2</sub>K**, it appears myoglobin (fluctuation domains) may quench the fluorescence of the phthalimides, because myoglobin intrinsically contains red color and a heme center.<sup>172</sup> In addition, 365 nm light is not the excitation maxima of **APIC<sub>2</sub>K** ( $\lambda_{\text{max}} \sim 395$  nm) and **DAPIC<sub>2</sub>K** ( $\lambda_{\text{max}} \sim 420$  nm), which could also explain why the in-gel fluorescent signal was not detected upon exposure to UV light.

#### **Incorporation of APIC<sub>2</sub>K and DAPIC<sub>3</sub>K into sfGFP.**

To further investigate the in-gel fluorescence of phthalimide lysines, sfGFP-**APIC<sub>2</sub>K** and -**DAPIC<sub>3</sub>K** were expressed in *E. coli*. The plasmids, pBAD\_sfGFP-Y151TAG-PylT and pBK\_PylRS-L274V\_C313V\_M315Q, were transformed into Top10 competent cells, and plated on LB agar containing tetracycline (25 µg/mL) and kanamycin (50 µg/mL). A colony was picked and grown in LB media (3 mL) overnight in a shaker (250 rpm, 37 °C). The cells from the starter culture were inoculated into 25 mL of LB media containing tetracycline and kanamycin with **APIC<sub>2</sub>K** and **DAPIC<sub>3</sub>K** (1 mM) for protein expression and were grown until an optical density of OD<sub>600</sub> ~0.6 was achieved. The cell culture was induced with arabinose (0.2%) and incubated overnight. The cells were collected by centrifugation (5,000 rpm, 10 min) and then stored at -80 °C until the purification was performed. After Ni-NTA purification, sfGFP-**APIC<sub>2</sub>K** and -

**DAPIC<sub>3</sub>K** were analyzed by 12% SDS-PAGE. Subsequently, the gel was imaged using a GE Typhoon scanner at 473 nm and in-gel fluorescence of sfGFP-**APIC<sub>2</sub>K** and -**DAPIC<sub>3</sub>K** was detected (Figure 4.12 bottom).

### **Incorporation of APIC<sub>2</sub>K and DAPIC<sub>2</sub>K into ubiquitin.**

First, Ub-wt and the Ub-K48TAG gene (Cropp lab, Virginia Commonwealth University) were inserted into the PyltRNA backbone to construct the plasmids pBAD-Ub-wt and pBAD-Ub-K48TAG, with the latter plasmid containing the TAG mutation for the incorporation of UAAs into ubiquitin. Using the newly constructed plasmids, wild type Ub and Ub-**AllocK** (control protein) were successfully expressed and purified in *E. coli* (Figure 4.13). Subsequently, in order to express Ub-**APIC<sub>2</sub>K** and **DAPIC<sub>2</sub>K**, the plasmid pBAD-Ub-K48TAG and pBK-PylRS3 were transformed into Top10 competent cells and plated on LB agar containing tetracycline (25 µg/mL) and kanamycin (50 µg/mL) as previously described. In order to improve the protein expression, different conditions were tested, including varying the UAA concentration and using a condensed cell culture method. After an overnight expression, the cells were collected by centrifuge (5,000 rpm, 10 min). After Ni-NTA purification, Ub-**APIC<sub>2</sub>K** was analyzed by 12% SDS-PAGE (Figure 4.13).

## 5.0 GENETIC ENCODING OF CAGED CYSTEINES

This chapter will describe the genetic incorporation of photocaged cysteines and photocaged homocysteine into proteins in bacterial and mammalian cells. Importantly, the genetic replacement of an active site cysteine with a caged cysteine for the regulation of enzyme function was demonstrated in live cells. These discoveries provide a new tool for the optochemical control of protein function in mammalian cells and expand the set of genetically encoded unnatural amino acids.

### 5.1 INTRODUCTION TO CAGED CYSTEINES

This material was reprinted, in part, with permission from [Uprety, R.; Luo, J.; Liu, J.; Naro, Y.; Samanta, S.; Deiters, A. \*Chembiochem\* 2014, 15\(36\), 13433-13439.](#)

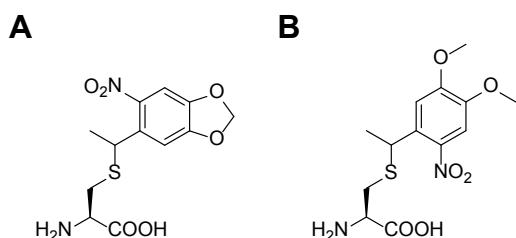
Cysteine is the least abundant amino acid in proteins, however, its importance is underlined by its presence in almost all proteins<sup>54</sup> and by the unique chemistry of its thiol group in cellular functions.<sup>55-58</sup> Active site cysteines have been found in proteases, kinases, metabolic enzymes, and transcription factors.<sup>39</sup> The diverse functional roles of cysteine include ubiquitination,<sup>59-60</sup> caspase-mediated apoptosis,<sup>61-62</sup> maintenance of cellular redox potential,<sup>63</sup> defense against oxidative stress,<sup>64, 67, 70, 73-74</sup> contribution to redox signaling pathways,<sup>75, 77</sup> as well as protein trafficking and localization.<sup>76, 78-80</sup> The chemistry of the cysteine side chain is

accountable for the aforementioned broad range of functional roles.<sup>81</sup> Intrinsic properties of the thiol group, such as the large atomic radius of sulfur, environmental sensitivity of the pKa, variable oxidation states,<sup>39, 64</sup> and metal chelating properties offer significant opportunities to accomplish the diverse functions of proteins.<sup>82</sup> Cysteine is also involved in various post-translational modifications, including alkylation, nitrosylation, and oxidation.<sup>39</sup> Furthermore, the thiol group can be oxidized to a disulfide bond in a dynamic and reversible fashion.<sup>63</sup> The cysteine-glutathione disulfide coupled system, a thiol-based redox switch, is a key component for the regulation of certain genes.<sup>57</sup> In addition, the formation of a disulfide bond between two cysteine residues results in intra- or intermolecular linkages that can induce substantial changes in protein folding, thereby altering protein function.<sup>83</sup>

Homocysteine, a homologue of cysteine, is a metabolite of the amino acid methionine to possess a thiol group, which is linked to an elevated risk of vascular disease,<sup>84</sup> and hence is being investigated as a biomarker for various cardiovascular diseases<sup>85</sup> and as a potential risk factor of Alzheimer's disease.<sup>86</sup> In contrast to cysteine it is not translationally incorporated into proteins, but can be introduced via *N*-homocysteinylation<sup>87</sup> and *S*-homocysteinylation,<sup>88</sup> which have been proposed as factors for certain diseases.<sup>89</sup> For example, incorporation of homocysteine through *N*-homocysteinylation results in protein misfolding<sup>90</sup> and autoimmune response.<sup>91</sup> Thus, further studies of the relationship between homocysteine incorporation and protein function is needed to elucidate the mechanism of hyperhomocysteinemia and related diseases.<sup>89</sup>

Caged Cys and Ser has previously been incorporated as direct nitrobenzylthioethers and nitrobenzylethers, respectively, through tRNA synthetase engineering,<sup>41, 95-98</sup> including the use of the pyrrolysyl-tRNA synthetase/tRNA<sub>CUA</sub> pair and the *Escherichia coli* leucyl-tRNA synthetase

(EcLeuRS)/tRNA<sub>CUA</sub> pair to generate photoactivatable Tobacco Etch Virus (TEV) protease and *Nostoc punctiforme* (*Npu*) DnaE intein using the caged cysteines shown in Figure 5.1.<sup>98-99</sup>



**Figure 5.1:** Chemical structures of the reported photocaged cysteines.

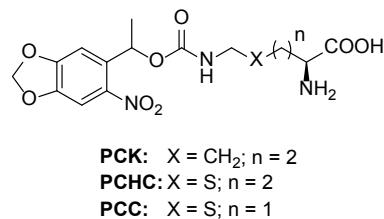
Structures of photocaged cysteines for photochemical control of TEV protease function (A)<sup>99</sup> and intein-mediated protein splicing (B).<sup>98</sup>

Here a different approach was applied to design and engineer the molecular structure of caged cysteines and caged homocysteine for incorporation into proteins in bacterial and mammalian cells by using an existing *M. barkeri* pyrrolysine tRNA synthetase (PylRS) mutant.<sup>100-104, 173-175</sup>

## 5.2 INCORPORATION AND APPLICATION OF NITROBENZYL-CAGED CYSTEINE ANALOGS

Given that important contacts for amino acid recognition in the PylRS active site exist, a “relay” caging/decaging strategy was developed by inserting an additional aminomethylene spacer between the caging group and the thiol group of the amino acid (Figure 5.2). Photolysis of the

*ortho*-nitrobenzyl group yields a carbamic acid. The carbamic acid rapidly decarboxylates, generating an unstable aminomethylene group which decarbonylates releasing the active Cysteine or homocysteine thiol. By applying this strategy, the caged homocysteine **PCHC** displays the exact same structure as the previously genetically encoded caged Lysine **PCK**, with the exception of a subtle  $\text{CH}_2 \rightarrow \text{S}$  modification. The caged Cysteine **PCC** has the same structure, shortened by just one methylene unit. It was most likely that **PCHC** and **PCC** would be substrates for the PylRS enzyme evolved for **PCK**. Therefore, the photocaged homocysteine (**PCHC**) and photocaged cysteine (**PCC**) were synthesized by Dr. Rajendra Uprety in our lab (Figure 5.2).



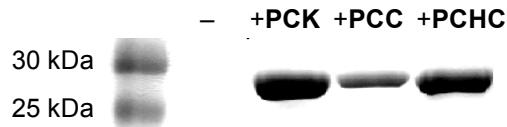
**Figure 5.2:** Structures of genetically encoded photocaged amino acids **PCK**, **PCHC**, and **PCC**.

These amino acids were synthesized by Dr. Rajendra Uprety in the Deiters lab.

A previously evolved pyrrolysyl-tRNA synthetase (CKRS), containing the four mutations M241F, A267S, Y271C, and L274M, was reported to successfully incorporate **PCK** in bacterial, yeast, mammalian cells, and mice.<sup>48</sup> The work presented here will use the evolved *M. barkeri* pyrrolysyl-tRNA<sub>CUA</sub>/aaRS pair for incorporation of the two unnatural amino acids, **PCHC** and **PCC**.

### 5.2.1 Incorporation of nitrobenzyl-caged cysteine analogs

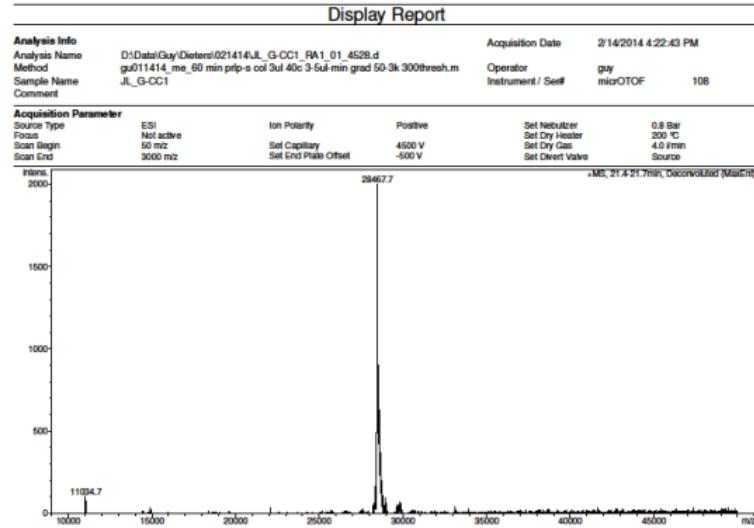
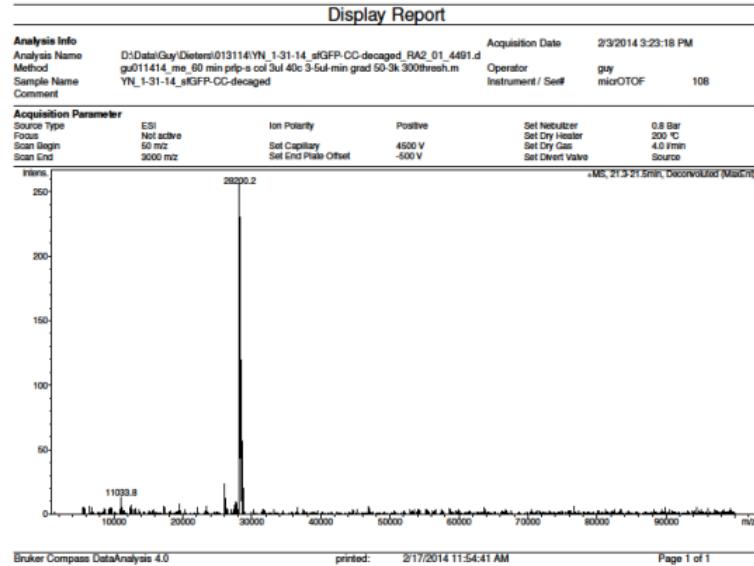
In order to determine the site-specific incorporation and analyze the incorporation efficiency of **PCK**, **PCHC**, and **PCC**, the plasmids pBAD-sfGFPY151TAG-pylT containing a hexahistidine-tagged sfGFP gene with an amber stop codon at the Y151 permissive site, and pBK-CKRS containing the gene encoding CKRS were co-transformed into *E. coli* Top10 cells. Cells were grown in Luria-Bertani (LB) medium in the presence of the designated unnatural amino acids (1 mM) and protein expression was induced with arabinose (0.2%) when the OD<sub>600</sub> reached 0.4. Expression was continued at 37 °C overnight, followed by Ni-NTA purification. Importantly, all caged amino acids could be incorporated into sfGFP at the Y151TAG mutation site (Figure 5.3). The expression yields of purified sfGFP-**PCK**, -**PCHC**, and -**PCC** in place of Y151 were 6 mg/L, 5.6 mg/L and 1.5 mg/L, respectively (Figure 5.3). Based on the protein expression yield, the incorporation efficiency of **PCHC** is similar to that of **PCK** (positive control), while that of **PCC** is lower.



**Figure 5.3:** Genetic incorporation of photocaged lysine **PCK**, photocaged homocysteine **PCHC**, and photocaged cysteine **PCC** in *E. coli*.

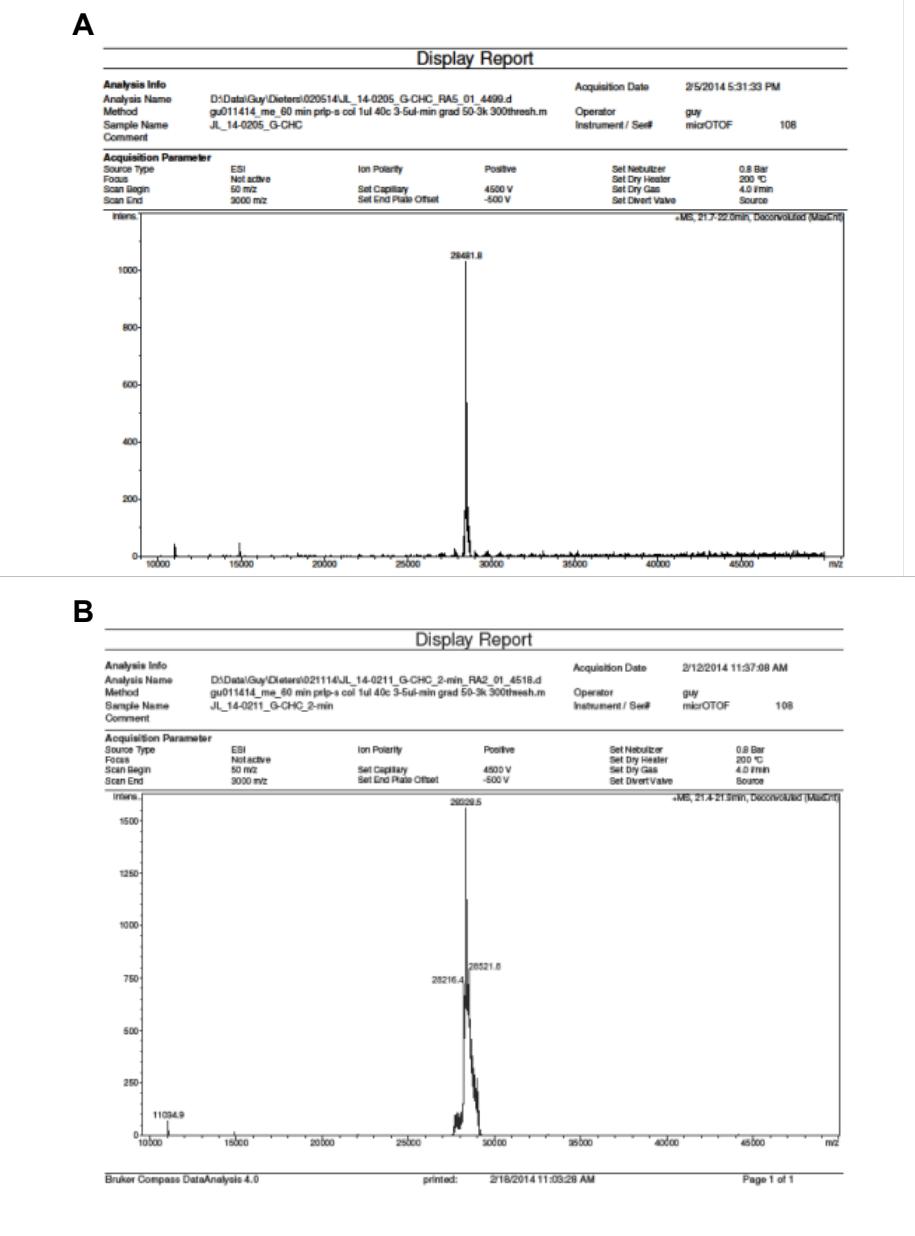
Coomassie stained gel of Ni-NTA purified sfGFP-Y151TAG expression in the absence and presence of **PCK**, **PCHC**, and **PCC**.

Incorporation of **PCHC** and **PCC** was further validated by electrospray ionization mass spectrometry (ESI-MS) of the purified caged-sfGFP (Figure 5.4A and Figure 5.5A), along with ESI-MS data for the decaged protein following light irradiation, confirming the expected fragmentation of the aminomethylene spacer (Figure 5.4B and Figure 5.5B). The ESI-MS of sfGFP-Y151-**PCHC** and -**PCC** showed molecular weights of 28481.8 Da (expected mass 28482.4 Da) and 28467.7 Da (expected mass 28468.2 Da), confirming the incorporation of **PCHC** and **PCC** into sfGFP, respectively (Figure 5.5A). Moreover, ESI-MS analyses were performed on the proteins irradiated with UV light. It showed that the decaged sfGFP-**PCC** had a mass of 28200.2 Da, in agreement with the expected mass of 28201.3 Da (Figure 5.4B), while the observed mass of the decaged sfGFP-**PCHC** was 28216.4 Da, corresponding to the expected mass, 28215.6 Da (Figure 5.5B). The ESI-MS results further demonstrated that **PCC** and **PCHC** could be decaged effectively with UV light.

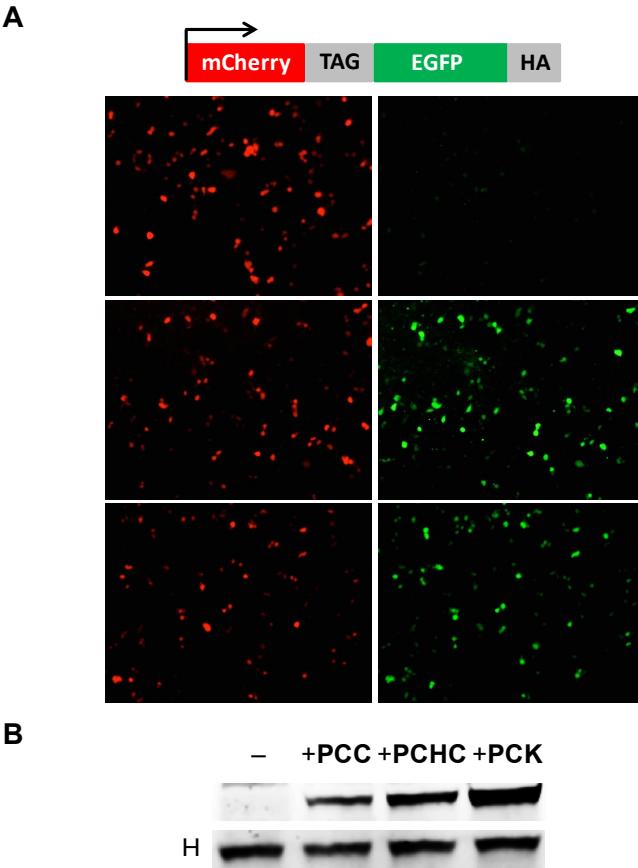
**A****B**

**Figure 5.4:** ESI-MS analysis of sfGFP-PCC and decaged sfGFP-PCC.

A) ESI-MS of sfGFP-PCC. Expected Mass: 28468.2 Da  $\pm$  2.0 Da, Found: 28467.7 Da B) ESI-MS of decaged sfGFP-PCC. Expected Mass: 28201.3 Da  $\pm$  2.0 Da, Found: 28200.2 Da.



co-transfected with pMbCKRS-mCherry-TAG-EGFP-HA and p4CMVE-U6-PylT in the presence or absence of **PCHC** and **PCC** (1 mM).<sup>176</sup> As expected, mCherry fluorescence was observed in the cells in presence or absence of **PCHC** and **PCC**, but EGFP fluorescence was detected only with the addition of **PCHC** and **PCC** (Figure 5.6A) – confirming the specificity of the synthetase. Western blot analysis further validated that the full-length mCherry-**PCK/PCHC/PCC**-EGFP-HA fusion protein was expressed in the presence of **PCK**, **PCHC** and **PCC**. The relative ratio of protein expression incorporating **PCK**, **PCHC**, and **PCC** is 3 : 1.5 : 1 (based on integration of the Western blot bands using GE ImageQuant). These experiments demonstrate the successful incorporation of **PCHC** and **PCC** into protein in response to the amber codon in mammalian cells (Figure 5.6B).

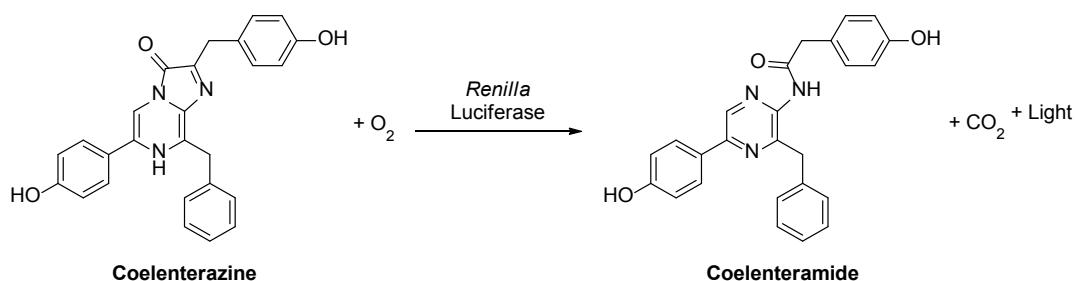


**Figure 5.6:** Genetic incorporation of photocaged lysine **PCK**, photocaged homocysteine **PCHC**, and photocaged cysteine **PCC** in mammalian cells.

A) Fluorescence micrographs of HEK293T cells expressing pMbCKRS-mCherry-TAG-EGFP-HA and PytRNA<sub>CUA</sub> in the absence and presence of **PCHC** and **PCC** (1 mM). B) Western blot of lysate from HEK293T cells expressing mCherry-**PCK/PCHC/PCC**-EGFP-HA with an anti-HA tag antibody and anti-GAPDH loading control confirms that the pCKRS/PytRNA<sub>CUA</sub> pair incorporates **PCK**, **PCHC** and **PCC** in response to the amber codon in mammalian cells.

### 5.2.2 Application of nitrobenzyl-caged cysteine analogs

Despite the lack of a cysteine residue within the active site of firefly *P. pyralis* luciferase, there is a cysteine residue within the active site of *Renilla* luciferase. Due to its ability to emit light, *Renilla* luciferase, a 36 kDa protein from *sea pansy* (*Renilla reniformis*, RLuc), is a versatile research tool for gene expression assays and biosensor applications in cell culture and in model organisms.<sup>105</sup> *Renilla* luciferase catalyzes the oxidation of coelenterazine (RLuc substrate) in the presence of oxygen to generate coelenteramide and CO<sub>2</sub> with the concomitant emission of blue light (480 nm peak, Figure 5.7).



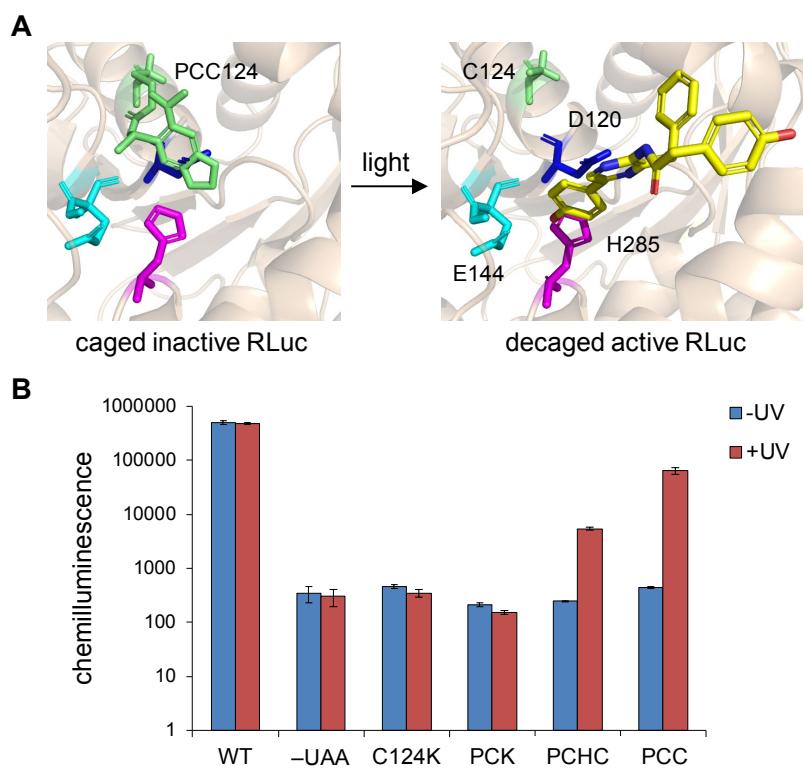
**Figure 5.7:** Bioluminescent reaction catalyzed by *Renilla* luciferase.

Crystal structures of *Renilla* luciferase reveal that the conserved catalytic triad of RLuc is directly involved in the oxidation of coelenterazine.<sup>105-106</sup> The catalytic triad residues are D120, E144, and H285.<sup>106</sup> The residue D120 acts as the “nucleophile elbow” (residues 118–122) in the catalytic triad of the  $\alpha/\beta$  hydrolase fold and plays an important role in catalysis.<sup>106</sup> C124 is located on the  $\beta$  sheet, which is closely associated with formation of the D120 nucleophile loop.<sup>105-106, 177</sup> Based on the structure of RLuc (Figure 5.8A), the caging group attached to C124

imposes steric hindrance on D120, and most likely interferes with the formation of the catalytic triad which is crucial for oxidation of the substrate, resulting in blocking luciferase activity. Therefore, the installation of a caging group at C124 might enable the photoregulation of *Renilla* luciferase activity.

In order to determine the efficiency of photolysis of **PCHC** and **PCC** and apply the caged amino acids to trigger light-induced protein function in live mammalian cells, a TAG mutation was introduced at position C124 in RLuc and the incorporation of **PCHC** and **PCC** was evaluated by co-transfected HEK 293T cells with the mutated *Renilla* luciferase plasmid (pRLuc-C124TAG) and plasmids encoding the *MbCKRS*/Py<sub>t</sub>tRNA<sub>CUA</sub> pair. After a 24-hour incubation in the presence (1 mM) or absence of **PCHC** and **PCC**, the cells were either irradiated for 4 min (365 nm, 25 W) or kept in the dark, followed by a *Renilla* luciferase assay. Incorporation of **PCHC** and **PCC** in RLuc caused complete enzyme inactivity before UV irradiation, compared to a negative control in the absence of any unnatural amino acid. This demonstrates that the caging group was capable of fully inhibiting the function of the catalytic triad. After UV exposure, RLuc-**PCHC** and -**PCC** were decaged, generating homocysteine and Cysteine in the active site and restoring *Renilla* luciferase activity. A profound luminescence increase of 150-fold was observed for **PCC**, as a wild-type active site containing Cys is obtained after photolysis of the caging group (Figure 5.8B). Only a 20-fold activity increase was seen in RLuc-**PCHC** decaging, despite the higher incorporation efficiency of **PCHC** in mammalian cells (see Figure 5.3), because homocysteine still induces a perturbation on the catalytic triad due to its CH<sub>2</sub>-extended side chain compared to Cys. The results showed that the incorporation of **PCC** upon illumination resulted in higher activity than that of **PCHC**, while still using the same synthetase. Accordingly, the incorporation of the caged Lys **PCK** did not allow for light-

activation of luciferase activity, as the resulting C124K mutant is also completely inactive (Figure 5.8B). These results confirm that decaging of **PCHC** and **PCC** is functional *in vivo*, that protein activity can be controlled with light in live cells using caged cysteine, and that the incorporation of homocysteine allows for precise probing (with single atom resolution) of protein active sites. These results further determined the fidelity of incorporation of **PCHC** and **PCC** by the *MbCKRS/PyltRNA<sub>CUA</sub>* pair. Therefore, the activity of *Renilla* luciferase can be tightly regulated in spatial and temporal resolution by incorporation of a photocaged cysteine residue into the associated active site of the constitutively active luciferase protein.



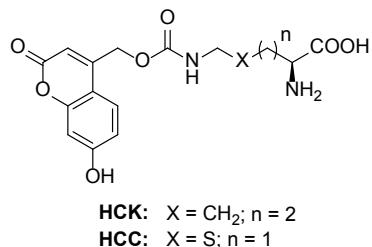
**Figure 5.8:** Optochemical control of *Renilla* luciferase.

A) Structural analysis of optochemical control of *Renilla* luciferase. The photocaged cysteine (**PCC**) at the C124 position (green) and the catalytic triad of D120, E144, and H285 are shown (PDB: 2PSJ). Light-induced decaging allows access of the coelenterazine substrate (yellow) to the active site. B) *In vivo* light-activation of *Renilla*

luciferase containing **PCK**, **PCHC**, and **PCC** at position C124. RLuc-WT, RLuc-C124TAG, and RLuc-C124K were performed as controls. Chemilluminescence is shown in a logarithmic scale, indicating no activity of the caged enzyme and a >150-fold activation after UV exposure of Rluc C124→**PCC**. Error bars represent standard deviations from three independent experiments.

### 5.3 INCORPORATION OF COUMARIN-CAGED CYSTEINE

Following the previously discussed approach, hydroxycoumarin cysteine **HCC** was designed because it is structurally similar to the previously genetically encoded hydroxycoumarin Lys **HCK** (see 2.0),<sup>107</sup> with less one methylene unit and a CH<sub>2</sub> → S modification. It could be possible to incorporate **HCC** by the same synthetase (*MbBHCKRS*) containing two mutations – Y271A and L274M.<sup>107</sup> If this is true, then this synthetase might be compatible to incorporate **HCC** in *E. coli* and in mammalian cells and thereby enable Cys decaging with 405 nm light. Therefore, the hydroxycoumarin cysteine (**HCC**) was synthesized by Dr. Subhas Samanta in the Deiters lab (Figure 5.9).

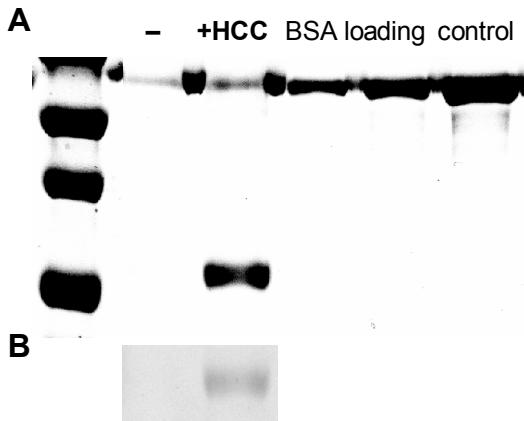


**Figure 5.9:** Structures of the genetically encoded hydroxycoumarin Lys (**HCK**) and Cys (**HCC**).

These amino acids were synthesized by Dr. Subhas Samanta in the Deiters lab.

To demonstrate the incorporation of HCC into proteins in *E. coli*, pBAD-6×HIS-sfGFP-Y151TAG\_PyltRNA and pBK-BHCKRS were co-transformed into Top10 competent cells. The cells were grown in 25 mL of LB media containing tetracycline, kanamycin, and **HCC** (1 mM) for protein expression and cell culture was induced with arabinose (0.2%) until an optical density of OD<sub>600</sub> ~0.6. After Ni-NTA purification, sfGFP-**HCC** was analyzed by 12% SDS-PAGE (Figure 5.10A). Upon visualization, in-gel intrinsic coumarin fluorescence of the expressed proteins containing **HCC** was observed, while no fluorescence was detected in the no UAA control (Figure 5.10B). These results demonstrated that **HCC** was specifically incorporated with the identified evolved BHCKRS in *E. coli*.

In addition, the protein expression yield of sfGFP-**HCC** was around 0.4 mg/L, which was significantly lower than that of sfGFP-**HCK** (8.0 mg/L). The two reasons resulted in the low incorporation yield of **HCC**. One is because of the structure difference – the shorter length of the cysteine side chain, compared with **HCK**. The other one is due to its impurity, including a certain percentage of methyl ester of **HCC**. The impure **HCC** will be purified in the future.



**Figure 5.10:** Genetic incorporation of photocaged cysteine **HCC** in *E. coli*.

A) Coomassie stained gel of Ni-NTA purified sfGFP-Y151TAG expressed in *E. coli* in the absence and presence of **HCC** (1 mM). The concentrations of BSA are 187.5, 375, and 750 ng/ $\mu$ L. B) Coumarin fluorescence gel imaged by ChemiDoc.

## 5.4 CONCLUSIONS

In summary, the genetic code of bacterial and mammalian cells has been expanded through the genetic incorporation of photocaged homocysteine **PCHC**, photocaged cysteine **PCC**, and hydroxycoumarin cysteine **HCC**. The design of the unnatural amino acids allowed for site-specific incorporation using *MbPylCKRS* synthetases that were evolved to incorporate the structurally similar caged lysine **PCK** and hydroxycoumarin lysine **HCK**. Cysteine has both structural and functional implications in many proteins and enzymes and photochemical control over its activity will offer precise spatial and temporal regulation of a wide range of cellular processes. As a proof-of-principle, photochemical control of *Renilla* luciferase was achieved by caging the critical residue C124 through incorporation of **PCHC** and **PCC** in live mammalian cells. Optochemical activation led to a 150-fold increase in enzymatic activity after a brief UV

illumination, demonstrating excellent OFF → ON switching behavior. Pyrrolysine-based genetic code expansion has also been successfully employed not only in bacterial and mammalian cells, but also in multicellular organisms.<sup>178-179</sup> Therefore, the genetically encoded caged cysteine analogs will have broad and diverse optogenetic applications in various biological systems. Beyond optochemical control, the site-specific incorporation of homocysteine into proteins will allow for a precise investigation of the geometric requirements for transition states of cysteine-containing active sites.<sup>180</sup>

## 5.5 EXPERIMENTAL

### Protein expression in *E. coli*.

The pBAD-sfGFP-Y151TAG-pyLT and pBK-CKRS<sup>176</sup> plasmids were co-transformed into *E. coli* Top10 competent cells. A single colony was grown in LB (supplemented with 25 µg/mL Tet and 50 µg/mL Kan) overnight and 250 µL of the overnight culture was added to 25 mL LB supplemented with 1 mM of the designated unnatural amino acid and 25 µg/mL of Tet and 50 µg/mL of Kan. Cells were grown at 37 °C, 250 rpm, and the protein expression was induced with 0.1% arabinose when OD<sub>600</sub> reached 0.4. After overnight expression at 37 °C, cells were harvested and resuspended in 6 mL phosphate lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 8.0), lysozyme (60 µL, 10%), and protease inhibitor cocktail for purification of His-tagged proteins (6 µL, Sigma). Triton X-100 (60 µL of a 10% solution in water) was added to the mixture. The lysate was incubated on ice for 1 hour, sonicated with six short bursts of 30 s on ice followed by 30 s intervals for cooling (Fisher Scientific 550 Sonic Dismembrator), and then centrifuged at 4 °C, 13,000 g, for 10 minutes. The supernatant was transferred to a 15 mL conical tube and 100 µL

Ni-NTA resin (Qiagen) was added. The mixture was incubated at 4 °C for 2 hours under mild shaking. The resin was then collected by centrifugation (1,000 g, 10 min), washed twice with 300 µL of lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, pH 8.0), followed by two washes with 300 µL of wash buffer (lysis buffer containing 20 mM imidazole). The protein was eluted with 300 µL of elution buffer (lysis buffer containing 250 mM imidazole). The purified proteins were analyzed by 10% SDS-PAGE, and stained with Coomassie Blue. The protein mass was obtained by electrospray ionization mass spectrometry (for caged homocysteine **PCHC**, found: 28481.8 Da, expected: 28482.4 Da; for caged cysteine **PCC**, found: 28467.7 Da, expected: 28468.2 Da).

### **Cell culture, plasmid transfection, and western blot.**

Human embryonic kidney (HEK) 293T cells were grown in DMEM (Dulbecco's Modified Eagle Medium, Gibco) supplemented with 10% FBS (Gibco), 1% Pen-Strep (Corning Cellgro) and 2 mM L-glutamin (Alfa Aesar) in 96-well plates (Costar) in a humidified atmosphere with 5% CO<sub>2</sub> at 37 °C. HEK 293T cells were transiently transfected with pMbCKRS-mCherry-TAG-EGFP-HA and p4CMVE-U6-PylT<sup>176</sup> (gifts from the Chin lab) at ~75% confluence in the presence or absence of **PCHC** and **PCC** (1 mM) in 96-well plates. Double transfections were performed with equal amounts of the two plasmids (100 ng of each plasmid) using linear polyethylenimine (LPEI; 1.5 µL, 0.323 mg/mL). After 37 °C incubation overnight, the cells were washed with PBS and imaged with a Zeiss Axio Observer.Z1 Microscope (10× objective). To confirm the expression and levels of the fusion protein, a Western blot was performed. HEK 293T cells were seeded at ~300,000 cells per well and co-transfected with pMbCKRS-mCherry-TAG-EGFP-HA and p4CMVE-U6-PylT (1.5 µg of each plasmid) using LPEI (15 µL, 0.323 mg/mL) at ~80% cell

confluence in the presence or absence of **PCHC** and **PCC** (1 mM) in 6-well plates. After 24 h incubation, the cells were washed once by chilled PBS (1 mL), lysed in mammalian protein extraction buffer (200 µL, GE Healthcare) with mammalian protease inhibitor cocktail (2 µL, Sigma, P8340) on the ice shaker, and the cell lysates were then quantified, isolated, and extracted at 13,200 rpm centrifugation (4 °C, 20 min). The protein lysate was boiled with SDS and then analyzed by 10% SDS-PAGE. After gel electrophoresis and transferring to nitrocellulose membrane (GE Healthcare), the membrane was blocked in TBS with 0.1% Tween 20 (Fisher Scientific) and 5% milk powder for 1 h. The blots were probed and incubated with the primary antibody – α-HA-probe (Y-11) rabbit polyclonal IgG (sc-805, Santa Cruz Biotech) overnight at 4 °C, and followed with the fluorescent secondary antibody – goat-α-rabbit IgG Cy3 (GE Healthcare) for 1 hr at room temperature. The membranes were washed with TBS with Tween 20 (0.1 %) three times. The Western blots were imaged using a GE Typhoon scanner.

#### **Photocaged *Renilla* luciferase assay.**

HEK 293T cells were cultured in DMEM (Dulbecco's Modified Eagle Medium, Gibco) supplemented with 10% FBS (Gibco), 1% Pen-Strep (Gibco) and 2 mM L-glutamin (Alfa Aesar) in 96-well plates (BD Falcon) in a humidified atmosphere with 5% CO<sub>2</sub> at 37 °C. At 80%-90% confluence, cells seeded on plates were transfected and medium was changed to fresh DMEM supplemented without or with **PCK**, **PCHC**, or **PCC** (1 mM). The plasmid pAG31-38 was constructed containing both pCMV-*MbCKRS* and p4CMVE-U6-PylT. A TAG amber stop codon was introduced into the C124 site using the QuikChange mutagenesis kit (Agilent Technologies) and primers For (gccacgactggggggtagctggcttcactactc) and Rev (gagtagtgaaaggccagctaagccccccagtcgtggc). A pRL-TK plasmid containing the gene encoding

*Renilla* luciferase with a TAG amber mutation at residue C124 (pRL-C124TAG) was co-transfected into cells with the plasmid pMbCKRS-4CMVE-U6-PylT using LPEI according to the manufacturer's protocols (Sigma). After double transfection and 24 h incubation, the medium was changed to DMEM without phenol red, and then cells were irradiated for 4 min using a 365-nm UV lamp (High performance UV transilluminator, UVP, 25 W). HEK293T cells expressing RLuc-C124TAG or RLuc-**PCK/PCHC/PCC** were lysed with 20 µL of lysis buffer (Promega) with and without UV irradiation on the shaker. The cell lysates (20 µL) were mixed with coelenterazine substrate solution (100 µL, Promega) in a 96-well plate (BD Falcon). The luciferase activity was performed using the *Renilla* luciferase assay system (Promega) and a Synergy 4 Multi-mode Microplate Reader with an integration time 2 s and sensitivity at 150.

### **Incorporation of HCC in bacteria.**

The pBAD-sfGFP-Y151TAG-pylT and pBK-BHCKRS<sup>107</sup> plasmids were co-transformed with into *E. coli* Top10 cells. A single colony was grown in LB (supplemented with 25 µg/mL Tet and 50 µg/mL Kan) overnight and 400 µL of the overnight culture was added to 25 mL LB supplemented with 1 mM **HCC** and 25 µg/mL Tet and 50 µg/mL Kan. Cells were grown at 37 °C, 250 rpm, and the protein expression was induced with 0.1% arabinose when OD<sub>600</sub> reached ~0.6. After overnight expression at 37 °C, cells were harvested and resuspended in 6 mL phosphate lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 8.0), lysozyme (60 µL, 10%), and protease inhibitor cocktail (6 µL, Sigma). Triton X-100 (60 µL of a 10% solution in water) was added to the mixture. The lysate was incubated on ice for 1 hour, sonicated with six short bursts of 30 s on ice followed by 30 s intervals for cooling (Fisher Scientific 550 Sonic Dismembrator), and then centrifuged at 4 °C, 13000 g, for 10 minutes. The supernatant was transferred to a 15 mL conical

tube and 100  $\mu$ L Ni-NTA resin (Qiagen) was added. The mixture was incubated at 4 °C for 2 hrs under mild shaking. The resin was then collected by centrifugation (1000 g, 10 min), washed twice with 300  $\mu$ L of lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, pH 8.0), followed by two washes with 300  $\mu$ L of wash buffer (lysis buffer containing 20 mM imidazole). The protein was eluted with 200  $\mu$ L of elution buffer (lysis buffer containing 250 mM imidazole). The purified proteins were analyzed by 10% SDS-PAGE, and stained with Coomassie Blue.

## 6.0 GENETIC ENCODING OF CAGED TYROSINES

This chapter will describe the genetic incorporation of three new caged tyrosine analogs with improved photochemical properties in bacterial and mammalian cells. The new tyrosine analogs were applied to the photoregulation of firefly luciferase and excellent OFF to ON light-switching was observed. This reporter was then used to evaluate the activation rates of the different light-removable protecting groups in live cells. The nitropiperonyl caging group was identified as an excellent compromise between incorporation efficiency and photoactivation properties. An important proteolytic enzyme, TEV protease, was engineered for optical control to demonstrate applicability of the new caged tyrosines. The ability to incorporate differently caged tyrosine analogs into proteins in live cells further expands the unnatural amino acid and optogenetic toolbox for synthetic biology.

### 6.1 INTRODUCTION TO CAGED TYROSINE ANALOGS

This material was reprinted, in part, with permission from [Luo, J.; Torres-Kolbus, J.; Liu, J.; Deiters, A. \*Chembiochem\* 2017, 18\(14\), 1442-1447.](#)

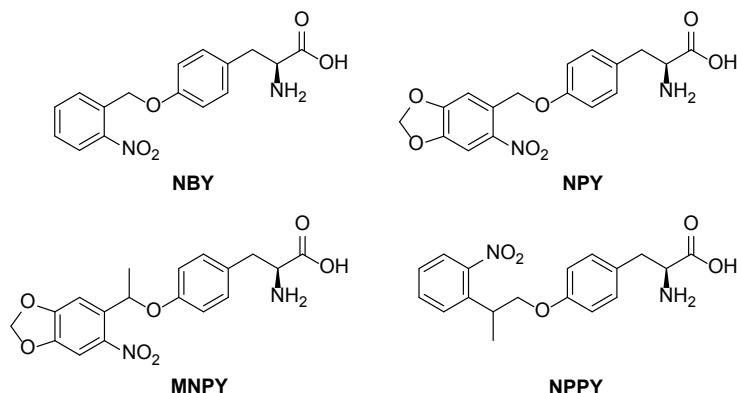
Previously, the *Methanococcus jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS)/tRNA<sub>CUA</sub> pair was developed,<sup>6</sup> and a wide range of tyrosine derivatives have been site-specifically incorporated into proteins in *E. coli* using evolved *Mj*Tyr synthetases.<sup>7, 181</sup> However,

*Mj*TyrRS/tRNA<sub>CUA</sub> could not be used in eukaryotic cells, because its tRNA<sub>CUA</sub> is recognized by endogenous eukaryotic aminoacyl-tRNA synthetases (aaRSs), while the PylRS/tRNA<sub>CUA</sub> pair has been successfully used in yeasts, mammalian cells, and different levels of cellular organisms, as previously described.<sup>40, 182-184</sup>

In order to resolve this problem and incorporate the tyrosine analogues in eukaryotes, the Chin lab previously developed a bio-orthogonal *Mb*PylRS/tRNA<sub>CUA</sub> pair from a library of *M. barkeri* PylRS mutants based on the pyrrolysine system.<sup>131</sup> The evolution of the *Mb*PylRS was done by creating  $\sim 10^8$  mutants with 5 randomized positions (L270, Y271, L274, N311, and C313) within the active binding site.<sup>131</sup> Based on this library, three rounds of alternating positive and negative selections were performed in *E. coli* in the presence and absence of photocaged tyrosine (*o*-nitrobenzyl-*O*-tyrosine, **NBY**) (Figure 6.1).<sup>131</sup> From the selection, the evolved pyrrolysyl-tRNA synthetase (NBYRS) was isolated, which has four mutations of L270F, L274M, N311G, and C313G. The NBYRS/tRNA<sub>CUA</sub> pair has successfully been used in bacterial and mammalian cells for the incorporation of **NBY**.

In the Liu lab (TAMU), the PylRS mutant N311A/C313A was reported to be successfully used for genetically incorporating a series of phenylalanine derivatives in *E. coli* through a rational design strategy.<sup>185</sup> Two mutations at N311 and C313 reported by the Liu lab<sup>185</sup> are consistent with those in the library constructed in the Chin lab. In the collaboration with the Cropp lab (VCU), directed evolution was performed based on a mutant library containing five positions (Y271, L274, N311, C313, and Y349), and two PylRS variants, Y271M\_L274A\_Y349F and Y271M\_L274A\_N311A\_C313A\_Y349F, were found to be good hits for the incorporation of **ONBY** and **ONBTY** in *E. coli*.

In the Deiters lab, to further investigate the incorporation of the tyrosine analogues using evolved *Mb*PylRS/tRNA<sub>CUA</sub> pairs in eukaryotes, a small synthetase mutant panel was rationally designed and created with specific mutations at six positions (L270, Y271, L274, N311, C313, and Y349) for the screening of synthetases for incorporating the individual tyrosine analogue in mammalian cells. A series of caged tyrosine analogues were designed and synthesized in the Deiters lab. Besides **NBY**, photocaged *o*-nitropiperonylmethyl tyrosine (**MNPY**), *o*-nitropiperonyl tyrosine (**NPY**), and *o*-nitrophenylpropyl tyrosine (**NPPY**) were synthesized by Dr. Jessica Torres in the Deiters Lab (Figure 6.1). The work presented here used the evolved *M. barkeri* pyrrolysyl-tRNA<sub>CUA</sub>/PylRS pair for the incorporation of these photocaged tyrosine analogues.

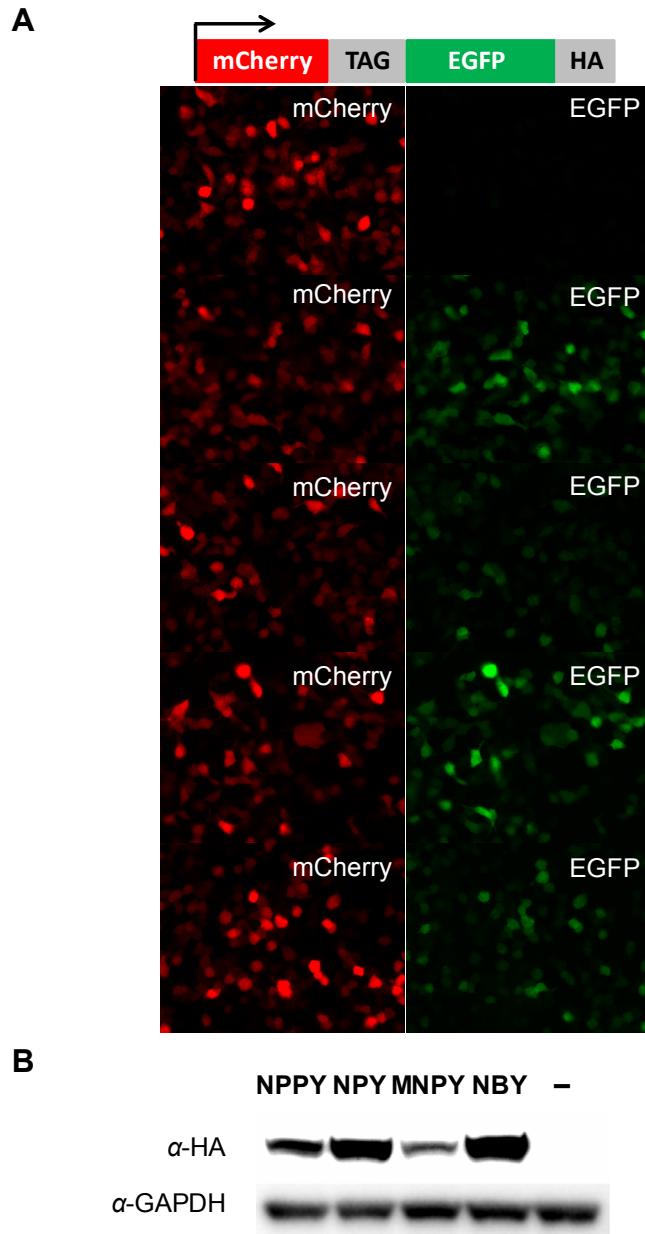


**Figure 6.1:** Structures of the caged tyrosine analogs **NBY**, **NPY**, **MNPY**, and **NPPY**.

## 6.2 INCORPORATION OF CAGED TYROSINE ANALOGS – **NBY**, **MNPY**, **NPY**, AND **NPPY**

A previously evolved pyrrolysyl-tRNA synthetase (NBYRS), containing the four mutations L270F, L274M, N311G, and C313G, accepts **NBY** as a substrate.<sup>186</sup> We found that **NPY** can be

incorporated into proteins by the same synthetase, while the incorporation efficiency of **MNPY** is very low, and **NPPY** could not be incorporated. Furthermore, screening a mutant PylRS panel did not lead to the identification of better synthetases. In order to improve the efficiency of the synthetase, an additional Y349F mutation was introduced, which has been shown to enhance enzymatic activity.<sup>187</sup> Thus, HEK 293T cells were co-transfected with pMbPylRS-mCherry-TAG-EGFP-HA, expressing an incorporation reporter and the engineered synthetase, and p4CMVE-U6-PylT, expressing multiple copies of the cognate tRNA, in the presence or absence of **NBY** (0.25 mM), **MNPY** (0.25 mM), **NPY** (0.25 mM), and **NPPY** (0.5 mM; due to enhanced solubility). As expected, mCherry fluorescence was observed in all cells, but EGFP fluorescence, as the result of unnatural amino acid incorporation, was only detected in cells treated with **NBY**, **MNPY**, **NPY**, and **NPPY** (Figure 6.2B). Gratifyingly, all the caged tyrosines can be successfully incorporated by this new synthetase (L270F, L274M, N311G, C313G, Y349F), termed NPYRS. Western blots (Figure 6.2C) further confirmed the following order of incorporation efficiency: **NPY** ≥ **NBY** > **NPPY** > **MNPY**. Importantly, none of the endogenous common 20 amino acids are being incorporated in the absence of the caged tyrosines. Additionally, the incorporation of **MNPY**, **NPY**, and **NPPY** was confirmed by electrospray ionization mass spectrometry.



**Figure 6.2:** Incorporation of new photocaged tyrosines into protein in mammalian cells.

A) Fluorescence micrographs of HEK 293T cells expressing mCherry-TAG-EGFP-HA, NPYRS, and PytRNA<sub>CUA</sub> in the absence and presence of **NBY**, **MNPY**, **NPY**, and **NPPY**. B) Western blot of mCherry-TAG-EGFP-HA confirms that the NPYRS/PytRNA<sub>CUA</sub> pair efficiently incorporates **NBY**, **MNPY**, **NPY**, and **NPPY** in response to an amber codon in mammalian cells. No protein expression was observed in the absence of a caged tyrosine. The anti-GAPDH antibody detection was included as a loading control.

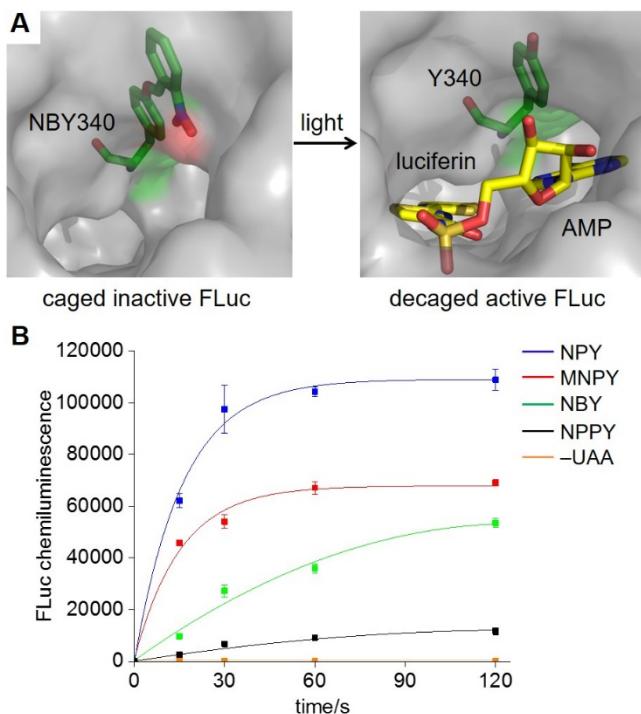
## 6.3 OPTICAL CONTROL OF PROTEIN FUNCTION

### 6.3.1 Optical control of firefly luciferase

The four caged tyrosine analogs were applied to a light-activated firefly luciferase enzyme as an initial target in order to realize the optical control of protein function in live cells. Chemiluminescence measurements have low background, high sensitivity, and allow for easy quantification, making them an attractive readout for protein function. On the basis of the *Photinus pyralis* firefly luciferase (FLuc) crystal structure, Y340 was identified as a key catalytic residue for the effective orientation of ATP in the binding pocket (Figure 6.3A).<sup>68</sup> If this critical tyrosine residue is replaced with a genetically encoded photocaged tyrosine analog, it may block ATP binding and thus interfere with the adenylation of luciferin catalyzed by FLuc (Figure 6.3A). Photolysis of the caged tyrosine through irradiation removes the caging group, produces a free tyrosine, and restores the catalytic activity of the FLuc enzyme.

Site-directed mutagenesis of the corresponding Y340 residue to the amber codon (TAG) enabled incorporation of **NBY**, **MNPY**, **NPY**, and **NPPY** into firefly luciferase in mammalian cells. HEK 293T cells were co-transfected with the mutated firefly luciferase plasmid and the *MbNPYRS/PyltRNA<sub>CUA</sub>* pair (p*MbNPYRS-4PylT*) in the presence of **NBY**, **MNPY**, **NPY**, and **NPPY** (0.25 mM). In order to demonstrate photoactivation of protein function in live cells, an irradiation time course with caged FLuc containing the different caged tyrosines was carried out (Figure 6.3B). Upon incorporation, caged FLuc activity was comparable to the –UAA negative control before UV irradiation, but increased dramatically afterwards, demonstrating optical

activation of firefly luciferase enzyme. Enzymatic activity gradually increased during the course of the irradiation, with a plateau starting to appear at 30-60 s and cells expressing FLuc-**NPY** showing the strongest bioluminescence, suggesting that **NPY** generates the most light-activated protein in cells. Both FLuc-**MNPY** and FLuc-**NPY** displayed rapid decaging. By contrast, the activities of Fluc-**NBY** and Fluc-**NPPY** did not plateau after even 120 s irradiation, thus, amino acid-dependent generation of light-activated Fluc follows: **NPY** > **MNPY** > **NBY** > **NPPY**. Given that the incorporation efficiency differs for the four caged tyrosine analogs (Figure 6.3C), the FLuc activity was correlated to protein expression, to more accurately compare the light-activation potential of the four caging groups.



**Figure 6.3:** Optical control of firefly luciferase activity.

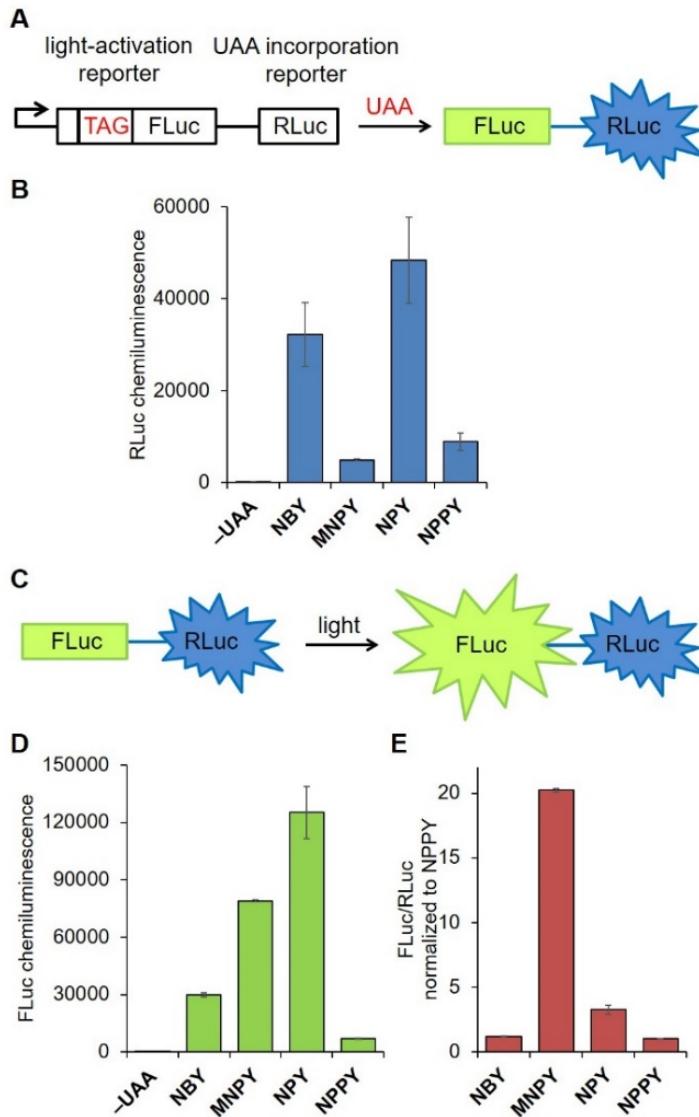
A) Light-activation of firefly luciferase activity through the Y340→**NBY** mutation based on PDB 2D1S. Before irradiation, the caging group blocks access of ATP to the active site until decaging through 365 nm exposure. B) Irradiation time-course luciferase assays for decaging of **NBY**, **MNPY**, **NPY**, and **NPPY** in live cells, showing

rapid decaging for **NPY** and **MNPY**, with virtually complete luciferase activation after 60 s. Error bars represent standard deviations from three independent experiments.

In order to determine protein activation relative to its expression, an incorporation reporter for caged tyrosines is needed in the caged FLuc system. Thus, *Renilla* luciferase (RLuc) was fused downstream to the caged firefly luciferase system as a transfection and expression control (pFLuc\_Y340TAG-RLuc plasmid; Figure 6.4A and Figure 6.4B). The caged Tyr analogs are incorporated into Y340TAG in firefly luciferase and formation of the FLuc-RLuc fusion protein is therefore dependent on the incorporation of the UAA into FLuc, but the enzymatic activity of RLuc is not optically regulated.

The four caged Tyr analogs have different decaging responses to a pulse of UV light due to the different properties of the chromophores. Although all four Tyr analogs were deprotected successfully, as shown by FLuc activation, the kinetics differed significantly. UV irradiation for 30 s produced near maximum activation of FLuc in the case of **NPY** and **MNPY**. Since the chemiluminescence signal plateaued starting from 30 s, but was still increasing at 15 s. Thus, the 15 s time-point was chosen to further showcase caging group differences after accounting for differences in caged tyrosine incorporation efficiency. HEK 293T cells were co-transfected with the pMbNPYRS/PyltRNA<sub>CUA</sub> expression plasmid and pFLuc\_Y340TAG-RLuc in the absence or presence of **NBY**, **MNPY**, **NPY**, and **NPPY**. After 24 h incubation, the cells were either irradiated for 15 s (365 nm) or kept in the dark, followed by a dual luciferase assay. As expected, without UAA, no RLuc activity was detected, while the presence of any of the caged tyrosines resulted in an increased RLuc signal (Figure 6.4B). The presence of **MNPY**, **NPPY**, **NBY**, and **NPY** resulted in a relative RLuc signal intensity of 1:2:7:10. Different incorporation efficiencies among the four caged Tyr analogs account for the variability in RLuc signal, in agreement with

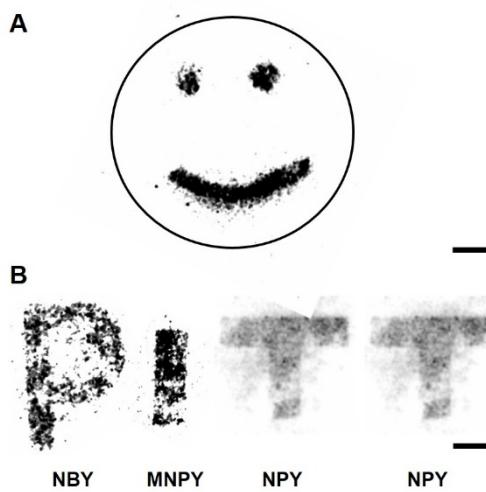
the Western blot results shown in Figure 6.2. As before, the incorporation of **MNPY**, **NPPY**, **NBY**, or **NPY** into the FLuc-Rluc fusion at position Y340 caused complete inhibition of firefly luciferase activity before UV irradiation, comparable to the negative controls (-UAA, Figure 6.3C). UV irradiation of FLuc-**NBY** led to a dramatic 371-fold increase in activity, while 982-fold, 1556-fold, and 88-fold increases were observed with FLuc-**MNPY**, FLuc-**NPY**, and FLuc-**NPPY**, respectively (Figure 6.4C). To compensate for differential incorporation efficiencies of the caged tyrosines, FLuc activity was normalized to the corresponding RLuc activity, as an expression reporter. The ratio of Fluc/Rluc activity in case of **NBY**, **NPPY**, **NPY**, and **MNPY** is 1:1:3:20, revealing that FLuc-**MNPY**, despite its modest expression level, provides the overall most efficient decaging and enzyme activation (Figure 6.4B). These results demonstrate that **NBY**, **MNPY**, **NPY**, and **NPPY** can be efficiently decaged *in vivo*, and that FLuc activity can be controlled with light. Therefore, this developed dual luciferase reporter allowed not only temporal control over the activation of FLuc, but also assessment of the photolysis efficiency among different caged amino acids *in vivo*. Overall, while **MNPY** leads to most efficient protein activation at short light exposure times, the also newly encoded **NPY** provides the best compromise between high incorporation efficiency and efficient decaging.



**Figure 6.4:** Dual luciferase assay for incorporation and decaging of **NBY**, **MNPY**, **NPY**, and **NPPY** *in vivo*.

A) Design of the FLuc-RLuc fusion reporter. After UAA incorporation, RLuc (the UAA incorporation reporter) gene expression is turned on, while FLuc is inactivated. B) In dual luciferase assays, RLuc serves as an incorporation control, showing incorporation of FLuc-**NBY**, -**MNPY**, -**NPY**, and **NPPY**. C) In the designed FLuc-RLuc fusion reporter, FLuc (the light-activation reporter) is activated upon UV irradiation. D) Firefly luciferase assay demonstrating decaging of FLuc-**NBY**, -**MNPY**, -**NPY**, and -**NPPY**. E) FLuc activity is normalized to RLuc activity (protein expression) and expressed as fold change from the FLuc/RLuc ratio for **NPPY** in order to determine the decaging efficiency of **NBY**, **MNPY**, **NPY**, and **NPPY** in cells. Error bars represent standard deviations from three independent experiments.

Furthermore, spatial control of light-activated FLuc was demonstrated via patterned irradiation of a layer of mammalian cells. HEK 293T cells expressing caged FLuc were exposed to 365 nm light ( $20 \text{ mW/cm}^2$ ) through a mask, followed by imaging of chemiluminescence (ChemiDoc MP, Bio-Rad). Bright luminescence was only observed in cells exposed to light and expressing any of the four caged FLuc enzymes (Figure 6.5). As above, **NBY**, **MNPY**, and **NPY** incorporation yielded the highest levels of luminescence after photoactivation, while **NPPY** incorporation again suffered from low expression levels. Taken together, these results demonstrate that the caged FLuc can be spatially and temporally activated in live mammalian cells.



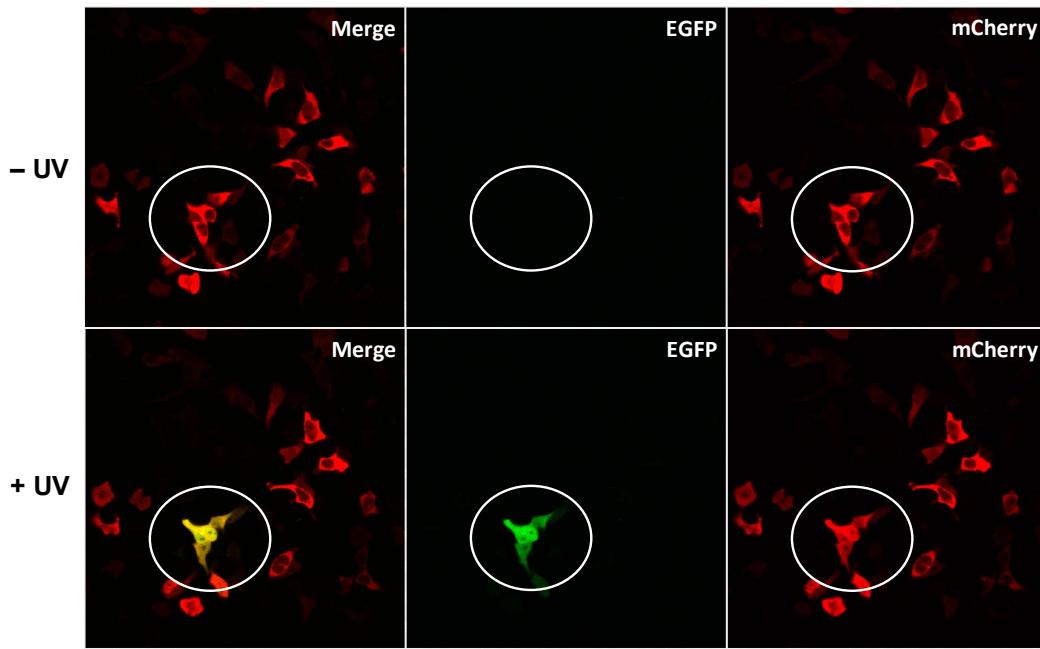
**Figure 6.5:** Spatial control of light-activated FLuc function.

A) Spatial activation of caged FLuc with **NPY** through patterned UV irradiation. HEK 293T cells expressing caged FLuc were treated with and without UV irradiation (365 nm, 5 min,  $20 \text{ mW/cm}^2$ ) followed by bioluminescence imaging of the live cells (ChemiDoc). B) Spatial activation of caged FLuc through patterned UV irradiation using masks and separate incorporation of **NBY**, **MNPY**, and **NPY**. HEK 293T cells expressing FLuc were irradiated for 5 min before imaging. Scale bar indicates 5 mm.

### 6.3.2 Optical control of EGFP maturation

The EGFP chromophore is formed by three amino acid residues, Thr65-Tyr66-Gly67, and it forms spontaneously through a self-catalyzed protein folding mechanism and intramolecular rearrangement. The chemical mechanism of the EGFP maturation has been determined, but the whole maturation process and dynamics *in vivo* (e.g., how long individual maturation steps take) is still unknown. We hypothesize that a photocaged amino acid would be a useful tool to investigate the maturation process with tight spatial and temporal control and that caged tyrosine analogues created in our lab can be successfully incorporated into EGFP and decaged *in vivo*.

To site-specifically incorporate the caged tyrosine analogues into EGFP in mammalian cells, HEK 293T cells were co-transfected with pEGFP\_Y66TAG-mCherry and the pMbNPYRS/PyltRNA<sub>CUA</sub> expression plasmid in the absence or presence of **NPY** using established transfection conditions. After a 24 hr incubation, the cells were either kept in the dark or irradiated for 15 seconds (DAPI filter, 358–365 nm). The cells were then imaged every 5 min for a total of 90 min (Figure 6.6). In the absence of UV irradiation, no EGFP fluorescence was detected, indicating that the caged EGFP was inactivated and the chromophore was not formed. However, upon decaging, the EGFP fluorescence gradually increased over time (Figure 6.6). This indicates that the photocaged tyrosine (**NPY**) at residue Y66 can block the formation of the EGFP chromophore in order to photo-regulate EGFP maturation, which is in agreement with the previous research.<sup>42</sup>



**Figure 6.6:** Optical control of EGFP maturation in live cells.

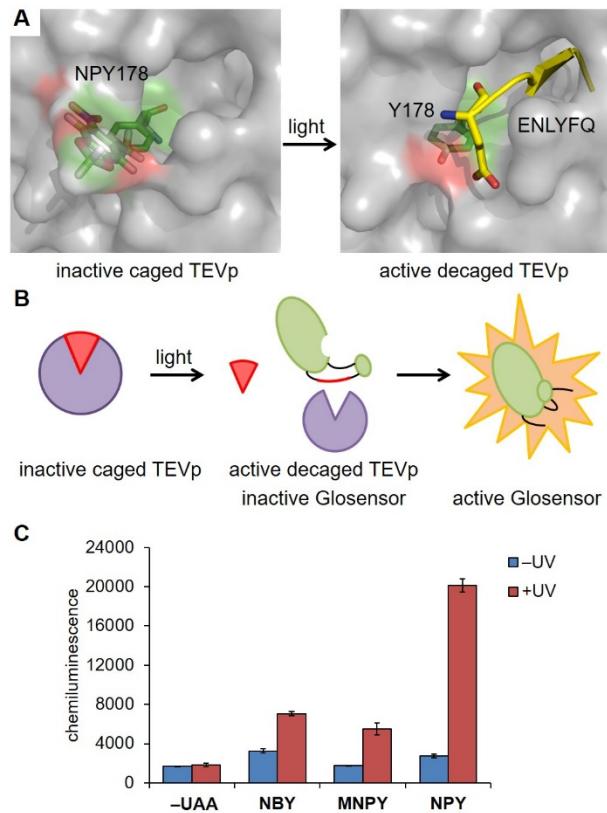
Fluorescence imaging of HEK 293T cells expressing EGFP-Y66TAG-mCherry, 90 min after irradiation by a 405 nm DMD system (30 s, in the red square) in the presence of **NPY** (0.25 mM). The fluorescent cells were imaged using Nikon A1R confocal microscope with 20 $\times$  objective (2-fold zoom, fluorescein (Ex = 488 nm) and Texas Red (Ex = 560 nm) channels).

### 6.3.3 Optical control of TEV protease

In order to further develop the utility of the new caged tyrosine **NPY** for the optical control of protein function in live cells, tobacco etch virus protease (TEVp) was selected as a target due to its broad utility in molecular and synthetic biology.<sup>188-189</sup> On the basis of the TEVp crystal structure, a critical tyrosine residue, Y178, was found to engage in a hydrophobic interaction with the side chain of P4-Leu in the peptide substrate.<sup>190</sup> Furthermore, P6-Glu, another substrate amino acid, is part of a cooperative network of three hydrogen bonds, one of which between the

Glu O-1 and the hydroxyl group of Y178.<sup>190</sup> Y178 is essential for catalytic activity as shown by a dramatic reduction in enzyme catalytic activity displayed in a Y178V mutant.<sup>191</sup> Therefore, we hypothesized that a caging group placed on Y178 would interrupt the favorable interaction with the peptide substrate and block the catalytic efficiency of the protease (Figure 6.7A, Figure 6.7B).

To demonstrate optical control of Y178-caged TEVp in live cells, HEK 293T cells were co-transfected with the mutated protease plasmid (pTEVp-Y178TAG), the *MbNPYRS/PyltRNA<sub>CUA</sub>* pair (pNPYRS-4PylT), and a proteolysis reporter plasmid (pGloSensor, Promega) in the presence of **NBY**, **MNPY**, and **NPY** (0.25 mM). After a 24 h incubation, the cells were either irradiated for 2 min (365 nm, 25 W) or kept in the dark. The expression of caged TEVp in the absence of UV irradiation showed very low reporter activity (similar to the – UAA negative control), indicating that the caging of Y178 effectively inhibited protease function. Importantly, upon UV irradiation, protease activity was restored, with the highest activity being observed with **NPY** due to its enhanced incorporation efficiency compared to **NBY** and **MNPY** and its high decaging efficiency in live cells (Figure 6.7C). Therefore, the catalytic activity of TEVp can be photochemically controlled, with excellent OFF to ON switching, by incorporation of a caged tyrosine residue at a critical site.



**Figure 6.7:** Optical control of TEV protease activity in live cells.

A) Structure of tobacco etch virus protease (TEVp) containing a peptide substrate in its active site (PDB: 1LVM). B) Optical activation of TEV protease in the GloSensor system. C) TEVp reporter cleavage assays demonstrated highly efficient optical OFF to ON switching for the Y178→**NPY** mutant.

## 6.4 CONCLUSIONS

In summary, we have demonstrated the application of an engineered pyrrolysyl-tRNA synthetase/tRNA<sub>CUA</sub> pair in the genetic encoding of three new photocaged tyrosine analogs, **MNPY**, **NPY**, and **NPPY**. These new analogs show improved photochemical properties (including red-shifted absorption spectra and enhanced light-activation rates) compared to the previously encoded **NBY**. Using firefly luciferase as a proof-of-principle target that enables

precise measurement of enzymatic activity, we found that the decaging of **MNPY** was the fastest and the most efficient in live cells, followed by **NPY** and **NBY**. However, expression levels of protein containing **MNPY** were modest, while **NPY** incorporation was significantly higher, thus currently representing an optimal compromise between efficient decaging at 365 nm and protein incorporation yields in mammalian cells. Also, we established a new method to reveal the activation kinetics of different light-removable protecting groups installed in a protein in live cells – an environment that is markedly different than that of test tube studies. We demonstrated that all four caged tyrosines can be used in the precise spatial and temporal control of protein function through patterned irradiation of cellular monolayers, and we utilized the genetically encoded **NPY** for efficient photoactivation of an important proteolytic enzyme, TEV protease. The ability to incorporate all four caged tyrosine analogs with the same PylRS/tRNA<sub>CUA</sub> pair further extends their applications in biological systems, and we expect the improved **NPY** to find broad applicability in light-triggered synthetic biology approaches.

## 6.5 EXPERIMENTAL

### **Plasmid cloning.**

(1) Construction of *MbNPYRS*. The NPYRS plasmid was obtained by making the following mutations to the wild-type PylRS: L270F, L274M, N311G\_C313G, and Y349F, using the primer sets P1/P2, P3/P4, P5/P6, and P7/P8, respectively, with the QuikChange site-directed mutagenesis method (Agilent).

(2) Construction of pNPYRS-4PyIT: The plasmid was obtained by ligating the p4CMVE-U6-PyIT fragment from pMbPyIT between the restriction sites *Nhe*I and *Mfe*I sites of pMbNPYRS.

**Table 6.1:** Primer list.

Primer No.	Sequence
P1	5'-CCACCCCTGTACAACACTACGCCCGAAACTGGACAGAACATC-3'
P2	5'-GTTTCCGCAGGTAGTTGTAGAAGGTGGGGCCAGCATAG-3'
P3	5'-CCACCTCTACAACATACATGCGGAAACTGGACAGAACATC-3'
P4	5'-GATTCTGTCCAGTTCCGCATGTAGTTGTAGAAGGTGG-3'
P5	5'-GAAGAGTTACAATGGTGGCTTGGCCAGATGGCAGCGGCTG-3'
P6	5'-CAGCCGCTGCCCATCTGGCAAAGCCCACCATTGTAAACTCTTC-3'
P7	5'-GCGACAGCTGCATGGTGGCGACACCCCTGGACATC-3'
P8	5'-GATGTCCAGGGTGTGCCAACACCATGCAGCTGCGC-3'
P9	5'-GGACCAATGTAAATATTGTATGAACTAGATCC-3'
P10	5'-ATCCAGGTTACGGATCTAGTTCATGACAATATT-3'
P11	5'-AGAGAGGCTAGCTCCAATTACTGACCGTACACCAA-3'
P12	5'-AGAGAGCTCGAGATGCCATCTTCAGCAGGCGCAC-3'
P13	5'-CTTACAAACACCAATAACTAGTTCACCTCAGTGCCAAAG-3'
P14	5'-CTTGGCACTGAGGTGAACTAGTTATTGGTGGTAAAG-3'

### Incorporation of caged tyrosine analogs in human cells.

Human embryonic kidney (HEK) 293T cells were grown in DMEM (Dulbecco's Modified Eagle Medium, Gibco) supplemented with 10% fetal bovine serum (FBS, Sigma), 1% Pen-Strep

(Corning Cellgro) and 2 mM L-glutamine (Alfa Aesar) in 96-well plates (Costar) in a humidified atmosphere with 5% CO<sub>2</sub> at 37 °C. HEK 293T cells were transiently transfected with the p*Mb*NBYRS-mCherry-TAG-EGFP-HA and p4CMVE-U6-PylT plasmids (100 ng of each plasmid) at ~75% confluence using linear polyethylenimine (LPEI; 1.5 µL, 0.323 mg/mL) in the presence or absence of **NBY**, **MNPY**, **NPY** (0.25 mM), or **NPPY** (0.5 mM) in 96-well plates. After overnight incubation at 37 °C, the cells were washed once with PBS (200 µL) and imaged with a Zeiss Axio Observer.Z1 Microscope (10× objective). To confirm the expression of the fusion protein and to differentiate between expression levels, a Western blot was performed. HEK 293T cells were co-transfected with p*Mb*NBYRS-mCherry-TAG-EGFP-HA and p4CMVE-U6-PylT (1.5 µg of each plasmid) using LPEI (15 µL, 0.323 mg/mL) in the presence or absence of **NBY**, **MNPY**, **NPY** (0.25 mM), or **NPPY** (0.5 mM) in 6-well plates. After 24 h of incubation, cells were washed with chilled PBS (2 mL) and lysed in mammalian protein extraction buffer (200 µL, GE Healthcare) with complete protease inhibitor cocktail (Sigma) on ice. Cell lysates were cleared through centrifugation (13,200 rpm, 4 °C, 20 min). The protein lysate was boiled with loading buffer and analyzed by 10% SDS-PAGE. After gel electrophoresis, proteins were transferred to a polyvinylidene difluoride (PVDF) membrane (GE Healthcare) and blocked with TBS with 0.1% Tween 20 (Fisher Scientific) and 5% milk for 1 h. The blots were probed and incubated with the primary anti-HA antibody (Y-11 rabbit polyclonal IgG, sc-805, Santa Cruz Biotech) overnight at 4 °C, followed by a goat anti-rabbit IgG-HRP secondary antibody (sc-2031, Santa Cruz Biotech) for 1 h at room temperature. The blots were washed with TBST three times, and then incubated in working solution (Thermo Scientific SuperSignal West Pico Chemiluminescent substrate) for 5 min. The blots were imaged by ChemiDoc (BioRad).

### **Expression of caged firefly luciferase and light activation.**

The plasmid pMbNBYRS-4PyLT was constructed containing both pCMV-MbNBYRS and p4CMVE-U6-PyLT. A TAG amber stop codon was introduced at the Y340 site using primers P9 and P10 with a QuikChange mutagenesis kit (Agilent Technologies), generating the pGL3-Y340TAG plasmid. Subsequently, the RLuc gene was amplified using primers P11 and P12, which contain *Hind*III and *Xba*I restriction sites for ligation into pGL3-Y340TAG to achieve an expression vector of fusion protein caged-FLuc-RLuc, pGL3-Y340TAG\_RLuc. HEK 293T cells were cultured in DMEM (Gibco) supplemented with 10% FBS (Sigma), 1% Pen-Strep (Corning Cellgro) and 2 mM L-glutamine (Alfa Aesar) in 96-well plates (BD Falcon) in a humidified atmosphere with 5% CO<sub>2</sub> at 37 °C. At 80-90% confluence, cells seeded on plates were co-transfected with pMbNBYRS-4PyLT and pGL3-Y340TAG\_RLuc (100 ng of each plasmid) using LPEI (1.5 µL, 0.323 mg/mL), and the medium was changed to fresh DMEM supplemented without or with **NBY**, **MNPY**, **NPY**, or **NPPY** (0.25 mM). After 24 h incubation, the medium was changed to DMEM without phenol red, and the cells were irradiated with UV light (365 nm) for 15 s using a 365-nm UV lamp (high-performance UV transilluminator, UVP, 25 W) or kept in the dark. Cells were lysed by addition of 20 µL of substrate solution (Promega) in a 96-well plate (Greiner) and luminescence was measured using a dual luciferase assay kit (Promega) and a microplate reader (Tecan M1000) with an integration time of 2 s.

### **Temporal control of activation of caged firefly luciferase.**

Transfected HEK 293T cells were treated as described above before light activation. After

changing to fresh DMEM, cells were irradiated with UV light (365 nm) for different durations (0, 15, 30, 60, and 120 s) using a 365-nm UV lamp (high-performance UV transilluminator, UVP, 25 W) or were kept in the dark. Cells were lysed by addition of 100  $\mu$ L of substrate solution (Promega) in a 96-well plate (Greiner), and luminescence was measured using a Bright-Glo luciferase assay kit (Promega) on a microplate reader (Tecan M1000) with an integration time of 1 s.

#### **Spatial control of activation of caged firefly luciferase.**

HEK 293T cells were co-transfected with pMbNBYRS-4PyIT and pGL3-Y340TAG (2  $\mu$ g of each plasmid) using LPEI (20  $\mu$ L, 0.323 mg/mL) with caged tyrosine analogs in 35-mm dishes (BD falcon). Petri dishes were exposed to UV irradiation for 5 min (Dual UV transilluminator, 20 mW/cm<sup>2</sup>) in a “smiley face” pattern covered with aluminum foil and imaged by ChemiDoc (Bio-Rad). Bright luminescence was only observed for cells containing activated FLuc in the patterned region, while no luminescence was detected in the non-irradiated regions. In addition, spatial control of light-activated FLuc in living cells was performed with **NBY**, **MNPY**, and **NPY** in “P”, “I”, and “T” letter patterns (forming “PITT”), respectively.

#### **Optical control of TEV protease.**

HEK 293T cells were co-transfected with pTEV-Y178TAG, pMbNPYRS-4PyIT, and pGloSensor (100 ng of each plasmid) using LPEI (2  $\mu$ L, 0.323 mg/mL) in the absence or presence of **NBY**, **MNPY**, and **NPY** (0.25 mM) in a 96-well plate. After a 24 h incubation, the cells were either irradiated for 2 min (365 nm, 25 W) or kept in the dark. Cells were lysed by

addition of 100 µL of Bright-Glo assay reagent (Promega) in the 96-well plate (Costar), and luminescence was measured on a Tecan M1000 microplate reader (integration time of 1 s).

## 7.0 GENETIC ENCODING OF PHOTOSWITCHABLE AMINO ACIDS

This chapter will describe the genetic incorporation of photoswitchable amino acids into proteins in both bacterial and mammalian cells, and discuss the application of photoswitchable azobenzenes in the reversible photocontrol of protein function through site-specific installation in live cells.

### 7.1 INTRODUCTION TO AZOBENZENE PHOTOSWITCHES

This material was reprinted, in part, with permission from Luo, J.; Samanta, S.; Convertino, M.; Dokholyan, N. V.; Deiters, A. *Chembiochem* 2018, 19(20), 2178–2185.

As a classical light-switchable motif, azobenzene has found extensive biological application, as it undergoes light-induced, wavelength-selective *cis/trans* isomerizations with a large quantum yield and significant structural differences between the two isomers.<sup>192–194</sup> Azobenzene derivatives have been shown to enable reversible photocontrol of the function of nucleic acids, peptides, enzymes, receptors, and ion channels in cells and in animals.<sup>195–203</sup> In order to improve the azobenzene photoswitches that can be isomerized upon irradiation with visible light, substitutions on both benzene rings have been modified. One approach is to red shift the  $\pi \rightarrow \pi^*$  band, or to obtain a splitting of the  $n \rightarrow \pi^*$  band of the *E-/Z*-isomer, resulting in

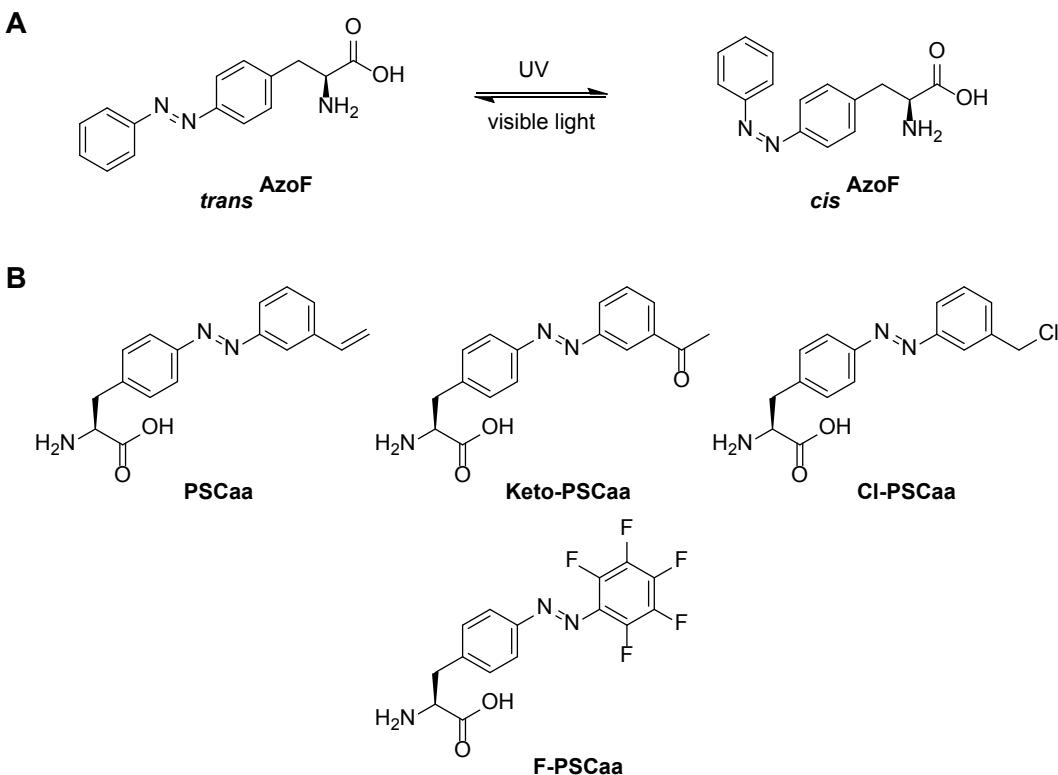
good switching properties.<sup>204-205</sup> Another way is to enhance the thermal stability of the Z-isomer through a red-shift of the  $n \rightarrow \pi^*$  band.<sup>206</sup>

The ability to site-specifically insert azobenzene photoswitches into proteins through genetic code expansion would facilitate reversible light-switching of protein functions with high spatiotemporal control in live cells. Through the application of an orthogonal *Mj*TyrRS/tRNA<sup>Tyr</sup> pair, the incorporation of photoisomerizable azobenzene (4-azophenylphenylalanine) was first reported by Schultz and co-workers (Figure 7.1A).<sup>207</sup> 4-Azophenylphenylalanine was incorporated into the *E. coli* catabolite activator protein (CAP) and was applied to the optical control of the binding affinity of CAP to its promoter. However, *Mj*TyrRS/tRNA<sub>CUA</sub> could not be used in eukaryotic cells, because its tRNA<sub>CUA</sub> is recognized by endogenous eukaryotic aminoacyl-tRNA synthetases (aaRSs). Due to this bioorthogonality limitation, its application was restricted to *E. coli*.

The Wang group reported genetic code expansion with three photoswitchable amino acids that can undergo bioconjugation reactions (Figure 7.1B) using an evolved *Mm*PylRS with 9 mutations and showed that calmodulin (CaM) conformation could be reversibly altered upon irradiation with UV light after generating an azo bridge.<sup>17</sup> In order to build a covalent protein bridge in proteins, a reactive functional group, such as an alkene, keto, and benzyl chloride group was introduced at the *para* position to the azo group for further chemical modification. Thus the azobenzene derivatives through selective bioconjugation reactions can be photoisomerized upon light illumination and altered the protein conformation. The alkene group can be reacted with the thiol of a nearby cysteine under mild thiol-ene conditions. The keto group is bio-orthogonal for protein modification through reaction with hydrazide or hydroxylamine derivatives. The benzyl

chloride group selectively reacts with the thiol functionality of a nearby cysteine, creating a covalent azobenzene protein bridge in CaM.<sup>17</sup>

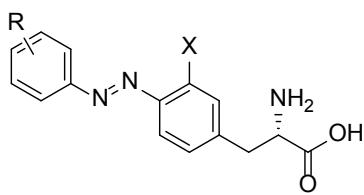
In order to render the protein function responsive to visible light, a pentafluoro photoswitchable click amino acid (F-PSCaa, Figure 7.1B) was developed, which enabled formation of an azobenzene bridge in CaM through an intramolecular nucleophilic aromatic substitution ( $S_NAr$ ) reaction with a cysteine residue.<sup>18</sup> The conformational changes of CaM driven by the photoisomerization of the azo bridge were detected using circular dichroism (CD), but reversible switching of protein activity upon illumination was not demonstrated. In fact, the reversible optical modulation of protein conformation through genetic encoding of PCSaa was exclusively carried out in test tubes, not in live cells. Also, this application was based on an intramolecular reaction and this design was dependent on the position and distance of suitable cysteine residues,<sup>208</sup> thereby limiting its general applicability.



**Figure 7.1:** Structures of photoswitchable phenylalanine analogs.

A) 4-azophenylphenylalanine (**AzoF**) reported by the Schultz lab. B) Structures of photoswitchable bioconjugation amino acids (PCSaa) reported by the Wang group.<sup>208-209</sup>

The Lin lab reported a series of red-shifted azobenzene amino acids and incorporated the azobenzene derivatives into sfGFP in *E. coli* using the same mutant *MmPylRS* as reported previously (Figure 7.2).<sup>210</sup> One of the sfGFP mutants showed robust photoswitching upon irradiation with alternating green-blue lights, but the results were limited to monitoring changes in absorbance at 340 nm ( $\pi \rightarrow \pi^*$  transition) and no regulation of protein function was demonstrated.<sup>210</sup>



X	R
H	2,6-F <sub>2</sub>
H	2,4,6-F <sub>3</sub>
H	2,3,4,5,6-F <sub>5</sub>
F	2,6-F <sub>2</sub>
F	2,4,6-F <sub>3</sub>
F	2,3,4,5,6-F <sub>5</sub>
Cl	2,6-F <sub>2</sub>
Cl	2,4,6-F <sub>3</sub>
Cl	2,3,4,5,6-F <sub>5</sub>

**Figure 7.2:** Structures of photoswitchable amino acids that were genetically encoded in *E. coli* by the Lin group.<sup>210</sup>

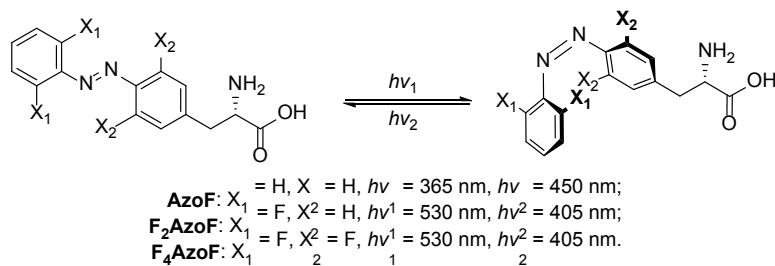
Adapted with permission from John *et al*, *Org. Lett.* 2015, 17, 6258–6261. Copyright 2015 American Chemical Society.

In this chapter, the efficient incorporation of a set of photoswitchable azobenzene amino acids into target proteins at a defined site using evolved *MbPylRS/tRNA<sup>Pyl</sup>* pairs in both bacterial and mammalian cells will be discussed. In addition, proteins will be engineered with UAAs to reversibly photocontrol their function in live cells.

## 7.2 INCORPORATION OF AZOPHENYLALANINE ANALOGS

The previously reported photoswitchable amino acid 4-azophenylphenylalanine (**AzoF**) was synthesized by Dr. Rajendra Upadhyay and Dr. Subhas Samanta in the Deiters lab (Figure 7.1). In addition, for the ability to reversibly control protein function with a minimum of light exposures,

two azobenzene derivatives were genetically encoded with greatly increased thermal stability of the *cis* isomer. The fact that fluoro substitutions at the *ortho*-position of azobenzenes significantly increase the stability of the *cis* form without increasing the overall size of the molecule guided us to synthesize *p*-azo(2,6-difluorophenyl) phenylalanine (**F<sub>2</sub>AzoF**) and *p*-azo(2,6,2',6'-tetrafluorophenyl)phenylalanine (**F<sub>4</sub>AzoF**, Figure 7.3). Here we show that these photoswitchable amino acids can be efficiently incorporated in both bacterial and mammalian cells, and demonstrate that site-specific installation of **AzoF**, **F<sub>2</sub>AzoF**, and **F<sub>4</sub>AzoF** can be applied to the reversible photocontrol of protein function with high spatiotemporal control in live cells.



**Figure 7.3:** Photoisomerization of azobenzene analogs **AzoF**, **F<sub>2</sub>AzoF**, and **F<sub>4</sub>AzoF**.

**AzoF**, **F<sub>2</sub>AzoF**, and **F<sub>4</sub>AzoF** were synthesized by Dr. Subhas Samanta in the Deiters lab.

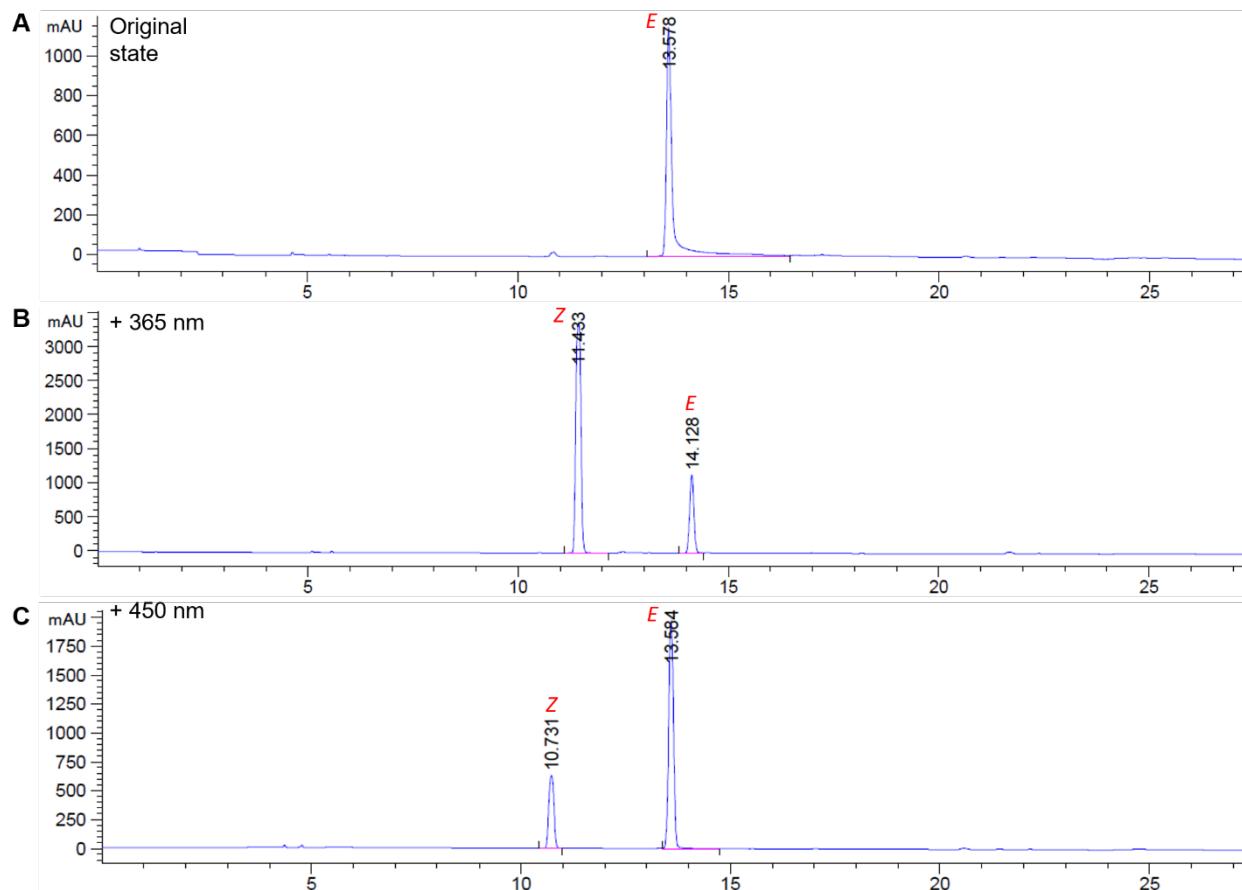
To determine the photoswitching efficiency, a solution of **AzoF** in DMSO/PBS buffer (pH 7.4) was treated consecutively at 365 nm UV light (by a UV transilluminator) or visible white light, and the *E/Z* ratios of the corresponding photostationary state (PSS) were determined by HPLC. Irradiation with UV light at 365 nm causes *E*→*Z* isomerization, resulting in a decrease of the  $\pi \rightarrow \pi^*$  band typically observed in azobenzenes, producing a PSS containing 78% of the corresponding *Z*-isomer (Figure 7.4B). Irradiation with visible light was used to isomerize **AzoF**, producing PSS containing 76% of *E*-isomer with 448 nm light (Figure 7.4C). In the visible part of

the UV/Vis spectrum, *E*→*Z* isomerization results in the characteristic increase of the n→π\* absorption compared to the parent *trans*-**AzoF** (Figure 7.7A–B). These results (absorption spectra and photostationary states) are consistent with previous reports of **AzoF**.<sup>193</sup>

Isomerization of the two amino acids **F<sub>2</sub>AzoF** and **F<sub>4</sub>AzoF** in response to green and blue light exposure was evaluated in DMSO/PBS buffer (pH 7.4), and the *E/Z* ratios of the corresponding PSS were determined by HPLC. Illumination of **F<sub>2</sub>AzoF** and **F<sub>4</sub>AzoF** with green light ( $\lambda = 530$  nm) results in *E*-to-*Z* isomerization yielding 82% and 91% of the *Z* isomer, respectively (Figure 7.5B and Figure 7.6B). Illumination using blue light ( $\lambda = 405$  nm) results in *Z*-to-*E* isomerization restoring 70% and 84% of the *E* isomer, respectively (Figure 7.5C and Figure 7.6C). In addition, the UV/vis spectra of **F<sub>2</sub>AzoF** and **F<sub>4</sub>AzoF** show typical features of azobenzene (Figure 7.7C–F). Two absorbance peaks were observed representing π–π\* and n–π\* transitions, and the π–π\* absorption decreased upon *E* to *Z* isomerization while the n–π\* absorption increased (Figure 7.7C–F). Notably, tetrafluoro substitution of **F<sub>4</sub>AzoF** leads to better separation in wavelength of both transition bands between the *Z* and *E* isomer than **F<sub>2</sub>AzoF**, and the separation at the n–π\* transition band (Figure 7.7D and Figure 7.7F) enabled the use of green light for the formation of the *Z* isomer. Notably, the photoisomerization yields induced by visible light are in agreement with the literature (Table 7.1).<sup>204</sup>

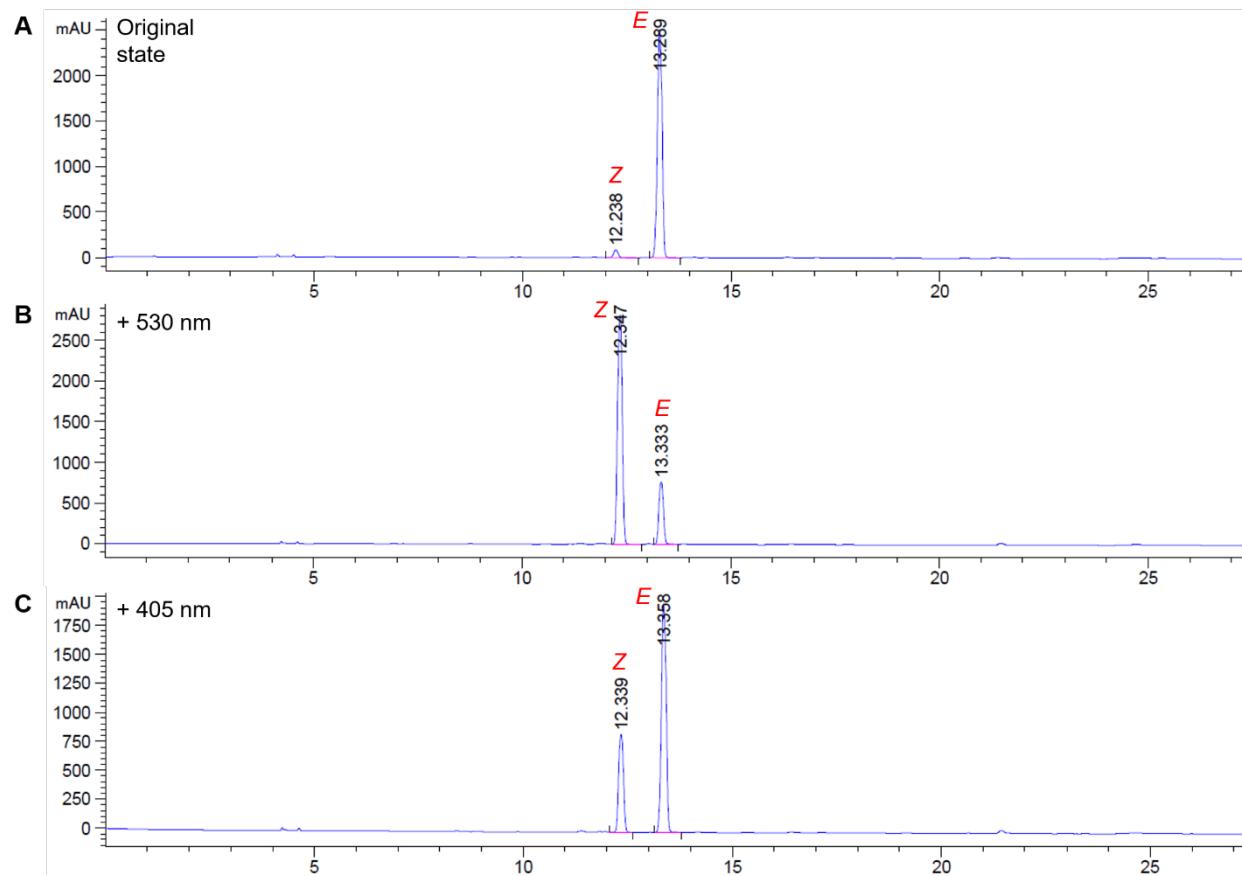
In order to identify the optimal wavelengths for reversible photoswitching of **F<sub>2</sub>AzoF** and **F<sub>4</sub>AzoF** in proteins, the absorption spectra upon illumination with visible blue/green light (405 nm/530 nm) were recorded (Figure 7.7). The 530 nm light was selected for *trans* to *cis* switching since the n → π\* transition in the visible region was well separated for the *trans*- and *cis*-isomer of **F<sub>4</sub>AzoF** (Figure 7.7E–F), while it didn't separate the *trans*- and *cis*-isomer of **F<sub>2</sub>AzoF** well (Figure 7.7C–D). The 405 nm light was used for *cis* to *trans* switching, but it seems this

wavelength was not able to fully convert the *cis* isomer of **F<sub>4</sub>AzoF** (green curve, Figure 7.7E-F) back to the *trans* isomer as compared to the parent **F<sub>4</sub>AzoF** stored in dark (blue curve).



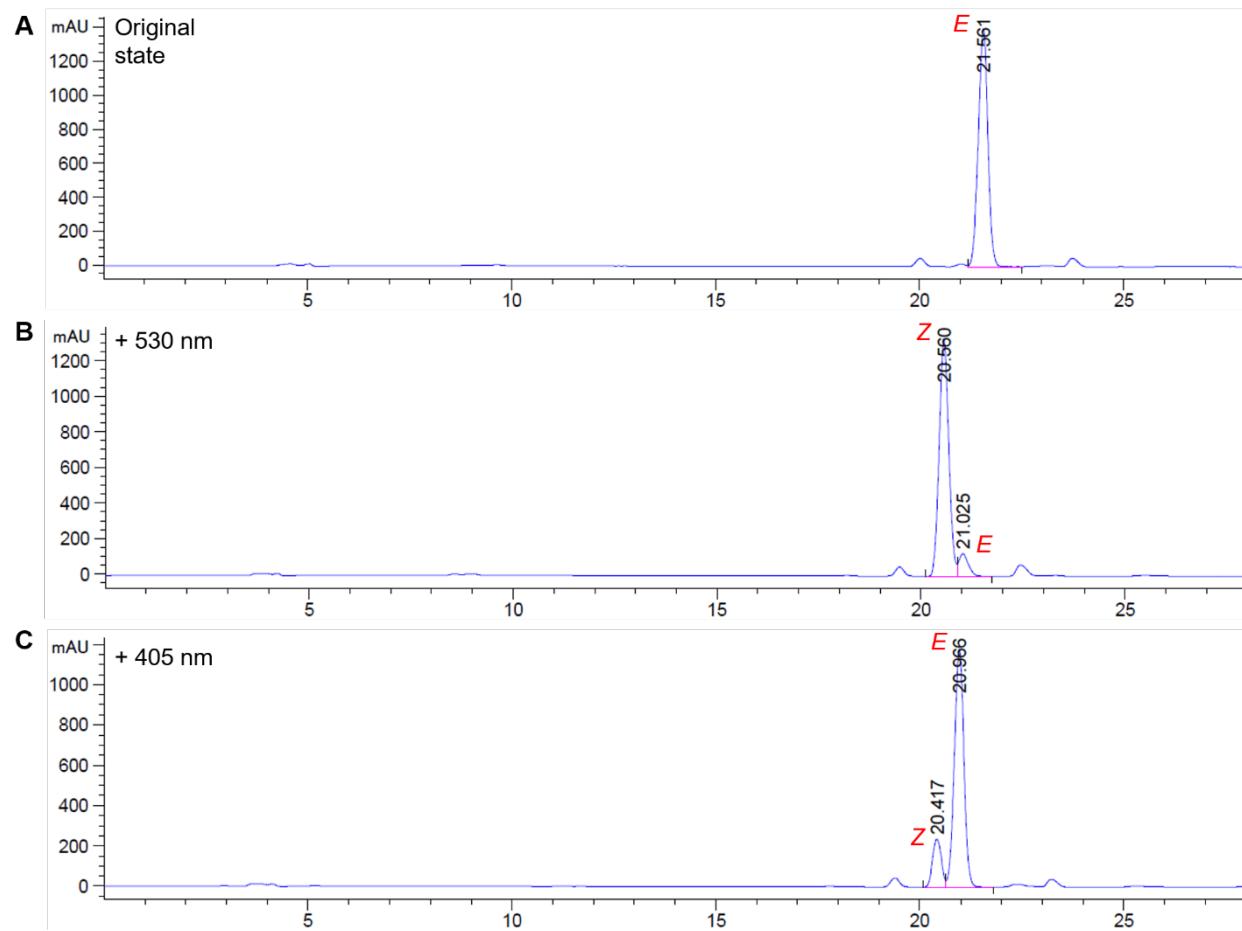
**Figure 7.4:** HPLC profiles of the photoequilibrium of **AzoF** after illumination.

HPLC profiles of the photoequilibrium after illumination of **AzoF** with UV light ( $\lambda = 365$  nm, **B**) and blue light ( $\lambda = 450$  nm, **C**).



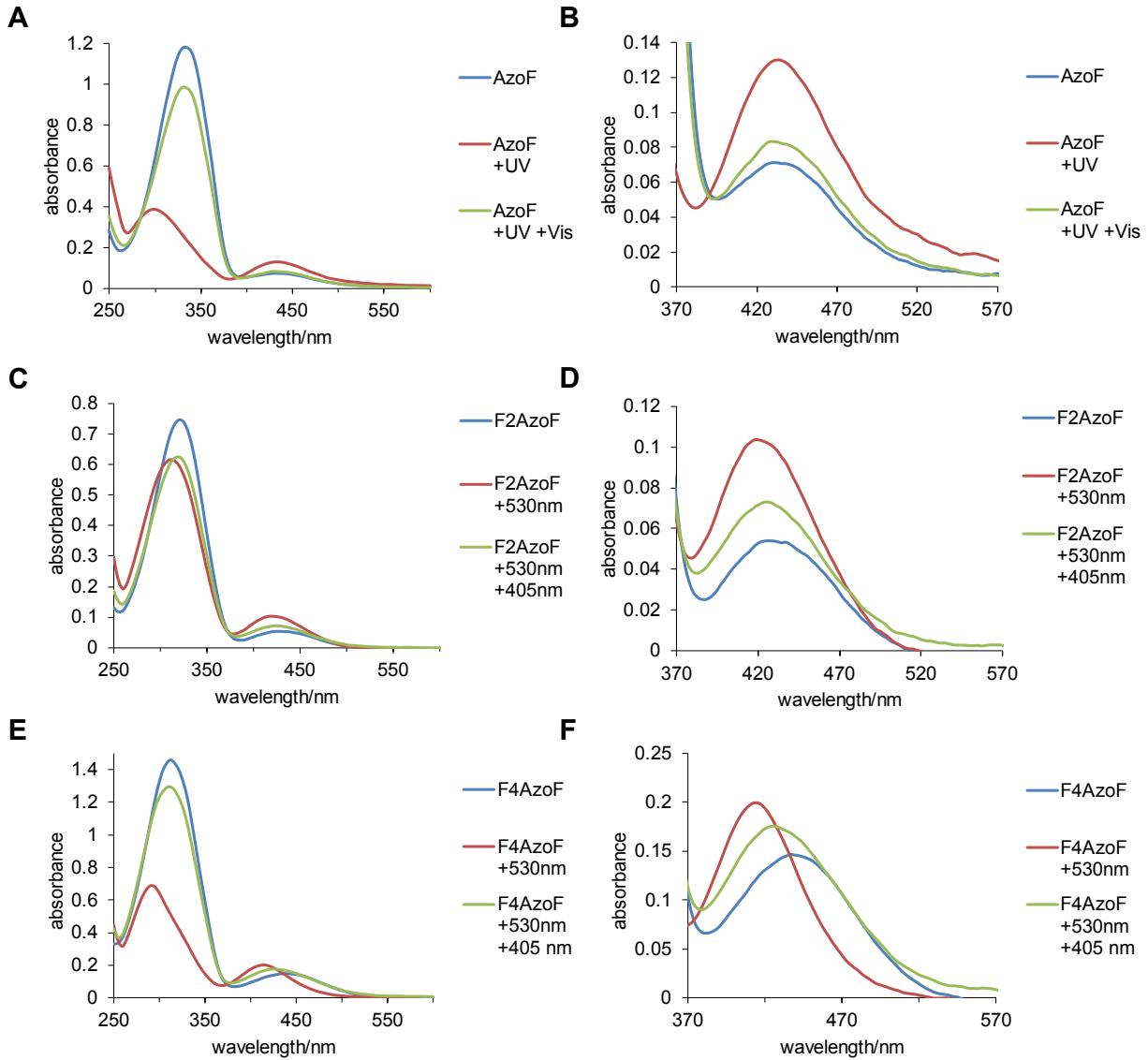
**Figure 7.5:** HPLC profiles of the photoequilibrium of  $\text{F}_2\text{AzoF}$  after illumination.

HPLC profiles of the photoequilibrium after illumination of  $\text{F}_2\text{AzoF}$  with green light ( $\lambda = 530 \text{ nm}$ , **B**) and blue light ( $\lambda = 405 \text{ nm}$ , **C**).



**Figure 7.6:** HPLC profiles of the photoequilibrium of **F<sub>4</sub>AzoF** after illumination.

HPLC profiles of the photoequilibrium after illumination of **F<sub>4</sub>AzoF** with green light ( $\lambda = 530$  nm, **B**) and blue light ( $\lambda = 405$  nm, **C**).



**Figure 7.7:** Absorption spectra of *trans*- and *cis*-photoisomers.

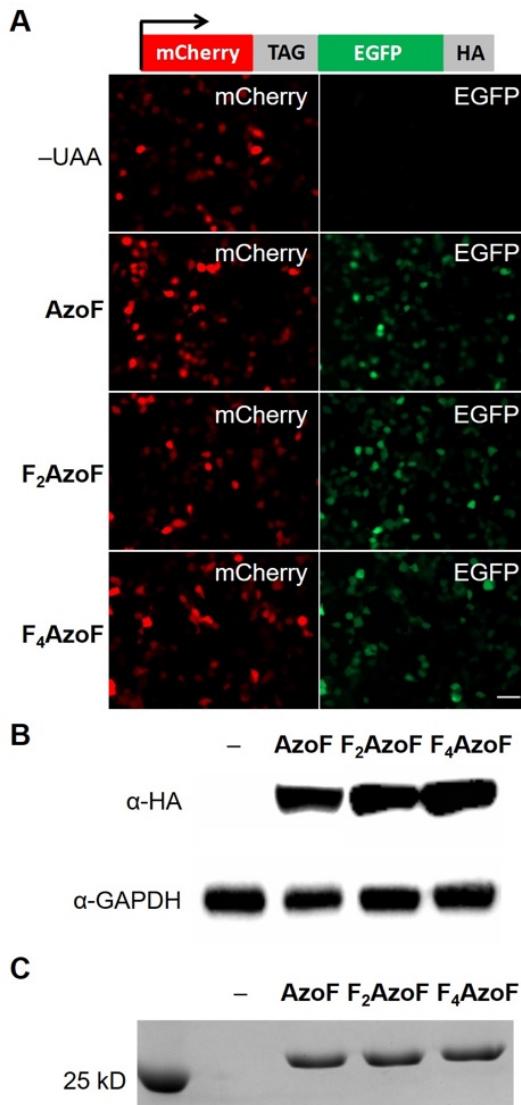
A) UV/Vis spectra of the *trans*-photoisomer (>99% *E*, blue, stock in the dark-adapted state), *cis*-photoisomer (78% *Z*, red, UV light), and *trans*-photoisomer (76% *E*, green, visible white light) of **AzoF** in PBS (pH 7.4). B) Zoom-in on the n- $\pi^*$  transition band. C) UV/Vis spectra of the *trans*-photoisomer (>99% *E*, blue, stock in the dark-adapted state), *cis*-photoisomer (82% *Z*, red, 530 nm), and *trans*-photoisomer (70% *E*, green, 405 nm) of **F<sub>2</sub>AzoF** in PBS (pH 7.4). D) Zoom-in on the n- $\pi^*$  transition band. E) UV/Vis spectra of the *trans*-photoisomer (>99% *E*, blue, in the dark-adapted state), *cis*-photoisomer (91% *Z*, red, 530 nm), and *trans*-photoisomer (84% *E*, green, 405 nm) of **F<sub>4</sub>AzoF** in PBS (pH 7.4). F) Zoom-in on the n- $\pi^*$  transition band.

**Table 7.1:** Measured absorption maxima ( $\lambda_{\text{max}}$ , orange) and photostationary states (blue) of azobenzene amino acids in DMSO/PBS (pH 7.4) at the indicated wavelengths.

UAA	AzoF	F <sub>2</sub> AzoF	F <sub>4</sub> AzoF
$\lambda_{\text{max}}$	335 nm ( $\pi \rightarrow \pi^*$ )	435 nm ( $n \rightarrow \pi^*$ )	447 nm ( $n \rightarrow \pi^*$ )
<i>trans</i> isomer			
$\lambda_{\text{max}}$	300 nm ( $\pi \rightarrow \pi^*$ )	420 nm ( $n \rightarrow \pi^*$ )	417 nm ( $n \rightarrow \pi^*$ )
<i>cis</i> isomer			
<i>trans:cis</i> ( $h\nu_1$ )	22:78 (365 nm)	18:82 (530 nm)	9:91 (530 nm)
<i>trans:cis</i> ( $h\nu_2$ )	76:24 (450 nm)	70:30 (405 nm)	84:16 (405 nm)

To genetically incorporate **AzoF** into proteins in mammalian cells, a screen was carried out using a panel of 10 rationally-designed *MbPylRS* mutants, which were created with select mutations at six positions (L270, Y271, L274, N311, C313, and Y349) and have previously been shown to accept structurally diverse phenylalanine derivatives as substrates.<sup>185, 211</sup> The synthetase panel was screened in mammalian cells using a plasmid encoding a dual fluorescent reporter gene – mCherry-EGFP containing a TAG mutation on the linker between mCherry and EGFP.<sup>212</sup> Two PylRS mutants led to EGFP expression in the presence of **AzoF** (Figure 7.8A). The first PylRS, named AzoFRS1, has five mutations – Y271M, L274A, N311A, C313A, and Y349F. The second one, termed AzoFRS2, contains L270F, L274M, N311G, C313G, and Y349F mutations. Fluorescence imaging revealed that mCherry expressions were observed as expected, AzoFRS1/2 showed no EGFP expression in the absence of **AzoF**, while EGFP expression was detected when **AzoF** (0.5 mM) was added to the media. This confirms the fidelity of AzoFRS1/2 for incorporation of **AzoF** in mammalian cells. Because fluorine has a similar size to hydrogen and the symmetric fluoro-substituents minimally change the planar geometry of the *trans* isomer,<sup>204, 213</sup> F<sub>2</sub>AzoF and F<sub>4</sub>AzoF do not significantly increase the steric demand on the synthetase. Thus, incorporation of F<sub>2</sub>AzoF and F<sub>4</sub>AzoF into proteins in mammalian cells was

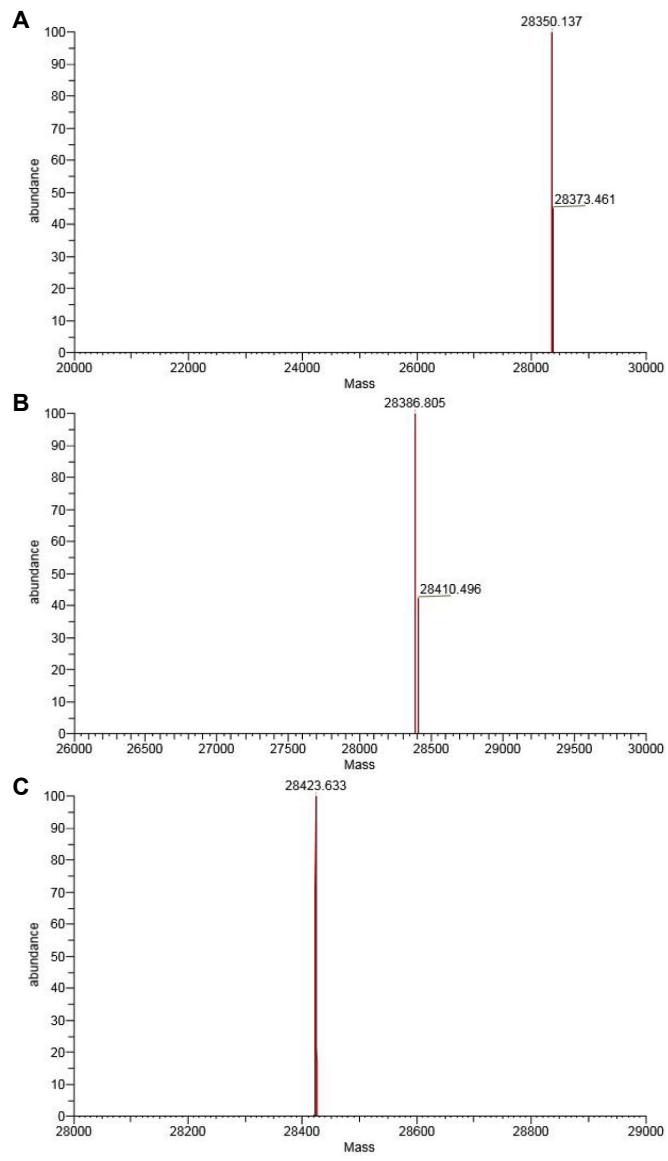
first tested using the same *MbAzoFRS1*/tRNA pair. Unexpectedly, very low EGFP expression was observed, suggesting that *MbAzoFRS1* did not allow for incorporation of **F<sub>2</sub>AzoF** and **F<sub>4</sub>AzoF** in mammalian cells. Gratifyingly, the AzoFRS2 synthetase was found to incorporate not only **AzoF**, but also **F<sub>2</sub>AzoF** and **F<sub>4</sub>AzoF** with high efficiency. Fluorescence imaging revealed that both mCherry and EGFP expression was detected with the addition of **AzoF**, **F<sub>2</sub>AzoF** and **F<sub>4</sub>AzoF** (0.5 mM) to the media, while no EGFP-expressing cells were observed in the absence of the UAAs (Figure 7.8A). In addition, full-length mCherry-EGFP protein was detected from HEK 293T cells using an HA-tag antibody, confirming full-length protein expression at sufficient levels (Figure 7.8B). Subsequently, incorporation of **AzoF**, **F<sub>2</sub>AzoF**, and **F<sub>4</sub>AzoF** in *E. coli* was investigated using *MbAzoFRS2*. As expected, it showed robust levels of sfGFP expression in the presence of **AzoF**, **F<sub>2</sub>AzoF** and **F<sub>4</sub>AzoF** (1 mM), while no sfGFP expression in the absence of azobenzene amino acids was observed (Figure 7.8C).



**Figure 7.8:** Genetically encoding of azobenzene amino acids in pro- and eukaryotic cells.

A) Fluorescence micrographs of HEK 293T cells expressing mCherry-EGFP in the presence of pMbAzoFRS2-mCherry-TAG-EGFP-HA and PyltRNA<sub>CUA</sub> incubated in the absence or presence of **AzoF**, **F<sub>2</sub>AzoF**, and **F<sub>4</sub>AzoF**. B) Western blots of mCherry-AzoF/F<sub>2</sub>AzoF/F<sub>4</sub>AzoF-EGFP-HA with an anti-HA tag antibody and an anti-GAPDH antibody proving the fidelity of the incorporation of **AzoF**, **F<sub>2</sub>AzoF**, and **F<sub>4</sub>AzoF** in response to a TAG codon in mammalian cells. C) Genetic incorporation of **AzoF**, **F<sub>2</sub>AzoF**, and **F<sub>4</sub>AzoF** using MbAzoFRS2 in *E. coli*. Coomassie stained gel of Ni-NTA purified sfGFP-Y151TAG expressed in *E. coli* in the absence and presence of **AzoF**, **F<sub>2</sub>AzoF**, and **F<sub>4</sub>AzoF**.

Furthermore, the incorporation was also confirmed by ESI-MS analysis of purified proteins, revealing a mass of [sfGFP-**AzoF**] and [sfGFP-**AzoF** + Na<sup>+</sup>] of 28350.137 Da and 28373.461 Da (expected MS: 28350.20 Da with Na<sup>+</sup>) for incorporation of **AzoF**, a mass of [sfGFP-**F<sub>2</sub>AzoF**] and [sfGFP-**F<sub>2</sub>AzoF** + Na<sup>+</sup>] of 28386.805 Da and 28410.496 Da (expected MS: 28386.20 Da with Na<sup>+</sup>) for incorporation of **F<sub>2</sub>AzoF**, and a mass of 28423.633 Da (calculated MS: 28422.80 Da) for incorporation of sfGFP-**F<sub>4</sub>AzoF** (Figure 7.9).



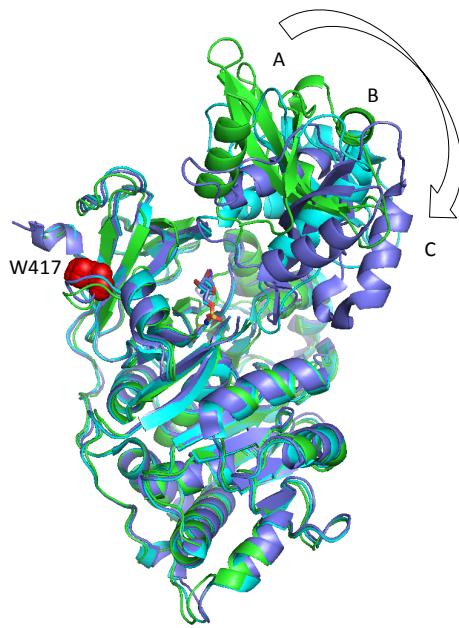
**Figure 7.9:** ESI-MS analysis of sfGFP-AzoF, -F<sub>2</sub>AzoF, and -F<sub>4</sub>AzoF.

A) sfGFP-AzoF: observed MS: 28350.137 Da, expected MS: 28350.20 Da; B) sfGFP-F<sub>2</sub>AzoF: observed MS: 28386.805 Da, expected MS: 28386.20 Da; C) irradiated sfGFP-F<sub>4</sub>AzoF: observed MS: 28423.633 Da, expected MS: 28422.80 Da.

### 7.3 REVERSIBLE OPTICAL CONTROL OF PROTEIN FUNCTION IN LIVING CELLS

Firefly luciferase was selected as a proof-of-principle target in order to demonstrate the utility of the photoswitchable amino acid to reversibly photocontrol protein function in live mammalian cells. Photoactivatable luciferases have been developed through the introduction of photocleavable caging groups by us and by the Chen lab.<sup>107, 214-215</sup> Yet, reversible photocontrol of a bioluminescence enzyme has not been achieved. A genetically encoded photoswitchable luciferase provides an additional tool for the cellular detection of ATP concentrations at different locations and different time points.<sup>216</sup>

To possibly predict potential residues for the incorporation of **AzoF** and achieve photoswitching of FLuc without the need for extensive mutational analysis, computational modeling and protein stability calculations were carried out by the Dokholyan lab at UNC for the three different conformational states of firefly luciferase in the two-step chemiluminescence reaction: the open inactive state (state A), the closed active state for the generation of luciferyl-AMP (state B), and the closed active state for the generation of oxyluciferin (state C), as shown in Figure 7.10.



**Figure 7.10:** Crystal structures of firefly luciferase in three states.

- 1) open inactive state (state A, PDB ID: 1LCI); 2) closed active state for the generation of luciferyl-AMP (state B, PDB ID: 4G36); 3) closed active state for the generation of oxyluciferin (state C, PDB ID: 4G37).

In order to evaluate the propensity of luciferase residues to adopt multiple conformational states in the apo-form or bound to its natural substrate [ $5'$ -O-(*N*-(dehydroluciferily)sulfamoyl] adenosine), a series of computational analyses were carried out using Eris, a methodology that efficiently and accurately computes the stability changes of proteins upon mutations.<sup>217</sup> A measure of protein stability is the difference between the free energies of the folded and unfolded states ( $\Delta G$ ). The extent to which mutations in specific residues perturb protein stability can be expressed as a  $\Delta\Delta G$  value for the mutation (the change in Gibbs free energy difference upon mutation between the wild-type and mutant forms of the protein:  $\Delta\Delta G = \Delta G_{\text{mutant}} - \Delta G_{\text{wild-type}}$ ) and it can be considered an estimate of protein stabilization/destabilization.<sup>218</sup> A lower (more negative)  $\Delta\Delta G$  value means that a mutation improves the thermodynamic stability of the protein. Not only does the magnitude of  $\Delta\Delta G$  provide useful information about proteins, but the variance

of estimated  $\Delta\Delta G$  values provides an estimate of conformational flexibility. Protein regions with multiple conformations will show a broad range of calculated  $\Delta\Delta G$  values, determined by modeling the local repacking of the protein through changes in side-chain conformations,<sup>219</sup> which can be used to predict if specific residues are likely to constitute an allosteric site.<sup>218</sup> The variance for the possible  $\Delta\Delta G$  values can be quantified by determining the standard deviation ( $\sigma\Delta G$ ) for a given set of  $\Delta\Delta G$ s. The higher this value is, the more conformational states are theoretically/computationally possible.

For each residue, the statistical significance of  $\sigma\Delta G$  was determined by its Zscore. This parameter compares the  $\sigma\Delta G$  for a given residue to the average  $\sigma\Delta G$  of all residues in a protein. For a given residue  $i$ , the Zscore( $i$ ) is defined as  $[\sigma\Delta G(i) - \text{average}(\sigma\Delta G(\text{all residues}))] / \text{standard deviation}(\sigma\Delta G(\text{all residues}))$ . Residues for which the Zscore values is greater than 2 have less than 97.7% chance to be randomly chosen. Residues identified in this way are quite flexible in their conformations, and therefore are likely to be part of an allosteric site.<sup>220</sup> Consequently, they are excellent candidates for substitution with **AzoF** in order to translate the configurational changes upon light-switching into changes in protein/enzyme function. Our computational modeling predicts that a series of allosteric sites in luciferase and we targeted aromatic residues at those sites for **AzoF** substitution in order to improve our chances for obtaining a photoswitchable luciferase enzyme. In the potential candidates (Zscore > 2.40, Table 7.2), F294 and F432 both have high Zscores (3.06 and 3.76, respectively); however, both are followed by sterically demanding F295 and F433 residues and based on the protein structure we expect severe steric hindrance in incorporating sterically demanding azophenylalanine analogs. Y340 (Zscore 3.46) is an essential catalytic residue in the ATP binding pocket of FLuc and thus was not modified.<sup>68</sup> W426 (Zscore 3.14) is very close to the flexible N-terminal of FLuc, thus

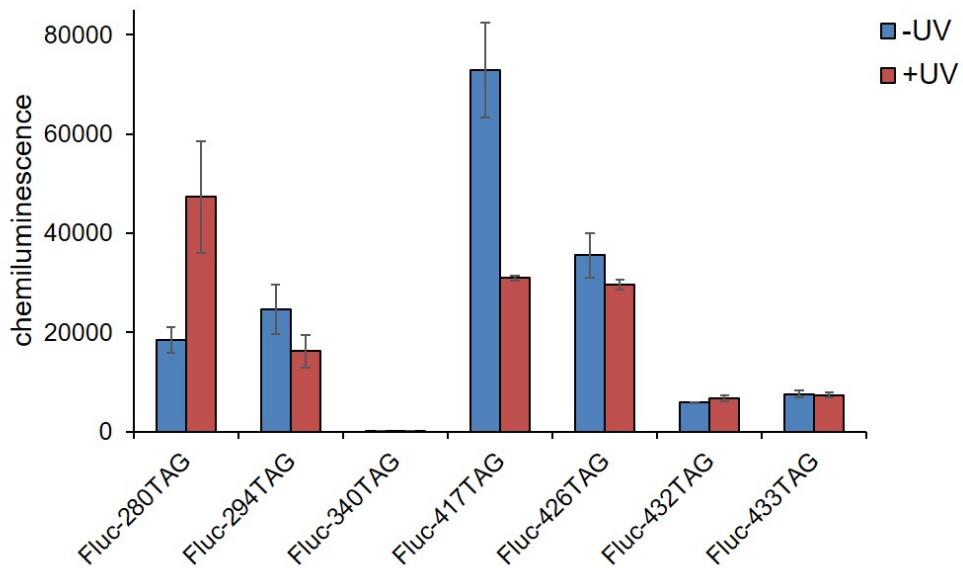
suggesting that it is not a good candidate to achieve optical switching of enzyme function. Of these possible sites, Y280TAG, F294TAG, Y340TAG, W417TAG, W426TAG, F432TAG, and F433TAG were selected for experimental validation.

**Table 7.2:** Potential residue candidates for azobenzene incorporation in luciferase (Zscore > 2).

Residue	$\sigma_{\Delta G}$ (kcal/mol)	Zscore
F268	1.98	2.15
F273	2.17	2.48
Y280	2.44	2.95
F292	2.39	2.85
F294	2.51	3.06
Y304	2.12	2.40
Y340	2.74	3.46
<b>W417</b>	<b>2.40</b>	<b>2.88</b>
W426	2.55	3.14
F429	2.91	3.76
F432	2.46	2.97
Y444	1.96	2.13
Y447	2.37	2.83
F465	2.25	2.62

Light-switching luciferase assays were carried out by incorporation of **AzoF** into FLuc at Y280TAG, F294TAG, Y340TAG, W417TAG, W426TAG, F432TAG, and F433TAG in HEK293T cells (Figure 7.11). To genetically incorporate **AzoF** into FLuc, HEK293T cells were co-transfected with the p*Mb*PylAzoFRS2/PyltRNA<sub>CUA</sub> expression plasmid and the corresponding pFLuc-TAG construct in the absence or presence of **AzoF**. After 24 h incubation, the transfected

cells were treated with UV light (5 min, 20 mW/cm<sup>2</sup>) or kept in the dark and FLuc activities were subsequently measured. Not surprisingly, most mutations either showed activity regardless of UV illumination or showed complete loss of activity, except for FLuc-W417→AzoF, which appears to be an allosteric site that allows for photoswitching of FLuc activity.



**Figure 7.11:** Mutant screening for reversible photoswitching of firefly luciferase containing **AzoF** in live mammalian cells.

Error bars represent standard deviations from three independent experiments.

As revealed by Eris calculations, the introduction of the azobenzene at position W417 may cause a decrease in the stabilization of the luciferase, particularly, more pronounced in state B and for the **AzoF** *cis*-isomer (Figure 7.4), which affects the allosteric rearrangement (conformational change) from state B to C or from state B to A (Figure 7.10). Thus, the light-triggered *cis/trans*-isomerization of azobenzene can affect the stability of the luciferase-luciferyl-AMP complex, which forms in the first step of luciferin oxidation, thereby interfering with the

oxidation process. Furthermore, the binding energy of D-luciferyl-AMP, as evaluated by MedusaScore,<sup>221</sup> an accurate force-field based scoring function which estimates binding energies by modeling physical interactions between proteins and small molecules, is slightly less favorable in the case of *cis*-FLuc-W417→**AzoF** in both the B and C states, making the formation of the protein-ligand complex less favorable for the bioluminescence reaction. Specifically, FLuc-W417→*cis*-**AzoF** shows higher interaction energy in states B and C ( $-30.9$  kcal/mol and  $-36$  kcal/mol, respectively) compared to wild-type FLuc ( $-31.9$  kcal/mol and  $-37.0$  kcal/mol, respectively) and FLuc-W417→*trans*-**AzoF** ( $-33.5$  kcal/mol and  $-41.9$  kcal/mol; Table 7.3). Based on the computational calculations, when W417→*trans*-**AzoF** isomerizes to the *cis* conformation, it may be possible that the amino acids within the luciferase binding site rearrange, resulting in overall destabilization of the protein, leading to higher energy. This may explain why the *cis*- and *trans*-W417→**AzoF** luciferase mutants display successive off-and-on switching behaviors upon light illumination. Overall, the protein stability complex formation energies (Table 7.2 and Table 7.3, respectively), calculated upon repacking around W417 mutants, are confirmed by the experimental results.

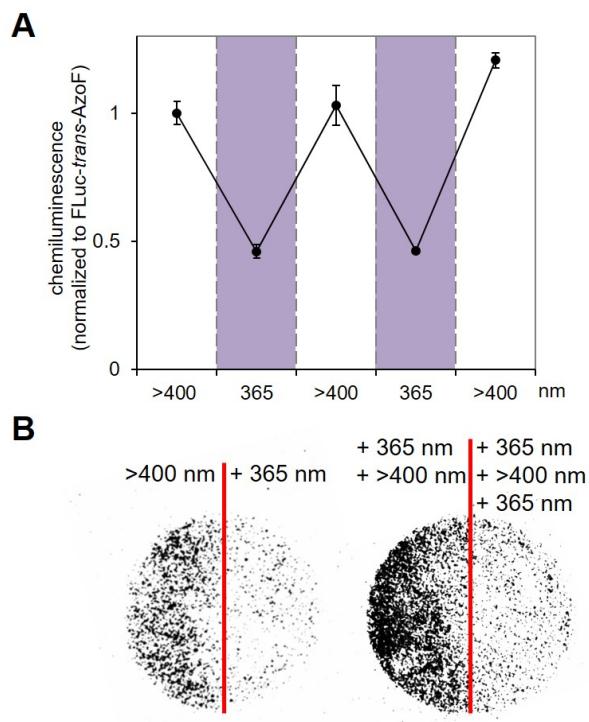
**Table 7.3:** Binding energy of D-luciferyl-AMP upon repacking at W417 for the wild-type enzyme in the **AzoF** mutant in the *trans* and *cis* geometry.

State	Geometry	MedusaScore (kcal/mol)
B	wt	-31.9
	<i>cis</i>	-30.9
	<i>trans</i>	-33.5
C	wt	-37
	<i>cis</i>	-36
	<i>trans</i>	<b>-41.9</b>

In order to demonstrate reversible photoswitching of protein function, HEK293T cells were subjected to alternating UV and white light exposures for five individual steps (5 min each). The activity of FLuc-W417→**AzoF** showed the expected reversible change with each step of light exposure, as activity decreased after UV-induced *trans* to *cis* photoswitching and was restored after *cis* to *trans* switching through visible light exposure (Figure 7.12A). The two different light treatments did not affect the activity of FLuc-TAG in the absence of **AzoF** (negative control; data not shown). These results support that **AzoF** is functioning as a reversible light-switch for protein activity in live human cells. The activity of the Fluc-W417→**AzoF** mutant was reduced in the case of the *cis*-isomer, while reversion to the *trans*-isomer led to recovery of activity.

To investigate whether the protein activity could be reversibly photoswitched for spatial control, two culture dishes of HEK293T cells expressing FLuc-W417→**AzoF** were irradiated alternately through a mask that covered half of each dish. In one dish (Figure 7.12B, left), the left side was illuminated with visible white light (>400 nm), while the right one was irradiated

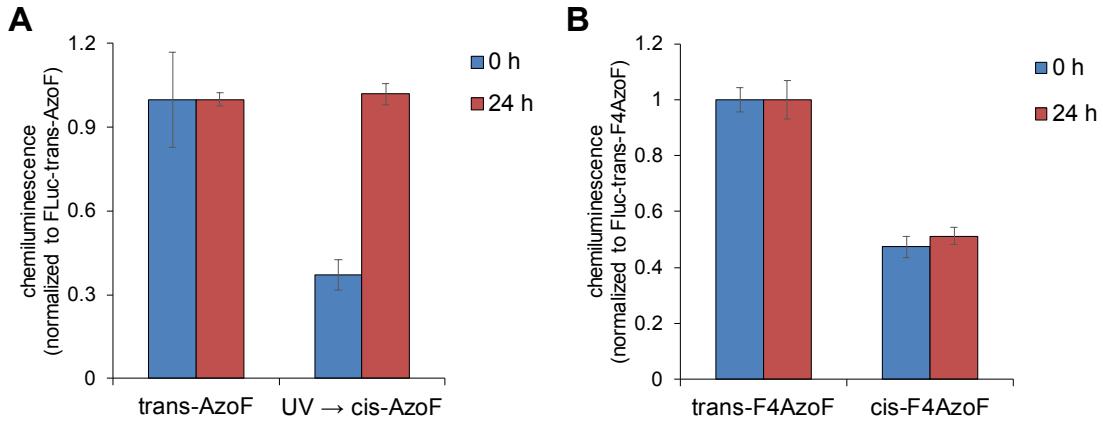
with UV (365 nm). In the other dish (Figure 7.12B, right), the left half was illuminated with UV (365 nm) followed by white light (>400 nm), while the right one was treated sequentially by UV (365 nm), white (>400 nm), and UV (365 nm) light. Thus, the half-dishes from left to right represent areas with increasing numbers of localized, reversible switching events. Cells that were exposed to visible white light (>400 nm) exhibited increased luminescence (*trans*-AzoF activates Fluc), while UV-exposed cells showed diminished luminescence (*cis*-AzoF inactivates Fluc), resulting in a visible “half-moon” pattern (Figure 7.12B). These spatial-control results are in agreement with the Bright-Glo luciferase assay results (Figure 7.12A) and demonstrate that AzoF can be efficiently and reversibly switched *in vivo*, allowing reversible control of firefly luciferase activity with temporal and spatial resolution.



**Figure 7.12:** Reversible photoswitching of Firefly luciferase containing a W417AzoF mutation in live mammalian cells.

A) HEK293T cells underwent five light-switching steps in total: >400 nm (visible light) → 365 nm (UV light) → >400 nm → 365 nm → >400 nm. Chemiluminescence was normalized to the activity of FLuc-W417→*trans*-AzoF that was kept in the dark. Error bars represent standard deviations from three independent experiments. B) Spatial control of reversible photoswitching of FLuc activity in live cells by exposing two halves of culture dishes containing transfected HEK293T cells to increasing numbers of switching events as indicated by the different wavelengths for visible light (>400 nm) and UV (365 nm) light.

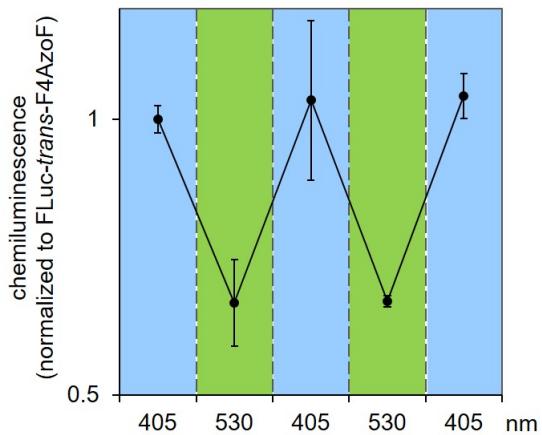
The drawbacks of AzoF include: 1) photoisomerization requires UV light, and 2) the *cis*-isomer is thermally unstable and reverts back to *trans* even in the absence of irradiation. To address these shortcomings, fluorine atoms were installed at the four *ortho* positions to induce a bathochromic shift that makes F<sub>4</sub>AzoF responsive to visible light, and also greatly improves the thermal stability of its *cis*-isomer and the isomer ratios in the photostationary states.<sup>204</sup> In order to demonstrate the thermostability of the *cis*-isomer of F<sub>4</sub>AzoF after incorporation into protein, the luciferase assays in live cells were either performed immediately or one day after light exposure to 530 nm (*trans* → *cis* switching) and 405 nm (*cis* → *trans* switching). Importantly, the enzymatic activity for FLuc-W417→*cis*-F<sub>4</sub>AzoF remains low due to stability of the *cis* isomer – even after 24 hours in live cells in a 37 °C incubator (Figure 7.13B). This is in contrast to *cis*-AzoF, which is fully converted back to *trans*-AzoF after 24 h incubation at 37 °C in the dark (Figure 7.13A). The ability to choose light-switchable amino acids with fast (AzoF) and slow (F<sub>4</sub>AzoF) thermal reversion rates enables switching experiments on short and long timescales, thereby further expanding the optochemical toolbox of light-controlled amino acids. In addition, the genetic encoding of F<sub>4</sub>AzoF provides more suitable switching wavelengths, red-shifted from the UV-A range.



**Figure 7.13:** Thermostability tests of **AzoF** and **F<sub>4</sub>AzoF**.

Switching experiment confirming that (A) *cis*-**AzoF** is not, but that (B) *cis*-**F<sub>4</sub>AzoF** is thermally stable when incorporated into protein in HEK293T cells. Cells were irradiated and then incubated for 24 h at 37 °C in the dark. Chemiluminescence was normalized to non-irradiated cells. Error bars represent standard deviations from three independent experiments.

The activity of FLuc-W417→**F<sub>4</sub>AzoF** was reversibly controlled through a *trans* → *cis* and *cis* → *trans* configurational change of **F<sub>4</sub>AzoF** upon 530 nm and 405 nm irradiation, respectively. Furthermore, photoswitching of FLuc activity was reproducible over five cycles (Figure 7.14), demonstrating that **F<sub>4</sub>AzoF** was functioning as a reversible photoswitch for FLuc activity in live cells. **F<sub>4</sub>AzoF** was irradiated with blue light (405 nm) in order to obtain a PSS of 84% *trans*-**F<sub>4</sub>AzoF** through photoisomerization, as reported in Table 7.1. Thus, in the first step, the dark-adapted *trans*-**F<sub>4</sub>AzoF** was exposed to 405 nm light so that all cells contained active FLuc at the same level. Cells were then exposed to green light (530 nm) followed by blue light (405 nm) two successive times, inducing isomerization of **F<sub>4</sub>AzoF**, effectively turning “off and on” FLuc function, allowing photoswitchable control of the enzymatic activity (Figure 7.14).



**Figure 7.14:** Reversible photoswitching of FLuc-W417→F<sub>4</sub>AzoF activity in live mammalian cells.

Cells underwent five light-switching steps in total: 405 nm → 530 nm → 405 nm → 530 nm → 405 nm. Error bars represent standard deviations from three independent experiments.

## 7.4 CONCLUSIONS

In summary, the three reversibly photoswitchable azobenzene amino acids **AzoF**, **F<sub>2</sub>AzoF**, and **F<sub>4</sub>AzoF** were site-specifically incorporated into proteins in both pro- and eukaryotic cells through genetic code expansion. The parent, unsubstituted **AzoF** requires UV irradiation for *trans* to *cis* photoswitching and quickly ( $t_{1/2}$  at 23 °C is 12.8 h) reverts back to the thermodynamically more stable *trans* isomer under physiological conditions, even when kept in the dark. This can be a useful property when biological processes on a similar timescale are being optically controlled; however, it may impact control of protein function at longer time scales and may complicate experimental designs and interpretation of experimental outcomes. In

contrast, the half-life of *cis*-F<sub>4</sub>AzoF ( $t_{1/2}$  at 25 °C is 2 years)<sup>222</sup> is dramatically increased, reducing thermal relaxation and showing complete stability for at least 24 h in proteins in live human cells. Due to the enhanced thermal stability of the *cis*-F<sub>4</sub>AzoF, it could be used to investigate long term biological events as well as for kinetics experiments in live animals. Moreover, the F<sub>4</sub>AzoF amino acid shows a red-shifted absorption spectrum enabling rapid photochemical *trans* to *cis* and *cis* to *trans* switching with visible light at 530 nm and 405 nm, respectively. This is the result of the electron-withdrawing effect of the fluorine substituents, which adjusts the energetics of the  $n \rightarrow \pi^*$  transition in F<sub>4</sub>AzoF, enabling photoswitching with visible light. The responsiveness of F<sub>4</sub>AzoF to the visible light avoids negative effects induced by UV light and makes it more suitable for *in vivo* applications. Moreover, F<sub>4</sub>AzoF displays excellent *trans* → *cis* isomerization (up to 91%). The two  $n \rightarrow \pi^*$  bands of *trans*- and *cis*-F<sub>4</sub>AzoF are well separated, enabling the selective isomerization of both isomers with enhanced photoconversions, compared to the overlap of the  $n \rightarrow \pi^*$  bands of the AzoF isomers (78% of *trans* → *cis* isomerization).

Site-specifically inserting azobenzene analogs into proteins in cells with an expanded genetic code enables the optical activation and deactivation of protein function in a reversible fashion using light illumination with two different wavelengths. However, selecting an amino acid residue for substitution is not trivial and usually involves extensive trial-and error mutagenesis. In order to place this process on a firmer, rational foundation, a computational protein modeling approach was developed. In order to narrow the set of potential installation sites on a given protein, we determined areas that provide a high degree of allosteric interaction and thus may have an impact on protein function when perturbed through azobenzene photoisomerization. Specifically, out of 7 predicted (and suitable) sites on firefly luciferase, they

were experimentally tested, and one (W417) provided reversible light-switching of enzymatic activity when used for azobenzene incorporation. FLuc activity can be controlled through the isomerization of azobenzene, incorporated at W417, upon illumination. The *cis* conformation of azobenzene reduces FLuc function, while the *trans* conformation fully restores FLuc activity, providing reversible, photoswitchable control over enzymatic activity. Reversible on and off switching of luciferase activity for up to five cycles was demonstrated. In addition to reversible temporal control of protein function, reversible spatial control of protein activation and deactivation was demonstrated as well. This approach of combining genetic code expansion with synthetic chromophore engineering and computational protein modeling adds a highly versatile system to the optogenetic toolbox for the light-regulation of protein function in cells and animals.

## 7.5 EXPERIMENTAL

### Cloning.

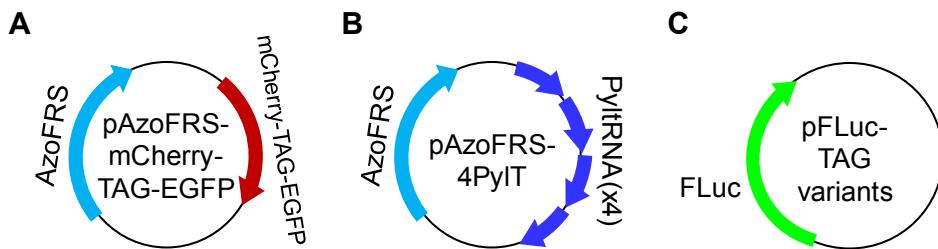
(1) Construction of pFLuc-Y340TAG, pFLuc-W417TAG, pFLuc-W426TAG, and pFLuc-F429TAG: These plasmids were obtained by introducing a TAG (amber stop codon) mutation into the wild-type FLuc (pGL3 plasmid) at the following residues: Y340TAG, W417TAG, W426TAG, and F429TAG, using the primer sets P1/P2, P3/P4, P5/P6, and P7/P8, respectively, and a QuikChange site-directed mutagenesis kit.

(2) Construction of pMbAzoFRS1: The pMbAzoFRS1 plasmid was obtained by making the following mutations to the wild-type *MbPylRS*<sup>48</sup>: Y271M\_L274A, N311A\_C313A,

and Y349F, using the primer sets P9/P10, P11/P12, and P13/P14, respectively, with the QuikChange site-directed mutagenesis method.

(3) Construction of pMbAzoFRS2: The pMbAzoFRS2 plasmid was obtained by making the following mutations to the wild-type MbPylRS: L270F, L274M, N311G\_C313G, and Y349F, using the primer sets P15/P16, P17/P18, P19/P20, and P13/P14, respectively, with the QuikChange site-directed mutagenesis method.

(4) Construction of pMbAzoFRS1-4PylT and pMbAzoFRS2-4PylT: Both plasmids were obtained by ligating the p4CMVE-U6-PylT<sup>48</sup> fragment from pMbPylT between the restriction sites *Nhe*I and *Mfe*I sites of pMbAzoFRS1 and pMbAzoFRS2, respectively.



**Figure 7.15:** Newly constructed plasmid maps for mammalian cell experiments.

**Table 7.4:** Primer list.

Primer No.	Sequence
P1	5'-AGAGAGGGCTAGCTCCAATTACTGACCGTACACCAA-3'
P2	5'-AGAGAGCTGAGATGCCATTTCCAGCAGGCGCAC-3'
P3	5'-CTTGATTGACAAGGATGGATAGCTACATTCTGGAGACATAG-3'
P4	5'-CTATGTCTCCAGAACATGTAGCTATCCATCCTGTCAATCAAG-3'
P5	5'-CTGGAGACATAGCTTACTAGGACGAAGACGAACACTTC-3'
P6	5'-GAAGTGTTCGTCTCGCCTAGTAAGCTATGTCTCCAG-3'
P7	5'-CATAGCTTACTGGGACGAATAGGAACACTTCTTCATCGTTG-3'
P8	5'-CAACGATGAAGAAGTGTTCCTATTGTCCCAGTAAGCTATG-3'
P9	5'-CTATGCTGGCCCCCACCTGTAACTACGCCGGAAACTGGACAGAAC-3'

P10	5'-GATTCTGTCCAGTTCCGGGCCTAGTCATCAGGGTGGGGGCCAGCATAG-3'
P11	5'-GAAGAGTTACAATGGTGGCCTTGCCCAGATGGCAGCGGCTG-3'
P12	5'-CAGCCGCTGCCCATCTGGCAAAGGCCACCATTGTAAACTCTTC-3'
P13	5'-GCGACAGCTGCATGGTGGCGACACCCCTGGACATC-3'
P14	5'-GATGTCCAGGGTGTGCCAACACCATGCAGCTGTCGC-3'
P15	5'-CCACCCCTGTACAACATGCCCGAAACTGGACAGAACATC-3'
P16	5'-GTTCCGCAGGTAGTTGAGAAGGTGGGGCCAGCATAG-3'
P17	5'-CCACCTTCTACAACATGCCGAAACTGGACAGAACATC-3'
P18	5'-GATTCTGTCCAGTTCCGCATGTAGTTGAGAAGGTGG-3'
P19	5'-GAAGAGTTACAATGGTGGCCTTGCCAGATGGCAGCGGCTG-3'
P20	5'-CAGCCGCTGCCCATCTGGCAAAGGCCACCATTGTAAACTCTTC-3'

### Incorporation of azobenzene analogs in mammalian cells.

HEK 293T cells were seeded at ~50,000 cells per well and grown in Dulbecco's Modified Eagle's Medium (200 µL, DMEM, Gibco) supplemented with FBS (Sigma, 10%), Pen-Strep (Corning Cellgro, 1%) and L-glutamine (Alfa Aesar, 2 mM) in 96-well plates (Greiner) under a humidified atmosphere containing CO<sub>2</sub> (5%) at 37 °C. Cells were transiently transfected with p*MbAzoFRS2-mCherry-TAG-EGFP-HA* and p4CMVE-U6-PyIT (100 ng of each plasmid) at ~80% confluence using linear polyethylenimine (LPEI; 1.5 µL, 0.323 mg/mL) in DMEM (200 µL) without any unnatural amino acid or with **AzoF**, **F<sub>2</sub>AzoF**, and **F<sub>4</sub>AzoF** (0.5 mM). After a 24 h incubation at 37 °C, the media were replaced with PBS and the cells were imaged with a Zeiss Axio Observer Z1 Microscope (10× objective) using EGFP (38HE: Ex 470/40; Em 525/50) and mCherry (43HE: Ex 550/25; Em 605/70) filter cubes. To further confirm the fidelity of incorporation and the expression of the fusion protein, Western blots were performed. HEK 293T cells were co-transfected with p*MbAzoFRS2-mCherry-TAG-EGFP-HA* and p4CMVE-U6-

PylT (1.5 µg of each plasmid) using LPEI (15 µL, 0.323 mg/mL) in the presence or in the absence of **AzoF**, **F<sub>2</sub>AzoF**, and **F<sub>4</sub>AzoF** (0.5 mM) in 6-well plates. After a 24 h incubation, the cells were washed with chilled PBS (2 mL) and lysed in mammalian protein extraction buffer (200 µL, GE Healthcare) with complete protease inhibitor cocktail (Sigma) through shaking at 4 °C for 15 min, and the cell lysates were cleared through centrifugation (14,000 g, 4 °C, 20 min). The protein lysates were boiled with loading buffer and analyzed by 10% SDS-PAGE (60 V for 20 min, and then 150 V for 1 h). After gel electrophoresis and transferring to a polyvinylidene difluoride (PVDF) membrane (80 V for 1.5 h, GE Healthcare), the membrane was blocked in TBS (5 mL) with Tween 20 (Fisher Scientific, 0.1%) and milk powder (5%) for 1 h. The membranes were probed and incubated with the primary α-HA antibody (Y-11) rabbit monoclonal IgG (sc-805, Santa Cruz Biotech) overnight at 4 °C, followed by a goat anti-rabbit IgG-HRP secondary antibody (sc-2031, Santa Cruz Biotech) for 1 h at room temperature. The membranes were washed with TBS with Tween 20 (0.1 %) three times, and then incubated in working solution (Thermo Scientific SuperSignal West Pico Chemiluminescent substrate) for 5 min. The Western blots were imaged using the “chemi” autofocus settings on a ChemiDoc (BioRad).

### **Photoswitching of firefly luciferase.**

HEK 293T cells (~50,000 cells per well) were cultured in DMEM (Gibco) supplemented with FBS (Sigma, 10%), Pen-Strep (Gibco, 1%) and L-glutamine (Alfa Aesar, 2 mM) in 96-well plates (Greiner) under a humidified atmosphere containing CO<sub>2</sub> (5%) at 37 °C. At ~80% confluence, cells seeded on plates were co-transfected with pMbAzoFRS2-4PylT and pFLuc-W417TAG (100 ng of each plasmid) using LPEI (1.5 µL, 0.323 mg/mL), and media was

changed to fresh DMEM supplemented without or with **AzoF** or **F<sub>4</sub>AzoF** (0.5 mM, 200 µL). After double transfection and incubation (24 h), the media was changed to DMEM without unnatural amino acids. Subsequently, cells were irradiated alternating between two light sources consecutively five times – alternating between 365 nm UV light for 5 min (Dual UV transilluminator) and visible white light for 5 min (JUST Normlicht) for **AzoF**, or between 530 nm light for 30 s per well (LUXEON) and 405 nm light 30 s per well (LEDENGIN) for **F<sub>4</sub>AzoF**. After light irradiation, cells were lysed by addition of 100 µL of substrate solution (Promega) in a 96-well plate (BD Falcon), and luminescence was measured using a Bright-Glo luciferase assay kit (Promega) and a microplate reader (Tecan M1000) with an integration time of 1 s.

### **Spatial control of firefly luciferase.**

HEK293T cells at ~80% confluence were co-transfected with pMbAzoFRS2-4PyLT and pFLuc-W417TAG (2 µg of each plasmid) using LPEI (15 µL, 0.323 mg/mL) in the presence of **AzoF** (0.25 mM) in 35-mm dishes (WillCo-Dish). After a 24 h incubation, the transfected cells were washed three times using PBS (pH 7.4) to remove excess **AzoF**. Plates were exposed to either UV (365 nm UV transilluminator, VWR) or white light (JUST Normlicht, RPimaging) irradiation for 5 min, while the other half of the plate was protected using aluminum foil, and then the plates were imaged using a ChemiDoc (Bio-Rad).

### **Protein expression in *E. coli*.**

The plasmid pBAD-sfGFP-Y151TAG-pyLT<sup>108</sup> was co-transformed with pBK-AzoFRS2 into chemically competent *E. coli* Top10 cells. A single colony was grown in LB media overnight and 400 µL of the overnight culture was added to 25 mL LB media supplemented with 1 mM of

**AzoF**, **F<sub>2</sub>AzoF**, and **F<sub>4</sub>AzoF** and 25 µg/mL tetracycline and 50 µg/mL kanamycin. Cells were grown at 37 °C, 250 rpm, and protein expression was induced with 0.1% arabinose when OD<sub>600</sub> reached ~0.5. After overnight expression at 37 °C, cells were harvested and washed with PBS. Cell pellets were re-suspended in 6 mL of phosphate lysis buffer (50 mM, pH 8.0), Triton X-100 (60 µL, 10%) and protease inhibitor (6 µL, Sigma), gently mixed, and incubated on ice for 1 h. Cell mixtures were sonicated (Fisher Scientific 550 Sonic Dismembrator) with six short bursts of 30 s on ice followed by 30 s intervals for cooling, and then the cell lysates were centrifuged at 4 °C and 13,000 g for 10 min. The supernatant was transferred to a 15 mL conical tube and 100 µL Ni-NTA resin (Qiagen) was added. The mixture was incubated at 4 °C for 2 h under mild shaking. The resin was then collected by centrifugation (1,000 g, 10 min) at 4 °C, washed three times with 300 µL of lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, pH 8.0), followed by two washes with 300 µL of wash buffer (lysis buffer containing 20 mM imidazole). The protein was eluted twice with 200 µL of elution buffer (lysis buffer containing 250 mM imidazole). The purified proteins were analyzed by 10% SDS-PAGE (60 V for 20 min, and then 150 V for 1 h), and stained with Coomassie Blue. The protein mass was obtained by electrospray ionization mass spectrometry (Figure 7.8).

### PylRS mutant panel generation.

**Table 7.5:** Panel of *MbPylRS* mutants screened for incorporation of azobenzene amino acids.

wt PylRS	L270	Y271	L274	N311	C313	Y349
AzoFRS1		M	A	A	A	F
AzoFRS2	F		M	G	G	F
PylRS3				A	A	
PylRS4				A	A	F
PylRS5		A		A	A	F

PylRS6		M	A	A	A	
PylRS7				G	G	F
PylRS8		M	A	G	G	
PylRS9	F		M	G	G	
PylRS10	F			G	G	F

### HPLC evaluation of photoswitching.

Photoisomerization was accomplished by illuminating separate solutions of **AzoF**, **F<sub>2</sub>AzoF**, and **F<sub>4</sub>AzoF** in acetonitrile (2.5 mM) with their corresponding light sources until their photostationary states were reached (~20 min). **AzoF** was exposed to UV (365 nm, UV transilluminator) and blue light (447.5 nm, LUXEON LED). **F<sub>2</sub>AzoF** and **F<sub>4</sub>AzoF** were exposed to green (530 nm, LUXEON) and blue (405 nm, LEDENGIN) light. To determine *E/Z* ratios, the illuminated solutions were analyzed by HPLC. Separation was carried out on a C18 column (xBridge, 5 µm particle size, 2.1 mm × 100 mm) using isocratic 70:30 acetonitrile:water on an Agilent 1220 HPLC system equipped with a diode array detector. *E/Z* ratios were determined by integrating corresponding peaks of the *E* and *Z* isomers ( $\lambda = 254$  nm).

### Absorbance spectra evaluation of photoswitching.

Solutions of **AzoF**, **F<sub>2</sub>AzoF**, and **F<sub>4</sub>AzoF** in DMSO/PBS (1 mM) were irradiated with UV (365 nm)/white (>400 nm), green (530 nm)/blue (405 nm), and green (530 nm)/blue (405 nm) light, respectively, until the photostationary state (PSS) was reached (~10 min). Absorbance spectra of *trans* and *cis* isomers of **AzoF**, **F<sub>2</sub>AzoF**, and **F<sub>4</sub>AzoF** were recorded using a Nanodrop spectrophotometer (Thermo Scientific).

## **8.0 GENETIC ENCODING OF PHOSPHINE-ACTIVATABLE AMINO ACIDS**

This chapter describes the genetic incorporation of phosphine-activatable amino acids into proteins in bacterial and mammalian cells. Genetic encoding of a set of azidobenzylloxycarbonyl amino acid using a pyrrolysyl tRNA synthetase/tRNA<sub>CUA</sub> pair in mammalian cells enables the site-specific introduction of a small-molecule-removable protecting group into the protein of interest. Strategic placement of this group renders the protein inactive until deprotection through a bioorthogonal Staudinger reduction delivers the active wild-type protein. This developed methodology was applied to the conditional control of several cellular processes, including bioluminescence, fluorescence, protein translocation, DNA recombination, post-translational modification, and gene editing (Cas9).

### **8.1 INTRODUCTION**

This material was reprinted, in part, with permission from [Luo, J.; Liu, Q.; Morihiro, K.; Deiters, A. \*Nat Chem\* 2016, 8, 1027-1034.](#)

Conditional control of protein function in live cells is essential for studies of the molecular details of biological processes. Several inhibitor-based loss-of-function strategies have been developed to reduce protein function and levels, for example, selective inhibition of kinase function using a bump-and-hole strategy,<sup>223</sup> selective induction of target protein degradation with

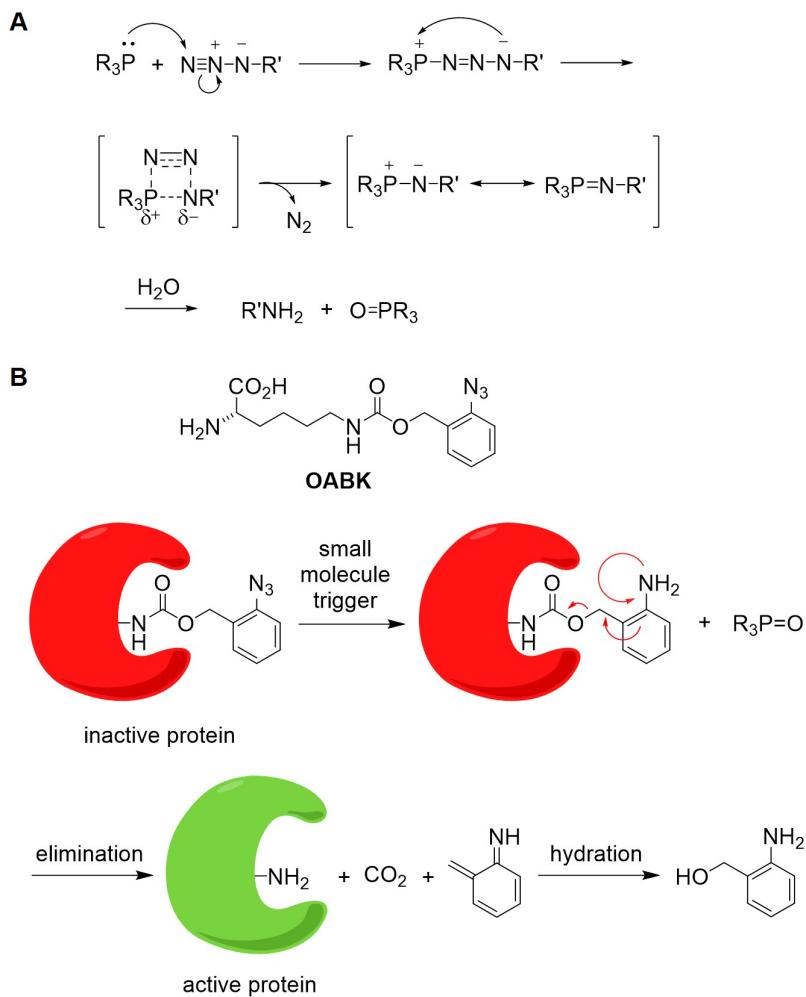
small molecule probes that recruit the ubiquitin proteasome system,<sup>224,225</sup> and selective small molecule-protein bioconjugations leading to the blocking of allosteric and active sites.<sup>226</sup> In contrast, the design and discovery of small molecule switches for activation of protein function has been more challenging,<sup>227-228</sup> with the rapamycin-mediated dimerization of FKBP12 and FRB domains being the most dominant example.<sup>9, 229-230</sup> This small molecule switch has been successfully applied to transcription factor domains, kinases, split enzyme systems such as TEV protease, cellular localization domains, and many others.<sup>231-234</sup> However, limitations include the need for the expression of two fusion proteins and the engineering of protein function to be controlled through FKBP12-FRB dimerization. Limited examples exist where single protein-rapamycin complex formations have been developed to control protein stability and kinase function; however, extensive protein engineering was required for the generation of these switches.<sup>235-236</sup> Other small molecule switches have been identified in nature and engineered for conditional control of protein function,<sup>237</sup> but examples of have been limited. In addition to small molecule activation of protein function, tools for optical activation of protein function have received significant interest in recent years.<sup>13, 238</sup> Small molecule control and light control complement each other and can be synergistically employed in a powerful fashion.<sup>239-240</sup>

We designed a small molecule switch through the inactivation of protein function via site-specific incorporation of an *ortho*-azidobenzylloxycarbonyl lysine (**OABK**, Figure 8.1) amino acid, followed by protein activation through bioorthogonal deprotection chemistry employing a Staudinger reduction. The Staudinger reduction-induced deprotection of amino acids adds an additional tool to the expanding set of chemical and photochemical peptide and protein activation approaches.<sup>141, 241-245</sup> Related reactions have been extensively used in bioconjugations *in vitro* and *in vivo*, and several types of “Staudinger ligations” have been

reported.<sup>246-248</sup> Azide groups installed on biological molecules are non-toxic and fully orthogonal to all cellular chemistries as demonstrated by numerous examples of bioconjugation reactions in cells and whole organisms.

## 8.2 GENETIC ENCODING OF *ORTHO*-AZIDOBENZYLOXYCARBONYL LYSINE

In the classic Staudinger reaction, phosphine reacts with the azide to generate a phosphazide intermediate. This intermediate undergoes intramolecular rearrangement via a four-membered ring transition state to lose N<sub>2</sub> and form an iminophosphorane intermediate. In the presence of water, it leads to the amine and stable phosphine oxide via hydrolysis (Figure 8.1A). Phosphine-induced Staudinger reduction of **OABK** yields an aniline derivative that undergoes a 1,4-elimination and decarboxylation, which was confirmed by LC/MS analysis (Figure 8.1B).<sup>249</sup> Also, efficient deprotection of **OABK** with 2-(diphenylphosphino)benzoic acid (2DPBA) was further confirmed in a time-course examination by Dr. Kunihiko Morihiro. This results in the formation of lysine and, when **OABK** is incorporated into a protein, the formation of active, wild-type protein.



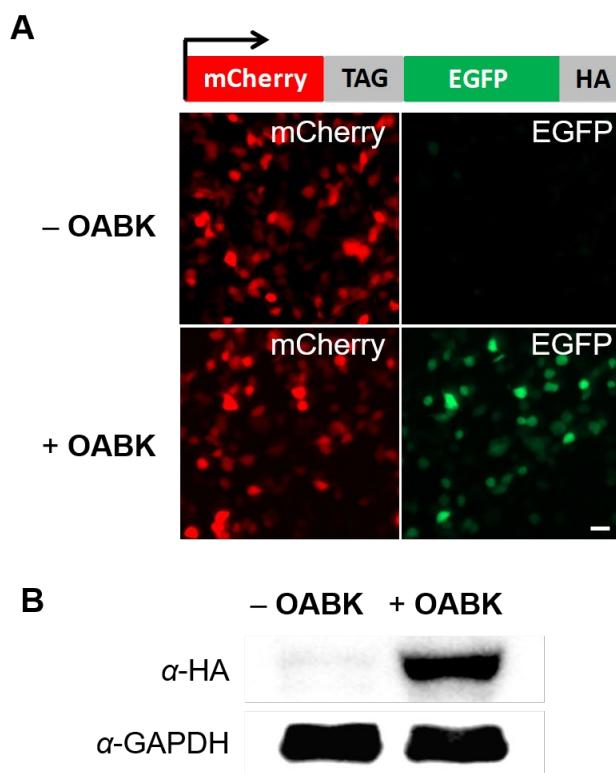
**Figure 8.1:** Schematic of the phosphine-triggered protein activation through protecting group removal via Staudinger reduction.

A) Schematic of mechanism of the Staudinger reduction between phosphines and azides. B) Schematic of the phosphine-triggered protein activation through protecting group removal via Staudinger reduction.

### 8.2.1 Incorporation of *ortho*-azidobenzyloxycarbonyl lysine (**OABK**) and deprotection

Based on a previously reported Y271A\_Y349F PylRS mutant (termed OABKRS),<sup>187</sup> incorporation of **OABK** into proteins was first demonstrated in mammalian cells (Figure 8.2A).

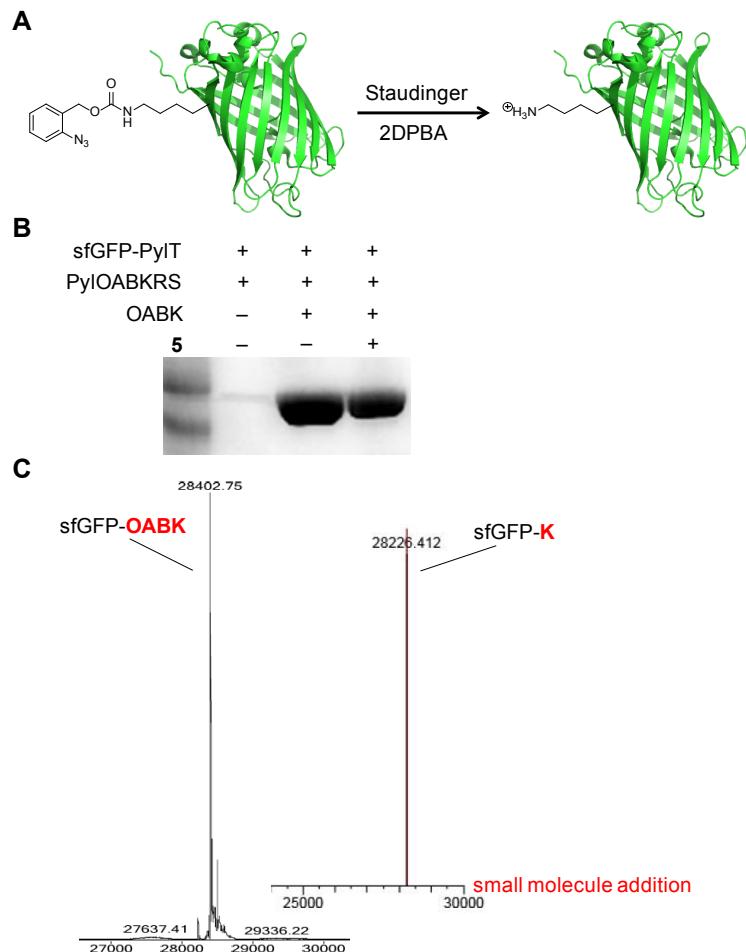
A plasmid encoding the synthetase and an amber-suppressor reporter (pOABKRS-mCherry-TAG-EGFP-HA), together with a plasmid encoding the tRNA (p4CMVE-U6-PylT), was transfected into HEK 293T cells. The cells were incubated for 24 h in the absence or presence of **OABK** (0.5 mM) and fluorescence imaging revealed EGFP expression only in the presence of the unnatural amino acid (Figure 8.2A). This was further confirmed by Western blot analysis with an anti-HA antibody (Figure 8.2B).



**Figure 8.2:** Genetic encoding of **OABK** in mammalian cells.

A) Micrographs confirming amino acid-dependent incorporation of **OABK** into mCherry-TAG-EGFP-HA in HEK293T cells. Scale bar represents 20  $\mu$ m. B) Confirmation of full-length protein expression through an anti-HA Western blot.

These results are in agreement with **OABK**-dependent protein expression in bacterial cells and confirm the high fidelity of the OABKRS in both pro- and eukaryotic systems (Figure 8.3). The efficient incorporation, a sfGFP-**OABK** protein yield of 12 mg/L was obtained in *E. coli*, was further confirmed by electrospray ionization mass spectrometry (ESI-MS), revealing an experimentally obtained mass of 28402.75 Da, in agreement with the expected mass of 28403.03 Da (Figure 8.3C). Furthermore, efficient *in vitro* protein deprotection through Staudinger reduction was confirmed by mass spec analysis of sfGFP-**OABK** after treatment with the small molecule trigger (the observed mass of 28226.41 Da for deprotected sfGFP matches the expected mass of 28227.89 Da; deprotection was quantitative and no protected sfGFP-**OABK** was detected (Figure 8.3C).

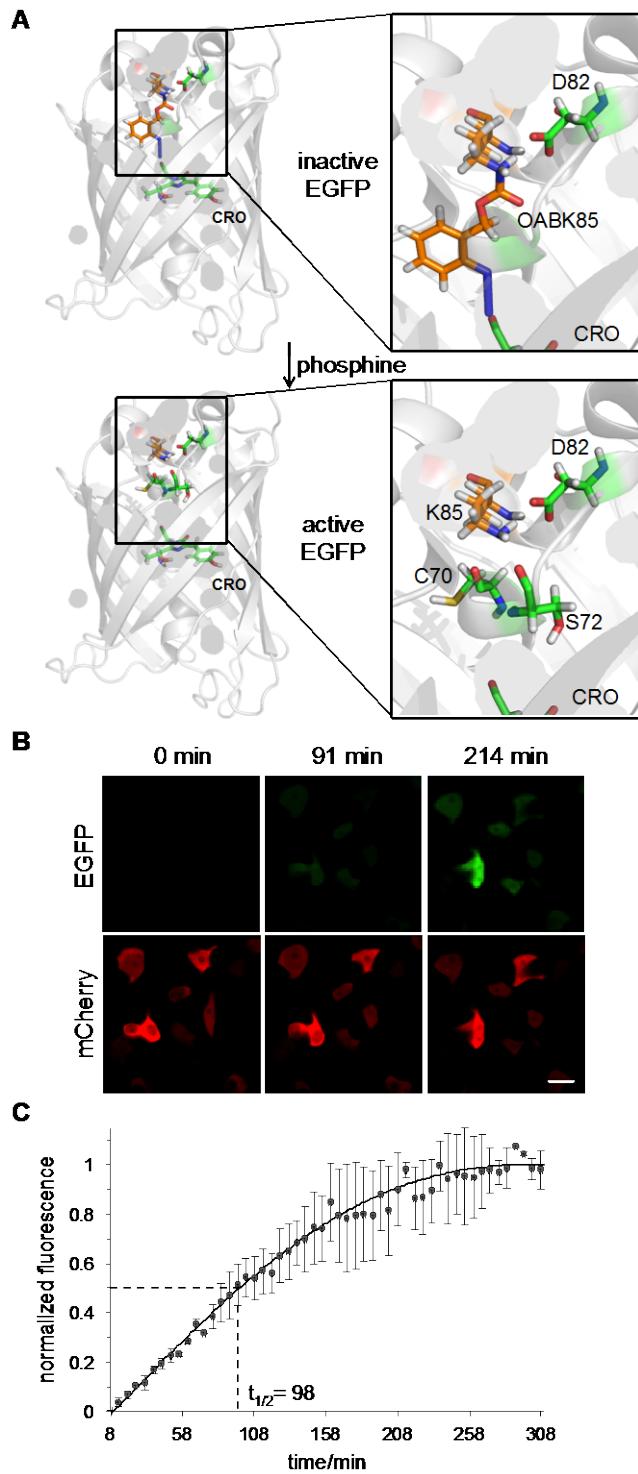


**Figure 8.3:** Staudinger reduction of sfGFP-OABK.

**A)** Staudinger reduction of sfGFP-OABK to sfGFP-K. **B)** SDS-PAGE confirmation of protein expression in *E. coli* and reaction with **5**. **C)** ESI-MS analysis of sfGFP-OABK and **5**-reduced sfGFP-OABK. sfGFP-OABK (left): observed MS: 28402.75 Da  $\pm$  0.28 Da, expected MS: 28403.03 Da; **5**-reduced sfGFP-OABK (right): observed MS: 28228.51 Da  $\pm$  0.11 Da, expected MS: 28228.62 Da. observed MS: 28226.41 Da  $\pm$  1.49 Da, expected MS: 28227.89 Da.

### 8.2.2 Small molecule-triggered fluorescent protein activation

Enhanced green fluorescent protein (EGFP) was selected as an initial target protein in order to test if this novel activation strategy would allow triggering of protein function in cells and since its activity can be measured in real time using live cell imaging.<sup>250</sup> Taking advantage of the small molecule-triggered deprotection feature of **OABK**, conditional control of EGFP maturation and fluorescence was performed (Figure 8.4A). HEK 293T cells were co-transfected with pEGFP-K85TAG-mCherry and pOABKRS-4PyIT in the presence of **OABK**. The mCherry serves as a transfection and expression control. After 24 h incubation, no EGFP fluorescence was observed in the absence of any small molecule trigger. However, addition of 2-(diphenylphosphino)benzoic acid to the cells led to EGFP activation (Figure 8.4B). Time-lapse imaging and quantification of cellular fluorescence revealed a  $t_{\frac{1}{2}}$  of 98 min for activation of EGFP (Figure 8.4C). Individual steps contributing to the kinetics of this activation include permeation of the phosphine into the cell, Staudinger reduction, 1,4-elimination of the lysine protecting group, EGFP folding, and fluorophore maturation through oxidation. Since individual rate constants of several of these steps are unknown, we decided to compare the overall kinetics to the optical activation (405 nm) of an EGFP-K85 mutant containing a caged coumarin-lysine (see Chapter 3). Here, in agreement with previously observed EGFP maturation kinetics,<sup>250-251</sup> a  $t_{\frac{1}{2}}$  of 35 min was observed, indicating that the fluorescent protein activation with the small molecule trigger is slightly slower than the light-activation as expected. Thus, genetically encoded **OABK** in conjunction with small molecule activation allows for the conditional regulation of intracellular protein maturation.



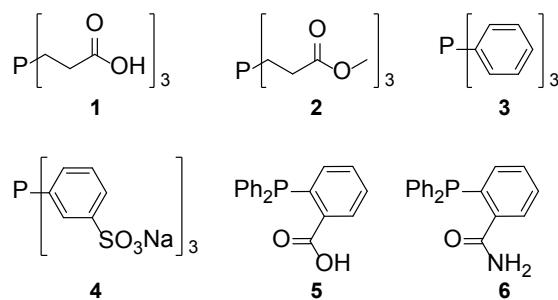
**Figure 8.4:** Small-molecule control of EGFP folding.

A) Incorporation of **OABK** (0.5 mM) at position K85 of EGFP inhibits fluorophore formation until the native lysine is generated through small molecule activation (model based on PDB 4EUL). B) Micrographs of HEK 293T cells

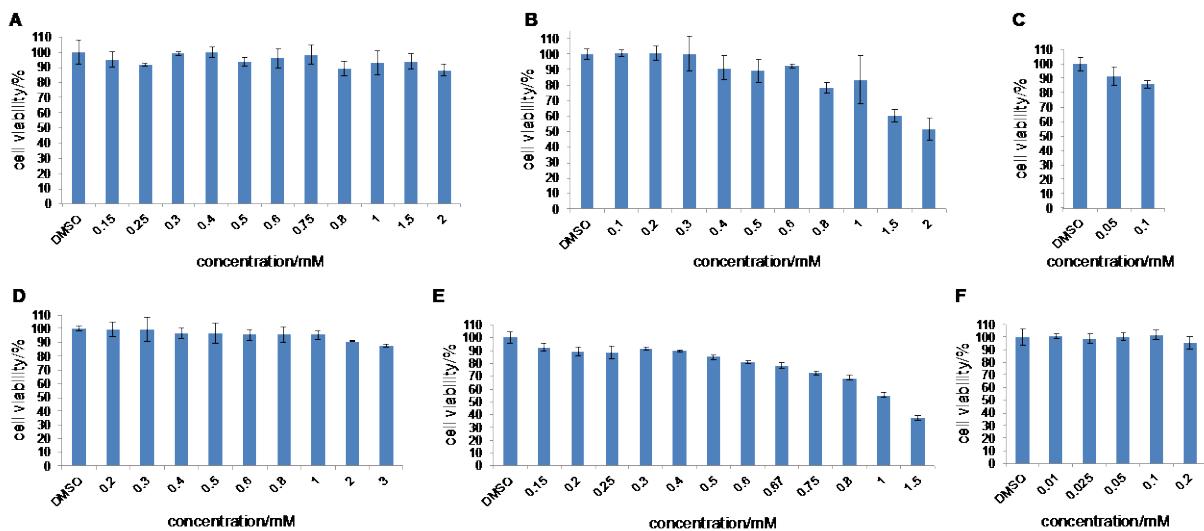
showing EGFP fluorescence only in the presence of 2-(diphenylphosphino)benzoic acid (500  $\mu$ M), but not in its absence. Scale bar represents 20  $\mu$ m. C) Time-lapse imaging and fluorescence quantification indicates a combined  $t_{1/2}$  of deprotection and EGFP maturation of 98 min. Error bars represent standard deviations from three cells.

### 8.2.3 Small molecule-triggered enzyme activation

Several different phosphine derivatives were selected, based on their availability, solubility, and use in Staudinger reductions (Figure 8.5). These include tris(2-carboxyethyl)phosphine (TCEP, **1**), the TCEP methylester **2**, triphenylphosphine **3**, phosphanetriyltris(benzenesulfonic acid) trisodium salt (TPPTS, **4**), 2-(diphenylphosphino)benzoic acid (2DPBA, **5**), and 2-(diphenylphosphino)benzamide (2DPBM, **6**). All phosphines were commercially available, besides **2** and **6** which were synthesized by Dr. Kunihiko Morihiro as previously reported.<sup>252-253</sup> At first, the solubility, cytotoxicity, and maximum concentration of phosphines were tested in living cells using XTT cell viability assays (Figure 8.6).



**Figure 8.5:** Structures of the six phosphine derivatives – TCEP (**1**), TCEP ester (**2**), TPP (**3**), TPPTS (**4**), 2DPBA (**5**), and 2DPBM (**6**).



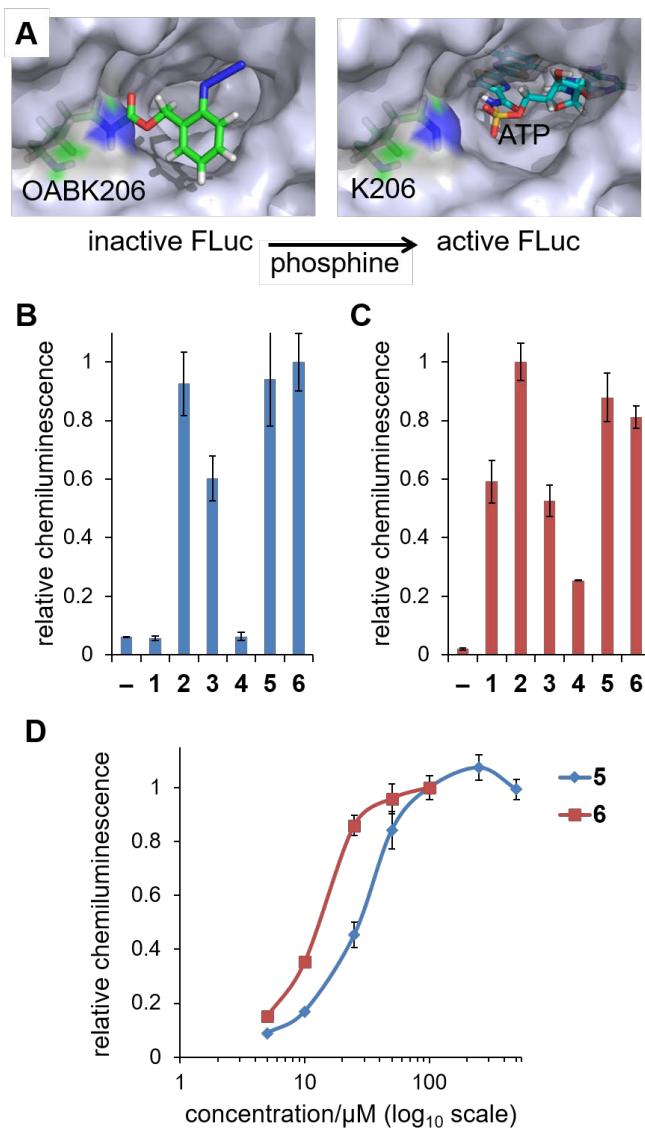
**Figure 8.6:** Cytotoxicity measurements of phosphine derivatives using XTT assays in HEK 293T cells.

**A)** tris(2-carboxyethyl)phosphine (TCEP, **1**), **B)** TCEP methylester (**2**), **C)** triphenylphosphine (TPP, **3**), **D)** phosphanetriyltris(benzenesulfonic acid) trisodium salt (TPPTS, **4**), **E)** 2-(diphenylphosphino) benzoic acid (2DPBA, **5**), **F)** 2-(diphenylphosphino)benzamide (2DPBM, **6**). Error bars represent standard deviations from three independent experiments.

To demonstrate broader applicability of this approach after these initial encouraging results and to obtain an easily quantified readout that can be used to screen a variety of different phosphines for their ability to activate protein function in cells and in cell lysates, the firefly luciferase (FLuc) was chosen as a second target protein. FLuc contains a catalytic site lysine, K206, which has previously shown to block catalytic activity when substituted with a sterically demanding hydroxycoumarin lysine (see Chapter 3).<sup>250</sup> Similarly, introducing an **OABK** at K206 inhibits FLuc enzymatic activity by restricting access of ATP to the active site, until the enzyme was deprotected and activated through phosphine treatment (Figure 8.7A).

To screen them, HEK293T cells were co-transfected with the mutated firefly luciferase plasmid (pGL3-K206TAG) and the *MbOABKRS*/PyLtRNA<sub>CUA</sub> pair (pOABKRS-4PyLT) in the

presence of **OABK** (0.5 mM). After a 24 h incubation, the cells were either treated with one of the six phosphine derivatives **1-6** or were continued to be incubated in the absence of the small molecule trigger. The incorporation of **OABK** into FLuc blocked luciferase activity in the absence of small molecule activation, as determined by a Bright-Glo luciferase assay (Figure 8.7B). However, after only a 3 h incubation with the phosphines **2** (500  $\mu$ M), **3** (50  $\mu$ M solubility limit in media), and **5** (500  $\mu$ M) firefly luciferase activity was increased by 43-fold, 11-fold, and 32-fold, respectively. This demonstrates, for the first time, that a small molecule protein switch based on a Staudinger reduction can be designed. Optimal OFF to ON switching was achieved with **2**, **5**, and **6**, while at a reduced concentration of the small molecule trigger ( $\leq$ 50  $\mu$ M) the phosphines **5** and **6** showed most efficient activation (Figure 8.7B and Figure 8.7C), and no activation was observed in **2** (data not shown). Not unexpected, the reactions with **1** and **4** did not show luciferase activation in live cells due to their diminished cell permeability (Figure 8.7B).<sup>254</sup> The pronounced differential activity of TCEP (**1**) in lysate versus cells (Figure 8.7B and Figure 8.7C) may provide an opportunity to selectively activate surface displayed or secreted proteins while not affecting intracellular proteins. A dose-dependent luciferase assay was conducted with **5** (5-500  $\mu$ M) and **6** (5-100  $\mu$ M) in live cells, indicating a maximal activation plateau starting at concentrations as low as 50  $\mu$ M (**5**) and 25  $\mu$ M (**6**) (Figure 8.7D). Thus, among the six phosphines, **5** and **6** were used for all subsequent experiments, including a time-course of cellular luciferase function (Figure 8.7E). The chemiluminescence intensity gradually increased over time, reaching a plateau at 2.5 h and a  $t_{1/2}$  of 55 min for activation of FLuc.



**Figure 8.7:** Small-molecule activation of firefly luciferase.

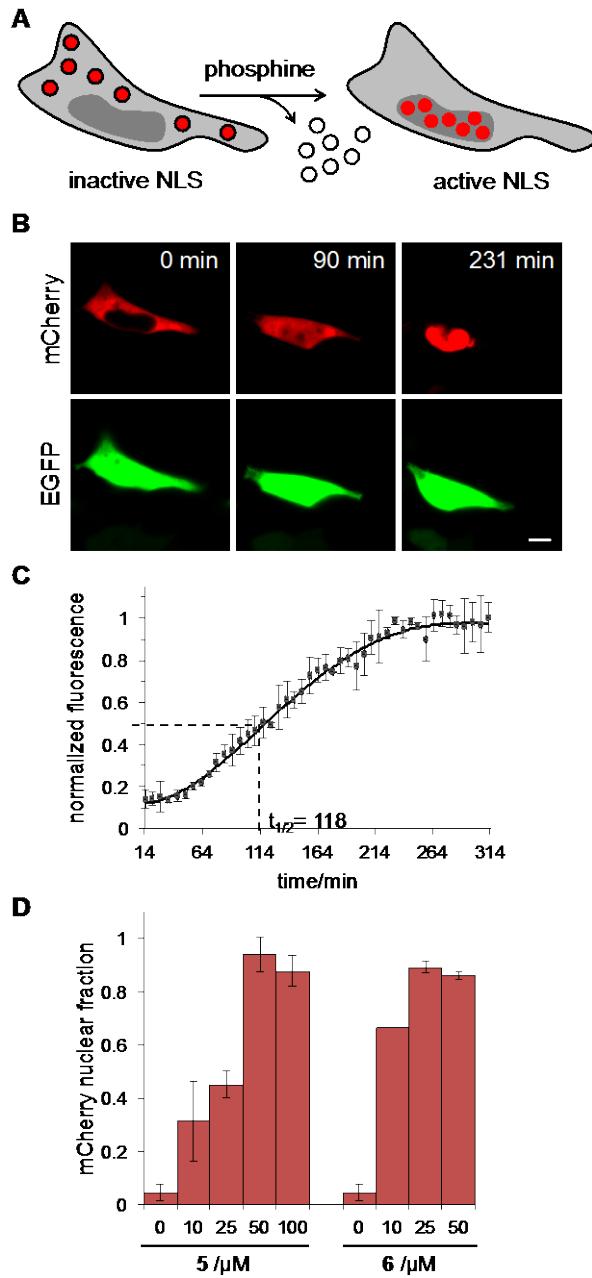
A) Model of FLuc containing **OABK** at position K206, thereby blocking access of ATP to the catalytic site until the azidobenzyloxycarbonyl group is cleaved by the small molecule trigger (based on PDB 2D1S). B) Small molecule activation of **OABK-FLuc** triggered by **1-6** (using maximum concentrations based on the solubility and cell viability, Figure 8.6) in HEK293T cells and in C) cell lysates. Error bars represent standard deviations from three independent experiments. E) In-cell dose-dependent luciferase assays of **5** (5, 10, 25, 50, 100, 250, 500  $\mu\text{M}$ ) and **6** (5, 10, 25, 50, 100  $\mu\text{M}$ ). The solubility of **6** in an aqueous environment limits its application to  $\leq 100 \mu\text{M}$ .

Thus, site-specific incorporation of **OABK** provides a novel approach to the small molecule activation of enzymatic function in live cells, complementing other protecting group strategies to conditionally control cellular processes, e.g., light-removable “caging groups”.<sup>14, 192</sup> Small molecule-activation of protein function is a complementary approach to light-activation in that it can be employed in cases where optical stimulation is not feasible, such as opaque, thick, or structurally complex specimens (most metazoans), systems that are not amenable to stimulation through irradiation, e.g., due to conflicting fluorescent reporters or UV toxicity. Moreover, bulk activation of large numbers of cells can be readily achieved, and, intriguingly, sequential activation of cellular processes is possible by using both light and small molecules as triggers, such as the engineering of dual-input mammalian genetic circuits that were applied to extremely stringent gene expression *in vivo*.<sup>239</sup> Thus, this Staudinger-reduction induced protein activation expands the toolbox for conditionally controlled protein activation in live cells. It compares favorably to palladium-mediated<sup>243</sup> and Diels-Alder reaction-mediated<sup>241</sup> protein deprotection methods, since benign phosphine-based reagents have been extensively used in biological studies, since low concentrations (as low as 25 μM) of the small molecule trigger were found to efficiently activate protein function, and since synthesis of **OABK** and its incorporation into proteins in both bacterial and mammalian systems is highly efficient. In addition, the protein triggering through Staudinger reduction is expected to be compatible and orthogonal to cycloaddition-based deprotection reactions.

#### 8.2.4 Small molecule-triggered protein translocation

In addition to the small molecule triggering of enzymatic function (FLuc) and protein maturation (EGFP), we wanted to demonstrate the small molecule control of a protein that is the substrate

for another protein. Thus, we investigated small molecule-activated nuclear translocation by genetically incorporating **OABK** into the nuclear localization signal (NLS) of the transcription factor SATB1 (Figure 8.8A).<sup>142</sup> The presence of the azidobenzyl group at position K29 completely inhibited translocation of an mCherry reporter as an NLS fusion, as observed in HEK 293T cells expressing SATB1K29**OABK**-mCherry. After 24 hr of incubation, the cells were treated with **5** (500 μM) and nuclear translocation of mCherry was triggered through Staudinger reduction-induced formation of the native SATB1-NLS (Figure 8.8B). Translocation was completed within 231 min, as shown by time-lapse imaging, and fluorescence quantification over time revealed a  $t_{1/2}$  of 118 min (Figure 8.8C). In comparison to the optical triggering of a caged SATB1-NLS ( $t_{1/2}$  of 54 min) (see Chapter 3), the small molecule activated process within a similar timeframe (approximately two vs. one hour). Importantly, complete protein translocation from the cytoplasm to the nucleus was observed, demonstrating complete **OABK**-SATB1 deprotection in live cells through small molecule treatment. Furthermore, a dose-dependent assays showed that concentrations as low as 50 μM (**5**) and 25 μM (**6**) were sufficient for fully deprotecting and activating the nuclear translocation signal, as all cytoplasmic mCherry quantitatively shuttled to the nucleus. The dose-response also indicated a correlation between the switch-on ratio and the phosphine concentration in the intracellular environment, allowing a precise control over the extend of a cellular action (Figure 8.8D).



**Figure 8.8:** Small-molecule activation of SatB1 translocation.

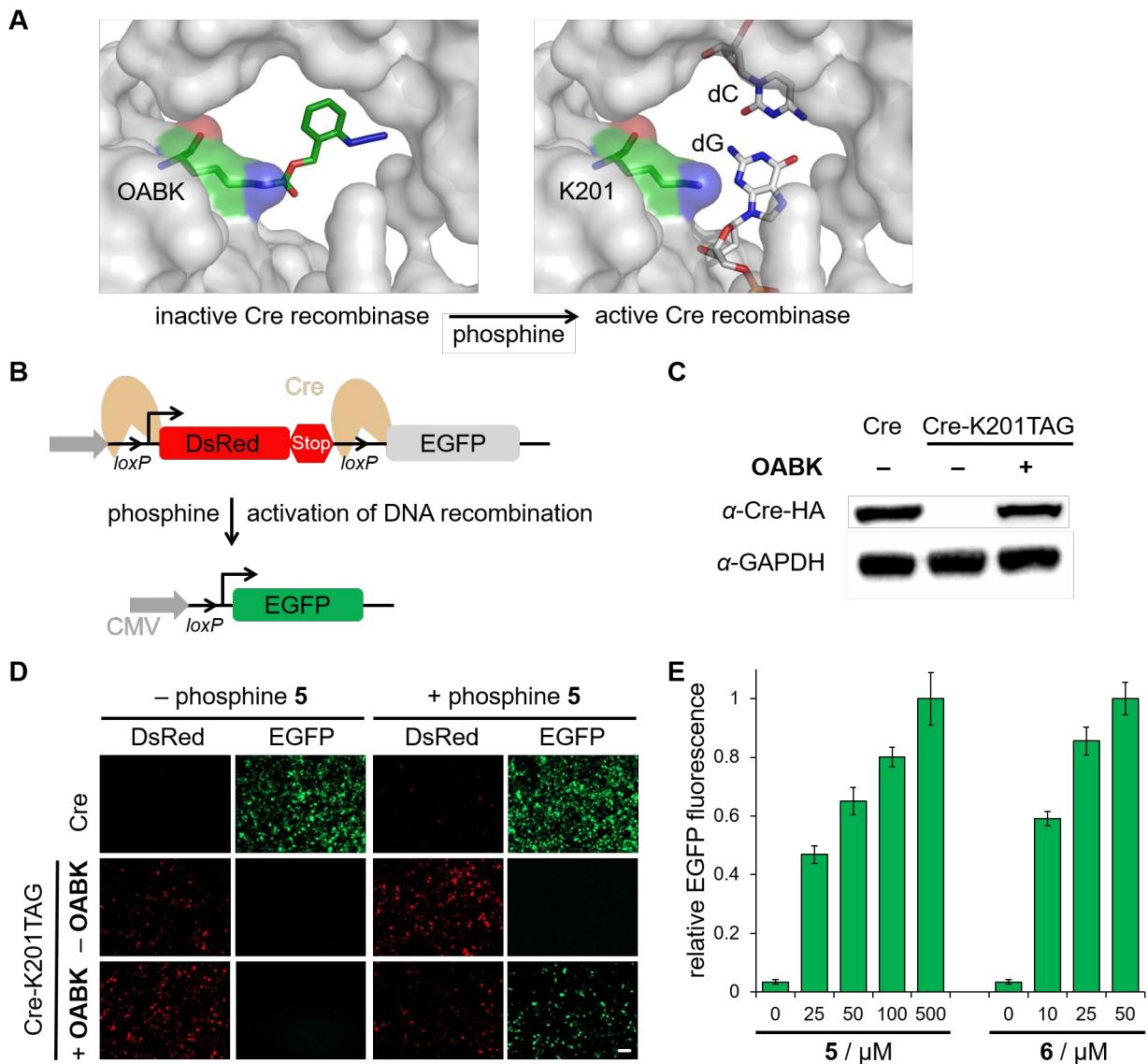
- A) Site-specific incorporation of **OABK** at K29 in the SATB1 nuclear localization sequence leads to complete exclusion of mCherry from the nucleus, until the wild-type NLS is generated through treatment with the phosphine **5**.
- B) Micrographs of HEK293T cells showing no nuclear localization of **OABK-SATB1-mCherry** and complete activation of translocation from the cytoplasm to the nucleus after activation with **5**. EGFP expression was used as a transfection and imaging control. Scale bar represents 10  $\mu\text{m}$ . C) A time-course of the normalized mCherry fluorescence in the nucleus indicates a combined  $t_{1/2}$  of deprotection and protein translocation of 118 min. Error bars

represent standard deviations from three cells. D) Dose-dependent mCherry nuclear import through activation of SatB1 translocation with different concentrations of **5** and **6**.

### 8.2.5 Small molecule-triggered DNA recombination

To further explore the general applicability of the developed small molecule switch of protein function, we investigated its function in two additional proteins that act upon DNA as their substrate. *Bacteriophage PI* Cre recombinase recognizes *loxP* sites on double-stranded DNA, induces phosphodiester cleavage, and either inserts or excises sequences through a double recombination event that proceeds via a holiday junction.<sup>255</sup> The Cre-*loxP* recombination system has been developed as a versatile genome manipulation tool and has been extensively applied in gene knock-out and knock-in studies.<sup>54</sup> To apply our phosphine-triggered switch, we incorporated **OABK** at position K201. A lysine in position 201 is essential for catalytic activity of Cre recombinase, because it is located in a loop extending into the DNA minor groove near the cleavage site of the DNA, and has interactions with the N3 of the +1 guanine base and the 5'-O and 4'-O of the -1 sugar, as well as a tightly bound water molecule within hydrogen bonding distance (Figure 8.9A).<sup>256</sup> The expression level of Cre-K201**OABK** is comparable to wild-type Cre recombinase expression in HEK293T cells (Figure 8.9C). Activation of Cre-**OABK** with **5** in live cells was explored using a Cre stoplight reporter assay (Figure 8.9B and Figure 8.9C).<sup>257</sup> Importantly, cells expressing Cre-**OABK** exclusively showed DsRed expression in the absence of the small molecule trigger **5**, demonstrating the complete catalytic inactivity of the recombinase enzyme containing the K201→**OABK** mutation. Through **5**-induced deprotection, the protein was activated very effectively and the functional Cre catalyzed DNA recombination

to excise DsRed and a transcriptional terminator, switching on EGFP expression (Figure 8.9B and Figure 8.9D). To identify the optimal phosphine activation conditions, the Cre recombination event was triggered with varying concentrations of **5** and **6**. Overall, activation of DNA recombination can be efficiently achieved at high levels through triggering with 25  $\mu$ M of **6** (Figure 8.9E). This is in full agreement with the previous luciferase activation and the activation of nuclear translocation, which indicated virtually complete intracellular protein deprotection at 25  $\mu$ M of **6**. These results demonstrate that conditional activation of DNA recombination was successfully accomplished through a small molecule-mediated Staudinger reduction.



**Figure 8.9:** Small molecule-triggered DNA recombination.

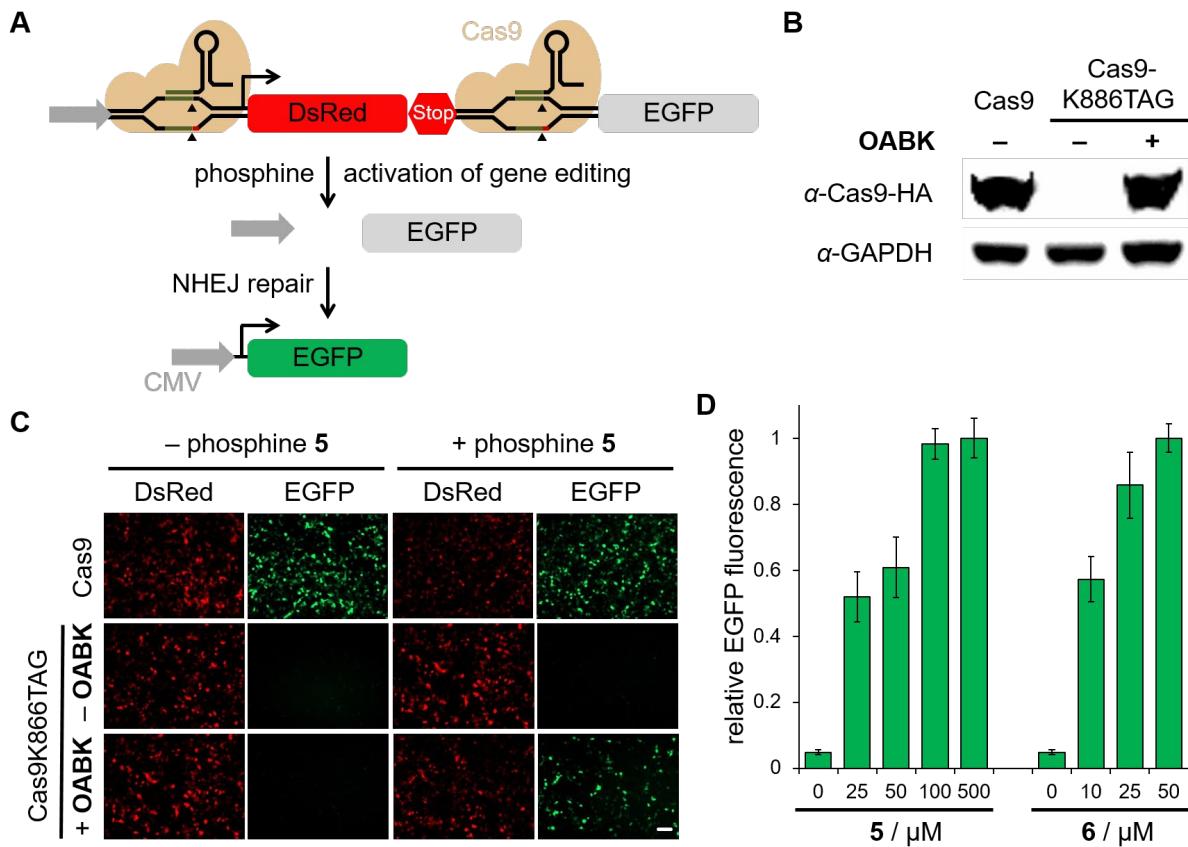
A) Model of the regulation of Cre recombinase activity through the K201→**OABK** mutation based on PDB 1CRX. Before addition of the phosphine trigger, the protecting group blocks the active site lysine from interacting with the DNA. Upon activation with phosphines, the caging group is removed and Cre recombination is restored. B) Schematic design of the Cre recombinase activity assay using a Stoplight dual reporter. In the absence of activated Cre, DsRed is exclusively expressed under control of the CMV promoter. When Cre-**OABK** is activated by the phosphine **5** or **6**, functional Cre excises the DsRed-terminator cassette between the *loxP* sites, and EGFP is placed under control of the CMV promoter and is expressed. C) Anti-HA-tag Western blot showing amino acid-dependent expression of modified Cre recombinase in HEK 293T cells in the absence or presence of **OABK** (0.25 mM), as

well as a GAPDH loading control. D) EGFP fluorescence imaging shows small molecule activation of Cre recombinase-catalyzed DNA recombination in HEK 293T cells using the dual fluorescence reporter. Scale bar represents 100  $\mu$ m. E) Analysis of EGFP expression using imaging cytometry reveals dose-dependent activation of Cre recombinase with increasing concentrations of **5** and **6**. The number of both DsRed and EGFP fluorescent cells were counted in four randomly selected fields of view per well and normalized to the total number of fluorescent cells. Data was analyzed using ImageJ. Error bars represent standard deviations from three independent experiments.

### 8.2.6 Small molecule-triggered CRISPR/Cas9 gene editing

The second DNA-editing enzyme we applied the engineered small molecule switch to was Cas9. The CRISPR/Cas9 system has been established as an extremely valuable tool for gene editing in cells and model organisms.<sup>258</sup> Most recently, conditional regulation of Cas9 enzymatic activity has been shown to provide spatial and temporal control over gene editing and to mitigate off-target effects observed with constitutively active Cas9.<sup>259-260</sup> To implement the Staudinger reduction switch, we incorporated **OABK** at position K866, a lysine residue that undergoes a large conformational change upon gRNA binding, thereby becoming exposed to the surface and possibly facilitating orientation of the DNA for nuclease cleavage.<sup>261</sup> This site has been successfully used for the optical control of Cas9 function.<sup>262</sup> The function of Cas9-**OABK** in the absence and presence of **5** was investigated using a DsRed/EGFP dual reporter assay (Figure 8.10A; encoded on pIRG). Cells expressing Cas9-**OABK** only showed DsRed fluorescence, because transcription terminates after DsRed in the absence of activated Cas9, indicating complete inhibition of enzymatic function. Once Cas9-**OABK** is activated by the small molecule phosphine, the complex of functional Cas9 and gRNAs induces excision of the DsRed-terminator cassette through DNA double-strand break, followed by NHEJ repair, and leading to EGFP expression. The expression level of Cas9-K866**OABK** is similar to wild-type Cas9

expression in mammalian cells (Figure 8.10B), and the activity after small molecule triggering almost reached native levels (Figure 8.10C). To optimize the concentration of phosphines for activation of gene editing in live cells, varying concentrations of **5** and **6** were investigated. Consistent with the other small molecule-triggered processes, dose-dependent results were obtained, and 25  $\mu$ M was identified as the universally effective concentration of **6** for activation of Cas9-**OABK** and other proteins in live cells (Figure 8.10D). These experiments successfully demonstrate that conditional activation of CRISPR/Cas9 gene editing can be accomplished via a Staudinger reduction.



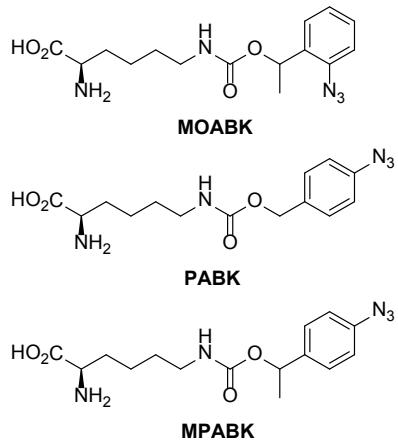
**Figure 8.10:** Small molecule-triggered CRISPR/Cas9 gene editing.

A) Schematic design of the CRISPR/Cas9 activity assay using the pIRG dual reporter. In the absence of activated Cas9, transcription terminates after the DsRed gene and no EGFP is expressed. When Cas9-**OABK** is activated by the phosphine **5** or **6**, the complex of functional Cas9 and gRNAs induces excision of the DsRed-terminator cassette, followed by DNA repair and EGFP expression. B) Anti-HA western blot showing amino acid-dependent expression of modified Cas9 in HEK 293T cells without or with **OABK** (0.25 mM), as well as a GAPDH loading control. C) Small-molecule control of CRISPR/Cas9 gene editing in HEK 293T cells using the pIRG reporter shows that EGFP fluorescence is only observed only in the presence of **5**. Residual DsRed fluorescence was still observed, due to reporter expression during the 24 h incubation before small molecule activation. Scale bar represents 100  $\mu$ m. D) EGFP expression, and thus Cas9 function, can be titrated in response to increasing concentrations of **5** and **6**. EGFP expression was analyzed using imaging cytometry by counting fluorescent cells in four randomly selected fields of view per well. Data was analyzed using ImageJ. Error bars represent standard deviations from three independent experiments.

## 8.3 GENETIC ENCODING OF AZIDOBENZYLOXYCARBONYL LYSINE ANALOGS

### 8.3.1 Incorporation of azidobenzyloxycarbonyl lysine analogs in *E. coli*

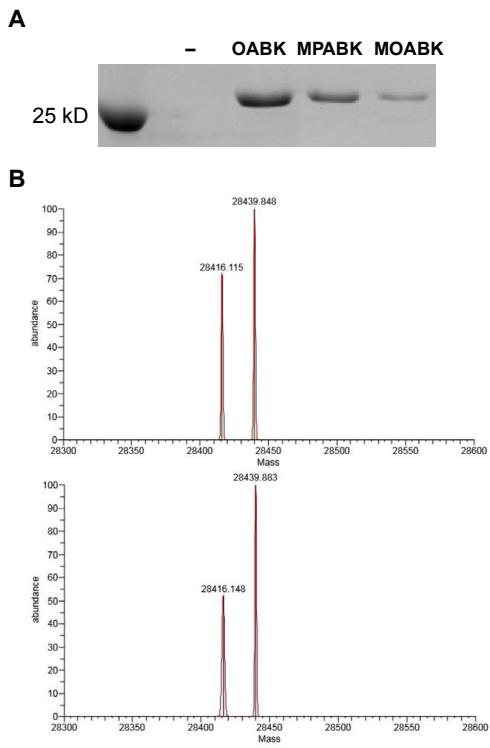
To optimize the kinetics of the small molecule decaging strategy, the UAA substrates were systematically surveyed bearing the azido group in *ortho*- and *para*-position for triggering the elimination process after the initial Staudinger reduction with 2DPBM. The photocaging approach in Chapter 6 guided us to design and test a series of azidobenzyl lysine analogs, which show significantly enhanced decaging rates and allow the instantaneous activation of protein in live cells. Herein, **PABK**, **MeOABK**, and **MePABK** (Figure 8.11) were synthesized by Dr. Kunihiko Morihiro, and **MePABK** was re-synthesized by Joshua Wesalo to improve the solubility.



**Figure 8.11:** Structures of genetically encoded azido amino acids – **PABK**, **MeOABK** and **MePABK**.

These amino acids were synthesized by Dr. Kunihiko Morihiro in the Deiters lab.

Given the structural similarity of azidobenzyloxycarbonyl lysine analogs to **OABK**, the engineered *MbOABKRS* (with the two mutations, Y271A and Y349F) was first used to test the amino acid-dependent expression of sfGFP in the presence of **MPABK** and **MOABK** in *E. coli*. ESI-MS data for sfGFP produced from psfGFP\_Y151TAG-PyIT-6xHis in the presence of **MPABK** and **MOABK** were consistent with their site-specific incorporation (Figure 8.12A), while mass spec data for **PABK** still needs to be obtained. Comparable to sfGFP-**OABK**, the expression of sfGFP-**MPABK** and sfGFP-**MOABK** displayed a protein yields of 5.6 mg/L and 1.4 mg/L, respectively. The incorporation was further confirmed by ESI-MS analysis of purified proteins see (Figure 8.12B).



**Figure 8.12:** Genetic incorporation of **MOABK** and **MPABK** in *E. coli*.

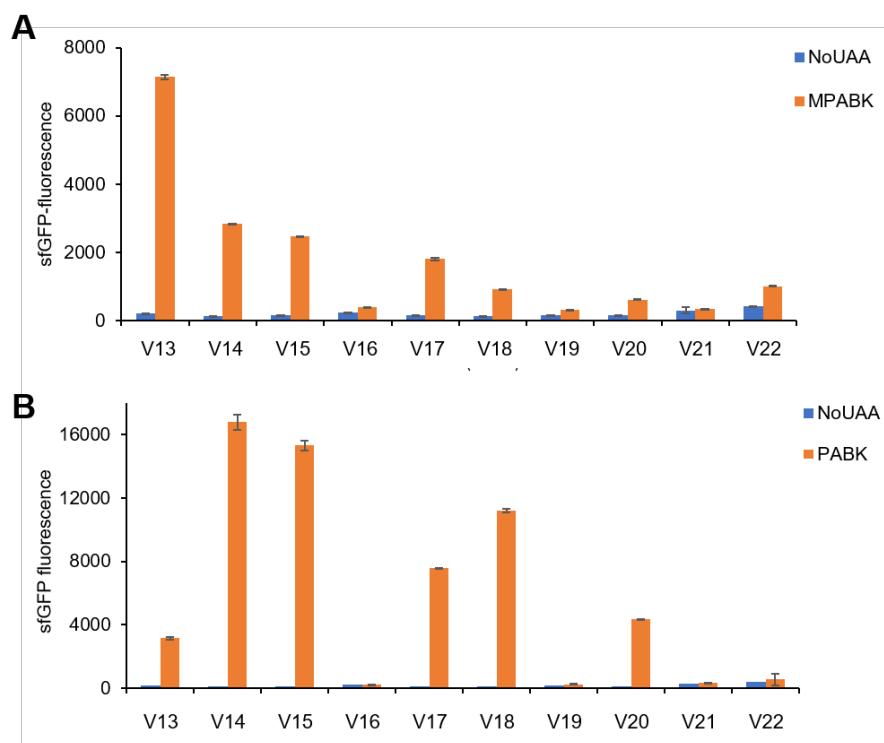
A) Coommassie stained gel of Ni-NTA purified sfGFP-Y151TAG expression in the absence and presence of **MOABK** and **MPABK**. B) ESI-MS analysis of sfGFP-**MPABK** and sfGFP-**MOABK**.

In order to obtain an optimal synthetase for enhanced incorporation efficiency, the *Mb*PylRS/tRNA<sub>CUA</sub> pair along with a panel of *Mb*PylRS mutants were screened using the sfGFP-Y151TAG reporter in *E. coli* to investigate the incorporation of **MPABK**, **PABK**, and **MOABK** (Figure 8.13). We first screened the PylRS panel for the incorporation of **MPABK** and **PABK** into sfGFP on a small-scale (200 µL) protein expression to conserve UAAs, and directly used overnight cell culture to measure sfGFP fluorescence ( $\lambda_{\text{ex}} = 488$  nm,  $\lambda_{\text{em}} = 510$  nm) and cell growth ( $\text{OD}_{600}$ ,  $\lambda_{\text{ab}} = 600$  nm) on a Tecan M1000 plate reader. The readout of the fluorescence was normalized to the density of the corresponding cell culture. The hit synthetase was discovered if the cells harboring corresponding synthetase showed sfGFP fluorescence in the

presence of **MPABK** or **PABK**, but no fluorescence detected without the UAA. Based on the structural similarity to **OABK** and the high promiscuity of PylRS,<sup>49</sup> Y271A\_Y349F (V13, OABKRS) was used as a starting PylRS and a positive control for small-scale expression. OABKRS (V13) shows a wide substrate specificity, including not only **OABK**, but also norbornenelysine,<sup>50</sup> *N* $\varepsilon$ -[2-(furan-2-yl)ethoxy]carbonyl lysine,<sup>51</sup> cyclooctynyl lysine,<sup>52</sup> diazirinylbenzyloxycarbonyl lysine,<sup>53</sup> etc. We chose to focus on four PylRS mutants (Y271, L274, C313, Y349) that may recognize and encode **MPABK** and **PABK**. The mutant L274A is present in most PylRS mutants that are evolved for lysine analogues, which might enlarge the cavity for the UAA substrate binding.<sup>49</sup> The Y349F mutation has shown to increase aminoacylation efficiency.<sup>31</sup> Also, mutations at Y271 and C313 were frequently observed for incorporation of bulky pyrrolysine derivatives in other PylRS library selections.<sup>50, 52</sup> Among ten synthetases tested in a cell-based fluorescent method allowing direct measurement of sfGFP fluorescence ( $\lambda_{\text{ex}} = 488$  nm,  $\lambda_{\text{em}} = 510$  nm),<sup>54-55</sup> V13 (OABKRS with Y271A and Y349F mutations) showed highest incorporation efficiency for **MPABK** (Figure 8.13A), while V14 (termed as PABKRS with L274A, C313A, and Y349F mutations) showed highest incorporation efficiency for **PABK** (Figure 8.13B). Therefore, we used OABKRS (V13) for incorporation of **MPABK** and PABKRS (V14) for incorporation of **PABK** in our future experiments.

**Table 8.1:** *MbPylRS* panel screening for incorporation of **MPABK** and **PABK**.

Code	Mutations			
	Y 271	L 274	C 313	Y 349
V13	A			F
V14		A	A	F
V15		A	S	F
V16			G	F
V17		T		F
V18		T	G	F
V19	A	A		F
V20	A	A	A	F
V21	A		G	F
V22	A		A	F

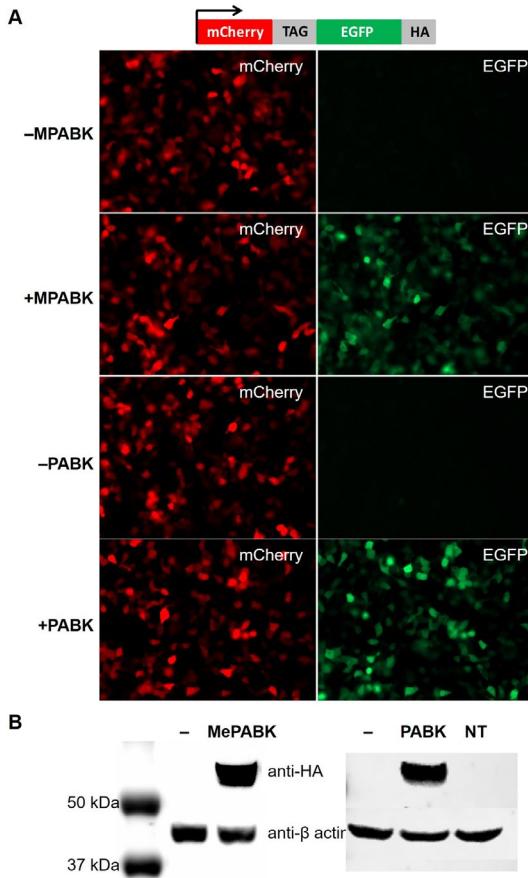


**Figure 8.13:** *MbPylRS* panel screening for incorporation of **MPABK** (A) and **PABK** (B) into sfGFP in *E. coli*.

sfGFP fluorescence = [fluorescence intensity]/[OD<sub>600</sub>×10]. Error bars represent standard deviation from three independent experiments.

### 8.3.2 Incorporation of the azidobenzyloxycarbonyl lysine analogs in mammalian cells

Based on a previously reported OABKRS, the first incorporation of **MPABK** into proteins was investigated in mammalian cells (Figure 8.14). A plasmid that encodes an amber-suppressor reporter (mCherry-TAG-EGFP-HA), together with a plasmid that encodes the synthetase and tRNA (pOABKRS-4CMVE-U6-PylT), was transfected into human embryonic kidney (HEK) 293T cells. The cells were incubated for 24 hrs with or without **MPABK** (0.5 mM) and fluorescence imaging revealed enhanced green fluorescent protein (EGFP) expression only in the presence of UAA. This was further confirmed by western blot analysis with an anti-hemagglutinin (HA) antibody (Figure 8.14). In the same method, incorporation of **PABK** into proteins using the newly-engineered synthetase (PABKRS) was demonstrated in mammalian cells through fluorescence imaging and Western blots (Figure 8.14).



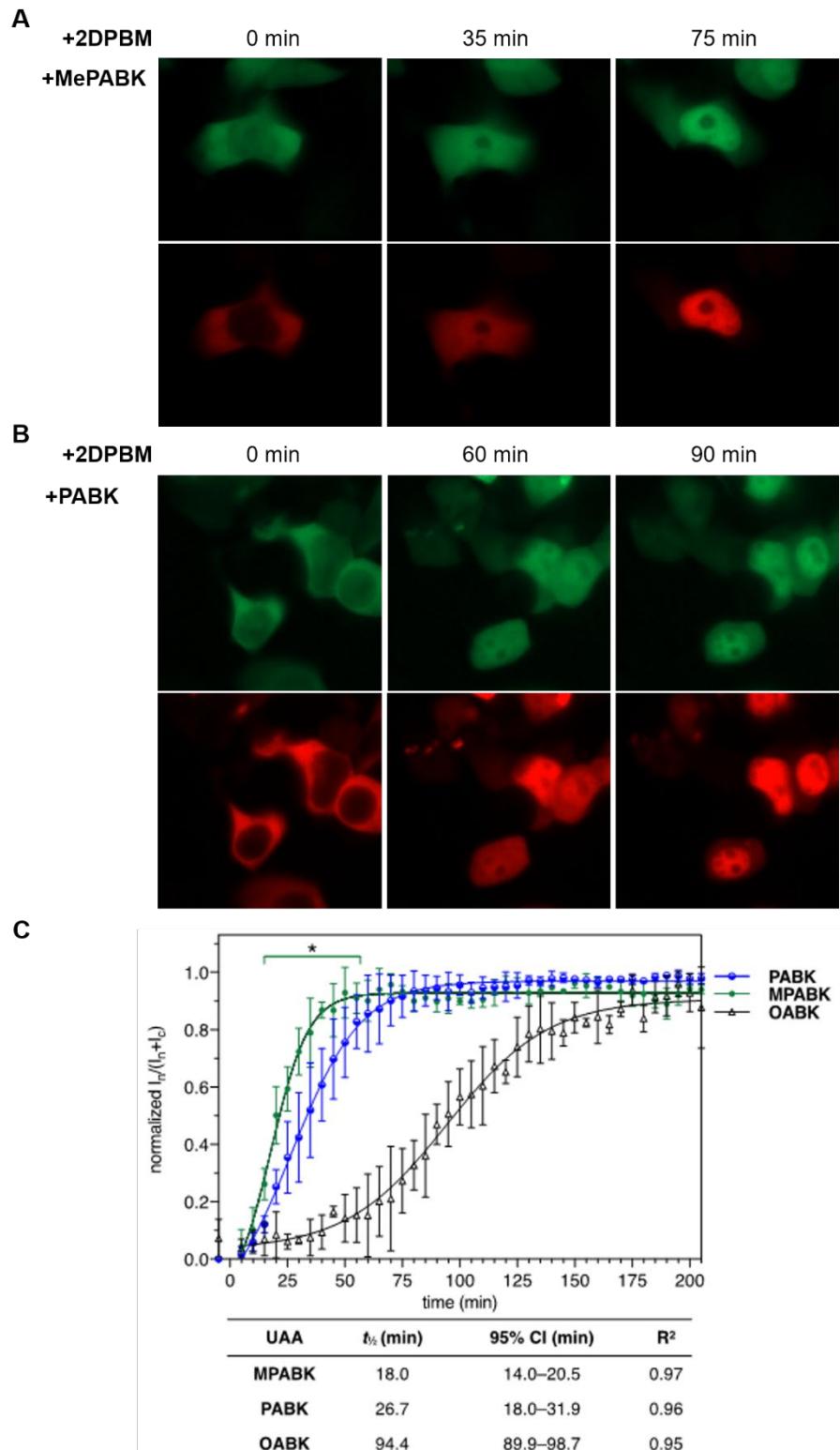
**Figure 8.14:** Genetically encoding of **MPABK** and **PABK** in mammalian cells.

A) Fluorescence micrographs of HEK293T cells expressing pMbOABKRS/PABKRS-mCherry-TAG-EGFP-HA and Pyl $tRNA_{CUA}$  in the absence and presence of **MPABK** (0.5 mM) and **PABK** (0.25 mM). B) Western blot of lysate from HEK293T cells expressing mCherry-**MPABK/PABK**-EGFP-HA with an anti-HA tag antibody and anti- $\beta$ -actin loading control confirms the fidelity of the incorporation.

### 8.3.3 Triggering of protein translocation through Staudinger reduction

In order to measure quantitative deprotection in live cells, small-molecule-activated nuclear translocation was investigated by genetically incorporating **MPABK** and **PABK** into the nuclear localization signal (NLS) of the transcription factor SATB1 (Figure 8.16). The presence of the

azidobenzyl group at position K29 completely inhibited the translocation of an mCherry reporter fused to the NLS, as observed in HEK293T cells that express SATB1\_K29**MPABK/PABK**-mCherry. After 24 hrs' incubation, the cells were treated with 2DPBM (100  $\mu$ M) and nuclear translocation of mCherry was triggered through a Staudinger reduction-induced formation of the native SATB1-NLS. Translocation was completed within 90 min, as shown by time-lapse imaging, and fluorescence quantification over time revealed a  $t_{1/2}$  of ~30 min, which is comparable to the light activation of the SATB1 nuclear translocation process. Based on the time-course imaging, the **MPABK**-driven decaging process is relatively faster than that induced by **PABK**. The fluorescence analyses were carried out by Josh Wesalo in the Deiters lab (Figure 8.16C).



**Figure 8.15:** Speeding up small-molecule control of protein translocation.

Micrographs of HEK293T cells show no nuclear localization of **MPABK(A)/PABK(B)-SATB1-mCherry** and complete activation of translocation from the cytoplasm to the nucleus after activation with 2DPBM. EGFP expression was used as a transfection and imaging control. C) Nuclear translocation kinetics for mCherry-SATB1-EGFP with SATB1-K29 mutated to **MPABK** and **PABK**(0.5 mM), compared to **OABK**.

### 8.3.4 Small-molecule control of post-translational modifications

Post-translational modifications dynamically alter the molecular function, localization, and stability of proteins, and thus regulate almost all aspects of biology.<sup>40</sup> There is extensive crosstalk between modifications, making it a persistent challenge to address the function of any specific modification and to understand the molecular effects of protein modifications by genetic approaches. It is often difficult to produce a homogeneously, site-specifically modified protein and to perform studies about the role of a post-translational modification. Genetic code expansion approaches, however, empower us to control post-translational modifications with a high degree of spatial and temporal resolution or with an increased level of molecular precision. As a first foray into understanding post-translational modifications using this technology, we chose to study the biology of Small Ubiquitin-Like Modifier (SUMO). SUMO has come under active investigation for its pleiotropic roles in cancer, cell signaling, and the stress response.<sup>263</sup> In order to begin our study of SUMO, we chose to focus on the prototypical SUMOylation substrate, Ran GTPase activating protein 1 (RanGAP1). This protein plays an essential role in nuclear transport by stimulating RanGTP hydrolysis in the cytoplasmic compartment,<sup>264</sup> and is one of the most efficient SUMOylation targets known.<sup>265</sup> In mammalian cells, native RanGAP1 is predominantly cytoplasmic, but modification by small ubiquitin-related modifier protein (SUMO)

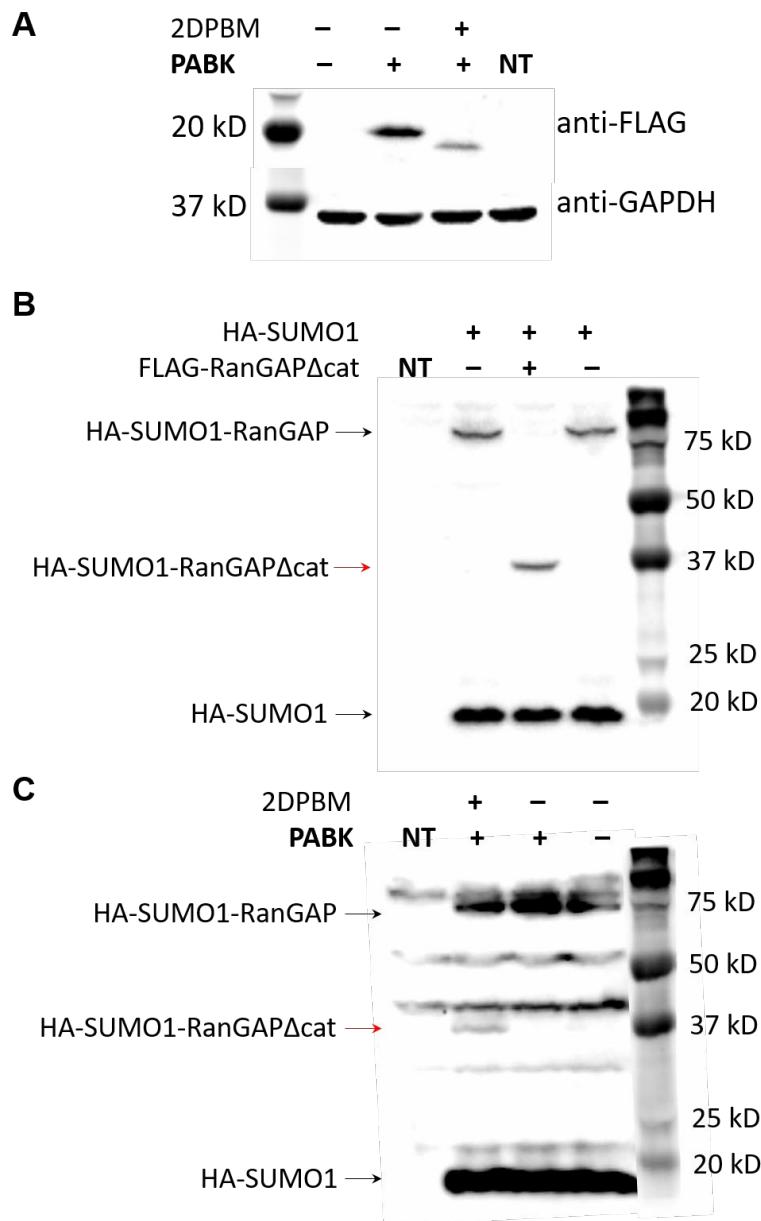
targets RanGAP1 to the cytoplasmic face of nuclear pore complex (NPC).<sup>264</sup> A single conserved lysine residue in RanGAP1, K526, is sumoylated.<sup>266</sup> By replacing the substrate lysine residue with a caged UAA (**PABK**), we can prevent cells from SUMOylating RanGAP1 until exposure to a phosphine trigger, helping us to elucidate the behavior of native RanGAP1 and the dynamics of RanGAP1 SUMOylation, which are currently unclear.<sup>264</sup> Further, validating this approach with RanGAP1, the prototypical SUMOylation substrate, will prove the principle of this technique and pave the way to applying it to other proteins that are conditionally SUMOylated in live cells.<sup>267</sup>

In order to demonstrate whether **PABK** could be incorporated into RanGAPΔcat-K526TAG, pFLAG-RanGAPΔcat-K526TAG and pPylPABKRS-4PylT were co-transfected in HEK 293T cells in the absence and presence of PABK for 24 hrs post-transfection. The transfected cells were then treated with 2DPBM (100 μM) for 12 hrs. Cells were lysed and proteins were analyzed using Western blots (Figure 8.16A). The results showed that incorporation of **PABK** into FLAG-RanGAP1Δcat-K526TAG was successful, and the amber suppression only occurs in the presence of **PABK** (Figure 8.16A). Importantly, the band intensity is reduced and the mobility shift is changed upon treatment with 2DPBM (Figure 8.16A), suggesting that 1) part of activated RanGAP1Δcat is SUMOylated and the SUMOylated RanGAP1Δcat is gel-shifted to a higher molecular weight, but it is not detected using anti-FLAG antibody; 2) non-modified activated RanGAP1Δcat runs faster than inactivated RanGAP1Δcat-K526**PABK**, which is not able to be targeted by SUMO1.

After successful incorporation, SUMOylation is the next step. For SUMOylation phenotype on Western blots, wild type pRanGAPΔcat was transfected with or without pH-A-SUMO1 in HEK 293T cells (Figure 8.16B). The Western blots clearly showed that wild-type

RanGAP1 $\Delta$ cat is successfully modified using HA-SUMO1. Transfected HA-SUMO1 also target the endogenous RanGAP1 substrate, and the endogenous SUMO1 modified RanGAP1 $\Delta$ cat as well.

In order to achieve the small-molecule control of SUMOylation, triple transfections of pFLAG-RanGAP $\Delta$ cat-KTAG, pH-A-SUMO1, and pFLAG-PylPABKRS-4PylT were carried out in HEK 293T cells in the absence and presence of PABK with or without the treatment of 2DPBM (Figure 8.16C). The Western blotting results demonstrates that RanGAP1 $\Delta$ cat-K526TAG is modified by HA-SUMO1 only when **PABK** is present and the cells are treated with 2DPBM. The results are very promising, but the strong background is observed, since there is crosstalk between exogenous SUMO1 and endogenous SUMO1. It suggests that the overexpressed SUMO1 is not needed. In order to get better signals, the target protein (RanGAP1) antibody will be used in the future studies.

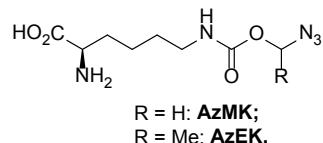


**Figure 8.16:** Small-molecule-triggered SUMOylation of SUMO1 and RanGAP1.

A) Western blots for incorporation of **PABK** into RanGAP $\Delta$ cat (top: anti-FLAG; bottom: anti-GAPDH as loading controls); B) Western blots for conjugation between wt-SUMO1 and wt- RanGAP $\Delta$ cat using anti-HA antibody; C) Western blots for 2DPBM-activated SUMOylation using anti-HA antibody.

## 8.4 GENETIC ENCODING OF THE AZIDOMETHYLENE AND THE AZIDOETHYL LYSINE ANALOGS

The azidomethylene (AzM) group has been used to control the fluorogenetic reporters in oligonucleotide detection. Given its small size, AzM modified UAAs are likely to be recognized by the wild-type PylRS for incorporation into proteins. An AzM lysine (**AzMK**) was synthesized by Dr. Qingyang Liu, as well as a more stable azidoethyl lysine (**AzEK**). Unfortunately, both of UAAs were extremely toxic to mammalian cells, so **AzMK** and **AzEK** were re-synthesized by Dr. Kunihiko Morihiro to improve the purity and reduce the cytotoxicity (Figure 8.17). Incorporation using a batch of **AzMK** and **AzEK** was performed in mammalian cells, and their incorporation in *E. coli* was tested by Josh Wesalo.



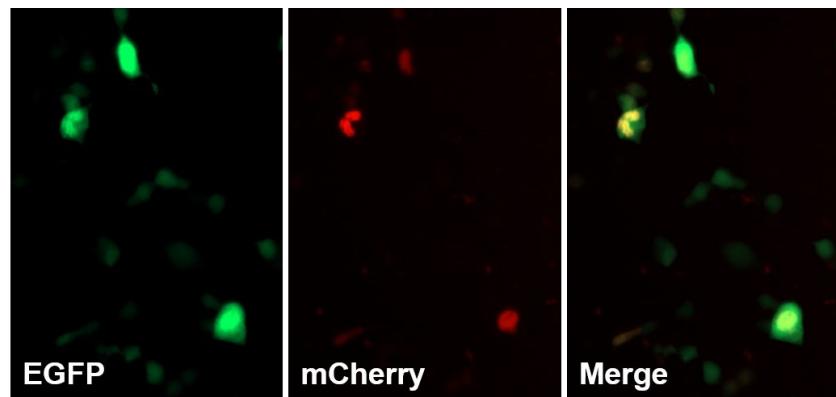
**Figure 8.17:** Structures of genetically encoded azido lysine analogs – **AzMK** and **AzEK**.

These amino acids were re-synthesized by Dr. Kunihiko Morihiro in the Deiters lab.

**AzMK** and **AzEK** were incorporated into sfGFP by the wild-type PylRS in *E. coli* (the experiments were performed by Joshua Wesalo in the Deiters lab). However, the ESI-MS results of both sfGFP-**AzMK** and -**AzEK** showed that the fully-reduced lysine was incorporated into sfGFP, which suggests that aliphatic azido groups might be degraded during bacterial protein expression, or reduced in the cytosol environment.

Thus, the incorporation and stability of **AzMK** and **AzEK** was further investigated in mammalian cells. SatB1 nuclear localization reporter is a good reporter to demonstrate both

incorporation and activation in one assay. HEK 293T cells were co-transfected with pEGFP-SatB1-KTAG-mCherry and pwtPylRS-4PylT in the presence of **AzMK** or **AzEK** (0.5 mM). After 24 hrs incubation, the cells were imaged by Zeiss microscope. The mCherry fluorescence was only localized in the nucleus (Figure 8.18), indicating that lysine is on SatB1, instead of **AzMK** or **AzEK**. Thus, neither **AzMK** nor **AzEK** was intact incorporated into proteins in mammalian cells. Taken together, **AzMK** and **AzEK** were not stable during incorporation in both bacterial and mammalian cells.

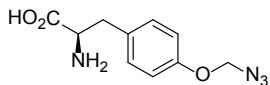


**Figure 8.18:** Small-molecule-triggered protein translocation.

Micrographs of HEK293T cells show no cytoplasmic localization of **AzEK-SATB1-mCherry** and complete activation of translocation from the cytoplasm to the nucleus before adding 2DPBM. EGFP expression was used as a transfection and imaging control.

## 8.5 GENETIC ENCODING OF THE AZIDOMETHYLENE TYROSINE

Due to the instability of **AzMK** and **AzEK** in both *E. coli* and mammalian cells, an AzM modified tyrosine was synthesized by Dr. Subhas Samantha (Figure 8.19).



**Figure 8.19:** Structures of genetically encoded azido tyrosine – **AzMY**.

It was synthesized by Dr. Subhas Samantha in the Deiters lab.

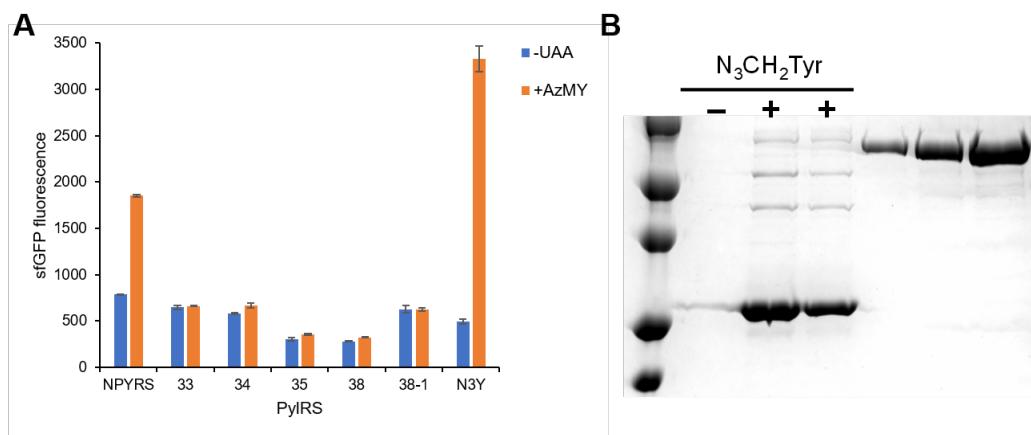
Through screening using a panel of engineered PylRS mutants for tyrosine derivatives (Table 8.2), synthetase V39 (termed as N<sub>3</sub>YRS) showed the highest sfGFP fluorescent signal relative to the “–UAA” control, suggesting that it is the hit synthetase(Figure 8.20A).

**Table 8.2:** *MbPylRS* panel screening for incorporation of **AzMY** in *E. coli*.

Code	Mutations								
	A 267	L 270	Y 271	L 274	N 311	C 313	M 315	Y 349	W 382
NPYRS		F		M	G	G		F	
V33		I	F	G	G	G		F	
V34		I	F	M	G	G		F	
V35		F	F	G	G	G		F	
V36						A		F	
V37			M	V		A		F	
V38	T			S	V	G		F	
V39	T				A	A		F	T

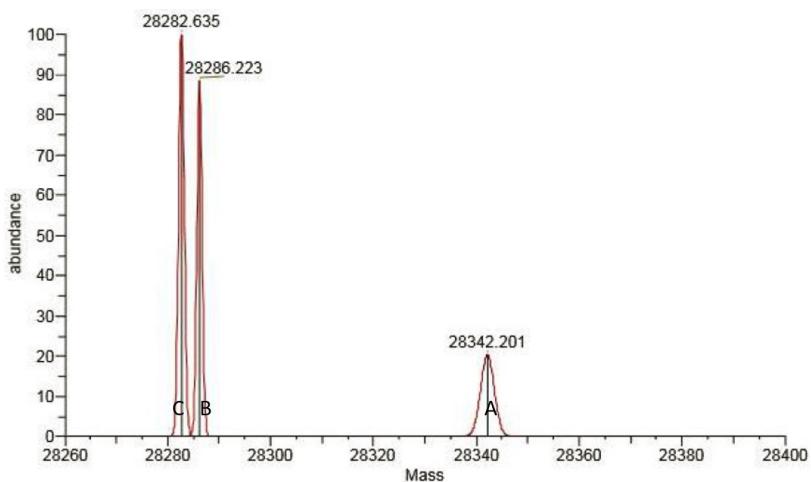
The incorporation of **AzMY** into sfGFP was further confirmed by SDS-PAGE with protein expression yield: 8.2 mg/L (Figure 8.20B), and then ESI-MS analysis was carried out in order to demonstrate the fidelity of incorporation and intact sfGFP-**AzMY**. The molecular weight of sfGFP-**AzMY** (Peak A) determined by ESI-MS, was in agreement with the expected

molecular weight (observed: 28,342.201 Da; expected: 28,341.66 Da). Specifically, ESI-MS analysis of purified proteins revealed a mass of [sfGFP-AzMY + Na<sup>+</sup>] of 28,342.201 Da (expected mass: 28,318.66 Da with Na<sup>+</sup>) for incorporation of AzMY. The mass of peak B and peak C corresponds to a mass of [sfGFP-Y with Na<sup>+</sup>] and [sfGFP-Y with H<sub>3</sub>O<sup>+</sup>], respectively (Figure 8.21). It indicated that part of AzMY was reduced to native Tyr during protein expression. The following investigation and application (small-molecule control of topoisomerase I through Staudinger reduction) will be carried out using AzMY in *E. coli* by Joshua Wesalo in the Deiters lab.



**Figure 8.20:** Genetic encoding of AzMY in *E. coli*.

A) Synthetase screening. B) Coomassie stained gel of Ni-NTA purified sfGFP-Y151TAG expression in the absence and presence of AzMY. (protein expression yield: 8.2 mg/L)



**Figure 8.21:** ESI-MS analysis of sfGFP-AzMY.

Observed mass of sfGFP-AzMY is 28,342.201 Da and expected mass is 28,341.66 Da.

## 8.6 CONCLUSIONS

In conclusion, the developed Staudinger reduction-based switch may provide a general method for the small molecule activation of proteins – as demonstrated in the conditional activation of protein functions as diverse as enzymatic activity (luciferase, Cre recombinase, and Cas9), protein maturation (EGFP), protein-protein interactions (SatB1 nuclear localization sequence), and protein-nucleic acid interactions (Cre recombinase and Cas9). While numerous small molecule inhibitors of protein function have been discovered through classical screening approaches (or have been engineered through bump & hole and targeted degradation strategies) and have been applied to the study of numerous biological processes, small molecule activators of specific protein function are rare. The small molecule activation approach reported here is highly specific since only a single, selected amino acid in only the protein of interest is modified.

It is easily implemented based on rational design following structural or mechanistic data, and, compared to existing methodologies that typically rely on the expression of fusion proteins or the conditional activation of protein expression, the site-specific introduction of a small protecting group at an essential site of the protein produces minimal structural perturbations while very effectively blocking protein function. Removal of the protecting group through a bioorthogonal Staudinger reduction activates protein function by generating the wild-type protein without any appended fusion domains. The timing of activation can be controlled on the minute time scale and thus is faster than small molecule activation of transcription, as in the case of the well-established Tet-On system, for example. In contrast to optical control of protein function, no specialized photostimulation equipment is required and the developed small molecule switch is orthogonal to optical switches in the visible spectrum, potentially enabling precisely staged protein activation events. If necessary, depending on the study that is being conducted, the endogenous, non-phosphine controlled copy of the protein could be deleted through gene editing (CRISPR/Cas9 or TALENs) or be knocked-down (RNAi or antisense).

Furthermore, we have demonstrated the genetic encoding of three new azidobenzyloxycarbonyl lysine analogues, **MPABK**, **PABK**, and **MOABK**, in bacterial and mammalian cells. These new analogues display enhanced chemical decaging rates compared to the previously encoded **OABK**. As a proof-of-principle, phosphine-activated nuclear translocation was achieved in live cells by caging a key lysine residue with **MPABK** and **PABK** in the NLS of the transcription factor SATB1, and fast, complete protein translocation from the cytoplasm to the nucleus was observed comparable to optical triggering of a caged SATB1-NLS. Importantly, we utilized the genetically encoded **PABK** for efficient phosphine activation of a prototypical SUMOylation substrate, RanGAP1,

to achieve small-molecule control of sumoylation in live cells, which will help us to elucidate the behavior of native RanGAP1 and the dynamics of RanGAP1 SUMOylation. Taken together, these results demonstrate that **MPABK** and **PABK** are new and valuable types of bioorthogonal Staudinger reduction-based probes that could be further utilized for the investigation and regulation of protein function and cellular processes in live cells.

In addition, we tested genetic encoding of **AzMK** and **AzEK** using wild type PylRS in both *E. coli* and mammalian cells, suggesting that the azido group is unfortunately reduced or cleaved during the incorporation process. Given the instability of **AzMK** and **AzEK**, we further developed an AzM modified tyrosine and discovered a new PylRS for **AzMY** incorporation through synthetase panel screening. Its applications will be further investigated by Joshua Wesalo in the Deiters lab.

## 8.7 EXPERIMENTAL

### Protein expression, purification, and reduction.

The plasmid pBAD-sfGFP-Y151TAG-pylT was co-transformed with pBK-OABKRS into chemically competent *E. coli* Top10 cells. A single colony was used to inoculate LB media overnight and 500 µL of the overnight culture was added to 25 mL of LB media, supplemented with 1 mM of OABK and 25 µg/mL of tetracycline and 50 µg/mL of kanamycin. Cells were grown at 37 °C, 250 rpm, and protein expression was induced with 0.2% arabinose when the OD600 reached ~0.5. After overnight expression at 37 °C, cells were pelleted and washed once with PBS. The cell pellet was re-suspended in 6 mL of phosphate lysis buffer (50 mM, pH 8.0) and Triton X-100 (60 µL, 10%), gently mixed, and incubated for an hour at 4 °C. The cell

suspensions were sonicated with 6 short bursts of 30 s followed by intervals of 30 s for cooling, and then the cell lysates were centrifuged at 4 °C, 13,000 g, for 10 minutes. The supernatant was transferred to a 15 mL conical tube and 100 µL of Ni-NTA resin (Qiagen) was added. The mixture was incubated at 4 °C for 2 h under shaking. The resin was then collected by centrifugation (500 g, 10 min), washed twice with 400 µL of lysis buffer, followed by two washes with 400 µL of wash buffer containing 20 mM imidazole. The protein was eluted with 400 µL of elution buffer containing 250 mM imidazole. After lysis of a bacterial expression culture, a Staudinger reaction was performed with **5** (1 mM) in cell lysates for 4 h at room temperature under shaking (250 rpm). The reaction mixture was centrifuged (5,000 rpm, 5 min) and the supernatant was transferred to a 15 mL conical tube followed by protein purification as described above. The purified proteins were analyzed by 10% SDS PAGE and stained with Coomassie Blue. ESI-MS analysis was carried out to show the incorporation of **OABK** into sfGFP and its reduction.

### **Incorporation of OABK in mammalian cells.**

HEK 293T cells were seeded into a 6-well cell culture microplate and grown to approximately 80% confluence in Dulbecco's modified eagle medium (DMEM) with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. Cells were transfected with two plasmids, pMbPylOABKRS-mCherry-TAG-EGFP-HA and p4CMVE-U6-PylT. Transfection was carried out using linear PEI (Millipore), according to the manufacturer's protocol. The growth media with 10% FBS contained either **OABK** (0.5 mM) or no unnatural amino acid. Cells were imaged on a Zeiss AxioObserver for mCherry and EGFP fluorescence after 20 h. Cells were then lysed using Lysis Buffer (Sigma) supplemented with complete protease inhibitor cocktail (Sigma).

After lysis, the cell debris was pelleted and the supernatant was added to 4X SDS sample loading buffer. Samples were analyzed by SDS-PAGE. Western blotting was carried out to detect reporter protein using mouse anti-HA antibody (Cell Signaling), visualized with an anti-mouse HRP conjugate (Santa Cruz). As a protein loading control, Western blotting was also carried out to detect the endogenous levels of total *GAPDH* protein using a mouse anti-GAPDH antibody (Santa Cruz) visualized with an HRP-conjugated anti-mouse secondary antibody (Santa Cruz).

#### **Live-cell fluorescence imaging and fluorescence measurement.**

HEK 293T cells were seeded into an 8-well chamber slide (Lab-Tek) and allowed to grow to approximately 70% confluence in DMEM (10% FBS). Co-transfection of pOABKRS-4PyIT and pEGFP-K29TAG-SatB1-mCherry or pEGFP-K85TAG-mCherry was carried out using the same protocol as above. After transfection, cells were grown for 20 h, the growth media with transfection mixtures was removed, and DMEM without phenol red but containing **OABK** was added into the well, followed by a 3 h incubation and treatment with **5** (0.5 mM) for 5 h while fluorescence imaging (a micrograph was collected every 5 min using a Nikon A1 confocal microscope with a Plan Apochromat 40X oil objective at 2X scan zoom). The mean fluorescence intensities in the nucleus, in case of activation of SatB1-mCherry, or the cytoplasm, in the case of EGFP activation, were quantified using NIS Elements software to enable the  $F_n/F_{n\max}$  or  $F_c/F_{c\max}$  ratio to be determined and normalized.

#### **Luciferase assay to screen different phosphines for protein activation in cells.**

Plasmids that encode pGL3-K206TAG and pOABKRS-PyIT were transfected in the presence of **OABK** using the same protocol as above. Cells were incubated in fresh medium for 2 h to

remove **OABK**, followed by phosphine treatment for 4 h. Luciferase activity was measured using a microplate reader (Tecan M100PRO) and a BrightGlo luciferase assay kit (Promega).

### **Cytotoxicity studies of phosphines.**

HEK 293T cells were seeded at 10,000 cells per well in clear bottom 96-well plates and incubated for 24 h. Cells were treated in triplicate with phosphine derivatives or 0.1% DMSO for 24 h. Cells were washed with 150 µL of fresh DMEM, and then 100 µL of DMEM and 40 µL of activated XTT-menadione reagent (1 mg/mL XTT in PBS containing 8 µL/mL menadione activation reagent, Alfa Aesar) were added to each well according to the manufacturer's instructions. Absorbance was immediately read out at 450 nm (XTT) and 630 nm (background) using a Tecan M1000 Pro microplate reader. After the addition of XTT reagents, the cells were incubated for 4 h in the 37 °C incubator. After 4 h incubation, absorbance was measured again. The cell viability was determined by subtracting 630 nm readings from 450 nm absorbances, then subtracting the averages of background wells (starting time point) from experimental wells (4 h), followed by normalization to the 0.1% DMSO control.<sup>268</sup>

### **Plasmid construction for mammalian cell applications.**

1) pM<sub>b</sub>PylOABKRS-mCherry-TAG-EGFP: The plasmid (pPylRS\_Y271A-mCherry-TAG-EGFP) was obtained by converting the TAC (Y271) codon of wild type PylRS to a GGC (A271) codon using primers P1 & P2 (table below) and a QuikChange site-directed mutagenesis kit (Agilent). The plasmid (pPylRS\_Y271A\_Y349F-mCherry-TAG-EGFP) was generated by converting the TAC (Y349) codon of PylRS\_Y271A into a TTC (F349) codon using primers P3 & P4 (table below) and the same mutagenesis method.

2) p*Mb*PylOABKRS: The *E. coli* codon-optimized plasmid (*Mb*PylOABKRS\_Y271A\_Y349F) was constructed using the same mutagenesis method as above. A and F mutations were introduced into wild-type *Mb*PylRS at Y271 and Y349 sites with two pairs of primers (P5 & P6 and P7 & P8; see table below), respectively.

3) p*Mb*PylOABKRS-PylT: The plasmid was obtained by ligating the p4CMVE-U6-PylT fragment from p*Mb*4PylT between the *Nhe*I and *Mfe*I restriction sites of p*Mb*OABKRS.

4) p*Mb*PylOABKRS-wtCre and p*Mb*PylOABKRS-Cre-K201TAG: The wtCre and Cre-K201TAG fragments were PCR amplified from p*Mb*PCKRS-wtCre/-CreK201TAG using the primers P9 & P10, digested with *Nhe*I and *Mfe*I, and ligated into the p*Mb*PylOABKRS-mCherry-TAG-EGFP vector in place of the mCherry-TAG-EGFP, generating the pOABKRS-wtCre and pOABKRS-Cre-K201TAG plasmids.

5) p*Mb*PylOABKRS-wtCas9 and p*Mb*PylOABKRS-Cas9-K866TAG: The pwtCas9 and pCas9-K866TAG fragments were generated from p*Mb*CKRS-wtCas9 and p*Mb*CKRS-Cas9-K866TAG<sup>68</sup> using the *Nhe*I and *Mfe*I restriction sites, respectively. p*Mb*PylOABKRS-mCherry-TAG-EGFP was digested with the same restriction enzymes (*Nhe*I and *Mfe*I) to remove the mCherry-TAG-EGFP-HA reporter. The wtCas9 and Cas9-K866TAG genes were ligated into the p*Mb*PylOABKRS-mCherry-TAG-EGFP backbone replacing the mCherry-EGFP using Quick ligase (NEB) to create p*Mb*PylOABKRS-wtCas9 and p*Mb*PylOABKRS-Cas9-K866TAG, respectively.

**Table 8.3:** Primer list.

Primer 1	ctatgctggcccccaccctggccaactacctgcggaaactg
Primer 2	cagttccgcaggtagtgtggccagggtggggccagcatag
Primer 3	gcgacagctgcatggtgtcggcgacaccctggacatc
Primer 4	gatgtccagggtgtcgccgaacaccatgcagctgtcgc
Primer 5	ctatgctggccgcaccctggccaactacctgcggtaactg
Primer 6	cagttaccgcaggtagtgtggccagggtggcggccagcatag
Primer 7	gcgacagctgcatggtgtcggcgacaccctggacatc
Primer 8	gatgtccagggtgtcgccgaacaccatgcagctgtcgc
Primer 9	gtcagatccgctagcacc
Primer 10	cgatcgatatcaattgtggttgtccaaactcatca

### Protein deprotection via Staudinger reduction.

After lysis of a bacterial expression culture (25 mL), a Staudinger reaction was performed with **5** (1 mM) in cell lysates (5 mL) for 4 h at room temperature under shaking (250 rpm). The reaction mixture was centrifuged (5,000 rpm, 5 min) and the supernatant was transferred to a 15 mL conical tube and 75 µL of Ni-NTA resin (Qiagen) was added. The mixture was incubated at 4 °C for 2 h under mild shaking. The resin was then collected by centrifugation (500 g, 10 min), washed three times with 400 µL of lysis buffer, followed by two washes with 400 µL of wash buffer containing 20 mM imidazole. The protein was eluted with 300 µL of elution buffer containing 250 mM imidazole. The purified protein was analyzed by 10% SDS PAGE and stained with Coomassie Blue. ESI-MS analysis confirmed that sfGFP-**OABK** could be reduced and deprotected via a Staudinger reaction, since the observed mass of 28226.41 Da is in agreement with the expected mass of 28227.89 Da of the deprotected protein.

### **Mass spec analysis of proteins.**

High-resolution ( $\geq$  10,000 R FWHM) exact mass measurement of sfGFP-**OABK** was carried out using Electrospray Ionization (ESI) on an Agilent Technologies (Santa Clara, California) 6210 LC-TOF mass spectrometer. Samples were analyzed via a 1  $\mu$ L flow injection at 300  $\mu$ L/min in a water:methanol mixture (25:75 v/v) with 0.1% formic acid. The mass spectrometer was operated in positive-ion mode with a capillary voltage of 4 kV, nebulizer pressure of 35 psig, and a drying gas flow rate of 12 L/min at 350 °C. The fragmentor and skimmer voltages were 200 and 60 V respectively. Reference ions of purine at *m/z* 121.0509 and HP-0921 at *m/z* 922.0098 were simultaneously introduced via a second orthogonal sprayer, and used as internal calibrants.

Thermo Q-Exactive Orbitrap mass spectrometer (Bremen, Germany), connected to a Dionex Ultimate 3000 UHPLC system was used for intact protein analysis (phosphine reduction and deprotection of sfGFP-**OABK**). The sample was analyzed through a ProSwift RP-10R, 1 mm by 5 cm column, flow rate of 200  $\mu$ L/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. Sheath gas, aux gas, and sweep gas flow rates were 35, 10, and 5, respectively. Source temperature was 250 °C. The S-lens RF level voltage was 50 V and the ion transfer tube temperature was 250 °C. The instrument was tuned and calibrated with standard thermo mix solution and the data was collected at the range of *m/z* 500-3000. The Xcaliber 3.0.63 and Protein Deconvolution 3.0 software was used for the data analysis.

### **Small molecule activation of Cre-**OABK** in live cells.**

HEK 293T cells (~20,000 cells/well) were plated into a 96-well cell culture microplate (Greiner),

and incubated overnight in 200 µL of DMEM growth media supplemented with 10% FBS and 1% penicillin/streptomycin at 37 °C in 5% CO<sub>2</sub>. At ~75% confluency, HEK293T cells were co-transfected with the p*MbOABKRS-wtCre* or p*MbOABKRS-CreK201TAG*, p4CMVE-U6-PylT,<sup>48</sup> and pC-SL<sup>257</sup> plasmids (100 ng of each) using linear PEI (2 µL per well, 0.323 mg/mL) in the presence or absence of **OABK** (0.25 mM) overnight at 37 °C. The media supplemented with **OABK** (0.25 mM) was replaced with 200 µL of fresh DMEM and incubation was continued for 2 h, followed by addition of **5** (0.5 mM) at 37 °C. Fluorescence imaging of the Cre activation reporter was performed after 24 h incubation. Media was replaced with DMEM without phenol red (Thermo Scientific) for imaging on a Zeiss Axio Observer Z1 microscope (10X objective, NA 0.8 plan-apochromat) with DsRed (E<sub>x</sub>: BP550/25; E<sub>m</sub>: BP605/70) and EGFP (E<sub>x</sub>: BP470/40; E<sub>m</sub>: BP525/50) filter cubes, then processed in Zen Pro 2012 imaging software. Fluorescent cell counting was performed on a Nikon A1 confocal microscope (10X objective) and analyzed using Elements software.

#### **Small molecule activation of Cas9-OABK in live cells.**

HEK 293T cells (~20,000 cells/well) were seeded into a 96-well cell culture microplate (Greiner), and incubated overnight in 200 µL of DMEM growth media supplemented with 10% FBS and 1% penicillin/streptomycin at 37 °C in 5% CO<sub>2</sub> humidified atmosphere. Before transfection, media was replaced with antibiotics-free DMEM supplemented with or without **OABK** (0.25 mM). Quadruple transfections were carried out with linear PEI (2 µL per well, 0.323 mg/mL) using p*MbOABKRS-wtCas9* or p*MbOABKRS-Cas9-K866TAG*, p4CMVE-U6-PylT,<sup>48</sup> gRNA plasmids,<sup>82</sup> and pRG dual reporter plasmid<sup>39</sup> (100 ng of each).

After a 24 h incubation, cells were washed with fresh DMEM and incubated for 2 h. For regulating gene editing in live cells, HEK 293T cells were treated with **5** (0.5 mM) or without **5**. After overnight incubation, media was replaced with 200  $\mu$ L of DMEM without phenol red, and cells were imaged on a Zeiss Axio Observer Z1 microscope (10X objective, NA 0.8 plan-apochromat) with EGFP ( $E_x$ : BP470/40;  $E_m$ : BP525/50) and DsRed ( $E_x$ : BP550/25;  $E_m$ : BP605/70) filter cubes. Fluorescent cell counting was performed on a Nikon A1 confocal microscope (10X objective) and analyzed using Elements software.

## **9.0 SUMMARY AND OUTLOOK**

### **9.1 CONCLUDING REMARKS**

Expanding the genetic code for site-specific incorporation of various biophysical probes and labels, such as fluorescent probes, photolabile caging groups, optical probes, photoswitches, bioorthogonal chemselective groups, offers a sophisticated tool for exploring protein structure and function, dissecting the cellular processes, and developing proteins with novel properties. Site-specific installation of distinct functional groups on a protein of interest is achieved using orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs in response to an amber stop codon (UAG) placed in the gene of interest. In this dissertation, genetic encoding of a series of diverse unnatural amino acids (UAAs), including photocaged lysines (**PCK/NPPK/NDBFK**), coumarin lysines (**HCK/BHCK/HC<sub>2</sub>K/ACK/AC<sub>2</sub>K**), phthalimide lysines (**APIC<sub>2</sub>K/DAPIC<sub>2</sub>K/DAPIC<sub>3</sub>K**), photocaged cysteines (**PCC/PCHC/HCC**), photocaged tyrosines (**NBY/NPY/MNPY/NPPY**), photoisomerizable phenylalanines (**AzoF/F<sub>2</sub>AzoF/F<sub>4</sub>AzoF**), and azido amino acids (**OABK/PABK/MOABK/MPABK/AzMK/AzEK/AzMY**), have been achieved through the discovery of the engineered pyrrolysyl-tRNA synthetase/tRNA pair from bacteria to eukaryotic cells. Strategic placement of these functional groups renders the protein inactive until deprotection through a bioorthogonal external trigger delivers the active wild-type protein. This developed methodology enables the conditional control of various cellular

processes with spatiotemporal precision in a non-invasive and tunable way, including enzymatic activation (FLuc), protein folding (EGFP), protein translocation (SATB1), gene repair (UvrD), post-translational modification (SUMO1), DNA recombination (Cre), RNA processing (Csy4), and gene editing (CRISPR/Cas9). These studies could provide new biological insights into protein interactions, protein conformational changes, and the role of post-translational modifications.

## 9.2 FUTURE WORK

Using this unnatural amino acid mutagenesis strategy, a large number of structurally diverse unnatural amino acids have been incorporated with high fidelity into proteins in response to nonsense and frameshift codons in our and other labs.

We propose to extend these studies in several directions:

- 1) Application of phage-assisted continuous evolution for selecting optimal synthetases for UAA incorporation.

Our lab has built a library of various UAAs over years, in order to further expand genetic toolbox, we could apply phage-assisted continuous evolution (PACE) selections,<sup>269</sup> instead of conventional directed evolution, to discover highly active and selective orthogonal PyLRSSs through hundreds of generations of mutation, selection, and replication.<sup>269</sup> Generally, the potential of genetic code expansion is limited to the low efficiency of UAA incorporation compared to endogenous protein expression. In combination with our screening methods, the development of the PACE selection would help us establish a rapid and effective approach to

enhancing aminoacylation activity and efficiency and improving amino acid specificity in *MbPylRS*.

2) Multiplexing – multiple incorporation and manipulation in a PTM system.

Development of UAA mutagenesis strategy offers the possibility to independently control the activity of several proteins within the same cell or at different extracellular locations. Multiplexing requires independent control of different protein activities by the parallel application of distinct UAAs to studying multiple proteins in a single system. For example, the development of coumarin-caged amino acids affords the engineering of wavelength-selective and orthogonally sequential activation of biological activity through incorporation of both nitrobenzyl- and coumarin-derivatized caging groups using different wavelengths of light to control orthogonal photoactivatable systems. Alternatively, both azido- and nitrobenzyl-caged amino acids could be simultaneously employed in combination of optical- and small-molecule-based strategies. This study would help elucidate confounding factors in a complex system.<sup>270</sup>

<sup>271</sup> Genetically encoding distinct unnatural amino acids into proteins synthesized in cells requires the introduction of evolved ribosomes and four-base blank codons has enabled the incorporation of more than one defined UAA in a protein of interest.<sup>272</sup> Recently, the Chin lab has engineered mutually orthogonal PylRS/tRNA pairs.<sup>273</sup> These technological developments facilitate multiple installation of UAAs at defined sites to make it applicable for future studies.

3) Application of UAA incorporation to mouse xenograft models for therapeutic investigations.

We have achieved adeno-associated viral (AAV) delivery of PylRS/tRNA pairs for efficient genetic code expansion in mammalian cells, and the Chin lab have further developed that the use of the AAV vectors enables the site-specific and efficient incorporation of UAAs

into neurons in primary neuronal culture, organotypic brain slices and the brains of live mice.<sup>274-276</sup> The mouse is a powerful mammalian, vertebrate model organism. This technological development allows us to extensively apply optochemical tools to probe and control signaling transduction in xenograft models for cancer therapeutic investigation.

Over the past few years, we have greatly expanded UAA mutagenesis using PylRS system and the range of applications to a variety of proteins in prokaryotes and eukaryotes. In this field, we are still facing major questions and challenges – 1) how to move photochemical tools beyond proof of principle to answer real biological questions; 2) how to transform this proof-of-concept UAA mutagenesis tool into routinely laboratory techniques and clinically-relevant therapy, which limits its utility and generality. Therefore, expanding the scope and impact of this technology is challenging but emerging and exciting.

## **10.0 EXPANDED METHODS**

### **10.1 GENERAL MOLECULAR AND CELL BIOLOGY TECHNIQUES**

#### **10.1.1 Protein expression and UAA incorporation in *E. coli***

The plasmid pBAD-POI-TAG-pylT (POI = protein of interest) was co-transformed with pBK-PylRS into chemically competent *E. coli* Top10 cells. A single colony was grown in LB media containing 25 µg/mL tetracycline and 50 µg/mL kanamycin overnight and 400 µL of the overnight culture was added to 25 mL LB media supplemented with 1 mM of UAA or without UAA (as a negative control) and 25 µg/mL tetracycline and 50 µg/mL kanamycin. Cells were grown at 37 °C, 250 rpm, and protein expression was induced with 0.1% arabinose when the OD<sub>600</sub> (Nanodrop) reached 0.4~0.6. After overnight expression at 37 °C, cells were harvested and washed with PBS. Cell pellets were re-suspended in 6 mL of phosphate lysis buffer (50 mM, pH 8.0), Triton X-100 (60 µL, 10%), and protease inhibitor (6 µL, Sigma), gently mixed, and incubated on ice for 1 h. Cell mixtures were sonicated (Fisher Scientific 550 Sonic Dismembrator) with six short bursts of 30 s on ice followed by 30 s intervals for cooling, and then the cell lysates were centrifuged at 4 °C and 13,000 g for 10 min. The supernatant was transferred to a 15 mL conical tube and 100 µL Ni-NTA resin (Qiagen) was added. The mixture was incubated at 4 °C for 2 h under mild shaking. The resin was then collected by centrifugation

(1,000 g, 10 min) at 4 °C, washed three times with 300 µL of lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, pH 8.0), followed by two washes with 300 µL of wash buffer (lysis buffer containing 20 mM imidazole). The protein was eluted twice with 200 µL of elution buffer (lysis buffer containing 250 mM imidazole). The elution time is varied based on the amounts of expressed proteins. The purified proteins were analyzed by 10% or 12% SDS-PAGE (60 V for 20 min, and then 150 V for 1 h), and stained with Coomassie Blue. The percentage of SDS-PAGE used is based on the molecular weight of the POI.

#### **10.1.2 Mammalian cell maintenance and passage**

Primary cell lines were maintained at 37 °C and 5% CO<sub>2</sub> in 10 cm culture plates, and manipulations were performed in a biosafety cabinet. The majority of the cell lines (including HEK 293T, HeLa, NIH 3T3, and COS-7) used here were grown in Dulbecco's Modified Eagle's Medium (DMEM, Gibco) supplemented with FBS (Sigma, 10%), and penicillin/streptomycin (100× P/S, Corning Cellgro, 1%). The DMEM solution was made by dissolving 5.89 g DMEM powder, 1.63 g sodium bicarbonate, and 0.145 g L-glutamine in 445 mL of Milli-Q purified water (Millipore). The pH was then adjusted to ~7.4 using 65% hydrochloric acid. Subsequently, 50 mL of FBS and 5 mL of a 100× P/S solution were added and the media was filter sterilized. A fresh media exchange (10 mL) was performed every 2 days for each 10 cm plate. The cells were passaged when confluence reached >90% (as determined by microscope) by removing the media and adding 1 mL of TrypLE Express reagent (Invitrogen), then incubated in the incubator (37 °C) for 2-3 min. Cells were lifted from the plate through pipetting and 9 mL of growth media was added.

For routinely passages into 10 cm plates, it was performed in a 1:10 dilution by adding 1 mL of the cell resuspension to 9 mL of media in a new 10 cm plate.

For passages into multi-well plates for transfections, the following volumes and cell numbers were used: 200  $\mu$ L ( $1.0\text{-}5.0 \times 10^4$  cells) per well in a 96-well plate (Greiner), 400  $\mu$ L ( $\sim 5.0 \times 10^4$  cells) per well in an 8-well chamber slide (Lab-Tek), and 2 mL ( $\sim 3.0 \times 10^5$  cells) per well in a 6-well plate (Corning). Cells were grown for  $\sim 24$  h prior to transfection. For 96-well plates (Greiner), black plates were used for fluorescence experiments, and white plates were used for chemiluminescence experiments.

#### **10.1.3 Cell transfection and UAA incorporation in mammalian cells**

Cell transfections were performed to introduce foreign DNA plasmid into the cells. When the cell confluence reaches 70%~80%, the cellular growth media was removed and replaced with P/S-free DMEM media containing UAAs for transfections. Transfection mixtures were then prepared by mixing the plasmids in Opti-MEM media (20  $\mu$ L per well for 96-well plate, 200  $\mu$ L per well for 6-well plate) and incubating with linear polyethylene imine (LPEI, 0.323 mg/mL) at 1-2  $\mu$ L per well for 96-well plates and 15-20  $\mu$ L per well for 6-well plates. Plasmid transfections were typically performed with 100 ng each plasmid in a 96-well plate (chemiluminescence or fluorescence readouts) and 1.5-2.0  $\mu$ g of each plasmid in a 6-well plate (for protein isolations). The transfection reagents were completely mixed with the plasmids, then incubated for 10-20 min at room temperature, and added to the cells dropwise. At this step, the amino acid was added at 0.25-1.0 mM concentration for expression of UAA-modified proteins.

For example, for UAA incorporation into the mCherry-TAG-EGFP fusion protein:

HEK 293T cells were seeded at ~50,000 cells per well and grown in Dulbecco's Modified Eagle's Medium (200 µL, DMEM, Gibco) supplemented with FBS (Sigma, 10%), Pen-Strep (Corning Cellgro, 1%) and L-glutamine (Alfa Aesar, 2 mM) in 96-well plates (Greiner) under a humidified atmosphere containing CO<sub>2</sub> (5%) at 37 °C. Cells were transiently transfected with pMbPylRS-mCherry-TAG-EGFP-HA and p4CMVE-U6-PylT (100 ng of each plasmid) at ~80% confluence using linear polyethylenimine (LPEI; 1.5 µL, 0.323 mg/mL) in DMEM (200 µL) without any UAA or with UAA (0.25-1.0 mM). After a 24 h incubation at 37 °C, the media were replaced with PBS and the cells were imaged with a Zeiss Axio Observer Z1 Microscope (10× objective) using EGFP (38HE: Ex 470/40; Em 525/50) and mCherry (43HE: Ex 550/25; Em 605/70) filter cubes.

#### **10.1.4 Protein isolation from mammalian cells**

Total protein isolation from ~10<sup>6</sup> cells in a single well of a 6-well plate was used to perform western blot analysis for confirmation of protein expression in mammalian cells. The cells were rinsed with 1 mL 4 °C PBS (pH 7.4), GE Healthcare mammalian cell lysis buffer (200 µL, Sigma-Aldrich) was added, and the cells were incubated by shaking on ice for 15 min. The cells were then collected via pipetting and centrifuged at 16,000 g for 20 min at 4 °C. Total protein lysates were separated from the cell debris and the supernatant was collected in a new tube, and then stored at -20 °C for future use.

### **10.1.5 Protein analysis (western blots)**

Standard western blot techniques were used for analysis of proteins. Total protein extract from a single well of a 6-well plate (20 µL) was denatured with protein loading dye at 95 °C for 5 min. The denatured protein solution was size separated on a 1.5-mm SDS-PAGE gel (10% or 12%), which was subsequently soaked in transfer buffers (25 mM Tris-base, 192 mM glycine, 20% methanol, 0.025% wt/vol SDS). The protein gel was then placed onto a GE Healthcare Hybond-LFP PVDF membrane (Fisher) and the following transfer was performed at 80 V for 90 min at 4 °C in a Bio-Rad Mini Trans-Blot module. Blocking of the membrane was performed with 5% milk in TBST buffer (10 mM Tris-base, 150 mM NaCl, pH 7.6, 10% W/V Tween 20) at room temperature for 1 h. Primary antibody (1:1000 dilution in 5 mL of TBST, according to the manufacturer's suggestions) binding was performed at 4 °C overnight or at room temperature for 1 h, followed by three times of washes with TBST. Secondary antibody (1:20,000 dilution in 5 mL of TBST, according to the manufacturer's suggestions) binding at room temperature was performed for 1 h, followed by three washes with TBST. All specific antibodies are based on the specific experiments (indicated in experimental sections). Chemiluminescent detection was performed and imaged on a ChemiDoc (Bio-Rad), according to the manufacturer's protocol.

For example, UAA incorporation into the mCherry-TAG-EGFP-HA fusion protein:

To confirm the fidelity of incorporation and the expression of the fusion protein, Western blots were performed. HEK 293T cells were co-transfected with pMbPylRS-mCherry-TAG-EGFP-HA and p4CMVE-U6-PylT (1.5 µg of each plasmid) using LPEI (15 µL, 0.323 mg/mL) in the presence or in the absence of UAA (0.25-1.0 mM) in 6-well plates. After a 24 h

incubation, the cells were washed with chilled PBS (2 mL) once and lysed in mammalian protein extraction buffer (200 µL, GE Healthcare) with complete protease inhibitor cocktail (Sigma) through gentle shaking at 4 °C for 15 min, and the cell lysates were cleared through centrifugation (14,000 g, 4 °C, 20 min). The protein lysates collected from the supernatant were boiled with loading buffer and analyzed by 10% SDS-PAGE (60 V for 20 min, and then 150 V for 1 h). After gel electrophoresis and transferring to a polyvinylidene difluoride (PVDF) membrane (80 V for 1.5 h, GE Healthcare), the membrane was blocked in TBS (5 mL) with Tween 20 (Fisher Scientific, 0.1%) and milk (5%) for 1 h. The membranes were probed and incubated with the primary α-HA antibody (Y-11) rabbit monoclonal IgG (sc-805, Santa Cruz Biotech) overnight at 4 °C, followed by a goat anti-rabbit IgG-HRP secondary antibody (sc-2031, Santa Cruz Biotech) for 1 h at room temperature. The membranes were washed with TBS with Tween 20 (0.1 %) three times, and then incubated in working solution (Thermo Scientific SuperSignal West Pico Chemiluminescent substrate) for 5 min. The Western blots were imaged using the “chemi” autofocus settings on a ChemiDoc (BioRad).

#### **10.1.6 Cell imaging**

Before imaging, growth media was replaced with 37 °C phenol red-free DMEM-high modified growth media (Thermo Scientific). The following filter cubes were used on a Zeiss Observer Z1 microscope: DAPI (filter set 68 HE; ex. BP377/28; em. BP464/100), EGFP (filter set 38 HE; ex. BP470/40; em. BP525/50), and DsRed/mCherry (filter set 43 HE; ex. BP575/25; em. BP605/70). The following Zeiss objectives were used: 5× (NA 0.16 plan-neofluora), 10× (NA 0.25 plan-apochromat), and 20× (NA 0.8 plan-apochromat). Cellular images were taken in brightfield as well as corresponding fluorescent channels, and then processed in Zen Pro 2011 imaging

software. Additional images were taken on a Nikon A1 confocal microscope and processed using Elements imaging software. Fluorescent cell counting was performed with ImageJ software (settings: threshold 5%-10%, size >200 pixels<sup>2</sup>, circularity 0-1).

#### **10.1.7 Cell irradiation**

The photochemical control of proteins was commonly performed in cell culture with UV exposure. Irradiation of the whole well was performed directly on top of a UV transilluminator ( $165.4 \pm 1.0 \text{ } \mu\text{mol}/(\text{m}^2 \cdot \text{s}^{-1})$ , photon flux measured at a distance of 1 cm from the UV light, 25 W, flux density  $6.3 \text{ mW/cm}^2$ ) for varying exposure times (2-4 min) at room temperature.

For spatial control, UV irradiations were performed with the UV transilluminator through precut designs using aluminum foils, which were taped at the bottom of the cell culture plates and left on during incubation periods to locate the area of interest.

In addition, localized irradiations were performed with a Zeiss Observer Z1 microscope (40 $\times$  objective, NA 0.75 plan-apochromat; Zeiss) and a DAPI filter set (68 HE ex. BP377/28) with a partially closed aperture to irradiate only a specific subset of cells, using the x:y location noted to locate the area of interest.

#### **10.1.8 Cell fixing**

Fixed cells were prepared in Nunc Lab-Tek II 4-well chamber slides (Thermo Scientific), which can be stained and stored for imaging. Cells were transfected and incubated for the required period of time, then washed twice with 500  $\mu\text{L}$  of PBS (pH 7.4) at 4 °C. The cells were then fixed on ice with 500  $\mu\text{L}$  of 3.75% formaldehyde in PBS for 15 min. After the fixing step, the

cells were washed with PBS three times, and permeabilized with 200 µL of 0.5% TritonX100 for 30 s at room temperature. The cells were washed and stained with specific reagents, including rhodamine-phalloidin actin staining (7 µL per mL of PBS containing 1% of BSA, 200 µL per well, 20 min room temperature, Life Technologies) and DAPI nuclear staining (15 µL of a 14.3 mM solution per mL PBS, 200 µL per well, 2 min room temperature, Invitrogen). After staining procedures, the cells were washed three times with 200 µL of PBS, the chamber slide top was removed carefully, and the mount was air-dried. Prolong Gold antifade reagent (Life Technologies) was then gently added dropwise, and a coverslip was evenly placed on the top to avoid air bubbles, and then dried overnight at room temperature.

### **10.1.9 RNA isolation**

Total RNA isolations from  $\sim 10^6$  cells in 6-well plates were performed using QIAzol reagent (Qiagen) following the manufacturer's protocol. The cellular media was removed, cells were rinsed with 1 mL of PBS (pH 7.4), and 700 µL of QIAzol reagent was added. The cells were incubated for 5 min at room temperature, and then lifted via pipetting into a microcentrifuge tube. Subsequently, chloroform was added (140 µL), mixed by shaking for 15 s, and incubated for 2-3 min at room temperature. The reaction was then phase separated by centrifugation at 12,000 g for 15 min at 4 °C, and the upper aqueous layer containing the RNA (around 350 µL) was transferred to a new tube. The RNA was precipitated from the reaction through addition of 525 µL of 100% ethanol, mixing by pipetting, and then purified through the Qiagen column. The column was washed by buffer twice and buffer once through centrifugation at 8,000 g for 15 s, and dried by centrifugation at 8,000 g for 1 min at room temperature. The total RNAs in the column were eluted using 30~50 µL RNase-free water, and the final concentrations (ng/µL) were

determined in triplicates with a NanoDrop spectrophotometer. The purified total RNAs were aliquoted and stored at a –80 °C freezer for future use.

#### **10.1.10 RNA analysis by quantitative real-time PCR (qRT-PCR)**

The first step of qRT-PCR was the reverse transcription of the RNA into DNA with the iScript cDNA Synthesis kit (Bio-Rad), using 100 ng of the RNA isolations (2 µL of 10 ng/µL dilutions) in 20 µL reactions with 1 µL of the reverse transcriptase enzyme, according to the manufacturer’s protocol. The RT reactions were heated to 25 °C for 5 min, then 42 °C for 45 min, and finally 95 °C for 5 min before being cooled down on ice. Quantitative PCR was then performed with the TaqMan® MicroRNA Kit (Bio-Rad) in 20 µL reactions using gene-specific primers, according to the manufacturer’s protocol. The cDNA product from the RT reaction (2 µL) was added to the qPCR reactions containing 10 µL of the 2× TaqMan reagent and 10 µM primers, one set for the target gene and one set for the internal control gene (GAPDH) in separate reactions in triplicate. Primer sequences are shown in each project-specific experimental section. The qPCRs were heated at 95 °C for 3 min then cycled 60 times from 95 °C for 10 s to 60 °C for 30 s. Amplification of the PCR product was monitored with Sybr Green fluorescence using a Bio-Rad CFX96 Real-Time System. The Cq values and relative gene expressions ( $\Delta(\Delta Cq)$ ) were determined with the Bio-Rad CFX Manager 3.1 software. Technical triplicates of the qPCRs were used to determine normalized gene expression relative to the GAPDH controls.

## **10.2 COMMON RECIPE**

### **10.2.1 DMEM media**

For 500 mL, combine:

5.89 g DMEM (Hyclone)

1.63 g sodium bicarbonate

0.145 g L-glutamine

440 mL milliQ water

Mix until everything is dissolved.

Adjust pH to 7.4 with addition of 65% HCl.

In the cell culture hood, adding

50 mL FBS,

5 mL Pen/Strep,

Sterilization using 0.22 µm filter.

Store media at 4 °C for up to 2 months.

### **10.2.2 PBS buffer (10×)**

For 500 mL, combine:

40 g NaCl

1 g KCl

7.2 g disodium phosphate

1.2 g monopotassium phosphate

500 mL milliQ water

Mix until everything is dissolved.

Adjust pH to 7.4 with addition of 60% HCl.

Autoclaved.

### **10.2.3 TBS buffer (10×)**

For 1 L, combine:

87.6 g NaCl

12.1 g Tris base (VWR)

1 L milliQ water

Mix until everything is dissolved.

Adjust pH to 7.6 with addition of 60% HCl.

### **10.2.4 TBST buffer (1×)**

For 500 mL, combine:

50 mL TBS buffer (10×)

0.5 g Tween

450 mL MilliQ water

Mix until everything is dissolved.

### **10.2.5 Transfer buffer (10×)**

For 1 L, combine:

30.3 g Tris

144 g glycine

1 L MilliQ water

Mix until everything is dissolved.

### **10.2.6 Transfer buffer (1×)**

For 1 L, combine:

100 mL Transfer buffer (10×)

200 mL methanol

0.25 g SDS

700 mL milliQ water

### **10.2.7 SDS running buffer (10×)**

For 1 L, combine:

144 g glycine

30.2 g Tris base

10 g SDS

1 L MilliQ water

Mix until everything is dissolved.

### **10.2.8 Staining buffer (coomassie blue)**

For 50 mL, combine:

5 mL glacial acetic acid

10 mL methanol

15 mg coomassie R250 (brilliant blue)

35 mL MilliQ water

### **10.2.9 Destaining buffer**

For 1 L, combine:

100 mL glacial acetic acid

200 mL methanol

700 mL milliQ water

### **10.2.10 10% or 12% SDS-PAGE**

For 10 mL, combine:

6.8 mL or 6.3 mL MilliQ water

2.5 mL or 3.0 mL 40% acrylamide

2.5 mL 1.5 M Tris buffer (pH 8.8)

0.1 mL 10% SDS

0.1 mL 10% APS

6 µL TEMED

### **10.2.11 4% SDS-PAGE stacking gel**

For 10 mL, combine:

7.3 mL milliQ water

1.25 mL 40% acrylamide

1.25 mL 1 M Tris buffer (pH 6.8)

0.1mL 10% SDS

0.1mL 10% APS

10 $\mu$ L TEMED

### **10.2.12 Lysis buffer for protein purification**

50 mM NaH<sub>2</sub>PO<sub>4</sub>

300 mM NaCl

10 mM imidazole

using NaOH to adjust pH to 8.0

### **10.2.13 Wash buffer for protein purification**

lysis buffer containing 20 mM imidazole:

50 mM NaH<sub>2</sub>PO<sub>4</sub>

300 mM NaCl

20 mM imidazole

#### **10.2.14 Elution buffer for protein purification**

lysis buffer containing 250 mM imidazole:

50 mM NaH<sub>2</sub>PO<sub>4</sub>

300 mM NaCl

250 mM imidazole

## APPENDIX A

### LIST OF UAA ABBREVIATION

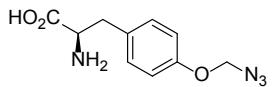
**Table A:** List of UAA abbreviation.

UAA Abbreviation	Chemical Structures	Descriptions
<b>PCK</b>		photocaged lysine
<b>NPPK</b>		<i>o</i> -nitrophenylpropyl lysine
<b>NDBFK</b>		two-photon <i>o</i> -nitrodibenzofuran caged lysine
<b>HCK</b>		hydroxycoumarin caged lysine

<b>BHCK</b>		two-photon coumarin caged lysine
<b>HC<sub>2</sub>K</b>		hydroxycoumarin fluorophore
<b>ACK</b>		aminocoumarin caged lysine
<b>AC<sub>2</sub>K</b>		aminocoumarin fluorophore
<b>APIC<sub>2</sub>K</b>		4-aminophthalimide lysine
<b>DAPIC<sub>2</sub>K</b>		4-( <i>N,N</i> -dimethylamino)phthalimide lysine
<b>DAPIC<sub>3</sub>K</b>		4-( <i>N,N</i> -dimethylamino)phthalimide lysine

<b>PCC</b>		photocaged cysteine
<b>PCHC</b>		photocaged homocysteine
<b>HCC</b>		hydroxycoumarin cysteine
<b>NBY</b>		photocaged tyrosine
<b>NPY</b>		<i>o</i> -nitropiperonyl tyrosine
<b>MNPY</b>		<i>o</i> -nitropiperonylmethyl tyrosine
<b>NPPY</b>		<i>o</i> -nitrophenylpropyl tyrosine

<b>AzoF</b>		4-azophenylphenylalanine
<b>F<sub>2</sub>AzoF</b>		<i>p</i> -azo(2,6-difluorophenyl) phenylalanine
<b>F<sub>4</sub>AzoF</b>		<i>p</i> -azo(2,6,2',6'-tetrafluorophenyl) phenylalanine
<b>OABK</b>		<i>o</i> -azidobenzyl lysine
<b>PABK</b>		<i>p</i> -azidobenzyl lysine
<b>MOABK</b>		<i>o</i> -azidobenzylmethyl lysine
<b>MPABK</b>		<i>p</i> -azidobenzylmethyl lysine
<b>AzMK</b>		azidomethylene lysine
<b>AzEK</b>		azidoethyl lysine

AzMY		azidomethylene tyrosine
------	---	-------------------------

## APPENDIX B

### LIST OF PUBLICATIONS

**Table B:** List of publications.

Published Papers	Corresponding Section
<p><i>Genetically encoded optical activation of DNA recombination in human cells.</i></p> <p>Luo, J.; Arbely, E.; Zhang, J.; Chou, C.; Upadhyay, R.; Chin, J. W.; Deiters, A. <i>Chem Commun</i> <b>2016</b>, 135(36), 13433-13439.</p>	2.3
<p><i>Genetically encoded optochemical probes for simultaneous fluorescence reporting and light activation of protein function with two-photon excitation.</i></p> <p>Luo, J.; Upadhyay, R.; Naro, Y.; Chou, C.; Nguyen, D. P.; Chin, J. W.; Deiters, A. <i>J Am Chem Soc</i> <b>2014</b>, 136(44), 15551-15558.</p>	3.0

<p><i>Optical control of DNA helicase function through genetic code expansion.</i></p> <p>Luo, J.; Kong, M.; Li, L.; Samanta, S.; Van Houten, B.; Deiters, A.</p> <p><i>Chembiochem</i> <b>2017</b>, 18(5), 466-469.</p>	3.4
<p><i>Genetic encoding of caged cysteine and caged homocysteine in bacterial and mammalian cells.</i></p> <p>Uprety, R.; Luo, J.; Liu, J.; Naro, Y.; Samanta, S.; Deiters, A.</p> <p><i>Chembiochem</i> <b>2014</b>, 15(36), 13433-13439.</p>	5.0
<p><i>Genetic encoding of photocaged tyrosines with improved light-activation properties for the optical control of protease function.</i></p> <p>Luo, J.; Torres-Kolbus, J.; Liu, J.; Deiters, A. <i>Chembiochem</i> <b>2017</b>, 18(14), 1442-1447.</p>	6.0
<p><i>Reversible and tunable photoswitching of protein function through genetic encoding of azobenzene amino acids in mammalian cells.</i></p> <p>Luo, J.; Samanta, S.; Convertino, M.; Dokholyan, N. V.; Deiters, A.</p> <p><i>Chembiochem</i> <b>2018</b>, 19(20), 2178–2185.</p>	7.0

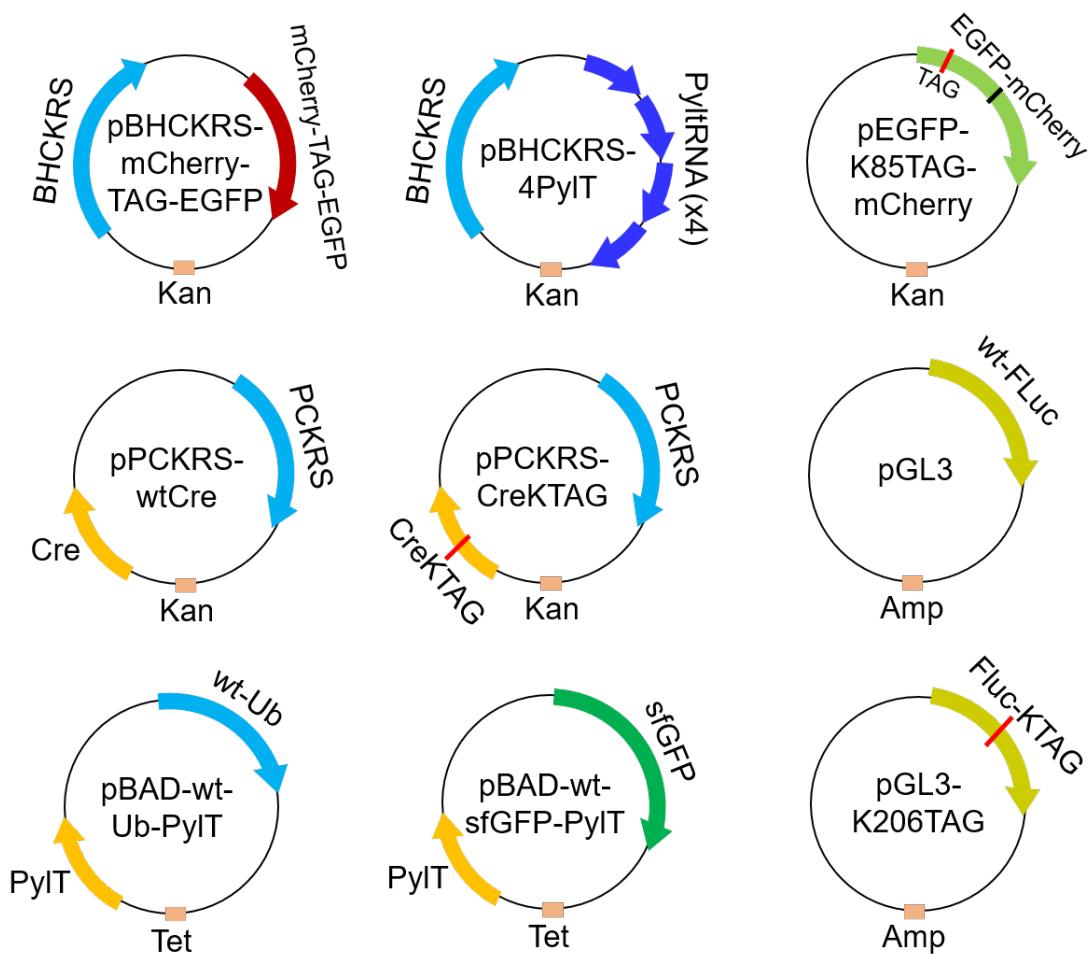
*Small-molecule control of protein function through Staudinger reduction.*

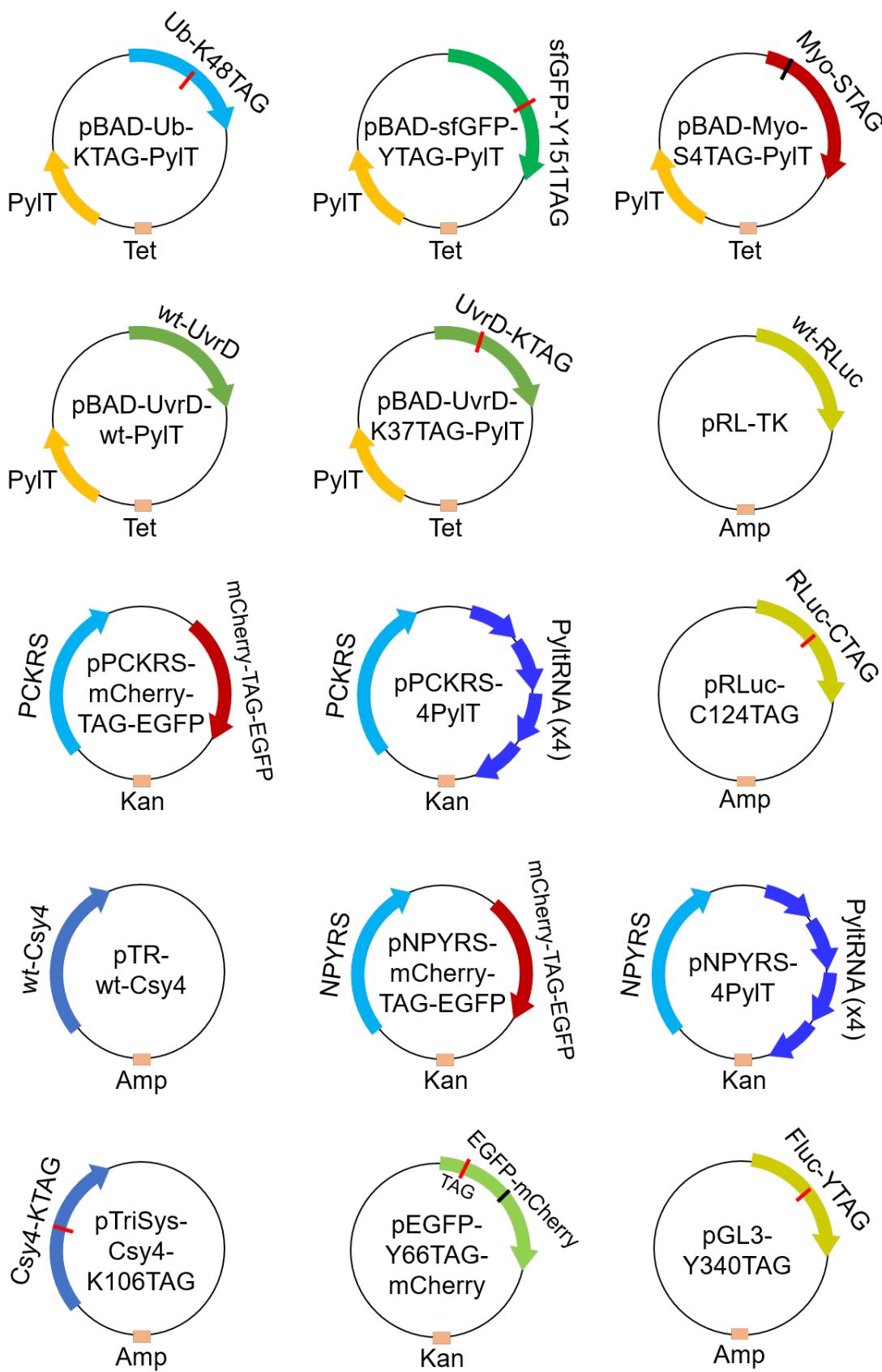
Luo, J.; Liu, Q.; Morihiro, K.; Deiters, A. *Nat Chem* **2016**, 8(11), 1027-1034.

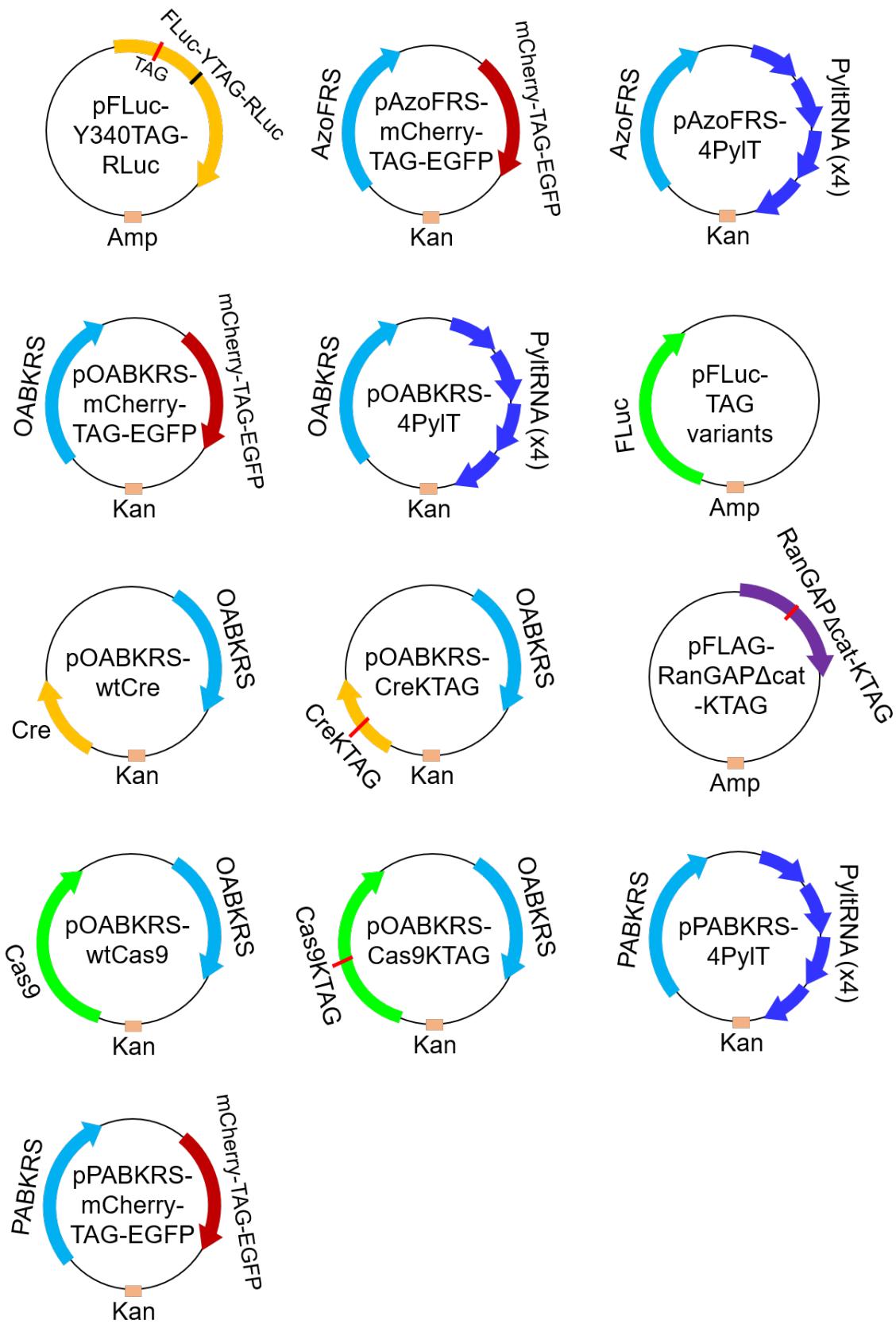
8.0

## APPENDIX C

### LIST OF PLASMID MAPS







## BIBLIOGRAPHY

1. Wang, L.; Xie, J.; Schultz, P. G., Expanding the genetic code. *Annual review of biophysics and biomolecular structure* **2006**, *35*, 225-49.
2. Wang, L.; Schultz, P. G., Expanding the genetic code. *Angewandte Chemie* **2004**, *44* (1), 34-66.
3. Muir, T. W., Semisynthesis of proteins by expressed protein ligation. *Annu Rev Biochem* **2003**, *72*, 249-89.
4. Noren, C. J.; Anthony-Cahill, 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), 182-8.
5. Young, T. S.; Schultz, P. G., Beyond the canonical 20 amino acids: expanding the genetic lexicon. *J Biol Chem* **2010**, *285* (15), 11039-44.
6. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., Expanding the genetic code of Escherichia coli. *Science* **2001**, *292* (5516), 498-500.
7. Wang, L.; Schultz, P. G., A general approach for the generation of orthogonal tRNAs. *Chem Biol* **2001**, *8* (9), 883-90.
8. 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), 1459-62.
9. Rakshit, R.; Navarro, R.; Wandless, T. J., Chemical Biology Strategies for Posttranslational Control of Protein Function. *Chem Biol* **2014**, *21* (9), 1238-1252.
10. Yu, H.; Li, J.; Wu, D.; Qiu, Z.; Zhang, Y., Chemistry and biological applications of photo-labile organic molecules. *Chem Soc Rev* **2010**, *39* (2), 464-73.
11. Fynno, L.; Yizhar, O.; Deisseroth, K., The development and application of optogenetics. *Annu Rev Neurosci* **2011**, *34*, 389-412.
12. Riggsbee, C. W.; Deiters, A., Recent advances in the photochemical control of protein function. *Trends in biotechnology* **2010**, *28* (9), 468-75.
13. Gautier, A.; Gauron, C.; Volovitch, M.; Bensimon, D.; Jullien, L.; Vriz, S., How to control proteins with light in living systems. *Nat Chem Biol* **2014**, *10* (7), 533-41.
14. Baker, A. S.; Deiters, A., Optical Control of Protein Function through Unnatural Amino Acid Mutagenesis and Other Optogenetic Approaches. *AcS Chem Biol* **2014**, *9* (7), 1398-1407.

15. Zhang, F.; Vierock, J.; Yizhar, O.; Fenno, L. E.; Tsunoda, S.; Kianianmomeni, A.; Prigge, M.; Berndt, A.; Cushman, J.; Polle, J.; Magnuson, J.; Hegemann, P.; Deisseroth, K., The microbial opsin family of optogenetic tools. *Cell* **2011**, *147* (7), 1446-57.
16. Christie, J. M.; Gawthorne, J.; Young, G.; Fraser, N. J.; Roe, A. J., LOV to BLUF: flavoprotein contributions to the optogenetic toolkit. *Mol Plant* **2012**, *5* (3), 533-44.
17. Tischer, D.; Weiner, O. D., Illuminating cell signalling with optogenetic tools. *Nature reviews. Molecular cell biology* **2014**, *15* (8), 551-8.
18. Guglielmi, G.; Falk, H. J.; De Renzis, S., Optogenetic Control of Protein Function: From Intracellular Processes to Tissue Morphogenesis. *Trends Cell Biol* **2016**, *26* (11), 864-874.
19. Zhou, X. X.; Chung, H. K.; Lam, A. J.; Lin, M. Z., Optical control of protein activity by fluorescent protein domains. *Science* **2012**, *338* (6108), 810-4.
20. Baker, A. S.; Deiters, A., Optical control of protein function through unnatural amino acid mutagenesis and other optogenetic approaches. *Acs Chem Biol* **2014**, *9* (7), 1398-407.
21. Deiters, A., Principles and applications of the photochemical control of cellular processes. *Chembiochem : a European journal of chemical biology* **2010**, *11* (1), 47-53.
22. Golynskiy, M. V.; Koay, M. S.; Vinkenborg, J. L.; Merkx, M., Engineering protein switches: sensors, regulators, and spare parts for biology and biotechnology. *Chembiochem : a European journal of chemical biology* **2011**, *12* (3), 353-61.
23. Bell, S. P.; Dutta, A., DNA replication in eukaryotic cells. *Annu Rev Biochem* **2002**, *71*, 333-74.
24. Miesenbock, G., The optogenetic catechism. *Science* **2009**, *326* (5951), 395-9.
25. Davey, J. A.; Wilson, C. J., Deconstruction of complex protein signaling switches: a roadmap toward engineering higher-order gene regulators. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **2017**, *9* (6).
26. Vale, R. D., The molecular motor toolbox for intracellular transport. *Cell* **2003**, *112* (4), 467-80.
27. Kennedy, M. J.; Hughes, R. M.; Peteya, L. A.; Schwartz, J. W.; Ehlers, M. D.; Tucker, C. L., Rapid blue-light-mediated induction of protein interactions in living cells. *Nat Methods* **2010**, *7* (12), 973-5.
28. Merrick, W. C., Mechanism and regulation of eukaryotic protein synthesis. *Microbiological reviews* **1992**, *56* (2), 291-315.
29. Levskaia, A.; Weiner, O. D.; Lim, W. A.; Voigt, C. A., Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature* **2009**, *461* (7266), 997-1001.
30. Shimizu-Sato, S.; Huq, E.; Tepperman, J. M.; Quail, P. H., A light-switchable gene promoter system. *Nat Biotechnol* **2002**, *20* (10), 1041-4.
31. Radu, I.; Bamann, C.; Nack, M.; Nagel, G.; Bamberg, E.; Heberle, J., Conformational changes of channelrhodopsin-2. *J Am Chem Soc* **2009**, *131* (21), 7313-9.

32. Renicke, C.; Schuster, D.; Usherenko, S.; Essen, L. O.; Taxis, C., A LOV2 domain-based optogenetic tool to control protein degradation and cellular function. *Chem Biol* **2013**, *20* (4), 619-26.
33. Crosson, S.; Rajagopal, S.; Moffat, K., The LOV domain family: photoresponsive signaling modules coupled to diverse output domains. *Biochemistry-US* **2003**, *42* (1), 2-10.
34. Lee, J.; Natarajan, M.; Nashine, V. C.; Socolich, M.; Vo, T.; Russ, W. P.; Benkovic, S. J.; Ranganathan, R., Surface sites for engineering allosteric control in proteins. *Science* **2008**, *322* (5900), 438-42.
35. Inoue, T.; Heo, W. D.; Grimley, J. S.; Wandless, T. J.; Meyer, T., An inducible translocation strategy to rapidly activate and inhibit small GTPase signaling pathways. *Nat Methods* **2005**, *2* (6), 415-8.
36. Christie, J. M.; Arvai, A. S.; Baxter, K. J.; Heilmann, M.; Pratt, A. J.; O'Hara, A.; Kelly, S. M.; Hothorn, M.; Smith, B. O.; Hitomi, K.; Jenkins, G. I.; Getzoff, E. D., Plant UVR8 photoreceptor senses UV-B by tryptophan-mediated disruption of cross-dimer salt bridges. *Science* **2012**, *335* (6075), 1492-6.
37. Pathak, G. P.; Vrana, J. D.; Tucker, C. L., Optogenetic control of cell function using engineered photoreceptors. *Biol Cell* **2013**, *105* (2), 59-72.
38. Liu, H.; Yu, X.; Li, K.; Klejnot, J.; Yang, H.; Lisiero, D.; Lin, C., Photoexcited CRY2 interacts with CIB1 to regulate transcription and floral initiation in Arabidopsis. *Science* **2008**, *322* (5907), 1535-9.
39. Giles, N. M.; Giles, G. I.; Jacob, C., Multiple roles of cysteine in biocatalysis. *Biochem Biophys Res Commun* **2003**, *300* (1), 1-4.
40. Davis, L.; Chin, J. W., Designer proteins: applications of genetic code expansion in cell biology. *Nature reviews. Molecular cell biology* **2012**, *13* (3), 168-82.
41. Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G., A genetically encoded photocaged amino acid. *J Am Chem Soc* **2004**, *126* (44), 14306-7.
42. Deiters, A.; Groff, D.; Ryu, Y.; Xie, J.; Schultz, P. G., A genetically encoded photocaged tyrosine. *Angewandte Chemie* **2006**, *45* (17), 2728-31.
43. Chen, P. R.; Groff, D.; Guo, J. T.; Ou, W. J.; Cellitti, S.; Geierstanger, B. H.; Schultz, P. G., A Facile System for Encoding Unnatural Amino Acids in Mammalian Cells. *Angew Chem Int Edit* **2009**, *48* (22), 4052-4055.
44. Schaeffer, H. J.; Weber, M. J., Mitogen-activated protein kinases: specific messages from ubiquitous messengers. *Mol Cell Biol* **1999**, *19* (4), 2435-44.
45. Shaul, Y. D.; Seger, R., The MEK/ERK cascade: from signaling specificity to diverse functions. *Biochim Biophys Acta* **2007**, *1773* (8), 1213-26.
46. Gautier, A.; Deiters, A.; Chin, J. W., Light-activated kinases enable temporal dissection of signaling networks in living cells. *J Am Chem Soc* **2011**, *133* (7), 2124-7.
47. Adams, S. R.; Tsien, R. Y., Controlling cell chemistry with caged compounds. *Annual review of physiology* **1993**, *55*, 755-84.

48. Gautier, A.; Nguyen, D. P.; Lusic, H.; An, W.; Deiters, A.; Chin, J. W., Genetically encoded photocontrol of protein localization in mammalian cells. *J Am Chem Soc* **2010**, *132* (12), 4086-8.
49. Young, D. D.; Deiters, A., Photochemical activation of protein expression in bacterial cells. *Angewandte Chemie* **2007**, *46* (23), 4290-2.
50. Chou, C.; Deiters, A., Light-activated gene editing with a photocaged zinc-finger nuclease. *Angewandte Chemie* **2011**, *50* (30), 6839-42.
51. Chou, C.; Young, D. D.; Deiters, A., A light-activated DNA polymerase. *Angewandte Chemie* **2009**, *48* (32), 5950-3.
52. Chou, C.; Young, D. D.; Deiters, A., Photocaged t7 RNA polymerase for the light activation of transcription and gene function in pro- and eukaryotic cells. *Chembiochem : a European journal of chemical biology* **2010**, *11* (7), 972-7.
53. Edwards, W. F.; Young, D. D.; Deiters, A., Light-activated Cre recombinase as a tool for the spatial and temporal control of gene function in mammalian cells. *Acs Chem Biol* **2009**, *4* (6), 441-5.
54. Nagy, A., Cre recombinase: The universal reagent for genome tailoring. *Genesis* **2000**, *26* (2), 99-109.
55. Kim, Y.; Ho, S. O.; Gassman, N. R.; Kirlann, Y.; Landorf, E. V.; Collart, F. R.; Weiss, S., Efficient site-specific labeling of proteins via cysteines. *Bioconjugate chemistry* **2008**, *19* (3), 786-91.
56. Basle, E.; Joubert, N.; Pucheault, M., Protein chemical modification on endogenous amino acids. *Chem Biol* **2010**, *17* (3), 213-27.
57. Antelmann, H.; Helmann, J. D., Thiol-based redox switches and gene regulation. *Antioxidants & redox signaling* **2011**, *14* (6), 1049-63.
58. Reddie, K. G.; Carroll, K. S., Expanding the functional diversity of proteins through cysteine oxidation. *Current opinion in chemical biology* **2008**, *12* (6), 746-54.
59. Spassier, L.; Brik, A., Chemistry and biology of the ubiquitin signal. *Angewandte Chemie* **2012**, *51* (28), 6840-62.
60. Tasaki, T.; Kwon, Y. T., The mammalian N-end rule pathway: new insights into its components and physiological roles. *Trends Biochem Sci* **2007**, *32* (11), 520-8.
61. Lamkanfi, M.; Festjens, N.; Declercq, W.; Vanden Berghe, T.; Vandenabeele, P., Caspases in cell survival, proliferation and differentiation. *Cell death and differentiation* **2007**, *14* (1), 44-55.
62. Riedl, S. J.; Shi, Y., Molecular mechanisms of caspase regulation during apoptosis. *Nature reviews. Molecular cell biology* **2004**, *5* (11), 897-907.
63. 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.
64. Netto, L. E.; de Oliveira, M. A.; Monteiro, G.; Demasi, A. P.; Cussiol, J. R.; Discola, K. F.; Demasi, M.; Silva, G. M.; Alves, S. V.; Faria, V. G.; Horta, B. B., Reactive cysteine in

proteins: protein folding, antioxidant defense, redox signaling and more. *Comparative biochemistry and physiology. Toxicology & pharmacology : CBP* **2007**, *146* (1-2), 180-93.

65. Woll, D.; Laimgruber, S.; Galetskaya, M.; Smirnova, J.; Pfleiderer, 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. *J Am Chem Soc* **2007**, *129* (40), 12148-58.
66. Woll, D.; Smirnova, J.; Galetskaya, M.; Prykota, T.; Buhler, J.; Stengele, K. P.; Pfleiderer, W.; Steiner, U. E., Intramolecular sensitization of photocleavage of the photolabile 2-(2-nitrophenyl)propoxycarbonyl (NPPOC) protecting group: photoproducts and photokinetics of the release of nucleosides. *Chemistry* **2008**, *14* (21), 6490-7.
67. 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.
68. Conti, E.; Franks, N. P.; Brick, P., Crystal structure of firefly luciferase throws light on a superfamily of adenylylate-forming enzymes. *Structure* **1996**, *4* (3), 287-98.
69. Fraga, H.; Fernandes, D.; Novotny, J.; Fontes, R.; Esteves da Silva, J. C., Firefly luciferase produces hydrogen peroxide as a coproduct in dehydroluciferyl adenylylate formation. *Chembiochem : a European journal of chemical biology* **2006**, *7* (6), 929-35.
70. Jones, D. P., Radical-free biology of oxidative stress. *American journal of physiology. Cell physiology* **2008**, *295* (4), C849-68.
71. Momotake, A.; Lindegger, N.; Niggli, E.; Barsotti, R. J.; Ellis-Davies, G. C., The nitrodibenzofuran chromophore: a new caging group for ultra-efficient photolysis in living cells. *Nat Methods* **2006**, *3* (1), 35-40.
72. Ellis-Davies, G. C., Caged compounds: photorelease technology for control of cellular chemistry and physiology. *Nat Methods* **2007**, *4* (8), 619-28.
73. Balaban, R. S.; Nemoto, S.; Finkel, T., Mitochondria, oxidants, and aging. *Cell* **2005**, *120* (4), 483-95.
74. Finkel, T., Radical medicine: treating ageing to cure disease. *Nature reviews. Molecular cell biology* **2005**, *6* (12), 971-6.
75. Collins, Y.; Chouchani, E. T.; James, A. M.; Menger, K. E.; Cocheme, H. M.; Murphy, M. P., Mitochondrial redox signalling at a glance. *J Cell Sci* **2012**, *125* (Pt 4), 801-6.
76. Bastin, G.; Singh, K.; Dissanayake, K.; Mighiu, A. S.; Nurmohamed, A.; Heximer, S. P., Amino-terminal cysteine residues differentially influence RGS4 protein plasma membrane targeting, intracellular trafficking, and function. *J Biol Chem* **2012**, *287* (34), 28966-74.
77. Delaunay, A.; Pflieger, D.; Barrault, M. B.; Vinh, J.; Toledano, M. B., A thiol peroxidase is an H<sub>2</sub>O<sub>2</sub> receptor and redox-transducer in gene activation. *Cell* **2002**, *111* (4), 471-81.
78. Greaves, J.; Chamberlain, L. H., Palmitoylation-dependent protein sorting. *J Cell Biol* **2007**, *176* (3), 249-54.
79. Smotrys, J. E.; Linder, M. E., Palmitoylation of intracellular signaling proteins: regulation and function. *Annu Rev Biochem* **2004**, *73*, 559-87.

80. Higdon, A.; Diers, A. R.; Oh, J. Y.; Landar, A.; Darley-Usmar, V. M., Cell signalling by reactive lipid species: new concepts and molecular mechanisms. *Biochem J* **2012**, *442* (3), 453-64.
81. 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.
82. Pace, N. J.; Weerapana, E., Diverse functional roles of reactive cysteines. *Acs Chem Biol* **2013**, *8* (2), 283-96.
83. Fass, D., Disulfide bonding in protein biophysics. *Annual review of biophysics* **2012**, *41*, 63-79.
84. Miller, A. L., The methionine-homocysteine cycle and its effects on cognitive diseases. *Alternative medicine review : a journal of clinical therapeutic* **2003**, *8* (1), 7-19.
85. Wierzbicki, A. S., Homocysteine and cardiovascular disease: a review of the evidence. *Diabetes & vascular disease research : official journal of the International Society of Diabetes and Vascular Disease* **2007**, *4* (2), 143-50.
86. Seshadri, S.; Beiser, A.; Selhub, J.; Jacques, P. F.; Rosenberg, I. H.; D'Agostino, R. B.; Wilson, P. W. F.; Wolf, P. A., Plasma homocysteine as a risk factor for dementia and Alzheimer's disease. *New Engl J Med* **2002**, *346* (7), 476-483.
87. Jakubowski, H., Metabolism of homocysteine thiolactone in human cell cultures. Possible mechanism for pathological consequences of elevated homocysteine levels. *J Biol Chem* **1997**, *272* (3), 1935-42.
88. Glushchenko, A. V.; Jacobsen, D. W., Molecular targeting of proteins by L-homocysteine: mechanistic implications for vascular disease. *Antioxidants & redox signaling* **2007**, *9* (11), 1883-98.
89. Jakubowski, H., Pathophysiological consequences of homocysteine excess. *The Journal of nutrition* **2006**, *136* (6 Suppl), 1741S-1749S.
90. Paoli, P.; Sbrana, F.; Tiribilli, B.; Caselli, A.; Pantera, B.; Cirri, P.; De Donatis, A.; Formigli, L.; Nosi, D.; Manao, G.; Camici, G.; Ramponi, G., Protein N-homocysteinylation induces the formation of toxic amyloid-like protofibrils. *Journal of molecular biology* **2010**, *400* (4), 889-907.
91. Undas, A.; Perla, J.; Lacinski, M.; Trzeciak, W.; Kazmierski, R.; Jakubowski, H., Autoantibodies against N-homocysteinylated proteins in humans: implications for atherosclerosis. *Stroke; a journal of cerebral circulation* **2004**, *35* (6), 1299-304.
92. 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 Chem Biol* **2009**, *4* (6), 409-27.
93. Gardner, L.; Deiters, A., Light-controlled synthetic gene circuits. *Current opinion in chemical biology* **2012**, *16* (3-4), 292-299.
94. Brieke, C.; Rohrbach, F.; Gottschalk, A.; Mayer, G.; Heckel, A., Light-controlled tools. *Angewandte Chemie* **2012**, *51* (34), 8446-76.

95. 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. *Nat Chem Biol* **2007**, *3* (12), 769-72.
96. Kang, J.-Y.; Kawaguchi, D.; Coin, I.; Xiang, Z.; O'Leary, Dennis D. M.; Slesinger, Paul A.; Wang, L., In Vivo Expression of a Light-Activatable Potassium Channel Using Unnatural Amino Acids. *Neuron* **2013**, *80* (2), 358-370.
97. Nguyen, D. P.; Mahesh, M.; Elsässer, S. J.; Hancock, S. M.; Uttamapinant, C.; Chin, J. W., Genetic Encoding of Photocaged Cysteine Allows Photoactivation of TEV Protease in Live Mammalian Cells. *J Am Chem Soc* **2014**, *136* (6), 2240-2243.
98. Ren, W.; Ji, A.; Ai, H. W., Light activation of protein splicing with a photocaged fast intein. *J Am Chem Soc* **2015**, *137* (6), 2155-8.
99. 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. *J Am Chem Soc* **2014**, *136* (6), 2240-3.
100. 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 (BBA) - Proteins and Proteomics* **2014**, *1844* (6), 1059-1070.
101. Neumann, H.; Peak-Chew, S. Y.; Chin, J. W., Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat Chem Biol* **2008**, *4* (4), 232-4.
102. 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.
103. 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.
104. 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. *J Am Chem Soc* **2009**, *131* (25), 8720-1.
105. Loening, A. M.; Fenn, T. D.; Gambhir, S. S., Crystal structures of the luciferase and green fluorescent protein from *Renilla reniformis*. *Journal of molecular biology* **2007**, *374* (4), 1017-28.
106. Woo, J.; Howell, M. H.; von Arnim, A. G., Structure-function studies on the active site of the coelenterazine-dependent luciferase from *Renilla*. *Protein science : a publication of the Protein Society* **2008**, *17* (4), 725-35.
107. Luo, J.; Upadhyay, 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. *J Am Chem Soc* **2014**, *136* (44), 15551-8.

108. 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 : a European journal of chemical biology* **2014**, *15* (12), 1793-9.
109. Bort, G.; Gallavardin, T.; Ogden, D.; Dalko, P. I., From one-photon to two-photon probes: "caged" compounds, actuators, and photoswitches. *Angewandte Chemie* **2013**, *52* (17), 4526-37.
110. Krueger, A. T.; Imperiali, B., Fluorescent amino acids: modular building blocks for the assembly of new tools for chemical biology. *Chembiochem : a European journal of chemical biology* **2013**, *14* (7), 788-99.
111. Goncalves, M. S., Fluorescent labeling of biomolecules with organic probes. *Chemical reviews* **2009**, *109* (1), 190-212.
112. Klan, P.; Solomek, T.; Bochet, C. G.; Blanc, A.; Givens, R.; Rubina, M.; Popik, V.; Kostikov, A.; Wirz, J., Photoremovable protecting groups in chemistry and biology: reaction mechanisms and efficacy. *Chemical reviews* **2013**, *113* (1), 119-91.
113. Deiters, A., Light activation as a method of regulating and studying gene expression. *Current opinion in chemical biology* **2009**, *13* (5-6), 678-686.
114. 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 Chem Biol* **2009**, *4* (6), 409-427.
115. Furuta, T.; Takeuchi, H.; Isozaki, M.; Takahashi, Y.; Kanehara, M.; Sugimoto, M.; Watanabe, T.; Noguchi, K.; Dore, T. M.; Kurahashi, T.; Iwamura, M.; Tsien, R. Y., Bhc-cNMPs as either water-soluble or membrane-permeant photoreleasable cyclic nucleotides for both one- and two-photon excitation. *Chembiochem : a European journal of chemical biology* **2004**, *5* (8), 1119-1128.
116. Furuta, T.; Wang, S. S.; Dantzker, J. L.; Dore, T. M.; Bybee, W. J.; Callaway, E. M.; Denk, W.; Tsien, R. Y., Brominated 7-hydroxycoumarin-4-ylmethyls: photolabile protecting groups with biologically useful cross-sections for two photon photolysis. *Proc Natl Acad Sci U S A* **1999**, *96* (4), 1193-200.
117. Ando, H.; Furuta, T.; Tsien, R. Y.; Okamoto, H., Photo-mediated gene activation using caged RNA/DNA in zebrafish embryos. *Nature Genetics* **2001**, *28* (4), 317-325.
118. Helmchen, F.; Denk, W., Deep tissue two-photon microscopy. *Nat Methods* **2005**, *2* (12), 932-40.
119. Arpino, J. A.; Rizkallah, P. J.; Jones, D. D., Crystal structure of enhanced green fluorescent protein to 1.35 Å resolution reveals alternative conformations for Glu222. *PLoS One* **2012**, *7* (10), e47132.
120. Kavran, J. M.; Gundlapalli, 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.
121. Mohn, W. W.; Tiedje, J. M., Microbial reductive dehalogenation. *Microbiological reviews* **1992**, *56* (3), 482-507.

122. Ormo, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J., Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **1996**, *273* (5280), 1392-5.
123. Stepanenko, O. V.; Stepanenko, O. V.; Kuznetsova, I. M.; Verkhusha, V. V.; Turoverov, K. K., Beta-barrel scaffold of fluorescent proteins: folding, stability and role in chromophore formation. *International review of cell and molecular biology* **2013**, *302*, 221-278.
124. Sokalingam, S.; Raghunathan, G.; Soundrarajan, N.; Lee, S. G., A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS One* **2012**, *7* (7), e40410.
125. Inouye, S.; Tsuji, F. I., Evidence for redox forms of the *Aequorea* green fluorescent protein. *FEBS Lett* **1994**, *351* (2), 211-4.
126. Pletnev, S.; Subach, F. V.; Dauter, Z.; Wlodawer, A.; Verkhusha, V. V., Understanding Blue-to-Red Conversion in Monomeric Fluorescent Timers and Hydrolytic Degradation of Their Chromophores. *J Am Chem Soc* **2010**, *132* (7), 2243-2253.
127. Iizuka, R.; Yamagishi-Shirasaki, M.; Funatsu, T., Kinetic study of de novo chromophore maturation of fluorescent proteins. *Analytical biochemistry* **2011**, *414* (2), 173-8.
128. Kutrowska, B. W.; Narczyk, M.; Buszko, A.; Bzowska, A.; Clark, P. L., Folding and unfolding of a non-fluorescent mutant of green fluorescent protein. *Journal of physics. Condensed matter : an Institute of Physics journal* **2007**, (28), 285223.
129. Zhao, Y.; Zheng, Q.; Dakin, K.; Xu, K.; Martinez, M. L.; Li, W. H., New caged coumarin fluorophores with extraordinary uncaging cross sections suitable for biological imaging applications. *J Am Chem Soc* **2004**, *126* (14), 4653-63.
130. Hemphill, J.; Chou, C.; Chin, J. W.; Deiters, A., Genetically encoded light-activated transcription for spatiotemporal control of gene expression and gene silencing in mammalian cells. *J Am Chem Soc* **2013**, *135* (36), 13433-9.
131. Arbely, E.; Torres-Kolbus, J.; Deiters, A.; Chin, J. W., Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *J Am Chem Soc* **2012**, *134* (29), 11912-5.
132. Fischer, C. J.; Maluf, N. K.; Lohman, T. M., Mechanism of ATP-dependent translocation of *E.coli* UvrD monomers along single-stranded DNA. *Journal of molecular biology* **2004**, *344* (5), 1287-309.
133. Lahue, R. S.; Au, K. G.; Modrich, P., DNA mismatch correction in a defined system. *Science* **1989**, *245* (4914), 160-4.
134. Husain, I.; Van Houten, B.; Thomas, D. C.; Abdel-Monem, M.; Sancar, A., Effect of DNA polymerase I and DNA helicase II on the turnover rate of UvrABC excision nuclease. *Proc Natl Acad Sci U S A* **1985**, *82* (20), 6774-8.
135. Ali, J. A.; Lohman, T. M., Kinetic measurement of the step size of DNA unwinding by *Escherichia coli* UvrD helicase. *Science* **1997**, *275* (5298), 377-80.
136. Kad, N. M.; Van Houten, B., Dynamics of lesion processing by bacterial nucleotide excision repair proteins. *Progress in molecular biology and translational science* **2012**, *110*, 1-24.

137. Brosh, R. M., Jr., DNA helicases involved in DNA repair and their roles in cancer. *Nat Rev Cancer* **2013**, *13* (8), 542-58.
138. Compe, E.; Egly, J. M., Nucleotide Excision Repair and Transcriptional Regulation: TFIIH and Beyond. *Annu Rev Biochem* **2016**, *85*, 265-90.
139. Gardner, L.; Deiters, A., Light-controlled synthetic gene circuits. *Curr Opin Chem Biol* **2012**, *16* (3-4), 292-9.
140. Shell, T. A.; Lawrence, D. S., Vitamin B12: a tunable, long wavelength, light-responsive platform for launching therapeutic agents. *Accounts of chemical research* **2015**, *48* (11), 2866-74.
141. Hemphill, J.; Chou, C. J.; Chin, J. W.; Deiters, A., Genetically Encoded Light-Activated Transcription for Spatiotemporal Control of Gene Expression and Gene Silencing in Mammalian Cells. *J Am Chem Soc* **2013**, *135* (36), 13433-13439.
142. 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.
143. 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. *Chem Commun (Camb)* **2016**.
144. 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.
145. Lee, J. Y.; Yang, W., UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell* **2006**, *127* (7), 1349-60.
146. Epshtain, V.; Kamarthapu, V.; McGary, K.; Svetlov, V.; Ueberheide, B.; Proshkin, S.; Mironov, A.; Nudler, E., UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* **2014**, *505* (7483), 372-7.
147. Takashi, R.; Putnam, S., A fluorimetric method for continuously assaying ATPase: application to small specimens of glycerol-extracted muscle fibers. *Anal Biochem* **1979**, *92* (2), 375-82.
148. Karginov, F. V.; Hannon, G. J., The CRISPR system: small RNA-guided defense in bacteria and archaea. *Molecular cell* **2010**, *37* (1), 7-19.
149. Marraffini, L. A.; Sontheimer, E. J., CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature reviews. Genetics* **2010**, *11* (3), 181-90.
150. Haurwitz, R. E.; Jinek, M.; Wiedenheft, B.; Zhou, K.; Doudna, J. A., Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **2010**, *329* (5997), 1355-8.
151. Haurwitz, R. E.; Sternberg, S. H.; Doudna, J. A., Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *The EMBO journal* **2012**, *31* (12), 2824-32.
152. Lee, H. Y.; Haurwitz, R. E.; Apffel, A.; Zhou, K.; Smart, B.; Wenger, C. D.; Laderman, S.; Bruhn, L.; Doudna, J. A., RNA-protein analysis using a conditional CRISPR nuclease. *Proc Natl Acad Sci U S A* **2013**, *110* (14), 5416-21.
153. Castello, A.; Fischer, B.; Eichelbaum, K.; Horos, R.; Beckmann, B. M.; Strein, C.; Davey, N. E.; Humphreys, D. T.; Preiss, T.; Steinmetz, L. M.; Krijgsveld, J.; Hentze, M. W.,

Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **2012**, *149* (6), 1393-406.

154. Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L., Sortagging: a versatile method for protein labeling. *Nat Chem Biol* **2007**, *3* (11), 707-8.
155. Uttamapinant, C.; White, K. A.; Baruah, H.; Thompson, S.; Fernandez-Suarez, M.; Puthenveetil, S.; Ting, A. Y., A fluorophore ligase for site-specific protein labeling inside living cells. *Proc Natl Acad Sci U S A* **2010**, *107* (24), 10914-9.
156. Fernandez-Suarez, M.; Baruah, H.; Martinez-Hernandez, L.; Xie, K. T.; Baskin, J. M.; Bertozzi, C. R.; Ting, A. Y., Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat Biotechnol* **2007**, *25* (12), 1483-7.
157. Duckworth, B. P.; Zhang, Z.; Hosokawa, A.; Distefano, M. D., Selective labeling of proteins by using protein farnesyltransferase. *Chembiochem : a European journal of chemical biology* **2007**, *8* (1), 98-105.
158. Keppler, A.; Gendreizig, S.; Gronemeyer, T.; Pick, H.; Vogel, H.; Johnsson, K., A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol* **2003**, *21* (1), 86-9.
159. Lukinavicius, G.; Reymond, L.; Johnsson, K., Fluorescent labeling of SNAP-tagged proteins in cells. *Methods in molecular biology* **2015**, *1266*, 107-18.
160. Gautier, A.; Juillerat, A.; Heinis, C.; Correa, I. R., Jr.; Kindermann, M.; Beaufils, F.; Johnsson, K., An engineered protein tag for multiprotein labeling in living cells. *Chem Biol* **2008**, *15* (2), 128-36.
161. Griffin, B. A.; Adams, S. R.; Tsien, R. Y., Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **1998**, *281* (5374), 269-272.
162. Hoffmann, C.; Gaietta, G.; Zurn, A.; Adams, S. R.; Terrillon, S.; Ellisman, M. H.; Tsien, R. Y.; Lohse, M. J., Fluorescent labeling of tetracysteine-tagged proteins in intact cells. *Nat Protoc* **2010**, *5* (10), 1666-77.
163. Chatterjee, A.; Guo, J.; Lee, H. S.; Schultz, P. G., A genetically encoded fluorescent probe in Mammalian cells. *J Am Chem Soc* **2013**, *135* (34), 12540-3.
164. 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.
165. Wang, J.; Xie, J.; Schultz, P. G., A genetically encoded fluorescent amino acid. *J Am Chem Soc* **2006**, *128* (27), 8738-9.
166. Curten, B.; Kullmann, P. H.; Bier, M. E.; Kandler, K.; Schmidt, B. F., Synthesis, photophysical, photochemical and biological properties of caged GABA, 4-[[2H-1-benzopyran-2-one-7-amino-4-methoxy) carbonyl] amino] butanoic acid. *Photochemistry and photobiology* **2005**, *81* (3), 641-8.
167. Simkovitch, R.; Pinto da Silva, L.; Esteves da Silva, J. C.; Huppert, D., Comparison of the Photoprotolytic Processes of Three 7-Hydroxycoumarins. *J Phys Chem B* **2016**, *120* (39), 10297-10310.

168. Kalderon, D.; Roberts, B. L.; Richardson, W. D.; Smith, A. E., A short amino acid sequence able to specify nuclear location. *Cell* **1984**, *39* (3 Pt 2), 499-509.
169. Zou, Y.; Mi, J.; Cui, J.; Lu, D.; Zhang, X.; Guo, C.; Gao, G.; Liu, Q.; Chen, B.; Shao, C.; Gong, Y., Characterization of nuclear localization signal in the N terminus of CUL4B and its essential role in cyclin E degradation and cell cycle progression. *J Biol Chem* **2009**, *284* (48), 33320-32.
170. Zhang, J.; Campbell, R. E.; Ting, A. Y.; Tsien, R. Y., Creating new fluorescent probes for cell biology. *Nature reviews. Molecular cell biology* **2002**, *3* (12), 906-18.
171. Loving, G.; Imperiali, B., A versatile amino acid analogue of the solvatochromic fluorophore 4-N,N-dimethylamino-1,8-naphthalimide: a powerful tool for the study of dynamic protein interactions. *J Am Chem Soc* **2008**, *130* (41), 13630-8.
172. Albani, J.; Alpert, B., Fluctuation domains in myoglobin. Fluorescence quenching studies. *European journal of biochemistry / FEBS* **1987**, *162* (1), 175-8.
173. Chou, C. J.; Uprety, R.; Davis, L.; Chin, J. W.; Deiters, A., Genetically encoding an aliphatic diazirine for protein photocrosslinking. *Chem Sci* **2011**, *2* (3), 480-483.
174. Kim, C. H.; Kang, M.; Kim, H. J.; Chatterjee, A.; Schultz, P. G., Site-Specific Incorporation of ε-N - Crotonyllysine into Histones. *Angewandte Chemie* **2012**, *51* (29), 7246-7249.
175. 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. *Nat Chem* **2012**, *4* (4), 298-304.
176. 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.
177. Loening, A. M.; Fenn, T. D.; Wu, A. M.; Gambhir, S. S., Consensus guided mutagenesis of Renilla luciferase yields enhanced stability and light output. *Protein engineering, design & selection : PEDS* **2006**, *19* (9), 391-400.
178. Wang, Y. S.; Fang, X.; Wallace, A. L.; Wu, B.; Liu, W. R., A rationally designed pyrrolylsyl-tRNA synthetase mutant with a broad substrate spectrum. *J Am Chem Soc* **2012**, *134* (6), 2950-3.
179. Greiss, S.; Chin, J. W., Expanding the genetic code of an animal. *J Am Chem Soc* **2011**, *133* (36), 14196-9.
180. Schwarzer, D.; Ludwig, C.; Thiel, I. V.; Mootz, H. D., Probing intein-catalyzed thioester formation by unnatural amino acid substitutions in the active site. *Biochemistry-US* **2012**, *51* (1), 233-42.
181. Zhang, D.; Vaidehi, N.; Goddard, W. A., 3rd; Danzer, J. F.; Debe, D., Structure-based design of mutant *Methanococcus jannaschii* tyrosyl-tRNA synthetase for incorporation of O-methyl-L-tyrosine. *Proc Natl Acad Sci U S A* **2002**, *99* (10), 6579-84.
182. Liu, C. C.; Schultz, P. G., Adding new chemistries to the genetic code. *Annu Rev Biochem* **2010**, *79*, 413-44.

183. 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 Chem Biol* **2012**, *7* (7), 1292-302.
184. Bianco, A.; Townsley, F. M.; Greiss, S.; Lang, K.; Chin, J. W., Expanding the genetic code of *Drosophila melanogaster*. *Nat Chem Biol* **2012**, *8* (9), 748-50.
185. 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. *J Am Chem Soc* **2012**, *134* (6), 2950-3.
186. Arbely, E.; Torres-Kolbus, J.; Deiters, A.; Chin, J. W., Photocontrol of Tyrosine Phosphorylation in Mammalian Cells via Genetic Encoding of Photocaged Tyrosine. *J Am Chem Soc* **2012**, *134* (29), 11912-11915.
187. Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S., Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode N(epsilon)-(o-Azidobenzylloxycarbonyl) lysine for Site-Specific Protein Modification. *Chem Biol* **2008**, *15* (11), 1187-1197.
188. Cesaratto, F.; Burrone, O. R.; Petris, G., Tobacco Etch Virus protease: A shortcut across biotechnologies. *Journal of biotechnology* **2016**, *231*, 239-49.
189. Waugh, D. S., An overview of enzymatic reagents for the removal of affinity tags. *Protein expression and purification* **2011**, *80* (2), 283-93.
190. Phan, J.; Zdanov, A.; Evdokimov, A. G.; Tropea, J. E.; Peters, H. K., 3rd; Kapust, R. B.; Li, M.; Wlodawer, A.; Waugh, D. S., Structural basis for the substrate specificity of tobacco etch virus protease. *J Biol Chem* **2002**, *277* (52), 50564-72.
191. Tozser, J.; Tropea, J. E.; Cherry, S.; Bagossi, P.; Copeland, T. D.; Wlodawer, A.; Waugh, D. S., Comparison of the substrate specificity of two potyvirus proteases. *Fefs J* **2005**, *272* (2), 514-523.
192. Brieke, C.; Rohrbach, F.; Gottschalk, A.; Mayer, G.; Heckel, A., Light-Controlled Tools. *Angew Chem Int Edit* **2012**, *51* (34), 8446-8476.
193. Beharry, A. A.; Woolley, G. A., Azobenzene photoswitches for biomolecules. *Chem Soc Rev* **2011**, *40* (8), 4422-4437.
194. Russew, M. M.; Hecht, S., Photoswitches: From Molecules to Materials. *Adv Mater* **2010**, *22* (31), 3348-3360.
195. Bonardi, F.; London, G.; Nouwen, N.; Feringa, B. L.; Driessens, A. J., Light-induced control of protein translocation by the SecYEG complex. *Angewandte Chemie* **2010**, *49* (40), 7234-8.
196. Liang, X. G.; Mochizuki, T.; Asanuma, H., A Supra-photoswitch Involving Sandwiched DNA Base Pairs and Azobenzenes for Light-Driven Nanostructures and Nanodevices. *Small* **2009**, *5* (15), 1761-1768.
197. Geoffroy, M.; Faure, D.; Oda, R.; Bassani, D. M.; Baigl, D., Photocontrol of Genomic DNA Conformation by Using a Photosensitive Gemini Surfactant: Binding Affinity versus Reversibility. *Chembiochem : a European journal of chemical biology* **2008**, *9* (15), 2382-2385.

198. Zhang, F. Z.; Zarrine-Afsar, A.; Al-Abdul-Wahid, M. S.; Prosser, R. S.; Davidson, A. R.; Woolley, G. A., Structure-Based Approach to the Photocontrol of Protein Folding. *J Am Chem Soc* **2009**, *131* (6), 2283-2289.
199. Schierling, B.; Noel, A. J.; Wende, W.; Hien, L. T.; Volkov, E.; Kubareva, E.; Oretskaya, T.; Kokkinidis, M.; Rompp, A.; Spengler, B.; Pingoud, A., Controlling the enzymatic activity of a restriction enzyme by light. *P Natl Acad Sci USA* **2010**, *107* (4), 1361-1366.
200. Banghart, M. R.; Mourot, A.; Fortin, D. L.; Yao, J. Z.; Kramer, R. H.; Trauner, D., Photochromic Blockers of Voltage-Gated Potassium Channels. *Angew Chem Int Edit* **2009**, *48* (48), 9097-9101.
201. Beharry, A. A.; Wong, L.; Tropepe, V.; Woolley, G. A., Fluorescence Imaging of Azobenzene Photoswitching In Vivo. *Angew Chem Int Edit* **2011**, *50* (6), 1325-1327.
202. Gorostiza, P.; Isacoff, E. Y., Nanoengineering Ion Channels for Optical Control. *Physiology* **2008**, *23* (5), 238-247.
203. Tsai, Y.-H.; Essig, S.; James, J. R.; Lang, K.; Chin, J. W., Selective, rapid and optically switchable regulation of protein function in live mammalian cells. *Nat Chem* **2015**.
204. Bleger, D.; Schwarz, J.; Brouwer, A. M.; Hecht, S., o-Fluoroazobenzenes as readily synthesized photoswitches offering nearly quantitative two-way isomerization with visible light. *J Am Chem Soc* **2012**, *134* (51), 20597-600.
205. Samanta, S.; Beharry, A. A.; Sadovski, O.; McCormick, T. M.; Babalhavaeji, A.; Tropepe, V.; Woolley, G. A., Photoswitching azo compounds in vivo with red light. *J Am Chem Soc* **2013**, *135* (26), 9777-84.
206. Scott, D. J.; Fuchter, M. J.; Ashley, A. E., Nonmetal catalyzed hydrogenation of carbonyl compounds. *J Am Chem Soc* **2014**, *136* (45), 15813-6.
207. Bose, M.; Groff, D.; Xie, J.; Brustad, E.; Schultz, P. G., The incorporation of a photoisomerizable amino acid into proteins in *E. coli*. *J Am Chem Soc* **2006**, *128* (2), 388-9.
208. Hoppmann, C.; Maslennikov, I.; Choe, S.; Wang, L., In Situ Formation of an Azo Bridge on Proteins Controllable by Visible Light. *J Am Chem Soc* **2015**, *137* (35), 11218-21.
209. 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.
210. John, A. A.; Ramil, C. P.; Tian, Y.; Cheng, G.; Lin, Q., Synthesis and Site-Specific Incorporation of Red-Shifted Azobenzene Amino Acids into Proteins. *Org Lett* **2015**, *17* (24), 6258-61.
211. 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.
212. 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-US* **2011**, *50* (11), 1894-1900.

213. Fliegl, H.; Kohn, A.; Hattig, C.; Ahlrichs, R., Ab initio calculation of the vibrational and electronic spectra of trans- and cis-azobenzene. *J Am Chem Soc* **2003**, *125* (32), 9821-7.
214. Zhao, J. Y.; Lin, S. X.; Huang, Y.; Zhao, J.; Chen, P. R., Mechanism-Based Design of a Photoactivatable Firefly Luciferase. *J Am Chem Soc* **2013**, *135* (20), 7410-7413.
215. Luo, J. T.-K., J.; Liu, J.; Deiters, A., Genetic Encoding of Photocaged Tyrosines with Improved Light-Activation Properties for the Optical Control of Protease Function. *Chembiochem : a European journal of chemical biology* **2017**.
216. Lee, M. S.; Park, W. S.; Kim, Y. H.; Ahn, W. G.; Kwon, S. H.; Her, S., Intracellular ATP Assay of Live Cells Using PTD-Conjugated Luciferase. *Sensors-Basel* **2012**, *12* (11), 15628-15637.
217. Yin, S.; Ding, F.; Dokholyan, N. V., Modeling backbone flexibility improves protein stability estimation. *Structure* **2007**, *15* (12), 1567-76.
218. Tobi, D.; Bahar, I., Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc Natl Acad Sci U S A* **2005**, *102* (52), 18908-13.
219. Ollikainen, N.; de Jong, R. M.; Kortemme, T., Coupling Protein Side-Chain and Backbone Flexibility Improves the Re-design of Protein-Ligand Specificity. *Plos Comput Biol* **2015**, *11* (9).
220. Panjkovich, A.; Daura, X., Exploiting protein flexibility to predict the location of allosteric sites. *Bmc Bioinformatics* **2012**, *13*.
221. Yin, S.; Biedermannova, L.; Vondrasek, J.; Dokholyan, N. V., MedusaScore: an accurate force field-based scoring function for virtual drug screening. *J Chem Inf Model* **2008**, *48* (8), 1656-62.
222. Knie, C.; Utecht, M.; Zhao, F. L.; Kulla, H.; Kovalenko, S.; Brouwer, A. M.; Saalfrank, P.; Hecht, S.; Bleger, D., ortho-Fluoroazobenzenes: Visible Light Switches with Very Long-Lived Z Isomers. *Chem-Eur J* **2014**, *20* (50), 16492-16501.
223. Bishop, A. C.; Ubersax, J. A.; Petsch, D. T.; Matheos, D. P.; Gray, N. S.; Blethrow, J.; Shimizu, E.; Tsien, J. Z.; Schultz, P. G.; Rose, M. D.; Wood, J. L.; Morgan, D. O.; Shokat, K. M., A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature* **2000**, *407* (6802), 395-401.
224. Buckley, D. L.; Crews, C. M., Small-Molecule Control of Intracellular Protein Levels through Modulation of the Ubiquitin Proteasome System. *Angew Chem Int Edit* **2014**, *53* (9), 2312-2330.
225. Winter, G. E.; Buckley, D. L.; Paulk, J.; Roberts, J. M.; Souza, A.; Dhe-Paganon, S.; Bradner, J. E., Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science* **2015**, *348* (6241), 1376-1381.
226. Warner, J. B.; Muthusamy, A. K.; Petersson, E. J., Specific Modulation of Protein Activity by Using a Bioorthogonal Reaction. *Chembiochem : a European journal of chemical biology* **2014**, *15* (17), 2508-2514.
227. Buskirk, A. R.; Liu, D. R., Creating small-molecule-dependent switches to modulate biological functions. *Chem Biol* **2005**, *12* (2), 151-61.

228. Zorn, J. A.; Wells, J. A., Turning enzymes ON with small molecules. *Nat Chem Biol* **2010**, *6* (3), 179-188.
229. Putyrski, M.; Schultz, C., Protein translocation as a tool: The current rapamycin story. *FEBS Lett* **2012**, *586* (15), 2097-105.
230. Fegan, A.; White, B.; Carlson, J. C.; Wagner, C. R., Chemically controlled protein assembly: techniques and applications. *Chemical reviews* **2010**, *110* (6), 3315-36.
231. Ho, S. N.; Biggar, S. R.; Spencer, D. M.; Schreiber, S. L.; Crabtree, G. R., Dimeric ligands define a role for transcriptional activation domains in reinitiation. *Nature* **1996**, *382* (6594), 822-826.
232. Gray, D. C.; Mahrus, S.; Wells, J. A., Activation of Specific Apoptotic Caspases with an Engineered Small-Molecule-Activated Protease. *Cell* **2010**, *142* (4), 637-646.
233. Karginov, A. V.; Ding, F.; Kota, P.; Dokholyan, N. V.; Hahn, K. M., Engineered allosteric activation of kinases in living cells. *Nat Biotechnol* **2010**, *28* (7), 743-U1756.
234. Jullien, N.; Sampieri, F.; Enjalbert, A.; Herman, J. P., Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. *Nucleic Acids Research* **2003**, *31* (21).
235. Bonger, K. M.; Chen, L. C.; Liu, C. W.; Wandless, T. J., Small-molecule displacement of a cryptic degron causes conditional protein degradation. *Nat Chem Biol* **2011**, *7* (8), 531-7.
236. Chu, P. H.; Tsygankov, D.; Berginski, M. E.; Dagliyan, O.; Gomez, S. M.; Elston, T. C.; Karginov, A. V.; Hahn, K. M., Engineered kinase activation reveals unique morphodynamic phenotypes and associated trafficking for Src family isoforms. *Proc Natl Acad Sci U S A* **2014**, *111* (34), 12420-5.
237. Miyamoto, T.; DeRose, R.; Suarez, A.; Ueno, T.; Chen, M.; Sun, T. P.; Wolfgang, M. J.; Mukherjee, C.; Meyers, D. J.; Inoue, T., Rapid and orthogonal logic gating with a gibberellin-induced dimerization system. *Nat Chem Biol* **2012**, *8* (5), 465-470.
238. Weitzman, M.; Hahn, K. M., Optogenetic approaches to cell migration and beyond. *Curr Opin Cell Biol* **2014**, *30*, 112-120.
239. Chen, X.; Li, T.; Wang, X.; Du, Z.; Liu, R.; Yang, Y., Synthetic dual-input mammalian genetic circuits enable tunable and stringent transcription control by chemical and light. *Nucleic Acids Res* **2015**.
240. Jeong, J. W.; McCall, J. G.; Shin, G.; Zhang, Y.; Al-Hasani, R.; Kim, M.; Li, S.; Sim, J. Y.; Jang, K. I.; Shi, Y.; Hong, D. Y.; Liu, Y.; Schmitz, G. P.; Xia, L.; He, Z.; Gamble, P.; Ray, W. Z.; Huang, Y.; Bruchas, M. R.; Rogers, J. A., Wireless Optofluidic Systems for Programmable In Vivo Pharmacology and Optogenetics. *Cell* **2015**, *162* (3), 662-74.
241. Li, J.; Jia, S.; Chen, P. R., Diels-Alder reaction-triggered bioorthogonal protein decaging in living cells. *Nat Chem Biol* **2014**, *10* (12), 1003-+.
242. Li, H.; Hah, J. M.; Lawrence, D. S., Light-mediated liberation of enzymatic activity: "Small Molecule" caged protein equivalents. *J Am Chem Soc* **2008**, *130* (32), 10474-+.
243. Li, J.; Yu, J. T.; Zhao, J. Y.; Wang, J.; Zheng, S. Q.; Lin, S. X.; Chen, L.; Yang, M. Y.; Jia, S.; Zhang, X. Y.; Chen, P. R., Palladium-triggered deprotection chemistry for protein activation in living cells. *Nat Chem* **2014**, *6* (4), 352-361.

244. Vila-Perello, M.; Hori, Y.; Ribo, M.; Muir, T. W., Activation of protein splicing by protease- or light-triggered O to N acyl migration. *Angew Chem Int Edit* **2008**, *47* (40), 7764-7767.
245. Sainlos, M.; Iskenderian-Epps, W. S.; Olivier, N. B.; Choquet, D.; Imperiali, B., Caged Mono- and Divalent Ligands for Light-Assisted Disruption of PDZ Domain-Mediated Interactions. *J Am Chem Soc* **2013**, *135* (12), 4580-4583.
246. Kohn, M.; Breinbauer, R., The Staudinger ligation-a gift to chemical biology. *Angewandte Chemie* **2004**, *43* (24), 3106-16.
247. van Berkel, S. S.; van Eldijk, M. B.; van Hest, J. C., Staudinger ligation as a method for bioconjugation. *Angewandte Chemie* **2011**, *50* (38), 8806-27.
248. Sletten, E. M.; Bertozzi, C. R., From mechanism to mouse: a tale of two bioorthogonal reactions. *Accounts of chemical research* **2011**, *44* (9), 666-76.
249. Alouane, A.; Labruere, R.; Le Saux, T.; Schmidt, F.; Jullien, L., Self-Immulative Spacers: Kinetic Aspects, Structure-Property Relationships, and Applications. *Angew Chem Int Edit* **2015**, *54* (26), 7492-7509.
250. Luo, J.; Upadhyay, R.; Naro, Y.; Chou, C. J.; 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. *J Am Chem Soc* **2014**, *136* (44), 15551-15558.
251. Iizuka, R.; Yamagishi-Shirasaki, M.; Funatsu, T., Kinetic study of de novo chromophore maturation of fluorescent proteins. *Anal Biochem* **2011**, *414* (2), 173-178.
252. Kawai, K.; Ieda, N.; Aizawa, K.; Suzuki, T.; Miyata, N.; Nakagawa, H., A Reductant-Resistant and Metal-Free Fluorescent Probe for Nitroxyl Applicable to Living Cells. *J Am Chem Soc* **2013**, *135* (34), 12690-12696.
253. Saneyoshi, H.; Ochikubo, T.; Mashimo, T.; Hatano, K.; Ito, Y.; Abe, H., Triphenylphosphinecarboxamide: An Effective Reagent for the Reduction of Azides and Its Application to Nucleic Acid Detection. *Org Lett* **2014**, *16* (1), 30-33.
254. Cline, D. J.; Redding, S. E.; Brohawn, S. G.; Psathas, J. N.; Schneider, J. P.; Thorpe, C., New water-soluble phosphines as reductants of peptide and protein disulfide bonds: Reactivity and membrane permeability. *Biochemistry-US* **2004**, *43* (48), 15195-15203.
255. Ho, P. S.; Eichman, B. F., The crystal structures of DNA Holliday junctions. *Curr Opin Struc Biol* **2001**, *11* (3), 302-308.
256. Gibb, B.; Gupta, K.; Ghosh, K.; Sharp, R.; Chen, J.; Van Duyne, G. D., Requirements for catalysis in the Cre recombinase active site. *Nucleic Acids Research* **2010**, *38* (17), 5817-5832.
257. 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-+.
258. Hsu, P. D.; Lander, E. S.; Zhang, F., Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **2014**, *157* (6), 1262-1278.

259. Davis, K. M.; Pattanayak, V.; Thompson, D. B.; Zuris, J. A.; Liu, D. R., Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat Chem Biol* **2015**, *11* (5), 316-U97.
260. Zetsche, B.; Volz, S. E.; Zhang, F., A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol* **2015**, *33* (2), 139-142.
261. Nishimasu, H.; Ran, F. A.; Hsu, P. D.; Konermann, S.; Shehata, S. I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O., Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **2014**, *156* (5), 935-949.
262. Hemphill, J.; Borchardt, E. K.; Brown, K.; Asokan, A.; Deiters, A., Optical Control of CRISPR/Cas9 Gene Editing. *J Am Chem Soc* **2015**, *137* (17), 5642-5645.
263. Martin, S.; Wilkinson, K. A.; Nishimune, A.; Henley, J. M., Emerging extranuclear roles of protein SUMOylation in neuronal function and dysfunction. *Nat Rev Neurosci* **2007**, *8* (12), 948-59.
264. Cha, K.; Sen, P.; Raghunayakula, S.; Zhang, X. D., The Cellular Distribution of RanGAP1 Is Regulated by CRM1-Mediated Nuclear Export in Mammalian Cells. *PLoS One* **2015**, *10* (10), e0141309.
265. Bossis, G.; Chmielarska, K.; Gartner, U.; Pichler, A.; Stieger, E.; Melchior, F., A fluorescence resonance energy transfer-based assay to study SUMO modification in solution. *Methods Enzymol* **2005**, *398*, 20-32.
266. Mahajan, R.; Gerace, L.; Melchior, F., Molecular characterization of the SUMO-1 modification of RanGAP1 and its role in nuclear envelope association. *J Cell Biol* **1998**, *140* (2), 259-70.
267. Hendriks, I. A.; Vertegaal, A. C., A comprehensive compilation of SUMO proteomics. *Nature reviews. Molecular cell biology* **2016**, *17* (9), 581-95.
268. Berridge, M. V.; Herst, P. M.; Tan, A. S., Tetrazolium dyes as tools in cell biology: new insights into their cellular reduction. *Biotechnol Annu Rev* **2005**, *11*, 127-52.
269. Bryson, D. I.; Fan, C.; Guo, L. T.; Miller, C.; Soll, D.; Liu, D. R., Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat Chem Biol* **2017**, *13* (12), 1253-1260.
270. Neumann, H.; Slusarczyk, A. L.; Chin, J. W., De novo generation of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs. *J Am Chem Soc* **2010**, *132* (7), 2142-4.
271. Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W., Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **2010**, *464* (7287), 441-4.
272. Chatterjee, A.; Xiao, H.; Schultz, P. G., Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli. *Proc Natl Acad Sci U S A* **2012**, *109* (37), 14841-6.
273. Willis, J. C. W.; Chin, J. W., Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nat Chem* **2018**.

274. Krogager, T. P.; Ernst, R. J.; Elliott, T. S.; Calo, L.; Beranek, V.; Ciabatti, E.; Spillantini, M. G.; Tripodi, M.; Hastings, M. H.; Chin, J. W., Labeling and identifying cell-specific proteomes in the mouse brain. *Nat Biotechnol* **2018**, *36* (2), 156-159.
275. 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. *Nat Chem Biol* **2016**, *12* (10), 776-778.
276. 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.